Summer 8-15-2019

Expression and Function of snoRNAs in Acute Myeloid Leukemia

Wayne Alsworth Warner
Washington University in St. Louis

WASHINGTON UNIVERSITY

Division of Biology and Biomedical Sciences

Program in Molecular Cell Biology

Dissertation Examination Committee:
Daniel Link, Chair
Kathleen Hall
Timothy Ley
James Skeath
David Spencer
Matthew Walter
Hani Zaher

Expression and Function of snoRNAs in Acute Myeloid Leukemia

by

Wayne Alsworth Warner

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2019

Saint Louis, Missouri

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 4

Supplement

ACKNOWLEDGEMENTS

I would like to thank Dr. Daniel Link, for his immeasurable support over the years. This work would not be possible without his encouragement and guidance. I greatly appreciate him willingness to let me test a new hypothesis, and I am grateful for his guidance.

I would like to thank Dr. Timothy Ley, Dr. David Spencer, Dr. Matthew Walter, Dr. Hani Zaher, Dr. James Skeath, and Dr. Kathleen Hall for serving on my thesis committee. Their support and suggestions have proved invaluable over the years. I would especially like to thank Dr. Tim Ley for serving as my committee chair.

I would like to acknowledge Chancellor's Graduate Fellowship Program (CGFP) and its prior directors, Dr. Rafia Zaher and Dr. Sheri Notaro for providing funding and mentorship. Additionally, I would like to acknowledge Amy Gassell, CGFP for her support over the years.

My graduate experience would not be the same without help from all members of the Link lab, both past and present. I would specifically like to thank Mark Wong for his assistance on the project over the past two years. Much appreciation is extended to Dr. Grazia Abou-Ezzi, and Dr. Jun Xia for their perspectives and many helpful suggestions. I especially would like to thank Amy Schmidt for managing the lab and helping with my project.

This work benefited greatly from assistance provided by Dr. John Edwards, Dr, Hamza Celik, Bill Eades and many others in the Hematology Department. I will like to single out Heidi Struthers for all the technical assistance she provided over the years. She has been a friend and supporter and I will always remember her willingness to assist

in the lab and then whip out a supply of pastries when life was challenging.

To Dr. Joseph Pangelinan, and Andwele Jolly, thank you for the many great conversations on leadership, life and more. I would also like to acknowledge Kim Lipsey for her assistance, support and listening ear. I could not have gone through the rough patches without her welcoming chair and cup of tea.

I would also like to acknowledge the phenomenal support given by Drs. Zaher and Spencer. They provided reagents and support this effort in every way. Thanks to Leo Yan/Zaher lab for lab and photography help. He was always willing to lend a hand and I'm happy to count him now as a friend.

To the Global Health community at WUSM, I appreciate your support and the many hours we spent thinking of how we could make a difference in the delivery of healthcare globally. I will see you on the trail.

Finally, I will like to thank my family and friends for their love and support. Their constant words of encouragement made all the difference. Too many to name. I love you all.

Lastly, I will like to acknowledge my girlfriend, Eileen. I appreciate all that you have done to assist with this body of work. Thank you for your friendship, support and love.

Wayne A. Warner
Washington University
August 2019

DEDICATION

This dissertation is dedicated to my late parents, Inez and Henry Warner

All I am and all I will be

I owe to you

You taught me so many lessons by example and words

To this day your words of wisdom are a beacon to my path

It was my hands that held the pipette

But it was your love and care that animated them

I miss you both dearly

With all my love, I dedicate this to you both and I share this doctorate with you

I still deeply love you

Wayne

ABSTRACT OF THE DISSERTATION

EXPRESSION AND FUNCTION OF snoRNAS IN ACUTE MYELOID LEUKEMIA

by

Wayne Alsworth Warner

Doctor of Philosophy in Biology and Biomedical Sciences

(Molecular Cell Biology)

Washington University in St. Louis,

Professor Daniel C. Link, Chair

2019

Small nucleolar RNAs (snoRNAs) are non-coding RNAs that contribute to ribosome biogenesis and RNA splicing by modifying ribosomal RNA and spliceosome RNAs, respectively. These modifications are critical for a variety of cellular processes, including ribosomal biogenesis and splicing of RNAs. Recent studies have suggested an expanded role for snoRNAs beyond ribosomal biogenesis and splicing, including, regulation of chromatin structure, metabolism, and neoplastic transformation. The contribution of snoRNAs to the regulation of normal and malignant hematopoiesis is largely unknown. The lack of a method to accurately and comprehensively assess snoRNA expression has limited research in this area. In particular, array-based methods only interrogate a subset of snoRNAs and cannot distinguish between mature and precursor snoRNAs. To better characterize the role of snoRNAs in the regulation of hematopoiesis, we therefore

developed a next-generation sequencing technique and bioinformatic approaches to quantify snoRNAs accurately.

We used our snoRNA sequencing pipeline to characterize snoRNA expression in acute myeloid leukemia (AML). We show that snoRNAs are regulated in a lineage- and development-specific fashion in normal hematopoiesis. Surprisingly, RNA splicing did not appear to be a major determinant of expression for most snoRNAs. Expression of most snoRNAs in AML cells was similar to that observed in normal $CD34^+$ cells. In contrast to a prior study, no increased in C/D box snoRNAs was observed in core binding factor AML. On the other hand, certain snoRNAs appear to be dysregulated in specific genetic subtypes of AML. In particular, snoRNAs in the imprinted *DLK1-DIO3* locus are markedly overexpressed in acute promyelocytic leukemia. We also observed reduced expression of *SNORA21* in several genetic subtypes of AML, most notably AML carrying spliceosome gene mutations.

SNORA21 mediates pseudouridylation of key nucleotides in the peptidyl transfer center region of rRNA. To better understand the impact of reduced *SNORA21* expression in AML, we generated *SNORA21* null K562 cells using CRISPR gene editing. Loss of *SNORA21* in K562 cells is associated with impaired ribosome biogenesis and reduced global translation. However, ribosome profiling/RNASeq suggests that only a few genes show significant changes in translational efficiency. Loss of *SNORA21* in K562 cells results in reduced cellular proliferation due to an increase in non-apoptotic cell death. Electron microscopy and mitochondrial assays of *SNORA21*$^{-/-}$ K562 cells suggest that reduced mitochondrial function may contribute to the increase in cell death.

Collectively these data provide compelling evidence for snoRNA involvement in leukemogenesis and highlights the importance of ribosomal biogenesis as a basis to better understand leukemogenesis.

CHAPTER 1

THE BIOLOGY AND FUNCTION OF snoRNAs

Discovery of snoRNAs

The term snoRNA was first used in 1981, more than a decade after the first small nuclear RNAs were detected and the first nucleolar species was identified[1]. The *U3* (*SNORD3*) species was the first to be described in detail after detection during fractionation of small RNA prepared from rat cell nuclei by gel electrophoresis[2]. It is ubiquitous among eukaryotes and is the most abundant of all snoRNAs with a key function in the endonucleolytic cleavage of precursor rRNA[3]. Sequence analysis of SNORD3 revealed several conserved 'box' sequences (A-D) where A and B box regions are upstream of the C and D boxes[4,5]. Box C/D elements were determined to be present in other small nuclear RNAs discovered some years later (1989) and this common feature formed the genesis of the box C/D nomenclature subsequently used to define one of the two large families of snoRNAs[6].

In 1983, gel electrophoresis fractionation strategies were used to describe the population of small nuclear yeast RNAs leading to the discovery of splicing snRNAs and other snoRNAs[7,8]. In 1996, the other large family of snoRNAs (box H/ACA) was identified in yeast from comparative sequencing of nuclear small RNAs[7,9-11]. Over the years, yeast, human, mouse and Xenopus cell models were used for expression, discovery, biosynthesis and function studies. snoRNAs have been identified in a variety of other metazoan organisms (e.g., mouse, rat, Xenopus and plants), and in protists.

snoRNA structure and classification

In humans, snoRNAs typically accumulate in the nucleolus, vary in size from 60-300nt (median 133nt) and are mostly encoded in the introns of translation associated host

genes. During splicing, snoRNAs are stabilized by formation of a small nucleolar RNA-protein complexes (snoRNPs). There are two main types of snoRNAs, box H/ACA (SNORA) and box C/D snoRNA (SNORD) which differ by the protein complex, secondary structure, conserved elements, function and size. A third class, Cajal body-specific RNAs (scaRNAs), which accumulate in the Cajal bodies are involved in the post-transcriptional modification of snRNAs and are usually characterized by composite H/ACA and C/D boxes[12,13].

H/ACA box snoRNAs

H/ACA box snoRNAs (SNORAs) contain 2 hairpins forming the evolutionarily conserved hairpin-hinge-hairpin-tail configuration that characterizes its secondary structure[14-16]. They have two short consensus sequence motifs critical for stability and accumulation, box H (ANANNA) located in the hinge region near the 5' end and the box ACA motif containing the triplet ACA exactly 3 nucleotides upstream of the 3′ end, in the tail region (Figure 1A)[7,17]. The H/ACA boxes influence the synthesis and function of both the snoRNAs and corresponding snoRNPs. Of importance, the H/ACA boxes are involved in providing structural stability and biological function, defining the ends of the mature snoRNA (ACA), localization to the nucleolus and Cajal bodies, and pseudouridylation[7,9-11,18-23].

Each hairpin contains an internal loop, the pseudouridylation pocket characterized by short bipartite recognition sequences (3-10 nts) complementary to the substrate pre-RNA containing the target uridine moiety that will be isomerized to pseudouridine (Ψ) (Figure 1A). The Ψ guide snoRNAs are specifically located in the bulge region of one or

both folded domains/internal loops[9]. For targeted specific site selection to occur there must be base pairing of the two guide sequences with substrate nucleotides that flank the uridine to be isomerized, and a distance of 14-16 nts between the target uridine and the corresponding H or ACA box element[9,23]. Interestingly, a single pseudouridine guide snoRNA can use one or both of the guide domains to target modification. This suggests that the actual modification capacity might be larger than currently known suggesting that SNORAs might have additional functions.

Complexing on the SNORA scaffold are four core proteins: Gar1, Nhp2 and Nop10 and the pseudouridine synthase, dyskerin (DKC1)[30,31]. Gar1 and Nhp2p are known to interact directly with H/ACA snoRNAs. Interestingly, Nhp2 is 38% identical and 61% similar at the amino acid level to the C/D snoRNP protein Snu13p/15.5 kDa, which binds to the C/D motif[32-36]. This could reflect the presence of common binding domains or a common ancestry[32-36]. Once the target RNA is selected through complementarity, pseudouridine is synthesized at the polyribonucleotide level through the action of the pseudouridine synthase, DKC1, which catalyzes the site-specific isomerization of the uridine to pseudouridine[24] (Figure 1.1B). This has been confirmed by the presence of a three-signature sequence element that is conserved among pseudouridine synthases and in yeast experiments where global disruption of Ψ synthesis occurred when *Cbfp5* (the yeast ortholog of *DKC1*) was mutated.

Pseudouridine was discovered in 1951 and is the most abundant post-transcriptionally modified nucleotide. It is formed by the isomerization of uridine under the action of DKC1. It is unique among modified nucleosides in that it has an inert C-C rather than the usual N-C glycosyl bond that links base and sugar (Figure 1.1B). The

irreversibility of this modification suggests distinct roles of pseudouridine in response to stimuli or stresses. Pseudouridylation results in an extra hydrogen-bonding donor site on the non-Watson Crick face (Figure 1.1B). The additional hydrogen bonding of Ψ-NH1 to the 5'-phosphate oxygen atoms increases the backbone rigidity, and the number of interactions critical for inter- and intramolecular binding. Additionally, it leads to **i**ncreased rotational freedom about the glycosyl bond, increased base pairing and enhanced local RNA stacking in both single-stranded and duplex regions resulting in a more stabilized and structurally rigid rRNA species that is critical for ribosomal stability and function[10,14,37-47].

Ribosomal RNA (rRNA) and spliceosomal RNA are the main substrate of SNORAs with around 100 pseudouridylation events documented in rRNA alone (Figure 1.2A-B) [25-27]. Modifications in rRNA can hypothetically affect any stage of ribosome synthesis, ribosome activity, and turnover of the ribosome. Changes can also be expected in RNA folding kinetics, conformational stability of individual and global folding domains, and in the activity of the final RNA. Of these, improved base stacking is considered to be the most important contributor to the stabilization of RNA structure[39,42,48,49]. Ribosome crystal structures have made it possible to catalog the known nucleotide modifications in rRNA in a three-dimensional sequence specific context[47,50,51]. The resulting 2/3-D maps show that ribosome regions known or predicted to be important for function are rich in pseudouridylation modifications. About half of the known SNORAs do not have known rRNA targets and are considered "orphan" suggesting functions other than rRNA-pseudouridylation[28,29].

It was recently shown that the physio-chemical properties of RNA and cellular functions can be altered with the incorporation of pseudouridine[52]. The same study showed that pseudouridylation could be induced by stress[52]. Other studies reported that the pseudouridination of multiple sites in synthetic RNA molecules results in an increased protein expression level[53-55]. Additionally, artificially incorporated pseudouridine in mRNAs mediates nonsense-to-sense codon conversion by facilitating unusual base pairing in the ribosome decoding center, thus demonstrating the potential regulatory role of pseudouridine and a new means of generating protein diversity[53-55].

C/D Box snoRNAs

The second major class of snoRNAs, C/D box snoRNAs (SNORDs) are typically 70–90 nt long. They provide a scaffold on which assembles a protein complex that includes NHP2L1 aka 15.5k, SNU13), NOP56, NOP58, as well as the methyltransferase fibrillarin, which facilitates the transfer of a hydroxyl group from S-adenosyl methionine (SAM) to its selected target nucleotide[56-60]. The secondary structure of SNORDs are characterized by the C (RUGAUGA) and D (CUGA) boxes located near the 5' and 3' ends of the snoRNA(Figure 1.1C)[33,61,62]. Additionally, short regions in up to two antisense boxes of approximately 5 nucleotides located upstream of the C box and downstream of the D box bind in a complementary manner to a rRNA target guiding its site specific 2′-O-methylation[57,58]. Interestingly, C/D snoRNAs are expressed as two distinct forms differing in their ends with respect to boxes C and D and in their terminal stem length. The long forms are more dependent than the short forms on the expression of the core snoRNP protein NOP58, and a subset of short forms are dependent on the splicing factor

RBFOX2. Structural analysis of the potential secondary structure suggests that the k-turn motif required for binding of NOP58 is less stable in short forms which are thus less likely to mature into a canonical snoRNP.

This 2′-O-methylation modification induces structural changes leading to altered steric properties, increased structural rigidity, altered solution energy, preferred 3′ endo configuration and reduced hydrolysis of internucleotide bonds that help ensure ribosomal stability and function[40,43,46,63]. This modification promotes increased stability in RNA conformations and alters the hydration sphere around the oxygen resulting in the blockage of sugar edge interactions, thus affecting the ability of the ribose to participate in hydrogen bonding interactions[64,65].

The majority of SNORDs do not have known rRNA targets based on sequence complementarity and are considered “orphan” suggesting functions other than ncRNA-methylation or cleavage[28,29]. Many of these orphan snoRNAs are in complexes devoid of the methylase fibrillarin[66]. Conversely, some of the SNORDs found in the fibrillarin-free fraction were known to act in rRNA methylation, indicating that SNORDs can have dual functions. The association of SNORDs with proteins other than fibrillarin, NOP56/58 and 15.5 has previously been demonstrated for *SNORD115*, which binds to hnRNPs and can be biochemically separated in fractions containing and lacking fibrillarin[67]. Together, these data indicate that a given SNORD can assemble into protein complexes containing a methylase (‘methylating complexes’) and into complexes lacking a methylase (‘non-methylating complexes’).

In both methylating and non-methylating complexes, the SNORDs are protected from degradation by the associated proteins. These different complexes might also

explain the occurrence of snoRNA fragments, which were first identified by RNAseq and later also validated by RNase protection[68-70]. These processed snoRNAs (psnoRNAs) are most prevalent for *SNORD115* and *SNORD116*, which play a role in Prader-Willi syndrome, and the SNORD113, -114 families as well as *SNORDs- 50, 19, 32B, 123, 111, 72 93, 23* and *85*[69-73]. Therefore, there are compelling reasons to suspect that SNORDs can fulfill new molecular functions different from rRNA modification. Interestingly, these are clustered primarily in two imprinted loci on chromosomes 14 (*DLK-DIO3* locus) and 15 (*SNURF-SNRPN* locus).

snoRNA biogenesis

Eukaryotic cells exhibit significant differences in genetic organization and therefore mechanisms of snoRNA biosynthesis. Only a select few vertebrate snoRNAs- the gene for the telomerase RNA component (*TERC*) and (*SNORD3* or *SNORD8*) that are primarily involved in rRNA processing (no modification function)- are transcribed as independent gene units with their own promoter[74,75]. Of note, in yeast nearly all snoRNAs identified are produced in this manner. In these cases, the precursor transcripts are m^7G-capped and also carry 3′ extensions. Maturation at the 3′-end requires exonucleases as in the case of intronic snoRNAs, while maturation at the 5′-end occurs either via endonucleolytic cleavage, or via the formation of an hypermethylated $m^{2,2,7}G$ cap (TMG) by the enzyme TGS1[76,77]. Hypermethylation is also coupled tightly with assembly since TGS1 is recruited by the core proteins NOP56/NOP58 for C/D snoRNAs, and DKC1 for H/ACA snoRNAs[78,79]. In the case of plants, snoRNAs are generated from polycistronic transcripts allowing for the processing of multiple snoRNAs. On the other hand, in

vertebrates, all of the snoRNAs that guide nucleotide modifications are located within introns of genes transcribed via RNA polymerase II. In fact, many of the snoRNAs are embedded within introns of genes that are involved in ribosome biogenesis[80,81].

Most snoRNAs are encoded in introns and the pre-snoRNA sequences are liberated from the primary transcript by exonucleases[7,82-84]. This entire process has not been fully characterized in humans. For human box C/D snoRNAs, there is a preferential intronic location situated 50 nucleotides upstream of the branch-point, and Aquarius, was proposed to couple recruitment of snoRNP core proteins with splicing of the host pre-mRNA[85]. For box H/ACA, two conserved proteins, NAF1 and SHQ1 are specifically required for the assembly and stability of box H/ACA RNAs, without being part of the mature particles[86-90]. These two factors are nucleocytoplasmic shuttling proteins that localize to the nucleoplasm and are excluded from the nucleoli and Cajal bodies[86-90].

Exonucleolytic processing begins after endonuclease cleavage, splicing and intron debranching to remove excess nucleotides from either end[91-94]. Poly(A)-specific ribonuclease, PARN, is a homodimeric 3′ exonuclease that reported plays a role in the processing of H/ACA box snoRNAs[76]. The noncanonical poly(A) polymerase PAPD5 adds oligo(A) tails to the last few nucleotides remaining after exonucleolytic degradation of the 3′ flanking intron and these oligoadenylated processing intermediates are then trimmed by PARN[76]. It is well established though that mature snoRNA sequences are protected from exonuclease digestion by binding to the snoRNP proteins and, if present, by the 5′ cap[9,11,18,95]. Signal sequences within the snoRNAs centered at boxes C and D or H and ACA direct binding of protein interacting partners that represent the functional

snoRNP complex[96]. Interaction and assembly of these protein-RNA complexes are required for the proper processing and localization of snoRNAs.

Expression of snoRNA

The expression of small nucleolar RNAs (snoRNAs) was determined using high-throughput quantitative PCR, microarray analysis or standard RNA-seq technologies[97-99]. Microarrays are unable to identify novel sequences or resolve the expression of snoRNAs in families with highly homologous members[100]. Standard RNA-seq is generally limited to RNA species >200 nucleotides in length, and thus does not reliably detect most sncRNAs, including snoRNAs. Additionally, these approaches are unable to distinguish the primary transcript from the mature snoRNA and as such the early reported snoRNA quantifications might not have been optimal.

Several studies have used these methods to characterize snoRNA expression in leukemia. Valleron et al., used microarray analysis to report that snoRNAs are differentially regulated in healthy samples compared with malignant T-cell populations and that in addition to a global down-regulation of these molecules, specific signatures may have prognostic significance in peripheral T-cell lymphoma (PTCL)[101]. Interestingly, they reported a specific signature in acute promyelocytic leukemia (APL) with ectopic expression of *SNORD112-114* snoRNAs located at the *DLK1-DIO3* locus[99].

Recently Zhou at al., used microarrays and reported that that fusion oncoproteins such as *RUNX1T1-RUNX1* (also known as *AML1-ETO*) and *MLLT3-KMT2A* (also known as *MLL-AF9*) contribute to AML development by upregulating the expression of a distinct subset of snoRNAs that promote leukemogenesis via enhanced formation of snoRNA

ribonucleoprotein complexes, induction of rRNA methylation, increased protein translation, and possibly other mechanisms[102].

Given the deficiencies of hybridization methods, we developed approaches to ensure that we addressed the shortcomings addressed above. As described in the next chapter, we took advantage of the fact that snoRNAs contain a free 3'-hydroxyl group, which allows for efficient ligation to sequencing adaptors allowing for their capture and efficient sequencing. We also developed a novel analysis pipeline that maps areas of contiguous alignment in the genome, forming *ab initio* "clusters" representing snoRNAs. Taken together, these approaches provide a reproducible and verifiable way to quantify snoRNAs. We used these optimized approaches to determine the expression of snoRNAs in AML and report our findings in subsequent chapters.

Role of SNORNAs in ribosome biogenesis.

Ribosomes are complex macromolecular machines that underlie the translation process, allowing the conversion of data encoded within mRNA into proteins. The human 80S ribosome (named for its apparent sedimentation velocity) is a ribonucleoprotein complex that comprises two ribosomal subunits, a large 60S subunit (LSU) (containing the 5S, 28S, 5.8S rRNA, and 46 ribosomal proteins (r-proteins)) and a small 40S subunit (SSU) (containing the 18S rRNA and 33 r-proteins)[103,104]. Ribosome biogenesis occurs in the nucleolus to form functional ribosomes critical for translation, cellular homeostasis and normal hematopoiesis (Figure 1.3). It is one of the most energetically demanding[105] of activities of the cell and appears to be a process of extraordinary complexity to generate

a machine capable of reading and interpreting the genetic code to produce functional proteins necessary for life[103,104].

Several regions of the ribosome are worth noting. The SSU contains the mRNA entry and exit sites, the path along which the mRNA travels and the decoding center at the heart of which codons are read[106]. The LSU is responsible for peptide bond formation and contains the polypeptide exit tunnel. It also has the ribozyme function required to catalyze peptide-bond formation and ensure mRNA decoding, as well as protein quality control[107]. Importantly, several contact points called inter-subunit bridges, formed of both rRNA and RPs, ensure the assembly of the 80S ribosome and the dynamic coordination between the subunits during translation[108-112]. Through all of this snoRNA modified rRNAs are essential for the function of the individual subunits and their ability to function collectively thus ensuring translation fidelity.

Three of the mature rRNA species, the 18S, 5.8S and 25S, are co-transcribed by RNA polymerase I (Pol I) as a single polycistronic transcript that is matured through a series of nucleolytic processing steps. Maturation of the rRNAs and recruitment of the r-proteins occurs within a series of precursor ribosomal particles, or pre-ribosomes within the nucleolus, nucleoplasm and cytoplasm. The systematic purification of pre-ribosomes in yeast models has allowed the protein and rRNA composition of multiple intermediates to be elucidated and ordered into a ribosome assembly map[113]. The plethora of assembly factors, ATPase, GTPase, helicase, kinase or nuclease activity, then orchestrate the ordered modification, folding and processing of rRNA, and the sequential recruitment of r-proteins[103,114,115]. A few snoRNAs (*SNORD3, SNORD8*) have been shown to play a

role in the processing of rRNA by base-pairing with the pre-rRNA which brings the critical sites into close proximity and conformation framework that supports cleavage[116-126].

As the transcript emerges, many snoRNPs mediate the co-transcriptional covalent modification of over 100 rRNA residues[127]. The function, dynamic folding and stability of these rRNA species are modulated by the post-transcriptional modifications, including pseudouridylation by H/ACA snoRNAs and 2'-O-methylation by C/D snoRNAs[60,128]. rRNA is the main substrate of snoRNAs and approximately 116 2′-O-methylation and around 107 pseudouridylation modifications have been reported in rRNA[25,26,129-131]. Modified rRNA nucleotides display altered steric properties and hydrogen bonding abilities that cumulatively act to stabilize the overall structure and conformation of the rRNAs and therefore the ribosome, allowing it to ensure translational fidelity[118-121,125,126].

Nuclear maturation of the pre-60S subunits, and pre-40S particles requires a large inventory of biogenesis factors, which associate and dissociate throughout the maturation process[132]. Ribosomal subunits are then transported to the cytoplasm for their final maturation. A series of stepwise processes occur in the cytoplasm under the action of biogenesis factors to convert the inactive pre-60S and pre-40S subunits, into functional subunits. An active surveillance system exists that recognizes aberrant or stalled pre-ribosomes and targets them for degradation thus excluding them from export to maintain translation fidelity[133]. Once fully matured, both cytosolic ribosomal subunits are competent to engage in the translation of mRNA[108].

Cells tightly regulate their quantity of ribosomes by controlling rDNA transcription through a number of pluripotency-associated factors that interact with RNA Pol I or with rDNA promoters (in stem cells), while in differentiated cells, rDNA transcription is

controlled by scarcely expressed co-transcription factors, such as lineage-specific factors[134]. Additionally, several oncogenic pathways and oncogenes have been reported as key activators of ribosome biogenesis. The mammalian target of the rapamycin complex (mTOR) is a protein kinase which is activated by nutrients, hormones and oncogenic signaling pathways and other stimuli. mTOR directly stimulates RNA Pol I and RNA Pol III by interacting with their promoters. In addition to regulating ribosome amount in cells, mTOR regulates translation of mRNAs that contain 5'-TOP elements in their 5'UTR[135,136]. Since all of the mRNAs coding for RPs contain a 5'-TOP, mTOR positively regulates synthesis of RPs allowing a coordinated synthesis of the different components of the ribosome[135,136]. This allows for coordination of ribosome biogenesis since most snoRNAs are encoded in RP host genes and some are thought to play a role in translation.

Myc is another regulator of ribosome biogenesis through its effect on the transcription of many genes involved in cell cycle, apoptosis, efficiency of RNA Pol I transcription, RNA Pol III and the synthesis of RPs (by stimulating the transcription of RNA Pol II)[137-139]. In contrast, some tumor suppressor genes (*p53, BRCA1* and *RB1*) have been identified as inhibitors of ribosome biogenesis[140,141]. Interestingly, *p53* represses the expression of *FBL* (methyltransferase responsible for rRNA 2'-O-ribose methylation) by binding directly to the *FBL* gene and inhibits the activity of the *FBL* promoter thus controlling the methylation status of rRNAs in ribosomes[142].

Role of rRNA modifications on ribosome-mediated regulation of translation

Translation results in the synthesis of proteins corresponding to specific informational codes that are contained within mRNAs. The regulation of the translational output is dependent on cell status, environment, development, and pathological conditions. This regulation occurs through a dynamic interplay between cis-regulators (mRNA structures and/or sequences) and the translational machinery composed of its critical effector, the ribosome, and trans-regulators, (eukaryotic factors involved in initiation (eIF), elongation (eEF) and termination (eTF))[143]. The role of the ribosome, as a key regulator of translation, with its hundreds of snoRNA induced modifications has recently become clearer leading to the idea of "specialized ribosomes".

rRNAs play a central role in translation, by directly underlying most of the key molecular interactions necessary for its fidelity. The role of these post-transcriptional modifications on mRNA decoding, peptide-bond formation, and inter-subunit bridges have been well established[106,144,145]. It has been reported that IRES-dependent translation initiation was directly affected by rRNA methylation. It is well-known that 2'-O-Me is more densely observed within the key functional rRNA domains such as the decoding center, the peptidyl transferase center (PTC), near sites involved in translational processes, including the A and P sites, the intersubunit bridges the peptide exit tunnel and intersubunit bridges implying that its presence might be playing a critical role[25,47,146-148]. 2'-O-Methylation can contribute to ribosome population heterogeneity thus preferentially affecting several functional domains of ribosomes and impacting ribosomal activity[25,47,146-148].

Pseudouridylation also plays a role on ribosome mediation of translation. As discussed previously, *DKC1* is the principal pseudouridinase, and in *DKC1*-depleted mice or X-DC patients with *DKC1* mutations there is a loss of rRNA pseudouridylation. The resulting alteration in ribosome function results in the decreased translation of mRNAs containing IRES elements, including the tumor suppressor p27 and the anti-apoptotic factors Bcl-xL and X-linked inhibitor of apoptosis protein (XIAP)[125]. In fact, several studies reported that *DKC1* inactivation, and thus rRNA pseudouridylation defects, are crucial for translational control and cell fate[149-152]. Analysis at the atomic level of ribosome structures indicate that pseudouridylation contributes to rRNA folding in the vicinity of sites critical to ribosome interaction with tRNA and mRNA and necessary for translation fidelity[153-157].

Role of snoRNAs in splicing

Spliceosomal snRNPs, *U1, U2, U4, U5* and *U6* are key components of the spliceosome and are absolutely required for pre-mRNA splicing (Figure 1.2A). They are extensively modified, particularly in regions of functional importance in pre-mRNA splicing, by small Cajal body-specific RNA (scaRNA) in Cajal bodies[13]. Some scaRNAs are structurally indistinguishable from snoRNAs, whereas others combine C/D and H/ACA box motifs with other RNA parts[158]. pre-mRNA splicing begins with a step-wise assembly of the spliceosome initiated by recognition of the 5' splice-site (5' SS) by complementary base-pairing interactions with the 5'-end of U1 snRNA. Through a series of coordinated events, the second step of splicing occurs, resulting in the production of mature mRNA and the

release of the excised intron (and possible snoRNA) and the U2, U5 and U6 snRNPs, which are then recycled for further rounds of pre-mRNA splicing.

Modification of U2 has a direct impact on splicing efficiency. Pseudouridylation of U35 within the branch site recognition region of U2 snRNA, has been demonstrated to direct the catalytic phase of pre-mRNA splicing and is required for pre-mRNA splicing. Interestingly, U35 is the nucleotide nearly opposite the branch-point adenosine. Biochemical, molecular data and biophysical data have clearly established links between spliceosomal snRNP biogenesis, spliceosome assembly, and splicing efficiency with the status of U2 snRNA modification indicating that U2 snRNA pseudouridylation is required for proper positioning of the 2'-OH of the branch-point adenosine so that it is accessible and exposed for recognition and nucleophilic activity[54,159-162].

SNORDs have also been implicated in alternative exon selection which controls the inclusion of exons and the retention of introns in mature mRNA thus increasing the diversity and complexity of the transcriptome[163-165]. By way of examples, the neuron-specific *SNORD115* promotes the inclusion of an alternative exon in the *serotonin receptor 2C* pre-mRNA and *SNORD88C* regulates alternative splicing of *FGFR3* pre-mRNA[66,68,72,166,167]. The mechanism underlying this process is not clear.

snoRNAs are critical mediators of metabolic stress

In a study aimed at identifying genes critical for the lipotoxic response, Schaeffer et al., performed a genetic screen in Chinese hamster ovary (CHO) cells using retroviral promoter trap mutagenesis to create single gene disruptions and positive selection for survival under lipotoxic growth conditions[168]. Unexpectedly, they discovered that three

highly conserved box C/D snoRNAs *U32a (SNORD32a), U33 (SNORD33), U34* (*SNORD34*), and *U35a (SNORD35a)* embedded within the *rpL13a* introns play a key role in the regulation of metabolic stress in mammalian cells possibly through their cytosolic function. Loss of these snoRNAs also altered mitochondrial metabolism and lowered reactive oxygen species tone[169]. The *rpL13a* locus is highly conserved across species and is a constituent of the 60S ribosomal subunit that contributes to regulation of peptide production and catalysis of peptide synthesis.

Separately, it was reported that inflammation stimulates secretion of the *Rpl13a* snoRNAs *SNORD32a, SNORD33, SNORD34,* and *SNORD35a* from cultured macrophages, mice, and human subjects[170]. These snoRNAs co-fractionate with extracellular vesicles and are taken up by recipient cells traveling through the circulation to function in distant tissues[170]. Previously snoRNAs were considered as housekeeping genes[169]. This study and others continue to expand the repertoire of snoRNAs. Here Schaeffer and co-workers provide evidence in support of a previously unappreciated link between inflammation and snoRNA secretion in mice and humans thus pointing to a potential role for secreted snoRNAs in cell-cell communication[170]. Building on this, it was reported that *SNORD60* is involved in intracellular cholesterol trafficking, which is independent on its suggested function in rRNA methylation[171]. These findings reinforce the emerging notion that snoRNAs may contribute to metabolic disease phenotypes in mammals.

snoRNAs play a functional role in hematopoiesis

AML is characterized by the rapid growth and accumulation of abnormal myeloid leukemia blasts in the bone marrow and blood[172]. It is well known that this differentiation block is facilitated by somatic gene mutations, recurring chromosomal transformations, metabolic impairment, epigenetic changes and aberrant ncRNA expression leading to increased self-renewal, increased cell survival and impaired differentiation[173]. It is unknown whether the snoRNA transcriptome plays a role in this process. Recently, Bellodi et al., reported that perturbations in H/ACA snoRNA modulated by DKC1 activity plays a functional role in hematopoietic stem cell differentiation[174]. Interestingly, they showed that *DKC1*-deficient human CD34 cells have an *in vitro* differentiation block relative to normal CD34+ even in the presence of a complete panel of cytokines that foster differentiation. Of note, this phenotype is rescued by $DKC1^{WT}$ expression and not by the catalytic mutant[174].

Hematopoietic stem cells have self-renewal capacity and can generate multipotent progenitors that can differentiate into lineage-committed myeloid and lymphoid progenitors giving rise to mature effector blood cells. These stem cells require a regulated protein synthesis rate and output and as such reports that ribosomal defects impair hematopoietic function are not surprising[175,176]. Intriguingly, aberrant ribosomal biogenesis and ribosomopathies are associated with increased cancer risk[177-180]. Despite its critical role in ribosomal biogenesis, the full role of snoRNAs in physiologic and malignant hematopoiesis remains unclear.

snoRNAs are Involved in neoplastic transformation.

Although snoRNAs are considered housekeeping genes, many of them have been recently identified with broader roles, including in genetic disorders, human variation, hematopoiesis, metabolism and neoplastic transformation[151,168,181-184]. snoRNAs have been studied and implicated in cancers of almost all major organs: the lung[182,185-187], colon and rectum[188,189], breast[190], B and T cell lymphoma[101,191] and leukemia[99,192] and the role of snoRNAs in cancer and disease has been extensively reviewed elsewhere[184,193-196].

Differently expressed snoRNAs were first linked to cancer in a human meningioma study where *SNORA h5sn2* was reported to be highly expressed in normal brain but was decreased dramatically in meningioma[197]. Other studies that looked at the expression of snoRNAs have demonstrated that their expression is altered in cancer and might be therapeutic targets. Oncogenic snoRNA42 has been identified at a locus commonly amplified in solid tumors and was subsequently reported to be commonly overexpressed in NSCLC [198-200]. In another NSCLC study, snoRNA silencing led to inhibition of cancer cell growth, proliferation and tumorigenesis by initiating caspase-3 dependent apoptosis while overexpression increased cell growth and colony formation[182]. Changes in the expression of *SNORD27*, *-30*, *-25*, and *-31* mark the progression of smoldering multiple myeloma[201], and *SNORD50* deletions are associated with prostate and breast cancer[202-204]. The differences are not just in the expression of a specific snoRNA. Genome-wide comparison of SNORD expression between cancer and normal cells showed the presence of two classes of SNORDs that differ in their terminal stem length but are made from the same SNORD hosting intron[28]. Ravo et al. described the aberrant expression

of the H/ACA snoRNA, SNORA3 in endometrial cancer and its presence as part of an sncRNA endometrial cancer signature[205].

Recently, the aberrant expression of snoRNAs along with altered *Rb/p16* regulation was reported in AML as compared to healthy blood samples[99,206]. Specific snoRNA prognostic and diagnostic signatures were noted in leukemic subgroups raising the possibility that snoRNA expression was regulated in a cell type specific manner [99,206]. All of this raises the possibility that snoRNAs might play an expanded role in AML.

Multiple Myeloma

Multiple myeloma (MM) is an incurable malignancy of antibody secreting plasma B cells characterized by the t(4;14)(p16.3;q32.3) translocation in 20% of cases and is associated with reduced patient survival[207]. The t(4;14) translocation upregulates the MMSET proteins which hosts a novel orphan box H/ACA class snoRNA within intron 19. In a recent study, t(4:14) positive and negative MM patient tumor cells were interrogated by RNA sequencing. The H/ACA snoRNA *ACA11* was found to be highly expressed and sufficient to down regulate ribosomal protein gene expression, suppress oxidative stress and increase proliferation of MM cells[207]. They reported a signature of up regulated genes involved in mitochondrial respiration and oxidative stress. They found that *ACA11* overexpression leads to the increased oxidative stress in primary splenic B-cells, established myeloma cell lines, as well as embryonic fibroblasts[207]. In particular, cells overexpressing *ACA11* showed enhanced proliferation, significantly larger colony sizes in soft agar assays, and elevated *ERK1/2* phosphorylation, a downstream consequence

of oxidative stress. Thus, the changes in gene express driven by *ACA11* overexpression may play a pivotal role in the development and/or progression of multiple myeloma[207].

In previous studies, a translocated allele (TKO) model system was used to prove the oncogenic effect of *MMSET* in myeloma thus shifting the pathogenic importance from *MMSET* to *ACA11.* A recent report suggested that *MMSET* and not *ACA11* is the key pathologic mediator in t(4;14) myeloma[208]. This study raises some interesting questions but it may be that *ACA11* has an important role in t(4;14) pathogenesis given that it is in the *MMSET* locus, and often, intronic RNA are found to be involved in the same biochemical pathway as their host gene[74,208]. Interestingly in another study, it was reported that *ACA11* promoted cell growth, migration and invasion through activation of the *PI3K/AKT* pathway, subsequently increasing cyclinD1 expression and inducing endothelial to mesenchyme transition[209].

Dyskeratosis Congenita

Dyskeratosis congenita (DC) is an inherited disorder characterized by bone marrow failure, cancer predisposition, additional somatic abnormalities, and shortened telomeres. The most common reported form of DC is the X-linked form, which is due to loss-of-function mutations of *DKC1*, encoding dyskerin[210,211]. Dyskerin is a 58kD nucleolar protein that is the catalytic component of the H/ACA snoRNP responsible for the isomerization of uridine to pseudouridine[166,212-214]. It was initially assumed when *DKC1* was identified as a cause of DC that a loss of fidelity in pseudouridylation or ribosome biogenesis was responsible for the disease phenotype. However, it was discovered that dyskerin binds to telomerase RNA, which is a critical part of the telomerase enzyme and acts as a

template for the synthesis of the TTAGGG repeats that are found at the telomeres of all chromosomes[215]. Of note, genetic alterations in autosomal dominant forms of DC also result in impaired telomere maintenance.

In humans, all H/ACA snoRNAs are encoded within introns of pre-mRNAs and the processed mature RNAs are devoid of a 5′ cap structure[212,216]. On the other hand, human telomerase RNA (TERC) is transcribed by RNA polymerase II (pol II) from its own promoter and the mature RNA (451 nt) bears a TMG-cap at the 5′ end. *TERC* contains a H/ACA snoRNA-like sequence at its 3'end that binds dyskerin and the 3 other H/ACA proteins and assists in the assembly of telomerase in Cajal bodies and its translocation to the telomeres[215,217-220]. While the H/ACA domain of *TERC* does not guide pseudouridylation, it is essential for its stability and accumulation in vivo[221,222]. Like scaRNA, telomerase RNPs are localized in Cajal bodies and contain a specific sequence motif, the CAB box, in the terminal loop of the conserved region 7 (CR7) that is necessary for their import into Cajal bodies[13,217-219,223]. The presence of DNA damage foci at the telomeres in mouse cells carrying *DKC1* mutations that are known to be pathogenic in humans has been reported[224,225]. This suggests that the presence of mutant dyskerin in the active telomerase complex may cause a transient capping defect at telomeres, leading to telomere loss[226].

Although there are no indications that ribosome biogenesis is affected in DC patients, studies of mice with *DKC1* mutations suggest that defects in ribosome biogenesis or pseudouridylation may contribute to the DC phenotype seen in humans[125,226-228]. The mechanism connecting the mutations in dyskerin to telomere shortening and lower levels of *TERC* has not been fully elucidated[228-231].

Role of C/D box snoRNAs in the pathogenesis of core binding factor APML.

Approximately 4%-12% of adult and 12%-30% of pediatric acute myeloid leukemia (AML) patients and up to 40% of those classified by the French-American-British system (FAB) as M2 subtype contain the t(8;21) translocation that gives rise to the *AML1/ETO* fusion oncoprotein (also known as *RUNX1-RUNX1T1*)[232]. It represents the most frequent chromosomal anomaly in leukemia (18–20%)[233,234]. The resulting fusion consists of the N-terminal DNA-binding domain of AML1(initial 177 amino acids of AML1), a required transcription factor for hematopoiesis, and almost all of ETO, a protein known to function as a corepressor for a variety of transcription factors[235]. These AML, together with AML cases carrying mutations in *CBFB*, are referred to as core binding factor AML and are thought to share a common pathogenetic mechanism.

It is well established that that *AML1/ETO* impacts multiple processes involved in normal myelomonocytic development where the fusion protein interferes with multiple signal transduction pathways, thus promoting early myeloid cell self-renewal and interfering with proper hematopoietic differentiation[236]. *AML1/ETO* also directly interferes with recruitment of essential cofactors by a number of crucial hematopoietic transcription factors such as C/EBPα and PU.1, thus potentially blocking differentiation of myeloid progenitors. *AML1-ETO* fusion is an active repressor that recruits a multi-protein complex including HDACs to AML1 target genes. Both human and mouse models of AML have clearly demonstrated that *AML1-ETO* is insufficient for leukemogenesis in the absence of secondary events[235].

A recent paper reported that leukaemogenesis by *AML1-ETO* required expression of the groucho-related amino-terminal enhancer of split (AES) which functioned by

inducing snoRNA/RNP formation through interaction with the RNA helicase DDX21[102]. They reported that expression levels of C/D box snoRNAs in AML patients correlated closely with in vivo frequency of leukaemic stem cells, indicating that C/D box snoRNAs contribute to the regulation of leukemia-initiating activity[102]. Of note, the increased in C/D box snoRNA expression was reportedly accompanied by increased ribosomal RNA methylation and protein translation. Interestingly, they also reported increased C/D box snoRNA expression in AML driven by *MYC* or *MLL-AF9*. The authors suggested that increased translation is a direct consequence of oncogene expression. This finding contrasts with other studies that report that disease stem cells exhibit reduced levels of translation[237,238]. However, as detailed in chapter 3, we show that C/D box snoRNA expression is not consistently elevated in AML expressing *AML-ETO* or MLL fusion genes.

Dysregulation of *DLK-DIO3* locus snoRNAs in acute promyelocytic leukemia

The *DLK-DIO3* locus located on human chromosome 14 (14q32) is an imprinted locus that contains the protein-coding genes *DLK1, Rtl1*, and *Dio3* expressed from the paternally inherited chromosome. The maternally inherited chromosome produces a large number of ncRNAs, including *MEG3, MEG8*, the largest miRNA cluster in the human genome (with 54 miRNAs), and a large cluster of C/D box snoRNAs. The genes in this locus are thought to be developmentally regulated and expressed in a range of embryonic and extraembryonic cells types with postnatal expression being found predominantly in the brain. The snoRNA genes within the *DLK-DIO3* locus are tandemly repeated C/D snoRNAs. They are thought to arise from the processing of a single long

transcript originating from the *MEG8* gene[239]. The *DLK-DIO3* locus C/D box snoRNAs are orphan snoRNAs, in that they have no known target mRNAs, although some targets have been ascribed based on nucleotide complementarity[240,241]. Allele-specific expression of the locus is mainly regulated by three methylated DMRs: the *IG-DMR*, DLK1-DMR, and the *MEG3-DMR*.

Using microarrays to analyze three cases of APML, Valleron and coworkers first reported that expression of *DLK-DIO3* snoRNAs was increased. This overexpression was in contrast to the reduced expression in normal karyotype AML. Two other studies have confirmed this finding[97,242]. Interestingly, increased expression of *DLK-DIO3* snoRNAs also is present in APML carrying variant *RARA*-fusion partners, suggesting that aberrant retinoic acid receptor signaling is likely responsible for the altered snoRNA expression. Although controversial, Manodoro et al., reported that altered DNA methylation in the *DLK-DIO3* locus results in a loss of imprinting and snoRNA overexpression[243].

SNORA21

As detailed in chapter 4, we have identified differential expression of *SNORA21* in AML. Here, we review literature, mostly derived from studies of yeast, implicating *SNORA21* as a key snoRNA regulating rRNA biogenesis and function.

In yeast, there are approximately 50 known Ψ sites occurring mainly in ribosome regions known to be functionally important, including the PTC[47,118]. To study the effect of Ψ modification on rRNA function in these critical regions, researchers characterized the impact of the genetic deletion/inactivation of individual or several snoRNAs on cell growth

and ribosome function. In a landmark study, King and coworkers targeted 5 snoRNAs modifying six positions within the PTC[118]. Deletion of *snR10*, the yeast ortholog of *SNORA21*, resulted in a loss of Ψ at position 2919 in the rRNA PTC and impaired cell growth[22,23]. Using luciferase protein expression and [^{35}S]methionine incorporation, they showed that the rate of protein synthesis was reduced by 20% in yeast lacking *snR10*. They also reported a reduced yield of polysomes and elevated 80S peak, reflecting a defect in the translation cycle, with fewer translating ribosomes at steady state[118].

Their summary finding that loss of *snR10* (*SNORA21*) results in slow growth, impaired production of small subunits, loss of Ψ2919 (Ψ 4470), an altered polysome pattern, and a reduced rate of protein synthesis has implications for humans given the evolutionary conservation and function of the ribosome. Strikingly, the patterns observed for snR10 were not observed with the deletion of snoRNAs individually.

Other studies have corroborated these findings[244] and extended[27,244] them to take a more intimate look at translation in the context of snoRNA loss. Using similar assays, Baxter-Roshek and coworkers reported changes in reading frame maintenance, nonsense suppression and altered aa-tRNA selection[128]. They concluded that certain rRNA modification fine tune the structure of the A-site region of the large subunit so as to assure correct positioning of critical rRNA bases involved in aa-tRNA accommodation into the PTC, of the eEF-1A•aa-tRNA•GTP ternary complex with the GTPase associated center, and of the aa-tRNA in the A-site[128]. These findings are in line with the known effects of pseudouridylation in altering folding patterns and stabilizing key RNA structures. Taken together, these findings represent a direct demonstration in support of the prevailing hypothesis that rRNA modifications mediated by snoRNAs serve to optimize

rRNA structure for production of accurate and efficient ribosomes critical to translation fidelity.

SNORA21 in human disease

The increasing evidence that snoRNA play a central role in oncogenesis prompted Yoshida and others to interrogate publicly available datasets to identify differentially expressed and commonly upregulated snoRNAs in colorectal cancer (CRC). They identified *SNORA21* (17q12) as a potential oncogenic snoRNA that is significantly upregulated in CRC[245,246]. They showed that knock-down of *SNORA21* in colorectal cell lines resulted in decreased cell proliferation and impaired migration[245,246].

Another study recently reported that *SNORA21* expression was significantly upregulated in human gastric cancer and associated with metastatic disease[247]. Using microarray data, Qin and coworkers recently reported that *SNORA21* expression correlated with the proliferation, migration and invasion of gallbladder cancer cells. Knock-down of SNORA21 in cell lines resulted in increased expression of E-cadherin, decreased expression of N-cadherin and vimentin, increased apoptosis, and G1 growth arrest[248]. In the non- small-cell lung cancer (NSCLC) setting Gao assessed *SNORA21* expression in 77 frozen NSCLC tissues and reported that high expression levels of *SNORA21* were associated with poor overall survival[185].

Taken together these studies report increased expression of *SNORA21* in several cancers with clinical consequences. In chapter 2, we present the expression of snoRNAs in normal and aberrant hematopoiesis. In chapter 3, we expand on this work by describing the expression of snoRNAs in fusion and mutated splicing factor AML. Finally, in chapter

4, we characterize the effect of *SNORA21* deletion on ribosome function, translation, and cell growth in leukemic cells.

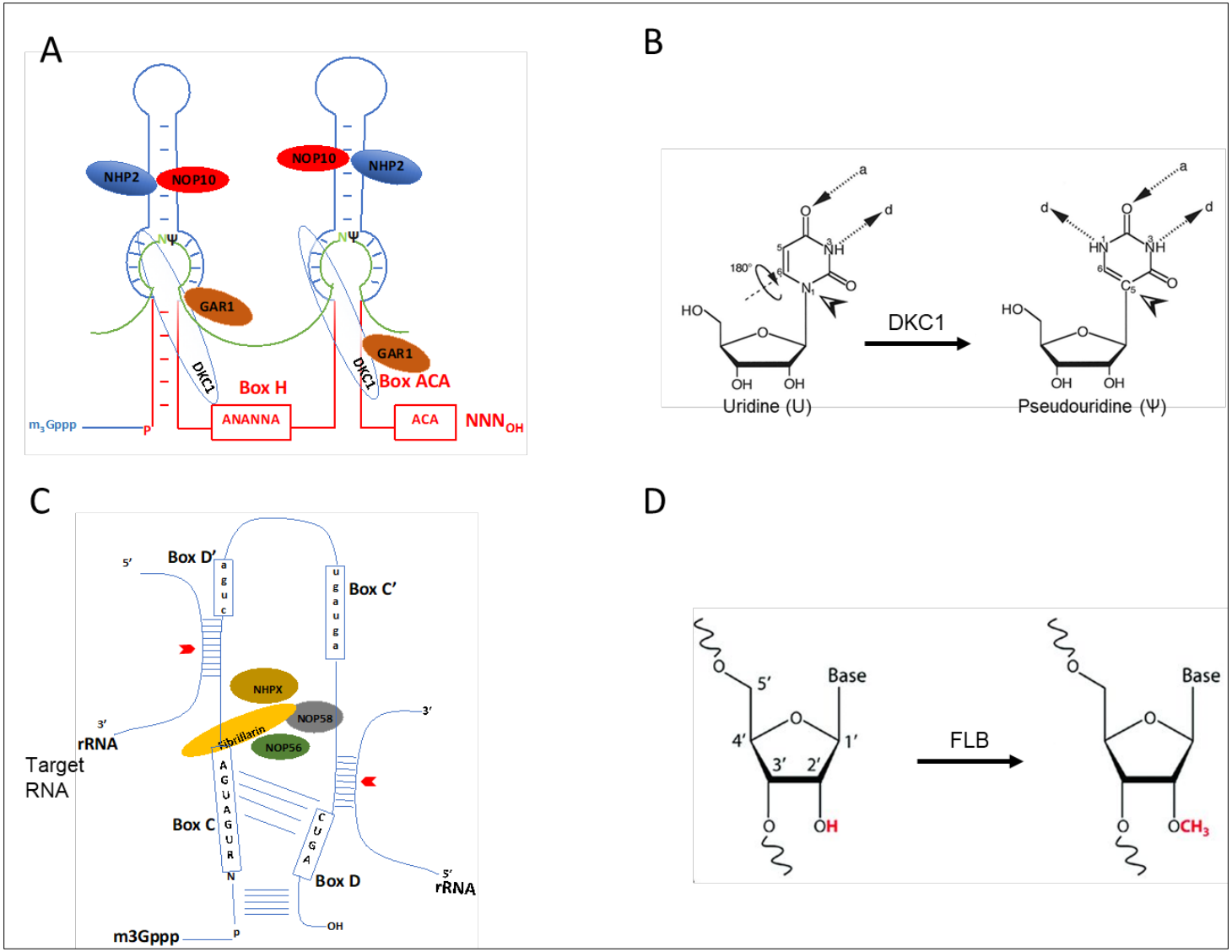


Figure 1.1 Modifications guided by box-C/D and box-H/ACA snoRNAs and the core associated proteins. A). The box H/ACA snoRNAs have an ACA motif at the 3′ end of the snoRNA and a hinge (H) box linking two stem structures. One or

both pseudouridylation pockets contains a bipartite antisense region of 9-13 nucleotides between its two strands. Watson-Crick base pairing selects and localizes the target rRNA to the pseudouridine synthase. Ψ represents the targeted uracil. All SNORAs associate with the four proteins, as shown. B). Isomerization of uridine to pseudouridine catalyzed by DKC1. The uracil base in uridine (left) is linked through its N-1 position (arrowhead) to the C-1^0 position of the ribose. The base in U possesses one hydrogen-bond acceptor and one donor (dashed arrows; a and d, respectively). Isomerization occurs when the uracil base is rotated $180^{\pm}$ through an N3–C6 diagonal axis (circular arrow). In pseudouridine (right), the C-5 position of uracil (arrowhead) is linked to the C-1^0 position of the sugar, resulting in an increase in hydrogen bonding capacity (to one acceptor and two donors) compared with that in U. C). Schematic structure of a C/D box snoRNAs. The box C/D snoRNAs have conserved C- and -D boxes, flanked by short inverted repeats at the 5′ and 3′ ends. Adjacent to box D or D′, there is an rRNA antisense element of 10–20 nucleotides complementary to a specific region of the rRNA. The SNORD forms a protein complex made of 15.5 (also known as SNU13 and NHP2L1), NOP56/58 and the methylase fibrillarin (Fib) that 2′-O-methylates (H_3CO-) rRNA at a defined position 5 nt upstream of the D or D′ box. D) Box C/D-dependent 2′-O-methylation where -OH is converted to -OCH_3.

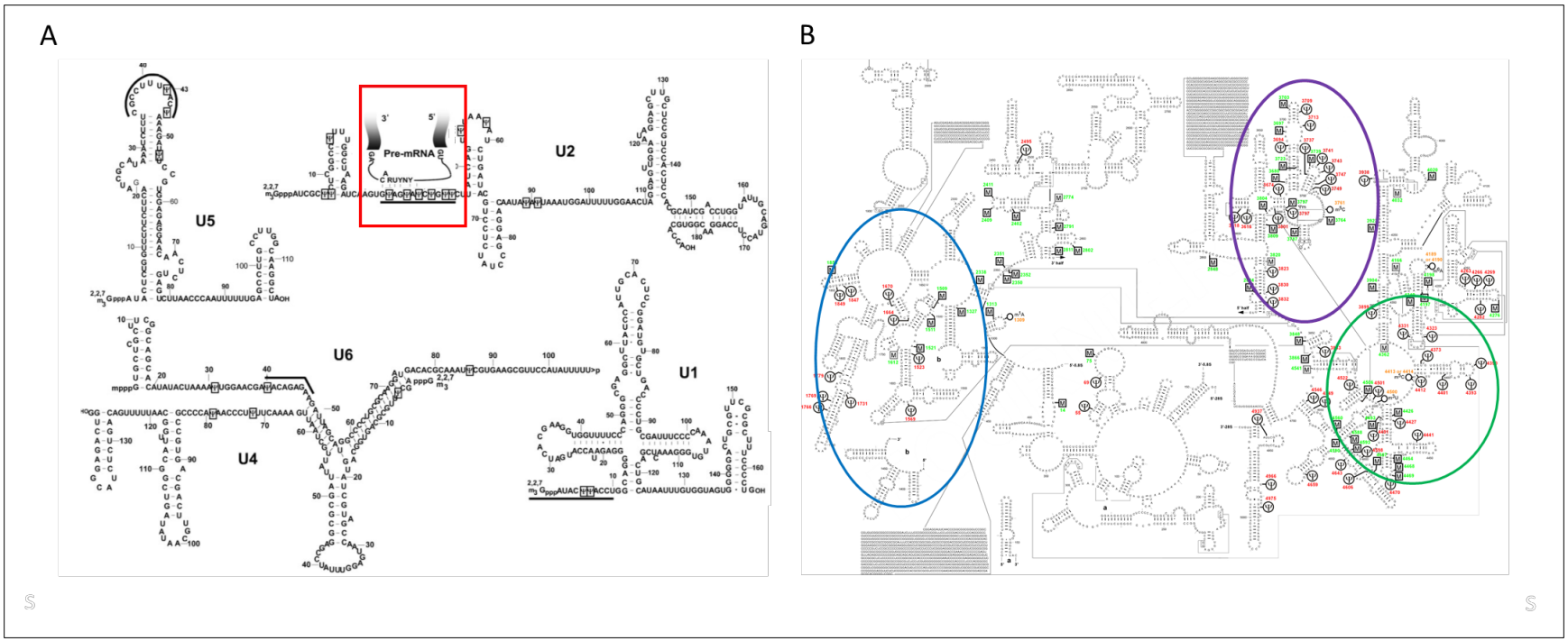


Figure 1.2. snoRNAs modify key functional regions. A) *U2* snRNA modification by snoRNAs is required for pre-mRNA splicing pseudouridines and 2'-O-methylated residues in human spliceosomal snRNAs. primary and secondary structures of human major spliceosomal snRNAs (*U1, U2, U4, U5* and *U6*) are shown. Pseudouridines are surrounded by rectangles; 2'-O-methylations are circled. The thick lines indicate the nucleotides participating in RNA-RNA interactions or involved in catalysis during pre-mRNA splicing. The 5' caps (2,2,7 trimethylated guanosine cap for *U1, U2, U4, U5* and γ-methylated guanosine cap for u6) are also depicted B) Modifications in the rRNA is critical for ribosome biogenesis. Blue oval- A site finger, purple- intersubunit bridge, green- peptidyl transferase center.

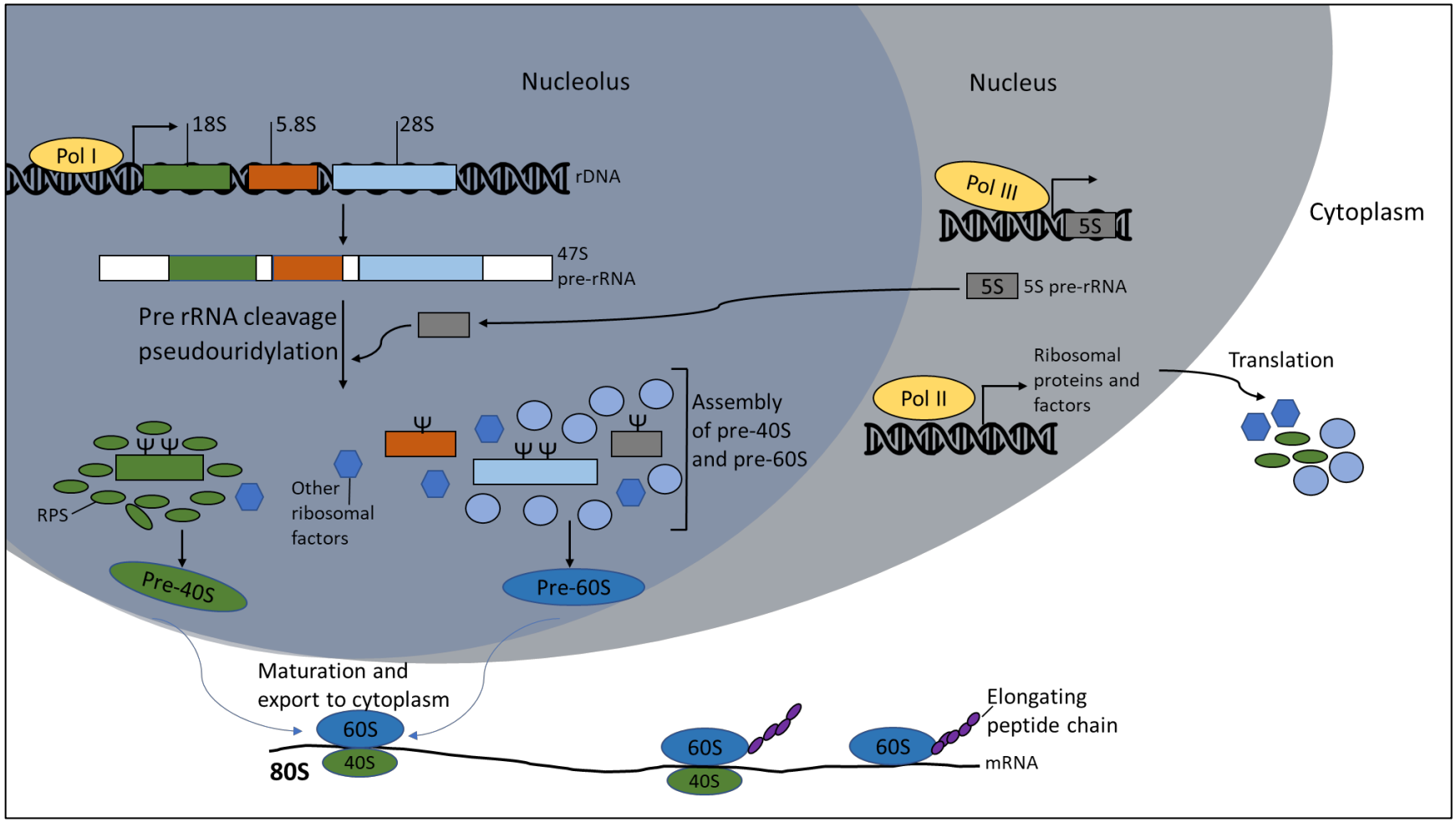


Figure 1.3. Ribosome biogenesis. The mature 80S ribosome is composed of a 40S subunit containing 18S ribosomal RNA (rRNA) and 33 ribosomal proteins (RPs) and a 60S subunit containing 5S, 5.8S and 28S rRNAs and 47 RPs. The majority of steps in ribosome biogenesis occur in the nucleolus, where RNA polymerase I (Pol I) transcribes the 47S precursor rRNAs (47S pre-rRNAs) from ribosomal DNA genes, which contain the sequences of 18S, 5.8S and 28S rRNAs.

RPs are assembled into pre-40S and pre-60S ribosomal subunits in the nucleoplasm and the cytoplasm. The complete process of ribosome biogenesis involves several hundred accessory factors, giving rise to the mature 80S ribosome.

CHAPTER 2

EXPRESSION PROFILING OF SNORNAS IN NORMAL HEMATOPOIESIS AND AML

Abstract

Small nucleolar RNAs (snoRNAs) are noncoding RNAs that contribute to ribosome biogenesis and RNA splicing by modifying ribosomal RNA and spliceosome RNAs, respectively. We optimized a next-generation sequencing approach and a custom analysis pipeline to identify and quantify expression of snoRNAs in acute myeloid leukemia (AML) and normal hematopoietic cell populations. We show that snoRNAs are expressed in a lineage- and development-specific fashion during hematopoiesis. The most striking examples involve snoRNAs located in 2 imprinted loci, which are highly expressed in hematopoietic progenitors and downregulated during myeloid differentiation. Although most snoRNAs are expressed at similar levels in AML cells compared with CD341, a subset of snoRNAs showed consistent differential expression, with the great majority of these being decreased in the AML samples. Analysis of host gene expression, splicing patterns, and whole-genome sequence data for mutational events did not identify transcriptional patterns or genetic alterations that account for these expression differences. These data provide a comprehensive analysis of the snoRNA transcriptome in normal and leukemic cells and should be helpful in the design of studies to define the contribution of snoRNAs to normal and malignant hematopoiesis.

INTRODUCTION

There has been increasing interest in the contribution of the noncoding transcriptome to the regulation of normal and malignant hematopoiesis. Noncoding RNA (ncRNA) species are classified into 2 groups based on their sizes. Long noncoding RNAs (lncRNAs) are >200 nucleotides, and they are expressed in a lineage-specific fashion in hematopoiesis[249]. Recent studies have implicated lncRNAs in hematopoietic lineage commitment and control of self-renewal[249]. Small noncoding RNAs (sncRNAs) are <200 nucleotides and include a heterogeneous group of RNA species. Best characterized are microRNAs (miRNAs), which are 19-26 nucleotide RNAs that repress translation of target RNAs by targeting them to the RNA-induced silencing complex. MicroRNAs are also expressed in a lineage-specific fashion and have been shown to play key roles in the regulation of hematopoiesis[250-252]. Other sncRNAs include small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), small interfering RNAs, and Piwi-interacting RNAs. With some exceptions, the expression of these other sncRNAs in hematopoietic cells and their contribution to the regulation of hematopoiesis are not well characterized.

snoRNAs are a subset of sncRNAs that are involved in the posttranscriptional modification of ribosomal RNAs (rRNAs) and snRNAs. These modifications are critical for a variety of cellular processes, including ribosomal biogenesis and splicing of RNAs. Classification of snoRNA species is based on the presence of highly conserved sequence elements that define 3 snoRNA families: H/ACA box (SNORAs), C/D box (SNORDs), or small Cajal body–specific RNAs (scaRNAs). H/ACA and CD box snoRNAs target specific ncRNA species with base pair complementarity for site-specific pseudouridylation[253] or 2'-O-methylation,[61] respectively. scaRNAs localize to RNA-containing Cajal bodies and

are responsible for the methylation and pseudouridylation of spliceosomal RNAs *U1, U2, U4, U5*, and *U12*. There are also orphan snoRNAs, which lack known complementarity to rRNAs or snRNAs and therefore largely have unknown functions. Recent studies have suggested an expanded role for snoRNAs beyond ribosomal biogenesis and modifications to snRNA. For example, emerging data suggest that snoRNAs may contribute to alternative splicing,[254] regulation of chromatin structure,[255] metabolism,[168] and neoplastic transformation.[204]

The contribution of snoRNAs to the regulation of normal and malignant hematopoiesis is largely unknown. Chu et al reported that over- expression of the H/ACA box snoRNA ACA11 in t(4;14)-associated multiple myeloma contributes to myeloma cell proliferation and resistance to chemotherapy[256]. Several groups have reported marked increased expression of snoRNAs contained in the *DLK-DIO3* locus in acute promyelocytic leukemia, although their contribution to leukemogenesis is unknown[97,99,242]. The lack of a method to accurately and comprehensively assess snoRNA expression has limited research in this area. Array-based methods only interrogate a subset of snoRNAs and cannot distinguish between mature and precursor snoRNAs[206,257]. To avoid sequencing very abundant rRNAs and transfer RNAs (tRNAs), most next-generation sequencing approaches to interrogate the transcriptome have focused on longer (>200 nucleotide) or very short (17-26 nucleotide) RNA species. Thus, there is a gap in current transcriptome sequencing that includes most snoRNAs. To address this gap, we developed a next-generation sequencing approach optimized to interrogate sncRNAs, including snoRNAs. We show that snoRNAs are expressed in a lineage- and development-specific fashion in human hematopoiesis with a subset of

snoRNAs that show consistent differential expression in acute myeloid leukemia (AML). We further show that expression of snoRNAs does not correlate with expression or splicing of host genes, suggesting that other factors are determining cellular levels of mature snoRNAs.

MATERIALS AND METHODS

Fluorescence-activated cell sorting of hematopoietic populations

Bone marrow aspirate samples were obtained from normal healthy donors after obtaining informed consent (institutional review board approval Washington University Human Studies Committee #01-1014). Samples were processed via ammonium–chloride–potassium red cell lysis, washed once in phosphate-buffered saline, and then stained for flow cytometry using the following antibodies: CD34-phycoerythrin (PE) (PE-pool, Beckman Coulter, IM1459U, Brea, CA), CD14-allophycocyanin (BD Biosciences, clone M5E2, San Jose, CA), CD15-fluorescein isothiocyanate (BD Biosciences, clone HI98), CD16-PE (BD Biosciences, clone 3G8), CD33-allophycocyanin (eBioscience, clone WM-53, San Diego, CA), CD3-V450 (eBioscience, clone OKT3), and CD19-PE (BD Biosciences, clone HIB19). Defined hematopoietic cell populations that were sorted included: promyelocytes ($CD14^-$, $CD15^+$, and $CD16^{low/-}$),[258] monocytes ($CD14^+$), neutrophils ($CD14^-$, $CD15^+$, and $CD16^-$),[258] and $CD34^+$ cells. Cells were sorted directly into lysis buffer, and RNA was isolated using the Quick-RNA Microprep Kit (Zymo Research, Irvine, CA).

Small RNA library construction and sequencing

The NEBNext Small RNA Library Prep Set for Illumina (New England BioLabs, Inc., Ipswich, MA) was used to prepare the libraries following the manufacturer's specifications using 100 to 500 ng of total RNA as input[259]. After adaptor ligation, reverse transcription, and polymerase chain reaction (PCR) amplification, the libraries were size selected on a Blue Pippin (Sage Science, Beverly, MA) to enrich for library molecules with inserts

between approximately 17 and 200 nucleotides. The resulting libraries were sequenced on a MiSeq instrument to generate 150 bp, single-end reads. All sequence data will be deposited in dbGaP.

Bioinformatic analysis

Sequencing data were trimmed to remove adapter sequences using cutadapt with the command "cutadapt -f fastq -a AGATCGGAAGAGCACACGTCT" and then mapped to the National Center for Biotechnology Information Build 37 human reference sequence using bwa mem[260] with the custom parameters "bwa mem -M -k 15 -T 17" to obtain short alignments that result from small RNA species. These alignments were then used in the following analyses to characterize the spectrum of RNA species captured by the library approach, identify novel RNA species, and quantify the expression of annotated snoRNAs.

We first defined the distribution of RNA species captured in the library by annotating the sequencing reads from all samples with RNA biotypes from GENCODE version 19,[261] mirBase version 21[262] and a previously described set of snoRNA annotations (snoRNAome[241]). Given that the library preparation method has a 3'-end bias, reads were assigned to a single RNA annotation in a strand-specific manner based on the proximity of the read start position to the 3' end of overlapping annotations. Reads were called "unannotated" if the alignments were uncertain (ie, mapping quality of 0), or they did not map to any annotation in a strand-dependent manner.

Next, we identified potentially novel RNA species using a custom Practical Extraction and Report Language (PERL) script designed to detect and annotate

aggregate read “clusters” using pooled sequence data from all normal hematopoietic cell and AML samples (N = 64). Briefly, mapped reads for all samples with a mapping quality >0 were merged into a single BAM file, and regions with a minimum strand-specific read depth of 50 were extracted. All reads spanning these regions were then merged to create strand-specific read clusters, which were trimmed such that the cluster edges were ≥ 20% of the maximum read depth (to separate closely spaced clusters that may have become merged by spurious “joining” reads), and subsequently filtered to retain those with an AT nucleotide content <80% to exclude low-complexity sequences. The total number of clusters that resulted from this procedure was 6231. Clusters were then annotated with read quality and mapping statistics (eg, mean mapping quality, number of unique read start positions, and mean number of mismatches with the reference sequence), the number of strand-specific read counts, and maximum depth across the cluster. Cluster coordinates were compared with the GENCODE, mirBase, and snoRNAome annotations, and “tagged” with the strand-specific transcript or gene annotation with the best reciprocal overlap. Potentially novel species from this set that demonstrated <50% reciprocal overlap with known annotations and total counts ≥ 500 (N = 340 clusters) were then manually analyzed with the programs snoGPS and snoSCAN, which identify H/ACA box and C/D box snoRNAs with reported rRNA targets,[263,264] and with snoReport for identification of all snoRNAs, including orphans[265] as well as a custom script. This produced a final list of 111 clusters that were identified as potential snoRNA species, which were manually reviewed using the Integrated Genome Viewer (version 2.3.40)[266] and the sno/miRNA track of UCSC Genome Browser[267] to exclude low-quality clusters or those that overlapped known snoRNAs.

Finally, expression levels for a comprehensive set of annotated sncRNA species were generated for each sample using annotations curated from snoRNAome[241] and miRBase[268] along with all GENCODE version 19 annotations with biotype "snoRNA". These annotation databases were combined to produce a set of 4931 nonoverlapping annotations, with snoRNAome and mirBase entries superseding those from GENCODE version 19 with overlapping coordinates. Overlapping annotations from snoRNAome and mirBase were reviewed, and a single species was selected based on the correspondence between the sequencing reads at the locus and annotation; the other annotation was excluded. Expression values for these annotations were then obtained with the featureCounts program[269] using parameters for strand-specific counting and including only reads with a mapping quality of ≥1. These counts were normalized to the total mapped reads 3 x 10^6 for visualization and subsequent statistical analyses.

Quantitative reverse transcription PCR of selected snoRNAs

The extracted RNA was purified on a RNA Clean & Concentrator-5 column (Zymo Research) using the manufacturer's >17- nucleotide-long protocol and resuspended in 10.0 mL of nuclease- free water. The Qubit RNA HS Assay Kit (Life Technologies, Carlsbad, CA) and the TapeStation system (Agilent, Santa Clara, CA) were used for quantification and quality assessment, respectively, according to the manufacturer's instructions. The RNA was reverse transcribed using iScript Reverse Transcriptase (BioRad, Hercules, CA) at 42°C, according to the manufacturer's instructions. The complementary DNA (cDNA) was PCR amplified using forward and reverse primers containing sequences specific to the snoRNAs (Table 2.1). In a 20-mL reaction, 3.0 ng

of cDNA template, 0.5 mM each of forward and reverse primers (IDT, Chicago, IL), 10 mL iTaq Universal SYBR Green 23 Supermix, (BioRad), and nuclease-free water were cycled for 60 rounds at an annealing temperature of 60°C on a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, CA). The 5S rRNA was used to normalize snoRNA expression.

Differential expression analysis

Differential expression and hierarchical clustering analyses were performed with the Partek Genomic Suite (Partek, Inc., St. Louis, MO) using $\log_2$ (read count per million mapped reads 3 x 10^6 [RPM]) expression values for the curated sncRNA annotations as input[270]; only RNA species with mean normalized count values ≥5 were selected to produce reliable differential expression profiles. Data were first assessed for normality, and differential expression analysis was performed with the Partek Genomic Suite using 1-way analysis of variance (ANOVA) with estimation via the method of moments model[271]. The differential expression of snoRNAs and miRNAs in the AML patients vs normal hematopoietic stem/progenitors was based on a fold change >2 and $P < 0.05$).

Splicing analysis

Intron junction counts for annotations in GENCODE version 19 were obtained from aligned BAM files using Tablemaker and Ballgown[272] and normalized to the total number of junction reads observed 3 x 10^6. The linear regression between normalized snoRNA expression (RPM) and the normalized expression for "host gene" junctions spanning each snoRNA were assembled in R[273]. Correlations between all snoRNAs and junction

expressions were similarly performed.

Somatic mutation of snoRNAs

The coordinates of the 344 651 introns in the genome (GRCh37) and that of 402 snoRNAs were intersected with the coordinates of 367 904 prevalidation indels from 49 Cancer Genome Atlas AML patient samples using BEDtools[274] and R.

Statistical analysis

Statistical analysis and graphing were performed with GraphPad Prism version 8.00 for Mac, (GraphPad Software, La Jolla, California) and R. Error bars represent the standard error of the mean (SEM). Significance was determined by 1-way ANOVA followed by Tukey multiple comparisons test. Significance is denoted as: *P < .05; ** < .01; ***P < .001; ****P < .0001; and ns, not significant.

RESULTS

Small RNA-seq pipeline

We modified a previously described method for sequencing miRNAs to analyze more comprehensively the small RNA component of the transcriptome[259] .A key aspect of this approach is the method used for cDNA library generation, which includes the addition of an oligonucleotide adaptor to the 3'-end of RNA molecules before reverse transcription. Importantly, this requires the presence of a free 3' hydroxyl group on the RNA molecule. We then performed an expanded size selection to capture RNA species between 17 and 200 nucleotides, which includes miRNAs, snoRNAs, and other sncRNAs, but excludes most messenger RNA (mRNA) and lncRNA molecules. The sequence data obtained were analyzed using 2 complementary bioinformatic approaches to quantify both annotated and novel sno- and miRNAs (Figure 2.1A).

We used this approach to interrogate the small RNA transcriptome in human hematopoietic cell populations from normal hematopoietic stem/progenitors and from diagnostic AML samples. *CD34* cells, promyelocytes, neutrophils, monocytes, T cells, and B cells were sorted by flow cytometry from the bone marrow of 6 healthy individuals. Data from primary AML samples were generated from bulk leukemic cells from 33 treatment-naive patients with AML (Table 2.2). Most of these cases (97%) had normal cytogenetics, and all were classified as intermediate-risk AML. An average of 3.2×10^6 reads was obtained across both normal and leukemic samples (Table 2.3). Mapping of sequencing reads from all samples to annotation features from GENCODE version 19 and snoRNA and miRNA annotations in the human snoRNAome and miRBase (see "Materials and methods") demonstrated that snoRNAs were by far the most abundant

small RNA species present in our data (Figure 2.1B). C/D box snoRNAs represented 74.95% of all reads; H/ACA box snoRNAs and scaRNAs represented another 3.19% and 0.34% of total mapped reads, respectively. Small nuclear RNAs, which are involved in RNA splicing, were the next most abundant class of sncRNA, representing 10.19% of reads. miRNAs represented a relatively small percentage of all sequenced reads (1.48% of all mapped reads). Reads mapping to unannotated regions of the genome accounted for 0.04% of all sequences.

We next compared the expression of snoRNAs using data from our modified library protocol with expression levels obtained by standard total RNA sequencing (RNA-seq) (Illumina Tru-seq) and of the same tissue sample. Relevant to this analysis, the majority of snoRNAs are embedded in the introns of host genes. We observed that standard transcriptome sequencing cannot reliably distinguish unspliced primary host gene RNA from correctly processed snoRNA. Typical results are shown for *SNORA64*, which is located in the intron of its host gene, RP32 (Figure 2.1C). Whereas sequence reads corresponding to mature *SNORA64* were readily identified using our pipeline, only low-level reads that span the entire intron of *RP32* were detected using total RNA-seq. Accordingly, the correlation of snoRNAs quantified using these 2 RNA-seq pipelines was poor (Figure 2.1D). These data demonstrate the superiority of our small sequencing pipeline to quantify mature, correctly processed snoRNA expression.

To provide orthogonal validation of the snoRNA expression data, we used commercially available reagents to perform quantitative reverse transcription PCR (RT-qPCR) on a set of 9 snoRNAs with a wide range of expression across 11 primary AML samples. Although some variability was observed, a significant correlation between

snoRNA expression determined by small RNA Seq and RT-qPCR was observed (R^2 = 0.5002; P < .0001; Figure 2.2).

To determine whether our sequencing approach identified any novel RNA species, we formed read clusters by merging over-lapping reads and compared them with the RNA annotations as described above. The intersection of read cluster coordinates with our sncRNA annotation set demonstrated that a number of them did not overlap with known annotations and could therefore represent novel RNA species. The genomic regions spanned by these clusters were then analyzed for features of snoRNAs, including the presence of conserved sequence motifs and secondary structure. Eight putative novel snoRNAs were identified, including 5 in the SNORA family and 3 in the SNORD family. One of the putative SNORDs lacked sequence complementarity to rRNAs or snRNAs and was therefore classified as an orphan snoRNA (Table 2.4). There was some degree of overlap (≤50%) with annotated species, but our analysis supports the characterization of these snoRNAs as putatively novel.

Developmental- and lineage-specific expression of snoRNAs in human hematopoiesis

Because snoRNAs were the most abundant sncRNA detected, we focused our analysis on these RNA species. We first performed unsupervised hierarchical clustering of annotated snoRNAs with a normalized expression of >5 RPM (N = 378) to determine whether expression of snoRNAs is developmentally regulated during hematopoiesis. This demonstrated that snoRNAs exhibit lineage- and developmentally restricted expression patterns (Figure 2.3). The most striking examples were orphan snoRNAs

contained in the imprinted *DLK-DIO3* and *SNURF/SNRPN* loci. The *DLK-DIO3* locus contained a large number of maternally expressed ncRNAs, including 41 snoRNAs, 11 lncRNAs, and 53 miRNAs (Figure 2.4A). Expression of snoRNAs in this locus was highest in CD34 cells and rapidly decreased with granulocytic differentiation, becoming nearly undetectable in mature neutrophils (Figure 2.4B). Expression of these snoRNAs was also markedly reduced in B cells and T cells. Expression of snoRNAs in the *SNURF/SNRPN* locus showed a similar, but distinct, pattern of snoRNA expression. This locus contained 82 paternally expressed snoRNAs that were expressed at a high level in CD34 cells and rapidly downregulated during granulocytic differentiation (Figure 2.4C). However, in contrast to the *DLK-DIO3* locus snoRNAs, expression of these snoRNAs remained high in B and T cells (Figure 2.4D).

Expression of a subset of snoRNAs is decreased in AML

We next compared snoRNA expression in 33 de novo AML samples with normal CD34 cells. Analysis of expression across all annotated snoRNA species (N = 364) via unsupervised hierarchical clustering demonstrated that AMLs had distinct snoRNA expression patterns from normal CD34 cells (Figure 2.5A). We required a mean normalized expression of ≥5 counts across any AML and healthy donor samples to be considered for analysis. Differential expression analysis identified 102 snoRNAs that were differentially expressed (adjusted $P \leq 0.05$; absolute $\log_2$-fold change >1) (Table 2.5), all of which had decreased expression in the AML samples (Figure 2.5B). By comparison, 24 differentially expressed miRNAs were identified in a similar analysis using the same samples, which included 17 with increased expression in AML vs 7 that were decreased

(Figure 2.5C). Although differentially expressed snoRNAs in AML spanned all RNA species, a disproportionate number of box C/D snoRNAs were observed (69 of 102, 67.65%; Figure 2.5D), with 37 located in the *DLK-DIO3* or *SNURF- SNRPN* loci. There were 66 (64.71%) orphan snoRNAs with representation from all snoRNA classes. Of note, differential expression of snoRNAs that are known to play key roles in splicing as well as the modification of the peptidyl transferase center (PTC) and the intersubunit bridge (ISB) during ribosomal biogenesis was observed (Table 2.5). For example, expression of *SNORA21* and -36C, which target crucial nucleotides in the PTC and ISB, respectively, were decreased 2.69- and 2.56-fold, respectively, in AML compared with CD34 cells, and expression of *SCARNA15*, which targets a key nucleotide in the U2 spliceosomal RNA, was decreased 2.81-fold.

Somatic mutation of snoRNAs is uncommon in AML

Whole-genome sequencing data were available for 14 of the 33 analyzed cases in this study. No somatic single nucleotide variants or small indels were detected in the snoRNA genes. In addition, for those snoRNAs located in a host gene, no recurrent indels in the introns harboring the snoRNA or mutations in splice donor sites for that intron were identified. We expanded this analysis to an additional 35 AML cases with whole-genome sequencing data available from The Cancer Genome Atlas[275]. Again, no somatic single nucleotide variants or small indels were detected in snoRNA genes, suggesting that genetic alterations in snoRNAs are uncommon in AML with normal cytogenetics and are not the cause of their differential expression in this disease.

There is minimal correlation between host gene and snoRNA expression

Because many snoRNAs are located in the introns of host genes,[74] we next asked whether variation in snoRNA expression may be explained by differences in expression and/or processing of these host genes. We limited our analysis to the AML cases, where matching small RNA and total RNA-seq data were available. For most snoRNAs, there was minimal correlation between host gene and corresponding snoRNA expression, as illustrated by host gene *RPL7A* and its corresponding snoRNAs (Figure 2.6A). Across all 1379 snoRNAs contained in host genes, the average coefficient of determination (R^2) was 0.037 ± 0.102 (Figure 2.6B). Multiple snoRNAs are often located within different introns of a single host gene, as shown for the *C19orf48* gene (Figure 2.6C). If host gene expression is the primary determinant of snoRNA expression, then expression of each snoRNA located in a given multihost gene should be similar. However, we observed marked variability in the expression of snoRNAs contained within a single gene. For example, expression of the 3 snoRNAs hosted by *C19orf48* varied by >32-fold (Figure 2.6D). Indeed, marked variability in the expression of snoRNAs contained within the same host gene was observed in the majority of cases (Figure 2.6E). These data show that host gene expression is not the primary determinant of snoRNA expression in AML.

Alternative splicing of host genes is not the primary determinant of snoRNA expression

Mature snoRNAs are processed from excised introns after splicing of the host gene. Thus, we next asked whether alternative splicing of host genes is a major determinant of snoRNA expression. We assessed RNA splicing by measuring junction reads, as

previously described[276]. For example, the host gene *C19orf48* has 10 predicted splice events that involve introns containing embedded snoRNAs (Figure 2.7A). Expression of junction reads corresponding to each splice event showed minimal correlation with expression of the relevant snoRNA (Figure 2.7B-K). This analysis was extended to look at snoRNA expression and encompassing junction expression across 858 snoRNAs spanning 1616 junctions (Figure 2.7L). For most cases, junction reads correlated minimally with snoRNA expression. Collectively, these data show that alternative splicing of host genes is unlikely to be the primary determinant of snoRNA expression.

Discussion

The expression of snoRNAs has traditionally been determined by high-throughput techniques that rely on hybridization-based methods, such as microarray analysis or by standard RNA-seq technologies[97-99]. Microarrays for snoRNAs are unable to effectively capture novel sequences or resolve the expression of snoRNAs in families with highly homologous members[100]. Standard RNA-seq is generally limited to RNA species >200 nucleotides in length, and thus does not reliably detect most sncRNAs, including snoRNAs. In this study, we optimized both library preparation and bioinformatic analysis to address these challenges, which resulted in improved sensitivity for detecting novel sncRNAs, more accurate expression levels of annotated species, and efficient resolution of closely related snoRNA species, such as those in the *DLK-DIO3* and *SNURF-SNRPN* loci. In addition, for those snoRNAs embedded in host genes, this approach can distinguish between host gene primary transcripts and mature, fully processed snoRNAs.

The best method to normalize small RNA-seq expression data is uncertain. For miRNAs, several studies have compared normalization methods, suggesting that the upper quartile, median, the DESeq normalization offered in the DESeq Bioconductor package, and the trimmed mean of M values offered in the edgeR Bioconductor package may be superior to the RPM normalization method[277-279]. In the absence of a "gold standard" for snoRNA expression, a rigorous comparison of normalization strategies of our small RNA-seq snoRNA expression data was not possible. Thus, in this study, we normalized our small RNA-seq data with the widely used RPM method.

To our knowledge, this is the first study to comprehensively analyze snoRNA expression in human hematopoiesis. snoRNAs are the most highly expressed sncRNAs

in all of the hematopoietic cell populations tested. Although snoRNAs have been considered to be housekeeping genes,[280] we identified a subset of snoRNAs that exhibit marked differential expression in a lineage- and development-specific pattern. This is particularly true for orphan snoRNAs contained in the imprinted *DLK-DIO3* and *SNURF/SNRPN* loci. The *DLK-DIO3* locus contains 47 orphan CD box snoRNAs that are highly expressed in CD34 cells and downregulated during myeloid or lymphoid differentiation. This observation is consistent with previous reports showing high hematopoietic stem/progenitor expression of lncRNAs[281] and miRNAs[282] that are contained in the *DLK-DIO3* locus. The *SNURF/SNRPN* locus contains 2 large orphan CD box snoRNA clusters, *SNORD115* and *SNORD116*, that are expressed highly in CD34 cells and downregulated during myeloid differentiation. Loss of *SNORD116* in the *SNURF/ SNRPN* locus is thought to be key to the pathogenesis of Prader-Willi syndrome, a genetic disorder characterized by obesity and developmental delay[283,284]. Of note, SNORD115 has been shown to promote alternative splicing of the serotonin receptor 2C[72]. The contribution of the DLK-DIO3 and SNURF/SNRPN loci to hematopoietic stem/progenitor function is un-known, although it is interesting to note that expression of ncRNAs from the DLK-DIO3 locus correlates with pluripotency in both embryonic and inducible pluripotent stem cells[285].

We observed no recurring mutations of snoRNA genes in our cohort of cytogenetically normal AML, suggesting that genetic alterations that specifically target snoRNAs in AML are uncommon. A previous study reported that snoRNAs are globally suppressed in AML relative to CD34 cells from normal hematopoietic stem/progenitors[99]. Although we also observed a trend toward decreased expression in AML, this was limited

to a small subset of snoRNAs (102 of 364, 28%). The reasons for this discrepancy are not certain, but the previous study primarily used a microarray approach to assess expression of a more limited set of snoRNAs. Of note, of the 102 snoRNAs with significantly reduced expression in AML, 37 are located in the *DLK-DIO3* or *SNURF-SNRPN* loci. Because expression of these snoRNAs is suppressed during normal myeloid differentiation, it is possible that their decrease in AML reflects normal differentiation along the myeloid lineage. This is in sharp contrast to previous studies showing marked increased expression of *DLK-DIO3* snoRNAs in acute promyelocytic leukemia[97,99,242]. Of note, Valleron et al., showed that enforced expression of *SNORD114-1*, which is contained in the *DLK1-DIO3* locus, promotes cell growth in vitro, possibly by targeting the *Rb* pathway.[99]

We observed significant differential expression of snoRNAs that mediate pseudouridylation or 2'-O methylation of key sites in rRNA. Decreased expression of snoRNAs that target modifications of the PTC and ISB regions of the 60S ribosome was observed in AML patients vs normal hematopoietic stem/progenitors. The PTC is the catalytic site where peptide bonds are made during protein elongation and peptidyl-tRNAs are hydrolyzed during the termination of protein synthesis[286]. The ISB forms multiple interactions between the ribosomal subunits, which maintain ribosome stability and modulate dynamics that are critical for translation, such as that between the tRNA and mRNA[287]. Studies in yeast suggest that, although loss of pseudouridylation or 2'-O methylation at individual rRNA sites has only subtle effects on activity, loss at multiple sites is synergistic, resulting in reading frame changes, increased stop - codon read-through, and altered tRNA selection[118,128,288]. We also identified several snoRNAs

responsible for the pseudouridylation of snRNAs in regions critical for splicing. For example, scaRNA15, whose expression is reduced 2.81-fold in AML, targets the branch site recognition region of U2 snRNA. Studies in HeLa cells[289] and yeast[162] show that pseudouridylation at this site is required for the formation of early spliceosomal complexes and the catalytic phase of pre-mRNA splicing. Further study is needed to determine whether the observed decreases in snoRNA expression in AML are sufficient to induce biologically meaningful differences in translation or splicing.

The mechanisms regulating snoRNA expression are not well defined. Most snoRNAs and scaRNAs are embedded in the introns of host genes that produce proteins involved in nucleolar function, ribosome structure, or protein synthesis,[3] providing a potential mechanism for the coordinated expression of snoRNAs and proteins targeting common pathways. Interestingly, we observed that in AML, snoRNA expression correlates minimally with host gene expression. Recent studies in yeast and human brain samples have reported a similar uncoupling of host gene and snoRNA expression[290-292]. Indeed, we even observed striking variability in the expression of snoRNAs contained in the same host gene. Mature snoRNAs are produced from host genes by exonucleolytic processing of the debranched intron after splicing[293,294]. A recent study suggested that alternative splicing of host genes contributes to the regulation of snoRNA expression and accounts, in part, for the variability in the expression of snoRNAs contained with the same host gene[295]. However, in AML, snoRNA expression and alternative splicing correlate minimally. Thus, in AML, other mechanisms besides host gene expression or splicing are contributing to mature snoRNA expression. This may include alterations in snoRNA processing, secondary snoRNA structure stability,

maturation, trans-acting protein accumulation factors, and intranuclear trafficking of the maturing snoRNPs to the nucleolus or Cajal body[296]. Of note, many snoRNA host genes contain a characteristic terminal oligopyrimidine track in their 59-untranslated region that has been shown to modulate the differential production of mRNA vs snoRNAs from that host gene[297,298]. Given the critical role of snoRNAs in translation, the contribution of these various elements to the regulation of snoRNA expression warrants further study.

As outlined, array-based and qPCR-based approaches do not distinguish between mature snoRNAs and primary mRNA transcripts containing unprocessed snoRNAs. Without robust orthogonal validation technologies for generating gold standard expression values, optimal statistical procedures for expression normalization from count-based sequence data have not been established for snoRNAs. This contrasts with miRNA sequencing data, for which qPCR provides robust orthogonal validation that has made it possible to evaluate and optimize expression normalization methods[278]. In the absence of a consensus approach for snoRNA data, we used the total count method, which involves normalization of the read count for each snoRNA species for the total number of counts obtained for each experiment. Additional studies will be needed to determine the optimal normalization procedures for sequence data from this intermediate-sized RNA species.

In summary, we developed a small RNA-seq pipeline to quantify snoRNA and other sncRNA expression. We showed that a subset of snoRNAs are regulated in lineage- and development-specific expression. Although genetic alterations that specifically target snoRNA genes in AML appear to be uncommon, a subset of snoRNAs are differentially expressed. The contribution of these differentially expressed snoRNAs

to the regulation of normal and malignant hematopoiesis represents an exciting new area of investigation.

Acknowledgments

This work was supported by National Cancer Institute (NCI), National Institutes of Health (NIH) grant PO1 CA101937 (D.C.L. and T.J.L.); Washington University School of Medicine Graduate School of Arts and Sciences/Chancellor's Graduate Fellowship Fund 94028C (W.A.W.); NIH, NCI grant K12CA167540 and Clinical and Translational Award UL1 TR000448 from the NIH, National Center for Advancing Translational Sciences (B.S.W.); and by NIH, NCI grant K08CA190815 and an American Society of Hematology Scholar Award (D.H.S.).

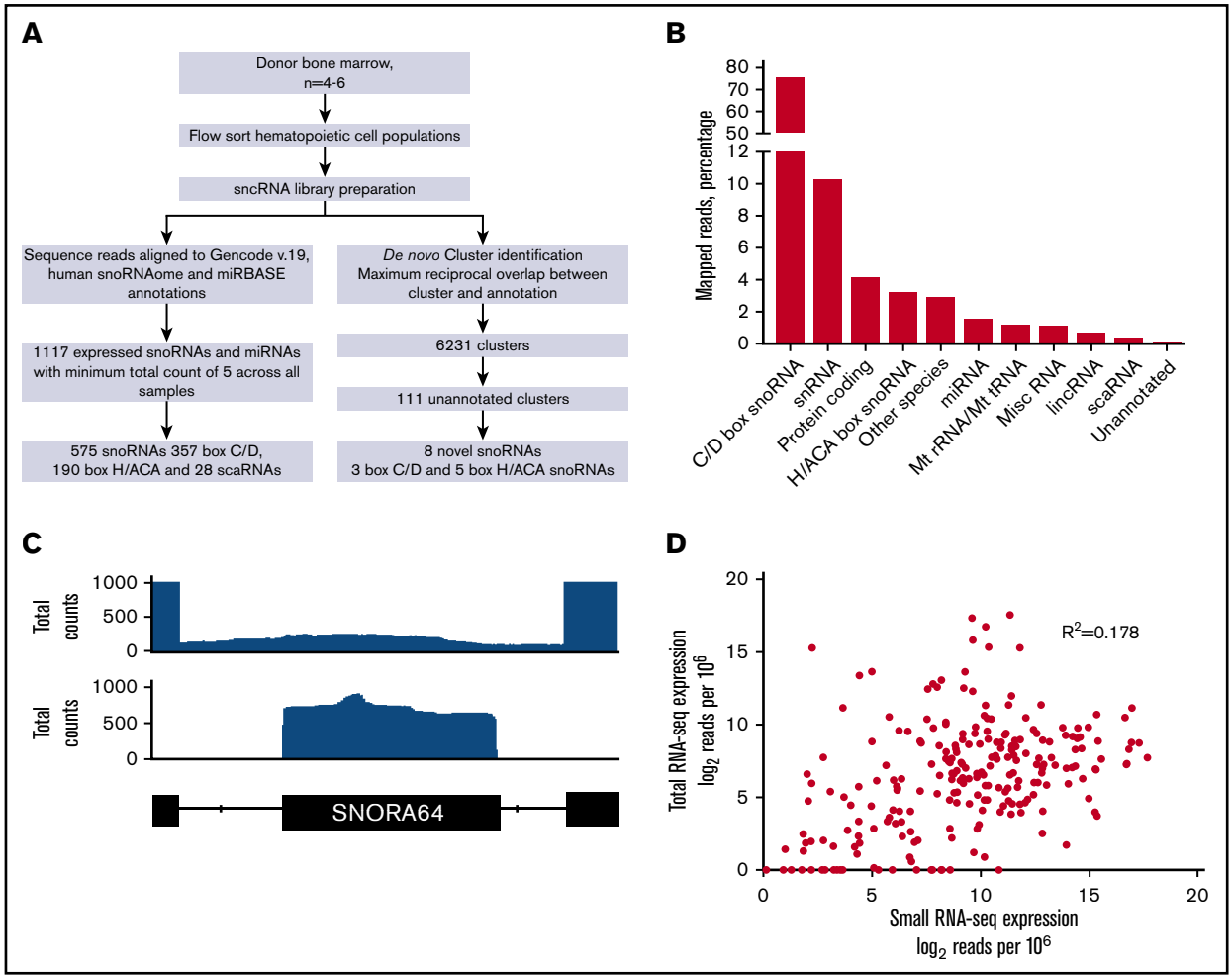


Figure 2.1 Small RNA-Seq pipeline. (A) Schematic representation of the small RNA-seq pipeline. Sequence reads were aligned to the GENCODE version 19, miRBase version 21, and human snoRNAome, and reads corresponding to annotated sncRNAs were quantified (left pathway). Alternatively, aligned sequence reads were organized into genomic clusters; a total of 111 unannotated clusters were identified, of which 8 were classified as novel snoRNAs (right pathway). (B) Graph showing the distribution of annotation biotypes in GENCODE version 19 for mapped reads from all samples. The percentage of all mapped reads (mapping quality >0) is shown on the y-axis, and the annotated species is shown on the x-axis. Miscellaneous (Misc) snoRNAs include rRNAs, other species include unprocessed pseudogenes, immunoglobulin genes, T-cell receptor genes, sense introns, antisense transcripts, sense overlapping transcripts, retained introns, and processed transcripts. Protein-coding genes include nonsense mediated decay, and nonstop decay biotypes. (C) Integrated Genome Viewer browser view of *SNORA64*, which is embedded in an intron of the coding gene *RPS2.* The top

panel shows the alignment of reads generated using total RNA-seq. The middle panel shows the alignment of reads from a small RNA-seq library produced from the same sample. The genomic boundaries of *SNORA64* and exons 4 and 5 of host gene *RPS2* are shown in the bottom panel. (D) Representative scatter plot showing log-transformed normalized read counts of annotated snoRNAs for a CD34 sample analyzed using total RNA-seq (y-axis) or our small RNA-seq pipeline (x-axis).

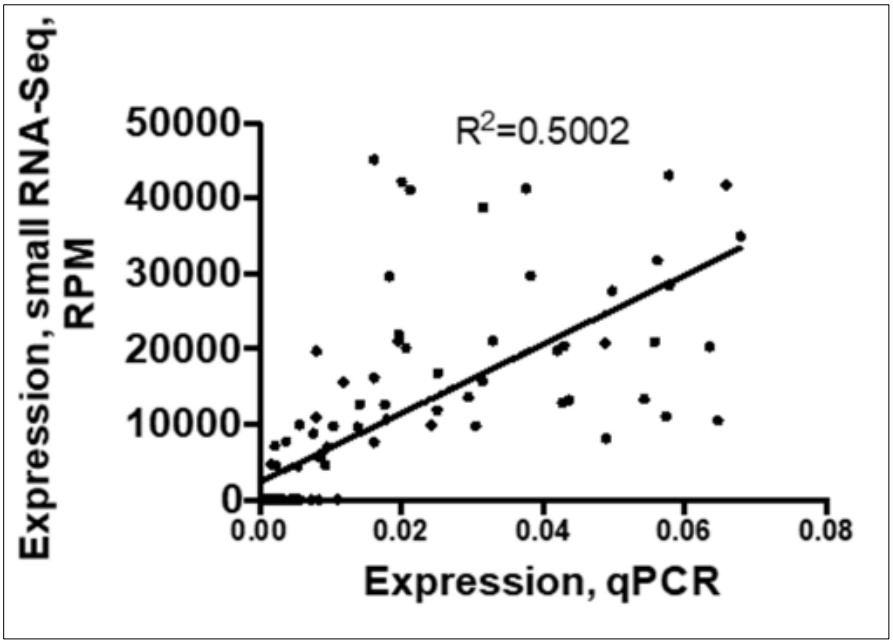


Figure 2.2. Correlation between by small RNASeq and RT-qPCR.

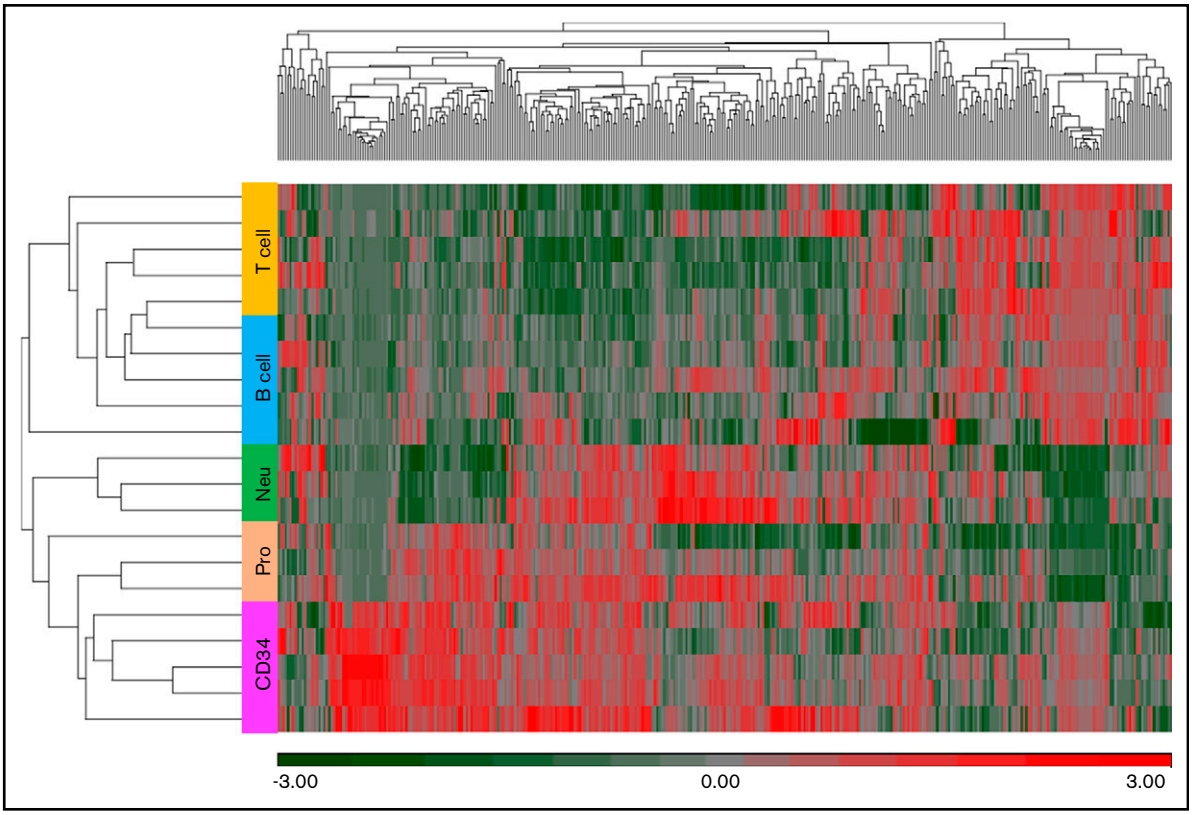


Figure 2.3 Unsupervised hierarchical clustering of snoRNA expression. Small RNA-seq was performed on the indicated sorted hematopoietic cell populations obtained from 4-6 normal hematopoietic stem/progenitors. snoRNA expression (z-scored $\log_2$ RPM) is shown, with red indicating high expression, and green indicating low expression. Each column represents a unique snoRNA, and each row represents a sample.

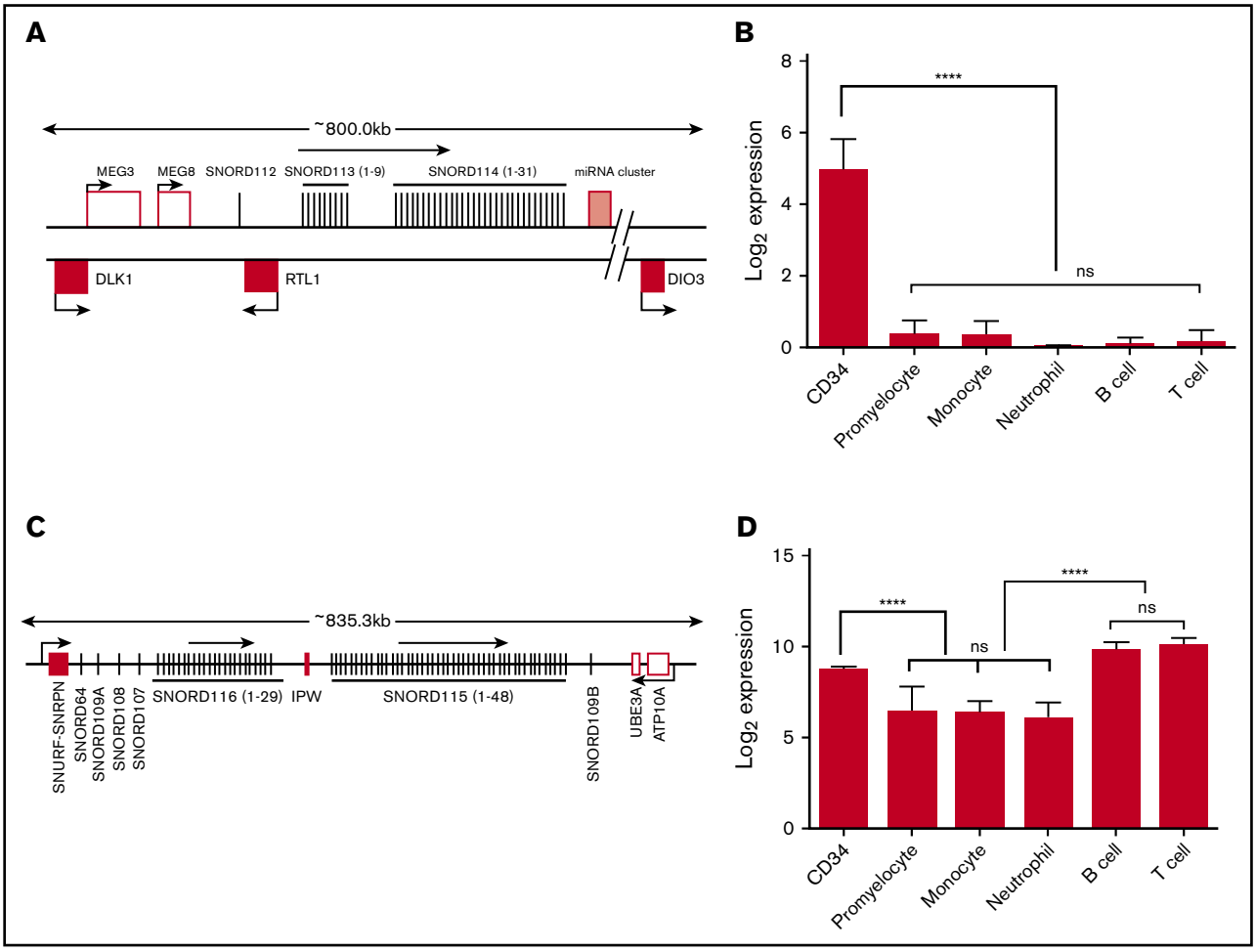


Figure 2.4. Expression of snoRNAs in the *DLK-DIO3* and *SNURF/SNRPN* loci. (A) Organization of the *DLK-DIO3* locus. Maternally expressed genes (open boxes) and paternally expressed genes (filled boxes) are shown. The miRNA cluster contains 54 miRNAs. (B) Normalized read counts for *SNORD113-6* are shown; its expression is representative of other SNORDs in this locus. Significance was determined by 1-way ANOVA followed by Tukey multiple comparisons test. (C) Organization of the human *SNURF-SNRPN* locus (drawing is not to scale). Paternally expressed genes, including SNORDs, are shown as n, and maternally expressed genes are shown as N. (D) Normalized read counts for SNORD116-1, representative of SNORDs in the *SNURF-SNRPN* locus, are shown. ****P < .0001. ns, not significant.

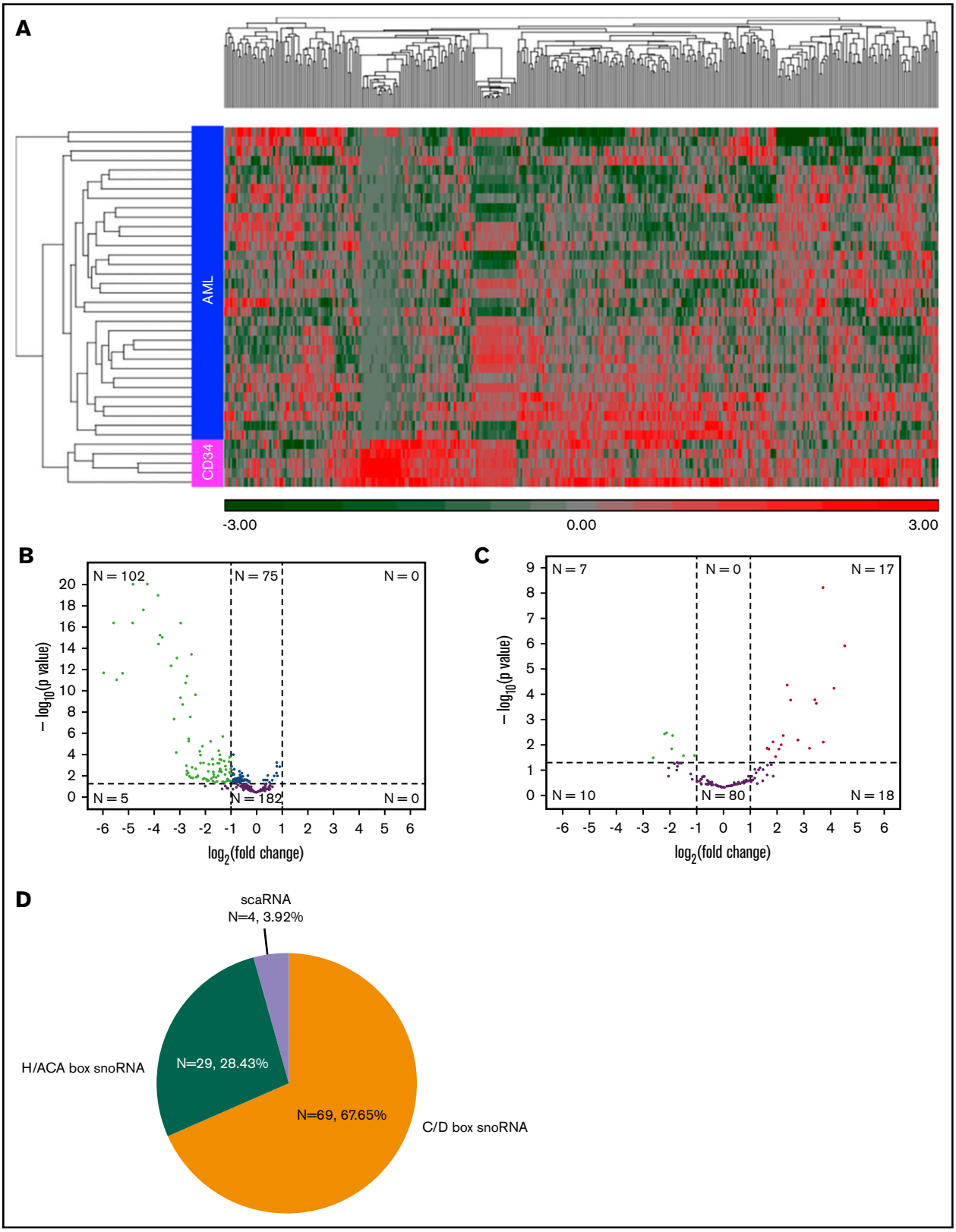


Figure 2.5. Differentially expressed snoRNAs. (A) Unsupervised hierarchical clustering of snoRNA expression in AML and CD34 from healthy adult donors based on

z scores, with red indicating high expression, and green indicating low expression. Each column represents a specific snoRNA, and each row represents an individual sample. (B-C) Volcano plot illustrating significant differentially expressed snoRNAs (B) and miRNAs (C) The fold change difference in RNA expression between AML and normal CD34 samples is plotted on the x-axis, and P value significance is plotted on the y-axis. The horizontal line on the plot represents the P-level used for this analysis (0.05). Vertical lines represent the threshold for the $\log_2$ fold change (equivalent to a twofold difference). The y-axis reflects the $-\log_{10}$ (q value–corrected P value). Green and red colored dots represent snoRNAs showing significantly increased or decreased expression in AML, respectively. (D) Distribution of differentially expressed snoRNAs across snoRNA subtypes. H/ACA snoRNA distribution includes 4 AluACA snoRNAs.

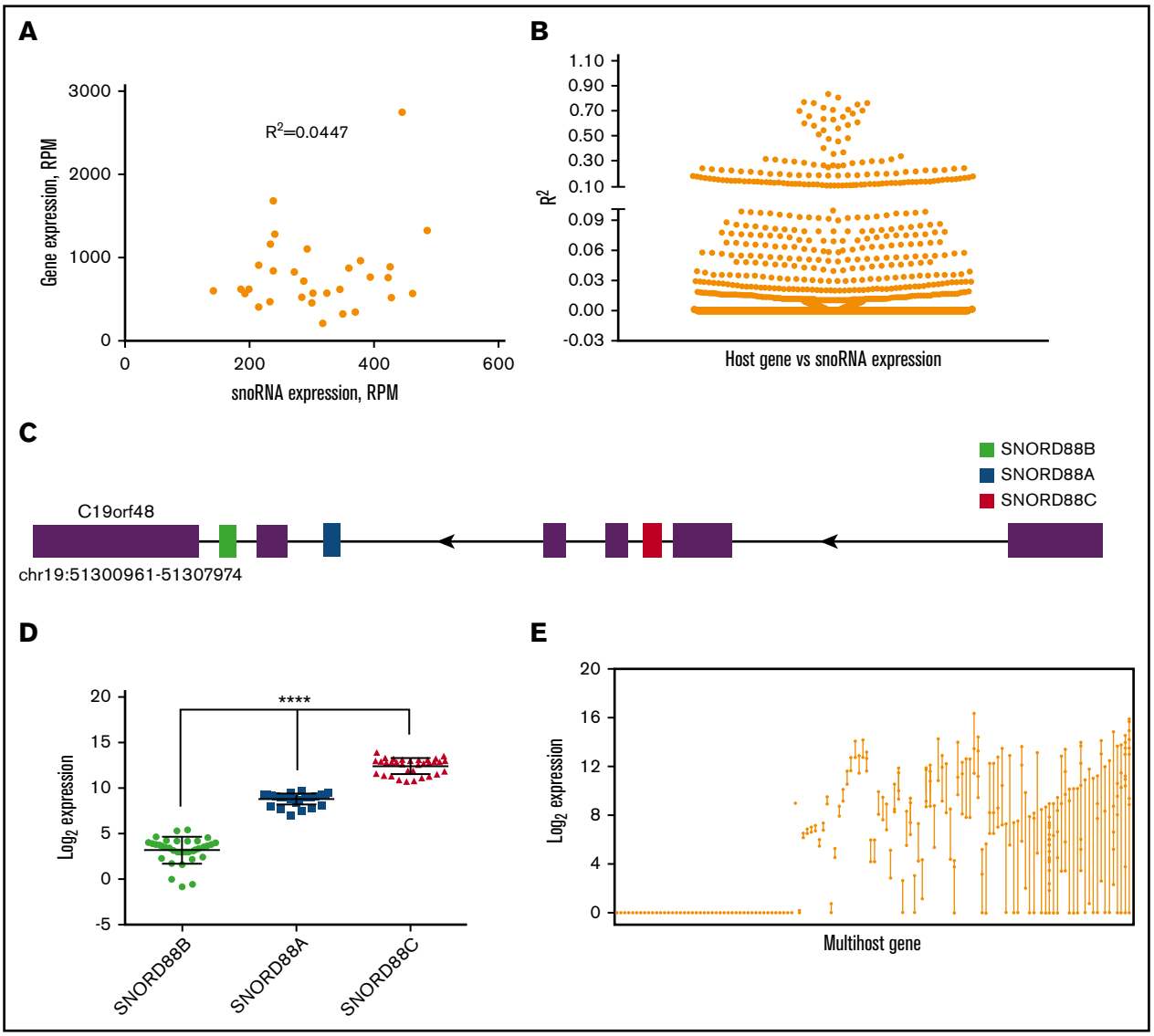


Figure 2.6. Correlation between host gene and snoRNAs expression. (A) Representative plot showing the correlation between the normalized expression of the host gene *RPL7A* and its encoded *SNORD36B* in AML. (B) Summary scatter plot showing the coefficient of determination (R 2) of all host gene/snoRNAs pairs (N = 1379, mean ± SEM = 0.037 6 ± 0.102). (C) Organization of the human gene *C19orf48* gene (drawing not to scale); exons are shown as a box and *SNORD88B, -88A* and *-88C* are shown in green, blue-, or red-filled boxes, respectively. (D) Log-transformed normalized read counts in AML. Significance was determined by 1-way ANOVA followed by Tukey multiple comparisons test. (E) Plot showing the log-transformed median expression values for all expressed snoRNAs in a multi-host gene (N = 130). Each line represents a distinct multi-host gene.

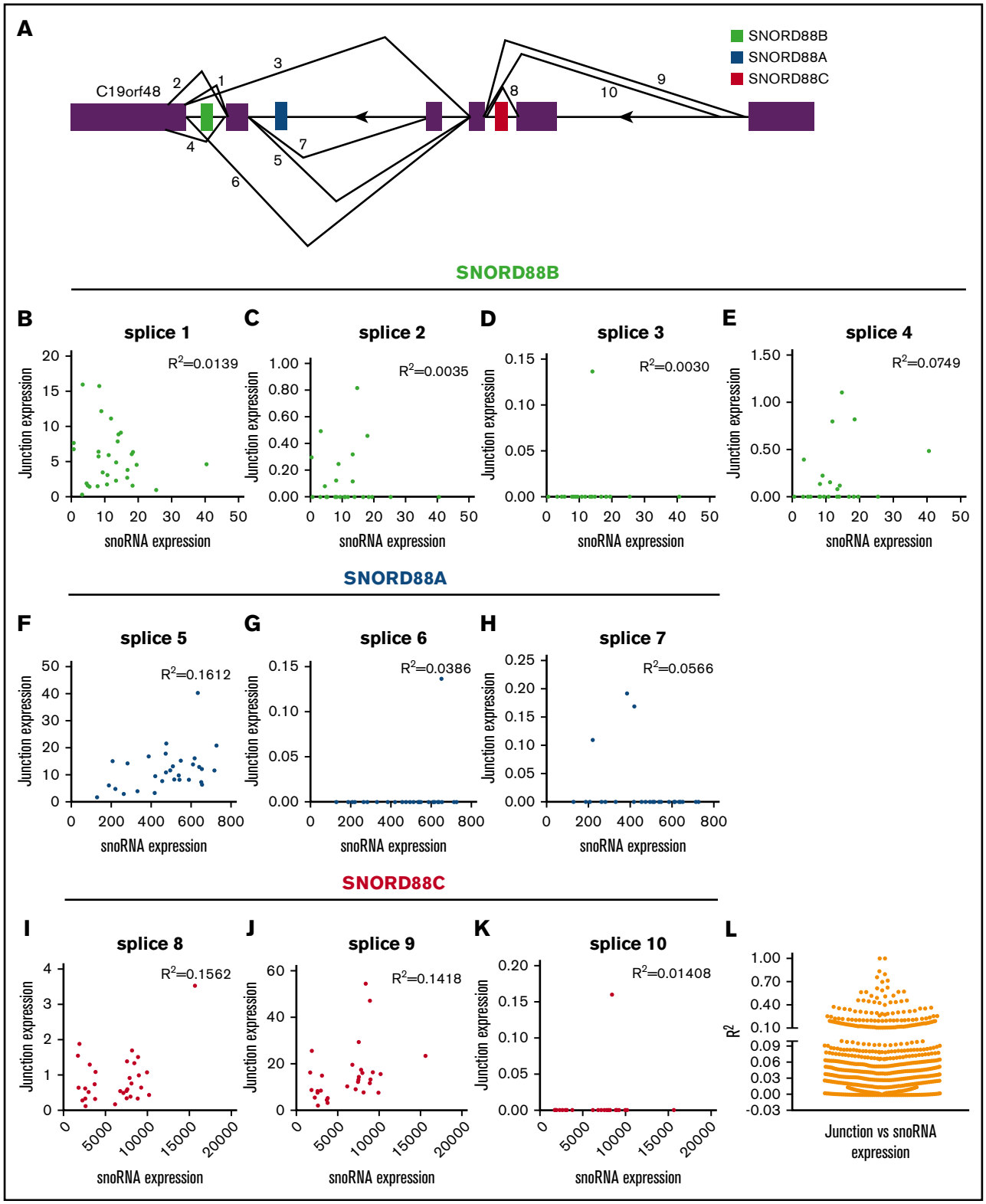


Figure 2.7. Correlation of host gene splicing to snoRNA expression. (A) Schematic of all splice junctions for C19orf48. Coding exons are shown as n and embedded

snoRNAs as color-filled boxes. Each numbered line represents a different splice event. Drawing not to scale. (B-K) RNA splicing was assessed by quantifying each of the junction reads (splice 1-10). Shown are scatter plots showing normalized junction reads (reads per million) vs snoRNA expression (reads per million). (L) Scatter plot showing R^2 values (mean ± SEM = 0.041 ± 0.093) between snoRNA expression and host gene splicing (total of 1616 junction reads).

CHAPTER 3

snoRNA EXPRESSION IN AML WITH FUSION GENES AND MUTATIONS IN SPLICING FACTORS

INTRODUCTION

AML is a complex and dynamic disease with a varied driver landscape providing discrete paths to leukemogenesis, disease pathophysiology, progression and recurrence[275,299,300]. This landscape includes fusion genes, and mutations in RNA-splicing regulators. snoRNA expression in these subtypes of AML is not well characterized. In the previous chapter, we looked at snoRNA expression in AML patient samples driven primarily by recurrent mutations in *NPM1, FLT3, DNMT3A, IDH1/2, NRAS* or *KRAS, CEBPA, TET2, WT1,* and *PTPN11*[192]. Here we expand on this work by looking at AML samples characterized by fusions, and mutations in splicing factors.

Fifteen percent of all AML cases and possibly up to 40% of M2 AML harbor the t(8;21)(q22;q22) translocation that leads to the *RUNX1-RUNX1T1* fusion oncoprotein (also known as AML1-ETO) (Figure 3.1). This oncoprotein includes the N-terminal DNA-binding domain of *RUNX1*, a DNA binding transcription factor essential for definitive hematopoiesis, and frequently mutated in *de novo* and therapy related AML, MDS, chronic myelomonocytic leukemia and acute lymphocytic leukemia[301,302]. A recent report suggested that the *AML1-ETO* oncoprotein alters ribosomal biogenesis by increasing expression of C/D box snoRNAs[102]. Increased expression of C/D box snoRNAs also was observed in cells expressing *MLL-AF9*. The authors propose that upregulation of C/D box snoRNA expression, by increasing 2′-*O*-methylation of ribosomal RNA, plays a key role in the induction of AML by *AML1-ETO* and possibly *MLL* translocations. However, microarrays were used to profile snoRNAs in this study, which may not be accurate, since they detect both the precursor and mature snoRNAs[192,303]. Moreover, it is not clear

whether other core binding factor AMLs, such as (16)/t(16:16) AML that express the *CBFB-MYH11* fusion gene, have altered snoRNA expression.

The t(15;17) translocation in acute promyelocytic leukemia (APML) generates a PML/RAR-α oncoprotein (Figure 3.1). Previous studies have reported marked increased expression of *DLK-DIO3* snoRNAs in APML[97,99,242]. The imprinted *DLK-DIO3* locus contains a large number of maternally expressed small non-coding RNAs, including 41 snoRNAs and 54 miRNAs. *DLK-DIO3* snoRNAs are highly expressed in primitive human CD34+ cells, and their expression is rapidly silenced during hematopoietic differentiation[192]. The function of *DLK-DIO3* snoRNAs is unknown. They are orphan CD-box snoRNAs that are computationally predicted, in one study, to target rRNA[241]. Whether other snoRNAs are dysregulated in APML is currently unknown.

Mutations in spliceosome genes are found in approximately 50% of MDS. The most commonly mutated spliceosome genes are *SF3B1, SRSF2,* and *U2AF1*. Mutations in each of these genes has been shown to result in altered splicing, although the specific genes that are affected are mostly distinct. This is relevant to snoRNAs, since the majority of snoRNAs are located in the introns of host genes and snoRNA biogenesis is dependent on splicing. Specifically, snoRNAs are processed exonucleolytically from spliced and debranched introns and released by endonucleolytic cleavages under the action of splicing factors[74]. Thus, we hypothesized that snoRNA biogenesis may be altered in myeloid malignancies carrying spliceosome mutations.

In this study, we sequenced the small RNA transcriptome of the following myeloid malignancies: core binding factor AML, *PML-RARa AML*, spliceosome gene mutated AML, and AML carrying an *MLL* rearrangement. We show that, in general,

snoRNA expression is decreased compared to CD34+ cells from healthy donors. Specific patterns of snoRNA expression were observed in each genetic AML subtype. However, we were not able to confirm an increase in C/D box snoRNA expression in core binding factor AML. Studies to assess the impact of splicing alterations on snoRNA expression are still in progress.

MATERIALS AND METHODS

We acquired AML samples characterized by fusions, and mutations in splicing factors as well as CD34 samples from healthy donors. The characteristics of these patients are fully described in Table 3.1, Figures 3.1-2 and elsewhere[304]. All patients in this study were enrolled in a single-institution tissue-banking protocol approved by the human studies committee at Washington University School of Medicine. Written informed consent for genome sequencing was obtained from all study participants.

All samples were interrogated to assess the snoRNA differential expression. Small RNA library construction was performed as previously described[192]. Sequencing was performed on an Illumina NextSeq 500 on a high output flowcell, targeting 400 million reads as a 2x150 configuration by the DNA Sequencing Innovation Lab at the Center for Genome Sciences and Systems Biology, Washington University, St. Louis, MO. The bioinformatic analysis pipeline has been previously described[192]. Differential expression analysis was carried out using Partek Genomic Suite as previously described[192].

Statistical significance was determined using Prism 6.0 software (GraphPad). Unpaired *t* test, 2-way ANOVA, or ANOVA with Tukey's Honest Significant Difference post hoc analysis was used to evaluate the significance of differences between multiple groups. All data are mean ± SEM.

RESULTS

Distinct patterns of snoRNA expression are associated with genetic subtypes of AML. We sequenced the small RNA transcriptome of 12 AML cases with an MLL translocation, 15 cases of core-binding factor AML, 16 AML cases with a spliceosome gene mutation, and 14 cases of APML (Table 3.1, Figures 3.1-2). These data were compared to CD34 cells isolated from 6 healthy donors.

Unsupervised hierarchical cluster analysis of expressed snoRNAs showed that differential snoRNA expression defined most of the AML subsets (Figure. 3.3). The number of differentially expressed snoRNAs (compared to normal CD34 cells) for each AML subset is shown in Table 3.2. For most of the AML subtypes, a similar pattern was observed, with decreased expression of C/D box snoRNAs representing the majority of differentially expressed snoRNAs (Table 3.2). The exception is AML carrying the *PML-RAR* oncogene, where increased expression of C/D box snoRNA was observed (see section on PML-RARa AML below).

We next asked whether there were snoRNAs that showed differential gene expression across all or multiple AML subtypes. The intersectionality of the differentially expressed snoRNAs revealed that there are 36 differentially expressed snoRNAs in common (Figure 3.4, Table 3.3). Of these, 22 (77.8%) were C/D box snoRNAs with 8 housed in the *DLK-DIO3* locus. Two of the snoRNAs (*SNORD1A*, and *SNORA10*) play a key role in the PTC modification and one (*SNORD36C*) is involved in modification of the ribosomal ISB.

Expression of snoRNAs from the imprinted *DLK-DIO3* and *SNURF-SNRPN* loci is markedly increased in APML.

The imprinted *SNURF-SNRPN* locus is the largest cluster of C/D box snoRNAs in the human genome. It includes 24 members of the SNORD116 family, 47 copies of the *SNORD115* family, along with *SNORD107, SNORD164, SNORD109A*, and *SNORD109B* (Figure 3.5A). Expression of many of the snoRNAs in this locus were markedly decreased in APML compared to normal CD34 cells (Figure 3.5B). For example, expression of *SNORD116-14* was increased nearly 500-fold. Additionally, expression of *SNURF-SNRPN* locus snoRNAs was decreased relative to CBF and MLL translocations and some samples with mutations in splicing factors. The *DLK-DIO3* imprinted locus contains another large cluster of C/D box snoRNAs (Figure 3.6A). Consistent with prior studies, we observed a marked increase in expression of *DLK-DIO3* snoRNAs in APML (Figure 3.6B)[97,99,242]. In contrast to APML, expression of snoRNAs in the *DLK-DIO3* locus in other types of AML was decreased related to CD34 cells, consistent with our prior study[192].

C/D box snoRNAs expression in Core Binding Factor AML and AML carrying MLL translocations is not globally increased.

A prior study reported that C/D box snoRNA expression was increased in AML with *RUNX1-RUNX1T1* t(8;21) expressing the *AML1-ETO* oncogene[102]. To confirm and extend this finding to other core binding factor AMLs, we analyzed 15 cases of CBF AML, including 10 cases with *MYH11* (inv(16)/t(16;16)) and four cases with *RUNX1-RUNX1T1* (t(8;21)(q22;q22)). Compared to normal CD34+ cells, a total of 109 differentially

expressed snoRNA were identified, with the majority being downregulated (Figure 3.7A). Limiting the analysis to C/D box snoRNAs, 67 differentially expressed snoRNAs were identified, of which only 6 were upregulated (Figure 3.7B). To directly compare our results to that of Zhou and colleagues, we limited our analysis to the 4 four cases of AML with *RUNX1-RUNX1T1*. A total of 50 differentially expressed C/D Box snoRNAs were identified, of which 5 were increased. Indeed, only one of the C/D box snoRNAs (*SNORD104*) reported as overexpressed by Zhou et al showed a similar pattern in our cohort (Supplementary Table 3.1)[102].

Zhou et al also reported that expression of *MLL-AF9* resulted in increased C/D box snoRNA expression. Thus, we also analyzed 11 cases of AML carrying translocations of *MLL* (Table 3.1). A total of 84 differentially expressed snoRNAs were identified, including 44 C/D box snoRNAs with decreased expression versus just 6 C/D box snoRNAs with increased expression (Supplementary Table 3.2).

Heterogeneity of snoRNA expression in AMLs with splicing factor mutations

We next asked whether splicing changes present in AML carrying spliceosome mutations affected snoRNA expression. A total of 16 AML cases were analyzed, including 11 with *U2AF1* mutations, 3 with *SRSF2* mutations, and 2 with *SF3B1* mutations. Comparing spliceosome gene-mutated AML as a group to normal CD34 cells, 107 differentially expressed snoRNAs were identified, with the majority (94%) showing decreased expression (Figure 3.8A, Supplementary Table 3.3). This pattern is held true for each spliceosome gene mutation subtype (Figure 3.8B). There is evidence that distinct splicing alterations are associated with the different spliceosome gene mutations. However,

unsupervised hierarchal clustering showed no distinct patterns for AML carrying *U2AF1*, *SRSF2*, or *SF3B1* mutations (Figure 3.9). Indeed, little overlap was observed in the differentially expressed snoRNAs observed in AMLs carrying (Figure 3.10, Table 3.4). Of note, AML with *U2AF1* splicing factor mutations decreased expression had the most unique DE snoRNAs.

To look in more detail at the effect of host gene splicing to snoRNA expression, we performed deep RNA sequencing on a subset of the AML samples carrying spliceosome gene mutations. The identification of splicing alterations in these samples is underway.

DISCUSSION

AML is a complex disease driven by a variety of factors that culminate in a block to differentiation and a build-up of leukemic blasts. Recently, there has been significant interest in snoRNAs and the role they play in AML. Many of these studies rely on less than optimal hybridization approaches to quantify snoRNAs. In the previous chapter, we reported on our development of optimal approaches for quantifying snoRNA. In the current study, we build on that work by sequencing the small RNA transcriptome of 12 AML cases with an MLL translocation, 15 cases of core-binding factor AML, 16 AML cases with a spliceosome gene mutation, and 14 cases of APML.

We found that distinct patterns of snoRNA expression are associated with genetic subtypes of AML and that these patterns provide a signature to differentiate AML subtypes. This varied and specific expression contradicts earlier thinking that snoRNAs are simply housekeeping genes. Intersecting the subsets of differentially expressed snoRNAs we found that a subset (36) were differentially expressed across all of the classes we interrogated. Whether these snoRNAs play a common role in maintaining the leukemic state is not clear and is worthy of further investigation. Interestingly, 8 of the common orphan snoRNAs are in the DLK-DIO3 locus and as such they are overexpressed in APL samples and under expressed in all other AML relative to CD34, suggesting that they might have a functional relevance. Additionally, several of these snoRNAs modify targets in the PTC suggesting that translation might be commonly impacted in these AML subtypes.

Recent work by Zhou *et al.,* reported that CD box snoRNAs are overexpressed in *AML-ETO* samples and that this plays a critical role in leukemogenesis. They point to C/D box snoRNA overexpression as a driver of leukemogenesis. However, we were

unable to confirm these findings in 15 CBF AML samples, including 4 AML cases expressing *AML-ETO*. These discrepant results may be secondary to the method used for snoRNA quantification. Whereas we used our small RNA sequencing pipeline, Zhou et al., used microarrays, which may be less accurate. Of note, Zhou and colleagues used RT-PCR to confirm overexpression of 9 C/D box snoRNAs in 4 *AML-ETO* AML samples. However, expression of only one of these, *SNORD104*, was increased in our cohort.

DLK-DIO3 snoRNA genes are tandemly repeated C/D snoRNAs located ~25 kb downstream of *Rtl1* and processed from a complex transcription unit mapping to the *MEG8* gene[239]. There have been numerous studies that used sub-optimal hybridization techniques to report on a *DLK-DIO3* APML signature in which the snoRNAs (*SNORD113-3, SNORD113-4, SNORD114-2* and *SNORD114-3*) are up-regulated relative to CD34 and NK AML[97,99,242]. We report similar findings here and provide a complete picture of the APML snoRNA landscape for the first time.

One area of significance is the *SNURF/SNRPN* locus which has a similar, but distinct, pattern of snoRNA expression. This locus contained 82 paternally expressed snoRNAs localized in the Prader-Willi critical region on chromosome 15[68]. This locus is one of the most complex transcriptional units in the human genome spanning more than 460 kb and containing at least 148 exons. The locus harbors a bipartite imprinting center (IC) that silences most maternal genes of the Prader-Willi critical region[68]. As we noted for all AML, the *SNORD115* family is not expressed. Here we report for the first-time suppressed expression of orphan *SNORD116* snoRNAs in APML samples. This is in contrast to the pattern observed for all other AMLs we have looked at. Clearly the

imprinted loci are dysregulated in APL. Further studies are needed to better understand the role of these orphan snoRNAs in this setting.

A recent study looking at Prader-Willi syndrome (PWS), a developmental imprinting disorder, unexpectedly found substantial upregulation of virtually all maternally expressed genes (MEGs) in the imprinted DLK1-DIO3 locus[305]. They report that IPW, a long noncoding RNA in the *SNURF/SNRPN* locus, is a regulator of the *DLK1-DIO3* region, as its overexpression in *SNURF/SNRPN* and parthenogenetic iPSCs resulted in downregulation of MEGs in this locus[305]. The reported role of *IPW* is to facilitate the methylation of H3K9, thus reducing expression of the MEGs in the *DLK1-DIO3* locus. The authors show that loss of imprinted genes in the *SNURF/SNRPN* locus leads to a trans effect where increased expression of imprinted genes in the *DLK1-DIO3* locus is observed. Our gene expression data seems to support this finding. Taken together, this cross talk and the expression patterns we report might lay the foundation to finally elucidate a role for the imprinted snoRNAs.

Given that snoRNAs are intronic[74] and are processed under the action of the spliceosome we looked at snoRNA patterns in AML samples with mutations in splicing factors, *SF3B1, SRSF2,* and *U2AF1.* Mutations in each of these genes has been shown to result in altered and distinct splicing patterns[306]. One commonality was the reduced expression of a subset of snoRNAs. It is noteworthy that intersecting the differentially expressed genes there were only 5 snoRNAs in common. This reinforces the idea that these splicing factors elicit different alterations in splicing. *U2AF1* recognizes the AG splice acceptor dinucleotide at the 3′ end of introns and mutations leads to exon skipping and/or retention. Mutations in this gene seem to have a dramatic effect on snoRNA

expression as 93 of the differentially expressed snoRNAs are unique to it. Several snoRNAs that modify critical PTC and ISB snoRNAs are differentially expressed suggesting that there might be some impact on ribosome biogenesis and thus translation. This is an area worthy of further investigation.

As a whole, our study expands on the previous chapter and provides new expression data for snoRNAs in AML with an MLL translocation, core-binding factor AML, AML cases with a spliceosome gene mutation, and cases of APML. We have ongoing analysis to look at alternative splicing patterns in the context of snoRNA expression.

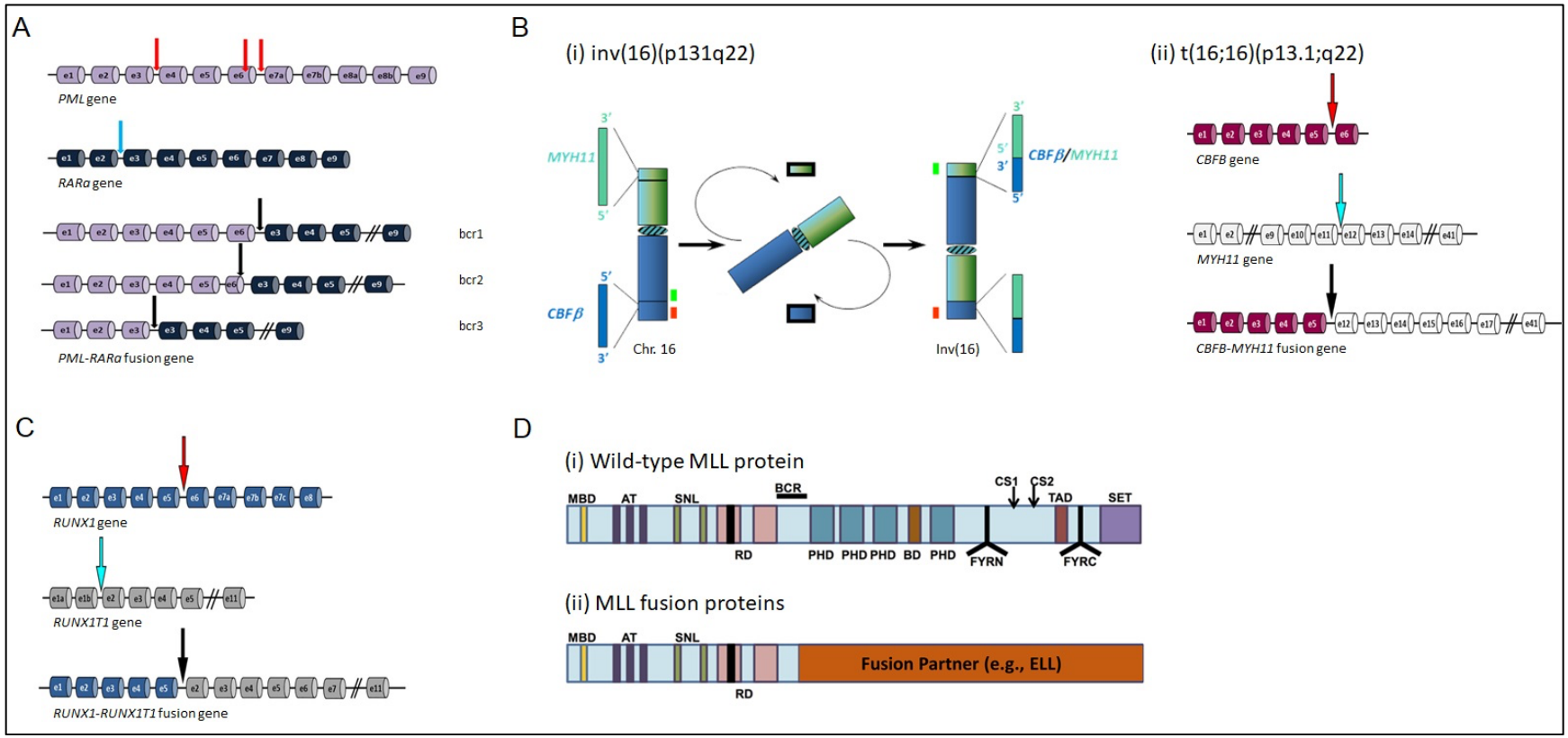


Figure 3.1. Genomic organization of the fusion loci. A) Schematic representation of *RUNX1, RUNXT1* and *RUNX1-RUNX1T1* genes. The red arrow indicates a breakpoint between exons e5 and e6 of the *RUNX1* gene. The blue arrow indicates a typical breakpoint within the *RUNX1T1* gene – between exons e1 and e2. The black arrow marks a junction between the *RUNX1* (behind exon e5) and *RUNX1T1* (in front of exon e2) genes. B) Schematic representation of MYH11, CBFB, and MYH11-CBFB genes. (i) The inversion fuses the 5′ end of the *CBFB* gene, located on 16q22, upstream of the 3′ end of *MYH11* on 16p13. (ii) The red arrow indicates a breakpoint between exons e5 and e6 of the *CBFB* gene. The blue arrow indicates a typical breakpoint within the *MYH11* gene – between exons e11 and e12. The black arrow marks a junction

between the *CBFB* (behind exon e5) and *MYH11* (in front of exon e12) genes. C) Schematic representation of the *PML, RARA* and *PML-RARA* genes. The red arrows indicate the three most common breakpoints in the *PML* gene: between exons e3 and e4, within exon e6 and between exons e6 and e7. The blue arrow indicates a typical breakpoint in the *RARA* gene – between exons e2 and e3. The black arrows mark junctions: i. bcr1 – behind exon e6 of the *PML* gene and in front of exon e3 of the *RARA* gene; ii. bcr2 – within exon e6 of the *PML* gene and in front of exon e3 of the *RARA* gene; iii. bcr3 – behind exon e3 of the *PML* gene and in front of exon e3 of the *RARA* gene. D) Schematic representation of the structure of mixed-lineage leukemia (MLL) and normal versus aberrant MLL complexes. (i) The structure of the wild-type MLL protein, emphasizing the functional domains. MBD, Menin-binding domain; AT, AT hooks; SNL, speckled nuclear localization domains; RD, repression domains (black box in first RD represents the CXXC domain); BCR, breakpoint cluster region; PHD, PHD fingers; BD, bromodomain. CS1 and CS2 are the taspase-1 cleavage sites, and FYRN and FYRC are the domains whereby MLL-N and MLL-C interact after cleavage. TAD, transactivation domain; SET, H3K4 histone methyltransferase domain. (ii) MLL fusion proteins are caused by chromosomal rearrangements leading to in-frame fusions between N-terminal MLL (to the BCR) and any of 80 different possible fusion partners. PHD domains, transactivation domains, and the SET domain are typically lost. Images and legend reproduced/adapted from Atlas of Genetics and Cytogenetics in Oncology and Haematology. http://atlasgeneticsoncology.org Accessed August 2, 2019.

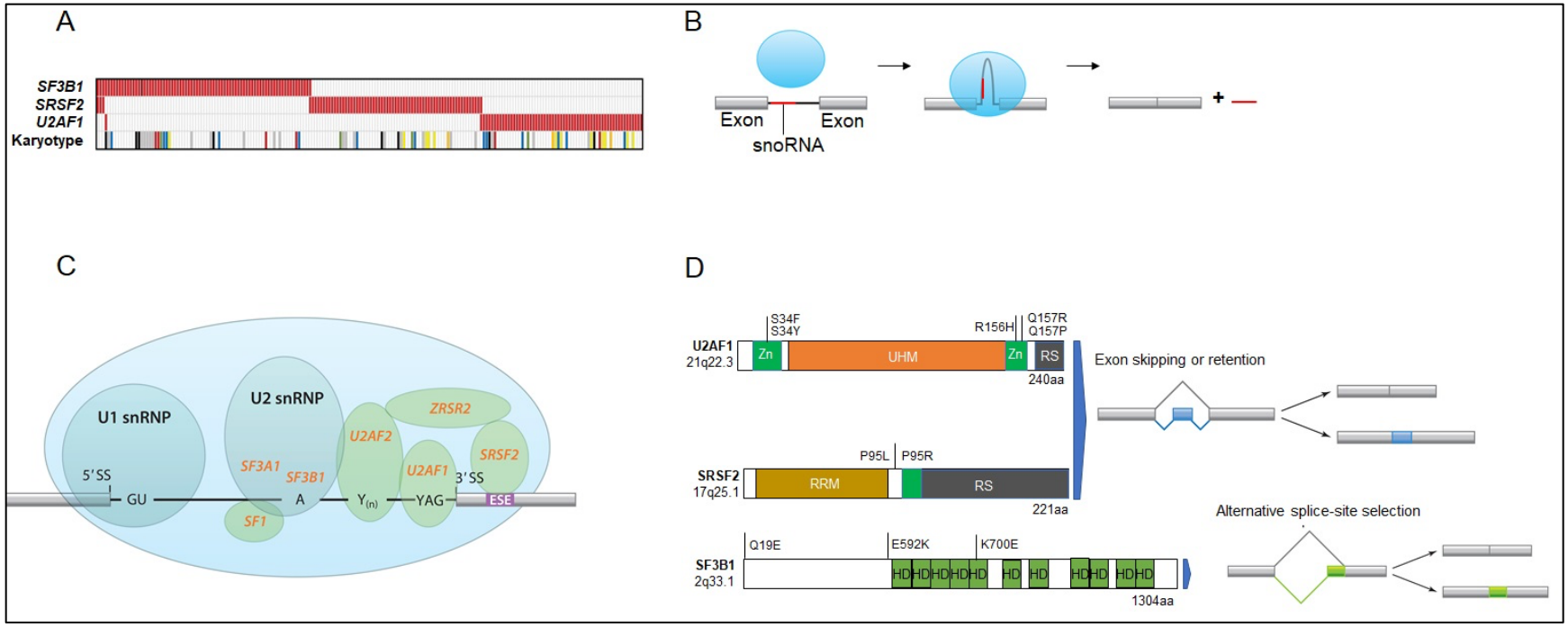


Figure 3.2. Mutations that affect RNA splicing machinery are common in myelodysplastic syndromes (MDS) and AML. A) Mutations in genes that encode key components of the RNA splicing machinery are mutually exclusive and occur in a variety of cytogenetic contexts. Each column represents one patient sample with mutations shown as red bars. Karyotypic abnormalities represented here are: red, isolated deletion on the long arm of chromosome 5 [del(5q)]; green, −7/del(7q) alone or +1 abnormality; blue, isolated +8; yellow, isolated del(20q); black, complex; white, normal or −Y; orange, unknown. B) Schematic of RNA splicing showing spliceosome-mediated precursor-messenger RNA processing that causes the excision of an intervening intron housing a snoRNA to generate a snoRNA and ligation of flanking exons. C) Selected components of the spliceosome are typically mutated in MDS/AML. The U1 small nuclear ribonucleoprotein (snRNP) binds to the 5' splice site (5' SS), and the U2 snRNP binds to the branch point. The U2 auxiliary factor (U2AF), which is composed of *U2AF1* and *U2AF2* subunits, binds to the 3' splice site (3' SS) and polypyrimidine tract [$Y_{(n)}$]. Serine- and arginine-rich proteins bind to the exonic splice enhancer (ESE). The genes encoding *SF3A1, SF3B1, U2AF1, U2AF2, SF1, SRSF2,* and *ZRSR2* are mutated in myeloid neoplasms and are shown in red. (d) Schematic of *U2AF1, SF3B1*, and *SRSF2's* mutational

landscape is shown. Outcomes of alternative splicing involving these splicing factors include alternative splice-site selection, and exon skipping. Images and legend reproduced/adapted from Lindsley RC, and Ebert BL. Annu Rev Pathol. 2013.

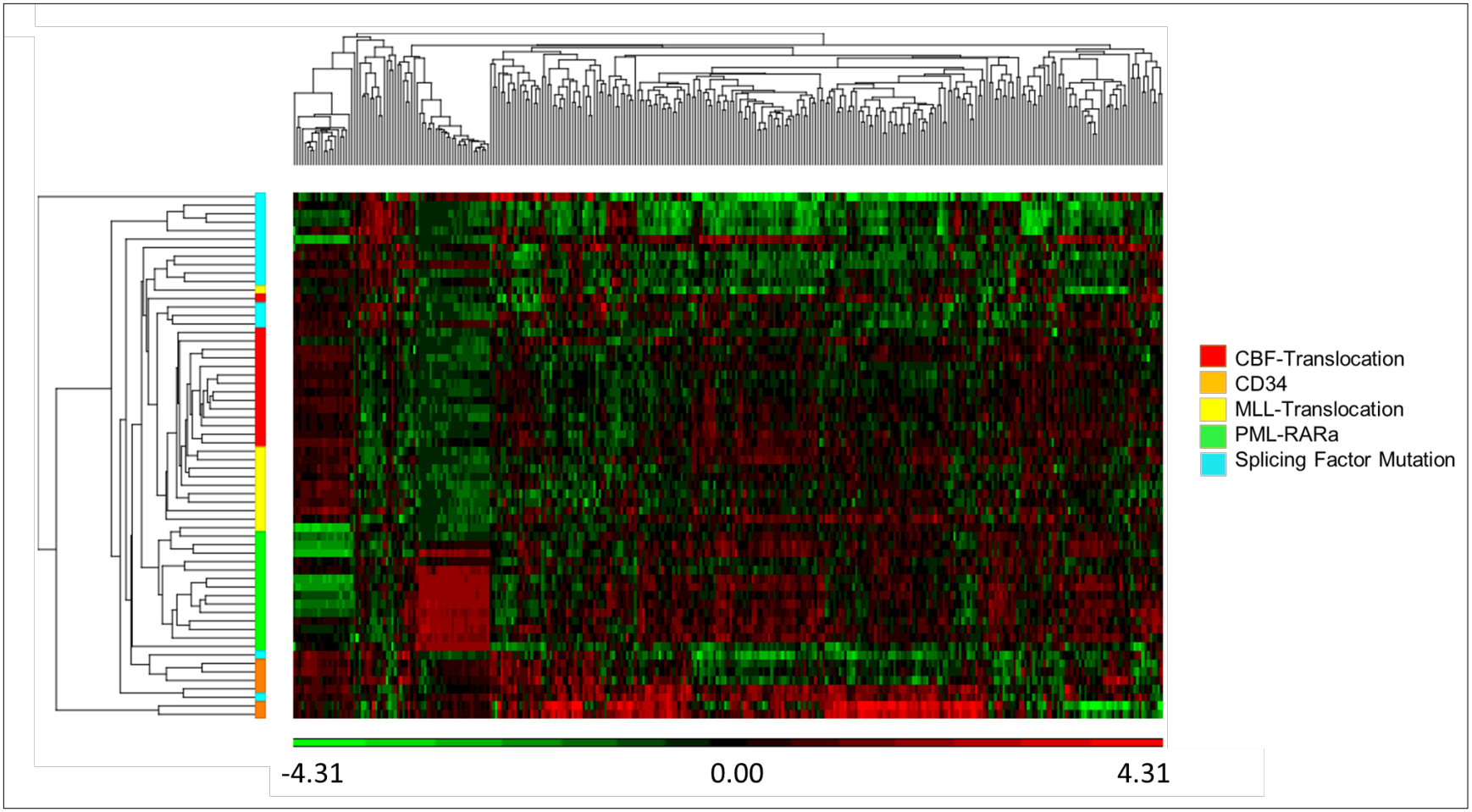


Figure 3.3. Dendrogram showing unsupervised hierarchical clustering of snoRNA expression in 58 AML samples. Small RNA-seq was performed on 58 AML samples with the indicated genomic background and CD34 samples from 6 healthy donors (Table 1). snoRNA expression (z-scored $\log_2$ RPM) is shown, with red indicating high expression, and green indicating low expression. Each column represents a unique snoRNA and each row represents a sample.

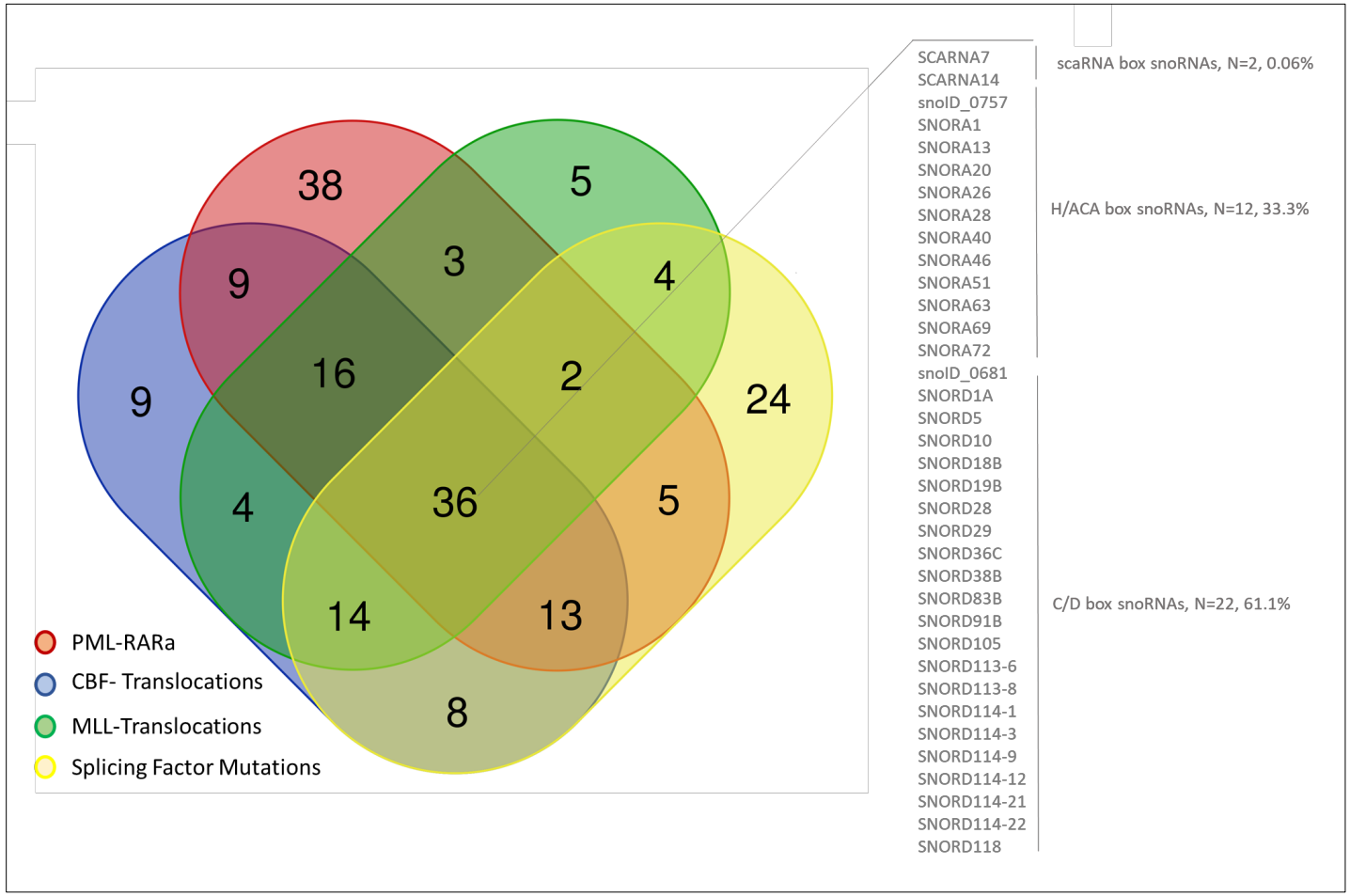


Figure 3.4. Venn diagram of differentially expressed snoRNAs within each dataset. Venn diagram is used to show the intersection of the numbers of the differentially expressed snoRNAs in AML samples as characterized. The numbers in the intersections represent the absolute numbers of the comprised differentially expressed overlapping/unique snoRNAs, and the listing on the right represents the 36 differentially expressed snoRNAs common to all.

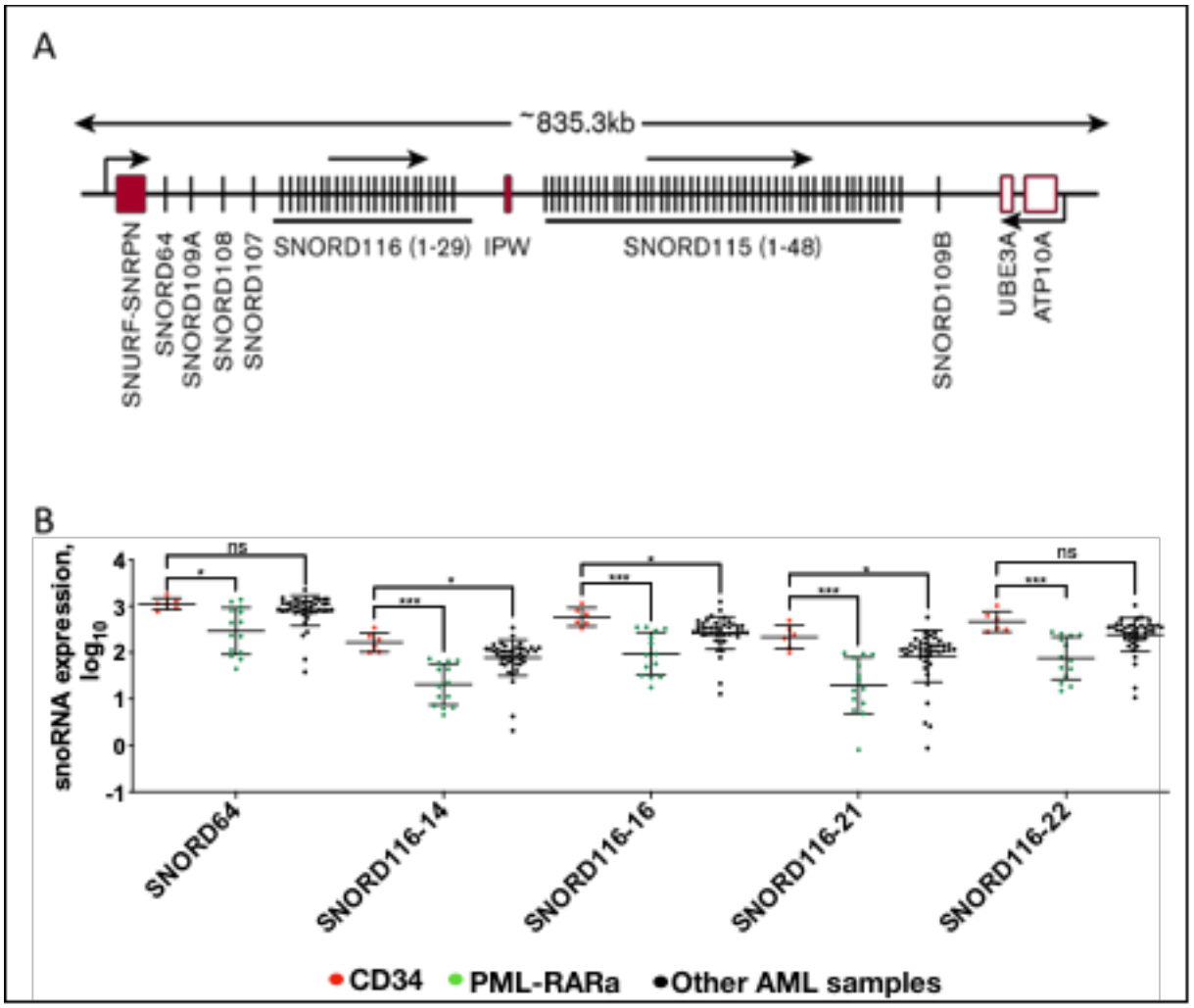


Figure 3.5. *SNURP-SNRPN locus* snoRNAs have suppression expression in *PML-RARa* samples. A) Schematic representation of the *SNURP-SNRPN* domain in humans showing genes expressed from the maternal chromosome and snoRNAs expressed from the paternally inherited chromosome. Not drawn to scale. B) Expression of representative *SNURP-SNRPN* snoRNAs in CD34, *PML-RARa* and all other AMLs in study.

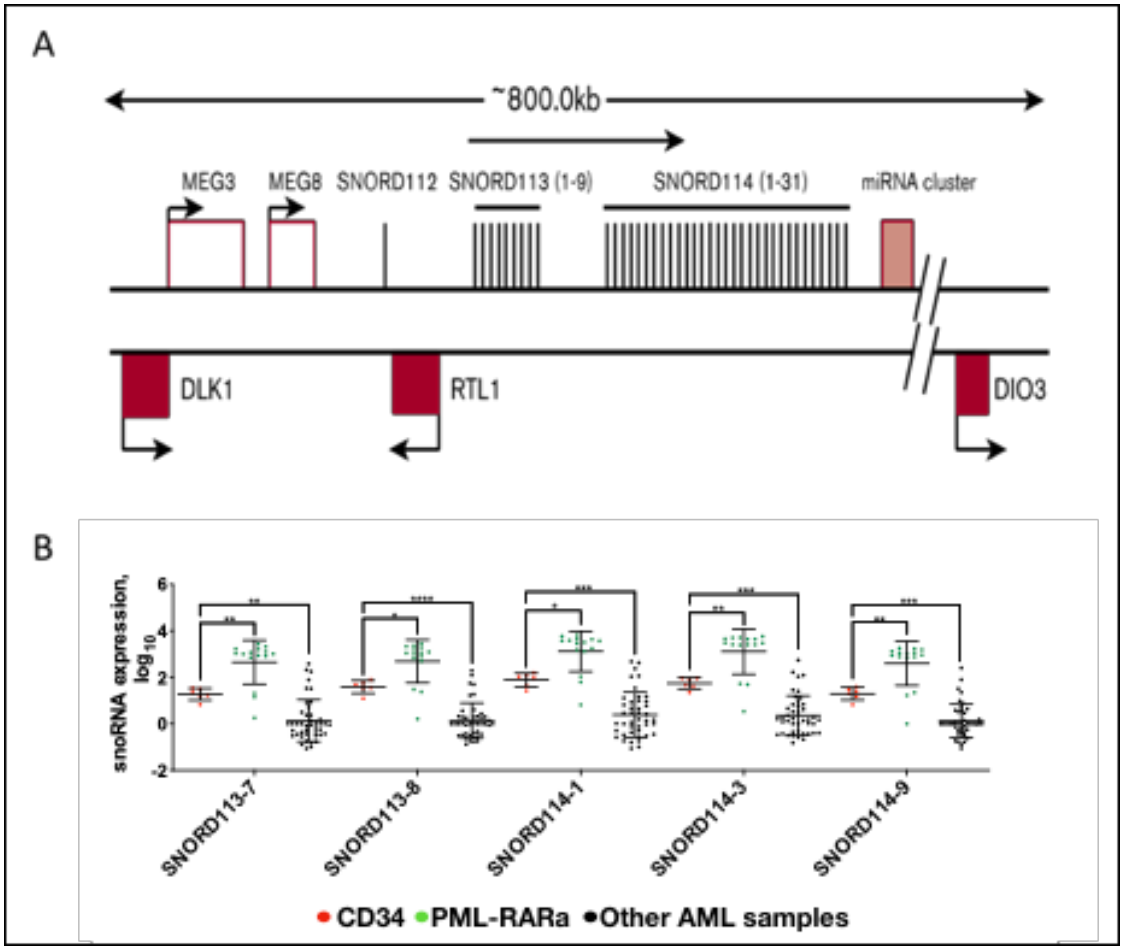


Figure 3.6. *DLK-DIO3* snoRNAs have increased expression in *PML-RARa* samples. A) Schematic representation of the *DLK1-DIO3* domain in humans showing genes expressed from the paternal chromosome and snoRNAs expressed from the maternally inherited chromosome. Not drawn to scale. B) Expression of representative *DLK-DIO3* snoRNAs in CD34, *PML-RARa* and all other AMLs in study.

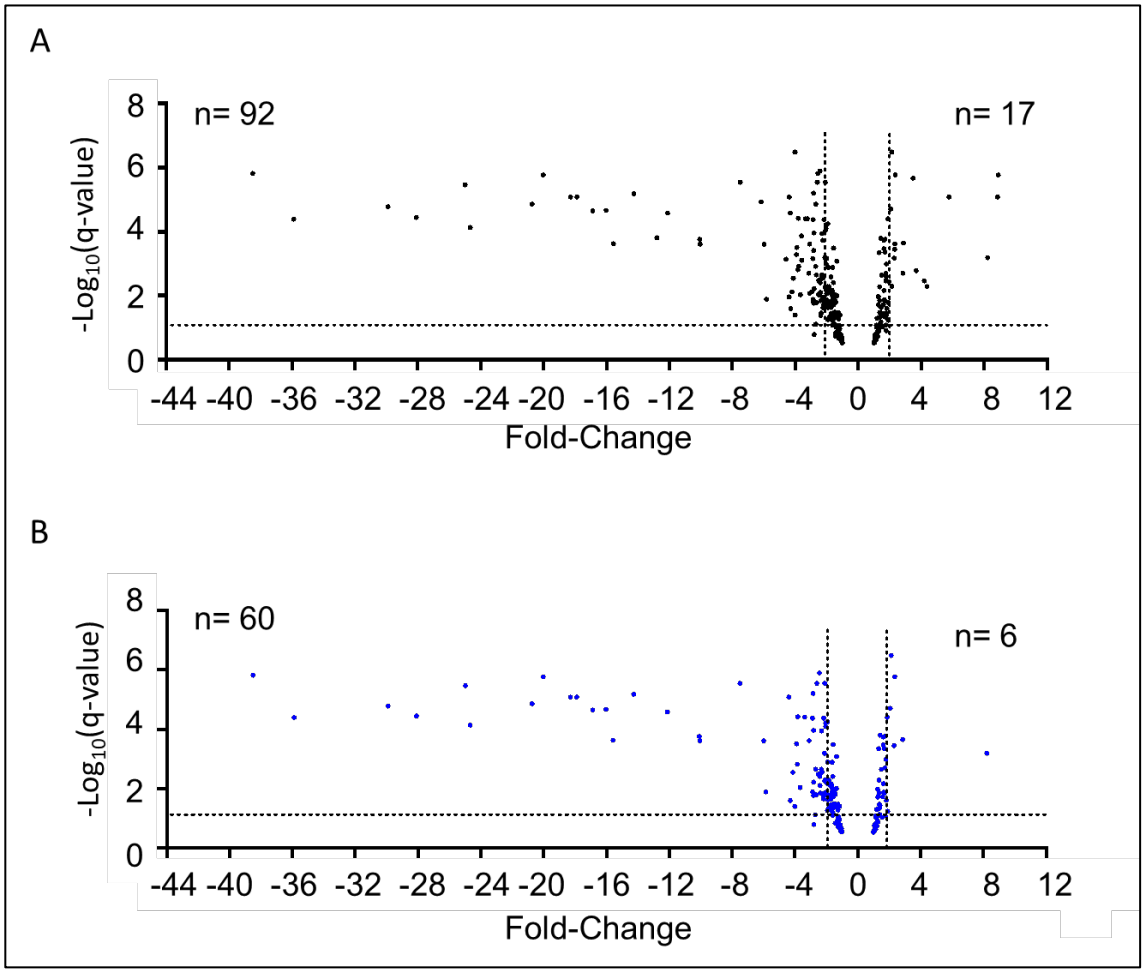


Figure 3.7. snoRNA differential expression plots for AML CBF (*RUNX1-RUNX1T1* and *MYH11-CBFB*) fusions. Small RNA-seq was performed on 15 AML samples with *RUNX1-RUNX1T1* (N = 4) and *MYH11-CBFB* (N = 11) fusions and CD34 from 6 healthy donors (Table 1). Volcano plots show the significance level of the expression difference between A) all snoRNAs and B) all C/D box snoRNAs in samples with CBF fusions versus CD34. Each dot represents one snoRNA. The

horizontal line on the plot represents the P-level used for this analysis (0.05). Vertical lines represent the threshold for a twofold difference. The y-axis reflects the $-\log_{10}$(q value–corrected P value). Dots in the upper right and left quadrants represent differentially expressed snoRNAs which have absolute fold change value >2 and $P \leq 0.05$. Differential expression and hierarchical clustering analyses were performed with the Partek Genomic Suite using $\log_2$ (read count per million mapped reads $\times 10^6$ [RPM]) expression values for the curated sncRNA annotations as input; only RNA species (N = 303) with mean normalized count values ≥5 were selected to produce reliable differential expression profiles.

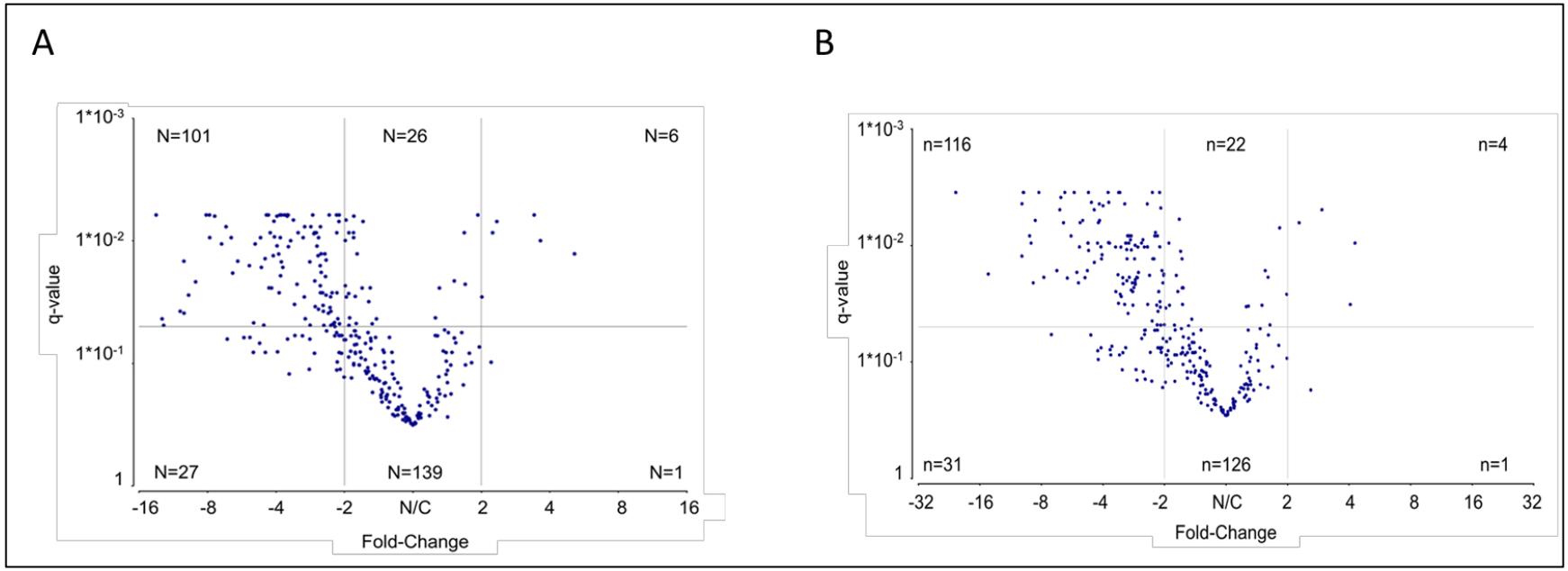


Figure 3.8. AML with splicing factor mutations have reduced expression of select snoRNAs. Volcano plot of the significance level of the expression difference between snoRNAs in samples with A) *SF3B1, SRSF2*, and *U2AF1* and B) *U2AF1* splicing mutations versus CD34. For A) small RNA-seq was performed on 16 AML samples with mutations in three splicing factors (*SRSF2* - P95L (2), P95R (1); *SF3B1* - E592K, Q19E (1), K700E (1); *U2AF1* - Q157P (1), Q157P ASXL1 (2), R156H (1), S34F (5), S34Y (1), and Q157R (1)) and CD34 from 6 healthy donors (Table 3.1 and Figure 3.1). For B analysis was between splicing factor mutation *U2AF1* - Q157P (1), Q157P ASXL1 (2), R156H (1), S34F (5), S34Y (1), and Q157R (1)) and CD34. Each dot represents one snoRNA. The horizontal line on the plot represents the P-level used for this analysis (0.05). Vertical lines represent the threshold equivalent to a twofold difference. The y-axis reflects the q value–corrected P value. Dots in the upper right and left quadrants represent differentially expressed snoRNAs which have absolute fold change value >2 and $P \leq 0.05$. Differential expression and hierarchical clustering analyses were performed with the Partek Genomic Suite using $\log_2$ (read count per million mapped reads × 10^6 [RPM]) expression values for the curated sncRNA annotations as input; only RNA species (N = 307) with mean normalized count values ≥5 were selected to produce reliable differential expression profiles.

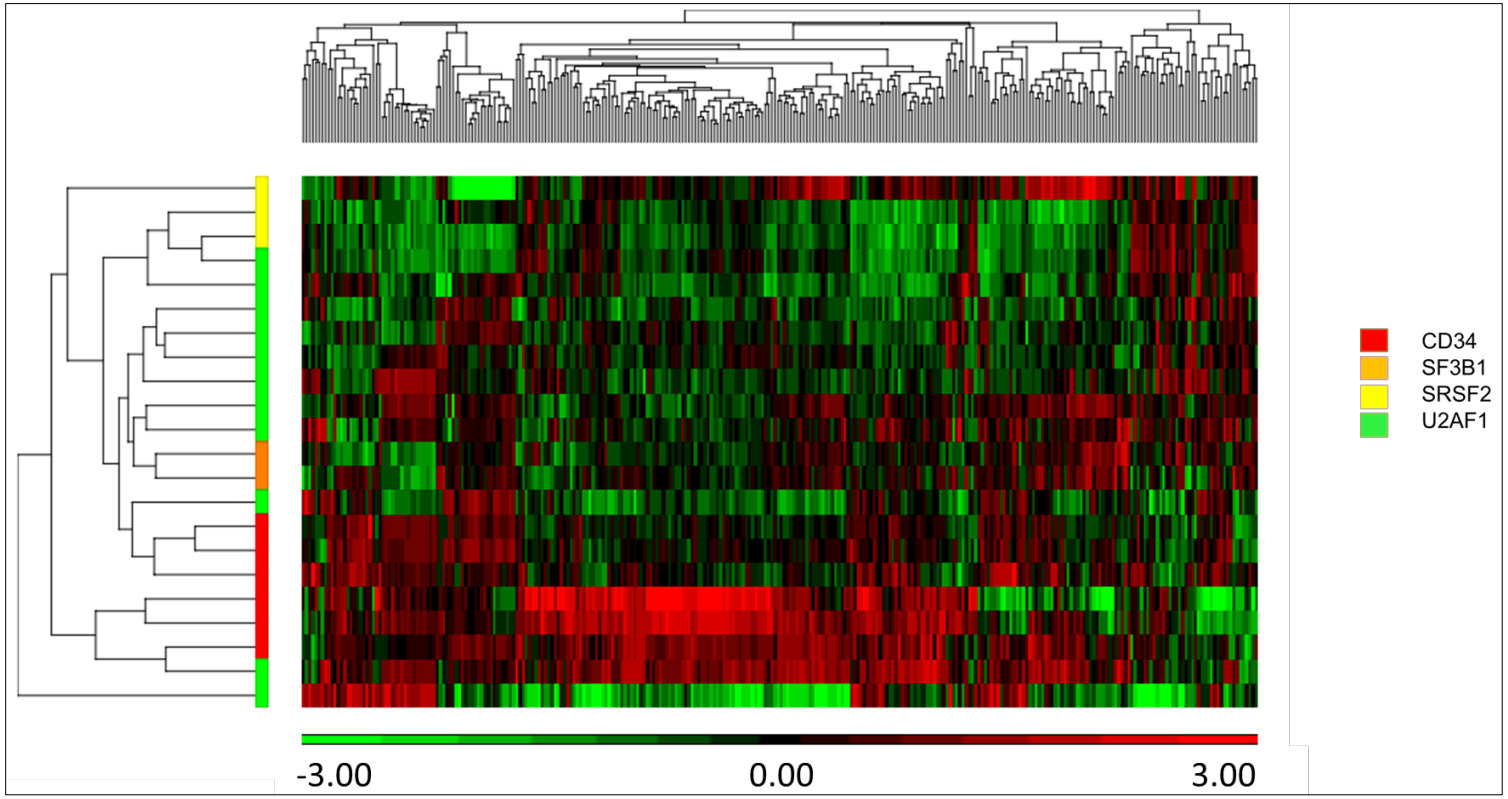


Figure 3.9. Unsupervised hierarchical clustering of snoRNA expression for AML with mutations in splicing factors. Small RNA-seq was performed on the indicated samples with mutations in splicing factors. snoRNA expression (z-scored $\log_2$ RPM) is shown, with red indicating high expression, and green indicating low expression. Each column represents a unique snoRNA, and each row represents a sample.

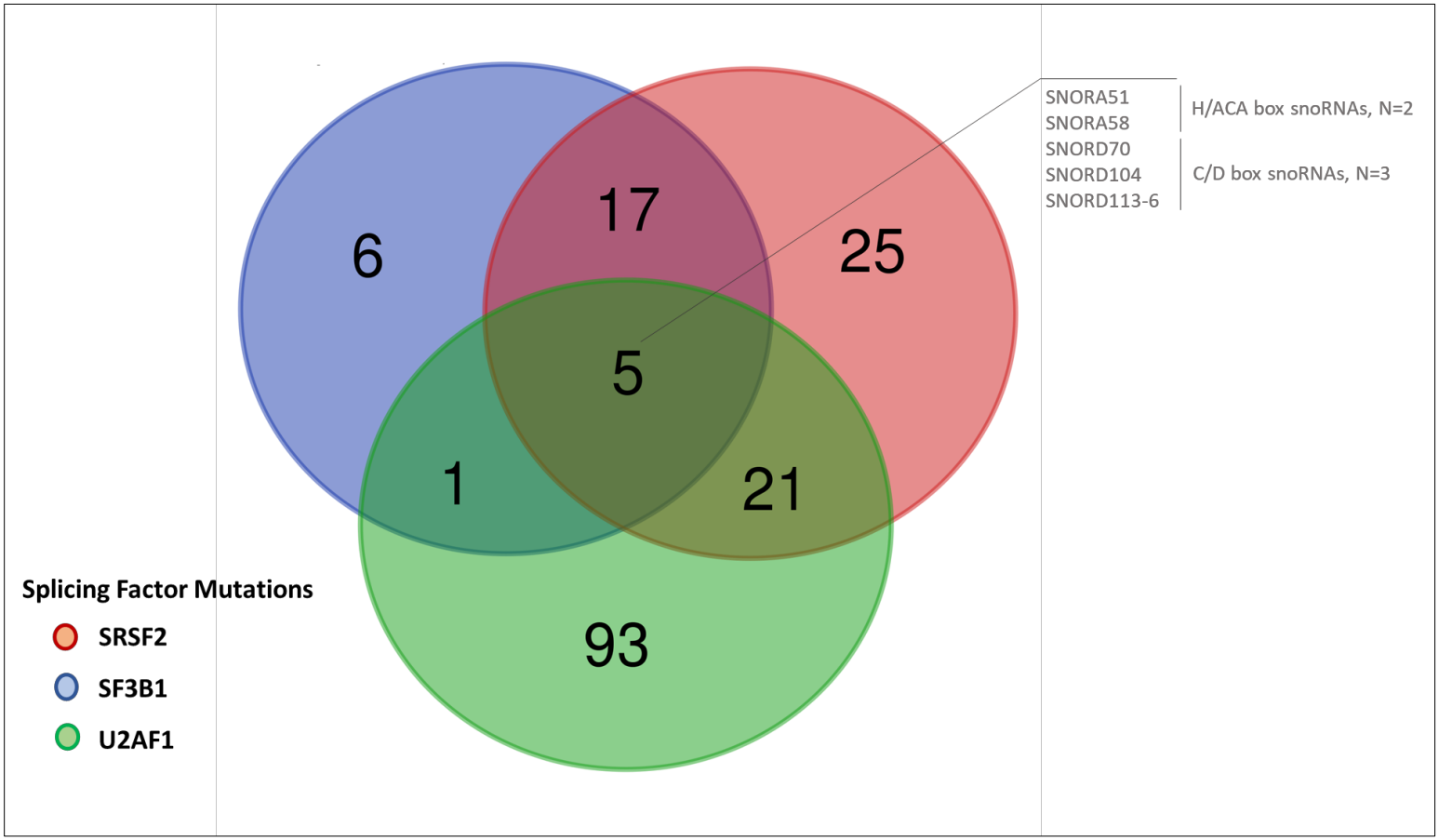


Figure 3.10. Venn diagram of differentially expressed snoRNAs in AML samples with mutations in the indicated splicing factors versus healthy donors. Venn diagram shows the intersection of the numbers of the differentially expressed snoRNAs in AML samples characterized by mutations in splicing factors versus CD34 from six healthy donors. The numbers in the intersections represent the absolute numbers of the comprised differentially expressed overlapping/unique snoRNAs.

Table 3.1 AML sample characteristics.

AML Subtype	Number analyzed	Genetic Alteration(s)
MLL translocations	12	*MLL-ELL (3), MLL-MLLT3 (2), MLL-MLLT4 (2), MLL-MLLT10 (2), MLLT10-MLL, MLLT10-CEP164 (1), MLLT10-MLL, MLLT10-PPP2R1B (1), MLLT10-PICALM (1)*
Core Binding Factor	15	*MYH11-CBFB (10), MYH11-CBFB, GPR128-TFG (1), RUNX1-RUNX1T1* (4)
PML-RARa	14	*t(15;17)*
Spliceosome	16	*SF3B1 E592K, Q19E (1), K700E (1)* *SRSF2 P95L (2), P95R (1)* *U2AF1 Q157P (1), Q157P (2), R156H (1), S34F (5), S34Y (1), Q157R (1)*
Healthy donors	6	

Table 3.2 AML Summary sequencing data

Genomic Alteration	Number of expressed snoRNAs	Number of differentially expressed snoRNA		DE snoRNAs by class, %			DE snoRNA from imprinted loci		snoRNAs that modify critical ribosomal regions	
		Suppressed expression	Increased expression	C/D Box	H/ACA Box	scaRNA	DLK-DIO3	SNURF-SNRPN	PTC	ISB
PML-RARa fusions (N = 15)	315	70 C/D-51, H/ACA-18, scaRNA-1	52 C/D-34, H/ACA-15, scaRNA-3	69.11	26.83	3.25	26	20	*SNORA71A, SNORA1, SNORD49B, SNORD29, SNORD1A*	*SNORD88A, SNORD36C, SNORD36A*
RUNX1-RUNX1T1 fusions (N = 4)	302	34 C/D-28, H/ACA-3, scaRNA-3	9 C/D-5, H/ACA-3, scaRNA-1	76.74	13.95	9.3	16	0	*SNORD1A*	
Core Binding Factor fusions (N = 15)	303	92 C/D-59, H/ACA-26, scaRNA-7	17 C/D-6, H/ACA-8, scaRNA-3	59.63	31.19	9.17	17	0	*SNORA1, SNORA1A*	*SNORD47, SNORD34, SNORD36A*
MLL fusions (N = 11)	304	72 C/D-44, H/ACA-23, scaRNA-5	12 C/D-6, H/ACA-4, scaRNA-2	57.95	35.23	6.82	17	1	*SNORD1A, SNORD29, SNORA62, SNORA 53, SNORA10, SNORA1*	*SNORD36A, SNORD36C*
Mutations in splicing factors (N =16)	307	101 C/D-62, H/ACA-33, scaRNA-6	6 C/D-4, scaRNA-2	61.68	30.84	7.48	8	6	*SNORA1, SNORD29, SNORA1, SNORD29*	*SNORD36C, SNORD34*

Table 3.3. Summary of the number of differentially expressed observed across all possible comparisons. Samples included *PML-RARa*, *CBF* fusions, *MLL* fusions and AML with mutations in splicing factors. (P-value ≤ 0.05 and absolute fold-change ≥ 2.00)

Names	Number of	Elements
CBF MLL PML SFM	36	*SNORD18B SNORD114-22 SNORA72 SNORD118 SNORD83B SNORD38B SNORD19B SNORA28 SNORD1A SNORA1 SNORD113-8 SNORA40 SNORA63 SNORD105 snoID_0681 SNORD36C SNORD10 SNORA69 SNORA46 SNORD114-12 SNORA20 SNORD113-6 SNORA26 SNORD114-9 SNORD114-21 SNORA51 SCARNA7 SNORD29 snoID_0757 SNORD28 SNORD114-1 SNORD91B SNORD114-3 SCARNA14 SNORD5 SNORA13*
CBF MLL PML	16	*SNORA6 SNORD36A snoID_0378 SNORD114-14 SNORD113-5 SNORA34 SNORA10 SNORD113-9 snoID_0758 SNORD114-26 SNORD113-7 SNORD114-23 SNORD114-17 SNORA62 SNORD114-28 SNORD114-25*
CBF PML SFM	13	*SNORD11B SNORD72 SNORD4B SNORA96 SNORA4 SNORD34 SNORD77 SNORD14B SNORA33 SNORA79 SNORD18C SNORD73A SNORD12B*
CBF MLL SFM	14	*SCARNA8 SNORA58 SNORD56 SCARNA16 SNORD22 SNORA31 SNORD91A SCARNA4 SNORA9 SNORD104 snoID_0409 SNORD138 14_70235681_70235822_+ SNORA75*
MLL PML SFM	2	*SNORD86 SNORD116-20*
CBF PML	9	*SNORD105B SNORD17 SNORA73B SNORA57 SNORA71A SNORD94 SCARNA20 SNORA71D SNORD76*
CBF MLL	4	*SNORA58B SCARNA1 SCARNA9 SNORD90*
CBF SFM	8	*SNORA36B SCARNA23 SNORD80 SNORD4A SCARNA9L SNORA97 SNORD70 SNORA32*
MLL PML	3	*SNORD14C SNORA77 SNORD145*
PML SFM	5	*SNORD116-27 SNORD116-12 SNORD116-14 SNORD116-16 SNORD116-15*
MLL SFM	4	*SNORD12 SNORA66 SNORA80 SNORD88B*
CBF	9	*SNORD97 SNORD117 SNORD47 SNORA76 SNORD84 snoID_0308 SNORD63 SNORD15A SNORD54*

PML	38	*SNORD88A snoID_0810 SNORD116-11 SNORD114-20 SNORA2A SNORA45 SNORD116-21 SNORD116-1 SNORD116-8 SNORD116-29 SNORD114-11 SNORD64 SNORA74B SNORA5C SNORD116-25 SNORD112 SNORD113.6-201 SNORD107 snoI_0409 SNORD113-4 SNORD116-24 SNORD89 SNORD25 SNORD116-2 SNORD116-13 SCARNA10 SNORD116-22 SNORD116-6 SNORD113-3 SNORD116-26 SNORA11F SNORA7B SNORD116-23 SNORD114-15 SNORD49B SNORD114-10 SNORD114-13 snord138*
MLL	5	*2_217366063_217366174_+ SNORD49A SNORD41 SNORA96(revised) SNORA53*
SFM	24	*SNORD24 SNORA49 snoID_0386 SCARNA11 SNORD124 SNORD136 SNORA71B SNORA67 SNORD45C SNORD14E SNORD53 SNORD11 SNORD121A SNORD51 SNORD48 snoID_0699 SNORA65 SNORD65 SNORD98 SNORD12C SNORA101B SNORA17 SNORD100 SNORA77B*

Table 3.4. **Summary of the number of differentially expressed observed across all possible comparisons.** The splicing factor mutations are *SRSF2, SF3B1*, and *U2AF1*. Absolute fold change value >2 and $P \leq 0.05$

Names	total	elements
SF3B1 SRSF2 U2AF1	5	*SNORA58 SNORD113-6 SNORD104 SNORA51 SNORD70*
SF3B1 SRSF2	17	*SNORD114-22 SNORD114-14 SNORD113-5 SNORD113-9 SNORD113-8 SNORD114-26 SNORD114-12 SNORD46 SNORD113-7 SNORD114-23 SNORD114-9 SCARNA9 SNORD114-21 SNORD114-1 SNORD114-28 SNORD114-3 SNORD114-25*
SF3B1 U2AF1	1	*SNORD76*
SRSF2 U2AF1	21	*SNORD24 SNORA72 SNORD136 SNORD83B SNORD118 SNORD4B SNORD1A SNORA40 SNORA96 SNORD53 snoID_0681 SNORA46 SNORD48 SNORD116-16 snoID_0409 SCARNA9L SNORD29 SNORD116-15 14_70235681_70235822_+ SCARNA14 SNORA13*
SF3B1	6	*SNORA74A SNORD16 SNORA73B SNORD63B SCARNA28 SNORD15A*
SRSF2	25	*SNORA42 SNORD116-11 SNORD116-21 snoID_0378 SNORD1C SNORD116-1 SNORD64 SNORD116-25 SNORD92 SNORD116-27 SNORD116-12 SNORD116-24 SNORD89 SNORD116-13 SCARNA2 SNORD114-17 SCARNA7 SNORD116-14 SNORD19C SNORD116-22 SNORD50A SNORD116-23 U8.4-201 SNORD102 SNORD116-20*
U2AF1	93	*SNORD14A SNORA58B SCARNA8 SNORD67 SNORA49 snoID_0386 SNORA8 SNORD11B SCARNA11 SNORD32A SNORD72 SNORD18B SNORA12 SNORD124 SNORA36B SNORA66 snoID_0728 SNORD45A SNORD56 SNORD38B SNORD19B SNORA71B SNORD14C SNORA28 SCARNA1 SNORA1 SNORA67 SCARNA23 SNORD45C SNORA74D SNORA44 SCARNA16 SNORD14E SNORD22 SNORA63 SNORD105 SNORA31 SNORD91A SNORD36C SNORD10 SCARNA4 SNORA4 SNORD11 SNORA50 SNORD34 SNORD146 SNORA69 SNORA9 SNORD121A SNORD51 SNORA20 SNORD36B SNORD80 SNORD25 SNORA18 SNORD77 SNORD14B SNORA53 SNORD86 SNORA33 SNORA26 SNORD43 SNORD4A SNORA24 SNORA65 SNORD65 SNORD70B SNORA79 SNORD98 SNORD54 SNORD18C SNORD8 SNORD12C snoID_0757 SNORA68 SNORA25 SNORD28 SNORA73A SNORA101B SNORD138 SNORA75 SNORA97 SNORD91B SNORA17 SNORD5 SNORD73A snoID_0680 SNORD100 SNORA14B SNORA77B SNORA21 SNORD12B SNORA32*

CHAPTER 4

ROLE OF SNORA21 in LEUKEMIA

INTRODUCTION

Mature blood cells are produced through a series of stepwise hierarchical progression of differentiation where hematopoietic stem cells and increasingly committed progenitors progress unidirectionally to mature circulating blood cells[307-309]. Considerable understanding has been acquired into lineage commitment at the transcriptional level[310,311], yet there remains much uncertainty about the molecular control of self-renewal, lineage determination and clonal integrity. The rate of protein synthesis varies considerably during hematopoietic differentiation[312-314]. Signer et al., reported that precise regulation of protein synthesis rates is required for hematopoietic stem cell (HSC) function[312]. Moreover, increased protein synthesis has been implicated in the development of certain cancers, including hematopoietic malignancies[315-317]. Recent human genetics, biochemical and proteomic studies have shown that ribosomes play a critical role in promoting lineage commitment from HSPCs[318,319]. Indeed, mutations that affect ribosome function are associated with impaired hematopoiesis[177,178]. Taken together, these observations highlight the importance of translational control in the regulation of normal and malignant hematopoiesis.

There is an emerging consensus that regulation at the level of the ribosome through modification of rRNAs play an important role in ribosome biogenesis, translational control, and the generations of specialized ribosomes[320]. rRNAs are highly modified, containing over 110 modifications, the most common of which are base pseudouridylation and ribose 2'-O-methylation mediated by snoRNAs[320]. On a molecular level, pseudouridylation leads to increased rotational freedom about the glycosyl bond and the creation of an extra hydrogen bond donor[118,321,322]. These changes have the effect of increasing base stacking and stabilizing double stranded base pairing interactions critical

for inter- and intramolecular binding properties in evolutionarily conserved regions[118,322]. One such region where these modifications are concentrated is the peptidyl transferase center (PTC) which is the ribosomal catalytic site where peptide bonds are made during protein elongation and peptidyl-tRNAs are hydrolyzed during the termination of protein synthesis[286]. Studies in yeast, suggest that loss of pseudouridylation or 2'-O methylation at individual and clusters of PTC sites results in reading frame changes, altered transfer RNA (tRNA) selection, increased stop codon read through, amino acid incorporation defects, altered polysome patterns, and reduced translational efficiency[118,128,288].

The importance of pseudouridylation in human hematopoiesis is illustrated by dyskeratosis congenita, a congenital bone marrow failure syndrome. The X-linked form of dyskeratosis congenita is due to loss of function mutations of *DKC1* which is a pseudouridine synthase that along with H/ACA snoRNAs carries out pseudouridylation. Loss of *DKC1* in murine cells is associated with reduced levels of pseudouridine in rRNA, impaired ribosome biogenesis, and reduced translation of genes with an internal ribosome entry site (IRES) motif in their 5'-untranslated region[125,228,323,324]. Patients with *DKC1* mutations also have impaired telomere maintenance, which likely contributes to the impaired hematopoiesis. Although there is strong evidence in mice that *DKC1* regulates hematopoiesis in a non-telomere dependent fashion, this is not proven in humans. Interestingly cancer incidence in patients with DC is 11 times higher than in the general population[320].

snoRNAs are one of the largest classes of noncoding RNAs in eukaryotes. They guide the site-specific 2′-*O*-ribose methylation (C/D box snoRNAs) and pseudouridylation of rRNAs (H/ACA box snoRNAs) via base complementarity and assist in the cleavage of

pre-rRNAs thus playing an essential role in ribosome biogenesis[3,240,298,325]. Originally thought of as simply playing a housekeeping role, they have now been implicated in alternative splicing,[254] regulation of chromatin structure,[255] metabolism,[168] and neoplastic transformation[204,326]. Specifically, numerous reports have recently implicated snoRNAs[99,102,192,327] in acute myeloid leukemia. For example, a recent paper looking at AML-ETO fusion AMLs reported that C/D box snoRNAs and rRNA 2'-O-methylation play a critical role in leukemic stem cell activity[102].

Our group recently profiled the snoRNA expression in de novo AML with normal karyotype[192]. We reported that a subset of snoRNAs is expressed in a developmental- and lineage-specific manner during human hematopoiesis[192]. Expression of most snoRNAs in AML was comparable to that observed in CD34 cells from healthy, consistent with a housekeeping function for snoRNAs. However, *SNORA21* expression was modestly, but significantly reduced in normal karyotype AML, and its expression varied more than 10-fold between AML samples. *SNORA21* modifies two positions with the PTC of the 28S rRNA (U4401 and U4470)[328,329]. Deletion of *snR10*, the orthologue of *SNORA21* in yeast[26,330] results in the complete loss of pseudouridylation at the orthologous ribosome position, as well as altered ribosome structure, biogenesis and activity[118]. The contribution of *SNORA21* in the regulation of ribosome function in human cells is largely unknown. However, increased *SNORA21* expression is associated with poor prognosis in patients with gastric, colorectal, or lung cancer.

Here we show that reduced *SNORA21* expression is reduced, but highly variable, in multiple genetic subtypes of AML. Loss of *SNORA21* in K562 cells is associated with impaired ribosome biogenesis and reduced global translation. It also induces a non-

apoptotic cell death characterized by impaired mitochondrial function. These observations suggest that alterations in *SNORA21* expression in human AML may contribute to translational efficiency by regulating ribosome function.

METHODS

Primary cell culture

K562 human chronic myelogenous leukemia (CML) cells were obtained from ATCC and grown in Iscove's Modified Dulbecco's Medium (ThermoFisher) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Atlas Biologicals, Larimer Colorado) and 1% (v/v) penicillin-streptomycin (Life Technologies, Inc.). Cells were maintained at a density of 1.0 - 1.5 x 10^6 cells/mL at a viability of 94 - 97% in a humidified atmosphere at 37 °C in 5% CO_2.

Generation of Cas9-expressing clonally derived cells

K562 cells stably expressing a constitutive rtTa were transduced with a lentivirus that expresses spCas9 under tetracycline-controlled promoter, where Cas9 expression is also linked to mCherry via an internal ribosome entry site (IRES). Once K562 cells were transduced, Cas9 expression was induced by doxycycline (500ng/mL) for 48h and clonally derived cells were established by sorting GFP+ cells into a well of a 96-well place using fluorescence-activated cell sorting (FACS) buffer. These cells were subsequently used to delete SNORA21.

Deletion of SNORD21 in K562 cells

Using the web interface of UCSC Genome Browser's CRISPR design software (https://genome.ucsc.edu/), multiple small guide RNAs (gRNAs) targeting spanning SNARDA21 gene. Offtarget activity of these gRNAs were further evaluated with Blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Briefly, two complementary oligos were

designed to generate double stranded gRNA with flanking BsmBI restriction sites. For example, one of the sequences of human sgRNAs targeting the 5' end of SNORD21 (5' -> 3') is ATTTTGAAGAAGGGCCGTCG. Briefly, two complementary oligos were designed to generate double stranded gRNA with BsmBI restriction sites (5' -> 3'): CACCG-ATTTTGAAGAAGGGCCGTCG and AAAC-CGACGGCCCTTCTTCAAAAT-C. These oligos were self-annealed and cloned into BsmBI digested LRB plasmid (Lenti_sgRNA_EFS_GFP; Addgene #65656). Clonally derived K562 cells with inducible Cas9 were treated with doxycycline (500ng/mL) for 48h, co-nucleofected LRG vector expressing aforementioned sgRNAs. 48-hours post-transduction, stable cell lines expressing both Cas9-GFP and sgRNA-BFP were established by sorting GFP+ BFP+ cells by FACS. Upon expansion of these cells, a PCR-based approach was performed to validate successful deletion of the target region.

Experimental validation of CRISPR/Cas9

Genomic DNA was extracted from the SNORA21$^{-/-}$ and wild type K562 using the PureLink Genomic DNA Kit, (Invitrogen) according to the manufacturer's instruction. A pair of primers (Primers F1: 5'..TGAGCTCTGGTTTGCCTTTCTTG..3'
R1: 5'..GTGAATTACAGCTTGAGTATGAG..3') that amplify a region spanning the gene deletion area (584nt region, wild type cells; 435nt, SNORA21$^{-/-}$ K562 cells) were designed using NCBI primer design tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). For PCR, 5 µl of GoTaq green master-mix (Promega, Madison, WI) and 0.2 µL of each of the forward and reverse primers (20 µM final conc; Integrated DNA Technologies) were added per PCR tube (0.2 mL; ThermoFisher). Finally, 2 µl of

genomic DNA sample and 3 µl of nuclease-free water were added to each tube for a final reaction volume of 10 µl. PCR was performed using a standard thermocycler (Bio-Rad T100 Thermal Cycler). The PCR conditions were as follows: Initial denaturation (95 °C for 2 minutes), followed by 35 cycles of denaturation (95°C for 30 secs), annealing (56 °C for 30 secs) and extension (72°C for 45 secs) and a final extension at 72°C for 5 minutes. PCR products were removed from the thermocycler and maintained at room temperature for a few minutes before loading on the 2% agarose gel. 7 µl of each PCR sample was loaded into each well and electrophoresis was performed for 1 hour and 20 minutes at 100 volts. The power supply (Thermo Scientific EC Electrophoresis Dual Power Supply EC135-90) was set to 100 V. A DNA size marker (100 bp DNA even ladder, EZ BioResearch, St. Louis, MO) was used. Gel images were acquired using a regular gel-documentation system (ChemiDoc MP Imaging System, Bio-Rad, Hercules, CA). Separately, *SNORA21* deletion *was* detected by PCR-based 2× bidirectional direct Sanger sequencing using the primers specified above. The sequencing results were interpreted using SerialCloner v6-3-1 software.

Cell proliferation assays

Cells were plated in triplicate in 6-well plates (100,000 cells/well) with growth medium. Cells were counted in triplicate at the indicated time points with a Bürker chamber cell chamber after 0.2% Trypan Blue staining to exclude apoptotic/necrotic cells. Separately, the cells were counted on a Cellometer (Nexcelom Bioscience, Lawrence, MA) and plated at 1 000 cells per well in 150 µL of media on a 96-well plate in triplicate. Each well was treated with 1 µM of IncuCyte® NucLight Rapid Red Reagent (Essen Biosciences, Ann

Arbor, MI). With an Incucyte live cell imager (Essen Biosciences), photomicrographs were recorded every 12 hours for a total of 156 hours. At each time point, each culture was scanned 5 times at a 10x objective. For each culture at each time point, the Incucyte software was used to measure the mean number of red-positive cells from the 5 non-overlapping images, and these counts were taken as the number of live cells in each culture. The counts for the cultures at the 0-hour time point were averaged to yield a baseline count for the starting point. All counts were normalized to the baseline and presented as a fold change from the baseline.

Caspase-3/7 activity assay

The caspase-3/7 activity in the cell extracts was determined using Caspase-Glo 3/7 Assay (Promega) according to the manufacturer's instructions[98]. In brief, K562 cells were seeded in the inner wells of a white, 96-well plate (Corning Costar, Corning, NY) at 10,000 cells per well (100µl) in replicates of three and incubated at 37°C with 5% CO_2. An equal volume of the Caspase-Glo® 3/7 Reagent was added directly to the cells in 96-well plates with brief agitation and incubated at 25°C for 1 hour. Caspase-Glo® 3/7 luminescent signal was read on the Xenogen IVIS 50 Imaging System (Caliper Life Sciences, Waltham, MA). Background fluorescence signals (media-only wells) was subtracted from all values, and averages and standard deviations were calculated from the three replicates. Graphs were created using GraphPad Prism® 6 Software.

Flow cytometric analysis of apoptosis.

Cells were washed once in PBS and resuspended in Annexin binding buffer (Biolegend) and stained with FITC-conjugated Annexin V and PI (Biolegend) per the manufacturer's instructions. After staining, cells were washed with Annexin binding buffer, resuspended in a fixing solution and stored at 4°C in the dark. Flow cytometry was performed on a FACScan (Beckman Coulter). All flow cytometry files were analyzed with FlowJo software (Tree Star, OR).

Cell cycle analysis

Cells were counted by trypan blue exclusion and resuspended to a concentration of 1 × 10^6 cells/ml in FACS buffer. They were then washed in ice-cold PBS followed by fixation and permeabilization using the Cytofix/Cytoperm Fixation Permeabilization Kit (BD Biosciences) followed by incubation at room temperature for 10 min in 1 x BD Perm/Wash buffer supplemented with 5% goat serum. Alexa Fluor® 700 anti-human Ki-67 Antibody (BioLegend) was then added and allowed to incubate on ice for 30 min followed by washing with 1x BD Perm/Wash buffer and resuspension in FACS buffer. Cells were stained for DNA content by FxCycle™ Violet DNA dye (Invitrogen) according to the manufacturer's protocol. Flow cytometry was performed on a FACScan. All flow cytometry files were analyzed with FlowJo software.

Proliferation measurement via EdU incorporation

A Click-iT EdU (5-ethynyl-2′-deoxyuridine) cytometry assay kit (Invitrogen) was used for the labeling and detection of newly synthesized DNA. The labeling depended on the

incorporation of the thymidine analog EdU followed by a copper catalyzed cycloaddition reaction analog with an azide-conjugated fluorophore for flow cytometry detection. The assay followed the manufacturer's protocol with some modification based on optimization for K562 cells. In brief, 1×10^6 WT and SNORA21$^{-/-}$ K562 cells were fixed and permeabilized 2 hours post-EdU administration. A copper catalyzed cycloaddition reaction between an alkyne of EDU and a dye labeled azide occurred followed by flow cytometric analysis. The background staining of K562 cells not incubated with EdU was used as a negative control[331,332].

Measurement of global protein synthesis

Ribosome activity was determined using puromycin labeling of nascent peptides that were detected at the single cell level[312,333]. After optimization, metabolic labeling assays were performed using $1x10^6$ K562 cells per assay. Cells were collected by centrifugation and washed in ice-cold phosphate-buffered saline (PBS). They were incubated in OPP, 20 mM (Click-iT™ Plus OPP Alexa Fluor™ 488 Protein Synthesis Assay Kit, Life Technologies) for 30 minutes. They were then washed in ice-cold Ca^{2+}- and Mg^{2+}-free PBS followed by fixation in cell fixation and permeabilization solution (BD Biosciences) for 20 minutes followed by a BD Perm/Wash™ Buffer (1x) wash (BD Biosciences). The azide–alkyne click chemistry cycloaddition was performed using the Click-iT Cell Reaction Buffer Kit (Thermo Scientific) as described in the manufacturer's instructions with the azide conjugated to Alexa Fluor 488. After 30 min, samples were washed with Click-iT® Reaction Rinse Buffer then with PBS/3% BSA and resuspended in fluorescence-activated cell sorting buffer before data acquisition using a FACS flow

cytometer (BD Biosciences). Data analysis was performed using FlowJo v10.1. To determine if protein synthesis is required for OPP incorporation into nascent polypeptides, cells were incubated for 30 mins at 37°C in a humidified 5% CO_2 atmosphere with culture medium supplemented with 100ug/ml cycloheximide (CHX). This concentration of CHX was determined *a priori* to essentially block protein translation in K562 cells. All assays were performed in triplicate. Data was expressed as mean signal intensity. A one-way analysis of variance (ANOVA, $p<0.001$) was applied to the data.

Ribosome Profiling

Equal amounts of total RNA from the supernatant was layered over a 10%–50% sucrose gradient (20 mM Tris-HCl pH 7.5; 140 mM KCl (Invitrogen/Thermo Fisher); 5 mM $MgCl_2$; 0.5 mM DTT) and centrifuged at 37,000 rpm for 160 min at 4ºC in a SW41Ti rotor (Beckman Coulter). Gradients were displaced from the bottom up with a BR-188 Density Gradient Fractionation System (Brandel, Gaithersburg, MD) by a chase solution (60% sucrose (IBI Scientific, Dubuque, IA) with bromophenol blue (Sigma, St. Louis, MO)) at a rate of 1.50 $ml.min^{-1}$ combined with continuous absorbance reading at A_{254} nm. Traces were recorded at a sensitivity setting of 2. Area under the curve was calculated using GraphPad Prism. Statistical significance was calculated with Student's t-test. P-values < 0.05 were considered significant.

RiboSeq

The ribosome profiling strategy was adapted from Mc.Glincy and Ingolia[334] and based on De Klerk et al.,[335] with some modifications. Supernatant from the cleared lysate

corresponding to 250ug of total RNA were incubated with 300 Units of RNase T1 (Thermo Scientific, Waltham, MA) and 500ng of RNase A (Invitrogen/Thermo Fisher) for 15 minutes at room temperature with gentle shaking at 350 rpm. Digestion was blocked with SuperaseIN (Ambion/Thermo Fisher) and the lysate was layered on 600ul of a 1M sucrose cushion. The sample was ultracentrifuged in an Optima Max-XP ultracentrifuge (Beckman Coulter) equipped with a MLA-130 rotor at 4ºC for 4 hours at 70,000 rpm. The pellet was resuspended in 700µl of TRizol reagent (Invitrogen/Thermo Fisher) and RNA extracted using the Direct-zol RNA MiniPrep (Zymo) per the manufacturer's instructions. Cytoplasmic and mitochondrial rRNAs were removed using NEBNext® rRNA Depletion Kit (NEBL) according to manufacturer's instructions, with modifications. RNA Clean and Concentrator-5 (Zymo) was used for cleanup post ribo depletion. Purified RNA was 3' end dephosphorylated using 10U of T4 PNK (NEBL) at 37ºC for 3 hours followed by cleanup with RNA Clean and Concentrator-5 (Zymo). The yield and quality were assessed using the Qubit microRNA assay (ThermoFisher) and the Agilent 2200 TapeStation system (Agilent Genomics).

Library preparation were carried out as previously described by Warner et al., 2018. Libraries were sequenced on an Illumina NextSeq 500 high output flowcell as a 2x40 by the DNA Sequencing Innovation Lab at the Center for Genome Sciences and Systems Biology, Washington University St. Louis, MO. Raw FASTQ files were demultiplexed using the Illumina bcl2fastq2 script. Illumina adapters were filtered using fqtrim. Reads were aligned to the human genome (hg19) using TopHat2, version 2.0.11 using a guided alignment strategy using Ensembl gene annotations (GRCh37 v82)[336]. TopHat2 has previously been shown to be an effective aligner for short RNA

sequencing libraries[334]. Mapped reads were assigned to Ensembl gene annotations and duplicate reads removed using featureCounts.

Transcriptome sequencing

Trizol extracted RNA was purified on RNA Clean & Concentrator-5 (Zymo) using the manufacturer's protocol. The Qubit RNA HS Assay Kit (Life Technologies) and the Agilent 2200 TapeStation system were used for quantification and quality assessment, respectively, according to the manufacturer's instructions. Library preparation was carried out using the KAPA Biosystems RNA Hyper Prep with Riboerase (KAPA Chicago, IL) according to manufacturer's recommendation. Cycle numbers were optimized using qPCR. NGS sequencing performed on an Illumina NextSeq 500 as a 2x150 configuration by the DNA Sequencing Innovation Lab at the Center for Genome Sciences and Systems Biology. Illumina adapters were filtered using fqtrim. Processed reads were aligned to the human genome (hg19) using HiSat2, Version 2.1.0[337]. Mapped reads were assigned to Ensembl gene annotations (GRCh37 v82) and duplicate reads removed using featureCounts. Separately, data analyses (cluster, principle component analyses, biofunctional analysis, differential expression) was carried out using the Partek Flow. Reads were mapped to the National Center for Biotechnology Information Build 37 human reference sequence by STAR (v. 2.5.3a) and normalized FPKM values were determined using Partek Genomic Suite. The Partek gene specific analysis (GSA) algorithm was utilized for differential expression analyses. Differentially expressed genes were selected at ≥2-fold difference (comparing wildtype and snorna21$^{-/-}$ K562 samples) and P< 5.00E-2. Results were validated with RT-PCR. We searched for genes with significantly

increased or decreased translational efficiency[338] (TE) (mRNA abundance vs ribosome occupancy) using RiboDiff[339]. We selected genes significant at an adjusted P< 5.00E-2 and showing $\log_2(TE_{SNORA21-/-}/TE_{WT})$ higher or lower than 0.00.

Analysis of pseudouridylation

The 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-p-toluenesulfonate (CMCT) treated samples were essentially prepared as previously described[51,119,120,328,340-345] with some exceptions. Incubation time, temperature, pH and the ratio of amounts of CMCT to sample were optimized. Ten µg of RNA was resuspended in a solution of freshly dissolved made CMCT in Bicine buffer (50 mM bicine pH 8.3, (Hampton Research, Aliso Viejo, CA), 4 mM EDTA (Invitrogen), and 7 M urea (Sigma) for a final concentration of 0.2 M CMCT. The mixture was incubated at 37°C for 20 minutes. The CMCT-RNA was ethanol purified and resuspended in 40 µL of 50 mM sodium bicarbonate (Sigma) and incubated at 37°C for 4 hours. Following ethanol purification, reverse transcription with SuperScript III RT (Thermo Scientific,) was conducted on 1 µg RNA in 8mM $MgCl_2$, 50mM Tris pH 8.3, 75 mM KCl, 10mM DTT, 0.5mM dNTPs and primers (RT primer 28S), IDT), at 42ºC for 3 hrs in a total volume of 10 µL. The samples were incubated at 90°C for 5 minutes to end the reverse transcription reaction. The cDNA was amplified in a 10 µL reaction (EconoTaq Plus Green 2X Master Mixes, Lucigen, Middleton, WI), primers, template) for 20 cycles. The indexes were added in a 20 µL reaction (EconoTaq Plus Green 2X Master Mixes, primers, template) with 8 cycles. The DNA was separated on a gel containing 2.0% agarose (MidSci) gel with 6X loading dye (NEBL). After gel electrophoresis, primer extension was visualized by ultraviolet light. The region of interest

of each lane was excised, and the DNA extracted using Zymoclean Gel DNA Recovery Kit (Zymo). The library was checked by Agilent 2200 TapeStation system and then sequenced in a MiSeq (Illumina) at the DNA Sequencing Innovation Lab at the Center for Genome Sciences and Systems Biology. Partek Flow was used for bioinformatic analysis.

Electron Microscopy

For an autophagy positive control, cells were amino acid starved by extensively washing with Hank's Balanced Salt Solution (HBSS) (Gibco) and further incubated in HBSS for 3 and 6 hrs as described elsewhere[346,347]. Cell suspensions were fixed in 2.5% glutaraldehyde (Electron Microscopy Sciences (EMS), Hatfield, PA), 2% paraformaldehyde (EMS) in 0.15 M cacodylate buffer: sodium cacodylate (EMS) with 4 nM $CaCl_2$ (EMS). Following several buffer rinses, cells were mixed with low temperature gelling agarose (MP Biomedicals, LLC, Solon, OH) and pelleted. The cell/agarose pellets were trimmed into 1-2 mm blocks for the remaining preparation steps. Cell blocks were then post fixed in a 1% OsO_4 (EMS)/1.5% FeKCN mixture (EMS) for 1 hr. in the dark, en block stained in 2% uranyl acetate (EMS) for 2 hrs and subsequently dehydrated in 100% ethanol (EMS) (3X 10min each), infiltrated in 50%, and 100% (3x) LX112 Resin (EMS) (microwave assisted), and cured at 60°C for 48 hrs. Sections were cut using a diamond knife on a Leica EM UCT7 ultramicrotome (Leica Microsystems GmbH, Wetzlar, Germany) then post stained in 2% uranyl acetate followed by Reynolds lead citrate (lead nitrate (EMS) and sodium citrate (EMS)[348] and viewed on a JEOL 1400 Plus transmission electron microscope (JEOL USA, Inc. Peabody, MA) operated at

120 kV. We collected more than one hundred images for both cell types obtained three independent experiments.

Ferroptosis induction and inhibition

K562 cell lines were exposed to varied concentrations (10, 5, 1 μM) of the ferroptosis inducer, erastin-glutathione depletion (Sigma) either alone or in combination with the lipid peroxidation inhibitor, ferrostatin-1 (Sigma, 10 μM). DMSO (0.1%) was used for vehicle. Cell viability was assessed by Incucyte.

Analysis of mitochondria mass and reactive oxygen species production

Cells were incubated with Mito-Tracker Green (MTG), (25nM, Invitrogen) at 37°C for 30 min, washed in 1X PBS/2% BSA as previously described[349]. MTG selectively accumulates in the mitochondrial matrix where it covalently binds to mitochondrial proteins by reacting with free thiol groups of cysteine residues[349]. For the lipophilic fluorophore C11-BODIPY$^{581/591}$ which localizes in subcellular membranes and "senses" lipid peroxidative damage in these compartments[350,351] the cells were incubated with 5 uM BODIPY for 30 min at 37 °C, as described[351]. For a positive control, cumene peroxide (100 uM) was added followed by incubation for 2 hrs at 37C. SYTOX blue was added and incubated for 15 min followed by flow cytometric cell sorting on an Attune flow cytometer (ThermoFisher). Data were analyzed using FlowJo 9.9.3 (TreeStar).

Data deposition

The sequencing data will be deposited to National Center for Biotechnology Information Gene Expression Omnibus database.

Statistical analysis

For statistical comparison of two groups, unpaired two-tailed Student's t test was used. When more than two groups were compared, one-way ANOVA followed by multiple comparisons test was performed using Prism version 7.00 for Mac (GraphPad). Differences among group means were considered significant when the probability value, p, was less than 0.05. Sample size (n) represents biological replicates. No statistical methods were used to pre- determine sample size. *, P <0.05, **, P <0.01, ***, P <0.001.

RESULTS

Differential expression of snoRNAs targeting the PTC domain of rRNA.

We previously reported that *SNORA21* expression is reduced approximately 2-fold in *de novo* AML compared to CD34 cells from healthy donors. To extend this observation, we sequenced the small RNA transcriptome of an additional 65 cases of AML, including AML carrying chromosomal translocations and deletion and/or mutations in spliceosome genes as reported in chapter 3. Compared to normal CD34 cells, *SNORA21* expression was significantly reduced in AML carrying mutations in spliceosome genes, core binding factor genes, and MLL translocations (Figure 4.1A). In addition, marked variability in *SNORA21* expression was observed with a given AML genetic subtype. For example, in spliceosome mutated AML, *SNORA21* expression varied more than 100-fold. Given the critical role that PTC snoRNAs play in translation fidelity, we next looked at the expression of other snoRNAs that target the PTC domain of the ribosome (circled in Figure 4.1B). Although considerably variability was observed, expression of most of these PTC-snoRNAs in AML was similar to that seen in normal CD34+ cells (Figure 4.1C-D).

Generation of *SNORA21*$^{-/-}$ K562 cells

To examine the contribution of *SNORA21* to ribosome biogenesis and function, we used CRISPR-gene editing to delete *SNORA21* from K562 cells, an erythroleukemia cell line. *SNORA21* is located in the intron of the host gene *RPL23*. We designed the guide RNA to target a region included in the mature *SNORA21* (Figure 4.2A). The *SNORA21* gRNA complexed to recombinant Cas9 was transfected into K562 cells and three independent

clones were obtained by limiting dilution cloning. Sanger sequencing confirmed disruption of the *SNORA21* gene (Figure 4.2B), and RNA sequencing confirmed the loss of *SNORA21* expression; however, expression of RPL23 was not affected (Figure 4.2C-D). We measured rRNA pseudouridylation at rRNA28S-4470 and rRNA28S-4401, sites which are targeted by SNORA21. We predict that our ongoing data analysis will reveal a loss of pseudouridylation at these sites.

Loss of *SNORA21* results in decreased ribosome function

In yeast, loss of pseudouridylation at U2919 in the 25S rRNA subunit is associated with impaired ribosome biogenesis, function, and altered translation[118]. To quantify ribosome biogenesis and assembly into polysomes, we performed polysome profiling on *SNORA21*$^{-/-}$ K562 cells. Steady-state levels of the small 40S and large 60S ribosomal subunits was comparable to control cells (Figure 4.3A-B). However, there was a significant decrease in 80S (monosomes) and a trend to a decrease in polysomes (Figure 4.3A-B). To assess global protein synthesis, we next employed an assay that uses O-propargylpuromycin (OP-Puro), an alkyne analog of puromycin, and click chemistry to label nascent polypeptide chains[333]. The amount of OP-Puro incorporation in *SNORA21*$^{-/-}$ K562 cells was reduced by approximately 40% compared to control cells. (Figure 4.3C-D). Collectively, these data show that loss of *SNORA21* results in impaired ribosome function and reduced global translation.

SNORA21 deletion in K562 cells has a modest impact on translational efficiency

There is evidence that altered ribosome function can result in selective changes in the translation of mRNAs[154,352,353]. Recent studies indicate that in *DKC1*-depleted cells with resultant reduction in pseudouridylation, the impact of the alteration in IRES-mediated protein synthesis is bifurcated: 1) reduced translation of anti-apoptotic factors (*Bcl-xL* and *XIAP*) and tumor suppressors (*p27* and *p53*)[152,354] or 2) increased translation of other IRES-containing transcripts such as VEGF and heat shock protein 70 mRNAs[149]. Thus, we next performed active mRNA translation sequencing (RiboSeq), which quantifies mRNA that are being translated. In brief, ribosomes complexed to mRNA are digested with ribonucleases to degrade RNA not protected by ribosomes. The protected RNA is subjected to next generation sequencing. In parallel, standard RNA sequencing is performed to allow for a calculation of translational efficiency, using RiboDiff[339]. The majority of genes showed no difference in translational efficiency (Figure 4.4A). Indeed, only 145 genes showed a difference in translational efficiency, as defined by an adjusted p-value of < 0.01 and a > 2-fold change (Supplementary Table 4.1). Interestingly, the great majority (90%) of these genes showed increased translational efficiency (Figure 4.4B). Of note, no enrichment for TOP mRNAs characterized by an oligopyrimidine tract at the 5' terminus was observed. These data show that loss of *SNORA21*, despite the overall decrease in translation, has a modest impact on the translational efficiency of individual genes.

SNORA21 deletion in K562 cells induces a non-apoptotic cell death

We next assessed the impact of the loss of *SNORA21* on cell growth and survival. Cell proliferation was measured using manual cell counts and validated using the IncuCyte® live-cell analysis system. In each case, cellular proliferation of *SNORA21*$^{-/-}$ clones was markedly reduced (Figure 4.5A-B). Cellular proliferation is a function of the rate of cell division and cell death. Cell cycle status was measured using the FxCycle™ Violet DNA dye. Surprisingly, compared with control cells no difference in the percentage of cells in the S or G2/M phases of the cell cycle were observed (Figure 4.5C-D). Likewise, the rate of DNA synthesis, as measured by EdU incorporation was similar between control and *SNORA21*$^{-/-}$ cells (Figure 4.5E-F). We next asked whether loss of *SNORA21* induced apoptosis. However, no increase in activated Caspase 3/7 or cell surface annexin-V expression was observed in *SNORA21*$^{-/-}$ cells (Figure 4.5G-H). The decrease in cell proliferation despite the normal rate of cell division and absence of apoptosis suggest that loss of *SNORA21* is inducing a non-apoptotic cell death.

SNORA21$^{-/-}$ K562 cells have altered mitochondrial morphology

Prior studies suggested that impaired ribosome function can induce autophagy[355-357]. Indeed, gene set enrichment and pathway analysis of transcriptome sequence data from SNORA21$^{-/-}$ K562 cells identified a molecular signature for phagosomes and translation (Figure 4.6 A-B). Thus, to determine whether loss of *SNORA21* induced autophagic cell death, we performed transmission electron microscopy on *SNORA21*$^{-/-}$ K562 cells. As a positive control, we also analyzed wildtype K562 cells that were serum starved for 3 and 6 hours to induce autophagy. A consistent finding in autophagy is the presence of large,

double-membrane enclosed vesicles[358]. Indeed, autophagosomes were readily observed in serum starved wildtype K562 cells (Figure 4.7A-B). However, they were not observed in *SNORA21*$^{-/-}$ K562 cells (Figure 4.7C). Instead, *SNORA21*$^{-/-}$ K562 cells displayed two consistent features. First, many of the cells demonstrated a loss of membrane integrity, which is indicative of cell necrosis (Figure 4.7D). Second, mitochondria in *SNORA21*$^{-/-}$ K562 cells were smaller and more rounded than control cells (Figure 4.7E-F). Of note, the mitochondrial morphology, but not loss of membrane integrity, has been reported in cells undergoing ferroptosis[359].

Loss of *SNORA21* is associated with impaired mitochondrial function

To test the hypothesis that loss of *SNORA21* induces cell death via ferroptosis, we first measured lipid peroxidation, which is the classic feature of ferroptosis. However, lipid peroxidation, as measured using fluorescent C11-BODIPY, was similar in *SNORA21*$^{-/-}$ K562 cells and control cells (Figure 4.8A-B). Prior studies have established that ferroptosis is induced by treatment with erastin and inhibited by treatment with ferrostatin-1[359,360]. However, treatment with ferrostatin-1 did not reverse the impaired cellular proliferation of *SNORA21*$^{-/-}$ K562 cells (Figure 4.8C). Moreover, *SNORA21*$^{-/-}$ K562 cells showed no increased sensitivity to erastin-induced cell death. Based on these data, we conclude that loss of *SNORA21* does not induce cell death through ferroptosis.

The altered mitochondrial morphology prompted us to characterize mitochondrial function in *SNORA21*$^{-/-}$ K562 cells. Mitochondria number, as measured using Mitotracker, was markedly reduced in *SNORA21*$^{-/-}$ K562 cells (Figure 4.8D-E).

Studies are underway to further assess mitochondrial function by measuring cellular ROS and cellular respiration using the Seahorse system.

Discussion

There have been numerous studies characterizing the transcriptome and genome in AML. However, the contribution of altered translation in AML pathogenesis has received little attention. As outlined in the Introduction, there is convincing evidence that snoRNAs play a key role in the regulation of ribosome function and translation. Our small RNA sequencing identified *SNORA21* as a snoRNA that is differentially expressed in AML compared with normal $CD34^+$ cells. All of the genetic subtypes AML analyzed in this study, except APML showed reduced *SNORA21* expression, with spliceosome gene-mutated AML showing the largest difference (a nearly 3-fold decrease in median expression). Moreover, we observed marked heterogeneity of snoRNA21 expression between AML cases with genetic subtypes. For example, in spliceosome gene-mutated AML, a 20-fold range in *SNORA21* expression was observed. Whether there is heterogeneity in *SNORA21* expression within an AML sample is an open question. Unfortunately, current single cell RNA sequencing approaches to do not efficiently capture snoRNAs due to their small size and lack of a poly-adenylation tail.

Prior studies in yeast have established that *snR10*, the yeast ortholog of *SNORA21*, is required for normal ribosome biogenesis, function, and translational control [118,128,288]. *SnR10*$^{-/-}$ yeast have impaired cellular growth and reduced global protein translation. The function of *SNORA21* in mammalian cells is less well studied. Qin et al., showed that enforced expression of *SNORA21* in a gall bladder cancer cell line resulted in impaired growth both in vitro and in vivo[248]. Here, we show that genetic loss of *SNORA21* in K562 cells results in impaired ribosome function as evidenced by a loss of 80S monosomes and polysomes and a global decrease in protein translation. Studies

are underway to confirm that *SNORA21*$^{-/-}$ in K562 cells have reduced pseudouridylation at U4470 and U4401 in rRNA. Yoon et al., reported that primary murine and human cells carrying loss-of-function mutations of *DKC1*, the major pseudouridine synthase in mammalian cells, have global reduction in pseudouridine[125]. Moreover, they showed that *DKC1*-deficient cells had a specific defect in the translation of mRNA species carrying an IRES motif. In contrast, in *SNORA21*$^{-/-}$ K562 cells, our ribosome sequencing results identified a limited number of genes with altered translational efficiency. Moreover, most of these genes showed increased translational efficiency and there was no specific effect on IRES-containing transcripts. Together, these data suggest that pseudouridylation at other rRNA sites is responsible for the selective defect in translation of IRES-containing transcripts.

Similar to yeast, we show that loss of *SNORA21* in K562 cells results in a marked decrease in cellular proliferation. Surprisingly, the rate of cell division, as assessed by cell cycle status and EdU incorporation was normal, suggesting increased cell death. However, no increase in apoptosis was observed. Thus, we conclude that loss of *SNORA21* in K562 cells resulted in a non-apoptotic cell death. We initially suspected that increased autophagy may have contributed to cell death. However, electron microscopy of *SNORA21*$^{-/-}$ K562 cells did not show the double-membrane phagosomes that are typical of autophagy. Instead, electron microscopy revealed alterations in mitochondrial morphology and plasma membrane disruption (which is suggestive of necrosis). The disruption of mitochondria was confirmed using Mitotracker, which showed that mitochondrial mass in *SNORA21*$^{-/-}$ K562 cells was significantly reduced. Of note, prior studies have linked impaired ribosome biogenesis and endoplasmic reticulum

stress to altered mitochondrial function[361]. Thus, our current working model is that decreased ribosome function in *SNORA21*$^{-/-}$ K562 cells results in impaired mitochondrial function, ultimately leading to cell death. Studies are underway to further characterize mitochondrial function in *SNORA21*$^{-/-}$ K562 using a Seahorse XF Analyzer and by measuring mitochondrial or cellular reactive oxygen species. A caveat of our current study is the use of K562 cells, which lack *TP53* and contain other chromosomal abnormalities. It will be important to confirm these findings using primary hematopoietic cells.

In summary, we provide evidence suggesting that the decreased expression of *SNORA21* in AML may contribute to the regulation of translation by modifying ribosome function. Studies to correlate *SNORA21* expression to global translation in primary AML cells are needed.

Summary and future directions

Next-generation sequencing has dramatically advanced our ability to interrogate the transcriptome of normal and malignant cells. However, initial studies focused on larger transcripts (>200 nucleotides) or miRNAs, resulting in a "sequencing gap" for small RNAs in the 25-200 nucleotide range. In this research, we developed a next-generation sequencing approach that accurately quantify small non-coding RNAs, closing this sequencing gap. We have used this assay to interrogate small non-coding RNA expression in normal hematopoiesis and acute myeloid leukemia. We focused our study on snoRNAs, the most abundant type of small non-coding RNA. Although snoRNAs have typically been considered to have housekeeping functions, we show that a subset of snoRNAs are regulated in a lineage- and development-specific fashion in hematopoiesis. Surprisingly, host gene expression and splicing correlated poorly with snoRNA expression suggesting that other mechanisms regulate the level of mature snoRNAs. One future direction would be to characterize these mechanisms. For example, there is evidence snoRNAs are protected from exonuclease digestion by binding to the snoRNP proteins, such as NOP56/NOP58. Studies to correlate snoRNP protein expression to the expression of specific snoRNAs may be informative.

We show that a subset of snoRNAs are differentially expressed in AML compared to normal CD34+ cells. The most striking example is the dysregulation of snoRNA expression in the imprinted *DLK-DIO3* locus in APML. Confirming prior reports, we show that expression of snoRNAs in this locus are increased up to 1000-fold in APML. In contrast, expression of snoRNAs from the other major imprinted locus in humans, the SNURF-SNRPN locus, is reduced in APML. This observation raises several important

unanswered questions. First, what are the mechanisms resulting in increased *DLK-DIO3* snoRNA expression. In contrast to a prior report, we have observed no change in methylation in the *DLK-DIO3* locus nor is imprinting altered (data not shown). Studies to characterize the chromatin organization of this locus in APML through Hi-C, or other chromosome conformation capture techniques may be informative. A second question is to characterize the contribution of increased *DLK-DIO3* snoRNA expression to APML pathogenesis or response to therapy. *DLK-DIO3* snoRNAs are considered orphan snoRNAs, with no known or convincing predicted target mRNAs. Studies to assess the impact of *DLK-DIO3* snoRNA expression of hematopoietic stem/progenitor cells may be difficult, since expression of multiple (>20) snoRNAs is increased in APML. Thus, expressing a single *DLK-DIO3* snoRNA may not be sufficient to induce changes in cell biology. RNA immunoprecipitation approaches have been used to identify miRNA targets. A similar approach could be used to pull down snoRNA/snRNP complexes to identify snoRNA targets.

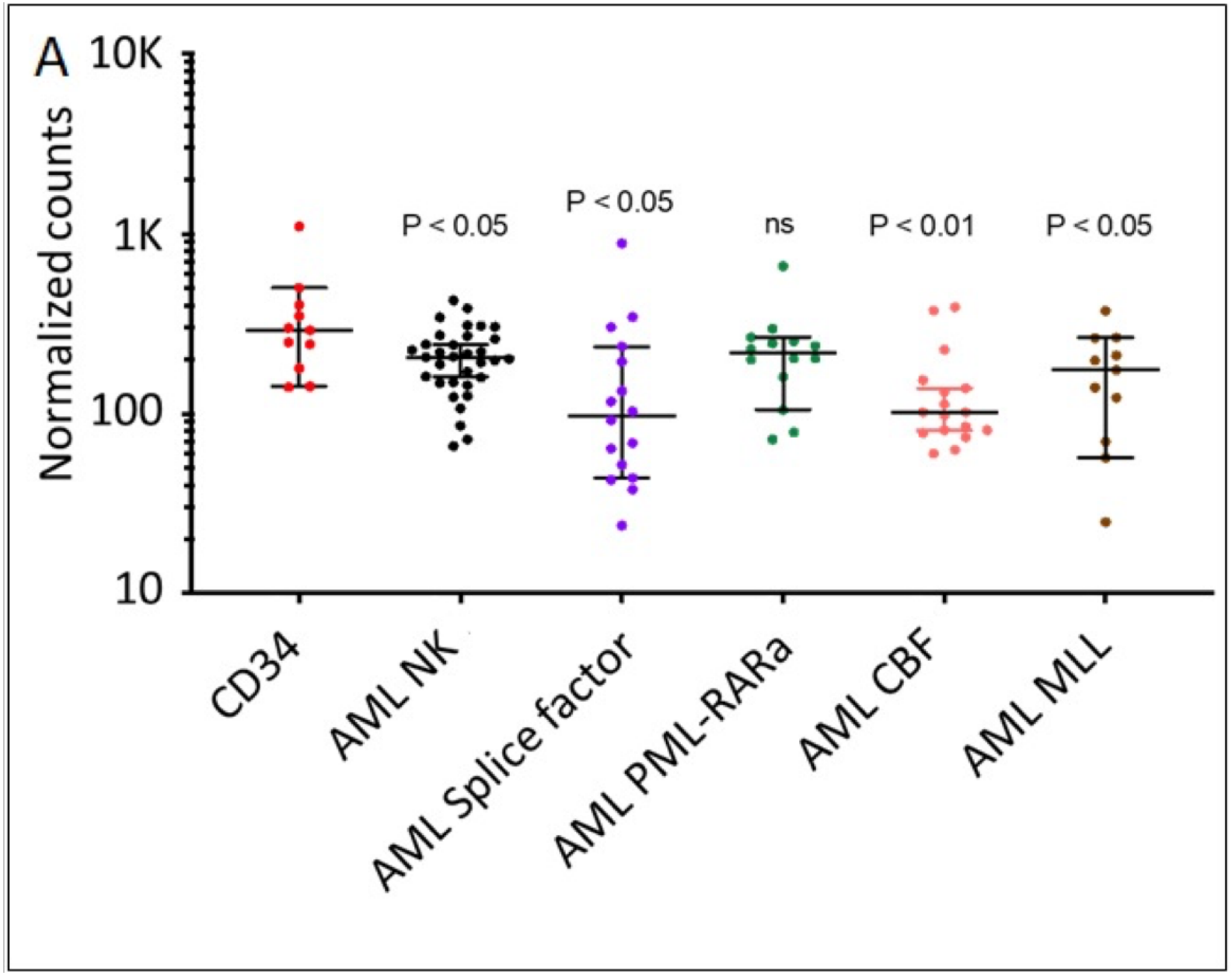
We also identified a limited number of differentially expressed snoRNAs in other genetic subtypes of AML. The analysis of additional AML samples is needed to confirm these findings and identify candidate snoRNAs for biologic validation. Of particular interest, are ongoing studies to assess the impact of altered RNA splicing in spliceosome-gene mutated AML on snoRNA expression. The focus of our analysis will be on host genes that contain snoRNAs in their introns. The major focus of future studies would be to model the impact of snoRNA expression loss (or gain) on hematopoietic stem progenitor proliferation, differentiation, and ribosome function. Another potential future direction would be to use our sequencing approach to interrogate the small RNA

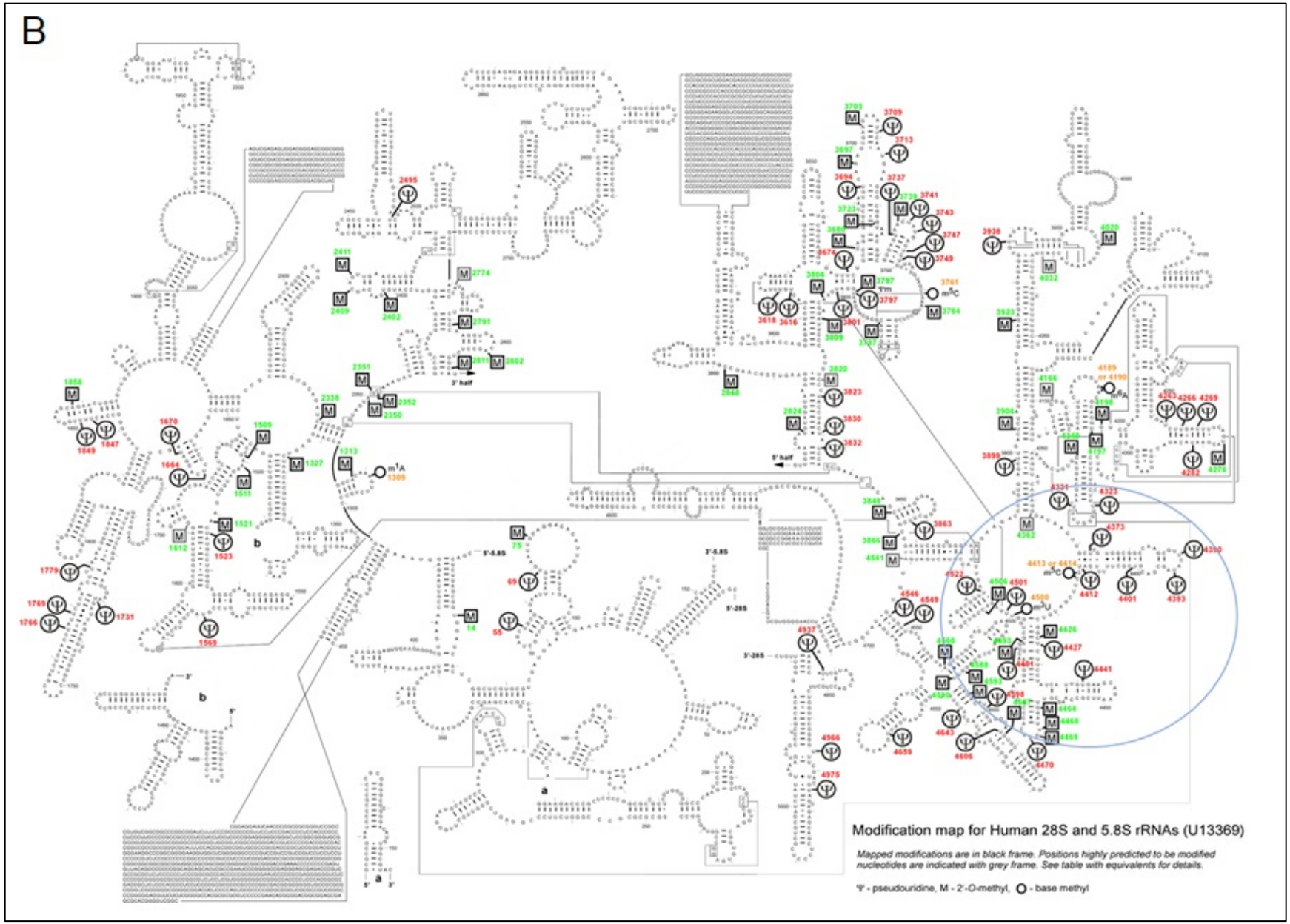
transcriptome of other cancers. Finally, it would worthwhile to modify single cell RNA sequencing approaches to interrogate snoRNAs and other small non-coding RNAs. Relatively modest changes to the library generation protocol that take advantage of the 3'-hydroxyl group of miRNAs and snoRNAs may be sufficient to allow for their detection by single cell RNA sequencing.

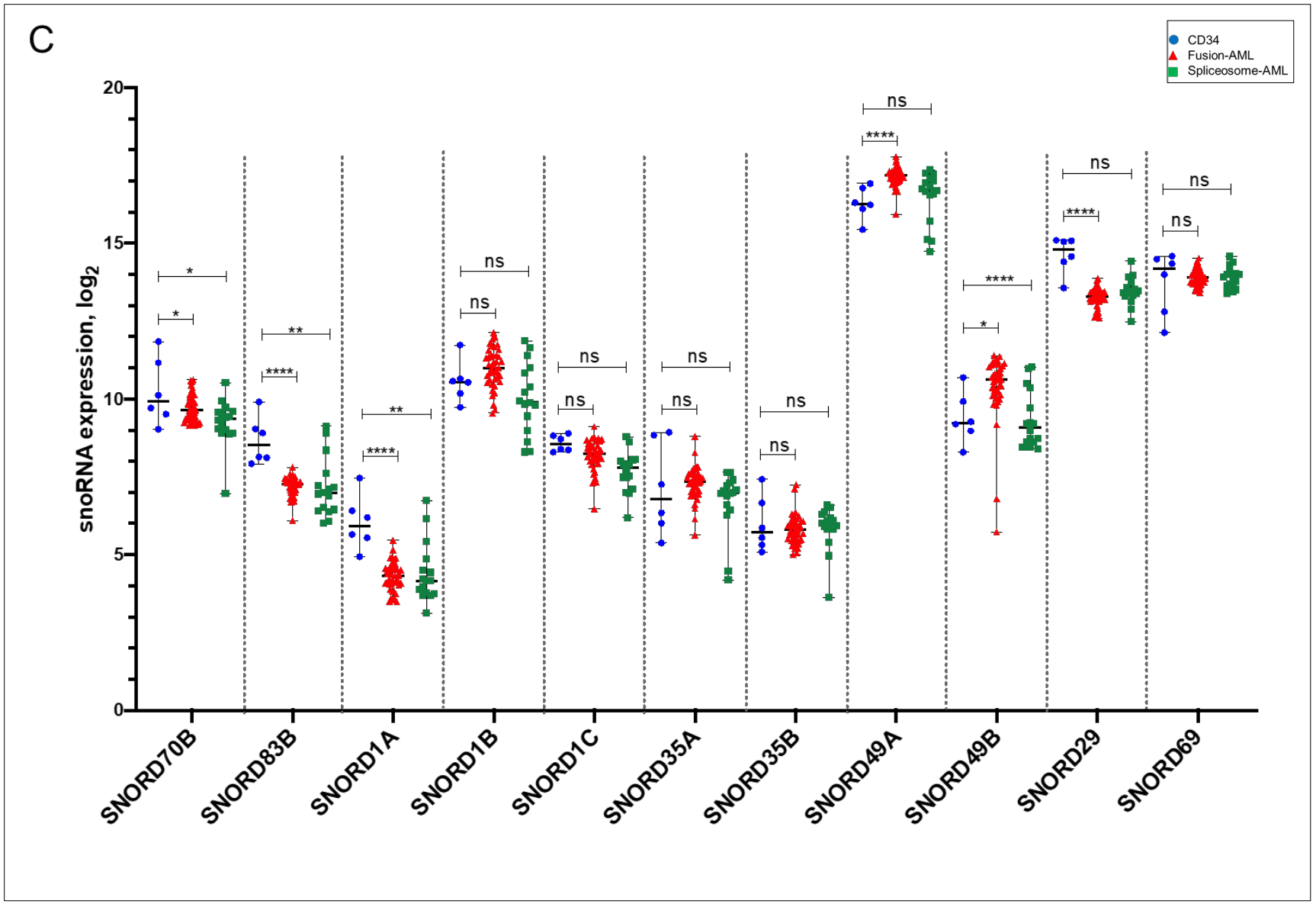
There is increasing evidence that regulation of translation plays an important role in the regulation of hematopoietic cell proliferation and differentiation. Our data suggest that regulation of snoRNA expression may be one mechanism controlling translation. In particular, our research suggests that *SNORA21* may play an important role in regulating ribosome function. *SNORA21* expression was modestly but significantly reduced in most of the AML genetic subtypes that were analyzed. Moreover, we observed marked variation in *SNORA21* expression within AMLs of a given genetic subtype. Our data show that the complete loss of *SNORA21* in K562 cells results in reduced cell growth, impaired ribosome biogenesis, and overall reduced global translation. This proof-of-principle experiment established the importance of SNORA21 in regulating ribosome function in human cells. However, it is not clear whether the more modest changes in *SNORA21* expression seen in AML is sufficient to alter translation. To address this concern, there are several experiments that could be done. First, we could study ribosome function in cell models where SNORA21 expression is reduced, but not completely lost. Second, we could analyze primary AML samples to determine whether protein translation rates correlate with SNORA21 expression. The second experiment will be limited by the number of AML cells that are typically available for analysis.

Our studies show that loss of *SNORA21* in K562 cells results in non-apoptotic cell death. A major future direction would to better characterize the type of cell death and define the pathways leading to cell death. As outline in Chapter 4, current evidence argues against autophagy or ferroptosis. Rather, our data suggest that altered mitochondrial function may be responsible. Ongoing studies to further assess mitochondrial function using the Seahorse CF flux analyzer and assays to measure mitochondrial potential and ROS may be informative. Of note, K562 cells lack *TP53* and *ARF*, key components of the ribosomal stress pathway. Thus, it also will be important to assess the impact of the loss of *SNORA21* in cells with intact TP53 and ARF. One approach would be to use CRISPR-gene editing to delete *SNORA21* from cord blood CD34 cells and assess the impact on cell proliferation and differentiation.

In summary, this research has provided new insight into the expression and function of snoRNAs in normal and malignant hematopoiesis. In particular, this research supports the hypothesis that snoRNAs, by regulating ribosome function and protein translation, may contribute to the regulation of hematopoietic proliferation, survival, and differentiation.


A
10K
1K
100
10
Normalized counts
P < 0.05
P < 0.05
ns
P < 0.01
P < 0.05
CD34
AML NK
AML Splice factor
AML PML-RARa
AML CBF
AML MLL


B
Modification map for Human 28S and 5.8S rRNAs (U13369)
Mapped modifications are in black frame. Positions highly predicted to be modified nucleotides are indicated with grey frame. See table with equivalents for details.
Ψ - pseudouridine, M - 2'-O-methyl, - base methyl
5' half
3' half
5'-5.8S
3'-5.8S
5'-28S
3'-28S
a
b


C
CD34
Fusion-AML
Spliceosome-AML
snoRNA expression, log2
20
15
10
5
0
SNORD70B
SNORD83B
SNORD1A
SNORD1B
SNORD1C
SNORD35A
SNORD35B
SNORD49A
SNORD49B
SNORD29
SNORD69
*
**
****
ns

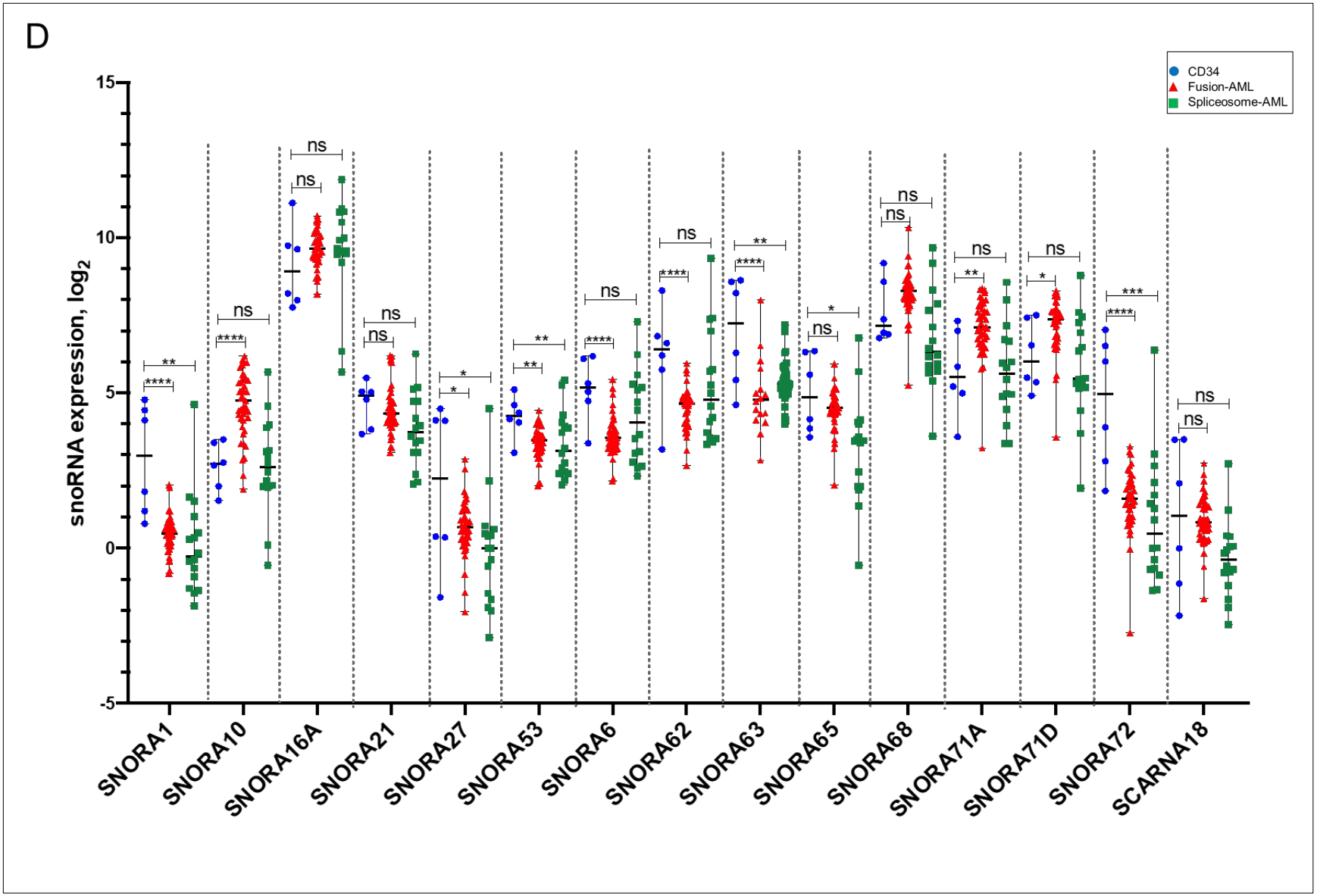

D
CD34
Fusion-AML
Spliceosome-AML
snoRNA expression, log2
15
10
5
0
-5
SNORA1
SNORA10
SNORA16A
SNORA21
SNORA27
SNORA53
SNORA6
SNORA62
SNORA63
SNORA65
SNORA68
SNORA71A
SNORA71D
SNORA72
SCARNA18

Figure 4.1. snoRNAs targeting the PTC domain of the ribosome have reduced expression in AML. A) *SNORA21* normalized expression across CD34 from healthy donors, AML samples with normal karyotype (NK), AML samples with mutations in splicing factors, AML samples with PML-RARa, core binding factor or MLL fusions. B) Overview of the 28 and 5.8S ribosome structure with rRNA modifications. Critical PTC region is circled in blue. Image courtesy Wayne Decateur. Log transformed expression in healthy donor CD34 (blue circle), AML fusions (table 3.1) (red triangles) and samples with mutations in splicing factors (green square) of all snoRNAs that target PTC. C) Expression of C/D box snoRNA that target the PTC. D) Expression of H/ACA snoRNAs that target the PTC.

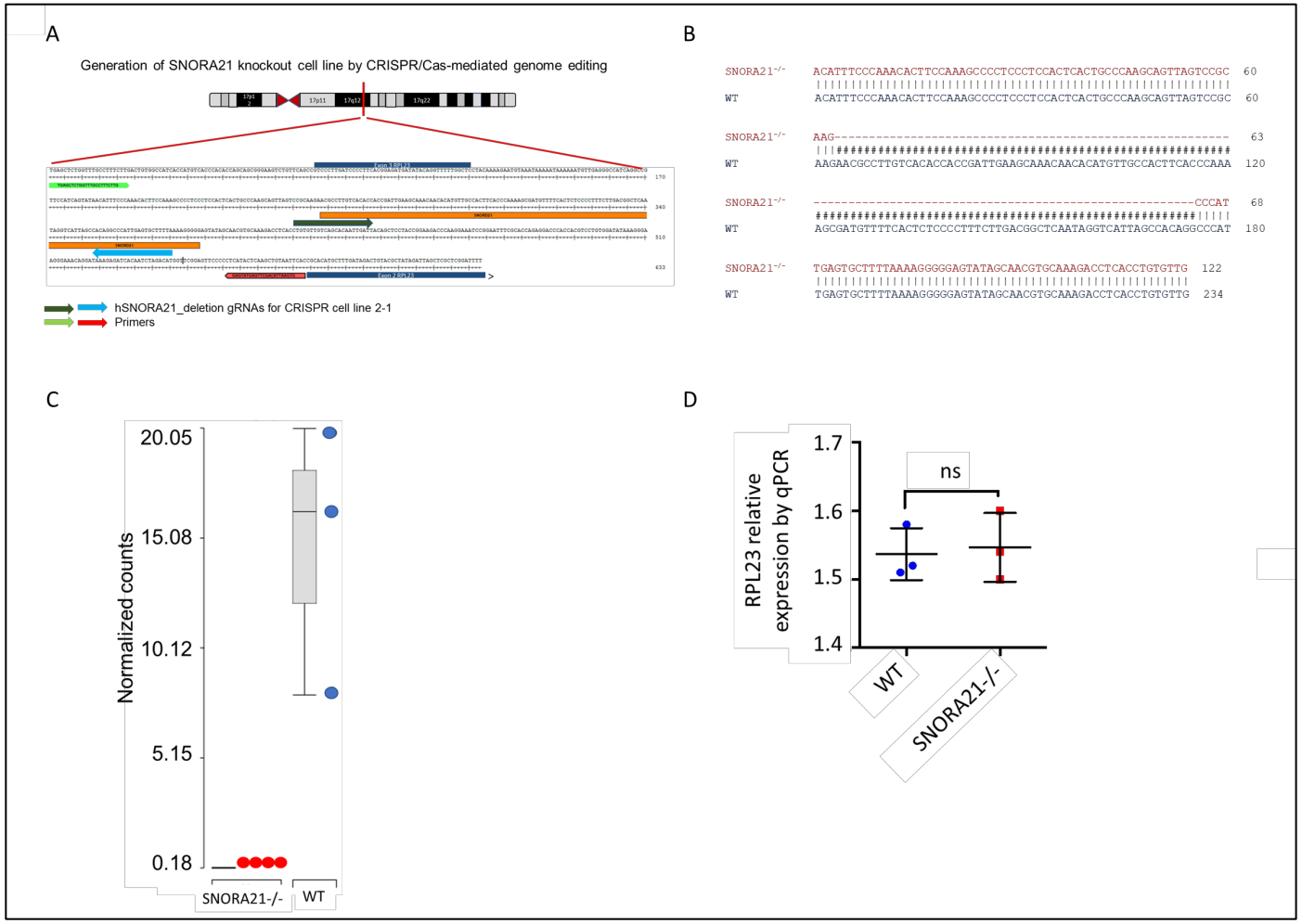


Fig 4.2. Generation of *SNORA21*. A) Schematic representation of genome-browser view and SNORA21 genomic locus is shown. gRNAs as well as primers for validation are shown. B) Sanger sequencing analysis of the CRISPR/Cas9-target site shows the alignment of the wild-type (blue) and SNORA21-/- (red) sequences indicating the 126 bp SNORA21 deletion (dashed line). C) Box plots of normalized SNORA21 expression in WT and SNORA21-/- samples. Analyzed using Partek Flow. D) Relative expression of the candidate genes analyzed by RT-qPCR. Relative expression of host gene RPL23 in WT and SNORA21$^{-/-}$ samples analyzed by RT-qPCR gene expression.

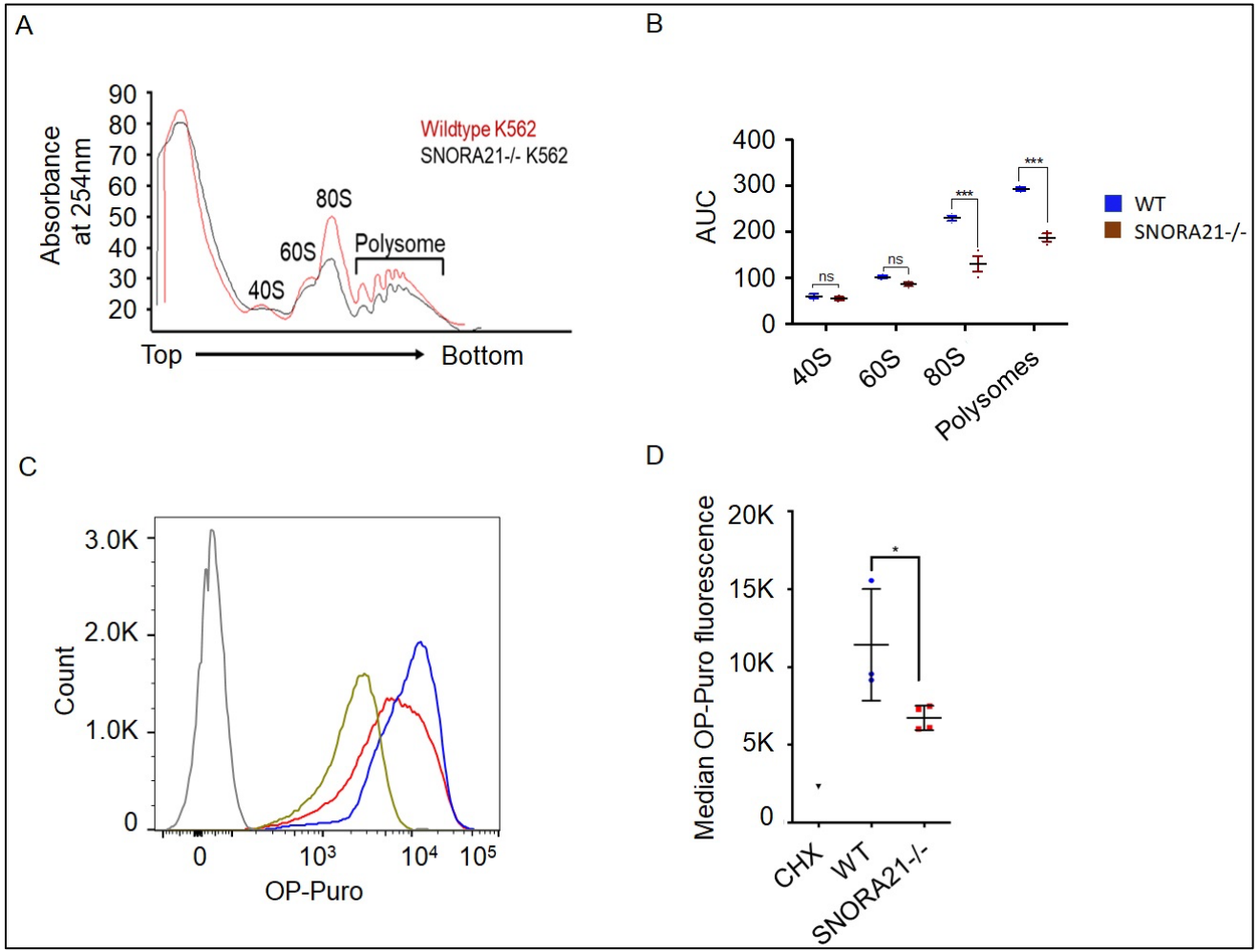


Figure 4.3. SNORA21 is required for ribosome biogenesis. A) Representative sucrose density gradient profiles of ribosomes from SNORA21$^{-/-}$ and WT K562 cell extracts showing the absorbance (A254 nm) of 40S, 60S, and 80S ribosomes. B) Summary data showing the 40S, 60S, 80S and polysome quantification. C) Representative FACS histograms of OP-Puro incorporation with control (grey), WT (blue), CHX treated (green) and SNORA 21$^{-/-}$ K562 cells (red). D) Mean intensity of OP-Puro fluorescence presented in control, CHX treated and SNORA 21$^{-/-}$ K562 cells.

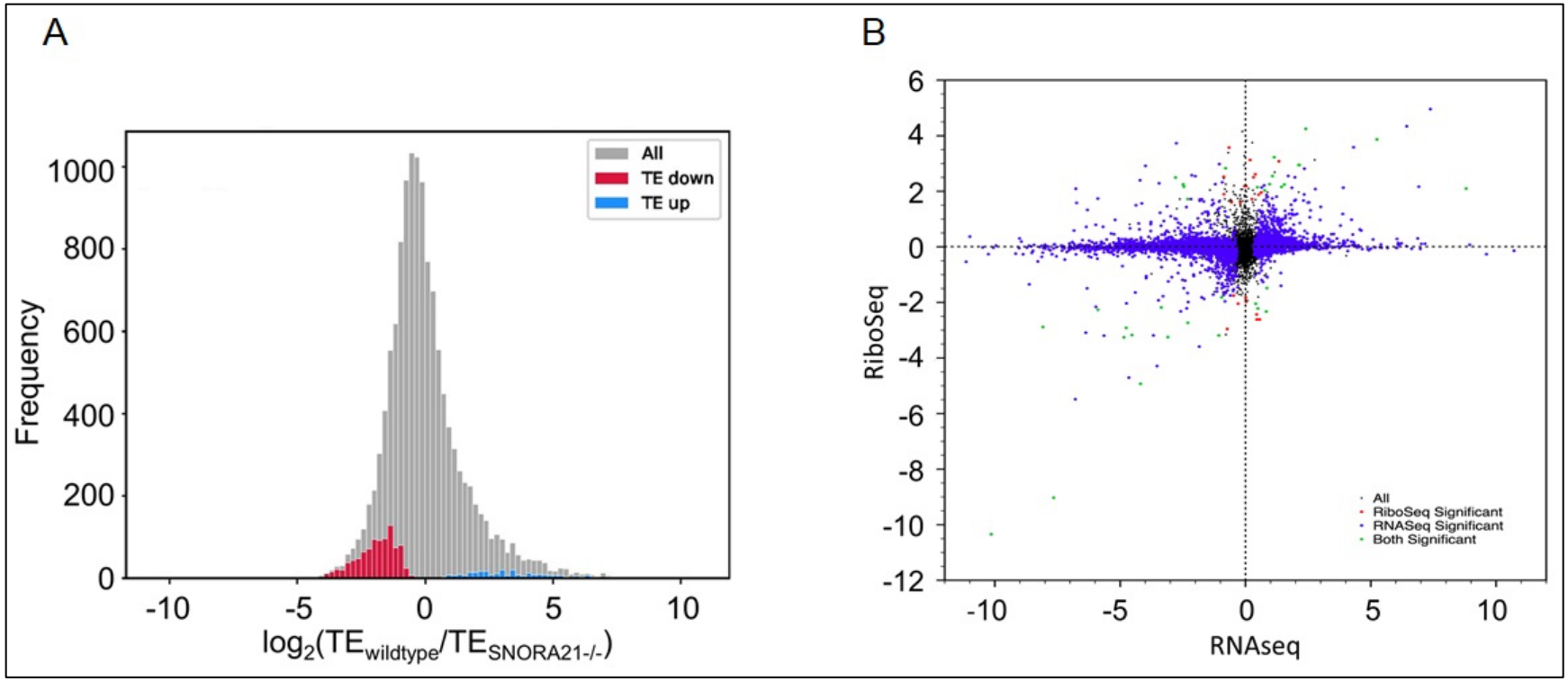


Figure 4.4. Translation efficiency. A) Plot show the frequency of log_2 transformed translation efficiency (TE) ratio where TE is which is the ratio of the abundances of translated mRNA (RiboSeq-ribosome occupancy and available mRNA (RNASeq-transcript levels). B) Although most transcripts showed no differential regulation between the WT and *SNORA21*$^{-/-}$ populations, several dysregulated genes were significantly (FDR ≤ 0.05) regulated at the level of mRNA abundance only (purple dots, RNASeq-only), ribosome occupancy only (red dots, RiboSeq-only), and at the level of both mRNA abundance and ribosome occupancy (green dots, RiboSeq+ RNASeq). Differences in translational efficiency were calculated in Ribodiff on transcripts with average RNA-seq DESeq2 normalized read counts ≥ 10.

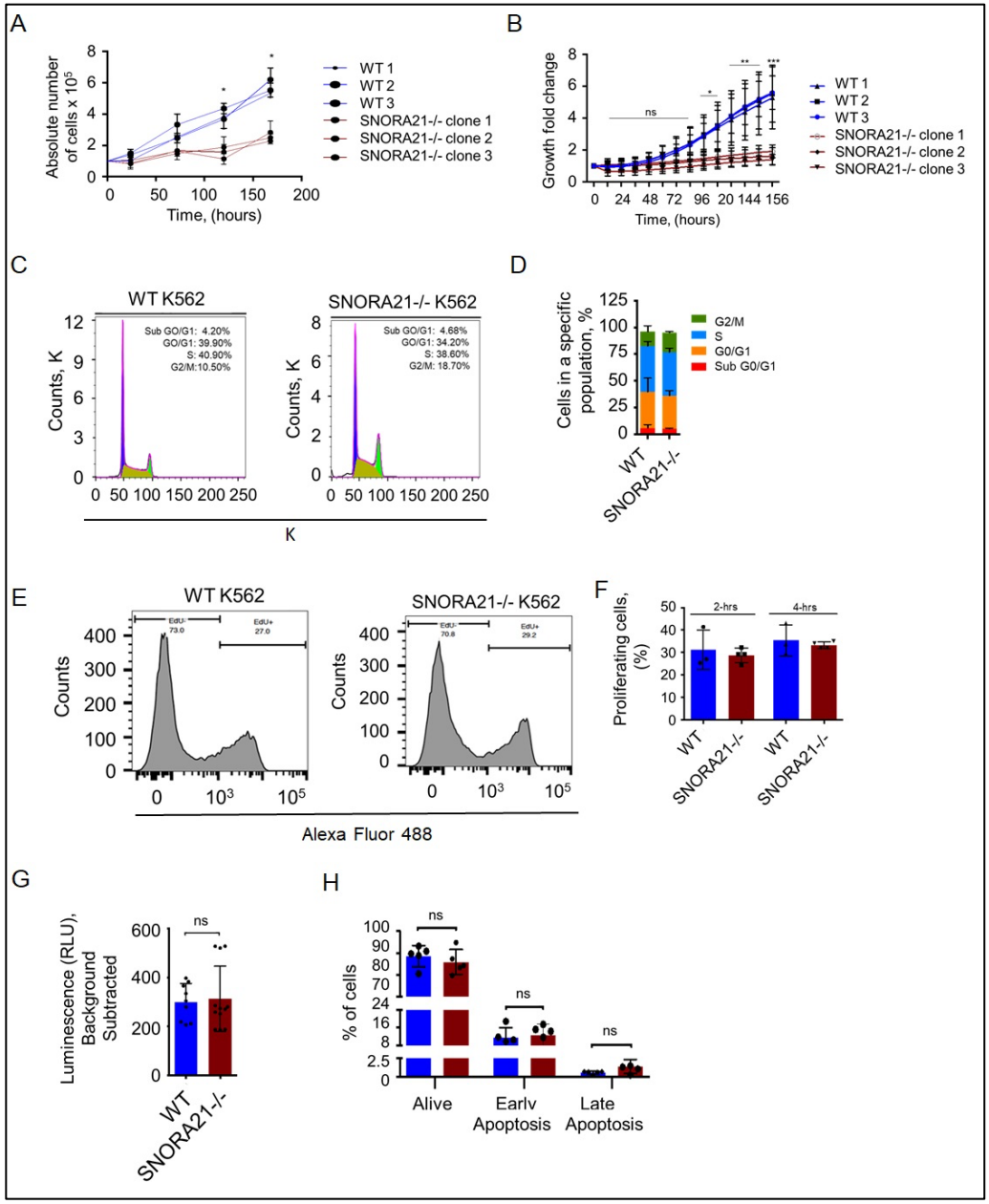

A
Absolute number of cells x 10^5
Time, (hours)
WT 1
WT 2
WT 3
SNORA21-/- clone 1
SNORA21-/- clone 2
SNORA21-/- clone 3
B
Growth fold change
Time, (hours)
ns
C
WT K562
SNORA21-/- K562
Sub GO/G1: 4.20%
GO/G1: 39.90%
S: 40.90%
G2/M:10.50%
Sub GO/G1: 4.68%
GO/G1: 34.20%
S: 38.60%
G2/M: 18.70%
Counts, K
K
D
Cells in a specific population, %
G2/M
S
G0/G1
Sub G0/G1
WT
SNORA21-/-
E
WT K562
SNORA21-/- K562
Counts
Alexa Fluor 488
F
Proliferating cells, (%)
2-hrs
4-hrs
G
Luminescence (RLU), Background Subtracted
ns
H
% of cells
Alive
Early Apoptosis
Late Apoptosis

Figure 4.5 Loss of *SNORA21* leads to altered cell growth. Cells were plated in triplicate in 6-well plates (100,000 cells/well) with growth medium. Cells were counted in triplicate at the indicated time points with a *Bürker chamber* cell chamber after 0.2% Trypan Blue staining to exclude apoptotic/necrotic cells. B) Real-time cell counting of cells labeled with NucLight Red. Cells were incubated at 37°C 5% CO_2 in the IncuCyte live-cell imaging system as described in "Methods." C) Cell cycle analysis of K562 cells. Effects of *SNORA21* deletion on the cell cycle distribution in K562 cells were analyzed using flow cytometry analysis. Representative DNA histograms display cell cycle phase distribution of WT and *SNORA 21*$^{-/-}$ K562 cells. D) Bar charts representing the percentage of various cell populations in WT and *SNORA 21*$^{-/-}$ K562 cells. E) Loss of *SNORA21* has no significant effect on proliferation. Analysis of the cell proliferation rate with Click IT EdU cell proliferation assays (Thermo Fisher) shown in 2-hour representative plots from wildtype and *SNORA21*$^{-/-}$ K562 cell lines, respectively having incorporated EdU and labeled with azide. F) Summary statistics for 2 and 4-hour incubation with EdU. G) *SNORA21* loss has no effect the extent of apoptosis in K562 cells. Intracellular caspase-3/7 activity was assessed by luminescence assay. Luminometer readings were taken 1 hour after adding reagents. Luminescence was proportional to caspase-3 activity. Background readings were determined from wells containing culture medium without K562 cells and subtracted from the experimental values. H) Annexin V/FITC-PI staining and flow cytometric analysis of apoptosis.

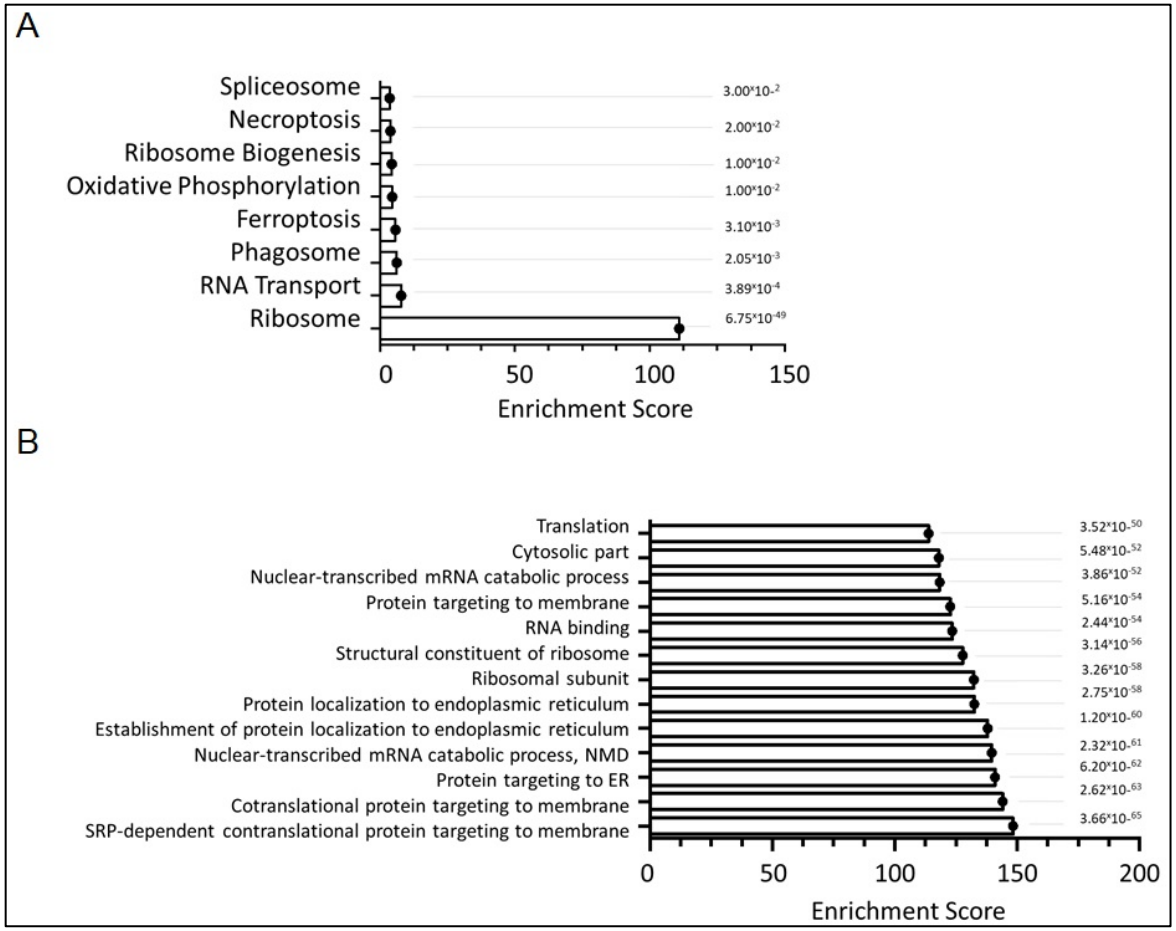


Figure 4.6 Loss of *SNORA21* impacts ribosome biogenesis A) KEGG pathway mapping analysis identified 8 pathways that were significantly enriched among the DEGs. B) Gene ontology enrichment analysis showing representative ontology categories. The enrichment p value is displayed on the side. The enrichment score is the negative natural log of the enrichment p-value derived from Fisher's Exact test.

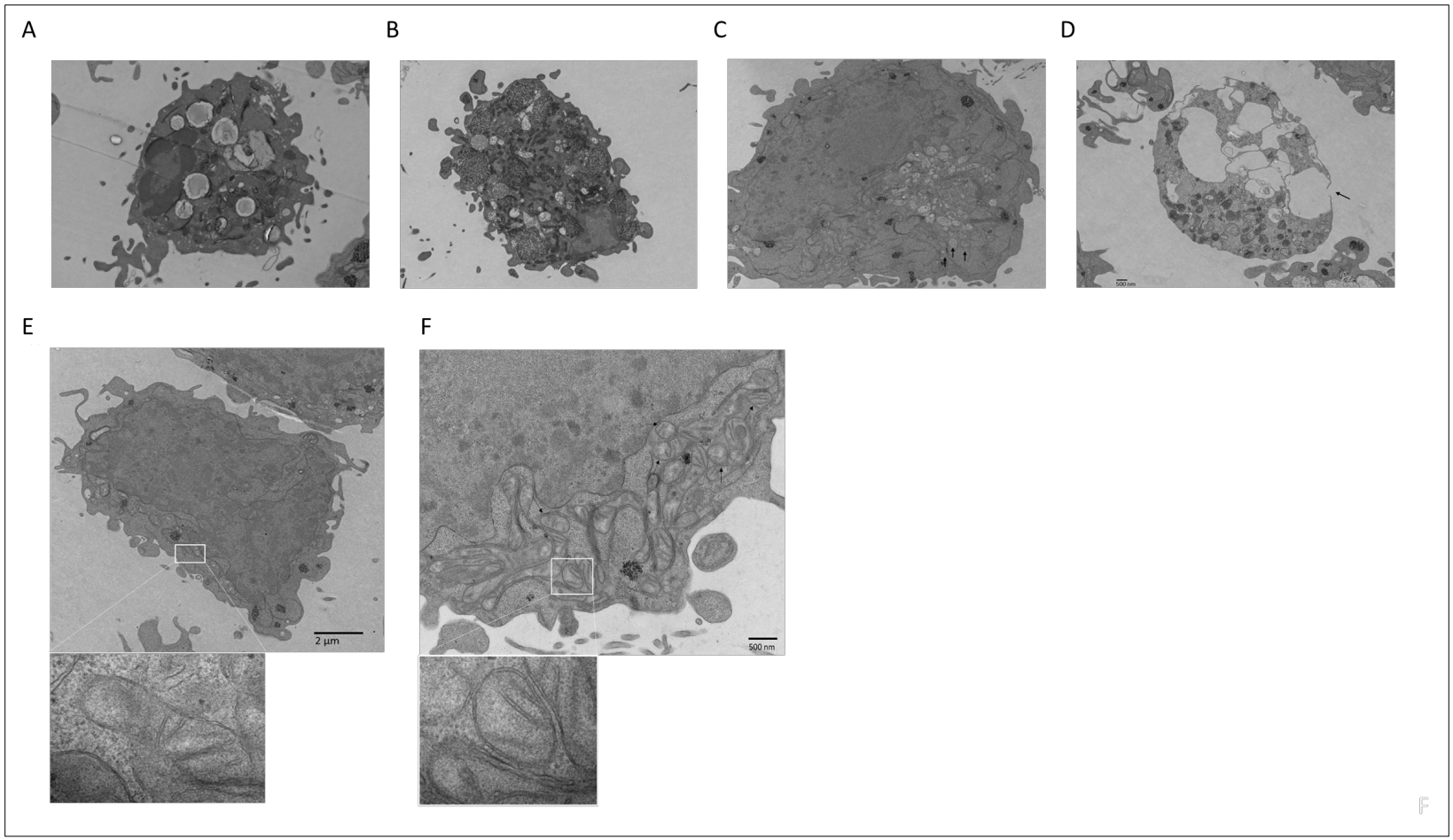


Figure 4.7. Transmission electron microscopy analysis of K562 cells. A) HBSS induced autophagy in K562 cells. Transmission electron microscopy (TEM) micrographs shows the ultrastructure of WT K562 cells incubated with HBSS for 3 hours. B) Cells after 6-hour HBSS incubation. Numerous autophagic vacuoles were observed in the HBSS-treated cells. C) *SNORA21*$^{-/-}$ cells incubated in growth media. D) Plasma membrane integrity loss in *SNORA21*$^{-/-}$ cell undergoing necrosis. (black arrow). E) Wildtype K562 cell with mitochondria highlighted. Inset shows micrograph of normal mitochondria with typical tubular cristae, large matrix, outer membrane, the intermembrane space, and inner membrane. Mitochondria is of normal shape and size. F) Electron micrograph of *SNORA21*$^{-/-}$ K562 cells demonstrating the presence of numerous

mitochondrial aggregates containing various variable numbers of misshapen - smaller rounded and oval shaped - mitochondria (arrow). Clusters of many misshapen mitochondria were common within these aggregates (inset). TEM images were captured at magnifications of A. x8000, B. x4000, C. x4000, D. x4,000. E. ×3,000 and ×60,000 (E, inset), and F ×10,000 ×1,310, x80,000 (F, inset). More than 105 micrographs from each cell type were viewed to select these representative images.

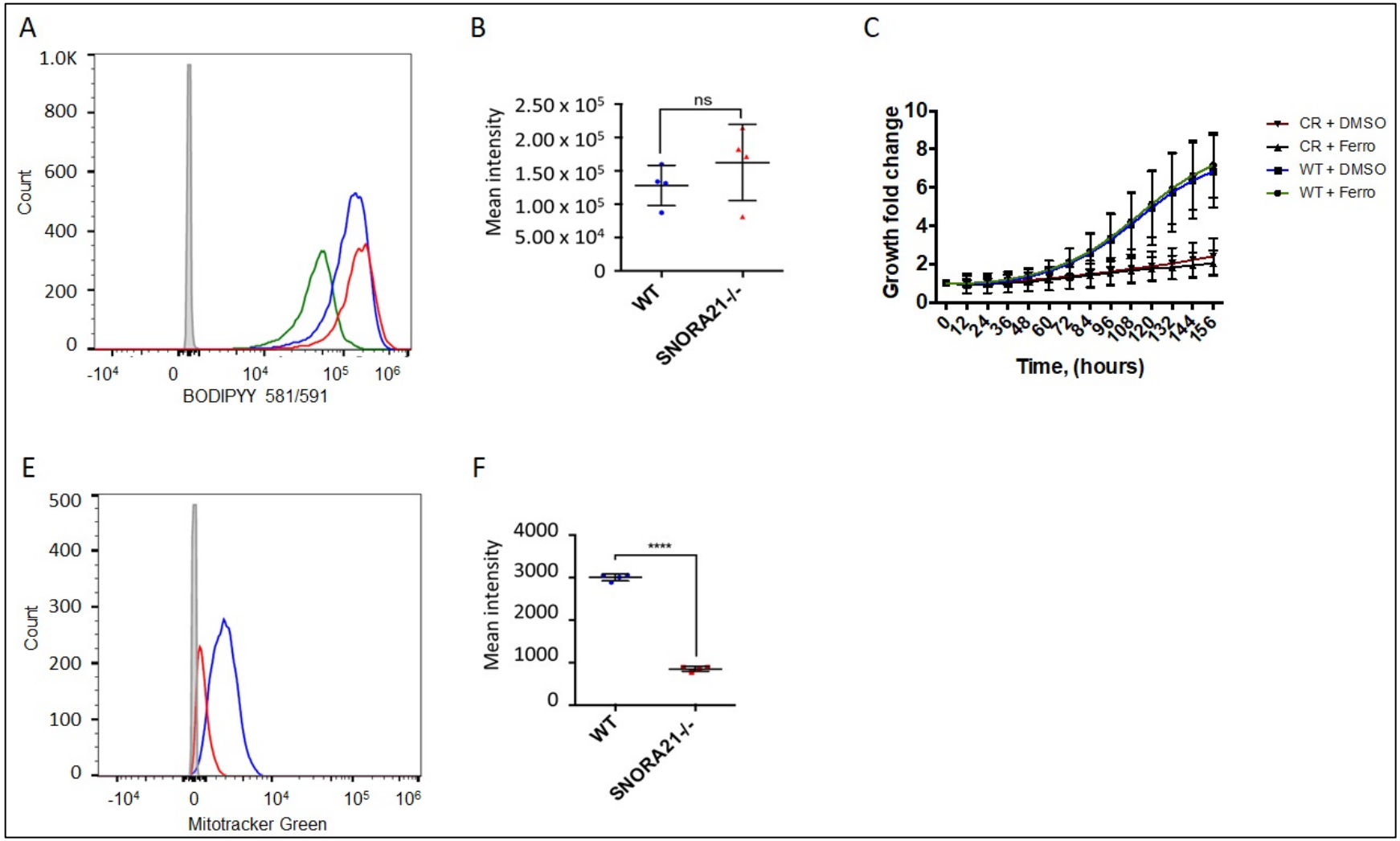


Figure 4.8 Loss of *SNORA21* leads to a loss of mitochondrial mass. A) Representative plot derived from flow cytometry analysis of lipid peroxidation using C11-BODIPY (581/591) probes. B) Results from assessment of lipid peroxidation on three samples. C) WT and *SNORA21*$^{-/-}$ cells were plated in 96 well plates in the presence of either DMSO or 10 μM ferrostatin-1, and survival was quantified after 7 d by Incucyte. Ferrostatin with ferrostatin-1 did not reverse the impaired

cellular proliferation of *SNORA21*$^{-/-}$ K562 cells. D-E) The Mitotracker assay followed by flow cytometry revealed a shift of fluorescence intensity in *SNORA21*$^{-/-}$ K562 cells, indicating decreased amount of mitochondria.

Acknowledgements

This work was supported by Washington University School of Medicine Graduate School of Arts and Sciences/Chancellor's Graduate Fellowship Fund 94028C to Wayne A. Warner. We would like to thank Jessica Hoisington-Lopez and MariaLynn Crosby from the DNA Sequencing Innovation Lab at The Edison Family Center for Genome Sciences and Systems Biology for their sequencing expertise. We gratefully acknowledge the electron microscopy expertise of Greg Strout and Dr. James Fitzpatrick at the Washington University Center for Cellular Imaging (WUCCI), which is supported by Washington University School of Medicine, The Children's Discovery Institute of Washington University and St. Louis Children's Hospital (CDI-CORE-2015-505) and the Foundation for Barnes-Jewish Hospital (3770). We appreciate experimental assistance and advice from Clifford Luke, PhD.

References

1 Busch, H., Reddy, R., Rothblum, L. & Choi, Y. C. SnRNAs, SnRNPs, and RNA processing. *Annu Rev Biochem* **51**, 617-654, doi:10.1146/annurev.bi.51.070182.003153 (1982).

2 Reddy, R., Sitz, T. O., Ro-Choi, T. S. & Busch, H. Two-dimensional polyacrylamide gel electrophoresis separation of low molecular weight nuclear RNA. *Biochemical and biophysical research communications* **56**, 1017-1022, doi:10.1016/s0006-291x(74)80290-1 (1974).

3 Maxwell, E. S. & Fournier, M. J. The small nucleolar RNAs. *Annu Rev Biochem* **64**, 897-934, doi:10.1146/annurev.bi.64.070195.004341 (1995).

4 Wormsley, S., Samarsky, D. A., Fournier, M. J. & Baserga, S. J. An unexpected, conserved element of the U3 snoRNA is required for Mpp10p association. *RNA (New York, N.Y.)* **7**, 904-919 (2001).

5 Mereau, A. *et al.* An in vivo and in vitro structure-function analysis of the Saccharomyces cerevisiae U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. *Journal of molecular biology* **273**, 552-571, doi:10.1006/jmbi.1997.1320 (1997).

6 Tyc, K. & Steitz, J. A. U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. *The EMBO journal* **8**, 3113-3119 (1989).

7 Balakin, A. G., Smith, L. & Fournier, M. J. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823-834 (1996).

8 Zagorski, J., Tollervey, D. & Fournier, M. J. Characterization of an SNR gene locus in Saccharomyces cerevisiae that specifies both dispensible and essential small nuclear RNAs. *Mol Cell Biol* **8**, 3282-3290 (1988).

9 Ganot, P., Bortolin, M. L. & Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799-809 (1997).

10 Ganot, P., Caizergues-Ferrer, M. & Kiss, T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11**, 941-956 (1997).

11 Ganot, P., Jády, B. E., Bortolin, M. L., Darzacq, X. & Kiss, T. Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Molecular and Cellular Biology* **19**, 6906-6917 (1999).

12 Tycowski, K. T., Shu, M. D., Kukoyi, A. & Steitz, J. A. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Mol Cell* **34**, 47-57, doi:10.1016/j.molcel.2009.02.020 (2009).

13 Darzacq, X. *et al.* Cajal body-specific small nuclear RNAs: A novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO Journal* **21**, 2746-2756, doi:10.1093/emboj/21.11.2746 (2002).

14 Kiss, T. Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *Embo J* **20**, 3617 - 3622 (2001).

15 Bachellerie, J. P., Cavaille, J. & Huttenhofer, A. The expanding snoRNA world. *Biochimie* **84**, 775-790, doi:S0300908402014025 [pii] (2002).

16 Kiss, A. M., Jady, B. E., Bertrand, E. & Kiss, T. Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol* **24**, 5797-5807, doi:10.1128/mcb.24.13.5797-5807.2004 (2004).
17 Ganot, P., Bortolin, M. & Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799 - 809 (1997).
18 Ganot, P., Jady, B. E., Bortolin, M. L., Darzacq, X. & Kiss, T. Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol* **19**, 6906-6917 (1999).
19 Bortolin, M. L., Ganot, P. & Kiss, T. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *The EMBO journal* **18**, 457-469, doi:10.1093/emboj/18.2.457 (1999).
20 Lange, T. S., Borovjagin, A. V. & Gerbi, S. A. Nucleolar localization elements in U8 snoRNA differ from sequences required for rRNA processing. *RNA (New York, N.Y.)* **4**, 789-800 (1998).
21 Narayanan, A. *et al.* Nucleolar localization signals of box H/ACA small nucleolar RNAs. *The EMBO journal* **18**, 5120-5130, doi:10.1093/emboj/18.18.5120 (1999).
22 Ni, J. *et al.* SnoRNAs as tools for RNA cleavage and modification. *Nucleic acids symposium series*, 61-63 (1997).
23 Ni, J., Tien, A. L. & Fournier, M. J. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565-573 (1997).
24 Cohn, W. E. Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *The Journal of biological chemistry* **235**, 1488-1498 (1960).
25 Birkedal, U. *et al.* Profiling of ribose methylations in RNA by high-throughput sequencing. *Angew Chem Int Ed Engl* **54**, 451-455, doi:10.1002/anie.201408362 (2015).
26 Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34**, D158-162, doi:10.1093/nar/gkj002 (2006).
27 Piekna-Przybylska, D., Przybylski, P., Baudin-Baillieu, A., Rousset, J. P. & Fournier, M. J. Ribosome performance is enhanced by a rich cluster of pseudouridines in the A-site finger region of the large subunit. *The Journal of biological chemistry* **283**, 26026-26036, doi:10.1074/jbc.M803049200 (2008).
28 Deschamps-Francoeur, G. *et al.* Identification of discrete classes of small nucleolar RNA featuring different ends and RNA binding protein dependency. *Nucleic Acids Res* **42**, 10073-10085, doi:10.1093/nar/gku664 (2014).
29 Smith, C. M. & Steitz, J. A. Sno storm in the nucleolus: new roles for myriad small RNPs. *Cell* **89**, 669-672, doi:10.1016/s0092-8674(00)80247-0 (1997).
30 Bousquet-Antonelli, C., Vanrobays, E., Gélugne, J. P., Caizergues-Ferrer, M. & Henry, Y. Rrp8p is a yeast nucleolar protein functionally linked to Gar1p and involved in pre-rRNA cleavage at site A2. *RNA (New York, N.Y.)* **6**, 826-843, doi:10.1017/S1355838200992288 (2000).
31 Henras, A. *et al.* Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO Journal* **17**, 7078-7090 (1998).

32 Watkins, N. J. *et al.* A common core RNP structure shared between the small nucleoar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* **103**, 457-466 (2000).
33 Reichow, S. L., Hamma, T., Ferre-D'Amare, A. R. & Varani, G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* **35**, 1452-1464, doi:gkl1172 [pii] 10.1093/nar/gkl1172 [doi] (2007).
34 Bagni, C. & Lapeyre, B. Gar1p binds to the small nucleolar RNAs snR10 and snR30 in vitro through a nontypical RNA binding element. *The Journal of biological chemistry* **273**, 10868-10873 (1998).
35 Henras, A., Dez, C., Noaillac-Depeyre, J., Henry, Y. & Caizergues-Ferrer, M. Accumulation of H/ACA snoRNPs depends on the integrity of the conserved central domain of the RNA-binding protein Nhp2p. *Nucleic Acids Res* **29**, 2733-2746 (2001).
36 Henras, A. *et al.* Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *The EMBO journal* **17**, 7078-7090, doi:10.1093/emboj/17.23.7078 (1998).
37 Weinstein, L. B. & Steitz, J. A. Guided tours: from precursor snoRNA to functional snoRNP. *Curr. Opin. Cell Biol.* **11**, 378-384 (1999).
38 Schattner, P., Barberan-Soler, S. & Lowe, T. A computational screen for mammalian pseudouridylation guide H/ACA RNAs. *RNA (New York, N.Y.)* **12**, 15 - 25 (2006).
39 Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23**, 5020-5026 (1995).
40 Charette, M. & Gray, M. W. Pseudouridine in RNA: what, where, how, and why. *IUBMB life* **49**, 341-351, doi:10.1080/152165400410182 (2000).
41 Kierzek, E. *et al.* The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res* **42**, 3492-3501, doi:10.1093/nar/gkt1330 (2014).
42 Davis, D. R. & Poulter, C. D. 1H-15N NMR studies of Escherichia coli tRNA(Phe) from hisT mutants: a structural role for pseudouridine. *Biochemistry* **30**, 4223-4231 (1991).
43 Arnez, J. G. & Steitz, T. A. Crystal structure of unmodified tRNA(Gln) complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry* **33**, 7560-7567 (1994).
44 Hayrapetyan, A. & Helm, M. in *DNA and RNA Modification Enzymes: Structure, Mechanism, Function, and Evolution* (ed Grosjean H) 550-563 (Landes Bioscience, 2009).
45 Auffinger, P. in *Modification and Editing of RNA* (ed H. Grosjean) 103-112 (AMS Press, 1998).
46 Auffinger, P. & Westhof, E. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA(Asp) anticodon hairpin. *Journal of molecular biology* **269**, 326-341, doi:10.1006/jmbi.1997.1022 (1997).
47 Decatur, W. A. & Fournier, M. J. RNA-guided nucleotide modification of ribosomal and other RNAs. *The Journal of biological chemistry* **278**, 695-698, doi:10.1074/jbc.R200023200 (2003).
48 Durant, P. C. & Davis, D. R. Stabilization of the anticodon stem-loop of tRNALys,3 by an A+-C base-pair and by pseudouridine. *Journal of molecular biology* **285**, 115-131, doi:10.1006/jmbi.1998.2297 (1999).

49 Yarian, C. S. *et al.* Structural and functional roles of the N1- and N3-protons of Ψ at tRNA's position 39. *Nucleic Acids Research* **27**, 3543-3549, doi:10.1093/nar/27.17.3543 (1999).
50 Decatur, W. A., Liang, X., Piekna-Przybylska, D. & Fournier, M. J. Vol. 425 283-316 (2007).
51 Ofengand, J. *et al.* Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harbor symposia on quantitative biology* **66**, 147-159 (2001).
52 Ge, J. & Yu, Y. T. RNA pseudouridylation: new insights into an old modification. *Trends Biochem Sci* **38**, 210-218, doi:10.1016/j.tibs.2013.01.002 (2013).
53 Karijolich, J., Yi, C. & Yu, Y. T. Transcriptome-wide dynamics of RNA pseudouridylation. *Nat. Rev. Mol. Cell Biol.* **16**, 581-585, doi:10.1038/nrm4040 (2015).
54 Karijolich, J. & Yu, Y. T. The new era of RNA modification. *RNA (New York, N.Y.)* **21**, 659-660, doi:10.1261/rna.049650.115 (2015).
55 Karikó, K., Muramatsu, H., Keller, J. M. & Weissman, D. Therapeutic potential of in vivo administered pseudouridine-containing mRNA. *Molecular Therapy* **20**, 48, doi:10.1038/mt.2012.179 (2012).
56 Filipowicz, W. & Kiss, T. Structure and function of nucleolar snRNPs. *Mol. Biol. Rep.* **18**, 149-156 (1993).
57 Filipowicz, W., Pelczar, P., Pogacic, V. & Dragon, F. Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification. *Acta biochimica Polonica* **46**, 377-389 (1999).
58 Filipowicz, W. & Pogacic, V. Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol* **14**, 319-327 (2002).
59 Hirose, T., Shu, M. D. & Steitz, J. A. Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol Cell* **12**, 113-123 (2003).
60 Watkins, N. J. & Bohnsack, M. T. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley interdisciplinary reviews. RNA* **3**, 397-414, doi:10.1002/wrna.117 (2012).
61 Kiss, T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145-148, doi:S0092867402007183 [pii] (2002).
62 Matera, A. G., Terns, R. M. & Terns, M. P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature reviews. Molecular cell biology* **8**, 209-220, doi:10.1038/nrm2124 (2007).
63 Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-351 (2002).
64 Auffinger, P. & Westhof, E. Rules governing the orientation of the 2'-hydroxyl group in RNA. *Journal of molecular biology* **274**, 54-63, doi:10.1006/jmbi.1997.1370 (1997).
65 Auffinger, P. & Westhof, E. Hydration of RNA base pairs. *J Biomol Struct Dyn* **16**, 693-707, doi:10.1080/07391102.1998.10508281 (1998).
66 Falaleeva, M. *et al.* Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E1625-1634, doi:10.1073/pnas.1519292113 (2016).

67 Stamm, S. & Lodmell, J. S. C/D box snoRNAs in viral infections: RNA viruses use old dogs for new tricks. *Noncoding RNA Res* **4**, 46-53, doi:10.1016/j.ncrna.2019.02.001 (2019).
68 Kishore, S. *et al.* The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Human molecular genetics* **19**, 1153-1164, doi:10.1093/hmg/ddp585 (2010).
69 Soeno, Y. *et al.* Identification of novel ribonucleo-protein complexes from the brain-specific snoRNA MBII-52. *RNA (New York, N.Y.)* **16**, 1293-1300, doi:10.1261/rna.2109710 (2010).
70 Taft, R. J. *et al.* Small RNAs derived from snoRNAs. *RNA (New York, N.Y.)* **15**, 1233-1240, doi:10.1261/rna.1528909 (2009).
71 Massenet, S., Bertrand, E. & Verheggen, C. Assembly and trafficking of box C/D and H/ACA snoRNPs. *RNA Biol*, 1-13, doi:10.1080/15476286.2016.1243646 (2016).
72 Kishore, S. & Stamm, S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**, 230-232, doi:10.1126/science.1118265 (2006).
73 Cavaille, J. Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: a complex interplay. *Wiley interdisciplinary reviews. RNA*, doi:10.1002/wrna.1417 (2017).
74 Dieci, G., Preti, M. & Montanini, B. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* **94**, 83-88, doi:10.1016/j.ygeno.2009.05.002 (2009).
75 Tycowski, K. T., Shu, M. D. & Steitz, J. A. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**, 464-466, doi:10.1038/379464a0 (1996).
76 Berndt, H. *et al.* Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA (New York, N.Y.)* **18**, 958-972, doi:10.1261/rna.032292.112 (2012).
77 Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J. & Bordonné, R. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell* **9**, 891-901, doi:10.1016/S1097-2765(02)00484-7 (2002).
78 King, T. H., Decatur, W. A., Bertrand, E., Maxwell, E. S. & Fournier, M. J. A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol Cell Biol* **21**, 7731-7746, doi:10.1128/mcb.21.22.7731-7746.2001 (2001).
79 Newman, D. R., Kuhn, J. F., Shanab, G. M. & Maxwell, E. S. Box C/D snoRNA-associated proteins: two pairs of evolutionarily ancient proteins and possible links to replication and transcription. *RNA (New York, N.Y.)* **6**, 861-879 (2000).
80 Terns, M. P. & Terns, R. M. Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene expression* **10**, 17-39 (2002).
81 Weinstein, L. B. & Steitz, J. A. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol* **11**, 378-384, doi:10.1016/s0955-0674(99)80053-2 (1999).
82 Terns, M. P., Grimm, C., Lund, E. & Dahlberg, J. E. A common maturation pathway for small nucleolar RNAs. *The EMBO journal* **14**, 4860-4871 (1995).
83 Tycowski, K. T., Shu, M. D. & Steitz, J. A. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes Dev* **7**, 1176-1190 (1993).

84 Watkins, N. J., Leverette, R. D., Xia, L., Andrews, M. T. & Stuart Maxwell, E. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA (New York, N.Y.)* **2**, 118-133 (1996).
85 Hirose, T. & Steitz, J. A. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12914-12919, doi:10.1073/pnas.231490998 (2001).
86 Darzacq, X. *et al.* Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *The Journal of cell biology* **173**, 207-218, doi:10.1083/jcb.200601105 (2006).
87 Dez, C., Noaillac-Depeyre, J., Caizergues-Ferrer, M. & Henry, Y. Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. *Mol Cell Biol* **22**, 7053-7065 (2002).
88 Fatica, A., Dlakic, M. & Tollervey, D. Naf1 p is a box H/ACA snoRNP assembly factor. *RNA (New York, N.Y.)* **8**, 1502-1514 (2002).
89 Fatica, A. & Tollervey, D. Insights into the structure and function of a guide RNP. *Nature structural biology* **10**, 237-239, doi:10.1038/nsb0403-237 (2003).
90 Hoareau-Aveilla, C., Bonoli, M., Caizergues-Ferrer, M. & Henry, Y. hNaf1 is required for accumulation of human box H/ACA snoRNPs, scaRNPs, and telomerase. *RNA (New York, N.Y.)* **12**, 832-840, doi:10.1261/rna.2344106 (2006).
91 Fragapane, P., Prislei, S., Michienzi, A., Caffarelli, E. & Bozzoni, I. A novel small nucleolar RNA (U16) is encoded inside a ribosomal protein intron and originates by processing of the pre-mRNA. *The EMBO journal* **12**, 2921-2928 (1993).
92 Ooi, S. L., Samarsky, D. A., Fournier, M. J. & Boeke, J. D. Intronic snoRNA biosynthesis in Saccharomyces cerevisiae depends on the lariat-debranching enzyme: intron length effects and activity of a precursor snoRNA. *RNA (New York, N.Y.)* **4**, 1096-1110 (1998).
93 Kiss, T. & Filipowicz, W. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev* **9**, 1411-1424 (1995).
94 Petfalski, E., Dandekar, T., Henry, Y. & Tollervey, D. Processing of the precursors to small nucleolar RNAs and rRNAs requires common components. *Mol Cell Biol* **18**, 1181-1189 (1998).
95 Caffarelli, E. *et al.* Processing of the intron-encoded U16 and U18 snoRNAs: the conserved C and D boxes control both the processing reaction and the stability of the mature snoRNA. *The EMBO journal* **15**, 1121-1131 (1996).
96 Cecconi, F., Mariottini, P. & Amaldi, F. The Xenopus intron-encoded U17 snoRNA is produced by exonucleolytic processing of its precursor in oocytes. *Nucleic Acids Res* **23**, 4670-4676 (1995).
97 Cohen, Y. *et al.* The increased expression of 14q32 small nucleolar RNA transcripts in promyelocytic leukemia cells is not dependent on PML-RARA fusion gene. *Blood Cancer J* **2**, e92, doi:10.1038/bcj.2012.39 (2012).
98 Liuksiala, T. *et al.* Overexpression of SNORD114-3 marks acute promyelocytic leukemia. *Leukemia* **28**, 233-236, doi:10.1038/leu.2013.250 (2014).

99 Valleron, W. *et al.* Specific small nucleolar RNA expression profiles in acute leukemia. *Leukemia* **26**, 2052-2060, doi:10.1038/leu.2012.111 (2012).
100 Head, S. R. *et al.* Library construction for next-generation sequencing: overviews and challenges. *BioTechniques* **56**, 61-64, 66, 68, passim, doi:10.2144/000114133 (2014).
101 Valleron, W. *et al.* Small nucleolar RNA expression profiling identifies potential prognostic markers in peripheral T-cell lymphoma. *Blood* **120**, 3997-4005, doi:10.1182/blood-2012-06-438135 (2012).
102 Zhou, F. *et al.* AML1-ETO requires enhanced C/D box snoRNA/RNP formation to induce self-renewal and leukaemia. *Nature cell biology* **19**, 844-855, doi:10.1038/ncb3563 (2017).
103 Henras, A. K. *et al.* The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cellular and Molecular Life Sciences* **65**, 2334-2359, doi:10.1007/s00018-008-8027-0 (2008).
104 Kressler, D., Hurt, E. & Baßler, J. Driving ribosome assembly. *Biochimica et Biophysica Acta - Molecular Cell Research* **1803**, 673-683, doi:10.1016/j.bbamcr.2009.10.009 (2010).
105 Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**, 437-440 (1999).
106 Anger, A. M. *et al.* Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80-85, doi:10.1038/nature12104 (2013).
107 Bashan, A. & Yonath, A. Correlating ribosome function with high-resolution structures. *Trends Microbiol* **16**, 326-335, doi:10.1016/j.tim.2008.05.001 (2008).
108 Green, R. & Noller, H. F. Ribosomes and translation. *Annu Rev Biochem* **66**, 679-716, doi:10.1146/annurev.biochem.66.1.679 (1997).
109 Moore, P. B. & Steitz, T. A. The structural basis of large ribosomal subunit function. *Annu Rev Biochem* **72**, 813-850, doi:10.1146/annurev.biochem.72.110601.135450 (2003).
110 Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-930, doi:10.1126/science.289.5481.920 (2000).
111 Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234-1242, doi:10.1038/nature08403 (2009).
112 Schmeing, T. M. *et al.* The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* **326**, 688-694, doi:10.1126/science.1179700 (2009).
113 Kos, M. & Tollervey, D. Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell* **37**, 809-820, doi:10.1016/j.molcel.2010.02.024 (2010).
114 Henras, A. K., Dez, C. & Henry, Y. RNA structure and function in C/D and H/ACA s(no)RNPs. *Current opinion in structural biology* **14**, 335-343, doi:10.1016/j.sbi.2004.05.006 (2004).
115 Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A. & Gleizes, P. E. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley interdisciplinary reviews. RNA* **6**, 225-242, doi:10.1002/wrna.1269 (2015).

116 Beltrame, M., Henry, Y. & Tollervey, D. Mutational analysis of an essential binding site for the U3 snoRNA in the 5' external transcribed spacer of yeast pre-rRNA. *Nucleic Acids Res* **22**, 5139-5147 (1994).
117 Beltrame, M. & Tollervey, D. Base pairing between U3 and the pre-ribosomal RNA is required for 18S rRNA synthesis. *The EMBO journal* **14**, 4350-4356 (1995).
118 King, T. H., Liu, B., McCully, R. R. & Fournier, M. J. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* **11**, 425-435 (2003).
119 Ofengand, J. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS letters* **514**, 17-25 (2002).
120 Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K. & Lane, B. G. The pseudouridine residues of ribosomal RNA. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **73**, 915-924 (1995).
121 Ofengand, J. *et al.* Vol. 66 147-159 (2001).
122 Peculis, B. A., DeGregorio, S. & McDowell, K. The U8 snoRNA gene family: identification and characterization of distinct, functional U8 genes in Xenopus. *Gene* **274**, 83-92 (2001).
123 Peculis, B. A. & Steitz, J. A. Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. *Cell* **73**, 1233-1245 (1993).
124 Peculis, B. A. & Steitz, J. A. Sequence and structural elements critical for U8 snRNP function in Xenopus oocytes are evolutionarily conserved. *Genes Dev* **8**, 2241-2255 (1994).
125 Yoon, A. *et al.* Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science* **312**, 902-906, doi:10.1126/science.1123835 (2006).
126 Yoon, J. H., Abdelmohsen, K. & Gorospe, M. Posttranscriptional gene regulation by long noncoding RNA. *Journal of molecular biology* **425**, 3723-3730, doi:10.1016/j.jmb.2012.11.024 (2013).
127 Fromont-Racine, M., Senger, B., Saveanu, C. & Fasiolo, F. Ribosome assembly in eukaryotes. *Gene* **313**, 17-42, doi:10.1016/s0378-1119(03)00629-2 (2003).
128 Baxter-Roshek, J. L., Petrov, A. N. & Dinman, J. D. Optimization of ribosome structure and function by rRNA base modification. *PloS one* **2**, e174, doi:10.1371/journal.pone.0000174 (2007).
129 Krogh, N. *et al.* Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.* **44**, 7884-7895, doi:10.1093/nar/gkw482 (2016).
130 Zhu, Y., Pirnie, S. P. & Carmichael, G. G. High-throughput and site-specific identification of 2'-O-methylation sites using ribose oxidation sequencing (RibOxi-seq). *RNA (New York, N.Y.)* **23**, 1303-1314, doi:10.1261/rna.061549.117 (2017).
131 Pienkowska, J. & Szweykowska-Kulinska, Z. [Small nucleolar RNA]. *Postepy Biochem.* **44**, 102-113 (1998).
132 Harnpicharnchai, P. *et al.* Composition and functional characterization of yeast 66S ribosome assembly intermediates. *Mol Cell* **8**, 505-515 (2001).

133 Cole, S. E., LaRiviere, F. J., Merrikh, C. N. & Moore, M. J. A convergence of rRNA and mRNA quality control pathways revealed by mechanistic analysis of nonfunctional rRNA decay. *Mol Cell* **34**, 440-450, doi:10.1016/j.molcel.2009.04.017 (2009).
134 Ruggero, D. Translational control in cancer etiology. *Cold Spring Harb Perspect Biol* **5**, doi:10.1101/cshperspect.a012336 (2013).
135 Iadevaia, V., Liu, R. & Proud, C. G. mTORC1 signaling controls multiple steps in ribosome biogenesis. *Seminars in cell & developmental biology* **36**, 113-120, doi:10.1016/j.semcdb.2014.08.004 (2014).
136 Tsang, C. K., Liu, H. & Zheng, X. F. mTOR binds to the promoters of RNA polymerase I- and III-transcribed genes. *Cell cycle* **9**, 953-957, doi:10.4161/cc.9.5.10876 (2010).
137 van Riggelen, J., Yetil, A. & Felsher, D. W. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nature reviews. Cancer* **10**, 301-309, doi:10.1038/nrc2819 (2010).
138 Arabi, A. *et al.* c-Myc associates with ribosomal DNA and activates RNA polymerase I transcription. *Nature cell biology* **7**, 303-310, doi:10.1038/ncb1225 (2005).
139 Dai, M. S. & Lu, H. Crosstalk between c-Myc and ribosome in ribosomal biogenesis and cancer. *J Cell Biochem* **105**, 670-677, doi:10.1002/jcb.21895 (2008).
140 Marcel, V., Nguyen Van Long, F. & Diaz, J. J. 40 Years of Research Put p53 in Translation. *Cancers (Basel)* **10**, doi:10.3390/cancers10050152 (2018).
141 Marcel, V., Catez, F. & Diaz, J. J. p53, a translational regulator: contribution to its tumour-suppressor activity. *Oncogene* **34**, 5513-5523, doi:10.1038/onc.2015.25 (2015).
142 Marcel, V. *et al.* p53 acts as a safeguard of translational control by regulating fibrillarin and rRNA methylation in cancer. *Cancer cell* **24**, 318-330, doi:10.1016/j.ccr.2013.08.013 (2013).
143 Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745, doi:10.1016/j.cell.2009.01.042 (2009).
144 Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256-259, doi:10.1038/nature10913 (2012).
145 Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. Structure of the human 80S ribosome. *Nature* **520**, 640-645, doi:10.1038/nature14427 (2015).
146 Sharma, S., Marchand, V., Motorin, Y. & Lafontaine, D. L. J. Identification of sites of 2'-O-methylation vulnerability in human ribosomal RNAs by systematic mapping. *Scientific reports* **7**, 11490, doi:10.1038/s41598-017-09734-9 (2017).
147 Monaco, P. L., Marcel, V., Diaz, J. J. & Catez, F. 2'-O-Methylation of Ribosomal RNA: Towards an Epitranscriptomic Control of Translation? *Biomolecules* **8**, doi:10.3390/biom8040106 (2018).
148 Erales, J. *et al.* Evidence for rRNA 2'-O-methylation plasticity: Control of intrinsic translational capabilities of human ribosomes. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 12934-12939, doi:10.1073/pnas.1707674114 (2017).
149 Rocchi, L. *et al.* Dyskerin depletion increases VEGF mRNA internal ribosome entry site-mediated translation. *Nucleic Acids Res* **41**, 8308-8318, doi:10.1093/nar/gkt587 (2013).

150 Bellodi, C. *et al.* Loss of function of the tumor suppressor DKC1 perturbs p27 translation control and contributes to pituitary tumorigenesis. *Cancer research* **70**, 6026-6035, doi:10.1158/0008-5472.can-09-4730 (2010).
151 Bellodi, C. *et al.* H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. *Cell reports* **3**, 1493-1502, doi:10.1016/j.celrep.2013.04.030 (2013).
152 Montanaro, L. *et al.* Novel dyskerin-mediated mechanism of p53 inactivation through defective mRNA translation. *Cancer research* **70**, 4767-4777, doi:10.1158/0008-5472.can-09-4024 (2010).
153 Penzo, M. *et al.* Dyskerin and TERC expression may condition survival in lung cancer patients. *Oncotarget* **6**, 21755-21760 (2015).
154 Penzo, M. *et al.* Human ribosomes from cells with reduced dyskerin levels are intrinsically altered in translation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **29**, 3472-3482, doi:10.1096/fj.15-270991 (2015).
155 Perdigones, N. *et al.* Clonal hematopoiesis in patients with dyskeratosis congenita. *American Journal of Hematology* **91**, 1227-1233, doi:10.1002/ajh.24552 (2016).
156 Sharma, S. & Lafontaine, D. L. J. 'View From A Bridge': A New Perspective on Eukaryotic rRNA Base Modification. *Trends Biochem Sci* **40**, 560-575, doi:10.1016/j.tibs.2015.07.008 (2015).
157 Natchiar, S. K., Myasnikov, A. G., Kratzat, H., Hazemann, I. & Klaholz, B. P. Visualization of chemical modifications in the human 80S ribosome structure. *Nature* **551**, 472-477, doi:10.1038/nature24482 (2017).
158 Tycowski, K. T., Aab, A. & Steitz, J. A. Guide RNAs with 5' caps and novel box C/D snoRNA-like domains for modification of snRNAs in metazoa. *Current biology : CB* **14**, 1985-1995, doi:10.1016/j.cub.2004.11.003 (2004).
159 Karijolich, J. & Yu, Y. T. Spliceosomal snRNA modifications and their function. *RNA Biol* **7**, 192-204 (2010).
160 Yu, Y. T. & Meier, U. T. RNA-guided isomerization of uridine to pseudouridine--pseudouridylation. *RNA Biol* **11**, 1483-1494, doi:10.4161/15476286.2014.972855 (2014).
161 Yu, Y. T. *et al.* Internal modification of U2 small nuclear (sn)RNA occurs in nucleoli of Xenopus oocytes. *The Journal of cell biology* **152**, 1279-1288 (2001).
162 Yu, Y. T., Shu, M. D. & Steitz, J. A. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *The EMBO journal* **17**, 5783-5795, doi:10.1093/emboj/17.19.5783 (1998).
163 Kelemen, O. *et al.* Function of alternative splicing. *Gene* **514**, 1-30, doi:10.1016/j.gene.2012.07.083 (2013).
164 Pan, Q. *et al.* Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends in genetics : TIG* **21**, 73-77, doi:10.1016/j.tig.2004.12.004 (2005).
165 Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476, doi:10.1038/nature07509 (2008).

166 Kishore, S. & Stamm, S. Regulation of alternative splicing by snoRNAs. *Cold Spring Harbor symposia on quantitative biology* **71**, 329-334, doi:10.1101/sqb.2006.71.024 (2006).
167 Scott, M. S. *et al.* Human box C/D snoRNA processing conservation across multiple cell types. *Nucleic Acids Res* **40**, 3676-3688, doi:10.1093/nar/gkr1233 (2012).
168 Michel, C. I. *et al.* Small nucleolar RNAs U32a, U33, and U35a are critical mediators of metabolic stress. *Cell metabolism* **14**, 33-44, doi:10.1016/j.cmet.2011.04.009 (2011).
169 Lee, J. *et al.* Rpl13a small nucleolar RNAs regulate systemic glucose metabolism. *The Journal of clinical investigation* **126**, 4616-4625, doi:10.1172/jci88069 (2016).
170 Rimer, J. M. *et al.* Long-range function of secreted small nucleolar RNAs that direct 2'-O-methylation. *The Journal of biological chemistry* **293**, 13284-13296, doi:10.1074/jbc.RA118.003410 (2018).
171 Brandis, K. A. *et al.* Box C/D small nucleolar RNA (snoRNA) U60 regulates intracellular cholesterol trafficking. *The Journal of biological chemistry* **288**, 35703-35713, doi:10.1074/jbc.M113.488577 (2013).
172 Chen, J., Odenike, O. & Rowley, J. D. Leukaemogenesis: more than mutant genes. *Nat Rev Cancer* **10**, 23-36, doi:10.1038/nrc2765 (2010).
173 Passegue, E., Jamieson, C. H., Ailles, L. E. & Weissman, I. L. Normal and leukemic hematopoiesis: are leukemias a stem cell disorder or a reacquisition of stem cell characteristics? *Proc Natl Acad Sci U S A* **100 Suppl 1**, 11842-11849, doi:10.1073/pnas.2034201100 (2003).
174 Bellodi, C. *et al.* Disease associated pertubations of H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. *Cell reports* **3**, 1493-1502, doi:10.1016/j.celrep.2013.04.030 (2013).
175 Signer, R. A. J., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**, 49-54, doi:10.1038/nature13035 (2014).
176 Ruggero, D. & Pandolfi, P. P. Does the ribosome translate cancer? *Nat Rev Cancer* **3**, 179-192 (2003).
177 Narla, A. & Ebert, B. L. Ribosomopathies: human disorders of ribosome dysfunction. *Blood* **115**, 3196-3205, doi:10.1182/blood-2009-10-178129 (2010).
178 Sakamoto, K. M., Shimamura, A. & Davies, S. M. Congenital disorders of ribosome biogenesis and bone marrow failure. *Biol Blood Marrow Transplant* **16**, S12-17, doi:10.1016/j.bbmt.2009.09.012 (2010).
179 Montanaro, L., Trere, D. & Derenzini, M. Nucleolus, ribosomes, and cancer. *The American journal of pathology* **173**, 301-310, doi:10.2353/ajpath.2008.070752 (2008).
180 Shiue, C. N., Berkson, R. G. & Wright, A. P. c-Myc induces changes in higher order rDNA structure on stimulation of quiescent cells. *Oncogene* **28**, 1833-1842, doi:10.1038/onc.2009.21 (2009).
181 Bhartiya, D., Talwar, J., Hasija, Y. & Scaria, V. Systematic curation and analysis of genomic variations and their potential functional consequences in snoRNA loci. *Human mutation* **33**, E2367-2374, doi:10.1002/humu.22158 (2012).
182 Mannoor, K., Shen, J., Liao, J., Liu, Z. & Jiang, F. Small nucleolar RNA signatures of lung tumor-initiating cells. *Molecular cancer* **13**, 104, doi:10.1186/1476-4598-13-104 (2014).

183 Sahoo, T. *et al.* Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nature genetics* **40**, 719-721, doi:10.1038/ng.158 (2008).
184 Williams, G. T. & Farzaneh, F. Are snoRNAs and snoRNA host genes new players in cancer? *Nature reviews. Cancer* **12**, 84-88, doi:10.1038/nrc3195 (2012).
185 Gao, L. *et al.* Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. *International journal of cancer. Journal international du cancer* **136**, E623-629, doi:10.1002/ijc.29169 (2015).
186 Liao, J. *et al.* Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer. *Molecular cancer* **9**, 198, doi:10.1186/1476-4598-9-198 (2010).
187 Mei, Y. P. *et al.* Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene* **31**, 2794-2804, doi:10.1038/onc.2011.449 (2012).
188 Okugawa, Y. *et al.* Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. *Gut* **66**, 107-117, doi:10.1136/gutjnl-2015-309359 (2017).
189 Okugawa, Y. *et al.* SnoRA42-an oncogenic small nucleolar rna, and a promising prognostic biomarker in human colorectal cancer. *Gastroenterology* **148**, S17 (2015).
190 Su, H. *et al.* Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene* **33**, 1348-1358, doi:10.1038/onc.2013.89 (2014).
191 Tanaka, R. *et al.* Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human B-cell lymphoma. *Genes to cells : devoted to molecular & cellular mechanisms* **5**, 277-287 (2000).
192 Warner, W. A. *et al.* Characterization of snoRNA expression in acute myeloid leukemia. *Blood* **126**, 3649 (2015).
193 Thorenoor, N. & Slaby, O. Small nucleolar RNAs functioning and potential roles in cancer. *Tumor Biology* **36**, 41-53, doi:10.1007/s13277-014-2818-8 (2014).
194 Stepanov, G. A. *et al.* Regulatory role of small nucleolar RNAs in human diseases. *BioMed research international* **2015**, 206849, doi:10.1155/2015/206849 (2015).
195 Liang, J. *et al.* Small Nucleolar RNAs: Insight Into Their Function in Cancer. *Frontiers in oncology* **9**, 587, doi:10.3389/fonc.2019.00587 (2019).
196 Abel, Y. & Rederstorff, M. SnoRNAs and the emerging class of sdRNAs: Multifaceted players in oncogenesis. *Biochimie* **164**, 17-21, doi:10.1016/j.biochi.2019.05.006 (2019).
197 Chang, L. S., Lin, S. Y., Lieu, A. S. & Wu, T. L. Differential expression of human 5S snoRNA genes. *Biochemical and biophysical research communications* **299**, 196-200 (2002).
198 Gebhart, E. Double minutes, cytogenetic equivalents of gene amplification, in human neoplasia - a review. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* **7**, 477-485 (2005).
199 Schwab, M. Oncogene amplification in solid tumors. *Semin Cancer Biol* **9**, 319-325, doi:10.1006/scbi.1999.0126 (1999).
200 Taguchi, T. *et al.* Combined chromosome microdissection and comparative genomic hybridization detect multiple sites of amplification DNA in a human lung carcinoma cell line. *Genes, chromosomes & cancer* **20**, 208-212 (1997).

201 Lopez-Corral, L. *et al.* Genomic analysis of high-risk smoldering multiple myeloma. *Haematologica* **97**, 1439-1443, doi:10.3324/haematol.2011.060780 (2012).
202 Dong, X. Y. *et al.* Implication of snoRNA U50 in human breast cancer. *Journal of genetics and genomics = Yi chuan xue bao* **36**, 447-454, doi:10.1016/s1673-8527(08)60134-4 (2009).
203 Dong, X. Y. *et al.* SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Human molecular genetics* **17**, 1031-1042, doi:10.1093/hmg/ddm375 (2008).
204 Siprashvili, Z. *et al.* The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. *Nature genetics* **48**, 53-58, doi:10.1038/ng.3452 (2016).
205 Ravo, M. *et al.* Small non-coding RNA deregulation in endometrial carcinogenesis. *Oncotarget* (2015).
206 Teittinen, K. J. *et al.* Expression of small nucleolar RNAs in leukemic cells. *Cellular oncology (Dordrecht)* **36**, 55-63, doi:10.1007/s13402-012-0113-5 (2013).
207 Chu, L. *et al.* Multiple myeloma-associated chromosomal translocation activates orphan snoRNA ACA11 to suppress oxidative stress. *The Journal of clinical investigation* **122**, 2793-2806, doi:10.1172/jci63051 (2012).
208 Mirabella, F. *et al.* MMSET is the key molecular target in t(4;14) myeloma. *Blood cancer journal* **3**, doi:10.1038/bcj.2013.9 (2013).
209 Wu, L., Zheng, J., Chen, P., Liu, Q. & Yuan, Y. Small nucleolar RNA ACA11 promotes proliferation, migration and invasion in hepatocellular carcinoma by targeting the PI3K/AKT signaling pathway. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **90**, 705-712, doi:10.1016/j.biopha.2017.04.014 (2017).
210 Heiss, N. S. *et al.* X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature genetics* **19**, 32-38, doi:10.1038/ng0598-32 (1998).
211 Knight, S. W. *et al.* 1.4 Mb candidate gene region for X linked dyskeratosis congenita defined by combined haplotype and X chromosome inactivation analysis. *Journal of medical genetics* **35**, 993-996, doi:10.1136/jmg.35.12.993 (1998).
212 Kiss, T., Fayet-Lebaron, E. & Jady, B. E. Box H/ACA small ribonucleoproteins. *Mol Cell* **37**, 597-606, doi:10.1016/j.molcel.2010.01.032 (2010).
213 Meier, U. T. The many facets of H/ACA ribonucleoproteins. *Chromosoma* **114**, 1-14, doi:10.1007/s00412-005-0333-9 (2005).
214 Watkins, N. J. *et al.* Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA (New York, N.Y.)* **4**, 1549-1568 (1998).
215 Mitchell, J. R., Wood, E. & Collins, K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* **402**, 551-555, doi:10.1038/990141 (1999).
216 Kiss, T. & Filipowicz, W. Small nucleolar RNAs encoded by introns of the human cell cycle regulatory gene RCC1. *The EMBO journal* **12**, 2913-2920 (1993).
217 Fu, D. & Collins, K. Human telomerase and Cajal body ribonucleoproteins share a unique specificity of Sm protein association. *Genes Dev* **20**, 531-536, doi:10.1101/gad.1390306 (2006).

218 Jady, B. E., Bertrand, E. & Kiss, T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *The Journal of cell biology* **164**, 647-652, doi:10.1083/jcb.200310138 (2004).
219 Venteicher, A. S. *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644-648, doi:10.1126/science.1165357 (2009).
220 Mitchell, J. R., Cheng, J. & Collins, K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol Cell Biol* **19**, 567-576 (1999).
221 Trahan, C. & Dragon, F. Dyskeratosis congenita mutations in the H/ACA domain of human telomerase RNA affect its assembly into a pre-RNP. *RNA (New York, N.Y.)* **15**, 235-243, doi:10.1261/rna.1354009 (2009).
222 Trahan, C., Martel, C. & Dragon, F. Effects of dyskeratosis congenita mutations in dyskerin, NHP2 and NOP10 on assembly of H/ACA pre-RNPs. *Human molecular genetics* **19**, 825-836, doi:10.1093/hmg/ddp551 (2010).
223 Theimer, C. A. *et al.* Structural and functional characterization of human telomerase RNA processing and cajal body localization signals. *Mol Cell* **27**, 869-881, doi:10.1016/j.molcel.2007.07.017 (2007).
224 Gu, B., Fan, J. M., Idol, R., Bessler, M. & Mason, P. J. The effects of pathogenic Dkc1 mutations on telomerase and ribosome biogenesis. *Blood* **116** (2010).
225 Gu, B. W. *et al.* Impaired Telomere Maintenance and Decreased Canonical WNT Signaling but Normal Ribosome Biogenesis in Induced Pluripotent Stem Cells from X-Linked Dyskeratosis Congenita Patients. *PloS one* **10**, e0127414, doi:10.1371/journal.pone.0127414 (2015).
226 Cohen, S. B. *et al.* Protein composition of catalytically active human telomerase from immortal cells. *Science* **315**, 1850-1853, doi:10.1126/science.1138596 (2007).
227 Bellodi, C., Kopmar, N. & Ruggero, D. Deregulation of oncogene-induced senescence and p53 translational control in X-linked dyskeratosis congenita. *EMBO Journal* **29**, 1865-1876, doi:10.1038/emboj.2010.83 (2010).
228 Mochizuki, Y., He, J., Kulkarni, S., Bessler, M. & Mason, P. J. Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10756-10761, doi:10.1073/pnas.0402560101 (2004).
229 Wong, J. M. & Collins, K. Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita. *Genes Dev* **20**, 2848-2858, doi:10.1101/gad.1476206 (2006).
230 Wong, J. M. Y. Disease mechanisms of X-linked dyskeratosis congenita. *Clinical and Experimental Pharmacology and Physiology* **40**, 30, doi:10.1111/1440-1681.12170_6 (2013).
231 Wong, J. M. Y. & Collins, K. Telomerase RNA level limits telomere maintenance in X-linked dyskeratosis congenita. *Genes and Development* **20**, 2848-2858, doi:10.1101/gad.1476206 (2006).
232 Licht, J. D. AML1 and the AML1-ETO fusion protein in the pathogenesis of t(8;21) AML. *Oncogene* **20**, 5660-5679, doi:10.1038/sj.onc.1204593 (2001).

233 Heim, S. & Mitelman, F. Cytogenetic analysis in the diagnosis of acute leukemia. *Cancer* **70**, 1701-1709, doi:10.1002/1097-0142(19920915)70:4+<1701::aid-cncr2820701609>3.0.co;2-s (1992).
234 Mitelman, F. & Heim, S. Quantitative acute leukemia cytogenetics. *Genes, chromosomes & cancer* **5**, 57-66 (1992).
235 Yuan, Y. *et al.* AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10398-10403, doi:10.1073/pnas.171321298 (2001).
236 Elagib, K. E. & Goldfarb, A. N. Oncogenic pathways of AML1-ETO in acute myeloid leukemia: multifaceted manipulation of marrow maturation. *Cancer letters* **251**, 179-186, doi:10.1016/j.canlet.2006.10.010 (2007).
237 Chung, S. S. *et al.* CD99 is a therapeutic target on disease stem cells in myeloid malignancies. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aaj2025 (2017).
238 McGowan, K. A. *et al.* Reduced ribosomal protein gene dosage and p53 activation in low-risk myelodysplastic syndrome. *Blood* **118**, 3622-3633, doi:10.1182/blood-2010-11-318584 (2011).
239 Cavaillé, J., Seitz, H., Paulsen, M., Ferguson-Smith, A. C. & Bachellerie, J. P. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Human molecular genetics* **11**, 1527-1538 (2002).
240 Bachellerie, J. P., Cavaillé, J. & Hüttenhofer, A. The expanding snoRNA world. *Biochimie* **84**, 775-790, doi:10.1016/S0300-9084(02)01402-5 (2002).
241 Jorjani, H. *et al.* An updated human snoRNAome. *Nucleic Acids Res* **44**, 5068-5082, doi:10.1093/nar/gkw386 (2016).
242 Liuksiala, T. *et al.* Overexpression of SNORD114-3 marks acute promyelocytic leukemia. *Leukemia* **28**, 233-236, doi:10.1038/leu.2013.250 (2014).
243 Manodoro, F. *et al.* Loss of imprinting at the 14q32 domain is associated with microRNA overexpression in acute promyelocytic leukemia. *Blood* **123**, 2066-2074, doi:10.1182/blood-2012-12-469833 (2014).
244 Liang, X. H., Liu, Q. & Fournier, M. J. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA (New York, N.Y.)* **15**, 1716-1728, doi:10.1261/rna.1724409 (2009).
245 Østrup, O. *et al.* Limitation of bovine ooplasm in reprogramming of porcine somatic cells. *Reproduction, Fertility and Development* **23**, 125, doi:10.1071/RDv23n1abs (2010).
246 Yoshida, K. *et al.* SNORA21 - An Oncogenic Small Nucleolar RNA, with a Prognostic Biomarker Potential in Human Colorectal Cancer. *EBioMedicine* **22**, 68-77, doi:10.1016/j.ebiom.2017.07.009 (2017).
247 Liu, C. X., Qiao, X. J., Xing, Z. W. & Hou, M. X. The SNORA21 expression is upregulated and acts as a novel independent indicator in human gastric cancer prognosis. *European review for medical and pharmacological sciences* **22**, 5519-5524, doi:10.26355/eurrev_201809_15812 (2018).

248 Qin, Y. *et al.* Overexpression of SNORA21 suppresses tumorgenesis of gallbladder cancer in vitro and in vivo. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* **118**, 109266, doi:10.1016/j.biopha.2019.109266 (2019).
249 Luo, M. *et al.* Long non-coding RNAs control hematopoietic stem cell function. *Cell Stem Cell* **16**, 426-438, doi:10.1016/j.stem.2015.02.002 (2015).
250 Georgantas, R. W., 3rd *et al.* CD34+ hematopoietic stem-progenitor cell microRNA expression and function: a circuit diagram of differentiation control. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2750-2755, doi:10.1073/pnas.0610983104 (2007).
251 Hu, W. *et al.* miR-29a maintains mouse hematopoietic stem cell self-renewal by regulating Dnmt3a. *Blood* **125**, 2206-2216, doi:10.1182/blood-2014-06-585273 (2015).
252 Raaijmakers, M. H. *et al.* Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature* **464**, 852-857, doi:10.1038/nature08851 (2010).
253 Reichow, S. L., Hamma, T., Ferré-D'Amaré, A. R. & Varani, G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Research* **35**, 1452-1464, doi:10.1093/nar/gkl1172 (2007).
254 Zhou, H. L., Luo, G., Wise, J. A. & Lou, H. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res* **42**, 701-713, doi:10.1093/nar/gkt875 (2014).
255 Schubert, T. & Langst, G. Changes in higher order structures of chromatin by RNP complexes. *RNA Biol* **10**, 175-179, doi:10.4161/rna.23175 (2013).
256 Chu, L. *et al.* Multiple myeloma-associated chromosomal translocation activates orphan snoRNA ACA11 to suppress oxidative stress. *Journal of Clinical Investigation* **122**, 2793-2806, doi:10.1172/JCI63051 (2012).
257 Ronchetti, D. *et al.* The expression pattern of small nucleolar and small Cajal body-specific RNAs characterizes distinct molecular subtypes of multiple myeloma. *Blood cancer journal* **2**, e96, doi:10.1038/bcj.2012.41 [doi]
bcj201241 [pii] (2012).
258 Elghetany, M. T., Patel, J., Martinez, J. & Schwab, H. CD87 as a marker for terminal granulocytic maturation: assessment of its expression during granulopoiesis. *Cytometry. Part B, Clinical cytometry* **51**, 9-13, doi:10.1002/cyto.b.10008 (2003).
259 Huang, X. *et al.* Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC genomics* **14**, 319, doi:10.1186/1471-2164-14-319 (2013).
260 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ARXIV* **eprint arXiv:1303.3997**, doi:2013arXiv1303.3997L (2013).
261 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research* **22**, 1760-1774, doi:10.1101/gr.135350.111 (2012).
262 Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research* **42**, D68-D73, doi:10.1093/nar/gkt1181 (2014).
263 Lowe, T. M. & Eddy, S. R. A computational screen for methylation guide snoRNAs in yeast. *Science* **283**, 1168-1171 (1999).
264 Schattner, P., Barberan-Soler, S. & Lowe, T. M. A computational screen for mammalian pseudouridylation guide H/ACA RNAs. *RNA (New York, N.Y.)* **12**, 15-25, doi:12/1/15 [pii]

10.1261/rna.2210406 [doi] (2006).
265 Hertel, J., Hofacker, I. I. & Stadler, P. F. SnoReport: Computational identification of snoRNAs with unknown targets. *Bioinformatics (Oxford, England)* **24**, 158-164, doi:10.1093/bioinformatics/btm464 (2008).
266 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature biotechnology* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
267 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006, doi:10.1101/gr.229102. Article published online before print in May 2002 (2002).
268 Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-144, doi:10.1093/nar/gkj112 (2006).
269 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
270 Partek Flow (Partek Inc., St. Louis, MO, USA, 2014).
271 Eisenhart, C. The assumptions underlying the analysis of variance. *Biometrics* **3**, 1-21 (1947).
272 Frazee, A. C. *et al.* Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature biotechnology* **33**, 243-246, doi:10.1038/nbt.3172 (2015).
273 R: a language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2013).
274 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
275 Ley, T. J. *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).
276 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).
277 Dillies, M. A. *et al.* A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in bioinformatics* **14**, 671-683, doi:10.1093/bib/bbs046 (2013).
278 Garmire, L. X. & Subramaniam, S. Evaluation of normalization methods in mammalian microRNA-Seq data. *RNA (New York, N.Y.)* **18**, 1279-1288, doi:10.1261/rna.030916.111 (2012).
279 Garmire, L. X. & Subramaniam, S. The poor performance of TMM on microRNA-Seq. *RNA (New York, N.Y.)* **19**, 735-736, doi:10.1261/rna.039271.113 (2013).
280 Galiveti, C. R., Rozhdestvensky, T. S., Brosius, J., Lehrach, H. & Konthur, Z. Application of housekeeping npcRNAs for quantitative expression analysis of human transcriptome by real-time PCR. *RNA (New York, N.Y.)* **16**, 450-461, doi:10.1261/rna.1755810 (2010).
281 Alvarez-Dominguez, J. R., Hu, W., Gromatzky, A. A. & Lodish, H. F. Long noncoding RNAs during normal and malignant hematopoiesis. *International journal of hematology* **99**, 531-541, doi:10.1007/s12185-014-1552-8 (2014).

282 Dostalova Merkerova, M. *et al.* Distinctive microRNA expression profiles in CD34+ bone marrow cells from patients with myelodysplastic syndrome. *European journal of human genetics : EJHG* **19**, 313-319, doi:10.1038/ejhg.2010.209 (2011).
283 de Smith, A. J. *et al.* A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Human molecular genetics* **18**, 3257-3265, doi:10.1093/hmg/ddp263 (2009).
284 Peters, J. Prader-Willi and snoRNAs. *Nature genetics* **40**, 688-689, doi:10.1038/ng0608-688 (2008).
285 Stadtfeld, M. *et al.* Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of all-iPS cell mice from terminally differentiated B cells. *Nature genetics* **44**, 398-405, s391-392, doi:10.1038/ng.1110 (2012).
286 Beringer, M. & Rodnina, M. V. The ribosomal peptidyl transferase. *Mol Cell* **26**, 311-321, doi:10.1016/j.molcel.2007.03.015 (2007).
287 Liu, Q. & Fredrick, K. Intersubunit Bridges of the Bacterial Ribosome. *Journal of molecular biology* **428**, 2146-2164, doi:10.1016/j.jmb.2016.02.009 (2016).
288 Liang, X. H., Liu, Q. & Fournier, M. J. rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol Cell* **28**, 965-977, doi:10.1016/j.molcel.2007.10.012 (2007).
289 Dönmez, G., Hartmuth, K. & Lührmann, R. Modified nucleotides at the 5′ end of human U2 snRNA are required for spliceosomal E-complex formation. *RNA (New York, N.Y.)* **10**, 1925-1933, doi:10.1261/rna.7186504 (2004).
290 Zhang, B. *et al.* Changes in snoRNA and snRNA Abundance in the Human, Chimpanzee, Macaque, and Mouse Brain. *Genome biology and evolution* **8**, 840-850, doi:10.1093/gbe/evw038 (2016).
291 Ge, J., Crosby, S. D., Heinz, M. E., Bessler, M. & Mason, P. J. SnoRNA microarray analysis reveals changes in H/ACA and C/D RNA levels caused by dyskerin ablation in mouse liver. *The Biochemical journal* **429**, 33-41, doi:10.1042/bj20091898 (2010).
292 He, H. *et al.* Profiling Caenorhabditis elegans non-coding RNA expression with a combined microarray. *Nucleic Acids Research* **34**, 2976-2983, doi:10.1093/nar/gkl371 (2006).
293 Allmang, C. *et al.* Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO journal* **18**, 5399-5410, doi:10.1093/emboj/18.19.5399 (1999).
294 van Hoof, A., Lennertz, P. & Parker, R. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol* **20**, 441-452 (2000).
295 Lykke-Andersen, S. *et al.* Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* **28**, 2498-2517, doi:10.1101/gad.246538.114 (2014).
296 Kiss, T., Fayet, E., Jady, B. E., Richard, P. & Weber, M. Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harbor symposia on quantitative biology* **71**, 407-417, doi:10.1101/sqb.2006.71.025 (2006).
297 de Turris, V. *et al.* TOP promoter elements control the relative ratio of intron-encoded snoRNA versus spliced mRNA biosynthesis. *Journal of molecular biology* **344**, 383-394, doi:10.1016/j.jmb.2004.09.049 (2004).

298 Smith, C. M. & Steitz, J. A. Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol Cell Biol* **18**, 6897-6909 (1998).
299 Walter, M. J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *The New England journal of medicine* **366**, 1090-1098, doi:10.1056/NEJMoa1106968 (2012).
300 Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264-278, doi:10.1016/j.cell.2012.06.023 (2012).
301 Cai, X. *et al.* Runx1 loss minimally impacts long-term hematopoietic stem cells. *PloS one* **6**, e28430, doi:10.1371/journal.pone.0028430 (2011).
302 Mangan, J. K. & Speck, N. A. RUNX1 mutations in clonal myeloid disorders: from conventional cytogenetics to next generation sequencing, a story 40 years in the making. *Crit Rev Oncog* **16**, 77-91 (2011).
303 Rao, M. S. *et al.* Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Frontiers in genetics* **9**, 636, doi:10.3389/fgene.2018.00636 (2018).
304 Network, C. G. A. R. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).
305 Stelzer, Y., Sagi, I., Yanuka, O., Eiges, R. & Benvenisty, N. The noncoding RNA IPW regulates the imprinted DLK1-DIO3 locus in an induced pluripotent stem cell model of Prader-Willi syndrome. *Nature genetics* **46**, 551-557, doi:10.1038/ng.2968 (2014).
306 Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-69, doi:10.1038/nature10496 (2011).
307 Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644, doi:10.1016/j.cell.2008.01.025 (2008).
308 Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120-136, doi:10.1016/j.stem.2012.01.006 (2012).
309 Maximow, A. A. *Der lymphozyt als gemeinsame stammzelle der verschiedenen blutelement in der embryonalen entwicklung und im postfetalen leben der säugetiere*. (1909).
310 van Galen, P. *et al.* The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature* **510**, 268-272, doi:10.1038/nature13228 (2014).
311 Boileau, M. *et al.* Mutant H3 histones drive human pre-leukemic hematopoietic stem cell expansion and promote leukemic aggressiveness. *Nature communications* **10**, 2891, doi:10.1038/s41467-019-10705-z (2019).
312 Signer, R. A., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **509**, 49-54, doi:10.1038/nature13035 (2014).
313 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).
314 Jovanovic, M. *et al.* Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038, doi:10.1126/science.1259038 (2015).

315 Barna, M. *et al.* Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature* **456**, 971-975, doi:10.1038/nature07449 (2008).
316 Hsieh, A. C. *et al.* Genetic dissection of the oncogenic mTOR pathway reveals druggable addiction to translational control via 4EBP-eIF4E. *Cancer cell* **17**, 249-261, doi:10.1016/j.ccr.2010.01.021 (2010).
317 Hsieh, A. C. *et al.* The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature* **485**, 55-61, doi:10.1038/nature10912 (2012).
318 Khajuria, R. K. *et al.* Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell* **173**, 90-103.e119, doi:10.1016/j.cell.2018.02.036 (2018).
319 Bulut-Karslioglu, A. *et al.* The Transcriptionally Permissive Chromatin State of Embryonic Stem Cells Is Acutely Tuned to Translational Output. *Cell Stem Cell* **22**, 369-383.e368, doi:10.1016/j.stem.2018.02.004 (2018).
320 Sulima, S. O., Hofman, I. J. F., De Keersmaecker, K. & Dinman, J. D. How Ribosomes Translate Cancer. *Cancer discovery*, doi:10.1158/2159-8290.cd-17-0550 (2017).
321 Meroueh, M. *et al.* Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of Escherichia coli 23S rRNA. *Nucleic Acids Res* **28**, 2075-2083 (2000).
322 Newby, M. I. & Greenbaum, N. L. A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA (New York, N.Y.)* **7**, 833-845 (2001).
323 Jack, K. *et al.* RRNA Pseudouridylation Defects Affect Ribosomal Ligand Binding and Translational Fidelity from Yeast to Human Cells. *Mol Cell* **44**, 660-666, doi:10.1016/j.molcel.2011.09.017 (2011).
324 Ruggero, D. *et al.* Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA modification. *Science* **299**, 259-262, doi:10.1126/science.1079447 (2003).
325 Venema, J. & Tollervey, D. RRP5 is required for formation of both 18S and 5.8S rRNA in yeast. *The EMBO journal* **15**, 5701-5714 (1996).
326 McMahon, M., Contreras, A. & Ruggero, D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley interdisciplinary reviews. RNA* **6**, 173-189, doi:10.1002/wrna.1266 (2015).
327 Siprashvili, Z. *et al.* The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. *Nature genetics* **48**, 53-58, doi:10.1038/ng.3452 (2015).
328 Ofengand, J. & Bakin, A. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *Journal of molecular biology* **266**, 246-268, doi:10.1006/jmbi.1996.0737 (1997).
329 Kiss, A. M., Jády, B. E., Bertrand, E. & Kiss, T. Human box H/ACA pseudouridylation guide RNA machinery. *Molecular and Cellular Biology* **24**, 5797-5807, doi:10.1128/MCB.24.13.5797-5807.2004 (2004).
330 Li, S. G., Zhou, H., Luo, Y. P., Zhang, P. & Qu, L. H. Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from Schizosaccharomyces pombe. *The Journal of biological chemistry* **280**, 16446-16455, doi:10.1074/jbc.M500326200 (2005).

331 Cappella, P., Gasparri, F., Pulici, M. & Moll, J. A novel method based on click chemistry, which overcomes limitations of cell cycle analysis by classical determination of BrdU incorporation, allowing multiplex antibody staining. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **73**, 626-636, doi:10.1002/cyto.a.20582 (2008).
332 Salic, A. & Mitchison, T. J. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2415-2420, doi:10.1073/pnas.0712168105 (2008).
333 Liu, J., Xu, Y., Stoleru, D. & Salic, A. Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 413-418, doi:10.1073/pnas.1111561108 (2012).
334 McGlincy, N. J. & Ingolia, N. T. Transcriptome-wide measurement of translation by ribosome profiling. *Methods* **126**, 112-129, doi:10.1016/j.ymeth.2017.05.028 (2017).
335 de Klerk, E. *et al.* Assessing the translational landscape of myogenic differentiation by ribosome profiling. *Nucleic Acids Res* **43**, 4408-4428, doi:10.1093/nar/gkv281 (2015).
336 Trapnell, C. & Schatz, M. C. Optimizing Data Intensive GPGPU Computations for DNA Sequence Alignment. *Parallel computing* **35**, 429-440, doi:10.1016/j.parco.2009.05.002 (2009).
337 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
338 Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (New York, N.Y.)* **324**, 218-223, doi:10.1126/science.1168978 (2009).
339 Zhong, Y. *et al.* RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics (Oxford, England)* **33**, 139-141, doi:10.1093/bioinformatics/btw585 (2017).
340 Bakin, A., Lane, B. G. & Ofengand, J. Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry* **33**, 13475-13483 (1994).
341 Bakin, A. & Ofengand, J. Four newly located pseudouridylate residues in Escherichia coli 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **32**, 9754-9762 (1993).
342 Bakin, A. & Ofengand, J. Mapping of the 13 pseudouridine residues in Saccharomyces cerevisiae small subunit ribosomal RNA to nucleotide resolution. *Nucleic Acids Res* **23**, 3290-3294 (1995).
343 Bakin, A. V. & Ofengand, J. Mapping of pseudouridine residues in RNA to nucleotide resolution. *Methods in molecular biology (Clifton, N.J.)* **77**, 297-309, doi:10.1385/0-89603-397-x:297 (1998).
344 Ofengand, J. A CHEMICAL METHOD FOR THE SELECTIVE MODIFICATION OF PSEUDOURIDINE IN. *Biochemical and biophysical research communications* **18**, 192-201 (1965).
345 Ofengand, J., Del Campo, M. & Kaya, Y. Mapping pseudouridines in RNA molecules. *Methods* **25**, 365-373, doi:10.1006/meth.2001.1249 (2001).

346 Cao, Y., Cai, J., Li, X., Yuan, N. & Zhang, S. Autophagy governs erythroid differentiation both in vitro and in vivo. *Hematology (Amsterdam, Netherlands)* **21**, 225-233, doi:10.1179/1607845415y.0000000027 (2016).
347 Wei, Y. *et al.* The stress-responsive kinases MAPKAPK2/MAPKAPK3 activate starvation-induced autophagy through Beclin 1 phosphorylation. *eLife* **4**, doi:10.7554/eLife.05289 (2015).
348 Reynolds, E. S. The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *The Journal of cell biology* **17**, 208-212, doi:10.1083/jcb.17.1.208 (1963).
349 Presley, A. D., Fuller, K. M. & Arriaga, E. A. MitoTracker Green labeling of mitochondrial proteins and their subsequent analysis by capillary electrophoresis with laser-induced fluorescence detection. *J Chromatogr B Analyt Technol Biomed Life Sci* **793**, 141-150 (2003).
350 Pap, E. H. *et al.* Ratio-fluorescence microscopy of lipid oxidation in living cells using C11-BODIPY(581/591). *FEBS letters* **453**, 278-282, doi:10.1016/s0014-5793(99)00696-1 (1999).
351 Drummen, G. P., van Liebergen, L. C., Op den Kamp, J. A. & Post, J. A. C11-BODIPY(581/591), an oxidation-sensitive fluorescent lipid peroxidation probe: (micro)spectroscopic characterization and validation of methodology. *Free radical biology & medicine* **33**, 473-490 (2002).
352 Kerr, C. H. *et al.* IRES-dependent ribosome repositioning directs translation of a +1 overlapping ORF that enhances viral infection. *Nucleic Acids Res* **46**, 11952-11967, doi:10.1093/nar/gky1121 (2018).
353 Mauro, V. P. & Matsuda, D. Translation regulation by ribosomes: Increased complexity and expanded scope. *RNA Biol* **13**, 748-755, doi:10.1080/15476286.2015.1107701 (2016).
354 Yoon, A. *et al.* Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science* **312**, 902-906, doi:10.1126/science.1123835 (2006).
355 Heijnen, H. F. *et al.* Ribosomal protein mutations induce autophagy through S6 kinase inhibition of the insulin pathway. *PLoS Genet* **10**, e1004371, doi:10.1371/journal.pgen.1004371 (2014).
356 Turi, Z., Lacey, M., Mistrik, M. & Moudry, P. Impaired ribosome biogenesis: mechanisms and relevance to cancer and aging. *Aging* **11**, 2512-2540, doi:10.18632/aging.101922 (2019).
357 Yang, K., Yang, J. & Yi, J. Nucleolar Stress: hallmarks, sensing mechanism and diseases. *Cell stress* **2**, 125-140, doi:10.15698/cst2018.06.139 (2018).
358 Anding, A. L. & Baehrecke, E. H. Cleaning House: Selective Autophagy of Organelles. *Dev Cell* **41**, 10-22, doi:10.1016/j.devcel.2017.02.016 (2017).
359 Dixon, S. J. *et al.* Ferroptosis: an iron-dependent form of nonapoptotic cell death. *Cell* **149**, 1060-1072, doi:10.1016/j.cell.2012.03.042 (2012).
360 Hou, W. *et al.* Autophagy promotes ferroptosis by degradation of ferritin. *Autophagy* **12**, 1425-1428, doi:10.1080/15548627.2016.1187366 (2016).
361 Henson, A. L. *et al.* Mitochondrial function is impaired in yeast and human cellular models of Shwachman Diamond syndrome. *Biochemical and biophysical research communications* **437**, 29-34, doi:10.1016/j.bbrc.2013.06.028 (2013).

362 Harger, J. W., Meskauskas, A. & Dinman, J. D. An "integrated model" of programmed ribosomal frameshifting. *Trends Biochem Sci* **27**, 448-454 (2002).
363 Meskauskas, A. *et al.* Delayed rRNA processing results in significant ribosome biogenesis and functional defects. *Molecular and cellular biology* **23**, 1602-1613, doi:10.1128/MCB.23.5.1602-1613.2003 (2003).
364 Pulk, A. & Cate, J. H. Control of ribosomal subunit rotation by elongation factor G. *Science* **340**, 1235970, doi:10.1126/science.1235970 (2013).
365 Zhou, J., Lancaster, L., Donohue, J. P. & Noller, H. F. Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation. *Science* **340**, 1236086, doi:10.1126/science.1236086 (2013).
366 Tourigny, D. S., Fernandez, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, 1235490, doi:10.1126/science.1235490 (2013).
367 Noeske, J. & Cate, J. H. Structural basis for protein synthesis: snapshots of the ribosome in motion. *Current opinion in structural biology* **22**, 743-749, doi:10.1016/j.sbi.2012.07.011 (2012).
368 Chen, J., Tsai, A., O'Leary, S. E., Petrov, A. & Puglisi, J. D. Unraveling the dynamics of ribosome translocation. *Current opinion in structural biology* **22**, 804-814, doi:10.1016/j.sbi.2012.09.004 (2012).