The missing piece

Winkle, Melanie

CHAPTER 6

Summary, discussion and future perspectives

Melanie Winkle

Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

1 Summary

Although current knowledge on long noncoding (lnc)RNA function in normal and disease states it limited, it is clear that these RNA molecules play crucial roles in almost all cellular processes. LncRNAs can act at the transcriptional, posttranscriptional and protein level by binding to DNA, RNA and proteins or combinations thereof. In this study, we aimed to characterize lncRNA expression and function in B-cell lymphoma.

In **CHAPTER 2**, we reviewed current literature on the expression patterns and function of lncRNAs in normal B-cell development and in B-cell malignancies. In addition, we studied their potential value as clinical biomarkers for disease activity, monitoring treatment response and as therapeutic targets.

In normal B cells, distinct lncRNA expression patterns have been observed, with higher numbers of cell type-specific lncRNAs in comparison to the number of cell type-specific protein coding genes. Several lncRNAs had expression patterns related to cell fate-associated transcription factors, suggesting potential regulatory roles in B-cell development and maturation.

The number of studies analyzing differential expression patterns in B-cell lymphoma are limited, with some studies being available for classical Hodgkin lymphoma (cHL), diffuse large B cell lymphoma (DLBCL) and Burkitt lymphoma (BL). Some lncRNAs with prognostic value have been reported in DLBCL, mantle cell lymphoma (MCL) and chronic lymphocytic leukemia (CLL) in either tissue or plasma samples. The in generally observed tissue- and cell-type specific expression patterns makes lncRNAs potentially valuable biomarkers to monitor tumor load or therapy response, as prognostic markers or as therapeutic targets. However, further research related to their clinical use is warranted.

A subset of the differentially expressed lncRNAs is located at genomic regions that are recurrently amplified in lymphoma, supporting their role in the pathogenesis. In addition, several of the lncRNAs were co-expressed with oncogenic B-cell lymphoma associated transcription factors. Of interest, several studies support a crucial role for Myc in the regulation of lncRNA expression in B-cell lymphoma. LncRNAs associated with Myc include HOTAIR, GAS5, PVT1 and MINCR. For a limited number of B-cell lymphoma associated lncRNAs, functional roles have been established. For example, decreased FAS-AS1 induced the production of sFAS and prevented FAS-ligand induced apoptosis. LincRNA-p21 regulated the expression of p21 in *cis* and ectopic expression of lncRNA-p21 caused an increase in p21 and G1 cell cycle arrest. LUNAR1 levels were linked to cell cycle progression, whereas MALAT1 knockdown caused increased apoptosis, cell cycle arrest and autophagy.

Altogether, is has become evident that lncRNAs are significantly involved in B-cell lymphoma biology. Some lncRNAs show potential as therapeutic (co-)targets, particularly those that are actively involved in the mediation of chemotherapeutic effects.

In **CHAPTER 3** we addressed lncRNA expression patterns in normal sorted B-cell populations and assessed the lncRNA expression pattern in HL cell lines. HL is a malignancy of germinal center (GC) B-cell origin, characterized by a low tumor cell percentage, which makes it complicated to study lncRNA expression patterns in primary tissue samples.

Within the three normal B cell subsets, we identified 401 significantly differentially expressed lncRNA probes with a fold change in expression level of at least 2. Most of these differences were based on the altered levels observed in GC-B cells, which showed significant different levels to naïve B cells for 377 probes and to memory cells for 378 probes. In contrast, only 2 probes were significantly different between naïve and memory B-cells. The nodular lymphocyte predominant Hodgkin lymphoma (NLPHL) cell line clustered in the same branch as the GC-B cells, while the 5 cHL cell lines clustered in a separate branch next to the GC-B cells. This is in line with the GC-B cell origin of the tumor cells of HL, with a loss-of-B-cell phenotype being more prominent in cHL as compared to NLPHL. The naïve and memory B cells clustered in a different branch.

Comparison of the expression profiles of cHL to GC-B cells revealed a significantly differential expression for 639 lncRNA probes, with a decreased expression level for 74% and an increased expression level for 26%. In contrast, mRNAs showed a more equal distribution between up and downregulated transcripts. For three of the upregulated lncRNAs we performed RNA-FISH to study their expression pattern in tissue sections of primary cHL cases. All three lncRNA probes showed a tumor cell-specific staining pattern. To further explore their relevance for B-cell lymphoma we analyzed the expression patterns in a panel of lymphoma cell lines. This revealed, besides high levels in HL, also high levels in DLBCL and EBV transformed lymphoblastoid cell lines for two lncRNAs. The third lncRNA showed besides high levels in HL cell lines also high levels in BL cell lines.

To identify lncRNAs that potentially affect the expression of nearby protein-coding genes in *cis* we determined which of the differentially expressed lncRNAs was close to a differentially expressed mRNA. A potential *cis*-regulatory role was observed for 51 lncRNA-mRNA pairs that both showed differential expression between the normal B-cell subsets and were localizing within a 60-kb region. Similar analysis between cHL vs GC-B cells revealed 57 of such lncRNA-mRNA pairs. Three of these putative *cis*-regulated lncRNA-mRNA pairs showed a positive correlation in both analyses. To support a putative *cis*-regulatory role we determined whether the lncRNAs were preferentially localized in the nucleus. In line with a *cis*-regulatory role, we indeed

showed enrichment in the nuclear fractions and depletion in the cytoplasmic fractions for two of the three lncRNAs. Analysis of the putative *cis*-regulated protein-coding genes by qRT-PCR confirmed the inverse expression pattern between *LINC00461* and the nearby upstream protein-coding gene, *MEF2C*. For FLJ42351, we could confirm the comparable expression patterns for this lncRNA and the neighboring protein-coding gene *SLC20A1*.

In summary, our comprehensive expression analysis showed dynamic regulation of lncRNA expression during the GC transition of B-cells and a widely deregulated expression in HL. The identified lncRNAs, together with the putative *cis*-acting data, is a valuable source for further studies aiming at understanding the role of lncRNAs in normal B-cell biology and in the pathogenesis of HL. Using RNA-FISH we could show for three lncRNAs a tumor cell-specific expression pattern in HL tissues, which might indicate their value as a disease biomarker.

Myc is a well-known transcription factor with important roles in cell cycle, apoptosis, and cellular transformation. Myc regulates a large number of protein coding genes and microRNAs. In **CHAPTER 4**, we explored to what extend Myc regulates lncRNAs using the model system P493-6, a B cell line which harbors a tetracycline-repressible *MYC* allele.

We performed two independent biologic replicates of the MycOFF and MycON state and Myc on for 4 and 24 hours to discriminate between early and late MYC-induced differences. We identified 1,244 lncRNA loci with a significant differential expression between MycOFF and any of the 3 MycON time points. GSEA and GO analysis of the Myc-regulated mRNAs revealed a significant enrichment of a previously identified Myc-responsive protein coding gene set and Myc-related gene ontologies. Unsupervised hierarchical clustering revealed a pairwise clustering of MycOFF with Myc on for 4 hours and of MycON with Myc on for 24 hours. An early Myc response was observed for 30% of the lncRNA probes. Similar Myc-response patterns were observed for protein coding genes. These results demonstrate that lncRNAs are a main component of the Myc-regulated transcriptional program. Most of the Myc-induced protein coding and lncRNA genes were already expressed in the MycOFF state, albeit at lower levels. These data are consistent with the notion that Myc acts as a general amplifier of expression rather than inducing expression of genes that are not expressed when Myc levels are low.

Myc binding site analysis using published Myc ChIP data sets showed a highly significant enrichment for binding sites in the Myc-induced lncRNA and protein coding gene sets. In contrast, Myc-repressed coding genes did not show any enrichment for Myc binding sites, whereas Myc-repressed lncRNAs did show a significant enrichment for Myc binding sites, especially in the 4- and 24-hour samples.

To study the subcellular localization of Myc-regulated lncRNAs, we analyzed cytoplasmic and nuclear fractions of P493-6 cells. LncRNAs significantly more often showed a specific subcellular localization compared to mRNAs. Of the lncRNAs with a specific subcellular localization, >60% showed enrichment in the nuclear fraction. This was most pronounced for the Myc-repressed lncRNAs. Next, we explored putative lncRNA-mediated regulation of nearby protein coding genes in *cis*. Using a probe-to-probe distance of 20kb, we identified 105 *cis*-regulatory lncRNA candidates. A sense orientation was observed for 38, a tail-to-tail for 36, a head-to-head for 18, an antisense for 12, and an intronic localization for one lncRNA-mRNA pair. The direction of expression was concordantly up- for 41, concordantly down for 32 pairs and inversed for 32 pairs.

To establish the potential relevance of these Myc-regulated lncRNAs we also studied their expression pattern in primary B-cell lymphoma with low and high Myc levels. For the vast majority of the lncRNAs we indeed observed an expression pattern consistent with the Myc levels, with low expression for Myc-repressed lncRNAs and *vice versa*, high expression for Myc-induced lncRNAs in BL compared to CLL. A Myc-dependent regulation in BL, was further supported by decreased expression of validated Myc-induced lncRNAs in two Myc-shRNA treated BL cell lines.

This study showed that lncRNAs are a main component of the transcriptional program regulated by Myc. Of the >1,200 Myc-regulated lncRNAs identified, 105 have a potential *cis*-regulatory function. In contrast to Myc-repressed mRNAs, Myc-repressed lncRNAs are potentially directly regulated by Myc.

In **CHAPTER 5** we further deepen the understanding of Myc-regulated lncRNAs by studying Myc knockdown in BL cell lines as an additional experimental approach to identify Myc target genes relevant for BL pathogenesis. In addition, we performed an analysis of their Myc binding profile employing Myc and Max ChIP datasets and CpG island coordinates.

BL cell lines were as expected dependent on Myc; as Myc knockdown resulted in a highly significant decrease in growth, with the most prominent effect in ST486 cells. Gene expression profiling was performed to identify Myc-dependent gene expression changes in ST486 cells. Efficiency of our knockdown approach was confirmed by gene set enrichment analysis (GSEA) showing a significant enrichment of Myc target gene sets in control over knockdown samples. The number of significantly Myc-induced and Myc-repressed coding genes was similar, whereas the number of Myc-induced lncRNA loci was much lower than the number of Myc-repressed lncRNA loci. A Myc-dependent regulation was confirmed by analyzing their expression in lymphoma samples with high and low Myc levels. For the vast majority of the lncRNAs the observed expression pattern was consistent with the expectation based on Myc levels in these samples. These differences were significant for more than 20% of them. Reanalysis of our

previous P493-6 data also showed the expected pattern with significant differences for 40% of the lncRNAs.

Analysis of all gene loci, revealed increased presence of CpG islands and Myc as well as Max binding sites near coding genes as compared to noncoding genes. The Myc-induced and Myc-repressed coding genes were significantly enriched for CpG islands, Myc and Max binding sites, with a more pronounced enrichment for Myc-induced coding genes. Analysis of Myc-regulated noncoding loci revealed enrichment of CpG islands, Myc and Max binding sites in Myc-repressed lncRNAs, while Myc-induced lncRNAs were enriched for CpG islands and Myc binding sites, but not for Max binding.

To identify high confidence Myc-regulated lncRNAs, we overlapped the lncRNAs up- and downregulated by Myc in ST486 cells with early response (within 4h after Myc induction, 445 lncRNA loci) lncRNAs as identified in P493-6 cells and noncoding gene loci with a proven Myc binding in BL. This resulted in six high confidence Myc-regulated targets including two Myc-induced and four Myc-repressed lncRNAs. Analysis of a broader cell line panel for the Myc-induced lncRNA, KTN1-AS1, revealed increased levels also in HL and DLBCL. In primary BL cases, KTN1-AS1 levels were as expected higher than in CLL cases. Within the DLBCL cases levels of KTN1-AS1 were higher in GCB type as compared to ABC type. In normal B-cell subsets, KTN1-AS1 expression levels were slightly increased in GC-B cells compared to naïve, but not compared to memory B cells.

To obtain insight in the relevance of KTN1-AS1 we knocked down its expression in BL. This induced a marked decrease in growth, which was not caused by an increase in apoptosis. To unravel the underlying mechanism we studied its subcellular localization. KTN1-AS1 showed a strong enrichment in the nuclear fraction, with preferential localization in the chromatin fraction hinting at a putative function in gene expression regulation. Analysis of the expression levels of KTN1, mapping close to KTN1-AS1, upon knockdown of KTN1-AS1 revealed no changes, excluding a regulation in *cis*. Genome wide gene expression changes upon KTN1-AS1 depletion revealed significant differences for 236 coding and 63 noncoding loci, with half of them being up and the other half being downregulated. GSEA and Enricher analysis indicated, similar to what was observed upon Myc knockdown, a clear enrichment of Myc-related gene sets and pathways related to metabolism. To establish a potential direct effect of KTN1-AS1 on Myc, we analyzed Myc protein levels in KTN1-AS1 knockdown cells. This revealed a decrease in Myc levels compared to control samples suggesting that KTN1-AS1 regulates Myc expression. Consistent with this potential positive feedback loop between Myc and KTN1-AS1 we also observed a highly significant enrichment for CpG islands as well as Myc and Max binding in the KTN1-AS1 protein coding gene set, similar to what we observed for Myc-regulated coding genes.

Altogether, our work shows that the Myc-induced lncRNA KTN1-AS1 is an oncogene in B-cell lymphoma. Our functional data suggest the presence of a reinforcing feedback

loop between MYC and KTN1-AS1 and suggest that KTN1-AS1 depletion induces a growth inhibitory phenotype mainly through downregulation of Myc and affecting a subset of the Myc target genes involved in metabolism.

2 Discussion

2.1 GC B and HL co-expression networks

Our custom array design broadly covers both coding and noncoding genes. This allows co-expression analysis and identification of lncRNAs with patterns similar to protein coding genes with known functions. These lncRNAs are likely to function in the same pathways or networks (guilt-by-association). In addition, identification of binding of known transcription factor binding to the lncRNA loci of interest, might further indicate their relevance in certain biological processes or disease states. For the data described in **CHAPTER 3** we performed gene set enrichment analysis with a focus on binding of transcription factors, pathways and gene ontologies to get a first indication of the putative functions of the differentially expressed lncRNAs (**TABLE1**). In accordance with the high proliferative phenotype of GC B cells compared to the quiescent naïve or memory B cells, genes differentially expressed between these two groups show a strong enrichment for cell cycle processes as well as for binding of cell cycle regulators (i.e. E2F4, FOXM1). In addition, we observe binding of SOX2, a gene expressed in activated B cells that have undergone class switch recombination and thought to repress AID expression[1]. DNA repair associated genes are also enriched, in line with the somatic hyper mutation machinery and class switch recombination processes that are actively involved in maturation of GC B cells in the GC reaction. Genes with a strong regulation at this step of B cell maturation might also contribute to malignant transformation by facilitating a state that allows introduction of oncogenic aberrations in the B cells. This is supported by the fact that the HL cells cluster in the same branch as the GC B cells as well as by the GC B cell origin postulated for many of the B-cell lymphoma subtypes.

Genes deregulated in HL cells compared to GC show enrichment of transcription factors implicated in B cell malignancy, i.e. TCF3 and ZBTB7A. TCF3 is located in a susceptibility locus for HL[2] and thought to be involved in HL pathogenesis[3]. Loss of TCF3 expression has been linked to the loss of B cell phenotype frequently observed in HL, and this might potential aid escape of apoptosis during the GC reaction of the HRS precursor cells. ZBTB7A (also LRF, Pokemon) is a known oncogene in B cells[4] and other cancers and plays essential roles in malignant transformation[5]. Remarkably, ZBTB7A is one of the two transcription factors regulating expression of TCF3. In addition, we find enrichment of PD-1 signaling, which is known to be hyperactive in classical HL and blocking of this pathway shows promising results in HL[6].

In summary, the protein coding genes differentially expressed between normal B-cell subsets and between HL and GC B cells are in good accordance with their known functions. The lncRNA loci deregulated in HL cells lines may be related to oncogenic signals provided by TCF3, ZBTB7A and PD-1 signaling. Further indications can be obtained by performing a more detailed co-expression analysis and by lncRNA expression analysis upon perturbing individual pathways or transcription factors.

TABLE 1 Analysis of regulation and function of protein coding genes specifically expressed in GC B cells and deregulated in HL cells.

GB B vs naïve/memory B	HL cell lines vs GC B
ChEA/ENCODE transcription factor binding sites (Top 5)	
E2F4_ENCODE (****)	UBTF_ENCODE (****)
FOXM1_ENCODE (****)	TCF3_ENCODE (****)
SOX2_CHEA (****)	ZBTB7A_ENCODE (****)
GATA1_CHEA (****)	GATA1_CHEA (****)
TCF3_CHEA (****)	RCOR1_ENCODE (****)
Reactome Pathways (Top 3)	
Cell Cycle_Homo sapiens_R-HSA-1640170 (****)	PD-1 signaling_Homo sapiens_R-HSA-389948 (*)
Cell Cycle, Mitotic_Homo sapiens_R-HSA-69278 (****)	Phosphorylation of CD3 and TCR zeta chains_Homo sapiens_R-HSA-202427 (*)
M Phase_Homo sapiens_R-HSA-68886 (****)	Immune System_Homo sapiens_R-HSA-168256 (*)
GO Biological Process (Top 3)	
negative regulation of transcription from RNA polymerase II promoter (GO:0000122) (****)	positive regulation of transcription initiation from RNA polymerase II promoter (GO:0060261) (***)
regulation of transcription involved in G1/S transition of mitotic cell cycle (GO:0000083) (****)	positive regulation of transcription from RNA polymerase II promoter in response to stress (GO:0036003) (***)
DNA repair (GO:0006281) (****)	positive regulation of pri-miRNA transcription from RNA polymerase II promoter (GO:1902895) (***)

For each category the top three to five significant hits (if available) are shown (cutoff $p<0.05$). 2,467 genes differentially expressed in GC B vs naïve/memory B cells, and 2,021 genes differentially expressed between GC B and HL cell lines were used for the analysis. (*) $p<0.05$, (**) $p<0.01$; (***) $p<0.001$; (****) $p<0.0001$, adjusted p-value as calculated by Enrichr.

2.2 Myc *in vitro* modeling systems

To identify Myc-regulated lncRNAs, we used two B-cell *in vitro* systems, namely the P493-6 model with conditional Myc expression (**CHAPTER 4**) and Myc knockdown in the BL cell line ST486 (**CHAPTER 5**). We detected more pronounced expression changes and a much larger number of coding and noncoding genes responding to Myc in the P493-6 model (6,555 mRNA and 2,014 lncRNA probes; 2-fold change) as compared to ST486 cells (1,153 mRNA and 387 lncRNA probes; 1,5-fold change). These differences in expression changes may be explained by the more pronounced difference between Myc levels in the Myc-off vs Myc-on conditions observed in P493-6 as compared to the incomplete knockdown (~70% reduction) achieved in ST486 cells. This is consistent with the notion that extremely high levels of Myc were shown to regulate a larger amount of genes, i.e. additional genes carrying non-canonical E-box sequences[7]. In addition, Myc-repression in P493-6 caused a significant G1 cell cycle arrest, indicating that genes differentially expressed between Myc-on and Myc-off states also include transcripts that show a dynamic regulation during the cell cycle. Cell cycle distribution was not affected by Myc knockdown in ST486 cells (data not shown).

Results of the enrichment of transcription factor binding, pathways and GO terms of three separate (protein coding) gene sets identified in P493-6 cells is shown in **TABLE 2**. These include: (**1**) 'P493-6 specific' (genes responding to Myc in P493-6 but not in ST486 cells), (**2**) 'P493-6 early responders' (genes responding within 4h after Myc re-induction), and (**3**) 'P493-6/ST486 shared' (genes Myc-responsive in both models). As expected, all sets showed a clear enrichment of Myc and Max binding sites. Early Myc-responders were also enriched for binding by Sin3a, a histone deacetylase and known co-repressor of Max[8], suggesting that gene repression is an active and immediate result of Myc induction. P493-6 specific genes were additionally enriched for NFYB, NFYA and E2F4. The heterotrimeric transcription factor Nf-y composed of NFYA, -B and -C subunits is involved in the regulation of cell cycle specific genes[8, 9] and thought to interact with Myc to cause gene repression[10]. E2F4 is a transcriptional activator of cell cycle and DNA replication associated genes. The other two sets show enrichment of E2F6 binding, a transcriptional repressor of G1 phase associated genes necessary to allow S-phase entry[11]. Lastly, shared genes were enriched for ZBTB7A, a known oncogene in B cells and an essential factor for Myc-driven malignant transformation[4, 5]. Pathway and GO term analysis gave a similar picture with a strong over-representation of cell cycle associated terms in P493-6 specific genes. P493-6 early responders show enrichment of rRNA processing and interferon signaling likely reflecting the influence of Myc on the rate of protein biosynthesis and the repression of B-cell specific functions, respectively. P493-6/ST486 shared genes were enriched for rRNA processing as well as nucleotide metabolism, arguing for roles in RNA and protein synthesis. Thus, the fraction of Myc-regulated genes that we identified as unique for P493-6 cells (**CHAPTER4**) is mainly involved in and/or influenced by the cell cycle. These genes were significantly occupied by Myc and Max, but also by other cell cycle regulators such as Nf-y subunits and

E2F4. Genes that respond to Myc more immediately (P493-6 early responders) do not show this enrichment, suggesting the strong cell cycle overrepresentation is in part an indirect effect. Genes shared between the models (as used in **CHAPTER 5**) also show enrichment of other oncogenic factors actively affecting proliferation (E2F6, ZBZB7A).

In summary, the intersection of Myc-responsive genes from the two *in vitro* models used in this work has resulted in a selection against cell cycle-driving and an enrichment of proliferation-driving Myc target genes. As the lncRNA transcripts co-expressed with these protein networks are likely involved in similar processes, our approach is highly likely to identify oncogenic or tumor suppressive lncRNAs. The in this thesis further studied lncRNA KTN1-AS1 is a good example of such a lncRNA and others will likely be identified by in-depth analysis of additional identified candidates.

TABLE 2 Different functional pathways and GO terms enriched in Myc-regulated genes unique to P493-6 cells or shared between P493-6 and ST486 cells.

(1) P493-6 specific	(2) P493-6 early responders	(3) P493-6/ST486 shared
ChEA/ENCODE transcription factor binding sites (Top 5)		
MAX_ENCODE (****)	MYC_ENCODE (****)	MAX_ENCODE (****)
MYC_ENCODE (****)	MAX_ENCODE (****)	MYC_ENCODE (****)
NFYB_ENCODE (****)	SIN3A_ENCODE (****)	MYC_CHEA (****)
E2F4_ENCODE (****)	MYC_CHEA (****)	ZBTB7A_ENCODE (****)
NFYA_ENCODE (****)	E2F6_ENCODE (****)	E2F6_ENCODE (****)
Reactome Pathways (Top 3)		
Cell Cycle_Homo sapiens_R-HSA-1640170 (****)	rRNA modification in the nucleus_Homo sapiens_R-HSA-6790901 (****)	Metabolism of nucleotides_Homo sapiens_R-HSA-15869 (**)
Cell Cycle, Mitotic_Homo sapiens_R-HSA-69278 (****)	rRNA processing_Homo sapiens_R-HSA-72312 (**)	
Gene Expression_Homo-sapiens_R-HSA-74160 (****)	Interferon alpha/beta signaling_Homo sapiens_R-HSA-909733 (**)	
GO Biological Process (Top 3)		
mitochondrial translational elongation (GO:0070125) (****)	maturation of LSU-rRNA (GO:0000470) (***)	maturation of LSU-rRNA (GO:0000470) (***)
DNA replication (GO:0006260) (****)	type I interferon signaling pathway (GO:0060337) (**)	cleavage involved in rRNA processing (GO:0000469) (**)
DNA synthesis involved in DNA replication (GO:0090592) (****)	rRNA processing (GO:0006364) (**)	maturation of 5.8S rRNA (GO:0000460) (**)

For each category the top three to five significant hits (if available) are shown (cutoff p<0.05). 4,903 P493-6 specific (1), 308 P493-6 early responders (2) and 678 P493-6/ST486 shared (3) protein coding genes were used for the analysis. (*) p<0.05, (**) p<0.01; (***) p<0.001; (****) p<0.0001, adjusted p-value as calculated by Enrichr.

2.3 Myc and Max co-occupancy and co-factors

Intrigued by the differences observed in Myc binding between coding and noncoding loci as observed in **CHAPTERS 4 and 5**, we further investigated co-occupancy of Myc binding with specific co-factors at the Myc-regulated loci. However, this analysis is hampered by the lack of publically available data for Myc co-factors as well as the comparison of data from different, though similar cell types (i.e. BL cell line ST486 expression data and GM12878 lymphoblastoid cell ChIP data).

A well-established binding partner of Myc is Max. Co-occupancy of binding sites with Myc / Max dimers leads to activation of gene expression. Less well-studied co-factors of Myc include Sp1 and Nf-Y members. Co-binding with these factors results in repressing of gene expression. Interaction of Sp1 with Myc leads to repression of p15 and p21, this process also involves Smad proteins[12, 13]. At the PDGF gene locus, Myc was shown to interact with two of the three Nf-y members, Nf-yb and Nf-yc, to repress expression[10]. Genome wide data about the relevance of these factors in relation to Myc-repressed genes are not available. No publicly available data are available for the best-characterized Max interactor, Miz1, this limits current analysis regarding binding of Max and its co-factors. Mxi1, a direct interactor of Max, competes with Myc for binding to Max. The Max / Mxi1 complex also leads to repression of gene expression. The Max / Mxi1 dimer was shown to act at a genome wide scale on multiple gene loci and competes for binding with the activating Myc / Max complex. Mxi1 additionally recruits the co-repressors Sin3a and Sin3b, which then attract chromatin modifiers to stably repress gene expression.

As we observed a much stronger enrichment of Myc binding in repressed noncoding compared to repressed coding loci, we hypothesized that Myc-dependent repression mechanisms may be different. To investigate this, we first checked the co-occupancy of Myc and Max at Myc- induced and Myc-repressed loci (**FIGURE 1A**). This revealed a moderate enrichment of Myc/Max co-occupancy at Myc-induced, but not at Myc-repressed coding loci, in line with the activating nature of this interaction. Conversely, for lncRNA loci we observed a depletion of Myc/Max binding at induced loci and enrichment at repressed loci. This is in line with the observed lack of Max binding enrichment for Myc-induced loci as shown in **CHAPTER 5**. However, it remains unclear why Myc-induced noncoding loci are depleted for Myc / Max co-occupancy. This suggests binding of Myc together with other activating co-factors. For the repressed noncoding loci, these findings suggest that despite the observed enrichment, it is unlikely that Myc / Max dimers are responsible for the decrease in expression.

A further analysis of co-occupancy for repressive factors revealed that Myc-repressed coding genes were more often co-enriched for Myc and Sp1 (75%) than for Myc and Nf-y (45%). This potential co-occupancy in ST486 cells was not much higher than expected by chance (control, 73 and 52% respectively; **FIGURE 1B**, left). Noncoding loci

were significantly more often co-occupied by Myc / Sp1 (75% vs. 49% expected by chance), while Myc / Nf-y co-occupancy was depleted rather than enriched at Myc-repressed noncoding loci (22% vs. 30% expected by chance; FIGURE 1B, right). At both coding and noncoding loci, Sp1 and Nf-y bound a larger number of genes alone than in combination with Myc, indicating additional co-factors may play a role in the repression of Sp1 and Nf-y bound gene loci.

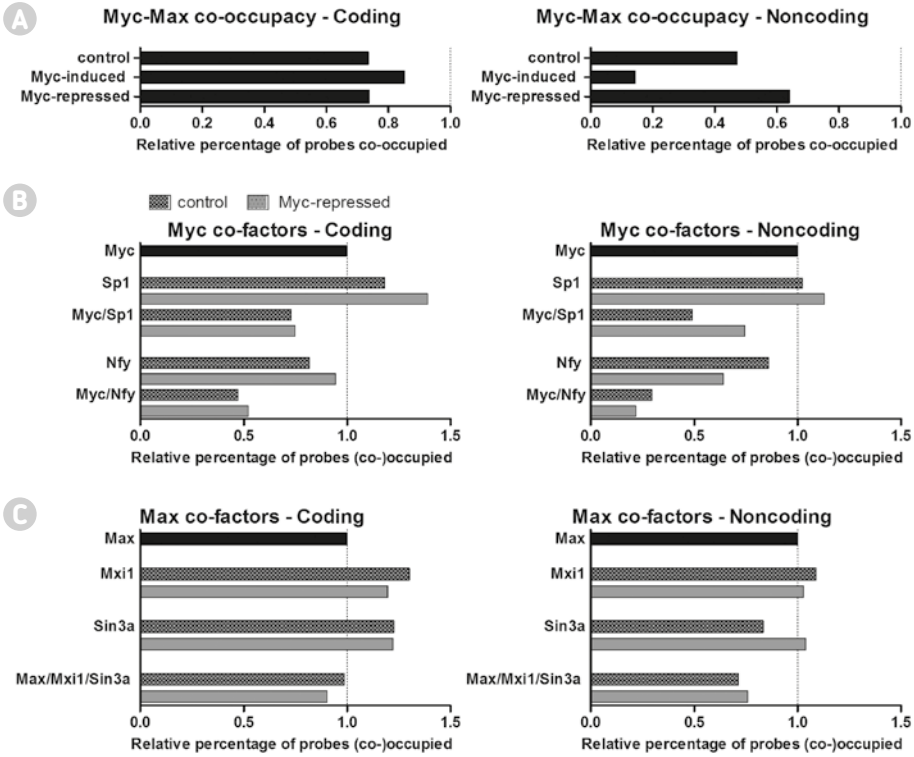

FIGURE 1 Co-occupancy patterns at Myc-repressed coding and noncoding loci. (A) Amount of control, Myc-induced or Myc-repressed probes co-occupied by Myc and Max relative to all genes occupied by MYC alone (= 1). **(B)** Binding of Myc co-factors alone or in combination with Myc at Myc-repressed coding and noncoding loci. The relative amount of genes bound by Myc alone at Myc-repressed or control loci is set to 1. **(C)** Same as in **(B)** for Max repressive interactors. The relative amount of genes bound by Max alone at Myc-repressed or control loci is set to 1.

Considering all repressed gene loci bound by Max, we observed for most of these loci co-occupancy by Max, Mxi1 and Sin3a. For coding genes (FIGURE 1C, left) this co-occupancy was slightly less than expected by chance (90% of Myc-repressed loci vs. 99% in control). In noncoding gene loci, Max/Mxi1/Sin3a co-occupancy was slightly higher at Myc-repressed loci (76%) vs. control (71%). Binding of Myc-repressed noncoding loci by Sin3a alone was observed more often than expected (104% vs 83%), suggesting additional recruitment mechanisms for Sin3a. Together, these data suggest

that Myc may repress noncoding loci more prominently through co-binding with direct Myc co-factors such as Sp1. Myc-repressed coding loci were most often co-occupied by Max/Mxi1/Sin3a and a Max-mediated repression mechanism would be in line with the stronger enrichment of Max binding compared to Myc binding at repressed coding loci. Of note, the binding of Myc/Max vs. Max/Mxi1/Sin3a occurs dynamically at the same loci depending on competition, therefore binding of both complexes may be observed in heterogeneous cell populations.

While these analyses may explain some of the observed differences in Myc binding between Myc-repressed coding and noncoding loci, they do not provide a complete answer. The factors considered here, are the few Myc and Max interactors for which ChIP data is publically available for a cell type similar to the cell types as used in our studies (i.e. GM12878 lymphoblastoid cells). In addition, the repressive mechanisms of Myc/Sp1 and Myc/Nf-y have been characterized only for specific loci such as p15, p21[12, 13] and PDGF[10], leaving it an outstanding question whether these mechanisms apply broadly or are specific to particular loci. As our data clearly indicate prevalent differences in Myc binding at Myc-repressed coding and noncoding loci, further studies on the precise repressive mechanisms are of great interest. ChIP sequencing for additional co-factors in the same cell lines as used here, as well as the disruption of canonical Myc-binding sites at repressed loci will provide further insight into the mechanisms involved.

3 Future Perspectives

3.1 Integration of lncRNAs into systems biology

The discovery of lncRNAs caused a vast expansion of our knowledge of cell biology, as we know it, adding additional pieces to an already elaborate puzzle. In order to systematically integrate these pieces, it is of utmost importance to map the human lncRNA repertoire, characterize their expression patterns and analyze their co-expression networks. The integration of lncRNAs into systems biology is in full progress and will ultimately support the understanding of their intricate roles inside the cell. There are three major challenges in this quest, namely the low, species (i.e. primate) specific and cell type specific expression levels of lncRNAs. These factors ask for unique approaches to successfully characterize lncRNAs and these differ from what is commonly applied for protein coding genes: (I) A higher sequencing depth is required to reliably detect low abundant transcripts via RNA sequencing; (II) Studies should focus on human samples rather than murine models due to the absence or limitated evolutionary conservation for a vast proportion of the lncRNAs; and (III) Carefully sorted, homogeneous cell populations rather than bulk samples are needed to reliably detect cell type specific transcripts.

Studies applying multiple approaches, including the lncRNA-specific re-analysis of available RNA sequencing data as well as the generation of new data for the identification of novel lncRNAs in bulk tissues and specific cell types are in progress. In addition, several lncRNA databases are being generated to facilitate knowledge distribution on lncRNA expression (e.g. LNCipedia) and deregulated expression patterns of lncRNAs in disease states (e.g. MiTranscriptome). Future efforts should focus on the systematic annotation of lncRNA expression patterns in the GENCODE project reference cell lines (i.e. GM12878, hESCs, HSMM, HUVEC, K562, NHEK, NHLF) to allow combination with all other -omics data sets. This will allow effective linking of e.g. epigenetic features and transcription factor binding profiles to lncRNA loci in a cell type specific manner. To this end, it is necessary to develop general protocols that should consider not only sequencing depth, but also the methods of RNA isolation and alignment strategies: LncRNAs can lack polyadenylation and thus require alternative methods for the depletion of ribosomal RNA. In addition, lncRNAs may be expressed at a low level and localize exclusively in the nucleus, making it difficult to distinguish native, intron-containing from mature transcripts. This as well as the high repeat content within lncRNA loci requires the adjustment of commonly used bioinformatic mapping protocols. To make a good distinction between native and mature transcripts, alternative techniques such as PacBio or Nanopore sequencing that generate reads up to 60kb are available. These techniques also allow for a better characterization of the different splice variants that may be expressed from each locus. Furthermore, strand specific sequencing is essential to reliably distinguish between sense / antisense transcripts. The application of global run-on (GRO-)sequencing[14] that specifically identifies nascent transcripts may furthermore be helpful to identify short-lived lncRNA species. Thus, optimally lncRNA expression profiles should be generated using multiple sequencing methods in parallel in a multi-center approach as also applied within the GENCODE project. In an era where single cell RNA sequencing is becoming a routinely applied technique, we can expand our knowledge and truly appreciate the complexity of what was previously considered to be a homogeneous cell population[15, 16]. Indeed, the first single cell sequencing studies focusing on lncRNAs indicate that transcripts being low abundant in bulk samples may be high abundant in individual cell types[17] or during specific cell development stages[18]. Such lncRNA transcripts should thus be considered as uncommon rather than lowly expressed[19]. Lastly, the function of chemical modifications on lncRNAs such as methylation is a completely unexplored field. The diversity of known (i.e. 'epi-transcriptomic') RNA modifications is likely to increase with the further characterization of lncRNAs and requires its own mapping and characterization process.

Next to the identification and characterization of cell type specific lncRNA expression patterns, analysis of their co-expression networks with protein coding genes is of major importance. Concordant expression changes of lncRNAs and proteins with known regulatory functions may indicate in which pathway the lncRNA acts and whether its deregulation might cause or contribute to maintain disease. To this end,

normal cellular processes as well as disease-related networks should be investigated. Given the predominantly translation-driven nature of current scientific research, it is not suprising that the first co-expression databases (e.g. LncMAP and lnCaNet) are related to oncogenic networks rather than to normal cellular functions. A comprehensive characterization of co-expression requires extensive and highly pure sample collections, preferably using single cell sequencing approaches. To generate cancer specific lncRNA-mRNA co-expression data sets matched normal and tumor cell samples should be analyzed in parallel.

Integration of lncRNA research into systems biology is an elaborate, time-consuming and costly quest that faces many hurdles. Nevertheless, a comprehensive integration of lncRNA data into standing data collections will be essential to generate a solid basis for future lncRNA research.

3.2 Evolutionary versus revolutionary lncRNAs

An outstanding question remains which of the huge number of lncRNAs currently mapped to the human genome represent functional entities, i.e. 'revolutionary' lncRNAs, as opposed to newly developing transcripts that have lost their function, i.e. 'evolutionary' lncRNAs. Three strategies may be helpful to discriminate between them: (**I**) Evolutionary conservation of sequence motifs, regulatory sequences and transcriptional status across shorter evolutionary distances: The lack of overall sequence conservation in lncRNA sequences does not imply a complete lack of selective pressure. Analysis of smaller sequence regions, splice sites, secondary structures as well as promoter integrity and transcriptional status may also reveal conserved features. In addition, evolutionary novel lncRNAs should be conserved to a higher degree with more closely related species such as chimpanzee and human[20, 21]. (**II**) Tight regulation of lncRNA expression by transcription factors and epigenetics: As a functional entity should be tightly regulated, the transcriptional and epigenetic regulation or the lack thereof can be informative to define functional lncRNAs. Accordingly, specific regulation of lncRNA genes is being studied broadly, particularly for well-known oncogenes and tumor suppressors, such as Myc[22-24], NOTCH1[25], STAT3[26] and p53[27-29]. In addition, epigenetic marks on lncRNA loci can distinguish between individual transcriptional units (H3K4me3 rich), enhancer associated lncRNAs (H3K4me1 rich) and inactive loci (e.g. H3K27me3). (**III**) Expression of transcriptional noise in permissive cell types: Many lncRNAs are expressed specifically in testis, an observation likely due to the transcriptionally permissive state in this tissue. In the post-meiotic stages of spermatogenesis, epigenetic features including DNA methylation and histone proteins are temporarily removed to allow genome re-organization. This may be considered an aberrant transcriptional program that leads to transcriptional noise. So, in general lncRNAs solely expressed in testis might represent nonfunctional lncRNAs. However, subsets of testis specific lncRNA transcripts can have active roles in spermatogenesis. Knockout of testis specific long noncoding RNA 1 (Tslrn1) expressed from the X chromosome resulted in a reduction of spermatozoa

production[30]. As cancer cells are also transcriptionally more active than normal cells, subsets of testis-specific lncRNAs can become deregulated in cancer tissues and this may cause genome instability and increase malignancy[31]. This was shown for the testis specific lncRNA THOR, whose expression lead to stabilization of IGF2BP1 expression in cancer cells thereby supporting proliferation[32].

Though these considerations can hint towards functionality, there is at this point no general straightforward approach to distinguish functional from nonfunctional lncRNA transcripts other than their experimental modulation and identification of the induced phenotypic changes. Large-scale genomic activation or repression studies need to be performed to explore the functionality of lncRNA, in the coming years.

3.3 Experimental approaches to study lncRNA function

To characterize the function of the human noncoding genome, functional knock out screening approaches can be highly informative. However, conventional screenings using RNA interference may not be optimal for lncRNAs as their predominant localization in the nucleus may prevent efficient downregulation. The more recently discovered CRISPR/Cas9 approaches for genome editing[33] has been a breakthrough also for lncRNA studies. CRISPR/Cas9 based approaches allow manipulation of selected genes at the DNA and RNA level in various ways. At the DNA level, these include the introduction of random small insertions or deletions at a specific site using one single guide [sg]RNA or insertion or deletion of specific sequences using two sgRNAs. At the level of transcription, gene expression can be modulated using nuclease-inactive or 'dead' Cas9 (dCas9) linked to an activating or repressive domain[34]. CRISPR/Cas9 systems are very suitable for large scale screening using sgRNA libraries and have been applied successfully for protein coding genes. However, for lncRNAs the question is whether simple random small indels at a specific site in a lncRNA as achieved with the standard CRISPR/Cas9 system lead to a reduction in its functionality. For these reasons the dCas9 system using a dead Cas9 protein fused to an activating or repressive domain may be preferred here. However, the magnitude of induction or repression is often limited and may not be strong enough to induce a phenotype. Another challenge is the close proximity of many lncRNAs with coding neighbors with shared regulatory elements, precluding lncRNA-restricted knockdown. A system to generate the early abortion of lncRNA transcription thereby preventing the generation of mature transcripts without affecting regulatory elements would be optimal for functional screenings. Although this has been achieved for two specific lncRNA genes using insertion of a premature polyadenylation signal, the efficiency to generate cells with the appropriate insertion leading to truncation of the transcript was low[35]. Alternatives including the insertion of other termination signals or targeting of termination-associated proteins using the dCas9 system have not been attempted yet.

Besides its utility in large-scale screens the CRISPR/Cas9 system offers endless opportunities to study the precise mechanics of a specific lncRNA in the cell type of interest. A specific lncRNA locus can be modified by removal of specific regions or by shortening the transcript through insertion of termination signals. Such experiments can provide information to whether the function of a lncRNA is determined by the transcript and whether specific sequence motifs are functionally important. Removal or mutation of regulatory sequences (e.g. promoter, enhancer, splice site) can provide information whether the functionality is related to the act of its transcription. Cells with a modified lncRNA locus can subsequently be analyzed for changes in its potential *cis*- or *trans*-regulatory effects and effects on the 3D genome organization. CRISPR/Cas9 can also be used to add specific sequences to the lncRNA, such as sequence-tags for pulldown or subcellular localization signals. Moreover, a variety of other protein domains in addition to transcriptional activators or repressors may be linked to dCas9 protein. This can be employed to change epigenetic marks at lncRNA loci (e.g. histone modifiers, DNA demethylases) or to allow spacio-temporal tagging of surrounding proteins (e.g. APEX).

More recently, it has been shown that RNA molecules may also be targeted using a modified CRISPR/Cas9 system[36, 37] or by utilizing the variant Cas13a[38]. This might provide even more opportunities for functional lncRNA research. Active Cas9/13a protein can be used to induce RNA degradation, while dCas9 linked to various proteins could be used to influence lncRNA splicing, stability, localization or interactions. Thus, the CRISPR/Cas system in all its variety is an essential tool for future lncRNA research.

Identification of lncRNA interacting partners is probably the most informative method to further define its function. To this end multiple lncRNA immunoprecipitation (IP) methods have been developed (ChIRP, CHART, RAP; reviewed in[39, 40]) that use streptavidin-coupled antisense oligonucleotides tiling the lncRNA of interest for pull down. Depending on the isolation method applied, the interacting DNA, RNA or protein molecules can be defined by performing sequencing or mass spectometry, respectively. These methods have served well in defining genomic binding sites as well as protein interactors of lncRNAs such as XIST[41], MALAT1[35] and MEG3[42]. However, multiple factors hamper the broad use of these techniques: Firstly, a large number of tiling oligonucleotides are necessary for an efficient pull down. As lncRNAs are generally shorter compared to mRNAs and often contain a relatively large proportion of repetitive sequences, the design of sufficient probes for efficient pull down can be challenging. Secondly, many lncRNAs are expressed at very low levels, making it difficult to capture sufficient amounts of the interactors for subsequent analysis. Thirdly, RNA-IP techniques are especially prone to background noise including artifacts due to off-target hybridization, capturing of the corresponding DNA, or identification of indirect RNA/DNA interactors. In addition, off-target hybridization may be more prominent for probes not capable of binding the target lncRNA due to obstruction by binding factors or secondary structure. For these reasons, IP of a protein of interest followed by the isolation of the associated lncRNAs is

a much more broadly used approach. Considerations to be made here are the strength of cross-linking (e.g. UV vs. formaldehyde), as well as the capturing method (direct vs. indirect using a tagged protein). Strong cross-linking and antibody capturing may lead to a higher background signal, due to washing of the IP fraction in non-denaturing conditions, allowing non-specific binding in solution. Lastly, most of these protocols have been developed using adherent cell types and may need optimization before successful application with cells growing in suspension.

Combination of IP methods with additional approaches can further broaden their potential value to unravel the function of the lncRNA. For example, the lncRNA repertoire in a specific subcellular compartment (e.g. in mitochondria, proximity of nuclear membrane) may be captured by APEX-RIP[43]. This can be achieved by coupling APEX to a protein with a known localization, followed by biotin-labeling of proteins in close proximity of APEX. The labeled proteins and associated RNAs (APEX cannot tag RNA directly) can then be isolated using a biotin-steptavidin mediated IP. By coupling APEX to different proteins with known subcellular localizations, this system can be adapted to identify lncRNAs enriched in different compartments. It would for example be interesting to investigate whether certain types of transposable elements enriched in lncRNA subsets have specific protein interactors. To this end, APEX could be coupled to the RNA-dCas9/13a system and guided to specific repeat sequences within lncRNAs followed by IP of the tagged proteins. Thus, the combination of new techniques with already established methods can greatly expand the toolbox to study lncRNA interactions in the coming years.

Visualization of lncRNAs by RNA fluorescent *in situ* hybridization (FISH) faces many of the same challenges as described above for RNA-IP procedures. RNA-FISH also requires multiple specific probes tiling the lncRNA, but in case sufficient probes can be designed can achieve single molecule resolution[44]. RNA-FISH thus allows detection of rough subcellular distribution patterns (e.g. cytoplasmic/nuclear, punctuate vs. diffuse) of lncRNAs. New developments such as RNA-targeting dCas9/13a promise an expansion of visualization techniques. Opposed to RNA-FISH, the exogenous expression of a dCas9/13a-fluorophore construct can potentially be used to visualize lncRNA transcripts in living cells and allows lncRNA expression-based cell sorting. Of note, this requires the development of a fluorophore system that only gives signal upon binding to the lncRNA[38]. dCas-based visualization techniques will facilitate sorting and in depth analysis of lncRNA-positive cell populations.

The secondary structure of lncRNAs is thought to be an important factor for their functionality. Multiple RNA folding prediction algorithms have been developed (based on e.g. base-pairing probabilities, minimal free-energy) and are available for public use, but currently fall short of means for functional interpretation of the putative structures. Experimental assessment of RNA structure relies mainly on RNAse degradation (distinguishing single vs. double stranded RNA) and the introduction of structure-

modifying mutations in e.g. stem-loop structures. Without the availability of methods similar to protein crystallography, a detailed visualization of the secondary structures of RNA transcripts is currently not possible. Outstanding questions include the role of specific RNA modifications in defining the structure, and whether RNA folding is regulated by protein components similar to chaperones regulating protein folding.

Eventually, any biological process is best studied *in vivo*. As the majority of lncRNAs appear to be primate-specific, this is probably the largest challenge faced in lncRNA research. Common animal models are often not useful and xenografts or (partially) humanized models are the few alternatives present at this time. In the absence of applicable whole-organism model systems, alternatives such as the use of organoid structures may prove important. Some advantages of organoid models include the much smaller time-investment compared to animal models, their generation from (patient-derived) human tissue thus closely mimicking the in vivo architecture and genetic signature, and their suitability for high-throughput studies. The use of organoids either generated from tissues or through differentiation of induced pluripotent stem cells allows not only the study of normal and diseased human tissue[45], but also of processes such as development[46], aging[47] and malignant transformation[48]. The germinal center reaction can also be modeled in this way[49, 50] and this may lead to further insights into lncRNA function in specific B-cell subsets (e.g. those expressing Myc) in a microenvironment mimicking the GC microenvironment.

In summary, many technological advancements to study lncRNA function are being made and will most certainly lead to a better understanding of lncRNA function and incorporation of lncRNAs into currently known biological processes within the cell. However, most of the described techniques need optimization, as they currently are time-consuming, technically challenging, and by far not sufficiently sensitive to successfully explore the full lncRNA repertoire.

3.4 The potential use of lncRNAs in the clinic

The high cell type specificity and broad deregulation in cancer cells suggest that lncRNAs have the potential to become valuable biomarkers. This lncRNA cell type specificity could also be further exploited by measuring levels of disease specific lncRNAs in the circulation[51] as a minimal-invasive assessment of disease load.

In cancers in which the driving mutations are currently unknown, investigation of possible causal lncRNA-mutations could provide additional insight. The first reports are now appearing that show recurrent mutations in lncRNAs[52-54]. But there is a large gap in our knowledge here as most studies so far used protein-coding gene focused whole exome sequencing approaches and ignored the non coding regions. Furthermore, tumors may contain very heterogeneous cell populations, subsets of which can develop resistance to therapy and cause disease relapse. The cell type restricted expression

patterns of lncRNAs could lead to a better understanding of tumor subpopulations and form a basis to further expand the options for personalized treatment.

Once lncRNAs with important functions in disease have been identified and characterized, ways to therapeutically change their levels or interactions are needed. MiRNA-based therapeutics are already on the rise and might also be applicable to inhibit lncRNA function. Moreover, small molecule inhibitors that disrupt specific lncRNA-protein interactions may be developed. In addition, novel possibilities may arise once we understand lncRNA biology in more detail, allowing the exploitation of their functional mode of action towards novel therapeutic approaches. Looking at the functional repertoire of lncRNAs including transcripts such as XIST, capable of inactivating an entire chromosome, the future of lncRNA research will surely bear exciting new developments.

4 References

1 DiMenna, L. J. *et al*. Cutting Edge: The Transcription Factor Sox2 Regulates AID Expression in Class-Switched B Cells. *J. Immunol.* **198**, 2244-2248 (2017).

2 Cozen, W. *et al*. A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. *Nat. Commun.* **5**, 3856 (2014).

3 Guan, H., Xie, L., Wirth, T. & Ushmorov, A. Repression of TCF3/E2A contributes to Hodgkin lymphomagenesis. *Oncotarget* **7**, 36854-36864 (2016).

4 Sakurai, N. *et al.* The LRF transcription factor regulates mature B cell development and the germinal center response in mice. *J. Clin. Invest.* **121**, 2583-2598 (2011).

5 Maeda, T., Hobbs, R. M. & Pandolfi, P. P. The transcription factor Pokemon: a new key player in cancer pathogenesis. *Cancer Res.* **65**, 8575-8578 (2005).

6 Merryman, R. W., Armand, P., Wright, K. T. & Rodig, S. J. Checkpoint blockade in Hodgkin and non-Hodgkin lymphoma. *Blood Adv.* **1**, 2643-2654 (2017).

7 Zeller, K. I. *et al*. Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 17834-17839 (2006).

8 Link, J. M. & Hurlin, P. J. The activities of MYC, MNT and the MAX-interactome in lymphocyte proliferation and oncogenesis. *Biochim. Biophys. Acta* **1849**, 554-562 (2015).

9 Caretti, G., Salsi, V., Vecchi, C., Imbriano, C. & Mantovani, R. Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters. *J. Biol. Chem.* **278**, 30435-30440 (2003).

10 Izumi, H. *et al.* Mechanism for the transcriptional repression by c-Myc on PDGF beta-receptor. *J. Cell. Sci.* **114**, 1533-1544 (2001).

11 Herlihy, A. E. & de Bruin, R. A. The Role of the Transcriptional Response to DNA Replication Stress. *Genes (Basel)* **8**, 10.3390/genes8030092 (2017).

12 Feng, X. H., Liang, Y. Y., Liang, M., Zhai, W. & Lin, X. Direct interaction of c-Myc with Smad2 and Smad3 to inhibit TGF-beta-mediated induction of the CDK inhibitor p15(Ink4B). *Mol. Cell* **9**, 133-143 (2002).

13 Gartel, A. L. *et al.* Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4510-4515 (2001).

14 Gardini, A. Global Run-On Sequencing (GRO-Seq). Methods Mol. Biol. 1468, 111-120 (2017).

15 Kumar, R. M. et al. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516**, 56-61 (2014).

16 Kolodziejczyk, A. A. *et al*. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell. Stem Cell.* **17**, 471-485 (2015).

17 Liu, S. J. *et al.* Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* **17**, 67-016-0932-1 (2016).

18 Kim, D. H. *et al.* Single-cell transcriptome analysis reveals dynamic changes in lncRNA expression during reprogramming. *Cell. Stem Cell.* **16**, 88-101 (2015).

19 Schonrock, N., Jonkhout, N. & Mattick, J. S. Seq and You Will Find. *Curr. Gene Ther.* **16**, 220-229 (2016).

20 Hezroni, H. *et al.* Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell. Rep*. **11**, 1110-1122 (2015).

21 Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **17**, 601-614 (2016).

22 Winkle, M. *et al.* Long noncoding RNAs as a novel component of the Myc transcriptional network. *FASEB J.* **29**, 2338-2346 (2015).

23 Hart, J. R., Roberts, T. C., Weinberg, M. S., Morris, K. V. & Vogt, P. K. MYC regulates the non-coding transcriptome. *Oncotarget* **5**, 12543-12554 (2014).

24 Doose, G. *et al.* MINCR is a MYC-induced lncRNA able to modulate MYC's transcriptional network in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5261-70 (2015).

25 Trimarchi, T. *et al.* Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* **158**, 593-606 (2014).

26 Zhang, J. *et al.* New insights into the roles of ncRNA in the STAT3 pathway. *Future Oncol.* **8**, 723-730 (2012).

27 Adriaens, C. *et al.* p53 induces formation of NEAT1 lncRNA-containing paraspeckles that modulate replication stress response and chemosensitivity. *Nat. Med.* **22**, 861-868 (2016).

28 Blume, C. J. *et al.* p53-dependent non-coding RNA networks in chronic lymphocytic leukemia. *Leukemia* **29**, 2015-2023 (2015).

29 Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409-419 (2010).

30 Wichman, L. *et al.* Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. *Biol. Reprod.* **97**, 313-323 (2017).

31 Qin, N. *et al.* Systematic identification of long non-coding RNAs with cancer-testis expression patterns in 14 cancer types. *Oncotarget* **8**, 94769-94779 (2017).

32 Hosono, Y. *et al.* Oncogenic Role of THOR, a Conserved Cancer / Testis Long Non-coding RNA. *Cell* **171**, 1559-1572.e20 (2017).

33 Cong, L. *et al.* Multiplex genome engineering using CRISPR / Cas systems. *Science* **339**, 819-823 (2013).

34 Du, D. & Qi, L. S. CRISPR Technology for Genome Activation and Repression in Mammalian Cells. *Cold Spring Harb Protoc.* **2016**, pdb.prot090175 (2016).

35 Engreitz, J. M. *et al*. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452-455 (2016).

36 O'Connell, M. R. *et al.* Programmable RNA recognition and cleavage by CRISPR / Cas9. *Nature* **516**, 263-266 (2014).

37 Nelles, D. A., Fang, M. Y., Aigner, S. & Yeo, G. W. Applications of Cas9 as an RNA-programmed RNA-binding protein. *Bioessays* **37**, 732-739 (2015).

38 Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. Nature 550, 280-284 (2017).

39 Machyna, M. & Simon, M. D. Catching RNAs on chromatin using hybridization capture methods. *Brief Funct. Genomics* (2017).

40 Simon, M. D. Insight into lncRNA biology using hybridization capture analyses. *Biochim. Biophys. Acta* **1859**, 121-127 (2016).

41 Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416 (2015).

42 Iyer, S., Modali, S. D. & Agarwal, S. K. Long Noncoding RNA MEG3 Is an Epigenetic Determinant of Oncogenic Signaling in Functional Pancreatic Neuroendocrine Tumor Cells. *Mol. Cell. Biol.* **37**, 10.1128 / MCB.00278-17. Print 2017 Nov 15 (2017).

43 Kaewsapsak, P., Shechner, D. M., Mallard, W., Rinn, J. L. & Ting, A. Y. Live-cell mapping of organelle-associated RNAs via proximity biotinylation combined with protein-RNA crosslinking. *Elife* **6**, 10.7554/eLife.29224 (2017).

44 Bayer, L. V., Batish, M., Formel, S. K. & Bratu, D. P. Single-Molecule RNA In Situ Hybridization (smFISH) and Immunofluorescence (IF) in the Drosophila Egg Chamber. *Methods Mol. Biol.* **1328**, 125-136 (2015).

45 Dutta, D., Heo, I. & Clevers, H. Disease Modeling in Stem Cell-Derived 3D Organoid Systems. *Trends Mol. Med.* **23**, 393-410 (2017).

46 Huch, M. & Koo, B. K. Modeling mouse and human development using organoid cultures. *Development* **142**, 3113-3125 (2015).

47 Hu, J. L., Todhunter, M. E., LaBarge, M. A. & Gartner, Z. J. Opportunities for organoids as new models of aging. *J. Cell Biol.* **217**, 39-50 (2018).

48 Unno, K. *et al.* Modeling African American prostate adenocarcinoma by inducing defined genetic alterations in organoids. *Oncotarget* **8**, 51264-51276 (2017).

49 Beguelin, W. *et al.* EZH2 enables germinal centre formation through epigenetic silencing of CDKN1A and an Rb-E2F1 feedback loop. *Nat. Commun.* **8**, 877-017-01029-x (2017).

50 Purwada, A. & Singh, A. Immuno-engineered organoids for regulating the kinetics of B-cell development and antibody production. *Nat. Protoc.* **12**, 168-182 (2017).

51 Jiang, X., Lei, R. & Ning, Q. Circulating long noncoding RNAs as novel biomarkers of human diseases. *Biomark Med.* **10**, 757-769 (2016).

52 Lanzos, A. *et al.* Discovery of Cancer Driver Long Noncoding RNAs across 1112 Tumour Genomes: New Candidates and Distinguishing Features. *Sci. Rep.* **7**, 41544 (2017).

53 Pan, W. *et al.* Whole exome sequencing identifies lncRNA GAS8-AS1 and LPAR4 as novel papillary thyroid carcinoma driver alternations. *Hum. Mol. Genet.* **25**, 1875-1884 (2016).

54 Fujimoto, A. *et al.* Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500-509 (2016).