Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Alena Blank-Giwojna

born in: Grodno, Belarus

Oral examination**:**

Long non-coding RNA *KHPS1* drives a regulatory cascade

which activates expression

of the protooncogene *SPHK1*

Referees: Prof. Dr. Ingrid Grummt

Prof. Dr. Dr. Georg Stoecklin

“For an idea that does not first seem insane, there is no hope.”

— Albert Einstein

Acknowledgments

First and foremost, I would like to express my sincere gratitude to Prof. Dr. Ingrid Grummt for the opportunity to complete my PhD study in her group. Dear Ingrid, thank you for giving me the most important thing for a young scientist – freedom to pursue my research, while non-intrusively guiding me, making sure that I am staying on course and do not deviate from the core of my research. I am eternally grateful for your constant willingness to help, continuous support and inspiration to think outside of the box, as well as for motivation to challenge own hypotheses. I truly appreciate your support in writing this thesis. Without your guidance this thesis would not have been possible. I am thankful for providing an amazing example of successful woman scientist and professor.

My appreciation also goes to Prof. Dr. Dr. Georg Stoecklin for valuable scientific advices during my PhD research.

My next gratitude goes to Dr. Ania Postepska-Igielska. Thank you for your insightful guidance, for your credible ideas and discussions that helped me to achieve my goal. It was a great experience to work with you in one team.

I would also like to thank Nina and Chrissy for their energy and positive attitude that they were sharing. Thanks for all the cakes and plus cm in waist. And of course, thank you for all the help in the lab and in private life.

I would also like to express my gratitude to Aishwarya and Nevcin. Thank you, girls, for your support, for good times, for scientific discussions, for all the team work. It was a great pleasure to grow with you.

My great appreciation goes to Anne and Monika. Thank you for being so helpful and understanding. Thank you for the department to run smoothly and orders being on time.

I would like to acknowledge the former and current members of A030. Dear Khelifa, Zhao, Bettina, Jeanette, Holger, Chenlin. Thank you for a great working atmosphere, for your support. Our laughters were helping to overcome the grief from the failed experiments. And thanks to our scientific discussions I could fix the failed experiments.

I would like to dedicate this work to my big family. Mom, dad, Marta, Max, Ania A., Grandmothers, Astrid, Bene, Opi - you were the greatest source of my strength. Your love, care and faith made tremendous contribution to my work. Whenever I didn't have enough power to continue, you were there for me to remind me about my dream of getting this degree. Мама, тата і Марта, я прысьвячаю гэтую працу вам. Толькі дзякуючы вашай падтрымцы я зрабіла гэта. Мы зрабілі гэта.

Table of Contents

Summary

Numerous long noncoding RNAs (lncRNAs) have been discovered, however only a small number of lncRNAs have been explored with respect to their function and little is known how they operate at chromatin. In this study, the function of an E2F1-regulated lncRNA, termed *KHPS1*, has been investigated. *KHPS1* is transcribed in antisense orientation from the *SPHK1* (Sphingosine kinase 1) promoter which in sense orientation directs transcription of SPHK1 mRNA. The results demonstrate that *KHPS1* activates *SPHK1* transcription by recruiting histone acetyltransferase p300 and the transcription factor E2F1 to the *SPHK1* enhancer. Binding of p300 and E2F1 is required for transcription of an enhancer-derived RNA, eRNA-Sphk1. Transcription of eRNA-Sphk1 evicts CTCF, a factor that insulates the *SPHK1* enhancer from the *SPHK1* promoter and facilitates *SPHK1* expression.

Importantly, the direct association of *KHPS1* with a homopurine stretch upstream at the *SPHK1* enhancer is essential for *SPHK1* expression. Binding of *KHPS1* to the *SPHK1* enhancer is mediated via Hoogsteen base pairing, forming RNA-DNA-DNA triplex structure. Tethering *KHPS1* and associated p300 and E2F1 to the *SPHK1* enhancer is a prerequisite for activation of eRNA-Sphk1 transcription and expression of *SPHK1*. The functional relevance of triplex formation was further studied using reporter plasmids which mimic *KHPS1*-dependent transcription activation of *SPHK1*. Deletion or mutation of the triplex forming region (TFR) attenuated recruitment of p300 and E2F1 and compromised transcription of eRNA-Sphk1. Replacement of the TFR by foreign triplex-forming motifs from lncRNAs *MEG3* or *Fendrr* functionally replaced the TFR of *SPHK1,* i.e. activated eRNA-Sphk1 transcription. Ectopic *KHPS1* with the TFR of *MEG3* targeted E2F1 and p300 to the *MEG3* target gene *TGFBR1*, underscoring the functional relevance of triplex-forming sequences. Genomic deletion of the TFR or intervention of *KHPS1* binding to DNA by ectopic TFR-containing RNA decreased *SPHK1* expression and impaired cell viability. Collectively, the results unravel a triplex-dependent regulatory feed-forward mechanism, involving lncRNA-mediated activation of eRNA which enhances expression of its target gene. The results underscore the pivotal role of triplex formation in transcription control, supporting a model whereby lncRNAs tethered to specific loci serve as sequence-specific molecular anchors.

Zusammenfassung

Trotz der Vielzahl an bereits bekannten langen nicht-kodierenden RNAs (lncRNA) wurden bislang nur die wenigsten umfangreich funktionell klassifiziert. In dieser Arbeit wurde die Funktion von *KHPS1*, einer E2F1-regulierten lncRNA, untersucht. *KHPS1* wird in Antisense-Orientierung vom *SPHK1* (Sphingosine kinase 1) Promoter transkribiert, der für SPHK1 mRNA kodiert. Die Ergebnisse zeigen, dass *KHPS1* die Transkription von SPHK1 mRNA durch Rekrutierung der Histon-Acetyltransferase p300 und des Transkriptionsfaktors E2F1 zum Enhancer des *SPHK1* Gens erhöht. Die Bindung von p300 und E2F1 führt zur Synthese einer von der Enhancer-Region kodierten RNA. Diese eRNA-Sphk1 verdrängt den Faktor CTCF, welcher den Enhancer und den *SPHK1* Promoter voneinander isoliert, und fördert damit die Expression von *SPHK1*. Dieser Mechanismus basiert auf der Assoziation von *KHPS1* mit einer Purin-reichen Region des Enhancers und der damit einhergehenden Bildung einer Triple-helikalen DNA-RNA Struktur. Die Bindung von *KHPS1* an den Enhancer ist für die *KHPS1*-abhängige Synthese von eRNA-Sphk1 und SPHK1 mRNA erforderlich. Die funktionelle Bedeutung dieser Triplex-Bildung wurde mithilfe von Reporterplasmiden, die die Situation am endogenen *SPHK1* Lokus simulieren, untersucht. Deletion oder Mutation der Triplex-bildenden Region (TFR) resultierte in einer abgeschwächten Bindung von p300 und E2F1 an den Enhancer und damit zu einer verminderten Transkription der eRNA-Sphk1. Genomweite Deletion der TFR oder Inhibierung der *KHPS1* Bindung durch ektopische Expression von TFR-enthaltender RNA verminderte sowohl die *SPHK1* expression wie auch die Lebensfähigkeit der Zellen. Austausch der TFR des Reporterplasmids mit bekannten TFR Sequenzen, wie zum Beispiel von *MEG3* oder *Fendrr,* führte ebenfalls zur Aktivierung der eRNA-Sphk1 Transkription und im Falle von *MEG3* zur Rekrutierung von *KHPS1* zusammen mit p300 und E2F1 zum MEG3 Zielgen *TGFBR1*. Zusammengefasst decken diese Ergebnisse einen Triplex-abhängigen "feed-forward" Mechanismus auf, zwischen lncRNAs, eRNAs und der Transkription ihrer Zielgene. Die Ergebnisse unterstreichen die zentrale Rolle von DNA-RNA Triplexen bei der Regulierung der Transkription und unterstützen die Hypothese, dass lncRNAs die Assoziation von Chromatin-modifizierenden Enzymen und Transkriptionskofaktoren sequenzspezifisch an designierte genomische Regionen leitet.

1. Introduction

1.1. Long non-coding RNAs implicated in regulation of cellular processes

Over the past decade, high-throughput sequencing has revealed that less than 3% of the genome encodes for proteins (Dunham et al., 2012). The remaining majority is transcribed into non-protein coding transcripts, termed noncoding RNA or ncRNA. Among those transcripts a distinct class of ncRNAs, which is more than 200 nt and up to 100 kilobases (kb) in length is defined as long non-coding RNAs (lncRNAs). The majority of lncRNAs is transcribed by RNA polymerase II (Pol II), spliced and polyadenylated. However, while mRNAs are predominantly localized in the cytoplasm, lncRNAs are often enriched in the nucleus (Derrien et al., 2012).

Despite a large number of lncRNA are transcribed in mammals (Harrow et al., 2012; Zhao et al., 2016), the function of most lncRNAs remains unknown. Recent reports provide evidence that lncRNAs are functionally relevant and implicated in control of cellular homeostasis (Salviano-Silva et al., 2018). The revealed regulatory function of lncRNAs led to abandoning the hypothesis of non-protein coding RNAs being "junk" or "transcriptional noise". LncRNAs were shown to be expressed in a tissue-specific and highly regulated manner upon different environmental stimuli, assuring cell- and tissue-specific regulation of gene expression without changing the transcriptional machinery of the cell (Dempsey and Cui, 2017). Many studies have revealed that lncRNAs act at multiple levels of gene expression, regulating transcriptional, post-transcriptional and post-translational processes and affecting diverse functions such as cell cycle, survival, immune response or pluripotency (Fang and Fullwood, 2016).

Due to the ability of lncRNAs to hybridize to complementary RNAs, some lncRNAs were shown to act at the post-transcriptional level to regulate splicing, mRNA turnover, protein translation or to serve as molecular decoys for microRNAs (Faghihi and Wahlestedt, 2009). For example, lncRNA Uchl1 (ubiquitin carboxy-terminal hydrolase L1) is transcribed in antisense orientation to the *Uchl1* gene and increases UCHL1 protein synthesis, thereby contributing to neuronal development and function. This is achieved by RNA-RNA interaction between Uchl1 mRNA and lncRNA Uchl1, which contains repetitive SINEB2 element that can activate polysomes for increased translation (Carrieri et al., 2012). Other studies demonstrated that some lncRNAs contain microRNA response elements (MREs), thus serving as microRNA sponge which competes for microRNA binding to protein-coding transcripts (Salmena et al., 2011). Likewise, lncRNAs TUG1 and CTB-89H12.4 have been shown to function as sponges

for PTEN-targeting microRNAs, therefore regulating expression of *PTEN* in prostate cancer (Du et al., 2016).

Cytoplasmic lncRNAs can modulate post-translational modifications of proteins. A recent study has identified lncRNA LINK-A (long intergenic non-coding RNA for kinase activation), which mediates HIF1-α phosphorylation. LINK-A-dependent activation of BRK (Breast-tumor Kinase) leads to phosphorylation and stabilization of HIF1-α, resulting in activation of HIF1-α transcriptional programs under normoxic conditions, thereby promotes breast cancer tumorigenesis (Lin et al., 2016). Furthermore, lncRNA DINO (Damage Induced Noncoding) regulates the levels of p53 protein. DINO stabilizes p53 through interaction with it, ensuring DNA damage response (Schmitt et al., 2016).

LncRNAs that are enriched in the nucleus can regulate transcription of specific genes by affecting the chromatin structure via recruitment or sequestration of epigenetic modifiers. Some lncRNAs can repress transcription, like PAPAS (promoter and pre-rRNA antisense), which guides the histone methyltransferase Suv4-20h2 and chromatin remodeling complex CHD4/NuRD to the rDNA promoter, thereby leading to the establishment of heterochromatin and repression of rDNA transcription in cells that underwent quiescence or heat shock (Bierhoff et al., 2014; Zhao et al., 2018). Another lncRNA, which is termed Dali and expressed in the central nervous system, regulates neural gene expression through controlling the interaction of DNA methyltransferase DNMT1 to affect DNA methylation at distal DNA regions (Chalei et al., 2014).

Important classification of lncRNAs comes from their ability to regulate transcription either in *cis* or in *trans*. Some lncRNAs work in *cis,* that is near their sites of synthesis, regulating expression of the nearby genes on the same chromosome from which they are transcribed. For example, *cis*-acting lncRNA Air silences the neighboring imprinted genes *Igf2r/Slc22a2/Slc22a2* (Nagano et al., 2008). *Trans*-acting lncRNAs, on the other hand, affect genes distant from their locus of expression. The lncRNA Jpx is transcribed from the active X chromosome and regulates transcription *in trans* of the *Xist* gene on the inactive X during inactivation of X chromosome (Sun et al., 2013b). Accumulated evidence of lncRNAs being engaged in almost every cellular process suggests that the functional complexity of the mammalian genome might be greatly attributed to lncRNAs.

1.2. Long non-coding RNAs control gene expression on multiple levels

Regulation of transcription is considered as interplay of chromatin modifying enzymes and DNA binding transcription factors (TFs), which occupy regulatory DNA elements and facilitate

the assembly of the transcription machinery at regulatory elements. Several lncRNAs have been demonstrated to interact with chromatin-remodeling complexes. Some lncRNAs have been implicated in the recruitment of transcriptional machinery or in architectural conformation of chromatin, functioning as repressors or activators (Balas and Johnson, 2018).

1.2.1. LncRNAs regulate gene transcription

Recent studies have demonstrated that lncRNAs can modulate binding of transcription factors (TFs) to DNA. According to the most recognized model, TFs are the DNA-binding proteins that recognize and bind specific genomic sequences. However, recent studies have reported that initially categorized DNA-binding transcription factors are being equally able to bind RNA (Cassiday and Maher, 2002; Sigova et al., 2015). In this regard, several lncRNAs have been reported to regulate gene expression through binding of previously known DBPs. For example, the lncRNA linc-YY1 has been shown to interact with the transcription factor yin yang 1 (YY1) to modulate its activity and contribute to the skeletal myogenesis (Zhou et al., 2015). Another lncRNA RMST has been shown to physically interact with sex-determining region Y-box 2 (SOX2), leading to co-regulation of genes involved in neurogenesis (Ng et al., 2013). The increasing number of lncRNAs that interact with TFs and modulate binding of these TFs to DNA suggests an important role of lncRNA in transcription regulation.

1.2.2. Active enhancers as a source of ncRNAs

The assembly of the transcription machinery at gene promoters is controlled by regulatory elements, for instance enhancer regions. Enhancers are classified as distal regulatory elements that bind transcription factors and promote the formation of the pre-initiation complex at their distal target genes (Bulger and Groudine, 2011; Ong and Corces, 2011). Recent advances in genomics suggest that the human genome contains millions of enhancers that can be activated at different developmental stages and in various tissues and cell types (Calo and Wysocka, 2013).

Enhancers can exert their regulatory function on distal promoters via promoter-enhancer interactions, which are mediated by CCCTC-binding factor (CTCF) (Ren et al., 2017b). Accordingly, depletion of CTCF decreases expression of the enhancer target genes. For instance, loss of CTCF binding reduced interaction of the *Sox2* gene with the ES-specific enhancer, leading to decreased transcription of the SOX2 mRNA (Zhou et al., 2014). Furthermore, ChIP-seq analyses of CTCF occupancy in the genome revealed that the enhancers that are involved in interactions with the promoters are highly enriched in CTCF as compared

to the enhancers without CTCF (Ren et al., 2017b). Importantly, CTCF can act as insulator that separates active chromatin from inactive, thereby inhibiting the communication between enhancers and their target genes. For example, CTCF is located between the promoter of Vascular endothelial growth factor *(VEGF*) and the hypoxia responsive element (HRE). Depletion of CTCF, which blocks HRE-enhancer activity, led to increased expression of VEGF during hypoxia and excessive VEGF-mediated angiogenesis (Lu and Tang, 2012). Furthermore, CTCF regulates transcription at the imprinted gene loci of the insulin-like growth factor 2 (Igf2)/H19. *H19* on the maternal allele and *Igf2* on the paternal allele are regulated by the same enhancer. Binding of CTCF blocks Igf2 gene activation by the enhancer, allowing expression of the *H19* (Bell and Felsenfeld, 2000). Taken together, these results demonstrate that CTCF acts as an enhancer blocker to restrain the repress the transcription of the target gene.

Despite a number of studies demonstrated versatile roles of CTCF in control of gene expression by either insulation of enhancers or by promoting enhancer-promoter interactions, the mechanism underlying regulation of CTCF remains elusive.

Analyses of global genomic studies revealed common features determining the distal regulatory regions, such as a high ratio of histone H3 lysine 4 monomethylation (H3K4me1) to trimethylation (H3K4me3), as well as the acetylation of histone H3 on lysine 27 (H3K27Ac) and occupancy of the histone acetyltransferase (HAT) p300 (Calo and Wysocka, 2013; Heintzman et al., 2009; Rada-Iglesias et al., 2011; Raisner et al., 2018). Importantly, H3K27ac distinguishes active enhancers from inactive/poised regulatory elements (Creyghton et al., 2010). Such poised enhancers lack H3K27ac but are enriched in H3K4me1 and H3K27me3. Poised enhancers were originally found near key developmental genes and were shown to have no impact on gene expression in pluripotent cells. Importantly, inactive enhancers were shown to acquire ability to drive the gene expression during differentiation, which is accompanied with loss of H3K27me3 and gain of H3K27ac (Rada-Iglesias et al., 2011). Such active state of chromatin allows the binding of transcription factors which in turn recruit Pol II to enhancers (Natoli and Andrau, 2012).

Recent studies revealed that a subclass of lncRNAs is transcribed from active enhancers, called enhancer-derived RNAs (eRNAs). A number of studies demonstrated the importance of eRNAs for enhancers activity (Kim et al., 2010; Orom et al., 2010; de Santa et al., 2010). eRNAs are characterized by nuclear localization, preferential lack of polyadenylation signal (polyA-) as well as low stability of the transcripts (Espinosa, 2016; Kim et al., 2010).

The key question remains whether and how eRNA contribute to enhancer function. Although

some studies demonstrated a strong correlation between eRNA synthesis and enhancer activity, it remains obscure if these two processes are mechanistically linked. On the one hand, eRNAs were thought to be just a byproduct of the transcription caused by recruitment of Pol II (Rahman et al., 2016). On the other hand, a substantial number of studies revealed functional importance of eRNAs. For instance, knock-down of 17β-estradiol (E2)-induced eRNAs compromised expression of E2-upregulated target genes and reduced enhancer-promoter interactions (Li et al., 2013). This work suggested a role of eRNAs in initiation or stabilization of enhancer-promoter interactions. Consistent with eRNAs mediating the interaction of distant genes, an eRNA, termed AS1eRNA, was shown to enhance transcription of the lncRNA DHRS4-AS1 by eRNA-driven looping of the enhancer and the *DHRS4-AS1* promoter (Yang et al., 2016). Knock-down of eRNAs transcribed from *myoD1* decreased recruitment of Pol II to the promoter of *MyoD1*, but not at the enhancer itself, suggesting that eRNA transcripts facilitate Pol II binding to the promoter of the target genes (Mousavi et al., 2013). Altogether, these studies underscore the functional relevance of eRNAs, thereby expanding the list of functions mediated by lncRNA to control gene transcription programs.

1.2.3. LncRNAs are important players in epigenetic regulation

DNA in the nucleus is compacted into chromatin with the organizational scaffold consisting of nucleosomes, each with two copies of histones H3, H4, H2A and H2B (Hayes and Hansen, 2001). Nucleosomes carry post-translational modifications of histone tails brought about by chromatin modifying enzymes. Histone modifications can determine the activity state of the promoter and influence binding of transcription factors (Luger and Richmond, 1998). Some modifications, such as H3 and H4 acetylation and trimethylation of histone H3 lysine 4 (H3K4me3) are characteristics of active promoters. Conversely, trimethylation of histone H3 lysine 27 (H3K27me3) leads to densely packed nucleosomes interfering with protein-DNA interactions and correlating with gene silencing (Handy et al., 2011).

Recent studies have revealed that lncRNAs can act as essential regulators of chromatin structure due to their ability to interact with chromatin modifiers and recruiting them to the gene promoter. For example, lncRNA MALAT1 directly interacts with EZH2, an enzymatic component of the Polycomb Repressive Complex 2, and represses transcription of cyclin-dependent kinases (CDKs) p21 and p27. This leads to the development of mantle cell lymphoma (Wang et al., 2016). In contrast, lncRNA HOTTIP (HOXA transcript at the distal tip) coordinates activation of several HOXA genes, which are required for animal development. HOTTIP guides the histone methyltransferase complex WDR5-MLL to the HOXA genes,

thereby establishes transcriptionally permissive chromatin structure and activates the HOXA locus (Wang et al., 2011).

A number of studies have shown that lncRNAs also play a role in DNA methylation. DNA methylation is established by the DNA methyltransferases (DNMTs) within the CpG sites leading to gene silencing. Some lncRNAs have been identified to physically interact with DNMTs to regulate DNA methylation. Promoter-associated RNA (pRNA), a lncRNA that is complementary to the promoter of ribosomal RNA (rRNA) genes, silences transcription of murine rRNA genes by targeting DNMT3B, which methylates the promoter of the *rDNA* locus (Schmitz et al., 2010). In contrast, lncRNAs also play a role in gene activation causing demethylation of silent promoters. For example, an antisense lncRNA, termed TARID (TCF21 antisense RNA inducing demethylation), activates transcription of *TCF21* by inducing demethylation of the gene promoter. Mechanistically, it is achieved via TARID-mediated recruiting of the DNA damage factor GADD45A, which then recruits TDG together with TET proteins to direct base excision repair for demethylation (Arab et al., 2014).

A number of lncRNAs have been implicated in epigenetic control of gene expression by recruiting chromatin modifiers, however how the enzymatic complexes are guided to their place of action remains elusive. It has been postulated that noncoding RNAs may target regulatory proteins to their sites of action, providing specificity in control of gene expression.

1.3. Direct interactions between lncRNAs and DNA: R-loops and RNA-DNA-DNA triple helices

The complexity and biochemical properties of RNAs allow lncRNAs to interact with both nucleic acids and regulatory proteins, enabling lncRNA-mediated transcription regulation in a sequence-specific manner. To exert the vast spectrum of regulatory functions, lncRNA utilize different archetypes of molecular functions (Wang and Chang, 2011). LncRNAs can serves as ‘guides’ to recruit regulatory factors to target genes and/or act as ‘scaffold’ to assemble proteins into ribonucleoprotein (RNP) complexes. The “guiding” or “scaffolding” mechanisms include direct interaction of lncRNAs with specific target DNA sequences. Such lncRNA-DNA interactions involve formation of structures such as RNA-DNA heteroduplexes (so called R-loops) or RNA-DNA-DNA triplexes (Figure 1). These interactions might be the key for target recognition by lncRNA and associated regulatory proteins.

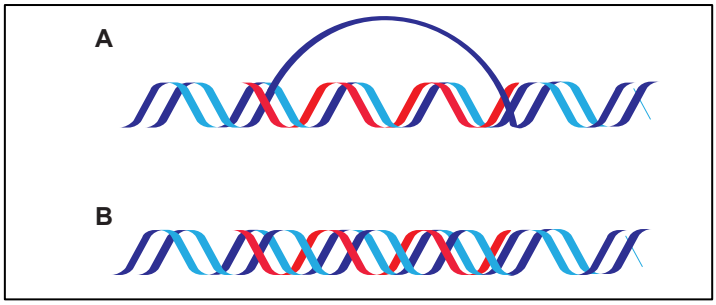

Figure 1. Schematic representation of R-loop and triple-helical structure. RNA (red) is associated with DNA (blue) either single-stranded forming RNA-DNA heteroduplex (**A**) or double stranded forming RNA-DNA-DNA triplex (**B**).

1.3.1. R loops are structures formed by single stranded DNA and RNA

R-loops are nucleic acid structures that form during transcription by reannealing of the nascent RNA to the single-stranded DNA template, giving rise to an RNA–DNA hybrid (Thomas et al., 1976). R-loop structures rely exclusively on Watson-Crick base-paring and form preferentially at C-rich sequences of DNA (Reaban et al., 1994; Roy and Lieber, 2009). Once formed, R-loops are thermodynamically stable and are removed by RNA-DNA helicases and RNase H (Cerritelli and Crouch, 2009; Song et al., 2017; Sugimoto et al., 1995). Due to ability of RNA to bind to DNA in a sequence-specific manner, R-loop structures could serve as a mechanism for lncRNA-mediated guidance of regulatory proteins.

R-loops formed by lncRNAs have been shown to regulate the expression of the neighboring protein-coding genes. For instance, the lncRNA VIM-AS1, which forms an R-loop at the promoter of the gene encoding vimentin (VIM), leads to activation of the NF-κB pathway via opening of chromatin which facilitates binding of transcriptional activators (Boque-Sastre et al., 2015). R-loops were also reported to repress transcription. In *Arabidopsis thaliana* expression of the FLOWERING LOCUS C (FLC) is negatively regulated by an lncRNA termed COOLAIR, that is transcribed into antisense orientation to the *FLC* gene. FLC RNA forms an R-loop in the promoter region of *COOLAIR*, which represses COOLAIR transcription. Consequently, FLC is expressed leading to induction of flowering (Sun et al., 2013).

Despite the increasing number of cis-acting lncRNA engaged into R-loop formation, the involvement of lncRNAs operating in *trans* via R-loops remains controversial (Cloutier et al., 2016). In addition, when unproperly resolved, R-loops can lead to DNA damage, causing double-strand breaks (DSBs) and leading to genome instability (Aguilera and García-Muse, 2012; Song et al., 2017). Thus, it is likely that a number of lncRNAs utilize other mechanisms to regulate transcription on the remote loci and to avoid exposing cells to genomic instabilities.

1.3.2. Triplex structures

Triple-helices, or triplexes, are structures composed of three nucleic strands of either RNA or DNA. The existence of triplexes was first proposed in 1957, when Felsenfeld and colleagues observed that a strand of a polyriboadenylic acid (polyrA) was able to associate with two strands of polyribouridylic acid (polyrU) forming a three-stranded structure in the presence of Mg^{2+} (Felsenfeld et al., 1957). The association of one strand with the duplex occurs via hydrogen bond interactions, that are distinct from classical Watson-Crick base-pairing rules, termed Hoogsteen (Hoogsteen, 1963). Hoogsteen basepairing engages a purine base from a duplex, in which the N7 exerts a function of a hydrogen bond acceptor and C6 group serves as a donor (Figure 2).

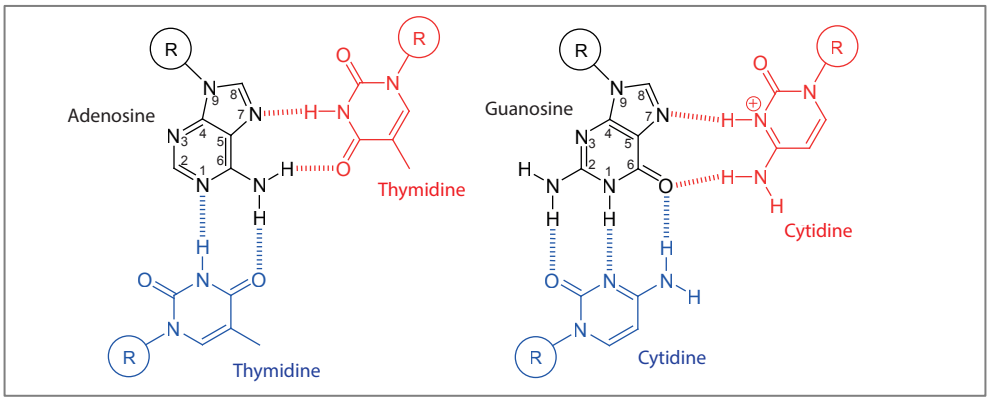

Figure 2. Triplex-helices are formed by Hoogsteen hydrogen bonds with the purine-rich strand. A Hoogsteen base pair engages the N7 position of the purine base as a hydrogen bond acceptor and C6 group as a donor. C6 and N1 groups simultaneously bind the N3-N4 of another pyrimidine via Watson-Crick basepairing. Hoogsteen base-pairing is indicated as red lines, Watson-Crick as blue lines.

Triple-helices are formed between the purine-rich strand of duplex DNA and either a pyrimidine-rich or a purine-rich third strand of nucleic acid in the major groove of DNA. The third strand, regardless whether RNA or ssDNA, can bind the duplex in two possible ways: parallel (forward Hoogsteen basepairing), that is in the same 5' to 3' direction to a purine-rich strand of the duplex, or antiparallel (reverse Hoogsteen basepairing) which is the opposite to the 5' to 3' direction of the purine-rich strand of dsDNA. If the third strand contains a polypyrimidine T(U)C motif, it preferably binds in parallel orientation. The polypurine AG motif is placed into anti-parallel direction, and GT(U) sequences form either parallel or anti-parallel triplex structures (Figure 3), indicating that triplexes can be formed by both purine- and pyrimidine-rich nucleic acids.

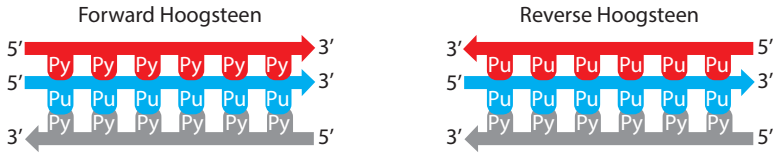

Figure 3. Orientation of triplex formation. If the third nucleotide strand (red) is pyrimidine-rich, it binds in parallel orientation to the purine stretch of duplex via forward Hoogsteen basepairing. If the third strand contains purine-rich sequence, it binds in anti-parallel orientation to the purine tract of the duplex via reverse Hoogsteen basepairing.

Triplexes are relatively unstable under physiological conditions, as binding of the third strand to the duplex creates unfavorable repulsion between three negatively charged strands (Duca et al., 2008). Thus, in order to increase the stability of triplexes, multivalent cations such as Mg^{2+} are used for triplex formation to neutralize the charge repulsion (Floris et al., 1999). Triplexes that are formed in parallel orientation particularly favor a low pH, which leads to cytosine protonation at the N3 position in order to deliver a hydrogen bond donor for the purine residue during formation of the CG-C triplet (Roberts and Crothers, 1996). In contrast, interaction of a purine-rich third strands with duplex forming antiparallel A-AT and G-GC triplexes are stabilized by monovalent cations such as K^+ (Cheng and van Dyke, 1993). These variable conditions for triplex formation create a challenge for the development of methods to monitor triple-helices both in vitro and in vivo.

Current knowledge about the structure, thermodynamics and kinetics of triple-helices has been mainly obtained from the studies on DNA-DNA-DNA triplexes (Plum et al., 1990; Protozanova and Macgregor, 1996; Roberts and Crothers, 1996). Intensive studies of triplexes led to therapeutic applications of triplex forming oligonucleotides (TFOs), which are able to interact with dsDNA in a sequence-specific manner. Due to binding specificity of TFOs, they are capable to target regulatory regions in the genome serving as sequence-specific regulators of gene expression.

TFOs have been shown to affect gene expression in vitro through targeted binding and prevention of binding of either transcription factor or initiation complex formation (Karympalis et al., 2004; Svinarchuk et al., 1996). Transcription of several genes, including the protooncogenes *c-myc* or *ets2,* was inhibited using TFOs. Binding of TFO to the purine-rich region within the *c-myc* promoter led to inhibition of *c-myc* promoter-driven reporter luciferase expression in HeLa cells (Kim and Miller, 1998). Transfection of Ets2-TFO led to impaired *Ets2* promoter activity followed by attenuated expression of the endogenous gene (Carbone et al., 2004). These studies unraveled the regulatory potential of triple-helix formation brought about by sequence-specific recognition and ability to associate with the major groove of DNA.

1.3.3. LncRNAs involved in regulation of gene expression are able to form RNA-DNA triplexes

The human genome contains a large number of polypurine stretches spanning about 15-20 nucleotides capable to form Hoogsteen hydrogen bonds. Bioinformatics analyses have revealed that potential triplex forming regions (TFRs) are present throughout the genome, particularly in regulatory regions, suggesting a possible functional role of triplexes in transcription regulation (Buske et al., 2011; Goñi et al., 2004; Soibam, 2017).

Interestingly, studies of the effect of the third strand composition on triplex stability revealed that the most stable structures are obtained if the two Watson-Crick paired strands are composed of DNA and the third strand is RNA, indicating that RNAs could form stable RNA-DNA triplexes (Van Dongen et al., 1999; Escudeé et al., 1993; Han and Dervan, 1993). In vivo, these observations support the notion that lncRNAs could form triplexes at the regulatory regions serving as molecular anchors for regulatory proteins to control expression of the target genes.

The first lncRNA that was shown to form a triplex structure and regulate transcription is synthesized in the *DHFR* locus. DHFR lncRNA interacts with TFIIB leading to dissociation of the pre-initiation complex from the major promoter, resulting in transcriptional repression of the *DHFR* gene. Importantly, *DHFR* lncRNA was shown to form RNA-DNA triplex at the G-rich stretch of the major *DHFR* promoter by *in vitro* EMSA (electrophoresis mobility shift assay), providing first *in vitro* evidence that lncRNAs, that are involved in transcription regulation, are able to form triplexes (Martianov et al., 2007).

The first evidence of *in vivo* RNA-DNA triplex formation was shown for murine pRNA (promoter-associated RNA), which forms stable RNA-DNA triplex at the promoter of *rRNA* genes (Schmitz et al., 2010). pRNA prevents binding of polymerase I-specific transcription termination factor (TTF-1) to the *rDNA* promoter by interaction with the region comprising a TTF-1 binding site. Moreover, pRNA targets DNMT3b to the *rDNA* promoter, which together with displacement of TTF-1 leads to transcriptional repression of *rRNA* genes. This study not only demonstrated the presence of lncRNA-DNA triplexes in vivo but also proposed a recruiting function of triplexes for regulatory proteins.

Another lncRNA, *Fendrr,* has been shown to associate with specific DNA sequences via triplex formation in vitro. *Fendrr* acts in *trans* by recruiting the PRC2 complex to developmental genes, thus facilitating tissue differentiation (Grote et al., 2013). Likewise, lncRNA *PARTICLE* serves as a scaffold for gene-silencing machineries in response to irradiation. *PARTICLE* was proven to directly interact with DNA (O'Leary et al., 2015). Finally, lncRNA *MEG3* has been

shown to form triplexes in vivo and regulate expression of the distal regulatory elements of the genes engaged into the TGF-β (transforming growth factor β) signaling pathway by recruitment of the PRC2 complex and downregulation of target gene expression (Mondal et al., 2015).
These studies demonstrate the ability of lncRNAs to control gene expression and form triplexes, suggesting that lncRNA-mediated recruitment of regulatory proteins to specific genomic sites may be a general mechanism that shapes chromatin and regulates gene expression.

1.4. *SPHK1* as model to study lncRNA-DNA triplex-mediated transcription regulation

To investigate the role of RNA-DNA triplexes in lncRNA-dependent epigenetic regulation of gene expression, a candidate gene *SPHK1* was chosen. *SPHK1* encodes sphingosine kinase 1, which phosphorylates sphingosine to yield sphingosine-1-phosphate (S1P), an important metabolite implicated in various cellular processes, including survival, proliferation, differentiation and apoptosis (Shida et al., 2008). In accord with SPHK1 being involved in maintenance of cellular homeostasis, a number of studies demonstrated that overexpression of SPHK1 is related to cancer metastasis, reduced survival and poor prognosis (Long et al., 2015; Pan et al., 2011). Although a number of stimuli, such as growth factors and cytokines, lead to increased levels of SPHK1, the molecular mechanism underlying regulation of *SPHK1* expression remains unknown (Spiegel and Milstien, 2003).
The *SPHK1* gene locus contains numerous isoforms, suggesting that multiple regulatory networks are involved in the control of *SPHK1*. Imamura and colleagues demonstrated that transcription of Sphk1 in rat is regulated by lncRNA-mediated epigenetic control. A non-coding antisense transcript, termed *KHPS1*, has been shown to mediate demethylation of the CG-island at the tissue-dependent differentially methylated region (T-DMR) at the Sphk1 promoter (Imamura et al., 2004a). Interestingly, comparison of *SPHK1/Sphk1* sequences between human, mouse and rat revealed multiple conserved regions, including the 200 bp region corresponding to T-DMR (Imamura et al., 2004b), suggesting that antisense RNA *KHPS1*/*KHPS1* is involved in epigenetic regulation of *SPHK1* among different species. Further analysis of the human *SPHK1* locus revealed an RNA *KHPS1* transcribed though the promoter of *SPHK1-B* isoform, which transcription if positively correlated with SPHK1 levels. Importantly, closer inspection of the sequence composition of the human *SPHK1-B* promoter revealed an E2F1 binding site and the presence of a purine-rich stretch that has potential to form Hoogsteen basepairing. These observations suggest a potential regulatory role of *KHPS1* in control of carried out by forming triplex structures at the *SPHK1-B* promoter.

1.5. Objectives

Long non-coding RNAs are increasingly recognized as important modulators of gene expression. Key questions to be answered concern the mechanisms by which lncRNAs exert their regulatory function in a sequence-specific manner. The ability of RNAs to directly interact with both DNA and proteins suggests that lncRNA-mediated recruitment of regulatory proteins to specific sites in genome could be a general mechanism by which lncRNAs fine-tune the expression of distinct genes. LncRNA can interact with double-stranded DNA forming RNA-DNA triplex structures. Based on the rules underlying triplex formation, several *in silico* studies have predicted triplex formation in the genome. However, experimental evidence for the existence of RNA-DNA triplexes in vivo and their biological relevance is limited.

The aim of the present study was to investigate the mechanism underlying lncRNA-mediated transcription regulation. A candidate lncRNA *KHPS1* was chosen, which regulates expression of the protooncogene *SPHK1*. Transcription of *KHPS1* positively correlates with transcription of two SPHK1 gene isoforms, indicating that *KHPS1* activates the *SPHK1* locus. Importantly, the *SPHK1-B* promoter harbours a putative triplex forming region (TFR), suggesting *KHPS1*-dependent transcription activation involved the formation of triplexes.

To investigate the mechanism underlying *KHPS1*-dependent transcriptional regulation the following topics had to be investigated:

- Investigation of the crosstalk between sense and antisense transcripts at the *SPHK1* locus by performing loss-of-function and gain-of-function experiments;
- Investigation of the role of *KHPS1* in epigenetic regulation of *SPHK1* using chromatin immunoprecipitation (ChIP) assays;
- Identification of the functional domains of *KHPS1* that interact with regulatory proteins with RNA immunoprecipitation (RIP) and cross-linking immunoprecipitation (CLIP) assays;
- Proof of RNA-DNA triplex formation at the *SPHK1* locus by performing in vitro and in vivo triplex capture assays;
- Validation of *KHPS1*-dependent triplex formation for activation of *SPHK1* transcription on reporter plasmids;
- Manipulation of the TFR to investigate the impact of triplex formation on *SPHK1* transcription;
- CRISPR/Cas9-mediated generation of a stable cell line lacking the *SPHK1* TFR to study the importance of the TFR for *SPHK1* expression and cell proliferation.

2. Results

2.1. LncRNA *KHPS1* is transcribed through the human *SPHK1* locus

The human *SPHK1* locus contains different gene isoforms (*SPHK1* A-C), which are expressed in tissue-dependent manner giving rise to different protein isoforms (Figure 4A) (Paugh et al., 2009). Previous studies in rat and transcriptome analyses in human cells revealed an antisense transcript, named *KHPS1,* which is transcribed into antisense orientation to *SPHK1-B* isoform. In human *KHPS1* is initiated about 1500 bp downstream of the transcription start site (TSS) of isoform B (SPHK1-B) from a bidirectional promoter, which in sense orientation gives rise to the SPHK1-C isoform and *KHPS1* in antisense direction (Figure 4B).

The region upstream of the TSS of *SPHK1-B* contains a binding site for the transcription factor E2F1 and a homopurine stretch, which has the potential to be engaged into triplex formation (Figure 4B). *KHPS1* is transcribed through the promoter of *SPHK1-B,* therefore has potential to regulate transcription of *SPHK1* by binding to the *SPHK1-B* promoter via Hoogsteen basepairing forming RNA-DNA triplexes.

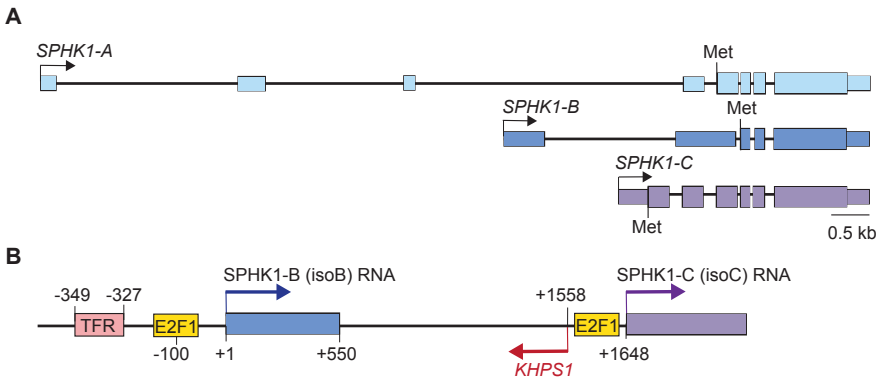

Figure 4. Scheme of the human *SPHK1* locus.
A) The transcription starts sites of *SPHK1-A, -B* and *-C* isoforms are marked by arrows. Exons of putative transcripts are presented as boxes, black lines represent introns, potential translation start sites depicted as Met.
B) The scheme showing zoom in of the *SPHK1-B* and *SPHK1-C* promoters. Transcription start sites of isoB, isoC and *KHPS1* are marked by blue, purple and red arrows, respectively. The potential triplex-forming region (TFR) (−327/−349) and E2F1 binding sites are indicated.

2.2. Transcription of *SPHK1-B* isoform is regulated by *KHPS1*

The *KHPS1* promoter contains putative binding sites for transcription factor E2F1, suggesting that members of the E2F family regulate *KHPS1* expression. In support of this notion, RNA sequencing revealed that *KHPS1* was among the lncRNAs that are upregulated upon E2F1 induction (Feldstein et al., 2013).

In order to validate E2F1-mediated activation of *KHPS1* and *SPHK1-B*, a U2OS cell line that expresses an estrogen receptor-E2F1 (ER-E2F1) fusion protein was used. ER-E2F1 is translocated to the nucleus upon treatment with 4-hydroxytamoxifen (4-OHT) and induces transcription of E2F1-responsive genes (Berkovich and Ginsberg, 2003). Treatment with 4-OHT led to increased levels of both *KHPS1* and SPHK1-B (isoB) RNAs in a time-dependent manner, demonstrating that *KHPS1* and *SPHK1-B* promoters are regulated by E2F1 (Figure 5).

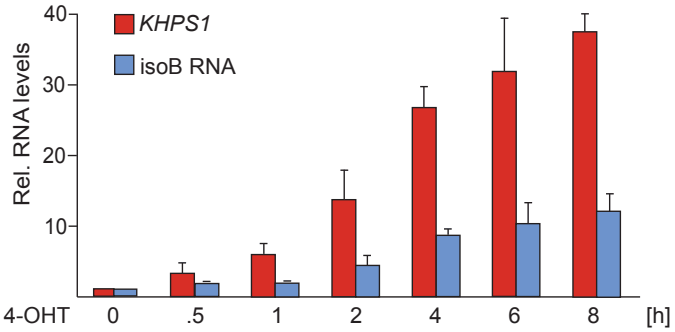

Figure 5. E2F1 activates transcription of *KHPS1* and isoB RNA.
Levels of *KHPS1* and isoB RNA were measured by RT-qPCR in U2OS/ER-E2F1 cells treated with 4-OHT for the indicated times (N=3).

The concurrent increase in *KHPS1* and isoB RNA levels upon E2F1 induction suggested that these two transcripts are either co-regulated or *KHPS1* regulates transcription of *SPHK1-B*. To distinguish between these possibilities, *KHPS1* was depleted using siRNA and levels of isoB RNA were monitored in E2F1-induced cells. Depletion of *KHPS1* led to attenuated activation of *SPHK1-B* transcription, demonstrating that E2F1 alone is not sufficient to activate *SPHK1* expression but requires the synthesis of *KHPS1* (Figure 6A).

To confirm that the diminished levels of isoB RNA are due to E2F1-mediated attenuated transcription of *KHPS1*, binding of polymerase II (Pol II) and E2F1 was monitored at the *SPHK1-B* promoter. If the cells treated with 4-OHT were transfected with siRNA targeting *KHPS1*, the association of Pol II and E2F1 at the *SPHK1-B* promoter was compromised. No change in Pol II and E2F1 occupancy was observed at the *CDC2* promoter, a well characterized E2F1 target gene (Figure 6B), indicating that *KHPS1* is required for E2F1-mediated activation of *SPHK1-B*.

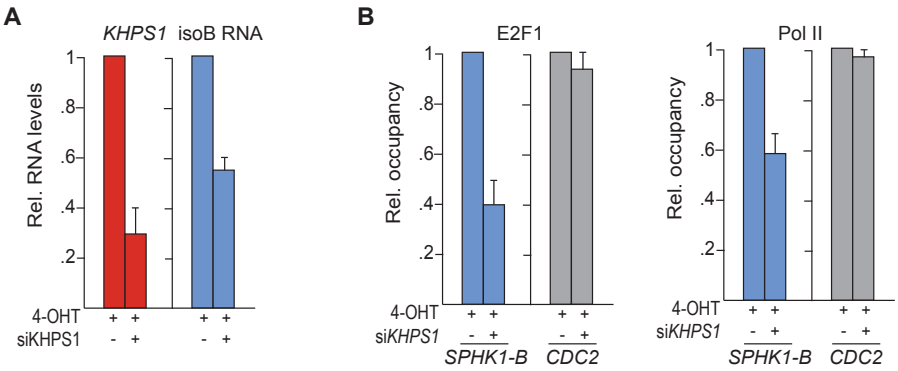

Figure 6. E2F1-dependent activation of *SPHK1-B* transcription requires *KHPS1*.
A) Levels of *KHPS1* and isoB RNA measured by RT-qPCR in U2OS/ER-E2F1 cells transfected with siRNA targeting *KHPS1* and treated with 4-OHT for 2 h (N=3).
B) ChIP showing E2F1 and Pol II occupancy at the *SPHK1-B* promoter in *KHPS1*-depleted (si*KHPS1*) U2OS/ER-E2F1 cells treated with 4-OHT for 2 h. Binding to the *CDC2* promoter was monitored as control (N=3).

2.3. *KHPS1* regulates transcription of *SPHK1-B* via chromatin remodeling

Studies in rat have shown that transcription of *Khps1* is accompanied by demethylation of the *Sphk1* promoter (Imamura et al., 2004a). However, no change in DNA methylation upon induction or knockdown of human *KHPS1* was observed (Postepska-Igielska et al., 2015), suggesting that *KHPS1*-dependent transcription activation in human is brought about by different epigenetic mechanisms.

To test whether human *KHPS1*-dependent activation of *SPHK1-B* transcription is brought about by chromatin remodeling, histone marks at the *SPHK1-B* promoter were monitored in U2OS/ER-E2F1 cells before and after E2F1 induction. Transcription of *KHPS1* coincided with an increase in active histone marks H3K9ac, H3K27ac, H4ac at the *SPHK1-B* promoter. Notably, the increase in euchromatic histone marks was attenuated when *KHPS1* was depleted prior to E2F1 induction (Figure 7). These results reinforce that transcription activation by *KHPS1* is achieved by the establishment of an active chromatin structure at the *SPHK1-B* promoter.

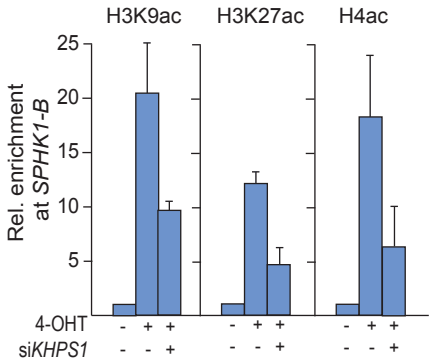

Figure 7. Establishment of active histone marks at the *SPHK1-B* promoter requires *KHPS1.*
ChIPs of the indicated histone marks monitored at the *SPHK1-B* promoter before and after 4-OHT treatment of U2OS/ER-E2F1 cells. Where indicated, cells were treated with *KHPS1*-specfic siRNA followed by 4-OHT induction (N=3).

To examine whether *KHPS1*-dependent increase in active histone marks is associated with an open chromatin structure at the *SPHK1-B* promoter, chromatin accessibility was monitored by FAIRE (formaldehyde-assisted isolation of regulatory elements) assays. FAIRE assay measures nucleosome-depleted DNA regions according to the extractability of crosslinked chromatin (Simon et al., 2012). For this, cells were crosslinked with formaldehyde and lysed. After chromatin shearing, DNA was isolated using phenol-chloroform extraction and measured by PCR. In accord with active histone modifications being established upon E2F1 induction, the fraction of nucleosome-free DNA comprising the *SPHK1-B* promoter was higher in tamoxifen-treated compared to uninduced cells. If elongation of Pol II transcription was inhibited by treatment with flavopiridol, E2F1-induced changes in chromatin extractability were compromised. Together, these results support the view that the establishment of an open chromatin structure at the *SPHK1-B* promoter is dependent on *KHPS1* transcription (Figure 8).

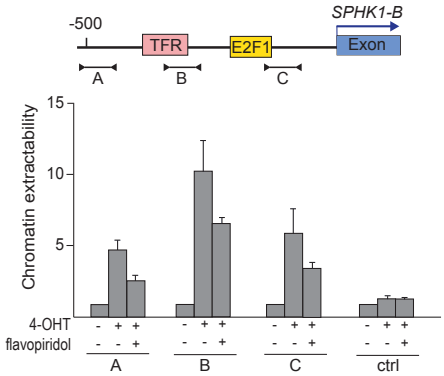

Figure 8. Transcription of *KHPS1* leads to chromatin decompaction at the *SPHK1-B* promoter.
FAIRE assay showing the levels of nucleosome-depleted *SPHK1*-B promoter region in U2OS/ER-E2F1 cells before and after E2F1 induction in the absence or presence of flavopiridol. Recovered DNA was measured by qPCR using primers covering the indicated regions. A: -592/-425; B: -406/-241; C: -175/-65; ctrl: r*DNA* promoter (N=3).

Modulation of histone acetylation is achieved by targeting of specific histone acetyltransferase (HAT) enzymes to defined gene loci. To identify the HAT that interacts with *KHPS1*, the association of *KHPS1* with the histone acetyltransferases p300, CBP, PCAF, and Tip60 was monitored by RNA immunoprecipitation (RIP) assays. Overexpressed proteins HA-, Flag-, or myc-tagged in U2OS and HepG2 cells were immunoprecipitated and protein-associated RNAs were measured by RT-qPCR. *KHPS1* was found to be preferentially associated with p300 and CBP in both cell lines, whereas no significant interaction was observed with PCAF or Tip60 (Figure 9A). RIP assays with endogenous p300 revealed that *KHPS1* was associated with p300. No interaction between p300 and HOTAIR, that served as a negative control, was observed (Figure 9B), suggesting that histone acetylation at the *SPHK1-B* promoter is brought about by p300/CBP associated with *KHPS1*.

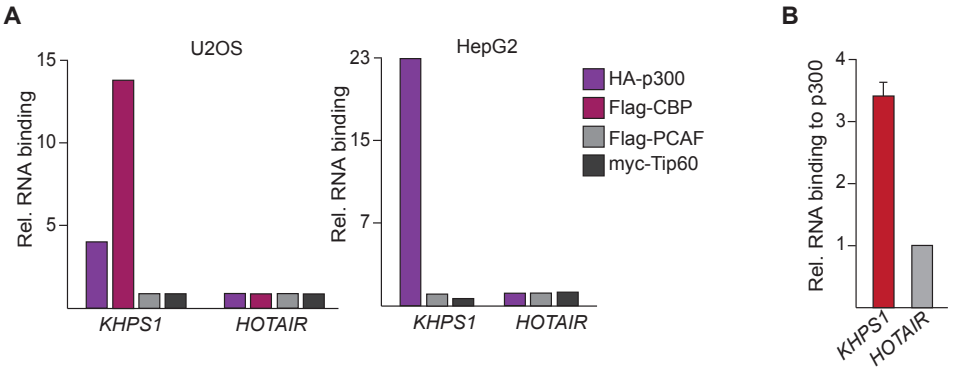

Figure 9. *KHPS1* interacts with p300/CBP.
A) RIP showing the levels of *KHPS1* and *HOTAIR* associated with the indicated HATs. Overexpressed tagged-HATs were immunoprecipitated from HepG2 and U2OS cells and RNA levels were analyzed by RT-qPCR.
B) RIP assay showing the association of *KHPS1* with p300 immunoprecipitated from 4-OHT-treated U2OS/ER-E2F1 cells. HOTAIR was used as control. Levels of RNAs were measured by RT-qPCR (N=2).

To validate *KHPS1*-mediated binding of p300 to the *SPHK1-B* promoter, p300 occupancy of was monitored upon E2F1 induction. Consistent with *KHPS1* interacting with p300/CBP, the association of p300/CBP with the *SPHK1-B* promoter was markedly enriched after E2F1-induction. If *KHPS1* was depleted prior to 4-OHT treatment, association of p300 with the *SPHK1-B* was compromised (Figure 10A). This result substantiates that *KHPS1* recruits p300 to the *SPHK1-B* promoter.

To investigate whether activation of *SPHK1* expression upon E2F1 induction was brought about by p300/CBP-dependent changes of chromatin structure or by secondary effects due to elevated E2F1 levels, cells were treated with curcumin, an inhibitor of the histone acetyltransferase p300, prior to 4-OHT treatment (Balasubramanyam et al., 2004). Inhibition of p300 activity using curcumin compromised E2F1-dependent increase of *KHPS1* and isoB

RNA transcription (Figure 10B), indicating that p300/CBP activity is required for *KHPS1*-dependent expression of *SPHK1-B*.

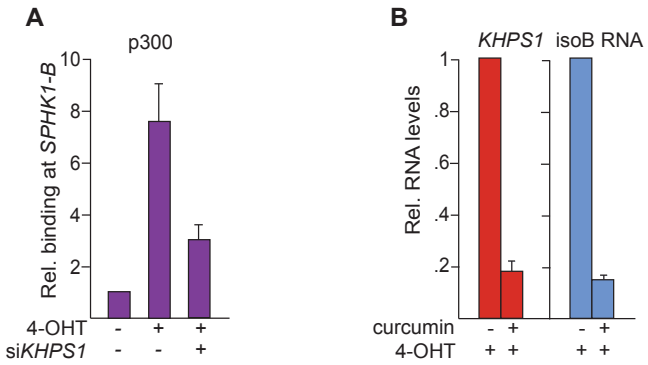

Figure 10. *KHPS1* recruits p300/CBP to the *SPHK1-B* promoter.
A) ChIP showing the binding of p300 to the *SPHK1-B* promoter in untreated U2OS/ER-E2F1 cells that were transfected with either control or siRNA targeting *KHPS1* or in the cells treated with 4-OHT for 2 h and *KHPS1*-depletion (N=3).
B) RT-qPCR showing levels of *KHPS1* and isoB RNA in U2OS/ER-E2F1 cells treated with curcumin followed by 4-OHT treatment for 4 h (N=3).

To delineate the region of *KHPS1* that conveys the interaction with p300, CLIP assays were performed. To this end, UV-crosslinked p300-RNA complexes were immunoprecipitated from 4-OHT-treated U2OS/ER-E2F1 cells and p300-bound *KHPS1* was monitored by RT-qPCR using primers that cover different regions of *KHPS1*. This experimental approach revealed that p300 preferentially bound to *KHPS1* sequences comprising the first exon of eRNA-Sphk1 (amplicon D) but not to sequences located upstream (amplicons A-C) or downstream (amplicons E-G) of exon 1 (Figure 11). No binding of p300 to HOTAIR and 18S rRNA was observed, demonstrating that p300/CBP binds to RNA in a sequence- and/or structure-dependent manner.

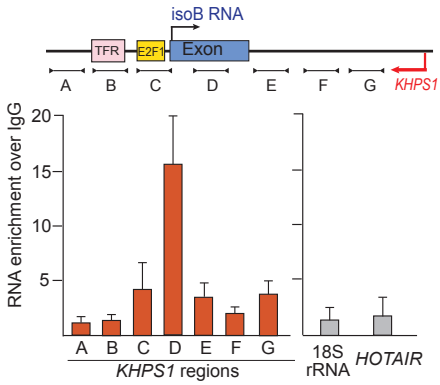

Figure 11. p300 binds to *KHPS1* at the region +108/+165 corresponding to the first exon of isoB RNA.
CLIP-qPCR showing interaction of p300 with different regions of *KHPS1*. U2OS/ER-E2F1 cells were treated with 4-OHT, UV crosslinked, lysed and p300 was immunoprecipitated. Levels of *KHPS1* associated with p300 were analyzed by RT-qPCR using primers: A:-592/-425, B: -373/-304, C: -137/-89, D: +108/+165, E: +630/+790, F:+930/+1124, G: +1132/+1242. Binding to *HOTAIR* and 18S rRNA was monitored as control. RNA enrichment was calculated as % of sample input and normalized to % of input of the IgG (N=4).

KHPS1 transcription starts about 1,500 bp downstream of the *SPHK1-B* promoter. This simultaneous bidirectional transcription would lead to head-to-head collision of RNA polymerases. To examine whether collision of polymerases is prevented by temporal separation of sense and antisense transcription, the levels of antisense and sense transcripts were monitored at different times upon E2F1 induction by 4-OHT. Treatment with 4-OHT led to rapid induction of both *KHPS1* and isoB RNA, *KHPS1* preceding activation of isoB transcription after 15 minutes of 4-OHT treatment (Figure 12A). Establishment of H3K27ac at the *SPHK1-B* promoter correlated with upregulation of *KHPS1* transcription, whereas increase of isoB RNA and binding of E2F1 to *SPHK1-B* occurred later, coinciding with activation of sense transcription (Figure 12B). These results suggest a temporal order of events leading to *SPHK1-B* transcription activation, that is, induction of antisense RNA and recruitment of chromatin modifiers preceding transcription factor binding and transcription of isoB RNA.

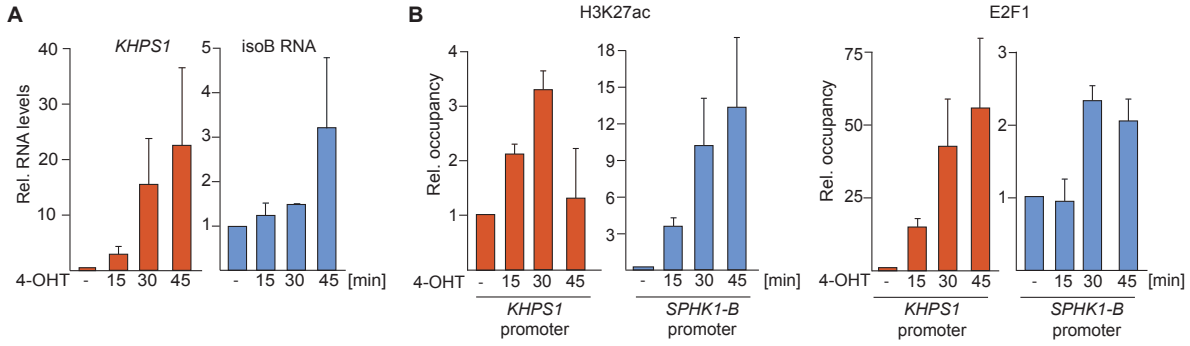

Figure 12. Induction of *KHPS1* and epigenetic changes at the *SPHK1-B* promoter precede E2F1 binding and upregulation of isoB RNA.
A) U2OS/ER-E2F1 cells were treated with 4-OHT for the indicated time points. Levels of *KHPS1* and isoB RNA were determined by RT-qPCR**.**
B) ChIPs showing occupancy of E2F1 and H3K27ac at the *KHPS1* and *SPHK1-B* promoters at the indicated time points of 4-OHT treatment (N=2).

2.4. *SPHK1-B* promoter is a poised enhancer activated by *KHPS1*

Inspection of available data sets revealed the presence of typical enhancer marks upstream of the TSS of *SPHK1-B*. High enrichment of H3K4me1 and H3K27ac suggested that *SPHK1-B* may function as a distal regulatory element (Figure 13A).

To verify whether the region upstream of the TSS of isoB exhibits enhancer-like features, the occupancy of enhancer-specific histone marks was monitored at the *SPHK1-B* promoter. In uninduced U2OS/ER-E2F1 cells, the *SPHK1-B* promoter displays high ratio (>2) of H3K4me1 H3K4me3, which distinguishes enhancers from promoters (Heintzman et al., 2007). The *SPHK1-C* promoter, on the other hand, exhibits low H3K4me1/H3K4me3 ratio (0.04), a characteristic of active promoters. These results support the view that *SPHK1-B* acts as an

enhancer. Interestingly, upon E2F1 induction, the occupancy of active histone mark H3K27ac was increased and repressive H3K27me3 mark was lost at the *SPHK1-B* promoter. Moreover, it coincided with the increased binding of p300 (Figure 13B). Taken together, establishment of H3K27ac by recruited p300, the high ratio of H3K4me1/H3K4me3 and the *KHPS1*-dependent activation of isoB transcription suggested that *SPHK1-B* is a poised enhancer that is activated by *KHPS1*. Therefore, the *SPHK1-B* promoter will thereafter be referred to as *SPHK1* enhancer (or e*SPHK1*) and the isoB transcript as eRNA-Sphk1.

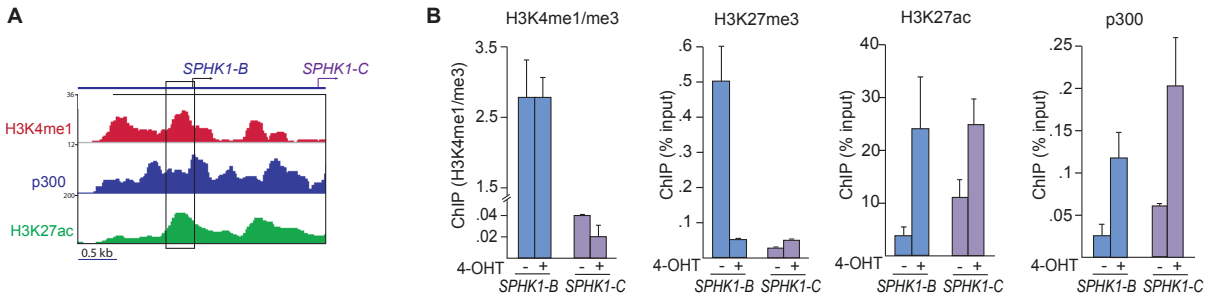

Figure 13. *SPHK1-B* promoter is a poised enhancer activated upon E2F1 induction.
A) ChIP-seq tracks of H3K4me1, H3K27ac and p300 available from osteoblasts (GEO accession numbers: H3K4me1 GSM733704, H3K27ac GSM733739, p300 GSM1003514) were visualized at the *SPHK1* locus. **B)** ChIPs showing occupancy of the indicated histone marks and p300 at the *SPHK1-B* and *SPHK1-C* promoters in uninduced U2OS/ER-E2F1 cells or after treatment with 4-OHT for 8 h (N=3)

To substantiate the requirement of *KHPS1* for activation of the *SPHK1-B* enhancer, the CRISPRa approach was utilized to upregulate transcription of endogenous *KHPS1* (Maeder et al., 2013). For this, a catalytically inactive dCas9 endonuclease fused to the VP64 activator (dCas9-VP64) was targeted to *KHPS1* promoter using sgRNA-mediated sequence recognition. Similar to upregulation of *KHPS1* by 4-OHT, induction of *KHPS1* transcription by dCas9-VP64 led to increased levels of eRNA-Sphk1 (Figure 14A). Elevated levels of sense transcription caused by dCas9-VP64-driven upregulation of *KHPS1* coincided with enhanced occupancy of E2F1, p300, RNA polymerase II (Pol II) and H3K27ac at *SPHK1-B* (Figure 14B), indicating that *KHPS1* is required for activation of the poised *SPHK1-B* enhancer.

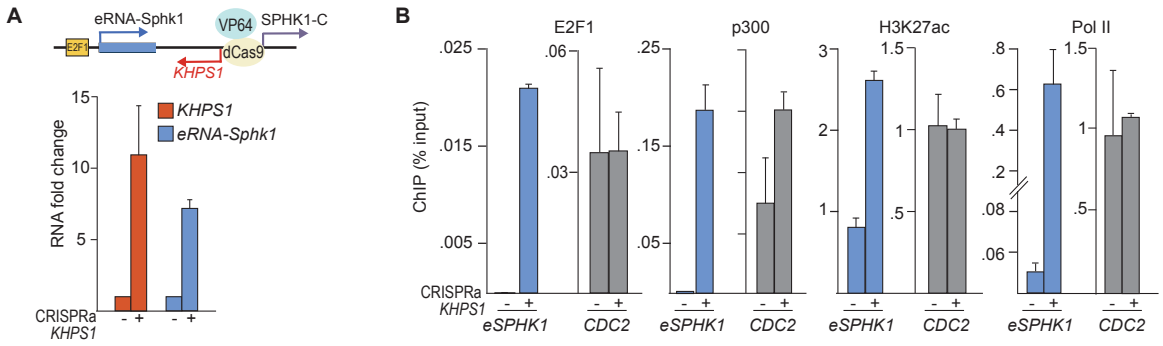

Figure 14. Upregulation of *KHPS1* leads to activation of *SPHK1* enhancer.
A) Levels of *KHPS1* and eRNA-Sphk1 in U2OS/ER-E2F1 cells expressing dCas9-VP64 targeted to the *KHPS1* promoter by sgRNAs (+) or co-transfected with a control sgRNA (-) (N=3).
B) ChIPs showing occupancy of E2F1, p300, H3K27ac and Pol II at *eSPHK1* in untreated U2OS/ER-E2F1 cells and after CRISPRa-mediated upregulation of *KHPS1*. Binding to *CDC2* promoter was monitored as control (N=3).

To confirm that the increase in *eSPHK1* transcription was brought about by the CRISPRa-upregulation of *KHPS1* rather than by spreading effects caused by dCas9-VP64 binding, the occupancy of H3K27ac was monitored 1,448 bp downstream of the transcription start site of *KHPS1*. Importantly, no increase in H3K27ac was observed (Figure 15A), supporting that dCas9-VP64-mediated changes in chromatin structure did not spread into adjacent gene regions. Moreover, if the cells expressing dCas9-VP64 targeted by sgRNA to the *KHPS1* promoter were treated with flavopiridol, increased occupancy of H3K27ac, E2F1, p300 and Pol II was compromised (Figure 15B), underscoring that of *KHPS1* triggers establishment of a transcription-permissive chromatin structure at the *SPHK1* enhancer.

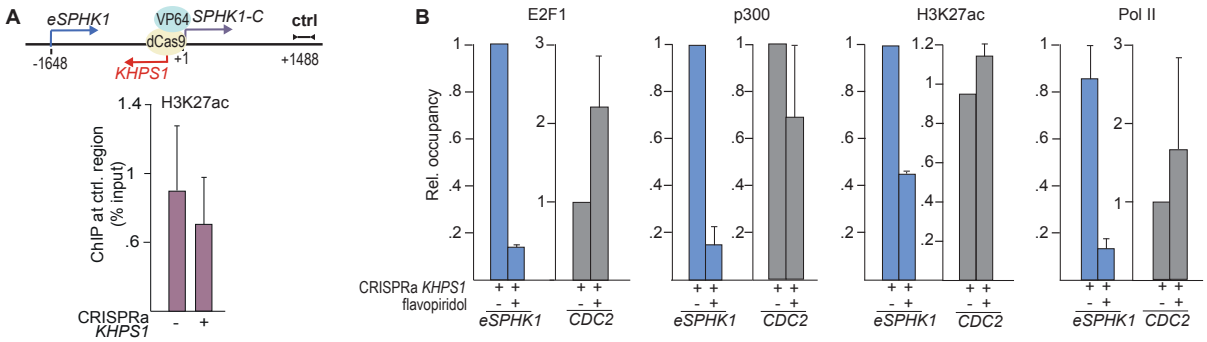

Figure 15. Transcription of *KHPS1* is required for activation of *SPHK1* enhancer.
A) ChIP showing occupancy of H3K27ac in the control region +1448 relative to the *SPHK1-C* transcription start site (+1) in U2OS/ER-E2F1 cells co-transfected with a dCas9-VP64 expression vector and either *KHPS1*-specific (+) or non-specific (-) sgRNAs (N=3).
B) ChIP showing occupancy of E2F1, p300, H3K27ac and Pol II at *eSPHK1* in U2OS/ER-E2F1 cells upon CRISPRa-mediated upregulation of *KHPS1* in the absence or presence of flavopiridol (3 h). Binding at *CDC2* promoter is monitored as control (N=3).

To corroborate the importance of *KHPS1*-mediated recruitment of transcriptional co-activators for induction of eRNA-Sphk1, a reporter plasmid was generated which drives *KHPS1* transcription under the control of a tetracycline-inducible promoter. Enhancer activation was

monitored by measuring expression of luciferase which was fused in frame with eRNA-Sphk1. Transfection of the tetracycline transactivator (tTA) led to a 30-40-fold increase in the luciferase signal. Enhanced luciferase expression was compromised if the cells were transfected with antisense oligo (ASO) targeting either *KHPS1* or eRNA-Sphk1, underscoring the requirement of *KHPS1* and eRNA-Sphk1 for luciferase expression. Increased luciferase expression was also attenuated if the cells were treated with curcumin, an inhibitor of p300/CBP activity, and by siRNA-mediated depletion of E2F1 or p300. Knockdown of the histone acetyltransferase PCAF or treatment with nicotinamide (NAM), a specific inhibitor of NAD^+-dependent deacetylases, did not affect *eSPHK1*-driven luciferase expression (Figure 16). These results reinforce that transcription of eRNA-Sphk1 requires *KHPS1*-dependent targeting of p300/CBP and E2F1 to the *SPHK1* enhancer.

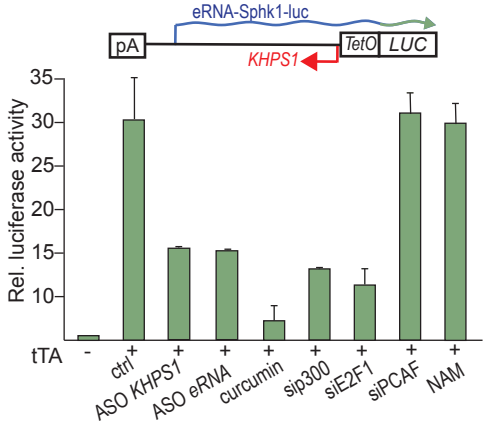

Figure 16. Transcription of eRNA-Sphk1 requires *KHPS1*-mediated recruitment of p300 and E2F1. Reporter assay monitoring *KHPS1*-dependent activation of *eSPHK1*-driven luciferase expression. U2OS/ER-E2F1 cells were co-transfected with the reporter plasmid pTet-*KHPS1*/+1448/-592-(isoB-luc) and a plasmid encoding tTA. Where indicated, cells were transfected with ASOs against *KHPS1* or eRNA-Sphk1, with siRNAs against E2F1, p300 or PCAF, or treated with curcumin (30 μM) or NAM (10 mM). Activation of eRNA-Sphk1 transcription was measured by expression of luciferase and presented in reference to tTA-untransfected cells (N=3).

2.5. The transcription factor E2F1 interacts with *KHPS1*

Loss-of-function experiments performed on *KHPS1* demonstrated that the association of E2F1 with the *SPHK1* enhancer requires transcription of *KHPS1*. Recent reports documenting the requirement of RNA for the function of transcription factors, such as CTCF (Kung et al., 2015) or YY1 (Sigova et al., 2015), suggested that binding to RNA may be a common mechanism involved in targeting of TFs to DNA. To get insight into the role of RNA in E2F1 function, the ability of E2F1 to directly interact with cellular RNAs was tested. For this, RNA-protein interactions were stabilized by UV light and RNA bound to immunopurified E2F1 was labeled with γ-ATP. Co-precipitated RNAs were visualized by autoradiography revealing a signal of

RNA-E2F1 complexes (Figure 17), indicating the presence of RNA associated with E2F1. Importantly, treatment of lysates with RNase I efficiently reduced the intensity of detected signal in a concentration-dependent manner, confirming that nucleic acids associated with crosslinked E2F1 were RNA and not DNA.

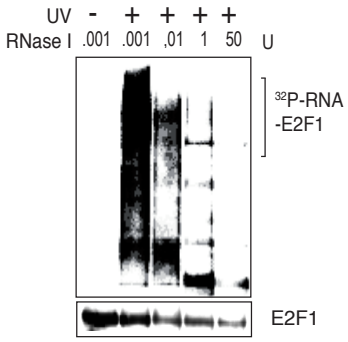

Figure 17. E2F1 is associated with RNAs in vivo.
U2OS/ER-E2F1 cells were induced with 4-OHT, UV-crosslinked and endogenous E2F1 was immunoprecipitated. Co-precipitated RNA was labeled with γ-ATP, and RNA-E2F1 complexes were subjected to SDS-PAGE and transferred to a membrane. Top: Autoradiography presenting RNaseI-sensitive signal of E2F1-associated nucleic acids. Bottom: Western Blot showing levels of immunoprecipitated E2F1.

To examine the relevance of RNA in E2F1 association with DNA, the level of E2F1 bound to chromatin was analyzed in untreated chromatin and chromatin treated with RNase A. The amount of E2F1 bound to chromatin was significantly lower after RNase A treatment, indicating that RNA contributes to the recruitment and/or stable binding of E2F1 to chromatin (Figure 18A). Furthermore, if cellular transcription was inhibited by flavopiridol prior to 4-OHT induction, the level of chromatin-associated E2F1 was markedly reduced, indicating that RNA enhances E2F1 occupancy at DNA (Figure 18B). This result suggests that E2F1 is an RNA binding protein that is capable of associating with specific RNAs. Binding to RNA in turn may contribute to the stable occupancy of transcription factors at regulatory elements.

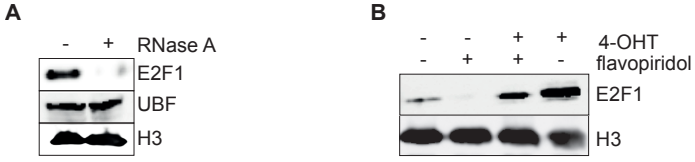

Figure 18. E2F1 association with chromatin requires RNA.
A) U2OS/ER-E2F1 cells were treated with 4-OHT (3 h) and chromatin fraction was prepared in the presence or absence of RNase A. E2F1, UBF and histone H3 levels were monitored by Western blot.
B) Western Blot showing the levels of E2F1 and histone H3 in chromatin prepared from U2OS/ER-E2F1 cells treated with flavopiridol (1 h) prior to incubation with 4-OHT.

The finding that *KHPS1* transcription facilitated binding of E2F1 to the *SPHK1* enhancer suggested that *KHPS1* might interact with E2F1 and recruit it to its binding site at the *SPHK1* enhancer. To test this, E2F1 was immunopurified from tamoxifen-treated U2OS/ER-E2F1 cells and the association of *KHPS1* was monitored by RT-qPCR. Interestingly, *KHPS1* was highly enriched in the fraction of RNAs associated with E2F1. No interaction of *HOTAIR* was observed, indicating that E2F1 is preferentially associated with *KHPS1* (Figure 19A). Furthermore, the region of *KHPS1* that dictates the interaction with E2F1 was identified by CLIP assay. For this, UV-crosslinked E2F1-RNA complexes were immunopurified from 4-OHT-treated U2OS/ER-E2F1 cells and *KHPS1* association was measured by RT-qPCR using primers that cover different regions of *KHPS1*. This experimental approach revealed that E2F1 is bound to *KHPS1* sequences -373/-241 (amplicon B) but not to further upstream or downstream regions (amplicons A and C-G) (Figure 19B). This result suggests that *KHPS1* recruits E2F1 to the *SPHK1* enhancer to activate transcription of eRNA-Sphk1.

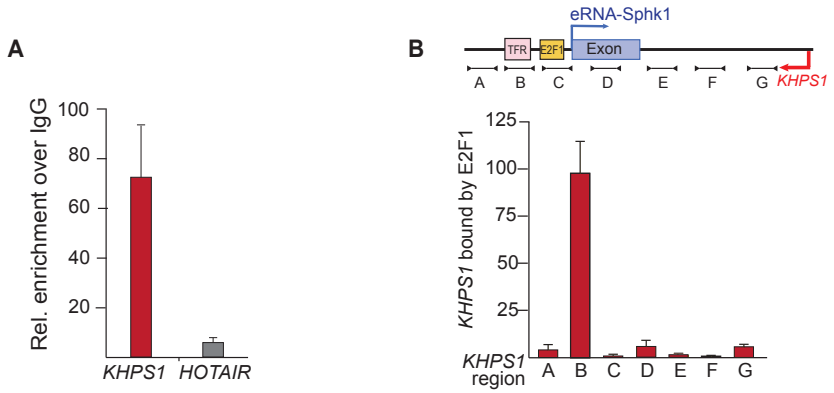

Figure 19. E2F1 interacts with *KHPS1*.
A) RIP experiment comparing the association of *KHPS1* and a control lncRNA (HOTAIR) with E2F1 immunoprecipitated from 4-OHT-treated U2OS/ER-E2F1 cells (N=3).
B) CLIP showing interaction of E2F1 with different regions of *KHPS1*. U2OS/ER-E2F1 cells were UV crosslinked, lysed and E2F1 was immunoprecipitated. Levels of *KHPS1* associated with E2F1 were analyzed by RT-qPCR. qPCR primer positions: A:-592/-425, B: -373/-304, C: -137/-89, D: +108/+165, E: +630/+790, F:+930/+1124, G: +1132/+1242 (N=4).

2.6. *KHPS1* is tethered to the *SPHK1* enhancer via RNA-DNA triplex

Having established that *KHPS1* mediates the recruitment of E2F1 and p300 to the *SPHK1* enhancer and is required for eRNA-Sphk1 transcription activation, the molecular mechanism underlying guidance of p300 and E2F1 to the enhancer had to be investigated. The presence of a polypurine stretch at the *SPHK1* enhancer which has potential to be engaged in triplexes, suggests that *KHPS1* may exert its regulatory function by directly binding to the region upstream of transcription start site of *eSPHK1*. To validate this hypothesis experimentally, *in vitro* and *in vivo* assays were performed to monitor triplex formation at the *SPHK1* locus.

The ability of different regions of *KHPS1* to interact with *SPHK1* enhancer was tested by in vitro triplex capture down assay. For this, a DNA fragment comprising a putative TFR (−592/+7) was incubated with biotinylated versions of *KHPS1*. After binding to streptavidin beads, RNA-associated DNA was monitored by PCR (Figure 20,scheme). RNAs that harbor TFR (−373/+7, version A) or (−406/−241, version C) were associated with the *SPHK1* enhancer, whereas *KHPS1* harboring the region around transcription start site of *eSPHK1* (−137/+7, version B) bound least efficiently to DNA (Figure 20A, left panel). If *KHPS1* contained sequences corresponding to the intronic region of eRNA-Sphk1, which lacks putative TFRs (+1050/+1239, version D), no interaction with DNA was observed. Moreover, a PCR fragment comprising the first intron of eRNA-Sphk1 was not able to associate with different *KHPS1* versions (Figure 20A, right panel), demonstrating the ability of *KHPS1* to interact with the *SPHK1* enhancer at the TFR region. To substantiate that *KHPS1* association with the *eSPHK1* is mediated by Hoogsteen basepairing, in vitro triplex pull down assay was performed using PCR fragment generated in the presence of 7-deaza-purine nucleotides, which modification at N7 position prevents hydrogen bonds formation with the third strand of nucleic acid. Importantly, *KHPS1* was not able to efficiently co-precipitated *eSPHK1* (Figure 20B), demonstrating the ability of *KHPS1* to bind to double-stranded DNA at the TFR sequence by RNA-DNA triplex formation.

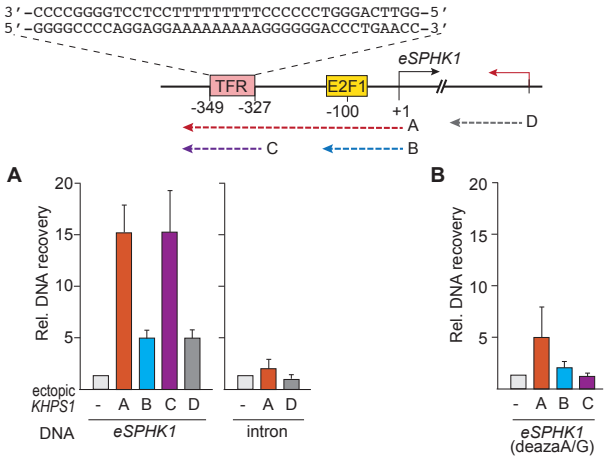

Figure 20. *KHPS1* binds to the triplex forming region (TFR) at the *SPHK1* enhancer via Hoogsteen base-pairing.

The scheme depicts the position and sequence of TFRs and synthetic RNAs used to pull-down DNA. E2F1 binding site is indicated. The arrowed lines depict synthetic *KHPS1* versions that were incubated with the *SPHK1* enhancer (A: 373/+7; B: 137/+7; C: 406/–241).

A) Biotinylated *KHPS1* versions (A, B, C, or D) were incubated with PCR fragments comprising either e*SPHK1* sequences (592/+7) intronic region (+596/+1242). After binding to streptavidin beads, RNA-associated DNA was monitored by qPCR using enhancer-specific (-592/–425, *eSPHK1*) or intron-specific (+1132/+1242) primers. Data represent the fold enrichment of DNA bound to the respective RNAs over the sample without RNA (N=3).

B) Biotinylated *KHPS1* versions (A, B, C) were incubated with e*SPHK1* fragment (592/+7) generated by PCR using 7-deaza-dATP/dGTP. After binding to streptavidin beads, recovered DNA was analyzed by qPCR using enhancer-specific (-592/–425, *eSPHK1*) primers. Data represent the fold enrichment of DNA bound to the respective RNAs over the sample without RNA (N=3).

To examine whether *KHPS1* is able to recognize and bind to the *SPHK1* enhancer at chromatin, biotinylated versions of *KHPS1* were incubated with isolated nuclei. After chromatin isolation, *KHPS1*-associated DNA was bound to streptavidin beads and analyzed by qPCR. Importantly, *eSPHK1* was found to be preferentially associated with ectopic *KHPS1* comprising TFRs but not with RNA corresponding to the intronic region of eRNA-Sphk1, which lacks TFRs. Consistently, no interaction was detected between *KHPS1* intronic regions and DNA corresponding to the intron of eRNA-Sphk1 (Figure 21A). These results indicate that *KHPS1* is able to specifically associate with the TFR sequence both on naked DNA and DNA assembled in nucleosomes.

To demonstrate that *KHPS1* forms triplexes at the *SPHK1* enhancer in vivo, biotinylated versions of *KHPS1* were transfected into cells, chromatin was isolated and *KHPS1*-associated DNA was co-precipitated using streptavidin beads. Analysis of recovered DNA revealed that *KHPS1* version harbouring the TFR was bound to the *SPHK1* enhancer (Figure 21B). In contrast, no association of intronic *SPHK1* region was observed with neither *KHPS1* containing TFR nor region corresponding to the intron of eRNA-Sphk1.

Collectively, the results from the in vitro and in vivo approaches indicate that *KHPS1* is anchored to the *SPHK1* enhancer via triplex formation at the TFR.

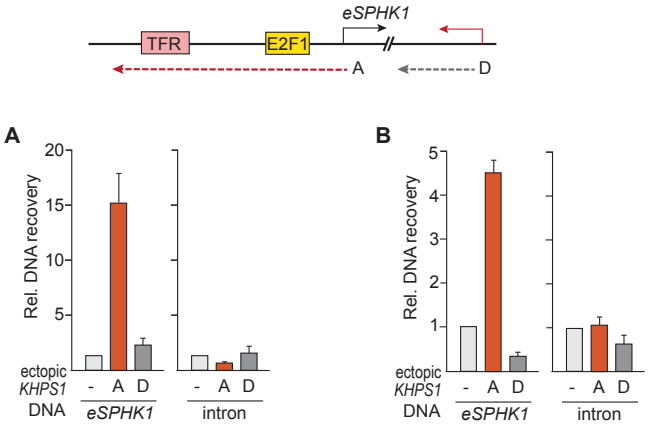

Figure 21. *KHPS1* associates with *SPHK1* enhancer in vivo.
A) Biotinylated *KHPS1* versions (A or D) were incubated with nuclei. After chromatin was isolated, RNA was bound to streptavidin beads, and associated DNA was quantified by qPCR using *eSPHK1* (-406/-304) or intron-specific (+1132/+1242) primers (N=3).
B) HeLa cells were transfected with biotinylated *KHPS1* versions (A or D) and DNA was recovered analyzed as described in (A) (N= 3).

To corroborate the physical association of *KHPS1* with the *SPHK1* enhancer occurring via RNA-DNA Hoogsteen basepairing, the homopurine stretch of the TFR was perturbed by either deletion of the TFR (ΔTFR) or by insertion of pyrimidine interruptions (mutant TFR), which do not support Hoogsteen basepairing. *KHPS1* versions comprising either wildtype TFR,

mutated or deleted TFR were incubated with the corresponding radiolabeled dsDNA followed by non-denaturing electrophoresis. There was no retardation in electrophoretic mobility of dsDNA when the homopurine stretches of RNA and DNA were perturbed by pyrimidines or deleted (Figure 22A). Similar result was observed in an in vitro triplex capture assay. For this, biotinylated *KHPS1* versions comprising wildtype, mutated or deleted TFRs were incubated with corresponding DNA fragments and RNA-associated DNA was captured on streptavidin beads. This experimental approach demonstrated that *KHPS1* was associated with DNA only if the TFR was intact, whereas no interaction was observed if the TFR was mutated or deleted (Figure 22B). Together these results demonstrate that sequence complementarity between RNA and DNA is not sufficient for binding of RNA to double-stranded DNA but requires specific purine-rich sequences that support Hoogsteen basepairing.

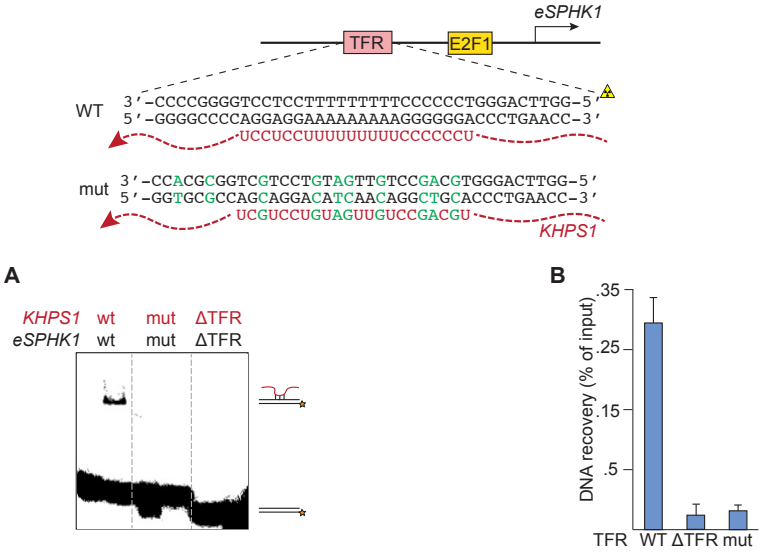

Figure 22. Purine-rich stretch at TFR is required for triplex formation.
A) EMSA showing the mobility of ^{32}P-labeled DNA oligonucleotide (-357/-319) comprising either wildtype (wt), mutated (mut) or deleted (ΔTFR) TFR after incubation with the corresponding synthetic *KHPS1* versions (-373/-241) harbouring either wildtype (WT), mutated (mut) or deleted (ΔTFR) TFR.
B) Biotinylated *KHPS1* (-373/-241) versions comprising the intact (WT), deleted (ΔTFR) or mutated (mut) TFR were incubated with a corresponding DNA fragment (-406/-65) and captured DNA was measured by qPCR using primers -137/-89 (N= 3).

2.7. Enhancer activation requires binding of *KHPS1* to *eSPHK1*

To investigate whether interaction of *KHPS1* with eSPHK1 is required for activation of eRNA-Sphk1 transcription, a reporter plasmid (pTet-*KHPS1*/+1448/-592) containing the 5'-terminal part of *KHPS1*, including the first exon of eRNA-Sphk1 and *eSPHK1* sequences (-592/+1) was generated. To activate *KHPS1* transcription from the reporter independent of E2F1, the endogenous promoter of *KHPS1* was replaced by a tetracycline-responsive element (Figure 23A). The plasmid was transfected into NIH3T3 cells to monitor exclusively reporter-derived RNAs rather than endogenous *KHPS1* and eRNA-Sphk1. Doxycycline-induced transcription

of *KHPS1* led to transcription activation of eRNA-Sphk1. Conversely, knockdown of *KHPS1* resulted in compromised eRNA-Sphk1 transcription activation (Figure 23B), indicating that the reporter assay mimics the in vivo situation, that is, transcription of sense eRNA-Sphk1 depends on transcription of the antisense RNA *KHPS1*.

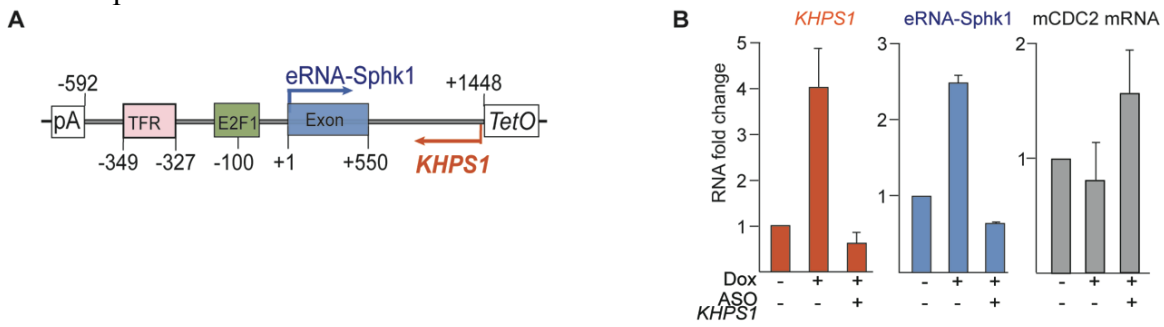

Figure 23. Reporter plasmid to monitor TFR- and *KHPS1*-dependent eRNA-Sphk1 transcription activation.
A) Scheme presents the structure of pTet-*KHPS1*/+1448/-592 reporter plasmid. Triplex forming region (TFR) and E2F1 binding site are indicated. Tet promoter and polyadenylation site are marked as TetO and pA, respectively.
B) NIH3T3 Tet-ON cells were transfected with the reporter plasmid pTet-*KHPS1*/+1448/-592 and *KHPS1*-specific or control ASO upon induction with doxycycline (3 µg/ml, 12 h). Levels of *KHPS1*, eRNA-Sphk1 and murine CDC2 mRNA measured by RT-qPCR. (N=3).

To test the effect of the triplex formation on *KHPS1*-dependent transcription activation, the triplex forming region at the reporter plasmid was either deleted (ΔTFR) or mutated (mutTFR) and the levels of eRNA-Sphk1 transcribed from the reporters was monitored (Figure 24A). Despite similar fold induction of *KHPS1* derived from the reporter plasmids, transcription was not induced if the TFR was deleted or mutated, indicating the importance of *KHPS1*-dependent triplex formation for activation of eRNA-Sphk1 (Figure 24B). These results demonstrate that anchoring of *KHPS1* to the TFR is required for *KHPS1*-mediated transcription of eRNA-Sphk1.

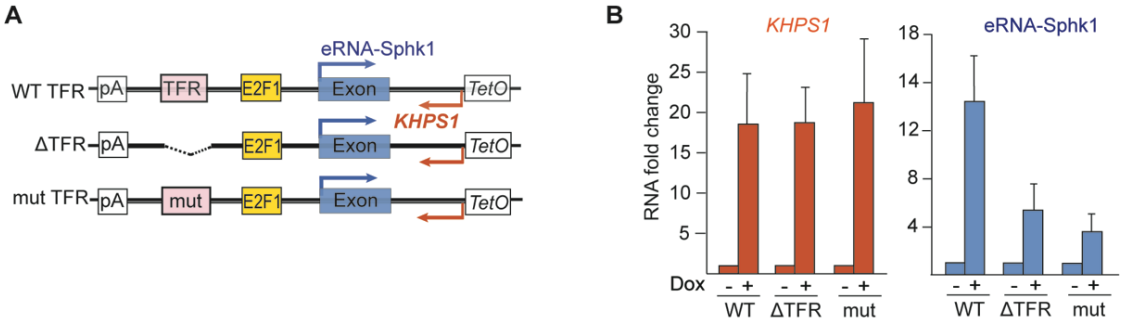

Figure 24. The triplex forming region (TFR) is necessary for *KHPS1*-dependent eRNA-Sphk1 transcription.
A) The scheme illustrates the structure of pTet-*KHPS1*/+1448/-592 reporter plasmid comprising the intact TFR of *eSPHK1* (WT TFR), a deleted (ΔTFR) or mutated TFR (mut). Tet promoter and polyadenylation site are marked as TetO and pA, respectively.
B) NIH3T3 Tet-ON cells were transfected with the reporter plasmids pTet-*KHPS1*/+1448/-592 from (A) and induced with doxycycline (5 µg/ml, 18 h) or left untreated. Levels of *KHPS1* and eRNA-Sphk1 were measured by RT-qPCR (N=4).

To test whether recruitment of regulatory proteins to the *SPHK1* enhancer requires the physical association of *KHPS1* with the TFR, the occupancy of E2F1 and p300 at the reporter plasmid was monitored by ChIP. Similar to the effect of *KHPS1* at the endogenous *SPHK1* locus, tetracycline-induced transcription of *KHPS1* led to increased binding of E2F1 and p300 to the reporter plasmid. However, if the TFR was mutated or deleted, elevated levels of *KHPS1* did not result in increased association of E2F1 and p300 with the *SPHK1* enhancer (Figure 25), emphasizing that tethering of *KHPS1* to *eSPHK1* via Hoogsteen base pairing is necessary for recruitment of E2F1 and p300, which activates eRNA-Sphk1 transcription.

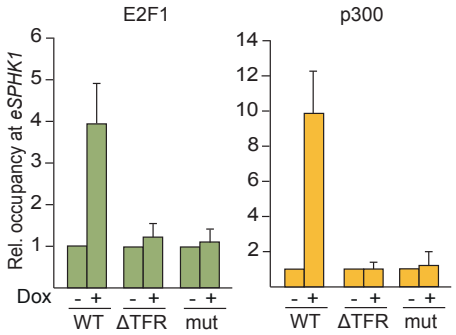

Figure 25. The TFR is required for recruitment of E2F1 and p300 to *SPHK1* enhancer.
NIH3T3 Tet-ON cells were transfected with the reporter plasmids pTet-*KHPS1*/+1448/-592 comprising the intact TFR of *eSPHK1* (WT), a deleted (ΔTFR) or mutated TFR (mut) followed by induction with doxycycline (5 µg/ml, 18 h) where it is indicated. Binding of E2F1 and p300 at pTet-*KHPS1*/+1448/-592 was assayed by ChIP and monitored by qPCR using primers -137/-89 (N=3).

To examine whether compromised binding of p300 and E2F1 to reporter plasmids lacking the TFR or comprising a mutated TFR was due to perturbation of triplex formation, the interaction of p300 and E2F1 with *KHPS1* was monitored by RIP experiments. For this, endogenous p300 or E2F1 were immunoprecipitated from cells, which were transfected with the respective reporter plasmids, and p300- or E2F1-associated with reporter-derived *KHPS1* was monitored by RT-qPCR. Consistent with cellular *KHPS1* being engaged into interaction with E2F1, *KHPS1* transcribed from reporter plasmid was associated with E2F1. Importantly, deletion or mutation of the TFR in *KHPS1* significantly reduced the association with E2F1. In contrast, deletion of the region of *KHPS1* corresponding to the first exon of eRNA-Sphk1 (+25/+487) did not affect E2F1 binding (Figure 26). These results indicate that the TFR is required for the interaction with E2F1 and recruitment to the *SPHK1* enhancer.

While p300 was associated with reporter-derived *KHPS1* regardless of whether TFR was intact, depleted or mutated, deletion of the *KHPS1* region corresponding to the first exon of eRNA-Sphk1 (Δ+25/+487) led to compromised association of *KHPS1* with p300 (Figure 26). These results demonstrate that the TFR of *KHPS1* is necessary for the recruitment of p300 to e*SPHK1* but not for the interaction of *KHPS1* with p300.

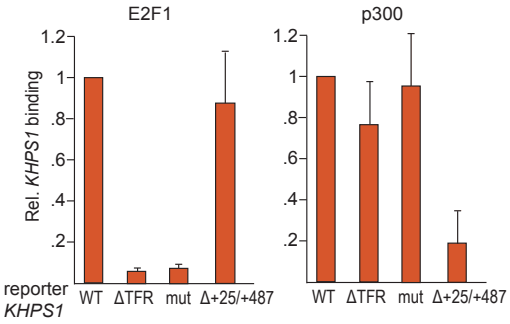

Figure 26. TFR is required for E2F1, but not for 300 association with *KHPS1*.
NIH3T3 Tet-ON cells were transfected with pTet-*KHPS1*/+1448/-592 comprising the intact TFR (WT), deleted (ΔTFR), mutated TFR (mut) or lacking nucleotides +25/+487 relative to the *eSPHK1* TSS (Δ+25/+487) followed by induction with doxycycline (5 µg/ml, 18 h). E2F1 and p300 were immunoprecipitated and levels of reporter-derived *KHPS1* associated with the proteins were measured by RT-qPCR. RNA binding was calculated as % of input of the sample and presented in reference to the sample with the wild type TFR.

2.8. Triplex motifs are responsible for site-specific targeting of lncRNA-associated proteins

Numerous studies have proposed that lncRNAs serve as readers of genomic address codes. To exploit the specificity of triplex-mediated lncRNA tethering, the capability of foreign sequences to functionally replace the TFR of e*SPHK1* was investigated. For this, the TFR of e*SPHK1* in the reporter plasmid pTet-*KHPS1*/+1448/-592 was replaced by TFR sequences of *Fendrr* (Grote et al., 2013), *MEG3* or the *MEG3* target gene *TGFBR1* (Mondal et al., 2015) or by sequences that do not support triplex formation, such as U2 snRNA- or luciferase-derived sequences (Figure 27A). Upon transfection of the chimeric constructs, transcription of eRNA-Sphk1 was monitored. Despite all chimeric constructs produced similar levels of *KHPS1* transcripts, sense transcription was activated at constructs that harbor the TFR of *Fendrr*, *MEG3* and *TGFBR1*. Constructs in which the TFR was replaced by U2 snRNA- or luciferase-derived

sequences did not activate of sense RNA synthesis (Figure 27B). These experiments suggest that triplex motifs are interchangeable can functionally substitute the TFR of *SPHK1*.

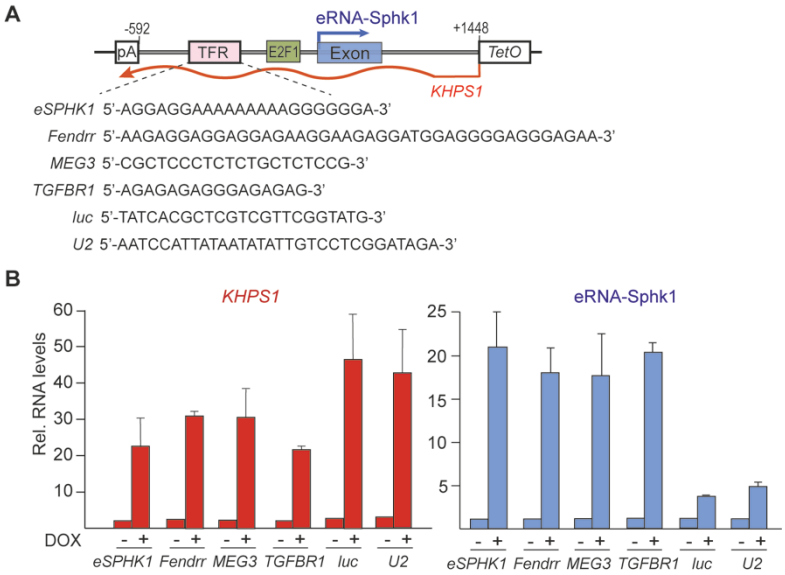

Figure 27. Poly-purine triplex forming region is required for activation of eRNA-Sphk1 transcription.
A) Sequences of foreign TFRs used to replace the TFR of *eSPHK1* in pTet-*KHPS1*/+1448/-592. Triplex forming region (TFR) and E2F1 binding site are indicated. Tet promoter and polyadenylation site are marked as TetO and pA, respectively.
B) NIH3T3 Tet-ON cells were transfected with the chimeric pTet-*KHPS1*/+1448/-592 plasmids comprising the indicated foreign TFRs. Cells were induced with doxycycline (5 µg/ml, 18 h) or left untreated. Levels of reporter-derived *KHPS1* and eRNA-Sphk1 were measured by RT-qPCR(N=3).

Consistent with transcription activation of eRNA-Sphk1, binding of E2F1 and p300 was enhanced at plasmids which contain a potential TFR but not at constructs in which the TFR was replaced by random sequences (Figure 28), emphasizing the relevance of triplex-mediated anchoring of RNA for targeting associated proteins to specific genomic loci. Collectively, these results underscore the relevance of triplex-mediated anchoring of either purine- or pyrimidine-rich RNAs for targeting associated proteins to the corresponding genomic loci.

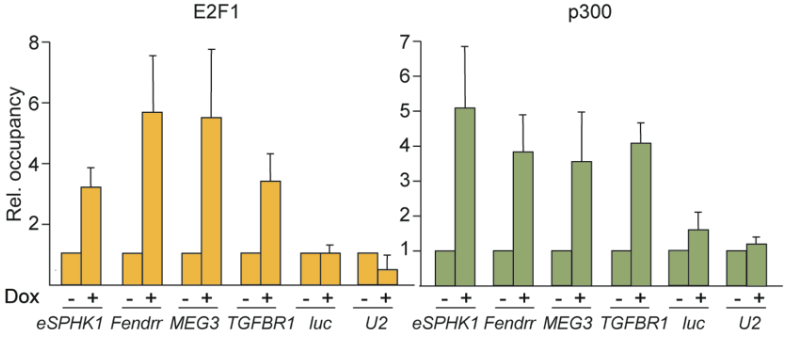

Figure 28. Potential triplex forming regions are required for E2F1 and p300 recruitment to *eSPHK1*.
NIH3T3 Tet-ON cells were transfected with the chimeric pTet-*KHPS1*/+1448/-592 plasmids comprising the indicated foreign TFRs. Cells were induced with doxycycline (5 µg/ml, 18 h) or left untreated. Binding of E2F1 and p300 to the reporter plasmid pTet-*KHPS1*(+1448/-592) was assayed by ChIP and monitored by qPCR using primers -137/-89 (N=3).

To confirm that the interaction of *KHPS1* with *eSPHK1* containing foreign TFRs is mediated via Hoogsteen base pairing, in vitro triplex capture assays was performed. Biotinylated RNAs comprising the TFR of *KHPS1*, *Fendrr*, *MEG3* and *TGFBR1*, as well as U2 snRNA- or luciferase-derived sequences were incubated with the corresponding DNA fragments. To distinguish Hoogsteen-mediated interaction from unspecific DNA binding, we used DNA fragments that were generated either in the presence of unmodified nucleotides or in the presence of 7-deaza-purine nucleotides which do not allow Hoogsteen base paring. After incubation of chimeric RNAs with corresponding DNA fragments and capture on streptavidin beads, RNA-associated DNA was monitored by PCR. Notably, the interaction between RNAs and DNA fragments containing the TFR of *eSPHK1*, *Fendrr, MEG3* or *TGFBR1* was abolished when the PCR fragments were generated in the presence of 7-deaza-purine nucleotides (Figure 29), validating that the association of chimeric *KHPS1* and *eSPHK1* is brought about by Hoogsteen base pairing.

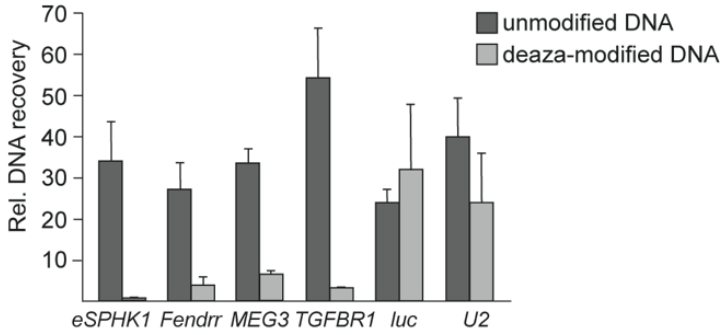

Figure 29. Chimeric *KHPS1* and *eSPHK1* form RNA-DNA triplexes.
Biotinylated chimeric *KHPS1* versions (-373/-241) were incubated with an *eSPHK1*-containing PCR fragment (-406/-65) harbouring foreign TFRs generated in the presence of either unmodified dATP/dGTP or 7-deaza-dATP/dGTP. After binding to streptavidin beads, RNA-associated DNA was measured by qPCR using primers -137/-89 (N=3).

In silico analyses have identified hundreds of lncRNAs with triplex-forming domains, which may interact with respective purine-rich DNA sequences forming RNA-DNA structures (Buske et al., 2012; Goñi et al., 2004; Soibam, 2017). Such specific structures may dictate how lncRNA-associated transcription regulators and chromatin modifying enzymes are guided to appropriate genomic sequences. Therefore, the ability of TFR motifs to target lncRNA-associated proteins to specific genomic loci was tested. For this, the TFR of *KHPS1* was either replaced by the TFR of *MEG3*, generating chimeric *KHPS1-MEG3* RNA, or the TFR was left intact. After transfection of synthetic RNAs *KHPS1* or *KHPS1-MEG3,* occupancy of H3K27ac and E2F1 at cellular *SPHK1,* and *MEG3's* target genes, *MEG3* and

TGFBR1, were monitored by ChIP assays. Ectopic *KHPS1* version comprising the TFR of *KHPS1* led to enhanced occupancy of E2F1 and H3K27ac at the *SPHK1* locus, whereas *KHPS1-MEG3* RNA did not affect occupancy of E2F1 and H3K27ac at *eSPHK1*. In contrast, increased occupancy of E2F1 and higher levels of H3K27ac at endogenous *TGFBR1* and *MEG3* were observed if the cells were transfected with ectopic *KHPS1-MEG3* RNA, but not upon transfection with wildtype *KHPS1* (Figure 30). These RNA transfection experiments indicate that the respective TFR determine the target specificity of lncRNAs, substantiating the importance of triplex-mediated anchoring of purine- or pyrimidine-rich RNAs for guidance of regulatory proteins to TFR-containing gene loci.

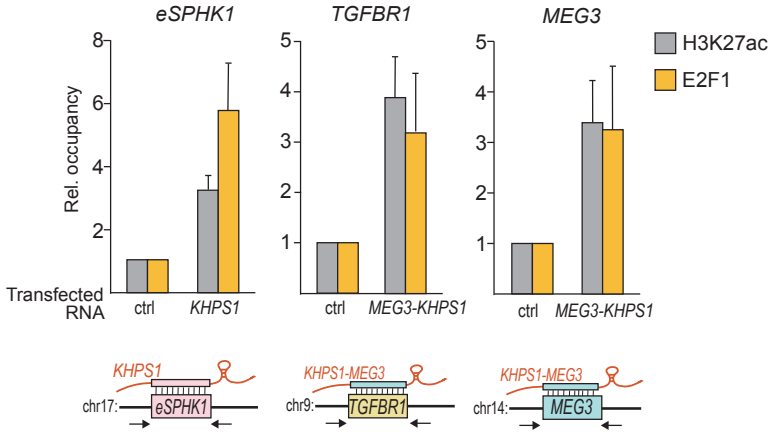

Figure 30. *KHPS1* containing *MEG3* triplex forming region is guided to the *MEG3* target gene *TGFBR1.*
HeLa cells were transfected with synthetic *KHPS1* versions (-406/+596) comprising either the TFR of *KHPS1* or *MEG3* (*KHPS1-MEG3*). Occupancy of H3K27ac and E2F1 at *eSPHK1, TGFBR1* and *MEG3* was monitored by ChIP. The schemes illustrate the genes that are targeted by the chimeric RNAs. Arrows indicate primer positions used for qPCR (N=3).

2.9. Transcription of eRNA-Sphk1 facilitates expression of *SPHK1*

Given that *SPHK1* enhancer gives rise to RNAs that are usually unstable, the features of eRNA-Sphk1 were examined. The half-lives of eRNA-Sphk1 and SPHK1 mRNA were measured after blocking transcription elongation with flavopiridol. Consistent with reports indicating that enhancer-derived transcripts are defined by rapid decay, eRNA-Sphk1 displayed short half-live, 50% of the eRNA being degraded after 15 min. In contrast, SPHK1 mRNA exhibits a half-life of about 3.5 hours (Figure 31A). In accord with low stability, the abundance of eRNA-Sphk1 coincided with low abundance as compared to SPHK1 mRNA (Figure 31B). Consistent with the nuclear localization of eRNAs, cellular fractionation experiments revealed that eRNA-Sphk1 resides mainly in the nucleus, in contrast to SPHK1 mRNA which is enriched in the cytoplasm (Figure 31C). These features of eRNA-Sphk1 support the view that the *SPHK1* enhancer produces enhancer RNA, which may serve a regulatory function.

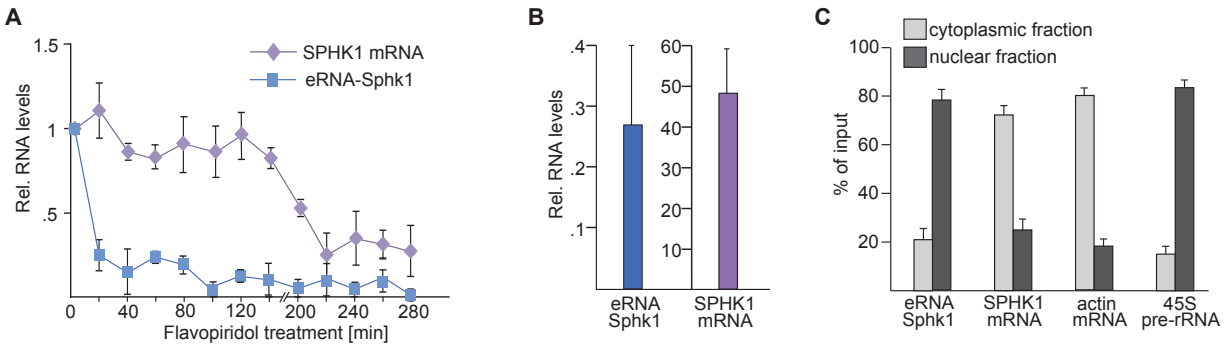

Figure 31. Transcripts from the *SPHK1* enhancer are unstable, low abundant and nuclear RNAs.
A) Different half-lives of eRNA-Sphk1 and SPHK1 mRNA transcripts in U2OS/ER-E2F1 cells treated with 4-OHT (5 h) followed by inhibition of transcription by treatment with flavopiridol (3 h). The levels of eRNA-Sphk1 and SPHK1 mRNA were measured by RT-qPCR at the indicated time point after addition of flavopiridol (N=3**).**
B) RT-qPCR showing steady-state levels of eRNA Sphk1 and SPHK1 mRNA in U2OS/ER-E2F1 cells (N=2).
C) Levels of eRNA-Sphk1 and SPHK1 mRNA in cytoplasmic and nuclear fractions prepared from 4-OHT-treated U2OS/ER-E2F1 cells and measured by RT-qPCR. Levels of actin mRNA and 45S pre-rRNA served as controls for cytoplasmic and nuclear localization, respectively (N=3).

Several studies have demonstrated that eRNAs are functionally important for target gene expression. Given that eRNA-Sphk1 is transcribed through the *SPHK1-C* promoter, the ability of eRNA-Sphk1 to affect *SPHK1-C* transcription was investigated. For this, reporter plasmids monitoring *SPHK1-C*-driven luciferase expression were generated. To test the relevance of the enhancer region for *SPHK1-C* transcription, luciferase expression was measured at the reporter plasmids which comprise or lack sequences upstream of the TSS of *eSPHK1*. Upon E2F1 induction, luciferase expression was higher in the cells transfected with the plasmid harboring *eSPHK1* sequences from -592 to +1795 (pREP4-SPHK1-592/+1795) than in cells transfected with the reporter lacking *eSPHK1* (pREP4-SPHK1+46/+1795) (Figure 32A). To investigate the role of enhancer-derived transcript RNA-Sphk1, a polyadenylation cassette was inserted downstream of the TSS of *eSPHK1*. This led to reduced luciferase expression, indicating the relevance of eRNA-Sphk1 for *SPHK1-C* transcription (Figure 32A). Importantly, the levels of *KHPS1* remained unchanged upon insertion of polyadenylation signal, underscoring that compromised *SPHK1-C* transcription was brought about by polyA-dependent termination of eRNA-Sphk1 (Figure 32B). Moreover, depletion of either *KHPS1* or eRNA-Sphk1 using ASO led to decreased luciferase expression, corroborating the importance of *KHPS1*-dependent induction of eRNA-Sphk1 for transcription of SPHK1 mRNA.

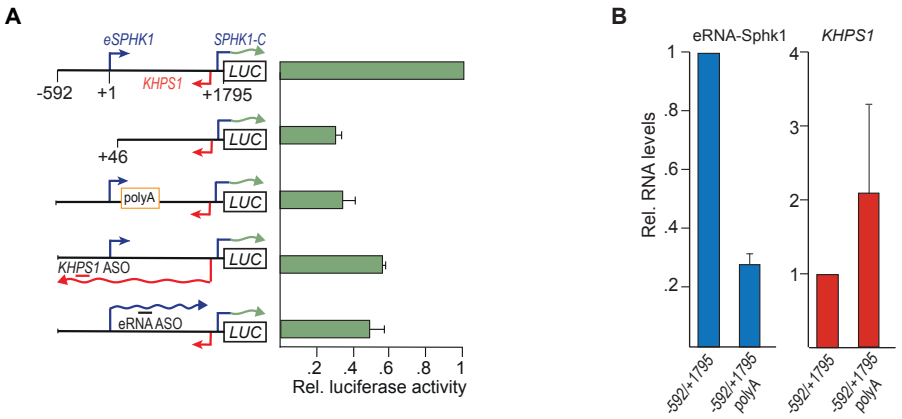

Figure 32. *SPHK1* enhancer sequences and eRNA-Sphk1 are required for *SPHK1-C* transcription.
A) Reporter assay measuring *SPHK1-C*-driven luciferase expression in U2OS/ER-E2F1 cells transfected with the depicted on the scheme pREP4-luciferase plasmids and an E2F1 expression vector. Where indicated, cells were co-transfected with ASOs targeting *KHPS1* or eRNA-Sphk1. The firefly luciferase expression was normalized to Renilla luciferase expression. Normalized luciferase signal of pREP4-SPHK1(-592/+1795)-luc was set to 1 (N=3).
B) RT-qPCR showing the levels of eRNA-Sphk1 and *KHPS1* transcribed from pREP4-SPHK1(-592/+1795)-luc and pREP4-SPHK1(-592/+1795polyA)-luc in NIH3T3 Tet-ON cells upon E2F1 transfection (N=3).

To investigate the impact of endogenous eRNA-Sphk1 on *SPHK1-C* transcription, cells were depleted of eRNA-Sphk1 using ASO. Knock-down of eRNA-Sphk1 markedly reduced the level of SPHK1 mRNA without affecting *KHPS1* levels (Figure 33). This result indicates the importance of eRNA-Sphk1 for activation of *SPHK1-C* transcription.

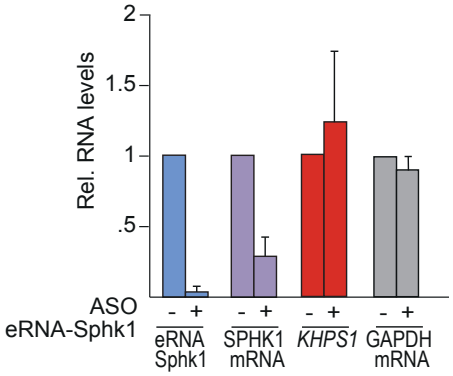

Figure 33. Knock-down of eRNA-Sphk1 leads to compromised transcription of SPHK1 mRNA.
Levels eRNA-Sphk1, SPHK1 mRNA, *KHPS1* and GAPDH mRNA were measured by RT-qPCR in U2OS/ER-E2F1 cells transfected with non-specific ASO (-) or ASO targeting eRNA-Sphk1 (at the region +352/+362 relative to *eSPHK1* TSS) (+) (N=3).

As a complementary approach to study the functional relevance of eRNA-Sphk1, the levels of SPHK1 mRNA were monitored upon downregulation of transcription from the *SPHK1* enhancer using the CRISPRi approach. For this, dCas9-KRAB co-repressor was targeted to the *SPHK1* enhancer using *eSPHK1*-specific sgRNA. This experimental approach resulted in decreased levels of eRNA-Sphk1. Compromised transcription of eRNA-Sphk1 led to

downregulation of SPHK1 mRNA (Figure 34A). Diminished transcription of *SPHK1-C* correlated with attenuated binding of initiating Pol II to the *SPHK1-C* promoter (Figure 34B), reinforcing that eRNA-Sphk1 transcription is required for enhanced *SPHK1* expression.

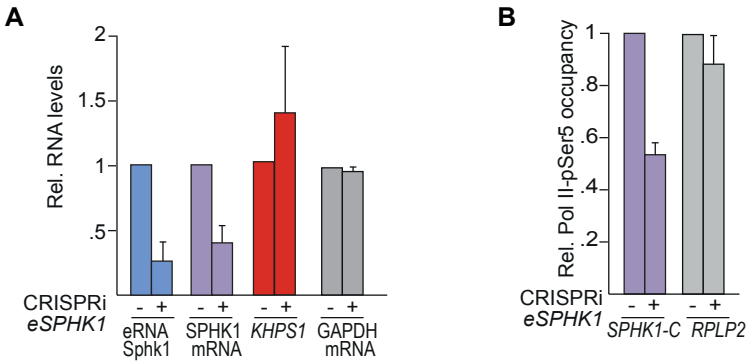

Figure 34. CRISPRi-mediated downregulation of eRNA-Sphk1 leads to decreased levels of SPHK1 mRNA and compromised binding of Pol II to the *SPHK1-C* promoter.
A) U2OS/ER-E2F1 cells were co-transfected with plasmids expressing dCas9-KRAB and sgRNA targeting either *eSPHK1* (-20/-1) (+) or a non-specific sgRNA (-). Levels of eRNA-Sphk1, SPHK1 mRNA, *KHPS1* and GAPDH mRNA were measured by RT-qPCR (N=3).
B) ChIP showing occupancy of initiating Pol II (Pol II-pSer5) at the *SPHK1-C* promoter normalized to total Pol II in U2OS/ER-E2F1 cells expressing dCas9-KRAB and either *eSPHK1*-specific sgRNA (+) or a control sgRNA (-) . Binding to the *RPLP2* promoter was monitored as control (N=3).

As a reciprocal approach, eRNA-Sphk1 transcription was upregulated by sgRNA-mediated targeting of dCas9-VP64 activating complex to *eSPHK1*. Increased transcription of eRNA-Sphk1 resulted in elevated levels of SPHK mRNA (Figure 35A). This coincided with increased binding of Pol II to the *SPHK1-C* promoter (Figure 35B) and led to increase in SPHK1 protein level (Figure 35C). These results substantiate that the *SPHK1* enhancer produces functional enhancer-derived transcripts that facilitate SPHK1 mRNA transcription.

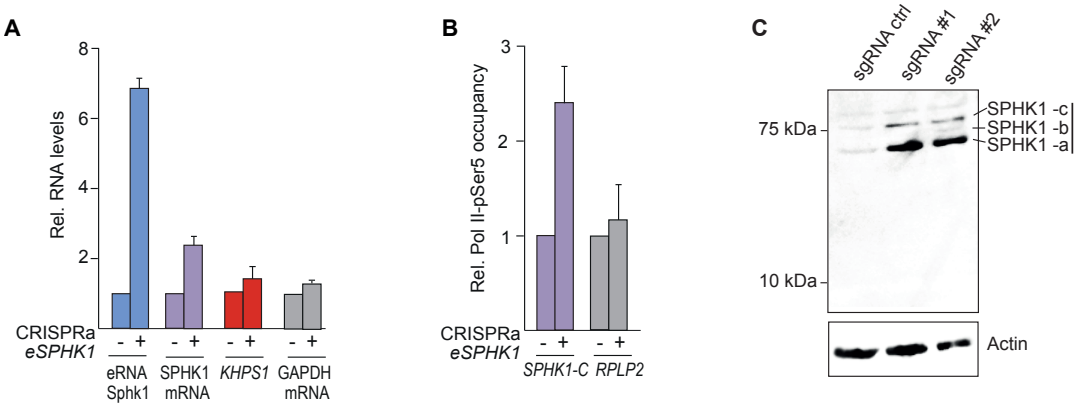

Figure 35. CRISPRa-mediate upregulation of eRNA-Sphk1 results in increased levels of SPHK1 mRNA, augmented binding of Pol II to the *SPHK1-C* promoter and increased expression **of SPHK1 proteins.**
A) RT-qPCR showing levels of eRNA-Sphk1, SPHK1 mRNA, *KHPS1* and GAPDH mRNA in U2OS/ER-E2F1 cells expressing dCas9-VP64 and sgRNAs targeting either *eSPHK1* (-20/-1) (+) or a non-specific sgRNA (N=3).
B) ChIP of initiating Pol II (Pol II-pSer5) at the *SPHK1-C* promoter normalized to total Pol II. U2OS/ER-E2F1 cells were co-transfected with plasmids expressing dCas9-VP64 and either *eSPHK1*-specific sgRNA (+) or a control sgRNA (-). Binding to the *RPLP2* promoter was monitored as control (N=3).
C) Western blot showing SPHK1 protein isoforms originating from alternatively spliced SPHK1-C transcripts. U2OS/ER-E2F1 cells were co-transfected with a plasmid expressing dCas9-VP64 and *eSPHK1*-specific (sgRNA #1 and #2) sgRNAs or control sgRNA (sgRNA ctrl).

2.10. Downregulation of *KHPS1* and eRNA-Sphk1 compromises the malignant phenotype of cancer cells

Having established the link of *KHPS1*, eRNA-Sphk1 and SPHK1 mRNA transcription, the mechanism underlying *KHPS1*-depenent activation of *SPHK1* expression was investigated. For this, the levels of SPHK1 mRNA were monitored after knock-down of *KHPS1*. Depletion of *KHPS1* reduced the levels of both eRNA-Sphk1 and SPHK1 mRNA (Figure 36). This result links two different classes of ncRNA, demonstrating the requirement of an lncRNA for activation of eRNA, which in turn augments mRNA transcription.

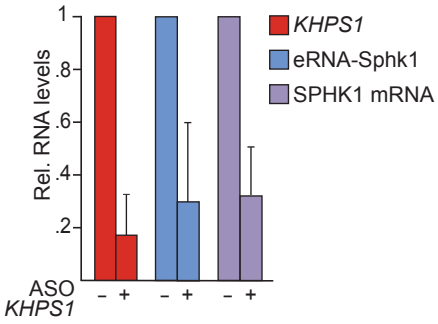

Figure 36. Downregulation of *KHPS1* transcription leads to decreased levels of eRNA-Sphk1 and SPHK1 mRNA.
RT-qPCR shows levels of *KHPS1*, eRNA-Sphk1 and SPHK1 mRNA in U2OS/ER-E2F1 cells transfected with control ASO (-) or *KHPS1*-specific ASO targeting nucleotides –101/-121 relative to the TSS of *eSPHK1* (+) (N=3).

Upregulated transcription of *SPHK1* is known to correlate with tumor development and cancer progression (Sarkar et al., 2005; Zhang et al., 2014; Zhu et al., 2015). Given that depletion of *KHPS1* and eRNA-Sphk1 leads to decrease in *SPHK1* expression, the tumorigenic potential of cells depleted of *KHPS1* and eRNA-Sphk1 was examined. To this end, ASOs against *KHPS1* or eRNA-Sphk1 were transfected into MDA-MB-231 breast cancer cells which express high levels of SPHK1 (Datta et al., 2014). Cell migration was monitored by wound healing assays. Depletion of either *KHPS1* or eRNA-Sphk1 led to a profound delay in gap closure. Retarded cell migration was similar to the phenotype observed in cells treated with the SPHK1 inhibitor SKI II (French et al., 2003) (Figure 37), demonstrating the functional relevance of *KHPS1* and eRNA-Sphk1 for enhancer-driven *SPHK1* expression.

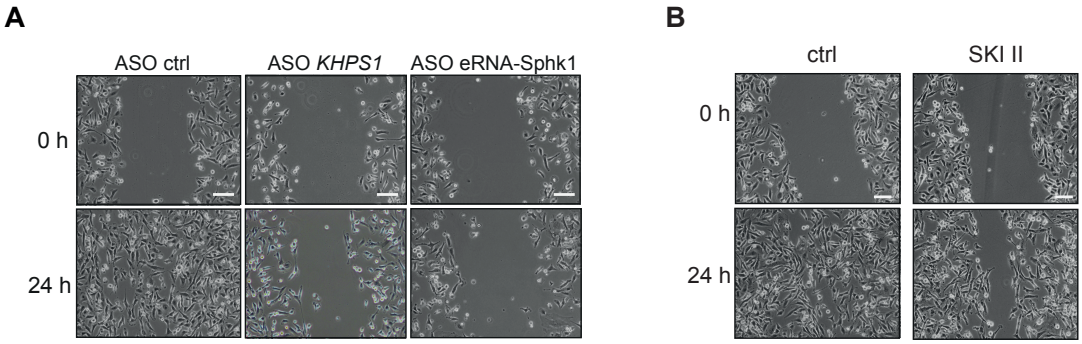

Figure 37. Depletion of *KHPS1* or eRNA-Sphk1 leads to reduced migration ability of the cancer cells.
A) Wound-healing assay in MDA-MB-231 cells transfected with a control ASO (ctrl) or with ASOs targeting *KHPS1* (–101/-121) or eRNA-Sphk1 (+352/+362). Gap closure was monitored 0 h and 24 h after scratching by bright-field microscopy. Scale bars, 100 µm.
B) Wound-healing assay in MDA-MB-231 cells treated with DMSO (ctrl) or with 10 µM of SKI II. Gap closure was monitored 0 h and 24 h after scratching by bright-field microscopy. Scale bars, 100 µm.

To further examine the impact of downregulation of *KHPS1* and eRNA-Sphk1 on the malignant behavior of cancer cells, the invasive capacity of control and ASO-treated MDA-MB-231 cells was compared using the Matrigel invasion assay. For this, transfected cells were added on top of the Matrigel coated transwell. Cells which migrated through the extracellular matrix after 24 h were fixed, stained with crystal violet and counted. The capability of cells to invade through the membrane was severely impaired after depletion of either *KHPS1* or eRNA-Sphk1, again resembling the phenotype of cells treated with the SKI II inhibitor (Figure 38A,B).

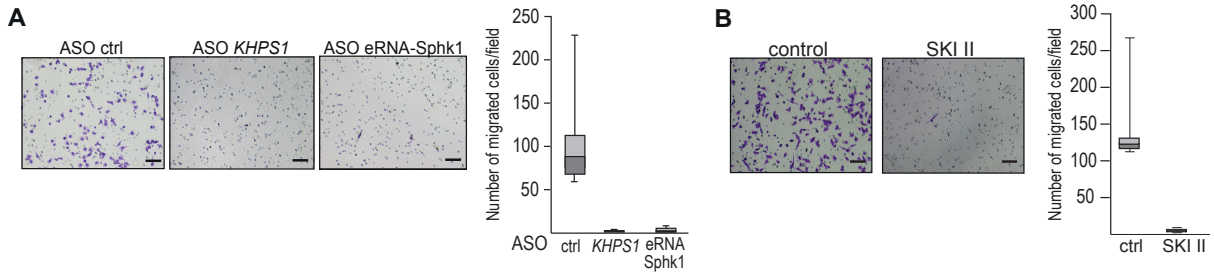

Figure 38. Knock-down of *KHPS1* or eRNA-Sphk1 decreases invasive capability of cancer cells.
A) Cell invasion assay of MDA-MB-231 cells transfected with a control ASO (ctrl) or with ASOs targeting *KHPS1* (–101/-121) or eRNA-Sphk1 (+352/+362). The images of the cells migrated through the matrigel were taken using bright-field microscopy. Scale bars, 100 µm. The number of migrated MDA-MB-231 cells is presented by box plot (n=7, whiskers: 5–95 percentile, light grey: mean value).
B) Cell invasion assay of MDA-MB-231 cells treated with DMSO (ctrl) or with 10 µM of SKI II. The images of the cells migrated through the matrigel were taken using bright-field microscopy. Scale bars, 100 µm. The number of migrated MDA-MB-231 cells is presented by box plot (n=7, whiskers: 5–95 percentile, light grey: mean value).

Furthermore, the effect of *KHPS1* and eRNA-Sphk1 depletion on the malignant phenotype was investigated by monitoring clonogenicity of MD-MBA-231 cells. ASO-mediated targeting *KHPS1* or eRNA-Sphk1 prevented colony formation of cells in soft-agar, reinforcing the stimulating effect of both regulatory RNAs on cancer cell tumorigenicity (Figure 39). These results demonstrate that downregulation of *KHPS1* or eRNA-Sphk1 suppresses metastatic features of cancer cells by compromising *KHPS1*- and eRNA-Sphk1-dependent expression of *SPHK1*.

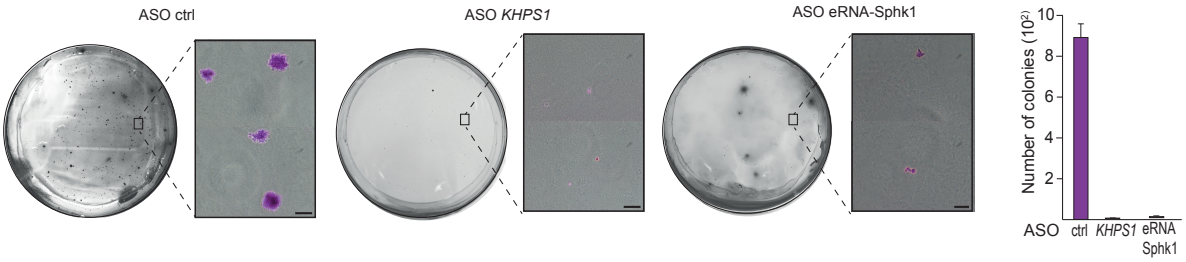

Figure 39. Decreased levels of *KHPS1* or eRNA-Sphk1 lead to impaired clonogenic potential of cancer cells.
Soft agar colony formation assay in MDA-MB-231 cells transfected with control ASO (ctrl) or ASOs against *KHPS1* or eRNA-Sphk1. Representative images of the colonies were taken by brightfield microscopy 3 weeks after transfection. Scale bars, 100 µm. The average number of colonies formed per plate is shown by bar diagram (N=2).

2.11. eRNA-Sphk1 evicts CTCF insulating *SPHK1* enhancer from *SPHK1-C* promoter

Having established the positive correlation between eRNA-Sphk1 transcription and expression of its target gene *SPHK1-C*, the mechanism contributing to eRNA function was investigated. Previous studies have demonstrated that eRNAs enhance transcription through stabilization of CTCF-mediated enhancer-promoter looping (Kim et al., 2015; Werner et al., 2017; Yang et al., 2016). Inspection of ChIP-seq data deposited in ENCODE data base revealed the presence of a CTCF binding site in the first intron of eRNA-Sphk1 (Figure 40A). To examine the effect of eRNA-Sphk1 transcription on CTCF binding, CTCF occupancy was monitored in U2OS/ER-E2F1 before and after 4-OHT treatment. E2F1-induced activation of eRNA-Sphk1 led to reduced binding of CTCF to its target site located between *eSPHK1* and *SPHK1-C* promoter (Figure 40B), suggesting CTCF-mediated molecular mechanism underlying eRNA-Sphk1 function.

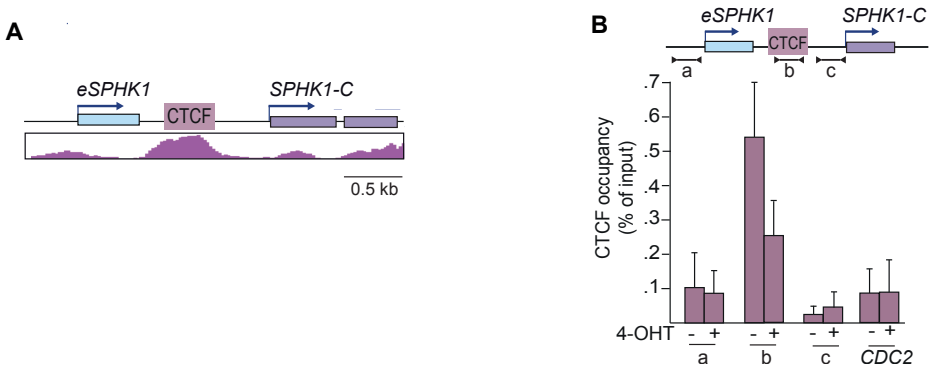

Figure 40. CTCF placed between *SPHK1* enhancer and promoter is evicted upon E2F1-induction.
A) ChIP-seq track showing CTCF present between the *SPHK1* enhancer and the *SPHK1-C* promoter in osteoblasts (GEO accession number GSM733784).
B) ChIP showing occupancy of CTCF in untreated and 4-OHT-induced U2OS/ER-E2F1 cells. The regions analyzed by PCR are: -464/-698 (a), +638/+790 (b), +1658/+1795 (c). Binding to the *CDC2* promoter was monitored as control (N=3).

To substantiate the link between eRNA-Sphk1 transcription and CTCF, the binding of CTCF was monitored upon CRISPRa-mediated activation of eRNA-Sphk1 transcription. Cells transfected with dCas9-VP64 targeted to *SPHK1* enhancer exhibited decreased CTCF binding at the *SPHK1* locus (Figure 41A). Importantly, eviction of CTCF led to enhanced H3K27ac occupancy specifically at *SPHK1-C*, reinforcing that transcription of eRNA-Sphk1 triggers displacement of the CTCF, thereby removing the repressive boundary between *eSPHK1* and *SPHK1-C* (Figure 41B).

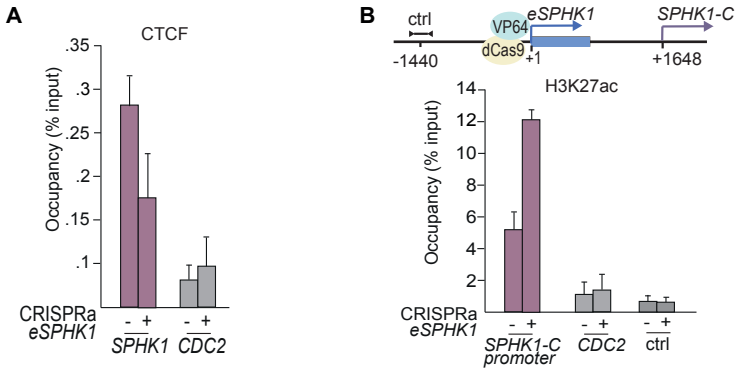

Figure 41. Transcription of eRNA-Sphk1 evicts CTCF at the *SPHK1* locus followed by increase occupancy in H3K27ac at the *SPHK1-C* promoter.
A) ChIP showing occupancy of CTCF in U2OS/ER-E2F1 cells co-transfected with a dCas9-VP64 expression vector and either e*SPHK1*-specific (+) or non-specific (-) sgRNAs. Binding to the *CDC2* promoter was monitored as control (N=3).
B) ChIP measuring H3K27ac at the *SPHK1-C* promoter in U2OS/ER-E2F1 cells expressing dCas9-VP64 and either e*SPHK1*-specific (+) or non-specific (-) sgRNAs. Occupancy at *CDC2* and a region upstream of the *eSPHK1* TSS (-1440) were monitored as control (N=3).

To corroborate the insulating function of CTCF, genomic deletion of the CTCF putative binding sites was performed using CRISPR/Cas9 approach (Figure 42A). In accord with eRNA-Sphk1-medaited eviction of CTCF, depletion of CTCF binding site led to increased level of SPHK1 mRNA. Importantly, levels of *KHPS1* and eRNA-Sphk1 remained unchanged, indicating that CTCF insulated *SPHK1* promoter from the enhancer, suppressing SPHK1 mRNA transcription (Figure 42B).

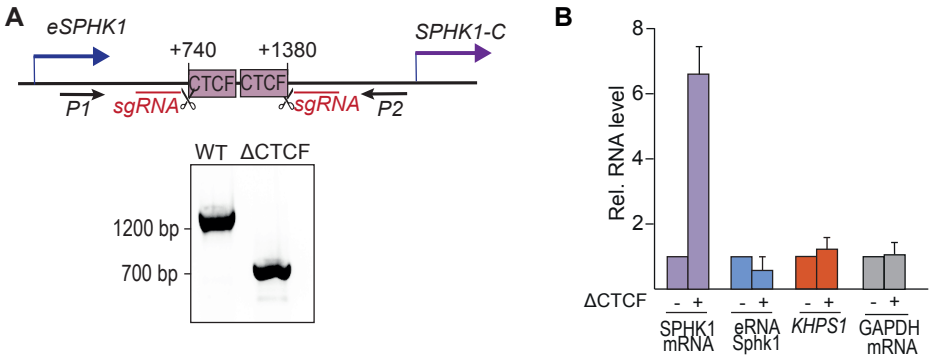

Figure 42. Genomic deletion of CTCF binding site between *SPHK1* enhancer and promoter leads to increased transcription of SPHK1 mRNA.
A) PCR showing the CRISPR-Cas9-mediated deletion of CTCF binding sites (Δ+737/+1390) indicated in the scheme) in U2OS/ER-E2F1 cells (ΔCTCF). PCR was performed using primers P1:+108F and P2: +1418R. Parental cells were used as control (WT).
B) Levels of SPHK1 mRNA, eRNA-Sphk1, *KHPS1* and GAPDH mRNA in parental U2OS/ER-E2F1 cells (WT) or cells lacking the CTCF binding sites between the TSS of *eSPHK1* and *SPHK1-C* (ΔCTCF) after induction with 4-OHT (3 h) (N=3).

Increased *SPHK1-C* transcription upon removal of CTCF suggested that CTCF acts an insulator that blocks the spreading of the active chromatin from the enhancer to the promoter at the *SPHK1* locus. To verify this, the occupancy of H3K27ac and Pol II at the *SPHK1-C* promoter was monitored in cells lacking the CTCF binding sites. In accord with effects observed upon eRNA-Sphk1-mediated removal of CTCF, enrichment of Pol II and H3K27ac at the *SPHK1-C* promoter was higher in cells lacking CTCF binding sites as compared to parental cells line. Interestingly, H3K27ac occupancy acquired uniform distribution between the *SPHK1* enhancer and *SPHK1-C* promoter (Figure 43), underscoring the insulating function of CTCF that prevents spreading of the chromatin state from the enhancer to the promoter at the *SPHK1* locus.

Collectively, these results indicate that transcription of eRNA-Sphk1 augments transcription of SPHK1 mRNA by evicting CTCF that insulates the *SPHK1* enhancer from the *SPHK1-C* promoter.

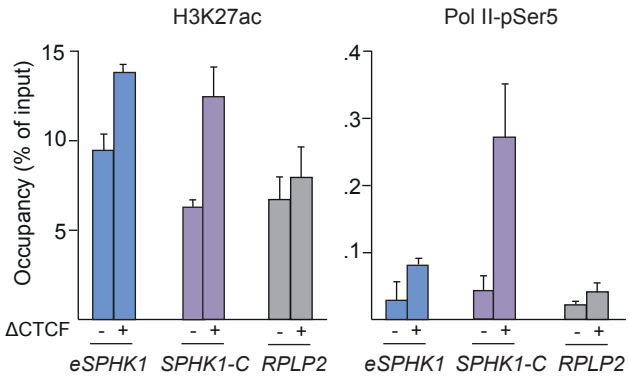

Figure 43. Genomic deletion of CTCF binding sites between *SPHK1* enhancer and promoter induces establishment of H3K27ac and increases binding Pol II at the *SPHK1-C* promoter.
ChIPs showing occupancy of H3K27ac and initiating Pol II at the *eSPHK1* and *SPHK1-C* promoter in U2OS/ER-E2F1 parental cells (-) or cells lacking the CTCF binding sites between the TSS of *eSPHK1* and *SPHK1-C* (ΔCTCF, +) after induction with 4-OHT. Binding to the *RPLP2* promoter was monitored as control (N=3).

2.12. Triplex formation is indispensable for *SPHK1* expression

Next, the in vivo relevance of triplex formation for *SPHK1* expression was investigated. For this, endogenous *KHPS1* was prevented from association with the *SPHK1* enhancer by competition with synthetic TFR-containing RNA. Transfection of ectopic RNA comprising the TFR of *KHPS1* (-373/-241) led to reduced transcription of eRNA-Sphk1 upon 4-OHT induction, indicating that the ectopic *KHPS1* competed for binding of cellular *KHPS1* to *eSPHK1*. Importantly, transfection of synthetic *KHPS1* versions in which the TFR was deleted (ΔTFR) or mutated (mutTFR). RNA comprising sequences corresponding to the intron of eRNA-Sphk1 (+638/+790), did not affect transcription activation of eRNA-Sphk1 (Figure 44A), underscoring the specificity of the competition approach. Furthermore, if *KHPS1* was prevented from binding to *eSPHK1* by TFR-containing RNA, ChIP experiments revealed reduced occupancy of Pol II and p300 at the *SPHK1* enhancer upon 4-OHT treatment. *KHPS1* complementary to the intron of eRNA-Sphk1 did not impact on the 4-OHT-induced increase in Pol II and p300 at the *SPHK1* enhancer (Figure 44B). These results reinforce the requirement of triplex formation for guidance of regulatory proteins to activate eRNA-Sphk1. Together, these results demonstrate that ectopic TFR-containing RNA competes for binding of *KHPS1* to *eSPHK1*, supporting that binding of endogenous *KHPS1* to DNA is required for p300 recruitment and transcription of eRNA-Sphk1.

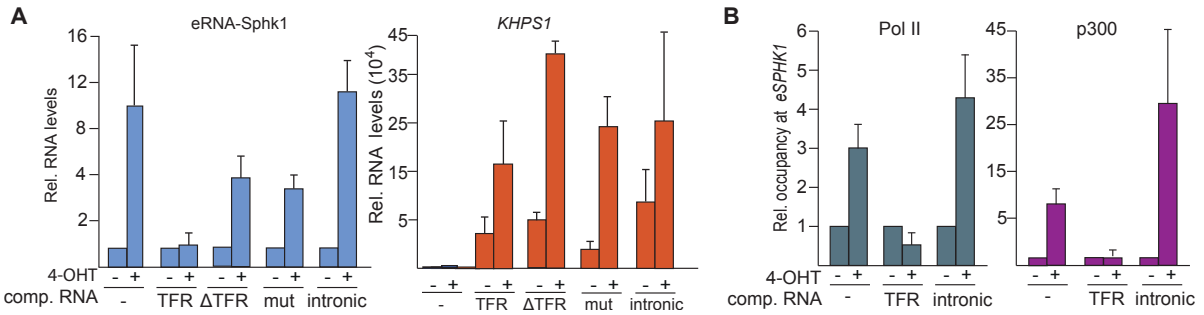

Figure 44. TFR-containing synthetic RNA prevents endogenous *KHPS1* from binding to the *SPHK1* enhancer.
A) Competition experiment showing the levels of eRNA-Sphk1 in untransfected (-) U2OS/ER-E2F1 cells or in cells transfected with synthetic *KHPS1* (-373/-241) harboring either intact TFR (WT), deleted (ΔTFR), mutated TFR (mut) or sequences corresponding to the intron of eRNA-Sphk1(+638/+790, intronic). eRNA-Sphk1 transcription was induced with 4-OHT (2 h) (N=3).
B) ChIPs showing binding of Pol II and p300 to *eSPHK1* in untransfected U2OS/ER-E2F1 cells (-) or upon transfection of short *KHPS1* comprising the TFR (-373/-241, TFR) or *KHPS1* sequences corresponding to the intron region of eRNA-Sphk1 +638/+790 (intronic). Where it is indicated cells were treated with 4-OHT for 2 h (N=3).

Next, the importance of triplex formation for SPHK1-mediated cellular tumorigenic potential was tested. For this, the association of *KHPS1* with *SPHK1* enhancer was prevented by transfection of ectopic TFR-containing *KHPS1* in highly metastatic MD-MBA-231 cells. Wound healing assay demonstrated reduced migration of the cells transfected with *KHPS1* harbouring the TFR as compared to the cells transfected with the *KHPS1* sequences corresponding to the intron of eRNA-Sphk1 (Figure 45A). Moreover, invasive abilities of the cells, measured by Matrigel invasion assays, were significantly decreased if the cells were transfected with the short TFR-containing RNA. Transfection with intronic control RNA did not affect cellular invasion (Figure 45B). Likewise, soft agar colony formation assay revealed decrease in colony formation of the cells transfected with TFR-containing RNA in contrast to intronic RNA (Figure 45C). Together, these experiments indicate that triplex formation at the

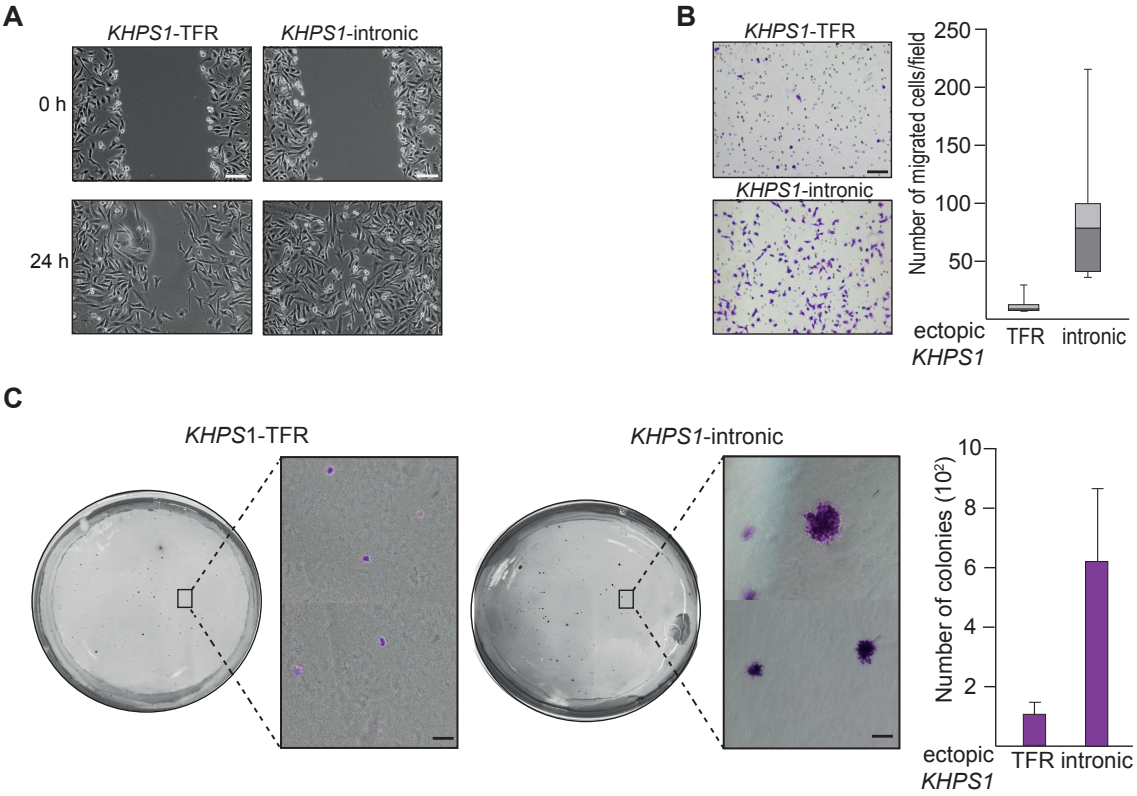

Figure 45. Ectopic TFR-containing *KHPS1* impairs migration, invasion and colony formation of cancer cells.
A) Wound-healing assay in MDA-MB-231 cells transfected with *KHPS1* versions comprising the TFR (-373/-241) or intronic sequences +638/+790 (intronic). Gap closure was monitored 0 h and 24 h after scratching by bright-field microscopy. Scale bars, 100 µm.
B) Cell invasion assay of MDA-MB-231 cells transfected with *KHPS1* versions comprising either the TFR (-373/-241) or intronic sequences +638/+790 (intronic). The images of the cells migrated through the matrigel were taken using bright-field microscopy. Scale bars, 100 µm. The number of migrated MDA-MB-231 cells is presented by box plot (n=7, whiskers: 5–95 percentile, light grey: mean value).
C) Soft agar colony formation assay in MDA-MB-231 cells transfected with *KHPS1* versions comprising the TFR (-373/-241) or region corresponding to intronic sequences of eRNA-Sphk1 +638/+790 (intronic). Representative pictures of the colonies were imaged 3 weeks after transfection by brightfield microscopy. Scale bars, 100 µm. The average number of colonies formed per plate is presented by bar diagram (N=2).

In order to substantiate the functional importance of *KHPS1*-mediated triplex formation for *SPHK1* expression, the TFR was deleted using the CRISPR-Cas9 approach. Deletion of a 66 bp of genomic sequence comprising the TFR in haploid HAP1 cells appeared to be lethal. To overcome the lethal phenotype caused by SPHK1 deficiency upon deletion of the TFR, MDA-MB-231 cells, which are triploid for chromosome 17 harboring *SPHK1,* were used. In accord with the results in HAP1 cells, screening of numerous clonal cell lines revealed no homozygous deletion, underscoring that lack of the TFR is lethal for cells. This result substantiates the importance of the TFR for *SPHK1* expression and cell viability. Consistent with this view, the proliferation rate of clones with mono- or biallelic deletions of the TFR ($TFR^{+/+/-}$ and $TFR^{+/-/-}$) was significantly reduced as compared to the parental cell line. Importantly, biallelic deletion

of the TFR (TFR$^{+/-/-}$) impaired cell proliferation to a higher extent and led to cell death (Figure 47A, B).

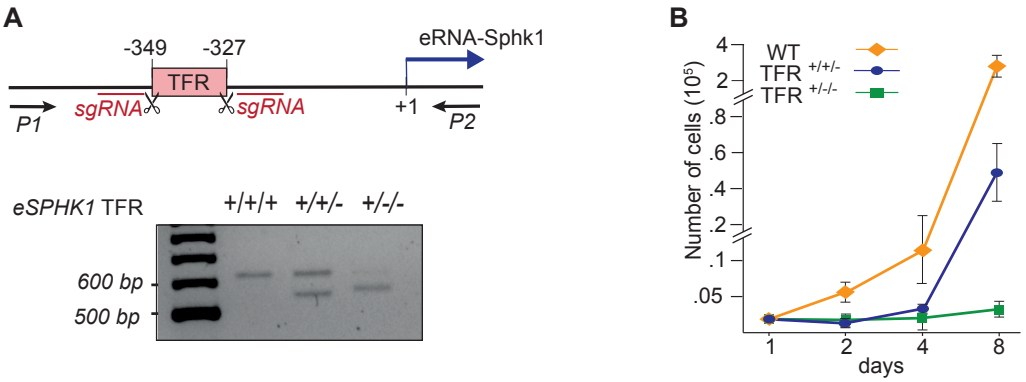

Figure 47. Genomic deletion of the TFR at *SPHK1* enhancer leads to impaired cellular proliferation.
A) PCR using primers P1:-592F and P2:+25R showing monoallelic (+/+/-) or biallelic (+/-/-) deletion of the TFR in MDA-MB-231 cells. Parental cells (+/+/+) were used as control.
B) Proliferation rate of parental (WT) and mutant MDA-MB-231 cells comprising a monoallelic (TFR+/+/-) or biallelic (TFR +/-/-) TFR deletions (N=3).

Further, the effect of genomic TFR deletion on *SPHK1* expression was tested. For this, the levels of sense and antisense RNAs transcribed at *SPHK1* were monitored in the parental and the TFR$^{+/+/-}$ cell line. Both eRNA-Sphk1 and SPHK1 mRNA levels were markedly decreased in TFR$^{+/+/-}$ cells (Figure 48). Importantly, the level of *KHPS1* was similar in all clones, indicating that deletion of the TFR did not impact on the synthesis of *KHPS1* but impaired its functionality, resulting in compromised transcription activation of eRNA-Sphk1 and SPHK1 mRNA. These results provided in vivo evidence that anchoring *KHPS1* to the enhancer TFR is pivotal for SPHK1-dependent functions.

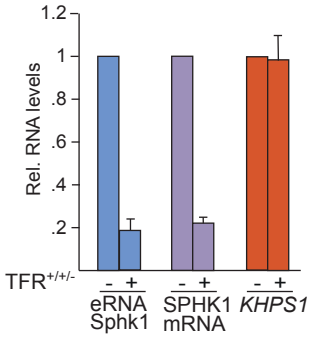

Figure 48. Monoallelic deletion of the TFR at the *SPHK1* enhancer compromises the activation of eRNA-Sphk1 followed by diminished SPHK1 mRNA transcription.
Levels of eRNA-Sphk1, SPHK1 mRNA and *KHPS1* in parental (-) and in mutant MDA-MB-231 cells comprising a monoallelic TFR deletion (ΔTFR$^{+/+/-}$) (+) (N=3).

To corroborate the effect of TFR deletion on cancerogenic properties of MD-MBA-231 cells, migration and invasiveness of the cells harbouring a monoallelic deletion of the TFR (TFR$^{+/+/-}$) compared to parental cells. Wound healing assay revealed considerable delay in gap closure

in TFR$^{+/+/-}$ cells as compared to cell line comprising three copies of the TFR (Figure 49A). Similarly, the Matrigel invasion assays showed decreased ability of the TFR$^{+/+/-}$ cells to migrate through the extracellular matrix (Figure 49B). Collectively, these experiments reveal a hierarchical regulatory cascade in which *KHPS1* tethered to the *SPHK1* enhancer promotes transcription of eRNA-Sphk1 which in turn is a prerequisite for upregulation of SPHK1 mRNA and cell proliferation.

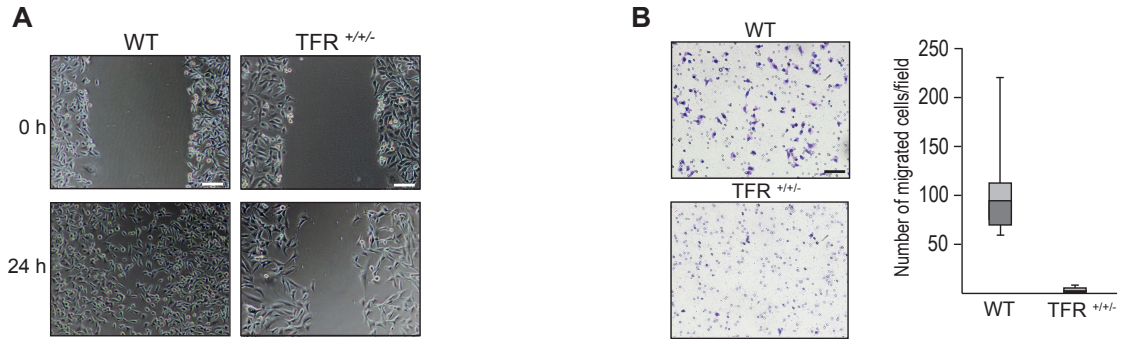

Figure 49. **Monoallelic deletion of the *eSPHK1* TFR compromises the malignant phenotype of cancer cells. A)** Wound-healing assay performed in wildtype (WT) and in mutant MDA-MB-231 cells comprising a monoallelic TFR deletion (ΔTFR$^{+/+/-}$). Gap closure was monitored 0 h and 24 h after scratching by bright-field microscopy. Scale bars, 100 µm.
B) Cell invasion assay in wildtype (WT) and ΔTFR$^{+/+/-}$ MDA-MB-231 cells. The images of the cells migrated through the matrigel were taken using bright-field microscopy. Scale bars, 100 µm. The number of cells migrated through Matrigel is presented by box plot (n=7, whiskers: 5–95 percentile, light grey: mean value).

3. Discussion

3.1. Antisense transcript *KHPS1* regulates transcription of its neighboring gene *SPHK1*

Regulatory lncRNAs have been shown to be involved in activation or silencing of gene transcription by serving as decoys, guides and scaffolds for functional proteins (Wang and Chang, 2011). Antisense transcript Lrp1-AS serves as molecular decoy for high-mobility group box 2 (Hmgb2) protein, which increases the accessibility of the chromatin for transcription factors, thereby leading to downregulation of Lrp1 mRNA (Yamanaka et al., 2015). In contrast, antisense ncRNA APO1-AS acts as a modular scaffold for histone-modifying enzymes, LSD1, SUZ12, and the PRC2 complex to establish transcriptionally repressive chromatin at the neighboring genes *APOA1, APOA4* and *APOC3* (Halley et al., 2014). Another antisense transcript, termed ANRIL, was shown to regulate expression of its neighboring gene cluster *INK4b-ARF-INK4a* by guiding PRC1 and PRC2 complexes and establishing repressive chromatin state at the promoters of *INK4b-ARF-INK4a* (Kotake et al., 2011). Moreover, an antisense transcript, named *KHPS1*, has been shown to activate transcription of its counterpart sense SPHK1 mRNA in both rat and human. However, rat and human *KHPS1* utilize different epigenetic mechanism to achieve activation of *SPHK1*. Rat *KHPS1* has been shown to mediate demethylation of the CG-island in the promoter of *Sphk1* (Imamura et al., 2004a), whereas human *KHPS1* has no effect on DNA methylation (Postepska-Igielska et al., 2015), but it guides E2F1 and p300 to the *SPHK1* locus. These data support a regulatory role for antisense transcripts in guiding epigenetic regulators to shape local chromatin state and control expression of the neighboring genes.

3.2. *KHPS1* activates a poised *SPHK1* enhancer by recruiting p300 and E2F1

Recent genome-wide studies have revealed thousands of distal regulatory elements, demonstrating that activity of some enhancers is switched on and off during development. The inactive enhancers, that harbor histones marks H3K4me1 and H3K27me3 and lack H3K27ac, are termed poised enhancers (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Rhie et al., 2014; Zentner and Henikoff, 2013). The knowledge about transition of poised into activate enhancers is gained primarily from ChIP-seq analyses, demonstrating dynamic establishment of the chromatin histone mark H3K27ac in expense of H3K27me3 and recruitment of p300 upon cellular differentiation (Bossen et al., 2015; Choukrallah et al., 2015; Khilji et al., 2018; Rada-Iglesias et al., 2011; Svotelis et al., 2011). However, the mechanistic details of how the

cellular stimuli lead to the enhancer activation remain unclear. Although lncRNAs have been implicated in a great number of nuclear signaling cascades, they have been mostly functionally linked to protein coding genes. This result is in accord with a recent *in silico* study which unraveled the potential of lncRNAs to target and activate super-enhancers. Bioinformatic analyses demonstrated that some of lncRNAs contain a triplex-forming region that can form RNA-DNA triplexes with a corresponding DNA site at the super enhancers (Soibam, 2017). In support of this study, here it has been demonstrated that *KHPS1* mediated transition of a poised into an active enhancer. *KHPS1* recruits p300 to the poised *SPHK1* enhancer, thereby establishes H3K27ac in expense of H3K27me3. This in turn activates transcription of enhancer-derived RNA, termed eRNA-Sphk1. Similarly, lncRNA LED was found to be associated with DNA at the *p21* enhancer. Transcription of LED correlated with deposition of H3K9ac at the enhancer, providing circumstantial evidence of lncRNA-mediated enhancer activation (Léveillé et al., 2015).

Most of lncRNAs exhibit weak evolutionary sequence conservation (Diederichs, 2014). Nevertheless, several lncRNAs have been shown to exhibit conservation of their regulatory function across species, predicting conserved biological functions of lncRNAs (Ulitsky, 2016). For instance, human and murine lncRNAs FIRRE are both transcriptionally induced by the NF-kB signaling pathway and regulate the same inflammatory response genes through interacting with hnRNPU (Lu et al., 2017). Another lncRNA *IL7-AS* is induced across human and mouse cell types and regulates the expression of interleukin-6 (IL6) (Roux et al., 2017). Here, rat and human *KHPS1* exhibit 88 % of sequence similarity and both have positive effect on *SPHK1* transcription, supporting that biologically relevant lncRNAs exhibit cross-species conservation.

A recent study revealed that enrichment of p300 correlates with enhancer activity, whereas lack of p300 characterizes inactive state of enhancers (Visel et al., 2009). The requirement of p300 for *SPHK1* enhancer activation was demonstrated by treatment with curcumin, a chemical that inhibits the activity of p300/CBP (Balasubramanyam et al., 2004). Addition of curcumin compromised *KHPS1*-induced activation of eRNA-Sphk1 transcription in vivo. Similar effect was observed at the reporter construct, which drove expression of luciferase fused in frame with eRNA-Sphk1. Treatment with curcumin or depletion of p300 attenuated luciferase expression, underscoring that activation of *SPHK1* enhancer is caused by *KHPS1*-mediated recruitment of p300.

3.3. *KHPS1* interacts with p300 and E2F1

LncRNA-mediated tethering of chromatin modifiers to gene promoters can be accomplished by direct interaction of RNAs with proteins or through other protein intermediates. Antisense transcript lncRNA-JADE has been shown to recruit p300 to the *Jade1* promoter upon DNA damage. However, recruitment is mediated by interaction of lncRNA-JADE with Brca1, which upon DNA damage is associated with p300 (Wan et al., 2013). Similarly, lncRNA PACER recruits p300 to the *COX-2* promoter without direct interaction with p300 (Krawczyk and Emerson, 2014). Results of this study demonstrate that *KHPS1* interacts with p300 directly. The binding of *KHPS1* with p300 was clearly detected by CLIP experiments performed after UV-crosslinking, a technique which fixes direct protein-RNA contacts (Zhang and Darnell, 2011). Immunoprecipitated p300-*KHPS1* complexes were resistant to washings with buffers containing 1M NaCl and 2M urea. These results exclude the possibility that the interaction of *KHPS1* with p300 is mediated by other proteins. The ability of p300 to interact with lncRNAs was supported by a study demonstrating that ZEB1-AS1 tethers p300 to the *ZEB1* promoter by direct association with p300 (Liu and Lin, 2016). However, none of the studies, including the present one, elucidated the role of lncRNA in p300 function. As p300 shares the same activity center as CBP (Wang et al., 2008), binding of RNA to p300 may stimulate its HAT activity, as shown for CBP (Bose et al., 2017).

In addition to p300, *KHPS1* is engaged in direct interaction with transcription factor E2F1. E2F1 is a well characterized DNA-binding protein, which induces transcription of the genes responsible for entry into the S phase and cell cycle progression (Chen et al., 2009; Huber et al., 1993; Slansky and Farnham, 1996). Recent studies revealed several lncRNAs that are bound to E2F1. For example, lncRNA GAS5 positively regulates expression of $P27^{Kip1}$ by that enhancing binding of E2F1 to the $P27^{Kip1}$ promoter (Luo et al., 2017). LncRNA EPEL, expressed in lung cancer cells, was shown to regulate expression of E2F1-responsive gene by interaction with E2F1 and increasing the binding efficiency of E2F1 to DNA (Park et al., 2018). Finally, lncRNA-HIT contributes to proliferation of NSCLC cells by binding to E2F1, thus regulating the interaction of E2F1 with its target genes (Yu et al., 2017). The identification of *KHPS1*-E2F1 interaction extends the list of lncRNAs involved in interaction with E2F1. Moreover, the results of this study revealed reduced association of E2F1 with chromatin upon RNase A or inhibition of transcription by flavopiridol, underscoring the functional relevance of RNAs in E2F1 recruitment/stabilization to DNA. However, further investigation is required to unravel the exact function of RNAs in transcription regulation by E2F1. Importantly, other genuine DNA binding factors were reported to be associated with RNAs as well. For example,

ncRNA inhibits STAT1 transcription factor in MHC expression, TLS/FUS factor interacts with RNA, and YY1 factor was shown to bind eRNAs (Perrotti et al., 1998; Peyman, 1999; Sigova et al., 2015; Zhou et al., 2015). These reports together with the present study support that association of lncRNA with transcription factors might be a general mechanism by which lncRNAs impact the regulatory network in the cell.

3.4. eRNA-Sphk1 facilitates transcription of SPHK1 mRNA

Large scale transcriptome profiling revealed specific features of ncRNA transcribed from enhancers, such as nuclear localization, rapid decay and a low copy number (Hsieh et al., 2014; Kim et al., 2010; Lam et al., 2014; Li et al.,2013). The present study experimentally demonstrates that eRNA-Sphk1 exhibits characteristic features of eRNAs, such as nuclear localization, short half-life and low abundance. Western blot analyses did not reveal polypeptides that correlate with eRNA-Sphk1 transcription, reinforcing that eRNA-Sphk1 is a non-coding eRNA.

Whether eRNAs are byproducts of transcription caused by Pol II binding or are functional players in gene regulatory networks is still a matter of dispute. However, numerous studies demonstrated the regulatory function of eRNAs. For example, depletion of eRNAs TFF1 or FOXC1 by siRNAs inhibited transcription of the adjacent coding regions (Li et al.,2013). Furthermore, siRNA-mediated depletion of p53-induced eRNAs led to reduced p53-dependent activation of target genes. Moreover, is has been shown that ectopic eRNA tethered to reporter is able to enhance gene expression, demonstrating the importance of eRNA *per se* rather than the process of enhancer transcription (Melo et al., 2013). The present study supports that eRNAs are functional, regulatory units that affect expression of the target protein-coding genes. Induction of eRNA-Sphk1 by dCas9-VP64 led to increased *SPHK1* expression. Conversely, ASO-mediated knock-down or dCas9-KRAB-mediated repression of eRNA-Sphk1 led to compromised transcription of SPHK1 mRNA, indicating that not only the transcription process but also eRNA-Sphk1 itself is required for expression of SPHK1.

In contrast to the activating role of eRNAs, intragenic enhancers have been shown to inhibit gene expression. The process of eRNA transcription creates an obstacle for Pol II which transcribes the host gene, thereby attenuating the transcription of target genes (Cinghu et al., 2017). *SPHK1* enhancer is located within the *KHPS1* locus. Thus, transcription of *KHPS1* from one strand can interfere with the eRNA-Sphk1 transcription on the opposite strand, which would lead to the collision of RNA polymerases. Investigation of the kinetics of *KHPS1* and

eRNA-Sphk1 transcription revealed that E2F1 activates *KHPS1* transcription before eRNA-Sphk1 is upregulated. Therefore, transcription of *KHPS1* precedes activation of sense eRNA-Sphk1, indicating that transcripts are expressed in temporal order, thus precluding the Pol II collision.

Overexpression of the protooncogene *SPHK1* was shown to promote tumorigenesis (Zhang et al., 2014). Downregulation of *SPHK1* expression using siRNAs led to compromised malignant phenotype of cancer cells by blocking proliferation, migration and invasion of cells (Maiti et al., 2017; Sarkar et al., 2005; Zhang et al., 2014). In accord with these results, knock-down of either *KHPS1* or eRNA-Sphk1 led to reduced migration, invasion and colony formation of single cells, suggesting that lncRNA-mediated downregulation of *SPHK1* expression could be used to suppress tumorigenesis.

3.5. Transcription of eRNA-Sphk1 evicts CTCF

eRNAs have been shown to promote the interaction of enhancers and promoters, thus activating transcription of protein-coding genes. For instance, knockdown of eRNA from the estrogen receptor alpha (ERα)-bound enhancers *NR1P1* or *GREB1* reduced enhancer–promoter interaction (Lam et al., 2014). Moreover, transcription of eRNA HIDALGO led to increased chromatin contacts of the enhancer with the target promoter of *HBG1* (Werner et al., 2017). Another eRNA, HPSE, has been shown to trigger the chromatin looping leading to interaction of the enhancer with the *HPSE* promoter (Jiao et al., 2018), supporting the role of eRNAs in bridging enhancers and promoters to facilitate expression of target genes. Importantly, CTCF functions as one of the core proteins engaged in chromatin looping thereby establishing interactions between enhancers and promoters (Kim et al., 2015; Splinter et al., 2006; Yang et al., 2016). During development, the interaction of the *Sox2* gene with the putative ES-specific enhancer was disrupted if CTCF binding was lost, resulting in decreased expression of SOX2 mRNA (Beagan et al., 2017). Another study demonstrated that CTCF mediates enhancer–promoter interactions which facilitates expression of telomerase reverse transcriptase catalytic subunit (TERT) (Eldholm et al., 2014). Recent study has shown that deletion of the CTCF binding site reduced looping of the enhancer with the the *c-myc* promoter, leading to decreased expression of *MYC* (Schuijers et al., 2018). As CTCF interacts with ncRNAs (Kung et al., 2015; Saldaña-Meyer et al., 2014), eRNA-mediated recruitment of CTCF might facilitate the interaction of enhancers with promoters

On the other hand, CTCF also represses transcription by blocking interaction between enhancers and promoter. For example, binding of CTCF at the imprinted *H19* and *Igf2* genes has been shown to prevent the communication between the enhancer and the *Igf2* promoter, leading to inactivation of *Igf2* (Hark et al., 2000). Another study demonstrated that CTCF at the chicken *β-globin* locus, blocks enhancer activity, thereby prevents spreading of the active chromatin to activate *β-globin* expression (Bell et al., 1999; Splinter et al., 2006). The present study is in accord with an enhancer-blocking role of CTCF. The results demonstrate that CTCF insulated the SPHK1 enhancer from the SPHK1 promoter. Importantly, CTCF occupancy between *SPHK1* enhancer and promoter was reduced upon eRNA-Sphk1 transcription. As the human genome contains thousands of CTCF insulator sites and eRNA regions, transcription of enhancer-derived RNA may represent a common mechanism allowing neighboring genes to be differentially regulated (Bell et al., 1999; Ren et al., 2017a; Xie et al., 2007).

3.6. *KHPS1* associates with the *SPHK1* enhancer by RNA-DNA triplex formation

LncRNA-mediated control of gene expression is achieved by recruiting chromatin modifiers to specific gene sequences. In contrast to transcription factors, many histone-modifying enzymes lack specific DNA-binding domains. Therefore, it has been postulated that lncRNAs might interact with ubiquitously expressed epigenetic remodelers providing the specificity in targeting to genomic sites. Although thousands of lncRNAs have been discovered, mechanistic details of how they act on chromatin remain unknown. A large number of non-coding RNAs has been shown to be associated with chromatin (Li et al., 2017; Nguyen et al., 2017). 25% of ncRNAs-chromatin interactions are not mediated by transcribing Pol II, suggesting that these lncRNAs can utilize different way of binding to DNA (Werner and Ruthenburg, 2015).

One of the mechanisms whereby lncRNA directly binds DNA involves formation of structures known as R-loops (Santos-Pereira and Aguilera, 2015). These RNA–DNA hybrid structures are usually formed during transcription by binding of nascent RNA to the DNA template. Given that a great number of regulatory lncRNAs act *in trans*, it is more likely that lncRNAs employ other mechanism of interaction with chromatin to recruit regulatory proteins. Direct interaction of lncRNAs with specific DNA sequences can be achieved by RNA-DNA triplex formation. Bioinformatical studies revealed genome-wide distribution of potential triplex forming regions across human genome, that can form Hoogsteen basepairing with lncRNAs (Jalali et al., 2017; Soibam, 2017). This *in silico* prediction was supported by studies demonstrated that ncRNAs can interact with genomic DNA by RNA-DNA triplexes. For instance, *trans*-acting lncRNA

HOTAIR as well as some miRNAs were shown to directly interact with DNA (Kalwa et al., 2016; Paugh et al., 2016). Presence of a large number of homopurine stretches in the human genome, which have the potential to form non-canonical Hoogsteen hydrogen bonds (Buske et al., 2012; Goñi et al., 2004) suggests that recruitment of chromatin-modifying enzymes by RNA-DNA triplexes may be a commonly used pathway for lncRNAs to shape the chromatin structure and regulate gene expression.

Until now, very few studies have demonstrated the occurrence and regulatory importance of RNA-DNA triplexes. Therefore, very little is known about triplex-mediated targeting of chromatin modifiers to specific gene sequences. The present study demonstrates that *KHPS1* directly interacts with a purine-rich DNA region located upstream of the transcription start site of the *SPHK1* enhancer forming an RNA-DNA triplex. Combination of in vitro and in vivo triplex capture assays provided experimental evidence for direct interaction of *KHPS1* with the *SPHK1* enhancer via Hoogsteen base-pairing. Importantly, interaction of *KHPS1* with the *SPHK1* enhancer was abolished if the PCR fragment was generated in the presence of 7-deaza-guanosine and 7-deaza-adenosine, which are not able to form Hoogsteen hydrogen bonds. The in vitro triplex capture assay has been validated by monitoring the ability of the reported TFRs, such as of *Fendrr*, *MEG3* and *MEG3* target gene *TGFBR1* (Mondal et al., 2015; Grote et al., 2013). The interaction of *KHPS1* containing the reported triplex forming motifs with the corresponding DNA fragments was lost if the purines in the DNA fragments were modified by 7-deaza-guanosine and 7-deaza-adenosine. The ability of *KHPS1* to recognize and bind to the *SPHK1* enhancer was corroborated by *in vivo* triplex capture experiments. Analysis of the DNA co-precipitated with biotinylated *KHPS1* versions, revealed that ectopic *KHPS1* version comprising the TFR was specifically bound to the *SPHK1* enhancer region, substantiating that *KHPS1* is bound to the *SPHK1* enhancer by formation of RNA-DNA triplexes. Interestingly, enrichment of lncRNA-associated DNA at the boundaries of topologically associated domains (TADs), suggests the interaction of lncRNAs with DNA plays an important role in the topological organization of the genome (Bell et al., 2018).

3.7. RNA-DNA triplex formation is indispensable for eRNA-Sphk1 transcription

Several studies have shown that lncRNAs that are able to form triplexes also engaged in transcription regulation. For example, transcript from the intergenic spacer region of rDNA (pRNAs) binds rDNA promoter and recruit DNMT3b leading to a transcriptional silencing of rDNA (Schmitz et al., 2010). Other examples of lncRNAs that form triplexes are PAPAS which

recruits CHD4/NuRD repressor to rDNA, *Fendrr* which recruits the PRC2 complex to developmental genes (Grote et al., 2013), *MEG3* which guides PRC2 to TGF-β pathway genes (Mondal et al., 2015) *PARTICLE* which regulates the expression of *MAT2A* (O'Leary et al., 2015). The impact of RNAs on transcription of the corresponding target genes has been correlated with their ability to form triplexes. However, none of the studies provided experimental proof that DNA-RNA triplexes are functionally relevant. This study experimentally demonstrated the pivotal functional role of triplexes in transcription activation of eRNA-Sphk1. Reporter-derived *KHPS1* was able to recruit p300 and E2F1 to the *SPHK1* enhancer and activate eRNA-Sphk1 transcription only if a purine stretch engaged in triplex formation remained intact. Deletion or mutation of the TFR region abrogated triplex formation, compromised *KHPS1*-dependent binding of p300 and E2F1 to the *SPHK1* enhancer and inhibited transcription of eRNA-Sphk1. Importantly, deletion or mutation of the *KHPS1* TFR abolished binding of p300 to the *SPHK1* enhancer, whereas interaction of *KHPS1* with p300 was not affected. This result suggests that the association of regulatory proteins with lncRNAs is not sufficient for the protein recruitment to the regulatory region but requires tethering of lncRNA to DNA.

Moreover, the present study demonstrated the requirement of RNA-DNA triplex formation for the transcription of the target genes. CRISPR-mediated monoallelic deletion of the endogenous triplex forming region at the *SPHK1* enhancer was sufficient to significantly reduce transcription of eRNA-Sphk1 and SPHK1 mRNA. Importantly, the levels of *KHPS1* remained unchanged, underscoring that *KHPS1* alone is not sufficient to activate the *SPHK1* enhancer, but needs to be associated with the TFR.

Recently, CRISPR/Cas9 has become a powerful tool for genome engineering in cancer research (Zhan et al., 2018). In addition to several promising results of targeted anti-cancer therapies (Jubair and McMillan, 2017; Maddalo et al., 2014). In this study, CRISPR-Cas9-mediated genome editing of the TFR at *SPHK1* locus turned out to be lethal. Consistently, mono- or biallelic deletion of the TFR severely impaired cell proliferation and viability. MD-MBA-231 cancer cells, lacking the triplex forming region on one of three chromosomes, harbouring the *SPHK1* gene, led to significant reduction in migration and invasion of the cells, presenting potential targeting point to downregulation of *SPHK1* for gene editing to treat cancer.

Triplex-forming oligonucleotides (TFOs) that bind in the major groove of double-stranded DNA and target specific sequences has been utilized as tool for gene regulation (Jain et al., 2008; Wu et al., 2007). For instance, TFO targeted to the promoter of cyclin D1 in HeLa cells inhibited transcription of the cyclin D1 mRNA (Kim and Miller, 1998). Additionally,

transfection of the cells with TFOs that form triplexes at the progesterone responsive elements (PRE) reduced transcription of the PRE-containing reporter genes (Ing et al., 1993). Likewise, a synthetic 27-base-long TFO, termed PU1, has been shown to bind to DNA at the human *c-myc P1* promoter and repress transcription (Postel et al., 1991). In accord with these studies, ectopic RNA comprising the TFR competed with the cellular *KHPS1* and led to reduced recruitment of the regulatory proteins and compromised eRNA-Sphk1 transcription.

Most chemotherapeutic drugs to fight cancer impact rapidly dividing cells and therefore have limited effectiveness for slowly-growing cancers (Lyle and Moore, 2011). Application of triplex forming oligonucleotides (TFOs) could allow for specific targeting of tumors with a population of quiescence cells. Consistent with reports demonstrating inhibition of protooncogenes using TFOs (Carbone et al., 2004; Kim et al., 1998), a TFO targeted to *c-MYC* in mice with human colon cancer xenografts led to inhibition of tumor growth (Boulware et al., 2014). Interference with activation of *SPHK1* enhancer by ectopic RNA that competed for *KHPS1*-dependent triplex formation, led to reduced malignant phenotype of MD-MBA-231 cancer cells by abolishing of migration, invasion and colony formation. The results of this study support the potential of TFOs for therapeutic applications in cancer treatment.

3.8. Triplex forming motifs serve as genomic address code

The intrinsic ability of RNA molecules to recognize and bind specific sequences support the notion that lncRNAs can form modular scaffolds in which different domains interact with DNA, RNA or proteins to form specific functional complexes. A number of studies have demonstrated that lncRNAs are able to simultaneously interact with different regulatory proteins and DNA (Jeon and Lee, 2011; McHugh et al., 2015; Wutz et al., 2002). Some lncRNAs, such as ANRIL or KCNQ1OT1, were shown to interact with different chromatin modifiers at the same time (Korostowski et al., 2012; Kotake et al., 2011; Mohammad et al., 2010; Yap et al., 2010). The sequences of *MEG3* RNA involved in PRC2 interaction and RNA–DNA triplex formation are functionally distinct (Mondal et al., 2015). This work is in accord with these studies, demonstrating that the regions of *KHPS1* engaged in interaction with p300 and E2F1 are distinct, substantiating the modular organization of lncRNAs. Interestingly, the region of *KHPS1* which interacts with E2F1 overlaps with the triplex-forming region. This suggests that E2F1 either interacts with the triplex structure or binds to adjacent sequences of the triplex forming region.

Several studies have proposed a model whereby lncRNAs are molecular address codes for associated regulatory proteins, which can be tethered to specific loci (Batista and Chang, 2013;

Nishikawa and Kinjo, 2017). The present study demonstrated the importance of lncRNA-triplex formation for targeting of regulatory proteins to remote genomic sites. Substitution of the reporter TFR of *SPHK1* with triplex-forming sequences of *Fendrr* (Grote et al., 2013), *MEG3* or its target gene *TGFBR1* (Mondal et al., 2015) functionally replaced the TFR of *SPHK1*. Although the knowledge about the rules underlying regulation of RNA-DNA triplex formation is limited, these swap TFR experiments showed that the triple helical structures do not depend on a defined length of the TFR and the orientation of Hoogsteen basepairing. The UC-rich triplex forming motif of *Fendrr* consists of 41 nucleotides, whereas *MEG3* TFR is AG-rich and contains 20 nucleotides. Two types of Hoogsteen basepairing can be formed based on the sequence of the third strand third. UC motifs form forward Hoogsteen, whereas AG motifs are bound to dsDNA via reverse Hoogsteen (Cheng and Montgomery Pettitt, 1992). The guiding and anchoring function of sequences involved in triplex formation was substantiated by replacement of the TFR of *KHPS1* by the TFR of *MEG3*. Chimeric *KHPS1-MEG3* led to increased occupancy of H3K27ac and E2F1 at endogenous *TGFBR1*, the gene that is regulated by *MEG3*. Together, these results emphasize the importance of triplex formation for lncRNA-mediated targeting of regulatory proteins to remote genomic sites and provide novel insights into how the specific recruitment of lncRNA-associated chromatin modifying enzymes is achieved.

4. Materials and methods

4.1. Materials

4.1.1. Chemicals and reagents

Table 1. Chemicals and reagents used in this study.

Name	Company
[γ-^{32}P]-ATP	PerkinElmer
4-12% Bis-Tris gel	Invitrogen
4-hydroxytamoxifen	Sigma Aldrich
Acrylamide Rotiphorese Gel 30 (37.5:1)	Carl Roth
Agarose	PeqLab
Ammonium Acetate ($NH_4CH_3CO_2$)	US Biologial
Ammonium persulfate (APS)	Carl Roth
Ampicillin	Carl Roth
anti-Flag M2 beads	Sigma Aldrich
ATP Lithium salt	Roche
Biotin-16-UTP	Roche
Bolt transfer Buffer	Invitrogen
Bovine Serum Albumin (BSA)	Sigma Aldrich
Bradford Reagent	Bio-Rad
Chloroform	Carl Roth
Curcumin	Sigma Aldrich
7-deaza-dATP	Tri-Link
7-deaza-dGTP	Sigma-Aldrich
Deoxyribonucleoside-triphosphates (dNTPs)	Roche
Dimethyl-Sulfoxide (DMSO)	Sigma Aldrich
Dithiothreitol (DTT)	Gerbu
DMEM GlutaMAX	Gibco
DNA Ladder, GeneRuler, Ladder mix	Thermo Scientific
Doxycycline	Sigma Aldrich
Dual color broad range protein marker	Bio-Rad
EDTA	Carl Roth

EGTA	Across Organics
Ethanol	Fisher Chemical
Fetal Bovine Serum FBS	Gibco
Flavopiridol	Sigma Aldrich
Formaldehyde	Sigma Aldrich
Glycine	Gerbu
Glycogen	Roche
HEPES	AppliChem
Immobilon Membrane ECL	Millipore
Isopropanol	Fisher Chemical
LB agar	Carl Roth
LiCl	Carl Roth
Lipofectamine 2000	Invitrogen
Lipofectamine 3000	Invitrogen
Lipofectamine RNAiMax	Invitrogen
Magnesium Acetate $Mg(CH_3COO)_2$	Carl Roth
Magnesium Chloride ($MgCl_2$)	AppliChem
Matrigel invasion chamber	CORNING
Methanol	Fisher Chemical
Micro Centrifuge Tubes (1.5 and 2 ml)	Eppendorf
Milk Powder (skimmed)	Sigma Aldrich
MyOne C1 Streptavidin Dynabeads	Invitrogen
Nicotinamide, NAM	Sigma Aldrich
Nitrocellulose membrane	Whatman
N-lauroylsarcosine	Sigma Aldrich
NP-40	MP Biomedicals
NuPAGE LDS loading buffer	Invitrogen
NuPAGE MOPS running buffer Bolt	Invitrogen
Opti-MEM	Gibco
PCR Tubes	Kisker Biotech
Penicillin/ Streptomycin	Lonza
Phase Lock Gel tubes	5PRIME, QuantaBio
Phenol/Chloroform/Isoamylalcohol (25:24:1)	Carl Roth

Phosphatase Inhibitor (PhosSTOP)	Roche
PIPES	Sigma Aldrich
Potassium Acetate (CH_3COOK)	Carl Roth
Potassium Chloride (KCl)	AppliChem
Protein G Dynabeads	Invitrogen
Proteinase K	Thermo Scientific
Puromycin	Sigma Aldrich
QuantiTect SYBR Green PCR kit	Qiagen
Random hexamer primer ($d(N)_6$)	Roche
SKI II	Sigma Aldrich
Sodium acetate (CH_3COONa)	Carl Roth
Sodium Chloride (NaCl)	Fisher Chemical
Sodium Deoxycholate (Na-DOC)	Sigma Aldrich
Sodium Dodecyl Sulfate (SDS)	Gerbu Biotechnik
Sodium hydrogen carbonate ($NaHCO_3$)	Carl Roth
Sucrose	Sigma Aldrich
T4 Polynucleotide Kinase	New England Biolab
TEMED (N,N,N',N'-Tetramethylethylenediamine)	Carl Roth
TRI Reagent	Sigma Aldrich
Triton X-100	Sigma Aldrich
Tween-20	Sigma Aldrich
Urea	Carl Roth
Western Lightning Plus-ECL Enhanced Luminol	Perkin Elmer
Whatman paper	Whatman

4.1.2. Buffers and solutions

Table 2. List of buffers and solutions used in this study.

Name	Composition
Blocking buffer for Western Blot	5% Milk powder 0.2% Tween 20 in 1xPBS
DMEM-GlutaMAX complete medium	500 ml DMEM-GlutaMAX

	50 ml FBS 5 ml Sodium pyruvate 5 ml Pen Strep
Laemmli-buffer (5x)	312.5 mM Tris-HCl, [pH 6.8] 10% SDS 50% Glycerol
LB medium	0.05% Bromophenol blue 1% Tryptone 0.5% yeast 1% NaCl pH adjusted to 7.0
PBS (10 x)	2.7 M NaCl 53.7 mM KCl 20 mM Na_2HPO_4 29.4 mM KH_2PO_4 pH adjusted to 7.4
SDS-PAGE electrophoresis Buffer (10 x)	250 mM Tris base 1.96 M Glycine 1% SDS
TBE (10 x)	1 M Tris base 1 M Boric acid 0.02 M EDTA [pH 8.0]
TE buffer	10 mM Tris-Cl [pH 8.1] 1 mM EDTA [pH 8]
Towbin buffer (10 x)	0.25 M Tris base 1.92 M Glycine
Trypsin/EDTA, [pH 7.7]	137 mM NaCl 5.4 mM KCl 0.7 mM Na2HPO4 25 mM Tris base 5 µM EDTA 0.25% Trypsin 0.5 mM Glucose

4.1.3. Enzymes

Table 3. List of enzymes used in this study.

Name	Company
BamHI	New England Biolab
BbsI	New England Biolab
BspEI	New England Biolab
CsiI	Thermo Scientific
EcoRI	New England Biolab
ExoI	New England Biolab

FastAP Thermosensitive Alkaline Phosphatase	Thermo Scientific
GoTaq DNA Polymerase	Promega
NheI	New England Biolab
PwoSuperYield DNA polymerase Kit	Roche
RNase A	Promega
RNase I	Thermo Scientific
SapI	New England Biolab
ScaI	New England Biolab
T4 Polynucleotide Kinase	New England Biolab
Transcriptor First Strand cDNA Synthesis Kit	Roche
TURBO DNase I	Ambion
XhoI	New England Biolab

4.1.4. Antibodies

Table 4. List of primary and secondary antibodies used in this study.

Name	Type	Company	Application
Anti-actin, clone AC-40	Mouse monoclonal	Sigma Aldrich	Western Blot
Anti-CTCF	Rabbit polyclonal	Active Motif, 61311	ChIP
Anti-E2F1	Mouse monoclonal	Santa Cruz, sc-251	ChIP, RIP
Anti-E2F1	Rabbit monoclonal	Abcam, ab179445	CLIP
Anti-Flag (M2)	Mouse monoclonal	Sigma-Aldrich	RIP
Anti-H3-pan	Rabbit polyclonal	Diagenode, C15310135	ChIP
Anti-H3K27Ac	Rabbit polyclonal	Abcam, ab4729	ChIP
Anti-H3K27me3	Rabbit polyclonal	Abcam, ab195477	ChIP
Anti-H3K4me1	Rabbit polyclonal	Millipore, 07-436	ChIP
Anti-H3K4me3	Rabbit polyclonal	Millipore, 07-473	ChIP
Anti-H3K9Ac	Rabbit polyclonal	Millipore, 07-352	ChIP
Anti-HA	Mouse polyclonal	E. Kremmer	RIP
anti-Mouse-HRP	Goat	Dianova	Western Blot
anti-myc	Mouse polyclonal	E. Kremmer	RIP
Anti-p300	Mouse monoclonal	Abcam, ab14984	ChIP, CLIP
Anti-Pol II	Rabbit polyclonal	Santa Cruz, N-20 sc-899	ChIP

Anti-pSer5-Pol II	Mouse monoclonal	Abcam, ab5408	ChIP
anti-Rabbit-HRP	Goat	Dianova	Western Blot
Anti-SPHK1	Rabbit monoclonal	Cell Signaling, 12071	Western Blot
Anti-UBF	Mouse polyclonal	Sigma Aldrich	WB

4.1.5. Generated plasmids

Table 5. List of plasmids that were generated in this study.

Generated plasmid	Recombinant product
pTet-*KHPS1*/+1448/-592	*KHPS1*(+1448/-592)
pTet-*KHPS1*/+1448/-592_mutTFR	*KHPS1*-mutTFR (+1448/-592)
pTet-*KHPS1*/+1448/-592_ΔTFR	*KHPS1*-ΔTFR (+1448/-592)
pTet-*KHPS1*/+1448/-592_*MEG3*_TFR	*KHPS1*- *MEG3* TFR (+1448/-592)
pTet-*KHPS1*/+1448/-592_*TGFBR1*_TFR	*KHPS1*- *TGFBR1* TFR (+1448/-592)
pTet-*KHPS1*/+1448/-592_*Fendrr*_TFR	*KHPS1*- *Fendrr* TFR (+1448/-592)
pTet-*KHPS1*/+1448/-592_U2_TFR	*KHPS1*- U2 TFR (+1448/-592)
pTet-*KHPS1*/+1448/-592_luc_TFR	*KHPS1*- luc TFR (+1448/-592)
pTet-*KHPS1*/+1448/-592-(isoB-luc)	*KHPS1* (+1448/-592) and eRNA-Sphk1-Luc mRNA
pTet-*KHPS1*/+1448/-592Δ+25/+487	*KHPS1* (+1448/-592Δ+25/+487)
pREP4-*SPHK1*(-592/+1795)-luc	*SPHK1*(-592/+1795)-driven firefly luciferase
pREP4-*SPHK1*(+46/+1795)-luc	*SPHK1*(+46/+1795)-driven firefly luciferase
pREP4-*SPHK1*(-592/+1795polyA)-luc	*SPHK1*(-592/+1795polyA)- driven firefly luciferase
lentiGuide-puro_sgRNA	Control single guide RNA for dCas9-KRAB
sgRNA(MS2)_*KHPS1-isoC*_sgRNA#1	single guide RNA #1 targeting dCas9-VP64 to the *KHPS1-SPHK1-C* promoter
sgRNA(MS2)_*KHPS1-isoC*_sgRNA#2	single guide RNA #2 targeting dCas9-VP64 to the *KHPS1-SPHK1-C* promoter
sgRNA(MS2)_*eSPHK1*_sgRNA#1	single guide RNA #1 targeting dCas9-VP64 to the *SPHK1* enhancer

sgRNA(MS2)_*eSPHK1*_sgRNA#2	single guide RNA #2 targeting dCas9-VP64 to the *SPHK1* enhancer
lentiGuide-puro _*eSPHK1*_sgRNA#1	Control single guide RNA#1 targeting dCas9-KRAB to the SPHK1 enhancer
lentiGuide-puro _*eSPHK1*_sgRNA#2	Control single guide RNA#2 targeting dCas9-KRAB to the SPHK1 enhancer
lentiCRISPR v2_ TFR_sgRNA#1	Control single guide RNA#1 targeting Cas9 to the *eSPHK1* TFR
lentiCRISPR v2_ TFR_sgRNA#2	Control single guide RNA#2 targeting Cas9 to the *eSPHK1* TFR
lentiCRISPR v2_ TFR_sgRNA#3	Control single guide RNA#3 targeting Cas9 to the *eSPHK1* TFR
lentiCRISPR v2_ TFR_sgRNA#4	Control single guide RNA#3 targeting Cas9 to the *eSPHK1* TFR
lentiCRISPR v2_CTCF_sgRNA#1	Control single guide RNA#1 targeting Cas9 to the *SPHK1* CTCF binding site
lentiCRISPR v2_CTCF_sgRNA#2	Control single guide RNA#1 targeting Cas9 to the *SPHK1* CTCF binding site

4.1.6. External plasmids

Table 6. List of obtained plasmids that were used in this study.

Plasmid	Source
dCas9-KRAB	Addgene #50917
dCas9-VP64_GFP	Addgene #61422
Flag-PCAF	From Y. Nakatani
HA-CBP	From M. Ott
HA-p300	From Y. Nakatani
lentiGuide-Puro	Addgene #52963
lentiCRISPR v2	Addgene #52961
MS2-P65-HSF1_GFP	Addgene #61423
Myc-Tip60	Thomas Hoffmann (DKFZ).
pBS (pBluescript)	Addgene, 212205
pGl4.10 [*luc2*]	Promega

pREP4	Invitrogen
pTet-7B-MS2bs (p2255)	from G. Stoecklin
pTKLP-tetA	Addgene #71325
pCR2.1-TOPO	Invitrogen
sgRNA(MS2)	Addgene #61424
TK-*Renilla*	Promega

4.1.7. Sequences of ASO and siRNAs

Table 7. Sequences of ASO and siRNA used for knock-down experiments
(indicated positions are relative to the TSS of *eSPHK1*)

Name	**(Sequence 5' to 3')**
ASO *KHPS1* #1 (-101/-121)	mU*mU*mU*mG*G*T*G*G*A*A*A*T*G*C*T*C*T*mC*mC*mG*mG
siRNA KHPS1	GGAGAGCAUUUCCACCAAATT
KHPS1 #2 (+174/+196)	mC*mC*mU*mC*G*T*T*C*C*T*G*T*T*T*C*T*mC*mG*mG*mA
isoB(eRNA-Sphk1)#1 (+355/+375)	mC*mC*mC*mC*G*G*G*C*G*G*G*T*C*G*G*G*mG*mG*mG*mC
isoB(eRNA-Sphk1)#2 (+388/+408)	mC*mC*mC*mA*A*C*A*C*T*T*G*G*G*G*G*mU*mC*mA*mG
Ctrl ASO	mU*mC*mA*mC*mC*T*T*C*A*C*C*C*T*C*T*mC*mC*mA*mC*mU
siOff target	siGENOME control pool (D-001206-14-20) Sequence is property od Dharmacon
siRNA p300	ON-TARGETplus HUMAN siRNA (Dharmacon) Sequence is a property od Dharmacon
siRNA E2F1	ON-TARGETplus HUMAN siRNA (Dharmacon) Sequence is a property od Dharmacon
siRNA PCAF	ON-TARGETplus HUMAN siRNA (Dharmacon) Sequence is a property od Dharmacon

4.1.8. Sequences of sgRNAs

Table 8. Sequences of sgRNAs used for CRISPR system
(indicated positions are relative to the TSS of *eSPHK1*)

Name	Sequence (5' to 3')
Ctrl sgRNA 1	CACCGAGCAGGCTCTGGCAACGCGCTGG
KHPS1-isoC promoter sgRNA #1 (+1563/+1583)	CACCGGCCGAGGCCGCGCCGGGCTC
KHPS1-isoC promoter sgRNA #2 (+1598/+1618)	CACCGGCGGAGCCAGGCCGGCGCCG
eSPHK1 sgRNA #1 (-20/-1)	GGGAGGAGGGGGCTCCGCGC
eSPHK1 sgRNA #2 (20/-39)	GCCTAGAAAACGCGAATCGG
SPHK1 TFR sgRNA 1 (-320/-300)	TCGTGGTGATAAAGCCCACCTGG
SPHK1 TFR sgRNA 2 (-303/-283)	AAAAGGGGGGACCCTGAACCAGG
SPHK1 TFR sgRNA 3 (-358/-378)	AGGGGGGACCCTGAACCAGGTGG
SPHK1 TFR sgRNA 4 (-388/-408)	GGGGGGACCCTGAACCAGGTGGG
SPHK1 CTCF sgRNA 1 (+694/+714)	CACCGCTAGGGCAGAGCGCGCAAAA
SPHK1 CTCF sgRNA 2 (+1189/+1209)	ACCGGATCTAACTCGAGGTGCTCG

4.1.9. Technical devices

Table 9. List of instruments and machines used in this study.

Name	Supplier
Agarose Gel Electrophoresis Chamber	CoreLifeSciences
Bioruptor, Pico	Diagenode
Centrifuge 5415R	Eppendorf
Centrifuge 5424	Eppendorf

CO_2 Incubator	Sanyo
Confocal microscopes	Leica
LAS-3000	GE Healthcare
Molecular Imager Gel-Doc XR+	Bio-Rad
Nano Drop 2000 Spectrophotometer	Thermo Scientific
NuPAGE chumber XCell SureLock	Invitrogen
PAGE Chamber	H.Holzel
pH Meter	Mettler Toledo SevenCompact
PhosphorImager	Fujifilm
Power Supply	Bio-Rad
Real time PCR system, LightCycler 480 II	Roche
Scintillation counter	Beckman coulter
96-Well Thermo MasterCycler	Eppendorf
Thermo Mixer Compact	Eppendorf
Trans-Blot SD semi-dry Transfer Cell	Bio-Rad
Wet Tank Mini Trans-Blot Cell	Bio-Rad
UV Crosslinker, Stratalinker	Stratagene
Zeiss cell Observer	Zeiss
Gel Dryer	Bio-rad

4.1.10. Kits

Table 10. Kits that were used in this study.

Name	Company
Gel extraction Kit	Qiagen
DNA ligation Kit	New England Biolab
Nucleotide removal Kit	Qiagen
PCR Purification Kit	Qiagen
Plasmid Miniprep Kit	Qiagen
Plasmid Maxiprep Kit	Qiagen
RNA Clean & Concentrator Kit	Zymo Research
MEGAscript T7 Transcription Kit	Ambion

4.1.11. Primers

Table 11. List of DNA oligonucleotides used for strand-specific RT, qPCR and PCR

Oligonucleotide	Sequence (5' to 3')
SPHK1 -1445F	TTGAGATGGAGTTTCGCCCT
SPHK1 -1300R	TCAAACCCCGTCTCTACACC
SPHK1 -698F	GGGAAAGGGGTCTGGAGAAA
SPHK1 -592F	AGGACAGAGGACCCATTGTG
SPHK1 -464R	CACGGGCTCTTTATCCAACG
SPHK1 -425R	GGGGTTGACTGCAAACTGAG
SPHK1 -406F	TCATTAGCCATAGAGTCTGAAGAGG
SPHK1 -373F	CAGGATCCGACCTTTCTGG
SPHK1 -241R	ACTCAAGGCTGGTGGTAGTGG
SPHK1 -241RT7	ACTCAAGGCTGGTGGTAGTGGGAAATTAATAC GACTCACTATAGG
SPHK1 -236F	GCTTTTCAGATTTGGTGGAAAATGCTCTCCG
SPHK1 -137F	CGGGGAGCACAGCCTCCGATTC
SPHK1 -89R	GAGGGACTTTGGTGCCTAGA
SPHK1 -65R	TTCTTCCCCTCCTCTTCCTC
SPHK1 -40F	CGGACGGATCTGGTCCTGG
SPHK1 +7R	CGGCTGTGGGAGGAGGGGGCTC
SPHK1 +25R	CTGCGGGGACGCGAG
SPHK1 +46F	GGATTCCTGGAGCAAGGGG
SPHK1 +108F	GGTGCAGGACCCATCATT
SPHK1 +208R	GGTCGAAAAACTCCGAGAAA
SPHK1 +596R	AGCCTCCCGCAGGGGCTTGGT
SPHK1 +694R	CTCGAGTTAGATCCCTGGGG
SPHK1 +841R	GCGTGAAGAAGGGGCGCCGGA
SPHK1 +1189F	CGACAGCACCCGCGGGGGCG
SPHK1 +1448R	CGTCCTAGGCGTCTGGCTA
SPHK1 +1678R	TTTTCGCTCAACTTCGCAGC
SPHK1 +1795R	CGTTCCCTACAGTGGCCTG
SPHK1 +2291F	GCCCTTCTCAGGGATTGTAGG

SPHK1 +3136F	AATCTCCTTCACGCTGATGC
SPHK1 +3235R	GACCGGCAGCTATCAGGAC
SPHK1 +3321R	TCCAGACATGACCACCAGAG
isoB RNA (eRNA-Sphk1) F	GGACCTGCCTCTTCTCGACT
isoB RNA (eRNA-Sphk1) R	CTGCCTTCAGCTCCTTATCG
SPHK1 mRNA F	TCCGCTCAAGTTCTGGGATT
SPHK1 mRNA R	GCCGTGTGACTAAGCCTACA
GAPDH mRNA F	CATGTTCCAATATGATTCCACC
GAPDH mRNA R	GATGGGATTTCCATTGATGAC
18S rRNA F	GCGACCTCAGATCAGACGTGG
18S rRNA R	CTGTTCACTCGCCGTTACTGAG
CDC2 F	CGCCCTTTCCTCTTTCTTTC
CDC2 R	ATCGGGTAGCCCGTAGACTT
mCDC2 F	ACTCCAGGCTGTATCTCATC
mCDC2 R	CAAGTCTCTGTGAAGAACTCG
MEG3 F	CCCTTCTGCGCCTCCATATA
MEG3 R	GATGCCGTCTTCCTTTTGCA
TGFBR1 F	TCCAGGCTTCCTCAAATCGT
TGFBR1 R	GCCTCCTTCTCTTTTCCGGT
RPLP2 F	CACCAAGGAGTCAAGGCGAG
RPLP2 R	CTCAACCTTTGCCAGCGAAC

4.1.12. DNA oligonucleotides used for cloning

Table 12. Sequences of oligonucleotides used for reporter plasmids.
Restriction sites used for cloning are marked in bold.

Name	Sequence (5' to 3')
Mutant TFR	***GGATCC***GACCTTTCTGGTGCGCCAGCACGACATCAA CAGCTGCGACCCTGA***ACCAGGT***
ΔTFR	***GGATCC***GACCTTTCTGGGGCCCCCCCTGA***ACCAGGT***
MEG3	***GGATCC***GACCTTTCTGGGGCCCCCGCTCCCTCTCTGC TCTCCGCCCTGA***ACCAGGT***
TGFBR1	***GGATCC***GACCTTTCTGGGGCCCCAGAGAGAGGGAG AGAGCCCTGA***ACCAGGT***

Fendrr	***GGATCC***GACCTTTCTGGGGCCCCAAGAGGAGGAGG AGAAGGAAGAGGATGGAGGGGAGGGAGAACCCTG A***ACCAGGT***
Scr U2	***GGATCC***GACCTTTCTGGGGCCCCAATCCATTATAAT ATATTGTCCTCGGATAGACCCTGA***ACCAGGT***
Scr luc	***GGATCC***GACCTTTCTGGGGCCCCTATCACGCTCGTC GTTCGGTATGCCCTGA***ACCAGGT***
PolyA signal	TAATTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCA GACATGATAAGATACATTGATGAGTTTGGACAAACC ACAACTAGAATGCAGTGAAAAAAATGCTTTATTT

4.2. Cell culture and related techniques

4.2.1. Cell culture and treatments

NIH3T3 Tet-ON, U2OS/ER-E2F1, MDA-MB-231 and HeLa cells were cultured in complete Dulbecco's modified Eagle's Medium (DMEM). Cells were kept at 37°C in a humidified atmosphere of 5% CO_2 in air. For activation of ER-tagged E2F1**,** U2OS/ER-E2F1 cells were treated with 100 nM 4-hydroxytamoxifen (4-OHT). To inhibit RNA polymerase II transcription cells were treated with flavopiridol for 1-3 h or 10 μM. To inhibit SPHK1 activity cells were treated with 10 μM of SKI II. To activate tetracycline-inducible gene expression, NIH3T3 Tet-ON cells were incubated with 5 μg/ml doxycycline for 16 h. To inhibit p300/CBP activity cells were treated with 10mM curcumin for 3 h.

4.2.2. Freezing and thawing of mammalian cells

To store generated cell clones in liquid nitrogen, $10x10^6$ cells were detached using trypsin-EDTA and mixed with 10 ml complete DMEM medium. Cells were collected by centrifugation (5 min, 800 rpm), the supernatant was removed, and the cells pellet was resuspended in 2 ml DMEM containing 20% FBS and incubated 15 min on ice. Then 2 ml of a solution, containing DMEM medium supplemented with 20% FBS and 10% DMSO, were slowly added to the cells, gently mixed and incubated on ice for 15 min. Aliquots of 1 ml were transferred into Cryotubes. For gradual freezing, vials with the cells were incubated for 24 h in methanol -80°C and transferred into liquid nitrogen.

4.2.3. Cell transfections with reporter plasmids

For reporter assays, 3×10^4 NIH3T3 Tet-ON cells in 12-well dish were co-transfected with 2 ng of pTet-*KHPS1*/+1448/-592 and 28 ng pBS using Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. 24 h after transfection, cells were replated on a 6-well dish, and after 6 h cells were induced with 5 µg/ml doxycycline for 16 h.

For luciferase reporter plasmids U2OS/ER-E2F1 or NIH3T3 TetON cells were seeded in 12-well plates and co-transfected with 50 ng of the respective firefly luciferase reporter plasmids, 5 ng of the TK-*Renilla* plasmid and 45 ng pBS using Lipofectamine 3000 (Life Technologies) according to the manufacturer's instructions.

4.2.4. Cell transfections with synthetic RNAs

U2OS/ER-E2F1, HeLa or MD-MBA-231 cells were seeded in 12 well plate and transfected with 100ng synthetic *KHPS1* versions using Lipofectamine™ RNAiMAX (Life Technologies). Cells were harvested 24 h after transfection.

4.2.5. Cell transfection with siRNA and ASO

To knockdown *KHPS1*, eRNA-Sphk1, p300 or E2F1, U2OS/ER-E2F1 cells were reverse-transfected twice with 20 µM ASO or 20nM siRNA using Lipofectamine™ RNAiMAX (Life Technologies). Cells were harvested 72 h after transfection and proceeded for RNA analysis. siRNA and ASO sequences are listed in **Table 7.**

4.2.6. Activation and inhibition of transcription by CRISPRi/CRISPRa

For CRISPRi, 8×10^4 U2OS/ER-E2F1 cells were transfected with 100 ng of dCas9-KRAB and 50 ng of each sgRNA#1 and sgRNA#2 expressed from lentiGuide-Puro plasmids that target *SPHK1* enhancer. Cells were harvested 24 h after transfection.

For CRISPRa, U2OS/ER-E2F1 cells were transfected with 50 ng of dCas9-VP64_GFP, 50 ng of MS2-P65-HSF1_GFP and 50 ng of sgRNA(MS2) expressing sgRNA#1 and sgRNA#2 which target *SPHK1* enhancer or *KHPS1* promoter. Cells were harvested 24 h after transfection. Sequences of sgRNAs are listed in **Table 8**.

4.2.7. CRISPR/Cas9-mediated generation of cell lines lacking the TFR or the CTCF binding site

sgRNAs targeting either the TFR or the CTCF binding sites at *SPHK1* (**Table 8)** were

cloned into lentiCRISPR v2. Plasmids were transfected into MDA-MB-231 or U2OS/ER-E2F1 and selected for 72 h using standard medium containing 0.2-0.5 µg/ml of puromycin. Single clones were isolated by serial dilutions. After clonal expansion, PCR-based screening was performed using primers -592F/+25R and +108F/+1448R. Deletion of the TFR and the putative CTCF binding sites was confirmed by sequencing.

4.2.8. Luciferase assay

U2OS/ER-E2F1 or NIH3T3 TetON cells seeded in 12-well plates were transfected with 50 ng of the respective firefly luciferase reporter plasmids and 5 ng of the TK-*Renilla* plasmid (Promega). Cells were lysed in 1x Passive Lysis Buffer (Promega) 24 h after transfection, and luciferase activity was measured by monitoring light emission with the in vivo imaging system of IVIS Lumina II instrument (PerkinElmer). Firefly luciferase activity was normalized to *Renilla* luciferase activity and presented in reference to expression of the control reporter vector

4.3. Generation of plasmids

4.3.1. Construct preparation for reporters

To generate a plasmid that expresses both *KHPS1* and eRNA-Sphk1 transcription, a PCR fragment comprising *SPHK1* sequences from −592 to +1448 (relative to *SPHK1* enhancer TSS) was inserted into pCR2.1-TOPO. Using BglII and XhoI restriction enzymes, the *SPHK1* fragment from pCR2.1-TOPO was cloned into pTet-7B-MS2bs to yield pTet-*KHPS1*/+1448/−592.

To generate reporter constructs containing foreign TFR sequences, oligonucleotides comprising TFR sequences were annealed and inserted between the *BamHI* and *CsiI* sites of pTet-*KHPS1*/+1448/-592. Oligonucleotide sequences are listed in **Table 12**.

To generate pREP4-luc, the luciferase gene from pGL4.10 was inserted into the *NheI-BamHI* site of the episomal vector pREP4. Luciferase reporter constructs were cloned by inserting PCR fragments comprising *SPHK1* sequences -592/+1795, +46/+1795 into pREP4-luc. For insertion of a polyA signal, *SPHK1* fragment comprising sequences from -247 to +439 and inserted polyA signal at the position +20 (relative to the *SPHK1* enhancer TSS) (**Table 12**) was inserted into pREP-(SPHK1+1795/-592)-luc using BglII and BspEI restriction enzymes.

To generate pTet-*KHPS1*/+1448/-592-(isoB-luc), a PCR fragment comprising nucleotides -592/+1448 was inserted into pTet-7B-MS2bs-luc. To produce pTet-7B-MS2bs-luc, the luciferase gene from pGL4.10 (Promega) was cloned into the *XhoI* site of pTet-7B-MS2bs. Primer sequences are listed in **Table 11.**

Generation of CRISPR plasmids was performed as described in Cong et al., 2013. Sequences of sgRNAs are listed in the **Table 8**.

4.3.2. Preparation of oligonucleotides for cloning

Oligonucleotides comprising sequences of the TFRs and sgRNAs and synthesized by Sigma Aldrich were annealed forming sticky end on both 5' and 3' ends (compatible with *BsmBI*-cut ends for sgRNAs, and *BamHI* and *CsiI* sites for TFR). Duplex formation of two complement oligonucleotides was performed by mixing 100 µM of each oligo with 20mM Tris-Acetate [pH7.5], followed by incubation using PCR program: 95 °C to 20 °C with 5 min incubation every 5 °C with 1% ramp. For 5' phosphorylation, annealed oligos were incubated with 5 U T4 polynucleotide Kinase (PNK), 80 mM ATP at 37 °C for 30 min. Sequences of oligonucleotides are presented in **Table 8 and 12**.

4.3.3. Transformation of bacteria

Competent Stbl *E.coli* and DH5α *E.coli* were used for *Cas9*-containing plasmids and for the rest of plasmids, respectively. Cells were incubated with plasmid DNA on ice for 15 min. After heatshock at 42°C for 90 sec and cooling on ice for 2 min, the cells were mixed with 1 ml LD medium and incubated for 1 h at 37°C with shaking. The transformed bacteria were plated on LB agar plates containing appropriate antibiotics.

4.4. DNA-related techniques

4.4.1. Isolation of genomic DNA from mammalian cells

Cells were pelleted in PBS and lysed in 500 µl of DNA-extraction buffer (10 mM Tris-HCl [pH 8], 0.5 mM EDTA, 0.5% SDS) and digested with 5 µl RNaseA (10mg/ml) for 1 h at 37°C. After Proteinase K was added, samples were incubated for 15 min at 56°C, followed by sonication using Pico Bioruptor (3 cycle, 15 sec on, 15 sec off). Then equal volume of phenol-chloroform was added, and phase separation was achieved by

centrifugation for 5 min at 8,000 rpm. The upper aqueous phase was collected and washed 1 time with an equal volume of phenol-chloroforms and 2 time with chloroform. DNA was precipitated from the aqueous phase using 1 µl glycogen, 1/10 7.5M ammonium acetate and equal volume of 100% ethanol and incubation at -20°C for 10 min. DNA was pelleted by centrifugation for 20 min at 13,200 rpm and washed once with 75% Ethanol, followed by solving in 50 µl TE buffer and incubation for 5 min at 65°C.

4.4.2. Small scale plasmid preparation

3 ml of *E.coli* (either Stbl or DH5α) cells were grown overnight at 37°C in LB medium containing appropriate antibiotic. Bacteria were centrifuged for 3 min at 4,000 rpm and DNA was isolated using 'plasmid Mini Kit'.

4.4.3. Large scale plasmid preparation

400-800 ml of *E.coli* (either Stbl or DH5α) cells were grown overnight at 37°C in LB medium containing appropriate antibiotic. Bacteria were centrifuged for 30 min at 4,000 rpm and DNA was isolated using 'plasmid Maxi Kit'.

4.4.4. Polymerase chain reaction (PCR)

PCR was used for:

1. Testing and optimizing primer pairs.
2. Generation of T7-conjugated DNA fragments as templates for *in vitro* transcription.
3. Generation of DNA fragments for *in vitro* triplex studies.
4. Validation of CRISPR/Cas9-mediated genomic deletions.

If not stated otherwise, PCR reactions were performed using 1x PCR master mix GoTaq Green, 500nM of forward and reverse primers and 10 – 30 ng of genomic DNA as a template. PCR was purified using PCR Purification Kit. PCR products were analyzed by agarose gel electrophoresis in 1xTBE buffer and visualized in UV light using Gel Doc XR System. Primers are listed in **Table 11.**

4.4.5. PCR with modified 7-deaza-2-deoxy-nucleotide-5'-triphosphates

7-deaza-2-deoxy-guanosine-5'-triphosphates (7-deaza-dGTP) or 7-deaza-2- deoxy-adenosine-5'-triphosphates (7-deaza-dATP) was used for incorporation into PCR products. PCR reactions were carried out using the PwoSuperYield DNA polymerase Kit. The final volume of reaction

was 50 µl, including 2.5 U of PwoPolymerase, 10 mM of each dNTPs and 14 pmol of each primer, using 15 ng of genomic DNA as a template. PCR products were purified using PCR Purification Kit according to the manufacturer's instructions. Primers are listed in **Table 11.**

4.4.6. Real-time quantitative PCR

Quantitative PCR (qPCR) was performed on gDNA or cDNA using 5 pmol of each primer and either LightCycler 480 SYBR Green 1 (Roche) or QuantiTect SYBR Green (Qiagen) 2x master mix in the total volume of 10 µl. For normalization of cDNA samples, 18S rRNA levels were measured. Primers are listed in **Table 11.** Measurement was performed under the following conditions using LightCycler 480:

Table 12. LightCycler programming.

Program	Step	Temperature	Time	Fluorescence acquisition
	Pre-incubation	95°C	10 min	-
Amplification 40 cycles	Denaturation	95°C	15 sec	-
	Annealing	60°C	15 sec	-
	Elongation	72°C	30 sec	Single
Melting	Denaturation	95°C		Continuous
	Cooling	40°C		-

4.5. RNA-related techniques

4.5.1. Preparation of total cellular RNA

To isolate RNA form cultured cells, medium was removed, and cells were lysed in 1 ml of TRI reagent on the plate. RNA was isolated according to the manufacturer's protocol and co-precipitated with 10- 20 µg glycogen. Isolated RNA pellet was recovered in 15-30 µl of H_20 and solved at 55°C for 15 min. RNA concentration and purity were determined using Nanodrop spectrophotometer.

4.5.2. Isolation of nuclear RNA

To isolate cellular nuclei, $7x10^6$ U2OS/ER-E2F1 cells were pelleted in 1x PBS (4°C) and resuspended in 300 µl of Nuclei isolation buffer (10mM KCl, 10mM HEPES [pH 8.0], 5mM $MgCl_2$, supplemented with 15 U RNasin). After incubation at 4°C for 5 min, 0.2%

NP40 and 0,5 mM DTT were added. Cellular lysis was monitored using bright-field microscope. Nuclei were centrifuged at 4°C, 1.3 x g for 3 min, cytoplasmic fraction was transferred in a new Eppendorf tube, and nuclei were washed 2 times in Nuclei isolation buffer without detergents. Then TRI reagent was added to the nuclear pellet.

4.5.3. Reverse transcription (RT)

For the reverse transcription (RT) using Random hexamer primers, 1 µg RNA was first treated with TURBO DNase I to remove the DNA and incubated with 2.5 µM random hexamer primers (dN6), 1 mM dNTPs, 0,2U Transcriptor Reverse Transcriptase and 1x Transcriptor RT Reaction Buffer and 10 Units of RNasin was added into the tube. The reaction was performed in PCR cycler using following programme: 10 min at 25 °C (annealing), 30 min at 70 °C (reverse transcription) and 5 min at 85 °C (heat inactivation of reverse transcriptase).

For primer-specific reverse transcription, 1 µg TURBO-DNase-treated RNA was incubated with 1 µM primer, RNasin and incubated at 65°C for 10 min. After cooling down the reaction, a mix of 1 mM dNTPs, 0,2 U Transcriptor Reverse Transcriptase and 1x Transcriptor RT Reaction Buffer was added, and Reverse transcription was performed in PCR cycler for 30 min at 70 °C and for 5 min at 85 °C. Primers are listed in **Table 11**.

4.5.4. In vitro transcription

To generate antisense RNA transcript, DNA templates for *in vitro* transcription were prepared by PCR on genomic DNA using reverse primer containing the T7 Polymerase recognition sequence (TAATACGACTCACTATAGGGAGG) and forward primer. PCR products were purified with PCR purification Kit according to the manufactures protocol. T7 transcription reaction was performed using MEGAscript T7 Transcription Kit with 30ng of purified DNA template. For biotinylated RNA 20% of biotin-16-UTP (Roche) and 80% UTP were used. *In vitro* transcription reaction mix was prepared according to the manufacturer's instructions. After incubation for 12 h at 37°C, the RNA was digested with TURBO DNase for 15 minutes at 37°C. RNA was purified by RNA Clean & Concentrator Kit. Amount of RNA were determined with Nanodrop spectrophotometer and the quality was assessed by agarose gel electrophoresis in 1x MOPS. Primers containing T7 sequence are in **Table 11**.

4.6. Protein related techniques

4.6.1. Chromatin isolation

Small scale chromatin isolation was performed according to the Wysocka et al., 2001. Briefly, $7x10^6$ U2OS/ER-E2F1 cells lysed in 200 µl of Buffer A (10 mM HEPES [pH 8.0], 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 mM sucrose, 10% glycerol, 1 mM DTT, 1 x Protease Inhibitor cocktail). After addition of 0.1% TritonX-100, cells were incubated at 4°C for 5 min. Cytoplasmic proteins were removed by centrifugation at 20,000 x g for 5 min at 4°C and pellet was washed in Buffer A. The nuclear pellet was resuspended in 100 µl Buffer B (3 mM EDTA, 0.2 mM EGTA, 1mM DTT, 1 x Protease inhibitory cocktail) and incubated for 30 min at 4°C. After centrifugation at 1,700 x g for 5 min at 4°C, 1x SDS sample buffer was added to the pellet followed by boiling for 10 min at 90°C.

4.6.2. Lysate preparation for western blot

Cells were scraped from the plates and washed once with 1 x PBS (4°C). Cell pellet was lysed by AM-300 buffer (10% Glycerol, 0.2 mM EDTA [pH 8.0], 5 mM $MgCl_2$, 20 mM Tris-HCl [pH 8.0], 300 mM KCl supplemented with proteinase inhibitors), followed by incubation for 10 min at 4°C. After sonication using Pico Biorupter (20 cycles, 30 sec on, 30 sec off), the cell debris were removed by centrifugation at 13,200 rpm 10 min at 4°C.

4.6.3. SDS-PAGE

Cell lysates or IP eluates were added to 5x Laemmli buffer to a final concentration of 1x Laemmli buffer. Samples were heated at 95 °C for 10 min, shortly centrifuged and loaded onto SDS-polyacrylamide gels. The stacking gel was run at 100 V then voltage was increased to 150 V.

4.6.4. Western blotting

For western blotting analysis, proteins from the gel were transferred onto a nitrocellulose membrane by a semi-dry transfer for approximately 1 h at 14 V using 1x Towbin buffer supplemented with 10% methanol (**Table 2**). After transfer, the nitrocellulose membrane was blocked for 1 h at room temperature in blocking buffer (**Table 2**). Then the membrane was then incubated with primary antibodies at appropriate dilutions overnight at 4 °C. After 3 times washing with PBST buffer (0.2% Tween-20 in 1 x PBS), membrane incubated with secondary antibodies (1:10,000 dilution) for 1 h at room temperature. After incubation with secondary

antibodies, the membrane was washed 3 time for 5 min in PBST and 1 time in 1x PBS. Proteins were visualized using ECL solution and developed in LAS 3000 machine. Images were analyzed using ImageGauge software.

4.7. Methods to study DNA-protein interactions

4.7.1. Chromatin immunoprecipitation (ChIP)

Cells were crosslinked with 1% formaldehyde for 10 min, quenched with 125 mM glycine. After harvesting, cell pellets were incubated in buffer A (100 mM Tris-HCl [pH 8.1], 10 mM DTT), buffer B (10 mM HEPES [pH 8.0], 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) and buffer C (10 mM HEPES [pH 8.0], 10 mM EDTA, 0.5 mM EGTA, 200 mM NaCl). Isolated chromatin was sonicated in a Bioruptor Pico (Diagenode) in buffer containing 1% SDS, 50 mM HEPES [pH 8.0], 10 mM EDTA to obtain an average fragment length of 200-500 bp. For ChIPs on reporter plasmids, the chromatin pellets were digested with a cocktail of restriction enzymes (*Xho*I, *Sca*I, *BspE*I, *Sap*I) prior to sonication. Upon dilution with 4 volumes of IP-buffer (20 mM HEPES [pH 8.0], 187.5 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100), chromatin was incubated with antibody-coupled Protein G-coated Dynabeads overnight at 4°C. Protein-DNA complexes were washed twice in buffer A (150 mM NaCl, 20 mM HEPES [pH 8.0], 0.5 mM EGTA, 1 mM EDTA, 0.1% SDS, 0.1% Na-deoxycholate, 1% Triton X-100), followed by two washes with buffer B containing 500 mM NaCl, with buffer C (250 mM LiCl, 0.5% Na-deoxycholate, 0.5% NP-40, 20 mM HEPES [pH 8.0], 0.5 mM EGTA, 1 mM EDTA), and with buffer D (20 mM HEPES [pH 8.0], 0.5 mM EGTA, 1 mM EDTA). After elution, reversal of the cross-link (65°C, 6 h) and digestion with proteinase K, DNA was purified and quantified by qPCR using gene-specific primers. Primer sequences are listed in **Table 11.** The ratio of DNA in the immunoprecipitants (upon subtraction of the IgG background) *versus* DNA in the input chromatin was calculated and normalized to control reactions.

4.7.2. Chromatin Accessibility (FAIRE) Assays

Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE) assays were carried out as described (Simon et al. 2012) with modifications. After crosslinking for 5 min with 1% formaldehyde and quenching with 125 mM glycine for 5 min, cells were lyzed in FAIRE buffer

1 (50 mM HEPES-KOH [pH 7.5], 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) at 4°C for 10 min. After centrifugation for 5 min, 13,000 x g at 4°C, pellets were incubated in FAIRE buffer 2 (10 mM Tris-HCl [pH 8.0], 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) for 10 min at room temperature. Then 400 ml FAIRE buffer 3 (10 mM Tris-HCl, [pH 8.0], 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine) was added to the chromatin and sonicated with a Pico Biorupter to an average length of 200-300 bps. Samples were cleared by centrifugation, extracted twice with phenol/chloroform and once with chloroform (10% was kept as input). Input and extracted DNA were de-crosslinked by incubation at 65°C for 6 h, purified with the PCR purification kit and analyzed by qPCR. The ratio of DNA *versus* input was normalized to control reaction from uninduced cells.

4.8. Methods to study RNA-protein interactions

4.8.1. RNA immunoprecipitation

Nuclei were lysed in RIP buffer (20 mM Tris-HCl [pH 8.0], 200 mM NaCl, 1 mM EDTA, 1 mM EGTA, 0.5% Triton X-100, 1% NP-40, 100 U RNAsin, and Roche Complete protease inhibitors) for 15 min at 4°C. After brief sonication and treatment with DNase I, lysates were sonicated, cleared by centrifugation, diluted 5-fold in RIP buffer without detergents and incubated with the respective antibodies coupled to Dynabeads Protein G (Life Technologies) for 3.5 h at 4°C. Immobilized protein-RNA complexes were washed 3 times in buffer containing 600 mM NaCl, 20 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1 mM EGTA, 100 U RNAsin, protease inhibitors, 0.2% NP-40, 0.1% Triton X-100. Co-precipitated RNA was eluted for 30 min at 56°C in buffer containing 20 mM Tris-HCl [pH 8.0], 30 mM NaCl, 2.5 mM EDTA, 0.4% SDS, 20 mg/ml proteinase K, purified with TRIzol and analyzed by RT-qPCR. The fraction of co-precipitated RNA is calculated as percentage of input normalized to the IgG signal and presented in the reference to the control reaction.

4.8.2. Cross-linking and immunoprecipitation (CLIP) assay

Cross-linking and immunoprecipitation followed by RT-qPCR (CLIP-qPCR) was performed in U2OS/ER-E2F1 cells that were crosslinked with 150 mJ/cm2 at 254 nm using a Stratalinker. Cells were washed in 1 x PBS and resuspended in 300 µl Nuclei

isolation buffer A (10 mM Pipes [pH 6.8], 100 mM NaCl, 3 mM $MgCl_2$, 300 mM sucrose, 0.5% TritonX-100, 1 x Proteinase Inhibitory cocktail (PIC) , 100 U RNasin), incubated at 4°C for 2 min and centrifuged 1.3 g for 5 min at 4°C. Then cells were resuspended in 300 µl Nuclei isolation buffer B (10 mM Tris-HCl [pH 7.5], 10 mM KCl, 0.5% NP-40, 1x PIC, 100 U RNAsin), incubated for 5 min at 4 °C and centrifuged at 1.3 g for 5 min at 4°C. Isolated nuclei were lysed in RIPA buffer (10 mM Tris-HCl [pH 7.5], 140 mM NaCl, 1% Triton X-100, 0.1% Na-deoxycholate, 100 U RNAsin, and protease inhibitors). RNA was digested for 3 min with 0.1 U RNase I and 4 U TURBO DNase. After clearing by centrifugation, the lysates were diluted 1:3 with RIPA buffer without SDS and incubated with antibody-coupled Protein G Dynabeads overnight at 4°C. After 2 times washing with the buffer containing 50 mM Tris-HCl [pH 7.5], 1 M NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 0.5% Na-deoxycholate and once in 50 mM Tris-HCl [pH 7.5], 600 mM NaCl, 1 mM EDTA, 1% NP-40, 0.1% SDS, 2 M urea, co-precipitated RNA was eluted by incubation at 37°C for 30 min in TE buffer containing 0.5% SDS and 200 ng/ml proteinase K and for another 30 min with the same buffer containing 7 M urea. RNA was analyzed by RT-qPCR, the fraction of co-precipitated *KHPS1* being presented as percentage of input normalized to the IgG signal.

For CLIP assay followed by SDS-PAGE, U2OS/ER-E2F1 cells were crosslinked with 180 mJ/cm2 at 254 nm using a Stratalinker. Nuclei isolation and lysis followed by binding to the Protein G Dynabeads and washing of bead-bound RNA-protein complexes were performed as described for CLIP-qPCR above. After washing, co-precipitated RNA was dephosphorylated using 5 U FastAP Thermosensitive Alkaline Phosphatase at 37°C for 20 min. After washing the beads in 1x T4 Polynucleotide Kinase Reaction Buffer, RNA was phosphorylated at 5'end using [γ-32P] ATP using 5 U T4 Polynucleotide Kinase.

Bead-bound protein-RNA complexes were resuspended in 1x NuPAGE loading buffer, incubated at 70°C for 10 min and supernatant was loaded onto a gradient 4-12% Bis-Tris gel. Electrophoresis was performed using NuPAGE MOPS Running Buffer. RNA-protein complexes from the gel were transferred onto a nitrocellulose membrane by a wet transfer for overnight at 4°C at 25 V using NuPAGE Transfer Buffer containing 10% methanol. Radioactive signal of membrane was visualized by PhosphorImager.

4.9. Methods to study RNA-DNA interactions

4.9.1. Electrophoretic mobility shift assay (EMSA)

0.3 pmol of a ^{32}P-labeled double-stranded oligonucleotide comprising *SPHK1* sequences -357/-319 (**Table 11**) were incubated with an 50-fold molar excess of synthetic *KHPS1* (-373/-304) for 1 h at room temperature in 40 mM Tris-acetate [pH 7.5], 20 mM KCl, 10 mM MgAcetate, 10% glycerol and 1 x Phosphatase Inhibitor (PhosphoSTOP). The reaction mix was loaded onto 10% polyacrylamide gels containing 40 mM Tris-acetate [pH 7.5] and 10 mM MgAcetate. After electrophoresis, gel was dried at 85 °C for 1 h in gel dryer. The migration of the nucleic acids was visualized by PhosphorImaging.

4.9.2. *In vitro* Triplex Capture Assay

100 fmoles of PCR-fragments containing *eSPHK1* sequences (either -592/+7 or -406/-65) were digested with exonuclease I and incubated with 1 pmol of biotin-labeled *KHPS1* versions in 10 mM Tris-HCl [pH 7.5], 20 mM KCl, 10 mM $MgCl_2$, 0.05% Tween 20, and 100 U of RNasin (Promega) for 1.5 h at room temperature. RNA-DNA complexes were captured on MyOne Streptavidin C1 Dynabeads, washed three times with a buffer containing 150 mM KCl, 10 mM Tris-HCl [pH 7.5], 5 mM $MgCl_2$, 0.5% NP-40, and once with buffer containing 15 mM KCl, 10 mM Tris-HCl [pH 7.5] and 5 mM $MgCl_2$. RNA-associated DNA was eluted with RNase A (50 ng/ml, 30 min at 37°C), analyzed by qPCR, normalized to input DNA and presented in reference to control sample. To monitor recovery of *eSPHK1* sequences, primers -406/-304 were used.

4.9.3. Native In Vivo Triplex Capture Assay

$8x10^6$ HeLa cells were resuspended in 250 µl of cell lysis buffer (10 mM HEPES [pH 7.9] 10 mM KCl, 1,5 mM $MgCl_2$, 0,34 M sucrose, 10% glycerol) and 0,1% TritonX-100 was added, mixed and incubated 5 min on ice. At this point, the efficiency of cell lysis was checked microscopically. Samples were centrifuged 1,3 x g for 5 min at 4°C. Pellet was washed 2 times in cold cell lysis buffer, each time followed by centrifugation (1,3 x g for 5 min at 4°C). Isolated nuclei were incubated with 8 pmol of biotinylated *KHPS1* versions in triplex buffer (10 mM Tris-HCl [pH 7.5], 20 mM KCl, 10 mM $MgCl_2$, and 100 U of RNAsin) for 1 h at room temperature. RNA excess was removed by centrifugation through 0.88 M sucrose. Alternatively, nuclei isolated from HeLa cells that were transfected with *KHPS1* versions were proceeded directly to the nuclei lysis. Nuclei pellet was resuspended

in triplex buffer supplemented with 0,5% SDS and 100 ng of proteinase K, and samples were incubated for 15 min at room temperature for nuclei lysis. After sonication using Pico Bioruptor (6 pulses, 15 s on/ 30 s off) to obtain 500-600 bp DNA size, chromatin was centrifuged in 1,5 ml Eppendorf tubes for 5 min 6,000 rpm at 4°C. Supernatant was transferred into a fresh tube and MyOne Streptavidin C1 Dynabeads were added, followed by incubation for 1 h at room temperature. Bead-bound RNA-DNA complexes were washed 3 times with buffer containing 150 mM KCl, 10 mM Tris-HCl [pH 7.5], 5 mM $MgCl_2$, 0,5% NP-40, 100 U RNasin and 1 time with buffer containing 15 mM KCl, 10 mM Tris-HCl [pH 7.5], 5 mM $MgCl_2$ and 100 U RNasin). RNA-associated DNA was eluted with RNase A (50 ng/ml, 30 min at 37°C) in 100 µl TE buffer. DNA was purified using PCR Purification Kit according to the manufacturer's instructions and eluted in 30 µl of Qiagen elution buffer. DNA was analyzed by qPCR and normalized to input DNA.

4.10. Methods to study tumor cell motility

4.10.1. Wound-healing

To monitor cell migration, confluent cells were wounded by manual scratching with a 10 µl pipette tip. Plates were photographed immediately and 24 h after scratching using Nikon microscope (Eclipse TE2000).

4.10.2. Invasion assay

For Matrigel invasion assay, 5×10^4 cells were suspended in 0.3 ml medium containing 10% serum, plated in the top chamber with a Matrigel-coated membrane (24-well insert; pore size, 8 µm, (Corning Biocat) with 0.5 ml medium containing 20% serum as an attractant at the lower chamber. 16 h after seeding, cells in lower part of the chamber were fixed with 100% methanol and stained with 0.5% crystal violet.

4.10.3. Soft agar colony formation assay

5×10^3 MDA-MB-231 cells in medium containing 0.2% agarose were plated on top of a bottom layer containing 0.5% agarose and 20% serum. After incubation for 20-30 days with medium changes every three days, colonies were stained with 0.5% crystal violet and images were taken.

5. Abbreviations

4-OHT	4-Hydroxytamoxifen
A	Adenine
Ac	Acetylation
dGTP	2'-Desoxyguanosin-5'-triphosphat
As	Antisense
ASO	Antisense oligonucleotide
ATP	Adenosine-5'-triphosphate
Bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
CBP	CREB-Binding-Protein
CDC2	Cell division cycle protein 2
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP-sequencing
CLIP	Cross-linking immunoprecipitation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTCF	CCCTC-binding factor
DNA	Deoxyribonucleic acid
dATP	2'-deoxyadenosine-5'- triphosphate
dCas9	Deactivated Cas9
dNTP	Deoxyribonucleoside triphosphate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNase	Deoxyribonuclease
DOX	Doxycycline
dsDNA	Double-stranded deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethyleneglycol-bis-(2-aminoethylether)-tetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
ER	Estrogen receptor
eSPHK1	SPHK1 enhancer
FAIRE	Formaldehyde-Assisted Isolation of Regulatory Elements

FBS	Fetal bovine serum
G	Guanine
gDNA	Genomic DNA
GFP	Green fluorescent protein
h	Hour
H3	Histone H3
H4	Histone H4
HAT	Histone acetyltransferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOTAIR	HOX Transcript Antisense RNA
HRP	Horseradish peroxidase
IgG	Immunoglobulin G
KRAB	Krüppel associated box
LB	Lysogeny broth
LDS	Lithium dodecyl sulfate
lncRNA	Long non-coding RNA
Me	Methylation
MOPS	3-Morpholinopropane-1-sulfonic acid
mRNA	Messenger RNA
Mut	Mutant
Na-DOC	Sodium deoxycholate
NAM	Nicotinamide
NP-40	Nonidep P-40
NSCLC	Non-Small Cell Lung Cancer
Nt	Nucleotide
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCAF	P300/CBP-associated factor
PIPES	1,4-Piperazinediethanesulfonic acid
PNK	Polynucleotide Kinase
Pol II	Polymerase II
PRC2	Polycomb repressive complex 2
pSer5	Phosphorylated serine 5
PTEN	Phosphatase and Tensin homolog
qPCR	Quantitative PCR

rDNA	Ribosomal DNA
RIP	RNA immunoprecipitation
RNA	Ribonucleic acid
RNase	Ribonuclease
RNasin	Ribonuclease Inhibitor
Rpm	Revolutions per minute
RT	Reverse transcription
SDS	Sodium dodecyl sulfate
sgRNA	Single guide RNA
siRNA	Small interfering RNA
SPHK1	Sphingosine kinase 1
T	Thymidine
T-DMR	Tissue-differentially methylated regions
TBE	Tris-EDTA-borate buffer
TE	Tris-EDTA-buffer
TFO	Triplex forming oligonucleotide
TFR	Triplex forming region
Tris	Tris(hydroxymethyl)aminomethane
TSS	Transcription start site
U	Uridine
UTP	Uridine-5'-triphosphate
UV	Ultraviolet
VP64	Viral Protein 16x4
WT	Wildtype

6. References

Aguilera, A., and García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. Mol. Cell. 46(2):115-24.

Arab, K., Park, Y.J., Lindroth, A.M., Schäfer, A., Oakes, C., Weichenhan, D., Lukanova, A., Lundin, E., Risch, A., Meister, M., et al. (2014). Long Noncoding RNA TARID Directs Demethylation and Activation of the Tumor Suppressor TCF21 via GADD45A. Mol. Cell. 55(4):604-14

Balas, M.M., and Johnson, A.M. (2018). Exploring the mechanisms behind long noncoding RNAs and cancer. Non-Coding RNA Res. 3(3) 108-117.

Balasubramanyam, K., Varier, R.A., Altaf, M., Swaminathan, V., Siddappa, N.B., Ranga, U., and Kundu, T.K. (2004). Curcumin, a novel p300/CREB-binding protein-specific inhibitor of acetyltransferase, represses the acetylation of histone/nonhistone proteins and histone acetyltransferase-dependent chromatin transcription. J. Biol. Chem. *279*, 51163–51171.

Batista, P.J., and Chang, H.Y. (2013). Long noncoding RNAs: Cellular address codes in development and disease. Cell. 152(6):1298-307.

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C. V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res. 27(7):1139-1152.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 405(6785):482-5.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387–396.

Bell, J.C., Jukam, D., Teran, N.A., Risca, V.I., Smith, O.K., Johnson, W.L., Skotheim, J.M., Greenleaf, W.J., and Straight, A.F. (2018). Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. Elife. e27024.

Belotserkovskii, B.P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K.M., and Hanawalt, P.C. (2007). A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. J. Biol. Chem. *282*, 32433–32441.

Berkovich, E., and Ginsberg, D. (2003). ATM is a target for positive regulation by E2F-1. Oncogene *22*, 161–167.

Bierhoff, H., Dammert, M.A., Brocks, D., Dambacher, S., Schotta, G., and Grummt, I. (2014). Quiescence-Induced LncRNAs Trigger H4K20 Trimethylation and Transcriptional Silencing. Mol. Cell. 54(4):675-82.

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc. Natl. Acad. Sci. 112(18):5785-90.

Bose, D.A., Donahue, G., Reinberg, D., Shiekhattar, R., Bonasio, R., and Berger, S.L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. Cell *168*, 135–149.e22.

Bossen, C., Murre, C.S., Chang, A.N., Mansson, R., Rodewald, H.-R., and Murre, C. (2015). The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. Nat. Immunol. *16*, 775–784.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell. 144(3):327-39.

Buske, F.A., Mattick, J.S., and Bailey, T.L. (2011). Potential in vivo roles of nucleic acid triple-helices. RNA Biol. *8*, 427–439.

Buske, F.A., Bauer, D.C., Mattick, J.S., and Bailey, T.L. (2012). Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. Genome Res. *22*, 1372–1381.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Carbone, G.M., Napoli, S., Valentini, A., Cavalli, F., Watson, D.K., and Catapano, C. V. (2004). Triplex DNA-mediated downregulation of Ets2 expression results in growth inhibition and apoptosis in human prostate cancer cells. Nucleic Acids Res. *32*, 4358–4367.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature. 491(7424):454-7.

Cassiday, L.A., and Maher, L.J. (2002). Having it both ways: transcription factors that bind DNA and RNA. Nucleic Acids Res. *30*, 4118–4126.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: The enzymes in eukaryotes. FEBS J. 276(6):1494-505.

Chalei, V., Sansom, S.N., Kong, L., Lee, S., Montiel, J.F., Vance, K.W., and Ponting, C.P. (2014). The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. Elife. e04530.

Chen, H.Z., Tsai, S.Y., and Leone, G. (2009). Emerging roles of E2Fs in cancer: An exit from cell cycle control. Nat. Rev. Cancer. 9(11):785-97.

Cheng, A. joy, and van Dyke, M.W. (1993). Monovalent cation effects on intermolecular purine-purine-pyrimidine triple-helix formation. Nucleic Acids Res. 21(24): 5630–5635.

Cheng, Y., and Montgomery Pettitt, B. (1992). Hoogsteen versus Reversed-Hoogsteen Base Pairing: J. Am. Chem. Soc. *114*, 4465–4474.

Choukrallah, M.A., Song, S., Rolink, A.G., Burger, L., and Matthias, P. (2015). Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. Nat. Commun. 19;6:8324.

Cinghu, S., Yang, P., Kosak, J.P., Conway, A.E., Kumar, D., Oldfield, A.J., Adelman, K., and Jothi, R. (2017). Intragenic Enhancers Attenuate Host Gene Expression. Mol. Cell *68*, 104–117.e6.

Cloutier, S.C., Wang, S., Ma, W.K., Al Husini, N., Dhoondia, Z., Ansari, A., Pascuzzi, P.E., and Tran, E.J. (2016). Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. Mol. Cell. 61(3):393-404.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science. 339(6121):819-23.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M. a, Frampton, G.M., Sharp, P. a, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Datta, A., Loo, S.Y., Huang, B., Wong, L., Tan, S.S.L., Tan, T.Z., Lee, S.-C., Thiery, J.P., Lim, Y.C., Yong, W.P., et al. (2014). SPHK1 regulates proliferation and survival responses in triple-negative breast cancer. Oncotarget *5*, 5920–5933.

Dempsey, J.L., and Cui, J.Y. (2017). Long non-coding RNAs: A novel paradigm for toxicology.

Toxicol. Sci. 155(1):3-21

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. Trends Genet. 30(4):121-3.

Van Dongen, M.J.P., Doreleijers, J.F., Van der Marel, G.A., Van Boom, J.H., Hilbers, C.W., and Wijmenga, S.S. (1999). Structure and mechanism of formation of the H-y5 isomer of an intramolecular DNA triple helix. Nat. Struct. Biol. 6(9):854-9.

Du, Z., Sun, T., Hacisuleyman, E., Fei, T., Wang, X., Brown, M., Rinn, J.L., Lee, M.G., Chen, Y., Kantoff, P.W., et al. (2016). in prostate cancer. Nat. Commun. *7*, 1–10.

Duca, M., Vekhoff, P., Oussedik, K., Halby, L., and Arimondo, P.B. (2008). The triple helix: 50 years later, the outcome. Nucleic Acids Res. 36(16):5123-38.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature. 489, pages 57–74.

Eldholm, V., Haugen, A., and Zienolddiny, S. (2014). CTCF mediates the TERT enhancer-promoter interactions in lung cancer cells: Identification of a novel enhancer region involved in the regulation of TERT gene. Int. J. Cancer. 134(10):2305-13.

Escudeé, C., Francçois, J. chritophe, Sun, J. sheng, Ott, G., Sprinzl, M., Garestier, T., and Heélene, J. chritophe (1993). Stability of triple helices containing RNA and DNA strands: Experimental and molecular modeling studies. Nucleic Acids Res. 21(24):5547-53.

Espinosa, M. (2016). Previews Revisiting lncRNAs : How Do You Know Yours Is Not an eRNA ? Mol.Cell. 62(1):1-2.

Faghihi, M.A., and Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat. Rev. Mol. Cell Biol. 10(9):637-43.

Fang, Y., and Fullwood, M.J. (2016). Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. Genomics, Proteomics Bioinforma. 14(1):42-54.

Feldstein, O., Nizri, T., Doniger, T., Jacob, J., Rechavi, G., and Ginsberg, D. (2013). The long non-coding RNA ERIC is regulated by E2F and modulates the cellular response to DNA damage. Mol. Cancer *12*, 131.

Felsenfeld, G., Davies, D.R., and Rich, A. (1957). Formation of a Three-Stranded Polynucleotide Molecule. J. Am. Chem. Soc. *79*, 2023–2024.

Floris, R., Scaggiante, B., Manzini, G., Quadrifoglio, F., and Xodo, L.E. (1999). Effect of cations on purine·purine·pyrimidine triple helix formation in mixed-valence salt solutions. Eur. J. Biochem. 260(3):801-9.

French, K.J., Schrecengost, R.S., Lee, B.D., Zhuang, Y., Smith, S.N., Eberly, J.L., Yun, J.K., and Smith, C.D. (2003). Discovery and evaluation of inhibitors of human sphingosine kinase. Cancer Res. *63*, 5962–5969.

Goñi, J.R., de la Cruz, X., and Orozco, M. (2004). Triplex-forming oligonucleotide target sequences in the human genome. Nucleic Acids Res. *32*, 354–360.

Grote, P., Wittler, L., Hendrix, D., Koch, F., Währisch, S., Beisaw, A., Macura, K., Bläss, G., Kellis, M., Werber, M., et al. (2013). The Tissue-Specific lncRNA Fendrr Is an Essential Regulator of Heart and Body Wall Development in the Mouse. Dev. Cell *24*, 206–214.

Halley, P., Kadakkuzha, B.M., Faghihi, M.A., Magistri, M., Zeier, Z., Khorkova, O., Coito, C., Hsiao, J., Lawrence, M., and Wahlestedt, C. (2014). Regulation of the apolipoprotein gene cluster by a long noncoding RNA. Cell Rep. ;6(1):222-30.

Han, H., and Dervan, P.B. (1993). Sequence-specific recognition of double helical RNA and RNA.DNA by triple helix formation. Proc. Natl. Acad. Sci. 90(9):3806-10.

Handy, D.E., Castro, R., and Loscalzo, J. (2011). Epigenetic modifications: Basic mechanisms and role in cardiovascular disease. Circulation. 4(2):169-74.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature. 405(6785):486-9.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: The reference human genome annotation for the ENCODE project. Genome Res. 22(9):1760-74.

Hayes, J.J., and Hansen, J.C. (2001). Nucleosomes and the chromatin fiber. Curr. Opin. Genet. Dev. 11(2):124-9.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 459(7243):108-12.

Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. Acta Crystallogr. 16, 907.

Hsieh, C.-L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.-S.M., Chen, S., et al. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. Proc. Natl. Acad. Sci. *111*, 7319–7324.

Huber, H.E., Edwards, G., Goodhart, P.J., Patrick, D.R., Huang, P.S., Ivey-Hoyle, M., Barnett, S.F., Oliff, a, and Heimbrook, D.C. (1993). Transcription factor E2F binds DNA as a heterodimer. Proc. Natl. Acad. Sci. U. S. A. 90(8):3525-9.

Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S., and Shiota, K. (2004a). Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem. Biophys. Res. Commun. *322*, 593–600.

Imamura, T., Miyauchi-senda, N., Tanaka, S., and Shiota, K. (2004b). Identification of Genetic and Epigenetic Similarities of SPHK1 / Sphk1 in Mammals. 66(11):1387-93.

Ing, N.H., Beekman, J.M., Kessler, D.J., Murphy, M., Jayaraman, K., Zendegui, J.G., Hogan, M.E., O'malley, B.W., and Tsai, M. jer (1993). In vivo transcription of a progesterone-responsive gene is specifically inhibited by a triplex-forming oligonucleotide. Nucleic Acids Res. 21(12):2789-96.

Jain, A., Wang, G., and Vasquez, K.M. (2008). DNA triple helices: Biological consequences and therapeutic potential. Biochimie. 90(8):1117-30.

Jalali, S., Singh, A., Maiti, S., and Scaria, V. (2017). Genome-wide computational analysis of potential long noncoding RNA mediated DNA:DNA:RNA triplexes in the human genome. J. Transl. Med. 15(1):186.

Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Jiao, W., Chen, Y., Song, H., Li, D., Mei, H., Yang, F., Fang, E., Wang, X., Huang, K., Zheng, L., et al. (2018). HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis. Oncogene. 37(20):2728-2745.

Jubair, L., and McMillan, N.A.J. (2017). The Therapeutic Potential of CRISPR/Cas9 Systems in Oncogene-Addicted Cancer Types: Virally Driven Cancers as a Model System. Mol. Ther. - Nucleic Acids. 15; 8: 56–63.

Kalwa, M., Hänzelmann, S., Otto, S., Kuo, C.C., Franzen, J., Joussen, S., Fernandez-Rebollo, E., Rath, B., Koch, C., Hofmann, A., et al. (2016). The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. Nucleic Acids Res. *44*, 10631–10643.

Karympalis, V., Kalopita, K., Zarros, A., and Carageorgiou, H. (2004). Regulation of gene expression via triple helical formations. Biochem. 69(8):855-60.

Khilji, S., Hamed, M., Chen, J., and Li, Q. (2018). Loci-specific histone acetylation profiles associated with transcriptional coactivator p300 during early myoblast differentiation. Epigenetics. 13(6):642-654.

Kim, H.G., and Miller, D.M. (1998). A novel triplex-forming oligonucleotide targeted to human cyclin D1 (bcl-1, proto-oncogene) promoter inhibits transcription in HeLa cells. Biochemistry. 37(8):2666-72.

Kim, H.G., Reddoch, J.F., Mayfield, C., Ebbinghaus, S., Vigneswaran, N., Thomas, S., Jones, D.E., and Miller, D.M. (1998). Inhibition of transcription of the human c-myc protooncogene by intermolecular triplex. Biochemistry *37*, 2299–2304.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kim, Y.W., Lee, S., Yun, J., and Kim, A. (2015). Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the β-globin locus. Biosci. Rep. *35*, 1–8.

Korostowski, L., Sedlak, N., and Engel, N. (2012). The Kcnq1ot1 Long Non-Coding RNA Affects Chromatin Conformation and Expression of Kcnq1, but Does Not Regulate Its Imprinting in the Developing Heart. PLoS Genet. e1002956.

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., and Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15 INK4B tumor suppressor gene. Oncogene. 30(16): 1956–1962.

Krawczyk, M., and Emerson, B.M. (2014). P50-associated COX-2 Extragenic RNA (pacer) activates human COX-2 gene expression by occluding repressive NF-κB p50 complexes. Elife. e01776.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., delRosario, B.C., Pinter, S.F., et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol. Cell *57*, 361–375.

Lam, M.T.Y., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. Trends Biochem. Sci. *39*, 170–182.

Léveillé, N., Melo, C.A., Rooijers, K., Díaz-Lagares, A., Melo, S.A., Korkmaz, G., Lopes, R., Moqadam, F.A., Maia, A.R., Wijchers, P.J., et al. (2015). Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. Nat. Commun. 6:6520.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., and Zhang, J. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. 3–10.

Li, X., Zhou, B., Chen, L., Gou, L.T., Li, H., and Fu, X.D. (2017). GRID-seq reveals the global RNA-chromatin interactome. Nat. Biotechnol. 35, pages 940–950.

Lin, A., Li, C., Xing, Z., Hu, Q., Liang, K., Han, L., Wang, C., Hawke, D.H., Wang, S., Zhang, Y., et al. (2016). The LINK-A lncRNA Activates Normoxic HIF1α Signaling in Triple-negative Breast Cancer HHS Public Access. Nat Cell Biol *18*, 213–224.

Liu, C., and Lin, J. (2016). Long noncoding RNA ZEB1-AZS1 acts as an oncogene in osteosarcoma by epigenetically activating ZEB1. Am. J. Transl. Res. 8(10): 4095–4105.

Long, J., Xie, Y., Yin, J., Lu, W., and Fang, S. (2015). SphK1 promotes tumor cell migration and invasion in colorectal cancer. Tumour Biol. 37(5):6831-6.

Lu, J., and Tang, M. (2012). CTCF-dependent chromatin insulator as a built-in attenuator of angiogenesis. Transcription. 3(2):73-7.

Lu, Y., Liu, X., Xie, M., Liu, M., Ye, M., Li, M., Chen, X.-M., Li, X., and Zhou, R. (2017). The NF-κB–Responsive Long Noncoding RNA FIRRE Regulates Posttranscriptional Regulation of Inflammatory Gene Expression through Interacting with hnRNPU. J. Immunol. 199(10):3571-3582.

Luger, K., and Richmond, T.J. (1998). The histone tails of the nucleosome. Curr. Opin. Genet. Dev. 8(2):140-6.

Luo, G., Liu, D., Huang, C., Wang, M., Xiao, X., Zeng, F., Wang, L., and Jiang, G. (2017). LncRNA GAS5 Inhibits Cellular Proliferation by Targeting P27 ^{Kip1}. Mol. Cancer Res. 15(7):789-799.

Lyle, S., and Moore, N. (2011). Quiescent, slow-cycling stem cell populations in cancer: A review of the evidence and discussion of significance. J. Oncol. Volume 2011.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C., Ogrodowski, P., Crippa, A., Rekhtman, N., Stanchina, E. De, et al. (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature. 516(7531):423-7.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat. Methods. 10, pages 977–979.

Maiti, A., Takabe, K., and Hait, N.C. (2017). Metastatic triple-negative breast cancer is dependent on SphKs/S1P signaling for growth and survival. Cell. Signal. 32:85-92.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666–670.

McHugh, C.A., Chen, C.-K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., et al. (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232–236.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Vrielink, J.A.F.O., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., et al. (2013). ERNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. Mol. Cell. 49(3):524-35.

Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development. 137(15):2493-9.

Mondal, T., Subhash, S., Vaid, R., Enroth, S., Uday, S., Reinius, B., Mitra, S., Mohammed, A., James, A.R., Hoberg, E., et al. (2015). MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA-DNA triplex structures. Nat. Commun. *6*, 7743.

Mousavi, K., Zare, H., Dell'Orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. Mol. Cell *51*, 606–617.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science. 322(5908):1717-20.

Natoli, G., and Andrau, J.-C. (2012). Noncoding Transcription at Enhancers: General Principles and Functional Models. Annu. Rev. Genet. 46:1-19.

Ng, S.Y., Bogu, G.K., Soh, B., and Stanton, L.W. (2013). The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. Mol. Cell. 51(3):349-59.

Nguyen, T.C., Cao, X., Hebert, L., Zhong, S., Sridhar, B., Rivas-astroza, M., Nguyen, T.C., Chen, W., Yan, Z., and Cao, X. (2017). Systematic Mapping of RNA-Chromatin Interactions In Vivo. Curr. Biol. 27(4):602-609.

Nishikawa, K., and Kinjo, A.R. (2017). Essential role of long non-coding RNAs in de novo chromatin modifications: the genomic address code hypothesis. Biophys. Rev. ;9(2):73-77.

O'Leary, V.B., Ovsepian, S.V., Carrascosa, L.G., Buske, F.A., Radulovic, V., Niyazi, M., Moertl, S., Trau, M., Atkinson, M.J., and Anastasov, N. (2015). PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. Cell Rep. *11*, 474–485.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: New insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. 12(4):283-93.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell *143*, 46–58.

Pan, J., Tao, Y.F., Zhou, Z., Cao, B.R., Wu, S.Y., Zhang, Y.L., Hu, S.Y., Zhao, W.L., Wang, J., Lou, G.L., et al. (2011). An novel role of sphingosine kinase-1 (SPHK1) in the invasion and metastasis of esophageal carcinoma. J. Transl. Med. 22;9:157.

Park, S.M., Choi, E.Y., Bae, D.H., Sohn, H.A., Kim, S.Y., and Kim, Y.J. (2018). The LncRNA EPEL Promotes Lung Cancer Cell Proliferation Through E2F Target Activation. Cell. Physiol. Biochem. 45(3):1270-1283.

Paugh, B.S., Bryan, L., Paugh, S.W., Wilczynska, K.M., Alvarez, S.M., Singh, S.K., Kapitonov, D., Rokita, H., Wright, S., Griswold-Prenner, I., et al. (2009). Interleukin-1 regulates the expression of sphingosine kinase 1 in glioblastoma cells. J. Biol. Chem. *284*, 3408–3417.

Paugh, S.W., Coss, D.R., Bao, J., Laudermilk, L.T., Grace, C.R., Ferreira, A.M., Waddell, M.B., Ridout, G., Naeve, D., Leuze, M., et al. (2016). MicroRNAs Form Triplexes with Double Stranded DNA at Sequence-Specific Binding Sites; a Eukaryotic Mechanism via which microRNAs Could Directly Alter Gene Expression. PLoS Comput. Biol. *12*.

Perrotti, D., Bonatti, S., Trotta, R., Martinez, R., Skorski, T., Salomoni, P., Grassilli, E., Iozzo, R. V., Cooper, D.R., and Calabretta, B. (1998). TLS/FUS, a pro-oncogene involved in multiple chromosomal translocations, is a novel regulator of BCR/ABL-mediated leukemogenesis. EMBO J. 17(15):4442-55.

Peyman, J.A. (1999). Repression of Major Histocompatibility Complex Genes by a Human Trophoblast Ribonucleic Acid. Biol. Reprod. 60(1):23-31.

Plum, G.E., Park, Y.W., Singleton, S.F., Dervan, P.B., and Breslauer, K.J. (1990).

Thermodynamic characterization of the stability and the melting behavior of a DNA triplex: a spectroscopic and calorimetric study. Proc Natl Acad Sci USA. 87(23):9436-40.

Postel, E.H., Flint, S.J., Kessler, D.J., and Hogan, M.E. (1991). Evidence that a triplex-forming oligodeoxyribonucleotide binds to the c-myc promoter in HeLa cells, thereby reducing c-myc mRNA levels. Proc Natl Acad Sci U S A. 88(18):8227-31.

Postepska-Igielska, A., Giwojna, A., Gasri-Plotnitsky, L., Schmitt, N., Dold, A., Ginsberg, D., and Grummt, I. (2015). LncRNA Khps1 Regulates Expression of the Proto-oncogene SPHK1 via Triplex-Mediated Changes in Chromatin Structure. Mol. Cell *60*, 626–636.

Protozanova, E., and Macgregor, R.B. (1996). Kinetic footprinting of DNA triplex formation. Anal. Biochem. 243(1):92-9.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Rahman, S., Zorca, C.E., Traboulsi, T., Noutahi, E., Krause, M.R., Mader, S., and Zenklusen, D. (2016). Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription. Nucleic Acids Res. 45(6):3017-3030.

Raisner, R., Kharbanda, S., Jin, L., Jeng, E., Chan, E., Merchant, M., Haverty, P.M., Bainer, R., Cheung, T., Arnott, D., et al. (2018). Enhancer Activity Requires CBP/P300 Bromodomain-Dependent Histone H3K27 Acetylation. Cell Rep. 24(7):1722-1729.

Reaban, M.E., Lebowitz, J., and Griffin, J.A. (1994). Transcription induces the formation of a stable RNA·DNA hybrid in the immunoglobulin α switch region. J. Biol. Chem. 269(34):21850-7.

Ren, C., Liu, F., Ouyang, Z., An, G., Zhao, C., Shuai, J., Cai, S., Bo, X., and Shu, W. (2017a). Functional annotation of structural ncRNAs within enhancer RNAs in the human genome: Implications for human disease. Sci. Rep. 15518.

Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D.R., and Zhao, K. (2017b). CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. Mol. Cell. ;67(6):1049-1058.

Rhie, S., Hazelett, D.J., Coetzee, S.G., Yan, C., Noushmehr, H., and Coetzee, G.A. (2014). Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. BMC Genomics *15*, 331.

Roberts, R.W., and Crothers, D.M. (1996). Prediction of the stability of DNA triplexes. Proc. Natl. Acad. Sci. 83(24):9373–9377.

Roux, B.T., Heward, J.A., Donnelly, L.E., Jones, S.W., and Lindsay, M.A. (2017). Catalog of Differentially expressed long non-coding RNA following activation of human and mouse innate immune response. Front. Immunol. 8: 1038.

Roy, D., and Lieber, M.R. (2009). G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. Mol. Cell. Biol. *29*, 3124–3133.

Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., and Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes Dev. *28*, 723–734.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: The rosetta stone of a hidden RNA language? Cell. 146(3):353-8.

Salviano-Silva, A., Lobo-Alves, S., Almeida, R., Malheiros, D., and Petzl-Erler, M. (2018).

Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis. Non-Coding RNA. 30;4(1).

de Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. PLoS Biol. *8*. 8(5):e1000384.

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: New modulators of genome dynamics and function. Nat. Rev. Genet. 16(10):583-97.

Sarkar, S., Maceyka, M., Hait, N.C., Paugh, S.W., Sankala, H., Milstien, S., and Spiegel, S. (2005). Sphingosine kinase 1 is required for migration, proliferation and survival of MCF-7 human breast cancer cells. FEBS Lett. *579*, 5313–5317.

Schmitt, A.M., Garcia, J.T., Hung, T., Flynn, R.A., Shen, Y., Qu, K., Payumo, A.Y., Peres-da-Silva, A., Broz, D.K., Baum, R., et al. (2016). An inducible long noncoding RNA amplifies DNA damage signaling. Nat. Genet. 1–9. 48(11):1370-1376.

Schmitz, K.-M., Mayer, C., Postepska, A., and Grummt, I. (2010a). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. *24*, 2264–2269.

Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010b). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. *24*, 2264–2269.

Schuijers, J., Manteiga, J.C., Weintraub, A.S., Day, D.S., Zamudio, A.V., Hnisz, D., Lee, T.I., and Young, R.A. (2018). Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. Cell Rep. 23(2):349-360.

Shida, D., Takabe, K., Kapitonov, D., Milstien, S., and Spiegel, S. (2008). Targeting SphK1 as a new strategy against cancer. Curr. Drug Targets *9*, 662–673.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science (80-.). *350*, 978–981.

Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nat. Protoc. 7(2):256-67.

Slansky, J.E., and Farnham, P.J. (1996). Introduction to the E2F family: protein structure and gene regulation. Curr. Top. Microbiol. Immunol. 208:1-30.

Soibam, B. (2017). Super-lncRNAs: identification of lncRNAs that target super-enhancers via RNA:DNA:DNA triplex formation. RNA *23*, 1729–1742.

Song, C., Hotz-Wagenblatt, A., Voit, R., and Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes Dev. 31(13):1370-1381.

Spiegel, S., and Milstien, S. (2003). Sphingosine-1-phosphate: An enigmatic signalling lipid. Nat. Rev. Mol. Cell Biol. 4(5):397-407.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and De Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. 20(17):2349-54.

Sugimoto, N., Nakano, S. ichi, Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995). Thermodynamic Parameters To Predict Stability of RNA/DNA Hybrid Duplexes. Biochemistry. 34 (35), pp 11211–11216.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013a). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science (80-.). *340*, 619–621.

Sun, S., Del Rosario, B.C., Szanto, A., Ogawa, Y., Jeon, Y., and Lee, J.T. (2013b). XJpx RNA activates xist by evicting CTCF. Cell *153*.

Svinarchuk, F., Cherny, D., Debin, A., Delain, E., and Malvy, C. (1996). A new approach to overcome potassium-mediated inhibition of triplex formation. Nucleic Acids Res. 24(19):3858-65.

Svotelis, A., Bianco, S., Madore, J., Huppé, G., Nordell-Markovits, A., Mes-Masson, A.-M., and Gévry, N. (2011). H3K27 demethylation by JMJD3 at a poised enhancer of anti-apoptotic gene *BCL2* determines ERα ligand dependency. EMBO J. *30*, 3947–3961.

Thomas, M., White, R.L., and Davis, R.W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc. Natl. Acad. Sci. *73*, 2294–2298.

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

Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. Nature. 461(7261): 199–205.

Wan, G., Hu, X., Liu, Y., Han, C., Sood, A.K., Calin, G.A., Zhang, X., and Lu, X. (2013). A novel non-coding RNA lncRNA-JADE connects DNA damage signalling to histone H4 acetylation. EMBO J. 32(21):2833-47.

Wang, K.C., and Chang, H.Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. Mol. Cell *43*, 904–914.

Wang, J., Pan, Y., Wu, J., Zhang, C., Huang, Y., Zhao, R., Cheng, G., Liu, J., Qin, C., Shao, P., et al. (2016). The association between abnormal long noncoding RNA MALAT-1 expression and cancer lymph node metastasis: A meta-analysis. Biomed Res. Int. Volume 2016.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature. 472, p. 120–124.

Wang, L., Tang, Y., Cole, P.A., and Marmorstein, R. (2008). Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function. Curr. Opin. Struct. Biol. 18(6):741-7.

Werner, M.S., and Ruthenburg, A.J. (2015). Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. Cell Rep. *12*, 1089–1098.

Werner, M.S., Sullivan, M.A., Shah, R.N., Nadadur, R.D., Grzybowski, A.T., Galat, V., Moskowitz, I.P., and Ruthenburg, A.J. (2017). Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription. Nat. Struct. Mol. Biol. 24(7):596-603.

Wu, Q., Gaddis, S.S., MacLeod, M.C., Walborg, E.F., Thames, H.D., DiGiovanni, J., and Vasquez, K.M. (2007). High-affinity triplex-forming oligonucleotide target sequences in mammalian genomes. Mol. Carcinog. 46(1):15-23.

Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat. Genet. *30*, 167–174.

Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc. Natl. Acad. Sci. *104*, 7145–7150.

Yamanaka, Y., Faghihi, M.A., Magistri, M., Alvarez-Garcia, O., Lotz, M., and Wahlestedt, C. (2015). Antisense RNA Controls LRP1 Sense Transcript Expression through Interaction with a Chromatin-Associated Protein, HMGB2. Cell Rep. 11(6):967-976.

Yang, Y., Su, Z., Song, X., Liang, B., Zeng, F., Chang, X., and Huang, D. (2016). Enhancer RNA-driven looping enhances the transcription of the long noncoding RNA DHRS4-AS1, a controller of the DHRS4 gene cluster. Sci. Rep. *6*, 20961.

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.M. (2010). Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. Mol. Cell. 38(5):662-74.

Yu, L., Fang, F., Lu, S., Li, X., Yang, Y., and Wang, Z. (2017). LncRNA-HIT promotes cell proliferation of non-small cell lung cancer by association with E2F1. Cancer Gene Ther. 24(5):221-226.

Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nat. Struct. Mol. Biol. *20*, 259–266.

Zhan, T., Rindtorff, N., Betge, J., Ebert, M.P., and Boutros, M. (2018). CRISPR/Cas9 for cancer research and therapy. Semin. Cancer Biol. (17)30274-2

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614.

Zhang, Y., Wang, Y., Wan, Z., Liu, S., Cao, Y., and Zeng, Z. (2014). Sphingosine kinase 1 and cancer: A systematic review and meta-analysis. PLoS One *9*. 9(2):e90362

Zhao, J., Song, X., and Wang, K. (2016). LncScore: Alignment-free identification of long noncoding RNA from assembled novel transcripts. Sci. Rep. 6:34838

Zhao, Z., Sentürk, N., Song, C., and Grummt, I. (2018). lncRNA PAPAS tethered to the rDNA enhancer recruits hypophosphorylated CHD4/NuRD to repress rRNA synthesis at elevated temperatures. Genes Dev. 32(11-12):836-848

Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes Dev. Dec 15;28(24):2699-711

Y., Zhang, S., Wang, X., Li, Y., Lu, L., Chen, X., Chen, F., Bao, X., et al. (2015). Linc-YY1 promotes myogenic differentiation and muscle regeneration through an interaction with the transcription factor YY1. Nat. Commun. Dec 11;6:10026

Zhu, L., Wang, Z., Lin, Y., Chen, Z., Liu, H., Chen, Y., Wang, N., and Song, X. (2015). Sphingosine kinase 1 enhances the invasion and migration of non-small cell lung cancer cells via the AKT pathway. Oncol. Rep. *33*, 1257–1263.