

University of Groningen

The missing piece

Winkle, Melanie

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Winkle, M. (2018). *The missing piece: Long noncoding RNAs in cancer cell biology*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

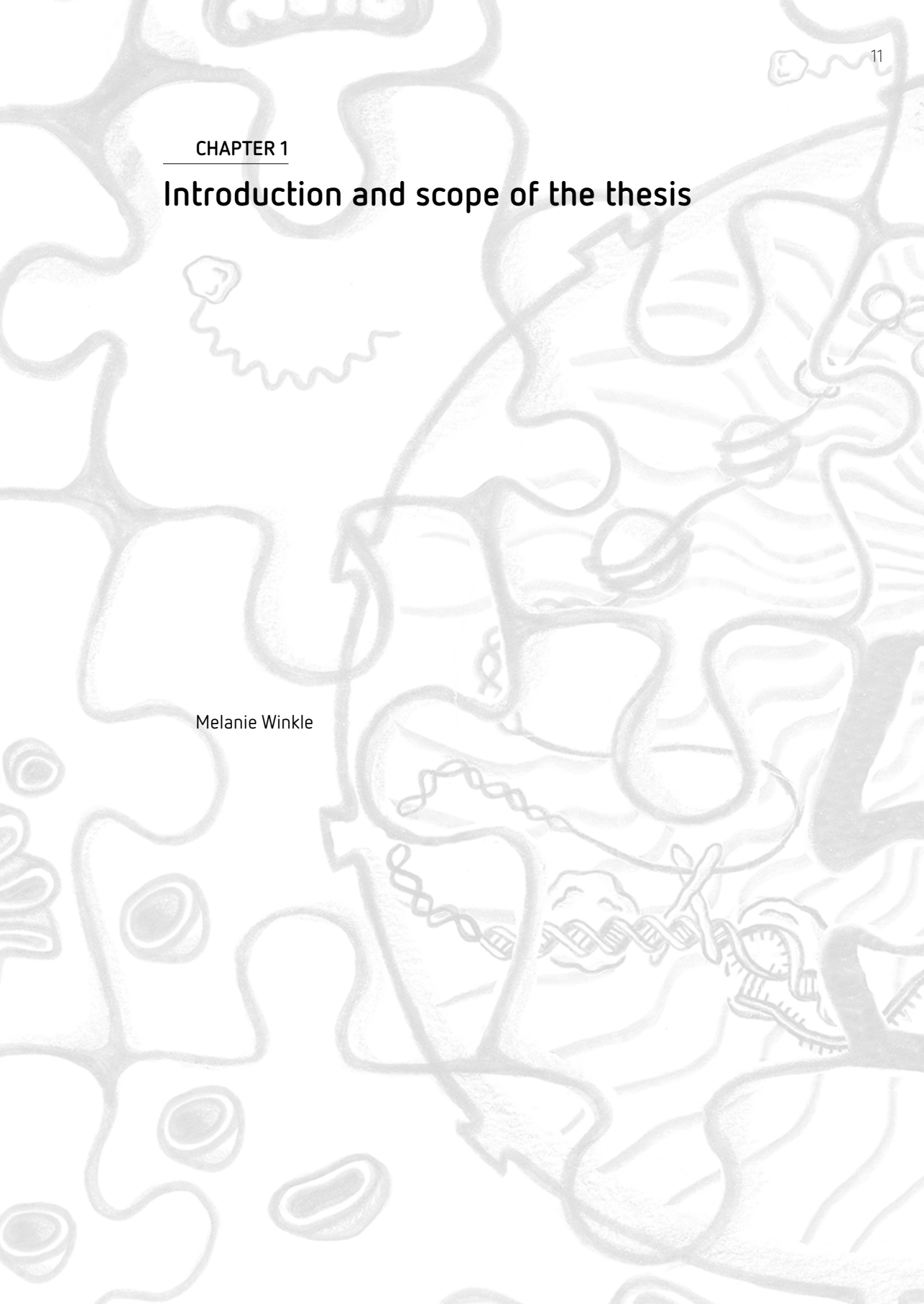
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CHAPTER 1

Introduction and scope of the thesis

Melanie Winkle



1 Introduction

1.1 Composition of the human genome

For a long time it has been a puzzling fact why the vast majority of the human genome does not encode for proteins; the main functional entities of cells. As shown in **FIGURE 1A**, coding exons make up less than two percent of the genome. A much larger part is occupied by noncoding sequences, either located within genes (i.e. intronic) and regulatory elements (ca. 24%) or within intergenic spaces (ca. 18%). The largest part of the genome (ca. 45%) is assigned to variable classes of transposable elements (TEs), mobile DNA sequences inserted into and moving within the genome, while around 11% contains simple repetitive sequence and duplications (estimations based on^{1,2}). However, the line between noncoding sequences and TEs or repeats is rather thin, as noncoding loci often contain these elements³. In addition, both repeats^{4,5} and TEs^{3,6-8} may act as functional entities in noncoding genes. Thus, although human cells go through great lengths to replicate their entire genome as fault-free as possible, a very minor part serves the production of proteins. This clearly suggests that there is a substantial piece of information missing in the jigsaw puzzle of the human genome.

1.2 The dawn of long noncoding RNA research

Major advances in sequencing technologies now provide a much more intricate view of the human transcriptome, showing that about two-thirds of the genome is actively transcribed in at least one cell type or tissue⁹. The presence of tightly regulated, independent transcriptional units lacking *coding potential* (**BOX 1**)⁹⁻¹¹ was followed by a plethora of publications discussing the significance of the noncoding genome for human life¹²⁻¹⁴. The identification and mapping of ten-thousands of long noncoding (lnc)RNA genes^{2,15-18} eventually lead to the emergence of an entire research field. Accordingly, the amount of studies addressing lncRNA function has risen from a single publication in 2006¹⁹ to over 7,000 in 2017 (**FIGURE 1B**). Of note, the highly abundant lncRNA XIST involved in X chromosome inactivation was identified as early

as 1991²⁰. The experimental support required for the hypothesis that lncRNAs have important functions was delivered by the first murine knockout study in 2013. For five of 18 lncRNAs knockout induced developmental errors, peri- or postnatal lethality²¹. Thus, the central dogma of RNA functioning as a mere intermediate between DNA and protein has been rattled since the discovery of lncRNAs as functional entities.

Box 1 The CODING POTENTIAL of a transcript can be calculated based on the quality and length of *putative open reading frames* (ORFs) as well as the *similarity* of the putative protein product to known proteins. To exclude the production of peptides from small ORFs that may be present within noncoding genes, *ribosome sequencing* data as well as *mass-spectrometry* data (peptide sequences) can be considered.

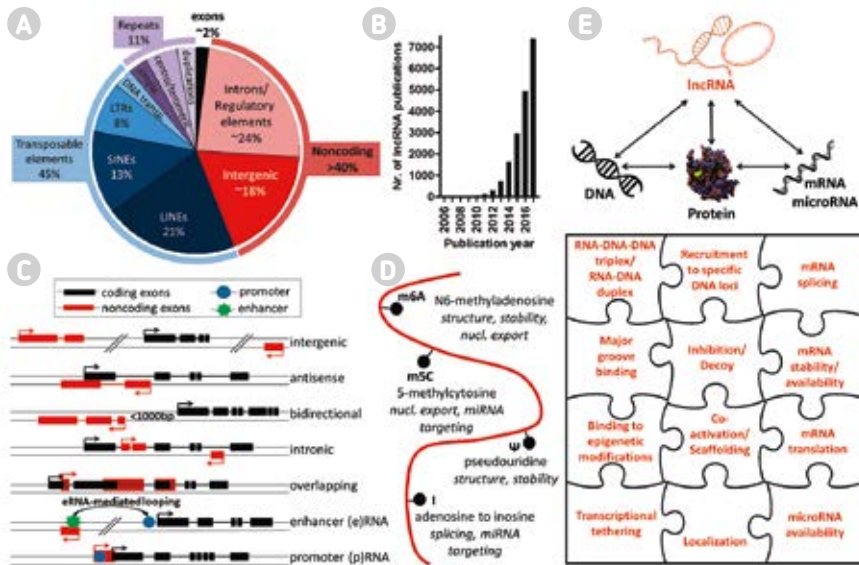


FIGURE 1 An exciting new world of RNA. **(A)** Composition of the human genome. While only 2% of genomic sequences encode for protein, the main part is composed of various classes of noncoding sequences including genic and intergenic sequences, transposable elements and repeats. **(B)** Number of PubMed listed publications on 'lncRNA' or 'long noncoding RNA' from 2006 to 2016. **(C)** LncRNA classes as defined by their genomic orientation with respect to protein coding genes. **(D)** The most common chemical modifications found on RNA molecules. The function of the respective modification in mRNAs is stated in italic. **(E)** The repertoire of lncRNA interactors includes DNA, protein and RNA. Experimentally supported functions of the possible lncRNA interactions are shown in the puzzle below.

1.3 LncRNA (R)evolution

Doubts about the functionality of lncRNAs are mainly based on the limited evolutionary conservation present in lncRNA as compared to coding genes. However, purifying selection does occur in approx. 5% of the genome, a much larger fraction than occupied by coding sequence (<2%). Closer inspection of lncRNA loci indeed identified evolutionary constraint, albeit mostly restricted to lncRNA promoters, splice sites and shorter stretches of primary sequence^{22, 23}. The definition of lncRNA orthologs (see also **BOX2**) accordingly applies less stringent sequence alignment cutoffs and may additionally depend on positional

Box 2 For assessment of evolutionary conservation of lncRNAs the following distinctions are made:

LncRNA orthologs are based on sequence similarity (~30% of exonic sequence can be aligned).

Syntenic lncRNA orthologs are based on genomic association (synteny) and thus contain conserved lncRNA-mRNA pairs.

Transcriptional conservation is present when the orthologous sequence also produces an RNA transcript. Orthologs that do not show transcriptional conservation are also referred to as 'pseudoconserved'.

Regulatory conservation is present when the lncRNA ortholog shows the same tissue specificity.

conservation (i.e. syntenic lncRNA orthologs or lncRNA-mRNA pairs). Furthermore, transcriptional conservation needs to be considered, i.e. whether the orthologous sequence is also actively transcribed. Considering these factors, protein coding genes are highly conserved: Comparison of human coding genes with chimpanzee and rat, showed high sequence (99 and 93%, respectively) as well as transcriptional conservation (92 and 90%, respectively)²⁴. This confirms that the coding genomic content is very static across large eras of evolution (also known as the G-value paradox). Interestingly, the amount of noncoding sequence (i.e. the ratio of noncoding sequence over total genome size) shows a linear correlation with the biological complexity of organisms^{12,25}. Accordingly, the conservation of lncRNAs decreases more rapidly with evolutionary distance: Of a set of 1,900 human lncRNAs 98 and 54% had orthologs in chimpanzee and rat, respectively, of which 80 and 35% also show transcriptional conservation²⁴. In other words, an estimated 70 to 80% of lncRNA genes have developed at a late point in evolution and are largely specific to the primate lineage²⁶⁻²⁸. Interestingly, of these primate-lncRNAs approx. 20% are brain-specific²⁶. Novel lncRNA genes are thought to originate from protein coding gene copies or pseudogenes that have lost their coding potential^{28,29}, or from the insertion and adaptation of TEs³⁰. TEs are able to acquire functional significance, e.g. as regulatory elements or by shaping splicing patterns, thereby creating novel transcriptional units with tissue specific expression^{6, 28,30}. These observations hint at a rapid turnover in lncRNA evolution and raise the question when these transcripts have undergone their 'revolution', i.e. at which point they have gained actual functional significance. The majority of lncRNAs that are conserved across multiple mammalian species, i.e. (syntenic) lncRNA orthologs, also show regulatory conservation (see BOX 2), i.e. expression in the same tissues²⁴. Of all lncRNAs initially mapped across 20 human tissues¹⁶, nearly 80% showed high tissue specificity. More than half of these were specifically expressed in testis^{16, 26}, a tissue with a highly permissive transcriptional landscape due to an open chromatin conformation³¹. Evolving lncRNA genes might respond to weak stimuli in such a permissive environment, while functional lncRNAs are more likely to be under tight regulation. In accordance with this, 'younger' lncRNA genes are more often testis specific than 'older' lncRNA genes, and brain specific lncRNAs show the highest grade of (regulatory) conservation²⁶. Together this shows that the noncoding part of the human genome is much more dynamic and subject to continuative evolutionary turnover.

1.4 lncRNA classes

lncRNAs are classified according to their size (>200 nucleotides) and their genomic position in relation to protein coding genes or regulatory elements. The major classes defined to date (FIGURE 1C) are intergenic, antisense, bidirectional, intronic and overlapping lncRNAs, as well as RNAs transcribed from enhancer (eRNAs) or promoter regions (pRNAs). In addition, a growing number of circular (circ)RNA transcripts generated by back-splicing reactions have been identified³². The majority of lncRNAs identified so far are intergenic or bidirectional lncRNAs (>10,000 transcripts), antisense

(~5,000) and intronic transcripts (~600), with just a few known overlapping lncRNAs (<200) (estimations based on²). However, these numbers likely also reflect the difficulty of distinguishing functional intronic, overlapping and promoter-derived lncRNAs from primary mRNA transcripts. Active transcription is often observed at enhancers producing enhancer (e)RNAs. Accordingly a significant number of lncRNA loci are marked by the typical enhancer chromatin mark H3K4me^{1,33}. Furthermore, about 4% of the lncRNA loci encode small RNAs such as micro (mi)RNAs or small nucleolar (sno) RNAs within their introns (compared to 7% of coding genes)².

1.5 lncRNA characteristics

Some main characteristics distinguish lncRNAs from mRNA transcripts: (1) a generally lower expression level^{2,16}, (2) a lower number of exons (mostly two to three)^{2,16}, (3) an expression pattern that is more often restricted to a single cell type or tissue^{2,33}, and (4) more frequent nuclear localization with a low amount or complete lack of transcripts in the cytoplasm³⁴. Similar to mRNAs, lncRNAs are (alternatively) spliced and stabilized through 5'cap structures and polyadenylation¹⁶. Exceptions to this are most eRNA transcripts (very short lived) and circRNAs (extremely stable structure). Moreover, novel triple-helix and snoRNA-mediated stabilization mechanisms of non-polyadenylated lncRNAs have been described³⁵, indicating that more yet unidentified mechanisms may be at hand. Chemical modification of lncRNA transcripts can potentially add flexibility to the system, as is observed in protein coding transcripts. The most prevalent types of RNA modifications, 5-methylcytosine (m⁵C), N6-methyladenosine (m⁶A), pseudouridine (Ψ), and adenosine-to-inosine editing (A-to-I) indeed occur in thousands of lncRNA transcripts^{36,37}. Though few studies have addressed the functional significance of these modifications, involvements in e.g. secondary lncRNA structure and/or binding interactions with proteins have been suggested (reviewed in³⁸). In addition, the functions of these modifications in lncRNAs may be similar to those described for mRNAs, e.g. regulating stability, nuclear export or miRNA targeting (FIGURE 1D)^{37,39}.

1.6 lncRNA functions

RNA molecules have the capacity to create complex molecular interaction networks. Linear lncRNAs can bind to DNA or RNA based on sequence homology thereby forming triplex structures⁴⁰, while lncRNAs folded into secondary structures may form various complex interactions with proteins and/or RNA/DNA⁴¹. Together, these factors contribute to a complicated jigsaw puzzle of possible interactions and functionalities (FIGURE 1E), including transcriptional and post-transcriptional control mechanisms (reviewed in^{42,43}). Interaction with DNA can occur not only through base-pairing with DNA, but also through duplex formation with nascent RNA transcripts, by interaction with DNA-bound proteins, chromatin marks or by transcriptional tethering (i.e. the lncRNA transcript being functional during its transcription)^{40,44-46}. This can lead to gene expression regulation in *cis* or *trans* through e.g. steric hindrance or recruitment of

transcription factors and/or co-factors or via recruitment of epigenetic activators and repressors. Furthermore, various post-transcriptional mechanisms can be at hand^{47, 48}, affecting mRNAs (splicing, stability, localization, availability), proteins (stability, localization, availability, activation/repression) and protein complexes (scaffolding) or microRNA availability. Thus, the flexibility and interaction capacity that is intrinsic to lncRNAs creates a broad functional repertoire. Furthermore, RNA secondary structures and chemical modifications can add additional diversity and fine-tuning to their functionality.

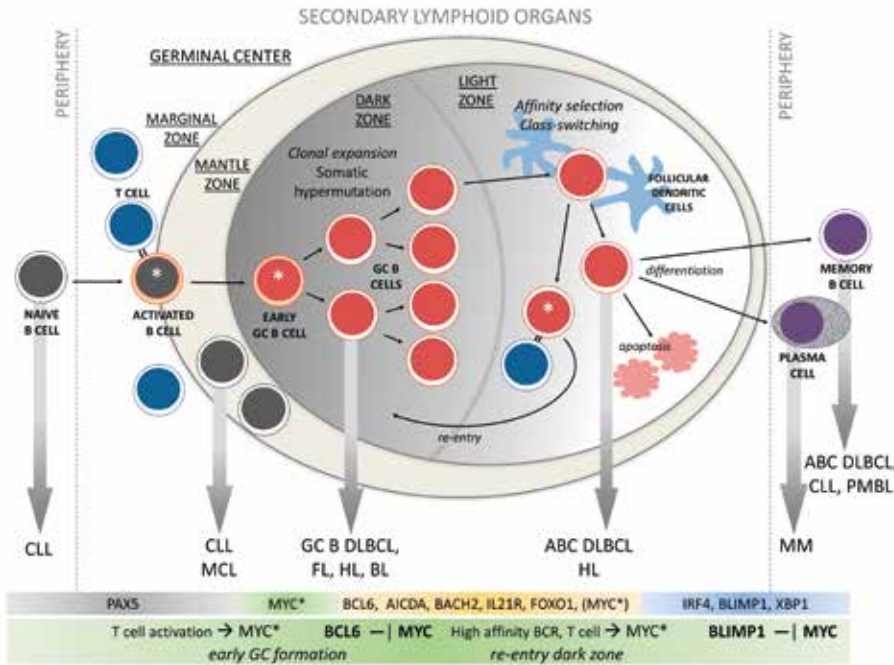


FIGURE 2 Schematic of normal B-cell maturation and lymphomagenesis. (left to right) Immature naïve B cells from the bone marrow enter the secondary lymphoid organs. B cells are activated upon encountering antigens and receiving activating stimuli from T helper cells, resulting in the formation of germinal centers (GCs). In the dark zone of the GC, B cells undergo clonal expansion and somatic hypermutation to diversify the B cell receptor repertoire. In the light zone, B cells are selected for antigen affinity. Apoptosis is induced in cells with sub-optimal antigen affinity, while high affinity B cells may differentiate into effector cells (i.e. memory B or antibody secreting plasma cells) or re-enter the dark zone for further expansion and germinal center maintenance. (grey arrows) B-cell lymphomas most commonly arise from GC B or post-GC B cells. ABC DLBCL – activated B cell-like diffuse large B-cell lymphoma; BL – Burkitt lymphoma; CLL – chronic lymphocytic leukemia; FL – Follicular lymphoma; GC B DLBCL – germinal center B cell-like diffuse large B-cell lymphoma; HL – Hodgkin lymphoma; PMBL – primary mediastinal B-cell lymphoma. (bottom) Major molecular components involved in the B cell maturation process with special emphasis on Myc. Myc (asterisk) is expressed upon T cell activation of B cells, during the early stages of germinal center formation, and in subsets of light zone B cells, while otherwise being repressed by BCL6 or BLIMP1.

1.7 B-cell lymphoma and Myc

B-cell lymphoma malignancies show an annual incidence of approx. 20 new cases per 100,000 persons. The most common subtypes include diffuse large B-cell lymphoma (DLBCL), Hodgkin lymphoma (HL), follicular lymphoma (FL) and chronic lymphocytic leukemia (CLL). Less common B-cell lymphomas include mantle cell lymphoma (MCL), marginal zone lymphoma (MZL), primary mediastinal B-cell lymphoma (PMBL) and Burkitt lymphoma (BL). Despite significant advancements in treatment strategies, lymphoma-associated deaths are still high with 5-year survival rates between 30-90% depending on lymphoma subtype and stage at diagnosis. Furthermore, successfully treated patients may suffer from secondary treatment effects including cardiac disease, secondary malignancies and infertility^{49,50}.

The majority of B-cell lymphomas arise during or after the germinal center reaction (FIGURE 2). GC B cells are especially susceptible to malignant transformation due to their high proliferative rate and the genomic hypermutation machinery active during B-cell receptor diversification⁵¹. Somatic hypermutation and class-switch recombination require the generation of double strand breaks within the immunoglobulin heavy/light chain gene (IgH/IgL) loci, which are induced by AID (encoded by AICDA). AID off-targeting may result in translocation of proto-oncogenes to the active Ig loci, resulting in constitutive oncogene expression^{52,53}. Such translocations are commonly observed in BL (MYC-IgH/IgL), DLBCL (BCL6-IgH, BCL2-IgH, MYC-IgH/IgL), FL (BCL2-IgH) and MCL (CCND1-IgH)⁵¹. In DLBCL, MYC rearrangements are seen in approximately 10% of the cases and a lower percentage is seen in FL, MCL and HL⁵⁴. Myc rearrangements have been associated with more aggressive lymphoma subtypes and a poor clinical outcome^{54,55}. A direct causative role for Myc-regulated genes and cancer development has been established for several protein coding genes and miRNAs^{56,57}. During normal B cell maturation, Myc is only expressed in specific subsets of cells (FIGURE 2): Myc is induced upon T cell activation and remains present during the initial stages of GC formation, to then be repressed by BCL6 in the active GC^{58,59}. A subset of light zone GC B cells with high affinity B-cell receptors express Myc upon interaction with T cells. These Myc⁺ light zone GC B cells (Myc⁺, BCL6⁻, NF- κ B⁺) are thought to re-enter the dark zone for further proliferation and are essential for GC maintenance^{58,59}. As MYC is frequently translocated in B-cell lymphomas⁵¹ and translocations generally require the active transcription of its partners^{52,53,58}, these Myc⁺ GC B cell populations may be especially prone to malignant transformation^{58,59}.

2 Aim and scope of the thesis

LncRNA research does not only greatly expand our understanding of human cell biology and its intricate, cell-type specific regulatory mechanisms, it also provides

an additional point-of-view on the deregulation of these cellular systems in cancer cells. In the future, fundamental research on lncRNA expression deregulation and functionality will likely provide a solid basis for a vast expansion of drugable targets in cancer therapy. This study aims at the characterization of lncRNA expression patterns in normal B cell subsets and different types of B-cell lymphoma and a further functional characterization of deregulated lncRNAs.

In **CHAPTER 2**, the topic is introduced more thoroughly in a review addressing the current state of lncRNA research in normal B-cell development and B-cell malignancies. **CHAPTER 3** addresses lncRNA expression patterns in normal sorted B-cell populations (i.e. naïve, GC and memory B cells) and assess lncRNA deregulation in HL cell lines compared to their presumed cell of origin. The study of HL is especially challenging due to a minor contribution of actual tumor cells (<1%) within an extensive inflammatory background. Studying lncRNAs deregulated in HL cell lines provided us with HL-specific candidate lncRNAs whose expression was further studied in primary HL tissue samples by RNA FISH. In **CHAPTER 4**, we focused on lncRNAs regulated by the oncogenic transcription factor Myc, using genome wide expression analyses in a B-cell model with a repressible *MYC* allele and in primary lymphoma samples characterized by high or low Myc expression. With these experiments we will for the first time establish the relevance of lncRNAs for the Myc transcriptional network. In **CHAPTER 5** we further deepen the understanding of Myc-regulated lncRNAs by using an additional *in vitro* model, i.e. Myc knockdown in BL cell lines. A reliable set of Myc regulated lncRNAs relevant to BL pathogenesis will be established by a further selection of candidates based on an early response to MYC induction and proven Myc occupancy near the transcription start site. Further studies on the top Myc-induced candidate KTN1-AS1 will be performed to establish its role in guiding the downstream effects of Myc in BL cells. In **CHAPTER 6** our findings are summarized, discussed and future perspectives are elaborated.

3 References

- 1 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
- 2 Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775-1789 (2012).
- 3 Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13**, R107-2012-13-11-r107 (2012).
- 4 Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756 (2008).
- 5 Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10654-10659 (2017).
- 6 Kapusta, A. *et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
- 7 Hadjiargyrou, M. & Delihias, N. The intertwining of transposable elements and non-coding RNAs. *Int. J. Mol. Sci.* **14**, 13307-13328 (2013).
- 8 Sytnikova, Y. A., Rahman, R., Chirn, G. W., Clark, J. P. & Lau, N. C. Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in *Drosophila* ovarian cell cultures. *Genome Res.* **24**, 1977-1990 (2014).
- 9 ENCODE Project Consortium *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
- 10 Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563-573 (2002).
- 11 Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-1563 (2005).
- 12 Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288-299 (2007).
- 13 Mattick, J. S. The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459 (2009).
- 14 Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629-641 (2009).
- 15 Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223-227 (2009).
- 16 Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25**, 1915-1927 (2011).
- 17 Bu, D. *et al.* NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res.* **40**, D210-5 (2012).
- 18 Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760-1774 (2012).
- 19 Shamovsky, I. & Nudler, E. Gene control by large noncoding RNAs. *Sci. STKE* **2006**, pe40 (2006).



- 20 Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38-44 (1991).
- 21 Sauvageau, M. *et al.* Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**, e01749 (2013).
- 22 Ponjavic, J., Ponting, C. P. & Lunter, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* **17**, 556-565 (2007).
- 23 Marques, A. C. & Ponting, C. P. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol.* **10**, R124-2009-10-11-r124. Epub 2009 Nov 6 (2009).
- 24 Washietl, S., Kellis, M. & Garber, M. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res.* **24**, 616-628 (2014).
- 25 Ponting, C. P. The functional repertoires of metazoan genomes. *Nat. Rev. Genet.* **9**, 689-698 (2008).
- 26 Necsulea, A. *et al.* The evolution of lincRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635-640 (2014).
- 27 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).
- 28 Ulitsky, I. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat. Rev. Genet.* **17**, 601-614 (2016).
- 29 Hezroni, H. *et al.* A subset of conserved mammalian long non-coding RNAs are fossils of ancestral protein-coding genes. *Genome Biol.* **18**, 162-017-1293-0 (2017).
- 30 Davis, M. P. *et al.* Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution. *EMBO Rep.* **18**, 1231-1247 (2017).
- 31 Soumillon, M. *et al.* Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell. Rep.* **3**, 2179-2190 (2013).
- 32 Chen, L. L. The biogenesis and emerging roles of circular RNAs. *Nat. Rev. Mol. Cell Biol.* **17**, 205-211 (2016).
- 33 Amin, V. *et al.* Epigenomic footprints across 111 reference epigenomes reveal tissue-specific epigenetic regulation of lincRNAs. *Nat. Commun.* **6**, 6370 (2015).
- 34 Cabili, M. N. *et al.* Localization and abundance analysis of human lincRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20-015-0586-4 (2015).
- 35 Wilusz, J. E. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. *Biochim. Biophys. Acta* **1859**, 128-138 (2016).
- 36 Shafik, A., Schumann, U., Evers, M., Sibbritt, T. & Preiss, T. The emerging epitranscriptomics of long noncoding RNAs. *Biochim. Biophys. Acta* **1859**, 59-70 (2016).
- 37 Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187-1200 (2017).
- 38 Jacob, R., Zander, S. & Gutschner, T. The Dark Side of the Epitranscriptome: Chemical Modifications in Long Non-Coding RNAs. *Int. J. Mol. Sci.* **18**, 10.3390/ijms18112387 (2017).
- 39 Xiong, X., Yi, C. & Peng, J. Epitranscriptomics: Toward A Better Understanding of RNA Modifications. *Genomics Proteomics Bioinformatics* **15**, 147-153 (2017).
- 40 Li, Y., Syed, J. & Sugiyama, H. RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell. Chem. Biol.* **23**, 1325-1333 (2016).

- 41 Goff, L. A. & Rinn, J. L. Linking RNA biology to lncRNAs. *Genome Res.* **25**, 1456-1465 (2015).
- 42 Kim, T. K. & Shiekhattar, R. Diverse regulatory interactions of long noncoding RNAs. *Curr. Opin. Genet. Dev.* **36**, 73-82 (2016).
- 43 Marchese, F. P., Raimondi, I. & Huarte, M. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol.* **18**, 206-017-1348-2 (2017).
- 44 Kornienko, A. E., Guenzl, P. M., Barlow, D. P. & Pauler, F. M. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* **11**, 59-7007-11-59 (2013).
- 45 Vance, K. W. & Ponting, C. P. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet.* **30**, 348-355 (2014).
- 46 Xing, Z. *et al.* lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* **159**, 1110-1125 (2014).
47. Kartha, R. V. & Subramanian, S. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. *Front. Genet.* **5**, 8 (2014).
- 48 Yoon, J. H., Abdelmohsen, K. & Gorospe, M. Posttranscriptional gene regulation by long noncoding RNA. *J. Mol. Biol.* **425**, 3723-3730 (2013).
- 49 Swerdlow, S. *et al.* in *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Fourth Edition* (IARC, Geneva, 2008).
- 50 Swerdlow, S. H. *et al.* The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* **127**, 2375-2390 (2016).
- 51 Kuppers, R. Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer.* **5**, 251-262 (2005).
- 52 Qian, J. *et al.* B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* **159**, 1524-1537 (2014).
- 53 Meng, F. L. *et al.* Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* **159**, 1538-1548 (2014).
- 54 Aukema, S. M. *et al.* Double-hit B-cell lymphomas. *Blood* **117**, 2319-2331 (2011).
- 55 Savage, K. J. *et al.* MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood* **114**, 3533-3537 (2009).
- 56 Chang, T. C. *et al.* Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat. Genet.* **40**, 43-50 (2008).
- 57 Klapproth, K. & Wirth, T. Advances in the understanding of MYC-induced lymphomagenesis. *Br. J. Haematol.* **149**, 484-497 (2010).
- 58 Calado, D. P. *et al.* The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nat. Immunol.* **13**, 1092-1100 (2012).
- 59 Dominguez-Sola, D. *et al.* The proto-oncogene MYC is required for selection in the germinal center and cyclic reentry. *Nat. Immunol.* **13**, 1083-1091 (2012).



