




Divergent lncRNA MYMLR regulates MYC by eliciting DNA looping and promoter-enhancer interaction

Taisuke Kajino^{1,2}, Teppei Shimamura³, Shuyi Gong¹, Kiyoshi Yanagisawa¹, Lisa Ida¹, Masahiro Nakatochi⁴ , Sebastian Griesing¹, Yukako Shimada^{1,2}, Keiko Kano¹, Motoshi Suzuki¹ , Satoru Miyano⁵ & Takashi Takahashi^{1,6,*} 

Abstract

Long non-coding RNAs (lncRNAs) function in a wide range of processes by diverse mechanisms, though their roles in regulation of oncogenes and/or tumor suppressors remain rather elusive. We performed a global search for lncRNAs affecting MYC activity using a systems biology-based approach with a K supercomputer and the GIMLET algorithm based on local distance correlations. Consequently, *MYMLR* was identified and experimentally shown to maintain MYC transcriptional activity and cell cycle progression despite the low levels of expression. A proteomic search for *MYMLR*-binding proteins identified PCBP2, while it was also found that *MYMLR* places a 557-kb upstream enhancer region in the proximity of the *MYC* promoter in cooperation with PCBP2. These findings implicate a crucial role for *MYMLR* in regulation of the archetypical oncogene *MYC* and warrant future studies regarding the involvement of low copy number lncRNAs in regulation of other crucial oncogenes and tumor suppressor genes.

Keywords enhancer; lncRNA; lung cancer; MYC; systems biology

Subject Categories Cancer; RNA Biology

DOI 10.15252/emj.201798441 | Received 17 October 2017 | Revised 13 June 2019 | Accepted 21 June 2019 | Published online 30 July 2019

The EMBO Journal (2019) 38: e98441

Introduction

Long non-coding RNAs (lncRNAs), broadly defined as non-coding transcripts longer than 200 nucleotides, are transcribed as antisense, divergent, intronic, and intergenic RNA transcripts (Rinn & Chang, 2012; Fatica & Bozzoni, 2014). Accumulating evidence indicates that

lncRNAs modulate protein-coding gene expression in *cis* or *trans* through a variety of molecular mechanisms, including such functions as molecular scaffold, guide to recruit proteins to a specific region of the genome or transcripts, and competing endogenous RNA. Although tens of thousands of lncRNAs are now thought to exist in the human genome, the molecular functions in only a small fraction have been well characterized, leaving the remaining vast majority yet to be studied in detail. It is also becoming increasingly clear that the expressions of numerous lncRNAs are altered in cancer, though mechanistic insight into how such cancer-related lncRNAs contribute to carcinogenesis and disease progression is rather limited.

Lung carcinoma is a leading cause of cancer death worldwide, resulting in more than 1.3 million deaths per year, of which more than 40% are from lung adenocarcinoma. *MYC* is regarded as an archetypical proto-oncogene, and its deregulated expression is frequently associated with poor prognosis in a broad range of human cancers including lung cancer, indicating a key role for *MYC* in cancer development (Little *et al.*, 1983, Cancer Genome Atlas Research, 2014). Several lines of evidence suggest the possibility that formation of a long-range loop structure involving the *MYC* promoter and enhancers may provide a functional link of risk SNPs in surrounding genomic regions to colorectal, prostate, and breast cancer susceptibility (Easton *et al.*, 2007; Haiman *et al.*, 2007; Tomlinson *et al.*, 2007). *MYC* functions as a transcription factor to regulate numerous genes involved in cell growth, proliferation, metabolism, and transformation (Dang, 2012; Kress *et al.*, 2015). Recent findings indicate that its transcriptional activity is controlled by multiple mechanisms, such as protein modification and MYC activity itself, while specific interactions with proteins, including epigenetic modifiers, transcriptional factors/coactivators, and ribonucleoproteins, have also been implicated. Furthermore, it is notable that its enhancer elements may be

1 Division of Molecular Carcinogenesis, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan

2 Division of Molecular Diagnostics, Aichi Cancer Center Research Institute, Nagoya, Japan

3 Division of Systems Biology, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan

4 Statistical Analysis Section, Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan

5 Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan

6 Aichi Cancer Center, Nagoya, Japan

*Corresponding author. Tel: +81 52 762 6111; E-mail: tak@aichi-cc.jp

involved in cell type/context-specific gene expression (Shlyueva et al, 2014).

In the present study, we attempted to identify lncRNAs that have effects on MYC transcriptional activity. For that purpose, we employed an integrative approach by combining both a systems biology-based genome-wide search conducted with a K supercomputer and findings from detailed molecular biologic experimentation to elucidate the underlying mechanism. We report here for the first time identification of a lncRNA required to maintain MYC activity in cooperation with the RNA-binding protein PCBP2.

Results

Identification of MYMLR as a lncRNA affecting MYC activity

We employed a systems biology-based search for lncRNAs with effects on MYC activity (Fig 1A). First, we downloaded 596 RNA-seq datasets of lung adenocarcinoma from the Cancer Genome Atlas (TCGA) and 935 RNA-seq datasets of tumor cell lines from the Cancer Cell Line Encyclopedia (CCLE). In parallel, we also defined a MYC module reflecting its transcriptional activity with a set of MYC downstream genes using Ingenuity Pathway Analysis. We then utilized the GIMLET program implemented on a K Super Computer (preprint: Shimamura et al, 2018), which allowed us to conduct a global search for lncRNAs based on genome-wide calculations of local distance correlations of each lncRNA with MYC module activity in the regulatory network (Fig 1B). Among the selected candidate lncRNAs with differential expression between normal and tumor tissues (Table EV1), one of the top-ranked lncRNAs was found to reside in close proximity (2 kb upstream) to the MYC gene locus with 3 exons, which attracted our attention (Figs 1C and EV1A). Those results were validated using GIMLET analysis with other experimentally defined MYC-targeting gene sets (Fig EV1B). It was shown that this lncRNA (herein termed *MYC-modulating lncRNA*, *MYMLR*) consisted of 3 exons, and the transcription start sites (TSSs) of *MYMLR* and *MYC* were 350 bp apart, as determined by 5'- and 3'-RACE findings (Fig EV1C). Furthermore, the *MYMLR* transcription start site was found to be in agreement with 5' cap analysis gene expression (CAGE) sequence data obtained from the Encyclopedia of DNA Elements (ENCODE) (Fig 1C). *MYMLR* was also found to be transcribed at a very low level (0.2 copies per cell) in the A549 lung adenocarcinoma cell line (Fig EV1D and Appendix Fig S1), in marked contrast to an abundant expression of *MYC* mRNA.

MYMLR-mediated sustainment of MYC activity and expression

Next, we conducted experiments to examine whether *MYMLR* has effects on MYC activity in lung cancer cells. To this end, gene set enrichment analysis (GSEA) was performed using a collection of hallmark gene sets, which revealed a significant association of *MYMLR* with multiple MYC-related gene sets (Fig 2A), including HALLMARK MYC TARGETS v2 and MYC UP.V1 UP (Figs 2B and EV2A). Those findings were confirmed by GSEA using another siRNA against *MYMLR* (Fig EV2B and Appendix Table S1). We also assessed changes in *miR-17-5p* expression, known to be regulated by MYC, in A549 cells treated with 3 independent siRNAs. siMYMLR

treatment significantly decreased *miR-17-5p* expression to a degree similar to that seen with siMYC treatment (Fig 2C). In addition, a luciferase assay with a reporter construct containing a promoter region of *C13orf25*, a MYC-regulated host gene of the *miR-17-92* cluster containing *miR-17-5p*, was conducted, which showed a significant decrease in luciferase activity by siMYMLR treatment, as also seen with siMYC (Fig 2D). Interestingly, knockdown of *MYMLR* markedly reduced MYC expression itself at both the mRNA and protein levels in A549 and ACC-LC-319 cells (Figs 2E and EV2C and D). Additionally, we confirmed that treatment with *MYMLR* antisense oligonucleotides similarly elicited a significant reduction of MYC (Appendix Fig S2). Together, these results indicated that *MYMLR* maintains MYC activity through sustainment of MYC expression.

Involvement of MYMLR in cell cycle progression and proliferation

MYC activity is known to play a role in cell cycle progression, while the present findings clearly indicated involvement of *MYMLR* in regulation of MYC expression. In our analysis of the role of *MYMLR* in cell cycle progression using thymidine-aphidicolin double-block treatment, both *MYMLR* knockdown and *MYC* knockdown similarly decreased phospho-histone H3-positive cells (Figs 3A and EV3A–C). Additionally, pathway analysis of genes down-regulated by siMYMLR introduction revealed significant enrichment of those tagged as related to the cell cycle (Fig EV3D). Colorimetric assay results also showed a significant reduction in cell proliferation of A549 and ACC-LC-319 cells knocked down for either *MYMLR* or *MYC* (Fig 3B). We then performed a colony formation assay and found that *MYMLR* knockdown markedly reduced the number of A549 cell colonies (Figs 3C and EV3E). In addition, siMYMLR treatment resulted in a marked decrease in tumorigenicity in mice administered A549 cells (Figs 3D and EV3F). Taken together, these findings demonstrated crucial involvement of *MYMLR* in cell cycle progression and cancer cell proliferation.

Identification of PCBP2 as MYMLR-binding protein

Accumulating evidence indicates that lncRNAs often function through an RNA–protein interaction. In order to gain mechanistic insight into how *MYMLR* regulates *MYC* expression, we conducted a global search for a *MYMLR*-binding protein(s). We first performed a nuclear fractionation assay, which showed subcellular localization of *MYMLR* in both nuclei and cytoplasm (Fig EV4A). An RNA pull-down experiment was performed by incubating *in vitro*-synthesized BrU-labeled RNA harboring exons 2 and 3 of *MYMLR* or *GFP* with either a cytosolic or nuclear cell extract; then, the resultant product was subjected to a proteomic search employing sequential window acquisition of all theoretical fragment-ion spectra-mass spectrometry (SWATH-MS; Fig 4A and Appendix Table S2). A specific interaction of BrU-labeled *MYMLR* with the top-ranked poly-C-binding protein 2 (PCBP2), but not another RNA-binding protein hnRNPH, was confirmed by Western blot (WB) analysis using an anti-BrU antibody (Figs 4B and EV4B). *MYMLR* preferentially bound to PCBP2c and PCBP2g, which were enriched in nuclear fraction (Fig EV4C), but not to PCBP2a or PCBP2b. Direct interaction of *MYMLR* with PCBP2c was verified by GST pulldown assay findings obtained using *in vitro*-synthesized *MYMLR* and recombinant GST-tagged PCBP2a,

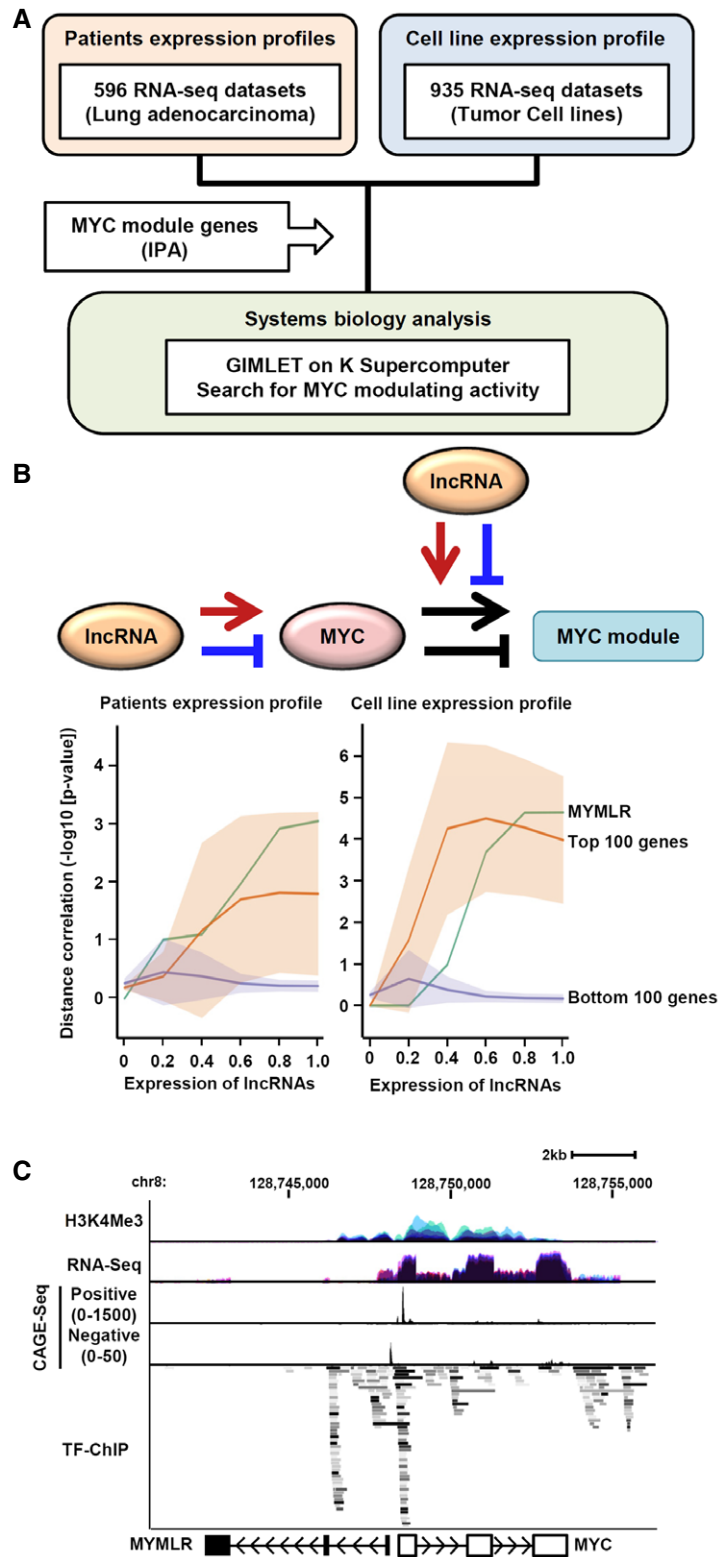


Figure 1. MYMLR lncRNA affecting MYC transcriptional activity was identified through a systems biology-based approach.

- A Schematic diagram to search for lncRNAs affecting MYC activity in lung adenocarcinoma specimens and tumor cell lines.
 B Results of GIMLET analysis based on distance correlation. Green, MYMLR; orange, average of top 100 genes; purple, average of bottom 100 genes.
 C Transcription-related information for a genomic region surrounding MYMLR and MYC.

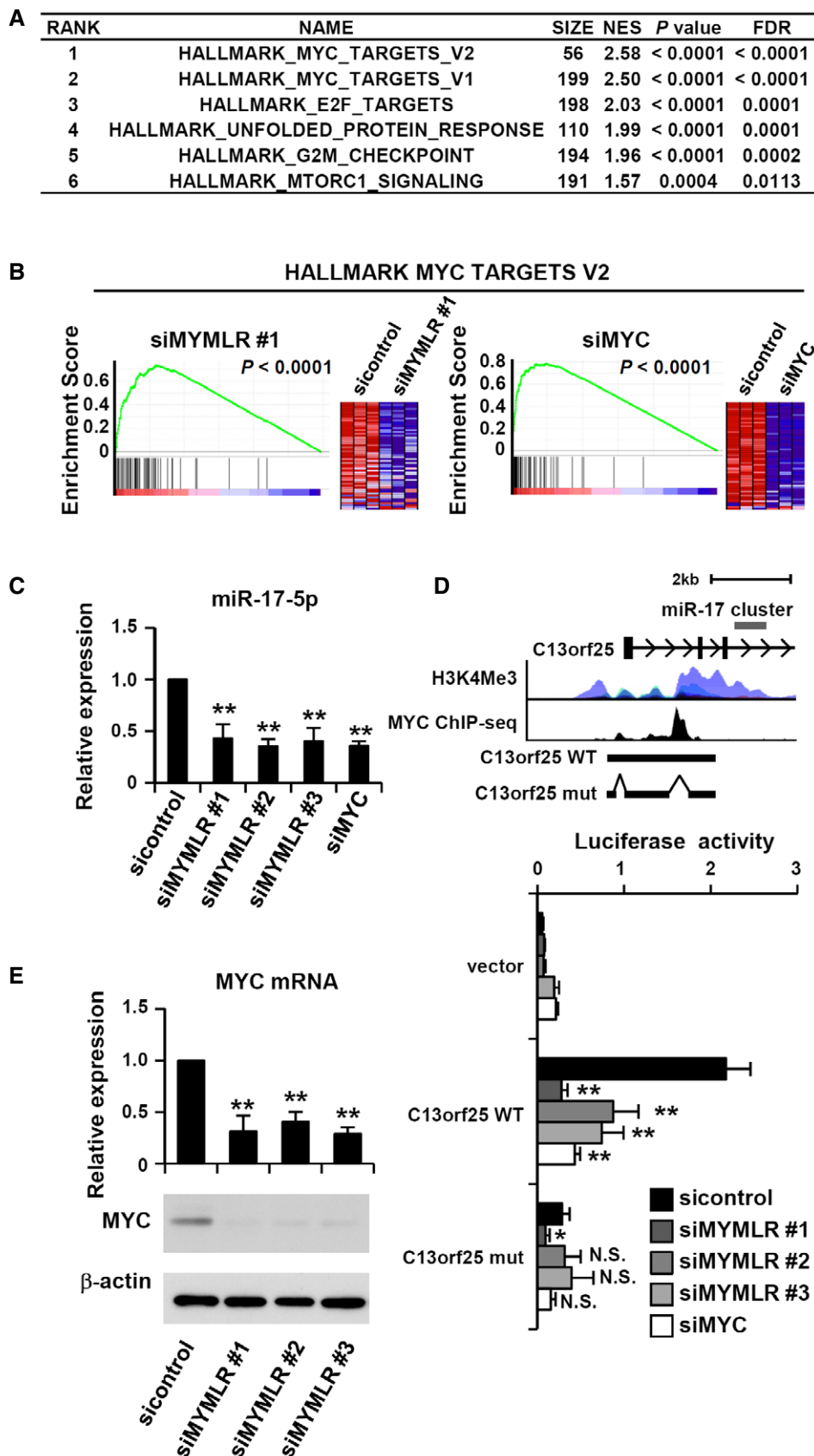


Figure 2.

Figure 2. MYMLR sustains MYC expression and its transcriptional activity.

- A Gene sets with a significant positive association with *MYMLR* in siMYMLR #1-treated A549 cells. *P*-value was calculated by permutation of genes ($n = 3$).
- B Representative gene sets significantly affected by introduction of either siMYMLR or siMYC. *P*-value was calculated by permutation of genes ($n = 3$).
- C qRT-PCR measurement results indicating reduced *miR-17-5p* expression by siMYMLR treatment. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $**P < 0.01$).
- D Luciferase assays of *C13orf25* promoter reporters carrying the wild type or deletion of MYC-binding regions showed suppression of *C13orf25* promoter activity following siMYMLR introduction. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $*P < 0.05$, $**P < 0.01$, N.S.; not significant).
- E qRT-PCR and Western blotting analysis of siMYMLR-mediated effects on MYC expression in A549 cells. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $**P < 0.01$).

Source data are available online for this figure.

PCBP2b, and PCBP2c (Fig EV4D). We further validated the interaction between *MYMLR* and PCBP2 using an RNA-immunoprecipitation (RIP) assay with an antibody against PCBP2 (Fig 4C), as well as an RIP assay of A549 cells transfected with PCBP2c (Fig EV4E). Furthermore, we investigated whether PCBP2 interacts with a specific region of *MYMLR* using RNA pulldown assays with various *MYMLR* deletion mutants. Consequently, the 5', but not the 3', poly-C motif of *MYMLR* was found to be required for its binding with PCBP2 (Fig 4D).

Next, we investigated whether *PCBP2* knockdown could phenocopy siMYMLR treatment with regard to MYC regulation, and observed a marked reduction in both MYC mRNA and protein in A549 cells (Fig 4E). Consequently, a significant reduction in *C13orf25* promoter activity was observed in siPCBP2-treated A549 cells (Fig 4F), while GSEA analysis clearly showed that MYC-targeted gene sets, such as HALLMARK MYC TARGETS V2 and MYC UP.V1 UP, were significantly enriched in response to *PCBP2* knockdown (Figs 4G and EV4F). Furthermore, *PCBP2* knockdown also resulted in impaired cell cycle progression (Figs 4H and EV4G).

MYMLR regulates MYC expression by binding to enhancer

To elucidate how *MYMLR* regulates MYC expression, we first examined MYC mRNA stability in siMYMLR-treated A549 cells and found no effects on the half-lives of MYC transcripts (Fig 5A). Next, MYC translational regulation was examined using a luciferase construct harboring MYC 5'-UTR upstream of luciferase cDNA, which showed a negligible decrease in response to siMYMLR treatment. In contrast, siPCBP2 clearly attenuated luciferase activity (Appendix Fig S3). We then performed RNA pulse-labeling analysis and observed that *MYMLR* knockdown significantly decreased incorporation of 5-ethynyl uridine into MYC mRNA (Fig 5B), indicating that *MYMLR* regulates *de novo* MYC mRNA synthesis. Since MYC expression is tightly controlled in part through engagement of multiple enhancers, we speculated that *MYMLR* might regulate MYC transcription in cooperation with such an enhancer(s). To examine this, we utilized

data obtained from ENCODE to select potential enhancers within a 600-kb region upstream of the MYC promoter based on the states of H3K4Me1 and H3K27Ac, enrichment of CTCF, and DNase hypersensitivity (Fig 5C). A chromatin isolation by RNA purification (ChIRP) assay was performed to investigate whether *MYMLR* binds to any of the potential enhancers, which showed that *MYMLR* was specifically purified with genomic DNA encompassing enhancer region 1 as well as that of the *MYMLR* locus itself (Fig 5D). It was also notable that knockdown of *PCBP2* significantly reduced *MYMLR* binding with enhancer region 1, whereas it did not affect binding between *MYMLR* and its own locus. In order to evaluate the chromatin configuration involving *MYMLR* and enhancer region 1, we conducted chromosome conformation capture (3C) analysis with a restriction enzyme XbaI (Appendix Fig S4A). Consequently, all possible combinations of re-ligated genomic fragments were formed between enhancer region 1 and the *MYMLR* locus (Fig 5E and Appendix Fig S4B). We also noted that treatment with siMYMLR or siPCBP2 abolished 3C product formation (Fig 5F). Together, the present findings indicate that *MYMLR*, a divergent lncRNA, transcriptionally regulates MYC by placing an MYC enhancer in proximity to the MYC promoter through direct interaction with PCBP2.

Discussion

MYC expression is tightly regulated to maintain robust control of cell cycle progression, cellular metabolism and proliferation, and malignant transformation (Dang, 2012; Kress et al, 2015). Dysregulation of MYC has been implicated in a number of human cancers including lung cancer. In the present study, we took an integrative approach by combining both systems biology and multi-faceted molecular biologic analyses, which led us to identify *MYMLR* as a lncRNA that sustains MYC expression.

A number of studies have described detailed molecular mechanisms related to MYC regulation and accumulated evidence points to involvement of non-coding RNAs in this regard. *miR-34a* and

Figure 3. MYMLR is required for cell cycle progression.

- A Flow cytometric analysis after release from thymidine-aphidicolin double-block, showing retarded cell cycle progression by introduction of siMYMLR or siMYC. PI, propidium iodide; pHH3, phospho-histone H3. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 4$, $**P < 0.01$).
- B Colorimetric assay findings showing reduced cell growth in response to treatment with siMYMLR or siMYC. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $*P < 0.05$, $**P < 0.01$).
- C Colony formation assay findings showing reduced number of colonies by introduction of siMYMLR or siMYC. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $*P < 0.05$, $**P < 0.01$).
- D Decreased *in vivo* tumor growth by siMYMLR-introduced A549 cells. *P*-value was calculated by Student's *t*-test ($n = 5$, $P = 0.019$).

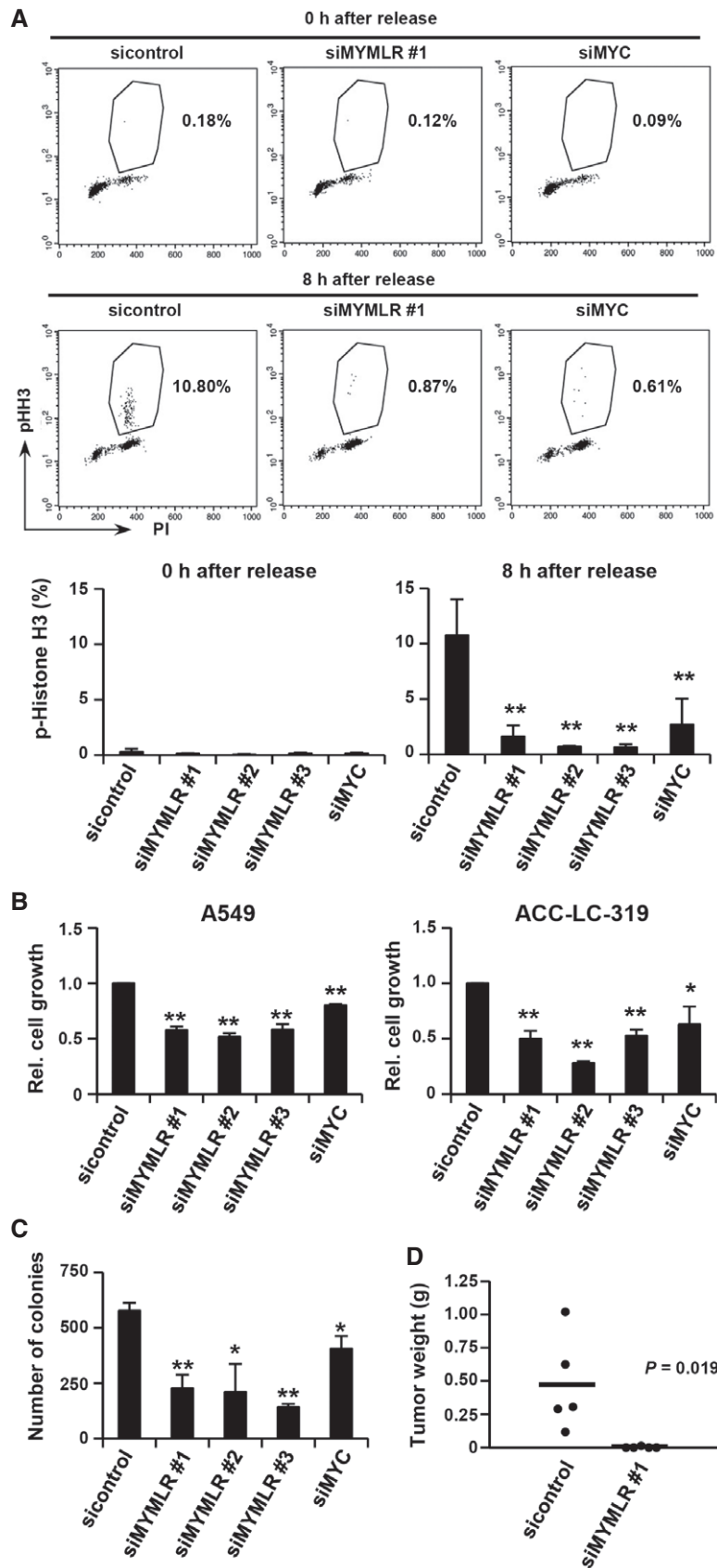


Figure 3.

Figure 4. PCBP2 was identified as a crucial MYMLR-binding protein.

- A Schematic diagram of RNA pulldown experiment to identify MYMLR-associated proteins. IVT-RNA, *in vitro*-transcribed RNA; BrU, 5-Bromo-UTP.
- B RNA pulldown assay findings showing specific association of PCBP2 with MYMLR.
- C RNA-immunoprecipitation (RIP) results showing endogenous association between PCBP2 and MYMLR. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $**P < 0.01$).
- D RNA pulldown assay findings showing MYMLR binding with PCBP2 through its 5' poly-C region.
- E qRT-PCR and Western blot analysis results showing siPCBP2-mediated suppression of MYC expression at both mRNA and protein levels. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $**P < 0.01$, N.S.; not significant).
- F Luciferase assay findings showing reduction of *C13orf25* promoter activity by siPCBP2 treatment. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $**P < 0.01$).
- G Gene set enrichment analysis results showing significant effect of siPCBP2 introduction on MYC-related gene sets. *P*-value was calculated by permutation of genes ($n = 3$).
- H Flow cytometric analysis results showing retarded cell cycle progression by introduction of siPCBP2. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $*P < 0.05$).

Data information: Solid arrowhead, PCBP2a and PCBP2b; solid arrow, PCBP2c; open arrowhead, PCBP2g.

Source data are available online for this figure.

let-7 were found to directly repress MYC expression by binding to their target sites at the 3' untranslated region of MYC mRNA (Sampson *et al*, 2007; Kong *et al*, 2008), while we recently reported that MYC activity is indirectly regulated by *miR-342-3p* via direct repression of *E2F1* (Tai *et al*, 2015). As for lncRNA, *PVT1* residing 60 kb downstream of MYC protects MYC protein from degradation (Tseng *et al*, 2014). *CCAT1/CARLo-5* is transcribed from a colorectal cancer-specific super enhancer region 515 kb upstream of MYC and promotes chromatin looping between an enhancer 335 kb upstream of MYC and the MYC promoter (Xiang *et al*, 2014). Furthermore, *PCAT1* residing 725 kb upstream of the MYC promoter was shown to post-transcriptionally induce MYC expression by acting as a sponge for *miR-34a* in prostate cancer (Prensner *et al*, 2014). Notably, MYMLR is distinct from these previously reported lncRNAs in that it is transcribed as a divergent lncRNA arising from a point only 350 bp upstream of the MYC promoter, and elicits long-range genome looping between the 557 kb upstream enhancer and MYC promoter (Fig 5G). Recently, Zhang *et al* reported that *CASC11*, which shares exons 2 and 3 of MYMLR but is transcribed from a distinct promoter, interacts with hnRNP-K in cytosol and stabilizes *hnRNP-K* mRNA, which has effects on various genes in the Wnt signaling pathway in colorectal cancer cell lines. On the other hand, we did not observe similar effects of MYMLR toward Wnt pathway genes, suggesting the functional complexity of this locus.

It is well known that lncRNAs often function through physical interaction with their protein-binding partners. For example, *HOTAIR* forms a complex with PRC2 and catalyzes H3K27 trimethylation, leading to epigenetic gene silencing (Rinn *et al*, 2007), while

cancer-related lncRNAs including *NBAT1* and *PINT* regulate gene expression through their interactions with PRC2 (Marin-Bejar *et al*, 2013; Pandey *et al*, 2014). In the present study, we showed that MYMLR sustains *de novo* synthesis of MYC mRNA in *cis* through its physical interaction with PCBP2, a member of the hnRNP-K family. PCBP2 is known to play multiple roles in mRNA processing, stabilization, and translation via interactions with single-stranded poly-C motifs in target mRNAs, including an IRES sequence in the 5' untranslated region of MYC mRNA (Choi *et al*, 2009). Although MYMLR and PCBP2 reside in both nuclei and cytoplasm, our findings indicate that MYMLR is not involved in translational regulation of MYC, in contrast to PCBP2. Instead, it binds to the MYC enhancer in a PCBP2-dependent manner, and both MYMLR and PCBP2 cooperatively regulate MYC expression at the transcriptional level by forming a looping structure between the MYC promoter and enhancer.

Interestingly, the present ChIRP assay results revealed that binding of MYMLR to an enhancer region 557 kb upstream of MYC requires its binding partner PCBP2, whereas the MYMLR locus itself is constitutively associated with MYMLR in a PCBP2-independent manner. Furthermore, 3C assay findings clearly demonstrated that the MYC promoter and upstream enhancer are placed together in close proximity by formation of a looped structure, which occurs in a MYMLR- and PCBP2-dependent manner. Along this line, a previous study found that *HOTTIP* lncRNA is transcribed from upstream of the *HOXA* locus and regulates *HOXA* cluster genes, despite its low level of expression at < 1 copy per cell, via recruitment of the MLL/WDR5 complex and consequential chromosomal looping (Wang *et al*, 2011). It is not inconceivable that MYMLR sustains

Figure 5. MYMLR transcriptionally regulates MYC expression.

- A Half-lives of MYC mRNA quantified by use of actinomycin D. Data represent the mean \pm SD ($n = 3$).
- B Decreased nascent MYC mRNA synthesis in siMYMLR-treated A549 cells. Data represent the mean \pm SD. *P*-value was calculated by Student's *t*-test ($n = 3$, $*P < 0.05$, $**P < 0.01$).
- C ENCODE ChIP-seq profiles and DNase-seq information for 8q24 harboring MYMLR and MYC from the UCSC genome browser. Open squares, ChIRP-PCR amplicons.
- D ChIRP-PCR analysis results of MYMLR binding to MYC enhancers in A549 cells. C, control probes; M, MYMLR probes.
- E Chromatin conformation capture (3C) assay findings showing association between MYMLR and genomic region 557 kb upstream of MYC.
- F Findings of 3C assay showing that the stem-loop structure formation between 5' of MYC enhancer locus and 3' of MYMLR locus was abolished by treatment with siMYMLR or siPCBP2. An internal region of the MYC promoter served as a loading control.
- G Schematic diagram of a model demonstrating how MYMLR functions to sustain MYC expression.

Source data are available online for this figure.

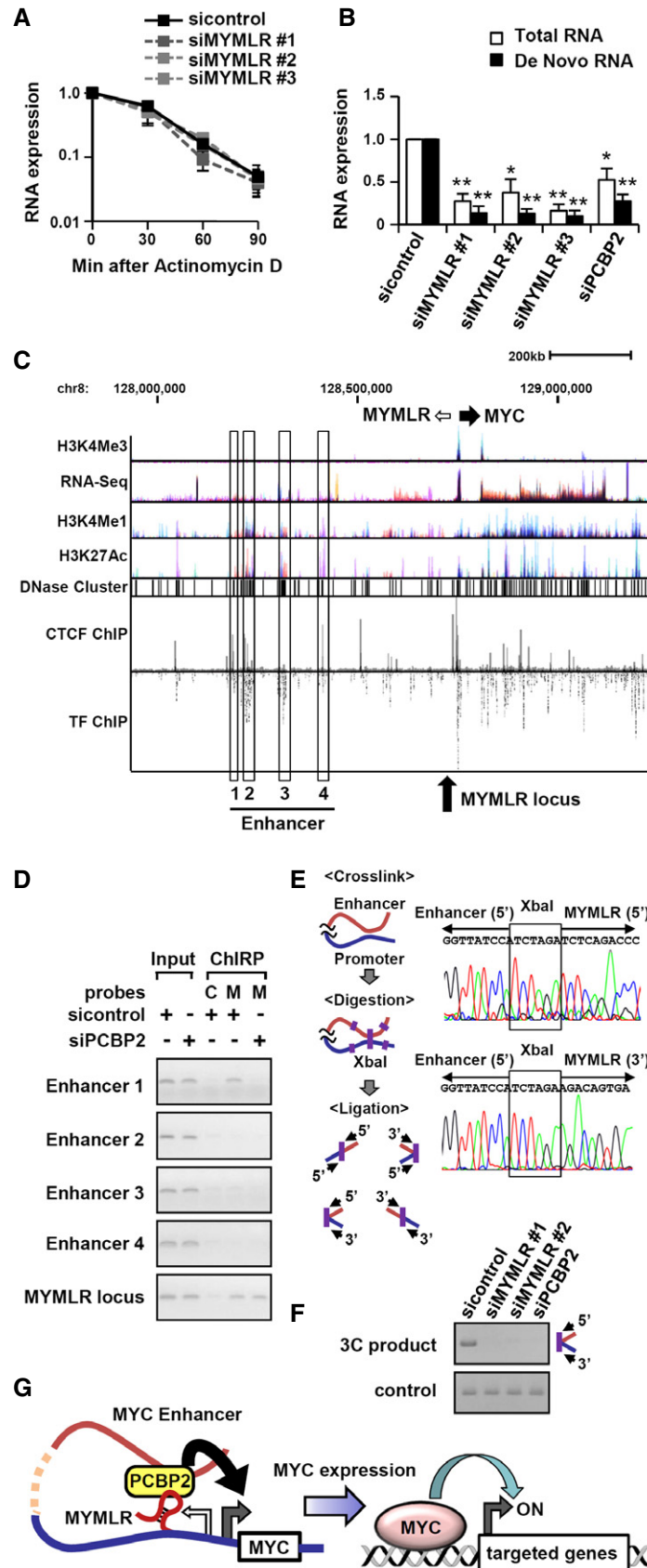


Figure 5.

MYC transcription in *cis* without the problem of diffusion, which may in turn explain why lncRNA with such a low copy number can regulate abundant *MYC* expression. In addition, a precedent example of an extremely low level of lncRNA *VELUCT*, which regulates the viability of lung cancer cell lines with as yet unidentified underlying molecular mechanism, has been reported (Seiler *et al*, 2017). We also noted that *MYMLR* expression is negatively regulated by *MYC* itself (unpublished observation), suggesting the existence of a negative feedback loop involving *MYC* and *MYMLR*, which possibly limits excessive *MYC* activation.

The present findings revealed crucial involvement of the low copy number lncRNA *MYMLR* in *MYC* transcription, implying that the underlying regulatory mechanisms of this archetypical oncogene have not been fully elucidated. Considering its low level of expression, *MYMLR* may be an attractive therapeutic target for repressing abundant *MYC* expression, which is known to be involved in cancer development. This study employed both systems biology and multi-faceted experimental analyses, and the findings obtained implicate the fruitfulness of such an integrative approach, in contrast to the nucleotide sequence of an lncRNA, which does not provide enough information to elucidate its molecular functions. While tens of thousands of lncRNAs are proposed to exist in the human genome, the present results clearly demonstrate a crucial role for *MYMLR* despite its very low level of expression. It is anticipated that future studies of lncRNA-mediated regulation of other crucial oncogenes and tumor suppressor genes will provide a foundation to increase understanding of the molecular processes of cancer development, in which low copy number lncRNAs must be taken into consideration.

Materials and Methods

Cell lines and chemical agents

The A549 lung adenocarcinoma cell line was purchased from American Type Culture Collection, while the ACC-LC-319 lung adenocarcinoma cell line was established by our group (Hida *et al*, 1993). Both were maintained in RPMI 1640 supplemented with 10% fetal bovine serum and confirmed to be free from mycoplasma contamination (MycAlert; Lonza). Verification of A549 and ACC-LC-319 was performed by short tandem repeat profiling in February 2015 and December 2016, respectively, at the Japanese Collection of Research Bioresources, National Institute of Biomedical Innovation of Japan. Actinomycin D was purchased from Sigma.

Plasmids and siRNAs

To generate a luciferase reporter construct carrying the promoter of *C13orf25*, a 2,977-bp fragment of that promoter was amplified with human genomic DNA (Promega), followed by cloning into a pGL4.10 vector (Promega). *MYMLR* cDNA was generated by reverse transcription-PCR and cloned into a pCMV-puro vector. Full-length human *PCBP2* cDNA (OriGene Technologies, MDs) was inserted into a pCMV-puro or pDEST20 (Thermo Scientific) vector. A pDEST53 vector (Thermo Scientific) was used to synthesize *GFP* RNA. *MYMLR* deletion constructs were generated by *in vitro* mutagenesis using KOD-plus-DNA polymerase (TOYOBO) and

thoroughly sequenced. Full-length human *MYC* cDNA was excised from pMXs-hc-myc (Addgene) and inserted into a pCMV-puro vector. *MYC* 5'-UTR was cloned into pGL3 vector. siMYMLR #1 and #2, siPCBP2, and sicontrol were purchased from Invitrogen. siMYMLR #3 and siMYC were purchased from Thermo Scientific Dharmacon. siRNAs were transfected at 20 nM into 3×10^4 cells per well of a 6-well plate using RNAiMAX (Thermo Scientific). Antisense oligonucleotides were introduced at 800 pmol per 1×10^6 cells using Amaxa Nucleofector (Lonza), according to the manufacturer's instructions. The sequences of all primers and siRNAs used are listed in Appendix Table S3.

RNA-seq data processing, systems biologic, and bioinformatic analyses

Raw RNA-seq datasets consisting of 596 lung adenocarcinoma patients of The Cancer Genome Atlas (TCGA) cohort and 935 cancer cell lines of the Cancer Cell Line Encyclopedia (CCLE) project were downloaded from the NCI Genomic Data Commons (<https://portal.gdc.cancer.gov/>) and then aligned using Salmon 0.4.2 (Patro *et al*, 2017). To identify cancer-associated lncRNAs affecting *MYC* activity, we applied a genome-wide identification of modulators with local energy statistical test (GIMLET), a permutation-based statistical test used for local distance correlation (preprint: Shimamura *et al*, 2018), to all of the processed RNA-seq data of 58,581 transcripts using the K supercomputer. Briefly, GIMLET is used to measure changes in energy statistics between *MYC* and its targets conditioned to the gene expression of the modulator (see Appendix Supplementary Methods). We included all data from the 58,581 processed transcripts as a candidate modulator and the *MYC* module was defined based on 246 target genes of *MYC* in the Ingenuity Knowledge Base (<http://www.ingenuity.com/science/knowledgebase/>, Appendix Table S4), as well as an experimentally defined *MYC*-related gene set (Liberzon *et al*, 2015; Tai *et al*, 2015). The analysis pipeline had three steps: (i) test whether the mean level of modulator expression in a lung adenocarcinoma was greater than that in normal lung tissues using Welch's *t*-test with the TCGA dataset, (ii) test whether the *MYC*-target energy statistic conditioned in the presence/absence of the modulator is greater than would be expected using the GIMLET algorithm, and (iii) test whether the *MYC*-target energy statistic changes when in the presence/absence of the modulator using the GIMLET algorithm. Gene set enrichment analysis (GSEA) and pathway enrichment analysis were also performed, as described in detail in the accompanying Appendix Supplementary Methods section.

Microarray analysis and qRT-PCR

Microarray analysis was conducted using a SurePrint G3 Human Gene Expression 8x60 K Microarray Kit version 2.0 and GeneSpring version 12.6 (Agilent), as previously described (Griesing *et al*, 2017). Quantitative reverse transcription-PCR (RT-PCR) analyses of individual RNAs were performed using the primers listed in Appendix Table S3, along with Power SYBR Green PCR Master Mix (Thermo Scientific) and an ABI Prism 7500 (Thermo Scientific). The expression levels were calculated using a standard curve method and normalized by the expression of *18S*. qRT-PCR analyses of miRNAs were performed with a TaqMan MicroRNA Reverse

Transcription Kit and TaqMan Universal Master Mix II (Thermo Scientific), according to the manufacturer's protocol. Data obtained were normalized to *RNA44*. RNA copy numbers were estimated by comparing with *in vitro*-synthesized RNA.

Rapid amplification of cDNA ends (RACE)

We performed the 5'- and 3'-RACE with a SMARTer RACE Kit (Clontech), according to the manufacturer's instructions. First, RNA was converted to cDNA using oligo (dT) primer; then, the 3'-UTR sequences were amplified with universal and *MYMLR*-specific primers, which are listed in Appendix Table S3.

Dual-luciferase reporter assay

A549 cells were transfected with an siRNA at 24 h prior to co-introduction of a luciferase reporter containing the *C13orf25* promoter with a pGL4.74-Renilla vector. A mutant vector lacking the MYC-binding regions was constructed using a QuikChange Site-Directed Mutagenesis Kit (Stratagene). Assays were performed at 48 h after luciferase reporter transfection using a Dual-Luciferase Reporter Assay (Promega). Firefly luciferase activity was normalized to that of Renilla. Each assay was performed in triplicate.

Western blot analysis

Western blot analysis was performed according to standard procedures using Immobilon-P filters (Millipore) and an Enhanced Chemiluminescence system (GE Healthcare). The primary antibodies used were anti-c-myc (9E10, Santa Cruz), anti- β -actin (Sigma), anti-PCBP2 (RN025P, MBL, Japan), anti-hnRNP-H (A300-511A, Bethyl), anti-GST (M071-3, MBL, Japan), and anti-Flag (M2, Sigma).

Flow cytometric analysis

To evaluate the effects on cell cycle progression, cells were transfected with sicontrol for 5 h and incubated with 2 mM of thymidine for 24 h and then released from S-phase arrest by replacing with fresh medium for 10 h. When treated with siRNAs against *MYMLR*, *MYC*, or *PCBP2*, cells were immediately treated with thymidine for 24 h and then incubated in fresh medium for 14 h to compensate for delayed cell cycle progression. Aphidicolin was then added at 5 μ M for the second block for 14 h. Cells were then released from the complete G1/S-phase block and harvested by trypsinization at 0 or 8 h later. All cells were collected, fixed with 70% ethanol and 30% PBS at -20°C and then immunostained with Alexa-647-conjugated anti-phospho-histone H3 (pS28; 558217, Becton Dickinson), followed by staining with propidium iodide (PI) in PBS containing Triton X-100 and RNase A. Phospho-histone H3 levels and DNA contents of individual cells were analyzed using a FACScalibur system (Becton Dickinson).

Colorimetric and colony formation assays

For colorimetric assays, 3×10^4 cells per well were plated in 6-well plates and transfected with siRNA; then, the number of viable cells was assessed 72 h later using a cell counting kit-8 (Dojindo, Japan), according to the manufacturer's instructions. A total of three

independent experiments were performed. For colony formation assays, 1,000 cells were plated in 10-cm dishes and transfected with siRNA; then, 14 days later the resultant colonies were fixed with formalin and subjected to Giemsa staining (Sigma). The number of colonies was quantified using ImageJ software 1.46r (National Institutes of Health, US). All assays were performed in triplicate with three independent transfections.

In vivo tumorigenicity assay

A549 cells were transfected with either siMYMLR or sicontrol, and then incubated for 72 h. After performing trypsinization, 3×10^5 cells were inoculated into a lower flank of 8-week-old athymic nude mice (Japan SLC, Japan). Tumor weights were measured 5 weeks after inoculation.

RNA pulldown assay and subcellular fractionation

BrU-labeled RNA was *in vitro* transcribed using a CUGA7 *in vitro* Transcription Kit (Nippon Gene, Japan); then, the samples were treated with DNase and purified using a mini Quick Spin RNA Column (Sigma). Subcellular fractionation and RNA pulldown assays were performed with a RiboTrap Kit (MBL, Japan), according to the manufacturer's instructions. Briefly, A549 cells were trypsinized, and cytosolic and nuclear fractions were isolated, and then, 600 μ g of cytosolic/nuclear lysates was mixed with 50 pmol of BrU-labeled RNAs and incubated at 4°C for 1 h. BrU-labeled RNA was immunoprecipitated with an anti-BrdU antibody, and RNA-binding proteins were eluted by addition of competing BrU.

Sequential window acquisition of all theoretical mass spectrometry (SWATH-MS)

SWATH-MS analysis was performed using a 5600 Triple-TOF mass spectrometer (ABSciex) equipped with Protein Pilot software version 4.0 (ABSciex) and PeakView software version 1.0 (ABSciex), which is a label-free quantification method used for mass spectrometry. The detailed experimental procedures are provided in the Appendix Supplementary Methods.

RNA immunoprecipitation (RIP)

A549 cells were lysed in NP-40 lysis buffer [20 mM Tris-HCl (pH 8.0), 137 mM NaCl, 2 mM EDTA, 1% NP-40, 10% glycerol, 1 mM Na_3VO_4] and incubated with 15 μ g of antibodies at 4°C for 2 h. The RNA/antibody complex was precipitated by incubation with protein A agarose beads. After washing four times, QIAzol (QIAGEN) was added and RNA was extracted using an miRNeasy Mini Kit (QIAGEN), according to the manufacturer's instructions.

In vitro-binding assay

GST-tagged PCBP2 was expressed in Sf9 insect cells using a Gateway system (Thermo Scientific), according to the manufacturer's instructions. Recombinant GST-tagged PCBP2 proteins were purified by glutathione-affinity chromatography. A total of 100 ng of *in vitro*-synthesized wild-type and various deletion mutants of *MYMLR* or *GFP* were incubated with 500 ng of GST-PCBP2 or GST

protein-bead complexes at room temperature for 30 min in PBS containing 2 mM MgCl₂, 0.2 mM ZnCl₂, 1 mM DTT, 100 U/ml RNase inhibitor, 0.1 mg/ml yeast tRNA (Sigma), 0.05% BSA, and 0.2% NP-40. The beads were washed five times with the same incubation buffer supplemented with an additional 150 mM of NaCl (total 300 mM NaCl) at room temperature.

Labeling of nascent RNA in live cells

Labeling of nascent RNA was performed using a Click-iT Nascent RNA Capture Kit (Thermo Scientific), following the manufacturer's instructions. Briefly, siRNA-transfected A549 cells were incubated with 0.2 mM 5-ethynyl uridine for 24 h. RNA was extracted using QIAzol (QIAGEN) then 5 µg of 5-ethynyl uridine RNA was used for biotinylation, followed by pulldown with streptavidin beads. Biotinylated RNA bound to the beads was serially washed and reverse-transcribed for quantification.

Chromatin isolation by RNA purification (ChIRP) assay

ChIRP assays were performed using an EZ-Magna ChIRP RNA Interactome Kit (Millipore). Briefly, antisense DNA probes against *MYMLR* were designed (Appendix Table S3) and 3'-end biotin-triethylene glycol (TEG)-modified oligonucleotide probes were synthesized by NIPPON GENE (Toyama, Japan). A549 cells were trypsinized and fixed with 1% glutaraldehyde for 2 min, and cross-linking was stopped by quenching with glycine at room temperature for 5 min. The cross-linked cells were then lysed and sonicated using a Bioruptor (Diagenode) at 4°C on the high setting with pulse intervals of 10-s on and 20-s off for a total of 90 min. Sonicated cell lysates were hybridized with a mixture of biotin-TEG-modified oligonucleotide probes against *MYMLR* in hybridization buffer at 37°C for 4 h, and then, the binding complexes were recovered by streptavidin-conjugated magnet beads and eluted with elution buffer. Bound DNA was PCR-amplified using the primers listed in Appendix Table S3.

Chromatin conformation capture (3C) assays

We performed 3C assays according to a protocol previously reported (Dekker *et al*, 2002) with minor modifications that are described in detail in the Appendix Supplementary Methods section. Briefly, A549 cells were trypsinized and fixed with 1% formaldehyde for 5 min at room temperature, and then, cross-linking was stopped by quenching with 0.125 M glycine on ice for 10 min. Nuclei were isolated, and chromatin was released and digested with 400 U of XbaI (Sigma) at 37°C overnight. After heat inactivation, chromatin ligation was performed by incubating with 100 U of T4 ligase (Promega) for 4 h at 16°C, followed by incubation with Proteinase K (200 µg) overnight at 65°C to reverse the cross-links. Following RNase A (Invitrogen) digestion, DNA was purified, PCR-amplified, and sequenced using the primers listed in Appendix Table S3.

Statistical analysis

The values are expressed as mean ± SD. The error bars (SD) shown for all results were derived from biological replicates, not technical replicates. Significant differences between two groups were

evaluated using a two-tailed, unpaired Student's *t*-test. Statistical significance of gene expression was determined by permutation tests.

Data availability

All microarray data obtained in this study are available at the Gene Expression Omnibus under the following accession number: GSE103085 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103085>). The *MYMLR* sequence information has been deposited in GenBank under the following accession number: MH782193.

Expanded View for this article is available online.

Acknowledgements

The authors thank Naoe Hotta for providing technical assistance and Ayumu Taguchi for helpful discussion during the process of revision. This work was supported in part by a Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (15H05910), Grants-in-Aid for Scientific Research (A) (16H02468) and for Young Scientists (B) (16K19050) from the Japan Society for the Promotion of Science, a grant from the Princess Takamatsu Cancer Research Fund, and Japan Agency for Medical Research and development (AMED, 18ck0106363h0002).

Author contributions

TK and TT conceived and designed the study. TK, SGo, KY, LI, SGr, YS, and KK performed the molecular biological and biochemical experiments. TK, KY, MS, and TT analyzed the data. TS, MN, and SM performed the systems biological and bioinformatic analyses. TK, TS, and TT wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Cancer Genome Atlas Research N (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511: 543–550
- Choi HS, Hwang CK, Song KY, Law PY, Wei LN, Loh HH (2009) Poly(C)-binding proteins as transcriptional regulators of gene expression. *Biochem Biophys Res Commun* 380: 431–436
- Dang CV (2012) MYC on the path to cancer. *Cell* 149: 22–35
- Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295: 1306–1311
- Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R *et al* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447: 1087–1093
- Fatica A, Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* 15: 7–21
- Griesing S, Kajino T, Tai MC, Liu Z, Nakatochi M, Shimada Y, Suzuki M, Takahashi T (2017) Thyroid transcription factor-1-regulated microRNA-532-5p targets KRAS and MKL2 oncogenes and induces apoptosis in lung adenocarcinoma cells. *Cancer Sci* 108: 1394–1404
- Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ *et al* (2007) Multiple

- regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39: 638–644
- Hida T, Ariyoshi Y, Kuwabara M, Sugiura T, Takahashi T, Takahashi T, Hosoda K, Niitsu Y, Ueda R (1993) Glutathione S-transferase pi levels in a panel of lung cancer cell lines and its relation to chemo-radiosensitivity. *Jap J Clin Oncol* 23: 14–19
- Kong YW, Cannell IG, de Moor CH, Hill K, Garside PG, Hamilton TL, Meijer HA, Dobbyn HC, Stoneley M, Spriggs KA et al (2008) The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. *Proc Natl Acad Sci USA* 105: 8866–8871
- Kress TR, Sabo A, Amati B (2015) MYC: connecting selective transcriptional control to global RNA production. *Nat Rev Cancer* 15: 593–607
- Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P (2015) The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* 1: 417–425
- Little CD, Nau MM, Carney DN, Gazdar AF, Minna JD (1983) Amplification and expression of the c-myc oncogene in human lung cancer cell lines. *Nature* 306: 194–196
- Marin-Bejar O, Marchese FP, Athie A, Sanchez Y, Gonzalez J, Segura V, Huang L, Moreno I, Navarro A, Monzo M et al (2013) Pint lincRNA connects the p53 pathway with epigenetic silencing by the Polycomb repressive complex 2. *Genome Biol* 14: R104
- Pandey GK, Mitra S, Subhash S, Hertwig F, Kanduri M, Mishra K, Fransson S, Ganeshram A, Mondal T, Bandaru S et al (2014) The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell* 26: 722–737
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14: 417–419
- Prensner JR, Chen W, Han S, Iyer MK, Cao Q, Kothari V, Evans JR, Knudsen KE, Paulsen MT, Ljungman M et al (2014) The long non-coding RNA PCAT-1 promotes prostate cancer cell proliferation through cMyc. *Neoplasia* 16: 900–908
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E et al (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129: 1311–1323
- Rinn JL, Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81: 145–166
- Sampson VB, Rong NH, Han J, Yang Q, Aris V, Soteropoulos P, Petrelli NJ, Dunn SP, Krueger LJ (2007) MicroRNA let-7a down-regulates MYC and reverts MYC-induced growth in Burkitt lymphoma cells. *Cancer Res* 67: 9762–9770
- Seiler J, Breinig M, Caudron-Herger M, Polycarpou-Schwarz M, Boutros M, Diederichs S (2017) The lincRNA VELUCT strongly regulates viability of lung cancer cells despite its extremely low abundance. *Nucleic Acids Res* 45: 5458–5469
- Shimamura T, Matsui Y, Kajino T, Ito S, Takahashi T, Miyano S (2018) GIMLET: identifying biological modulators in context-specific gene regulation using local energy statistics. *bioRxiv* <https://doi.org/10.1101/349928> [PREPRINT]
- Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* 15: 272–286
- Tai MC, Kajino T, Nakatochi M, Arima C, Shimada Y, Suzuki M, Miyoshi H, Yatabe Y, Yanagisawa K, Takahashi T (2015) miR-342-3p regulates MYC transcriptional activity via direct repression of E2F1 in human lung cancer. *Carcinogenesis* 36: 1464–1473
- Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W et al (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39: 984–988
- Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, Ronning P, Reuland B, Guenther K, Beadnell TC et al (2014) PVT1 dependence in cancer with MYC copy-number increase. *Nature* 512: 82–86
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA et al (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472: 120–124
- Xiang JF, Yin QF, Chen T, Zhang Y, Zhang XO, Wu Z, Zhang S, Wang HB, Ge J, Lu X et al (2014) Human colorectal cancer-specific CCAT1-L lincRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res* 24: 513–531