

Long-range interaction and correlation between *MYC* enhancer and oncogenic long noncoding RNA *CARLo-5*

Taewan Kim^{a,1}, Ri Cui^{a,b}, Young-Jun Jeon^a, Ji-Hoon Lee^c, Ju Hee Lee^d, Hosung Sim^e, Jong Kook Park^f, Paolo Fadda^a, Esmerina Tilli^{a,g}, Hiroshi Nakanishi^a, Man-Il Huh^h, Sung-Hak Kim^{i,j}, Ju Hwan Cho^k, Bong Hwan Sung^l, Yong Peng^a, Tae Jin Lee^a, Zhenghua Luo^a, Hui-Lung Sun^a, Huijun Wei^a, Hansjuerg Alder^a, Jeong Su Oh^m, Kang Sup Shim^a, Sang-Bong Koⁿ, and Carlo M. Croce^{a,2}

^aDepartment of Molecular Virology, Immunology and Medical Genetics and ⁹Department of Anesthesiology, Wexner Medical Center, and ¹Department of Neurological Surgery, The Ohio State University, Columbus, OH 43210; ^bDepartment of Pathology, Affiliated Hospital of Yanbian University, YanJi 133002, China; ^cDepartment of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75235; ^dDepartment of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064; ^eDepartment of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, MI 48109; ^fDepartment of Dermatology, Northwestern University, Chicago, IL 60208; ^hDepartment of Ophthalmology, School of Medicine, Kyungpook National University, Daegu 700-842, Korea; ⁱSchool of Life Sciences and Biotechnology, Korea University, Seoul 136-702, Korea; ^kDepartment of Radiation Oncology, The Ohio State University Medical Center, Arthur G. James Comprehensive Cancer Center, Columbus, OH 43210; ^lDepartment of Cancer Biology, Vanderbilt University Medical Center, Nashville, TN 37232; ^mDepartment of Genetic Engineering, Sungkyunkwan University, Suwon 110-745, Korea; and ⁿDepartment of Orthopedic Surgery, Daegu Catholic University Medical Center, Catholic University of Daegu School of Medicine, Daegu 705-718, Korea

Contributed by Carlo M. Croce, January 10, 2014 (sent for review December 9, 2013)

The mechanism by which the 8q24 *MYC* enhancer region, including cancer-associated variant rs6983267, increases cancer risk is unknown due to the lack of protein-coding genes at 8q24.21. Here we report the identification of long noncoding RNAs named cancer-associated region long noncoding RNAs (*CARLo*s) in the 8q24 region. The expression of one of the long noncoding RNAs, *CARLo-5*, is significantly correlated with the rs6983267 allele associated with increased cancer susceptibility. We also found the *MYC* enhancer region physically interacts with the active regulatory region of the *CARLo-5* promoter, suggesting long-range interaction of *MYC* enhancer with the *CARLo-5* promoter regulates *CARLo-5* expression. Finally, we demonstrate that *CARLo-5* has a function in cell-cycle regulation and tumor development. Overall, our data provide a key of the mystery of the 8q24 gene desert.

Several genome-wide association (GWA) studies have reported strong associations between the G allele of rs6983267 at 8q24.21 and increased risk of various types of cancers such as colorectal and prostate cancers (1–3). In addition, a recent *in vivo* study shows that deletion of the enhancer region, including rs6983267, results in resistance to intestinal tumors (4). Nevertheless, the biological function of rs6983267 remains unclear because the region is located in a gene desert that does not contain protein-coding genes. Several studies have shown that the region including rs6983267 has enhancer activity and interacts with the proto-oncogene *MYC* involved in early steps of colorectal cancer (CRC)–development (5–7). However, no significant correlation between rs6983267 and *MYC* expression was found in humans (5, 8–11).

Long noncoding RNAs (lncRNAs), nonprotein coding transcripts longer than ~200 bp, have recently attracted interest due to their functions in various cellular processes, development, and human diseases (12). For example, Guttman et al. (13) found that a cohort of lncRNAs controls embryonic stem cell state, and Klattenhoff et al. (14) recently revealed that the lncRNA *braveheart* (*Bvht*) has a role in the establishment of cardiovascular lineage, providing the evidence that lncRNAs are involved in embryo development (13, 14). In addition, it was reported that several lncRNAs including *lincRNA-p21* and *PANDA* are implicated in cellular process such as cell growth and apoptosis (15, 16). Several lncRNAs are known to be implicated in tumor development. The lncRNAs such as *HOTAIR*, *lincRNA-HEIH*, and *lincRNA-LALR1* also have a role in tumor initiation or progression (17–19). Thus, the studies of lncRNAs provide keys of the unsolved questions in current scientific knowledge.

Here, we found that the 8q24 gene desert harbors lncRNAs and that the enhancer region including rs6983267 regulates at least one

of them by directly interacting with the promoter region of the lncRNA. We also report that the lncRNA has a role in cell-cycle regulation and tumor development.

Results and Discussion

Identification of lncRNAs in the 8q24 Gene Desert. Manual annotation performed by the human and vertebrate analysis and annotation (HAVANA) team discovered putative lncRNAs in the human genome that are available in ENSEMBL. Using this information, we found seven lncRNAs that are located in the 8q24.21 gene desert [hereafter referred to as cancer-associated region long noncoding RNA (*CARLo*); Fig. 1*A*, *Top* and *Middle*]. To identify and characterize *CARLo*s, we conducted 5' and 3' RACE (*Materials and Methods* and Fig. 1*B*). All *CARLo*s and their variants were identified by RACE in CRC-derived cells,

Significance

Many cancer-associated variants have been found in the 8q24.21 region harboring enhancer activity. However, the functional mechanism of the variants is not clear due to the lack of protein-coding genes in the region and no significant correlation with the nearest oncogene *MYC*. We identified long noncoding RNAs (lncRNAs) named cancer-associated region long noncoding RNAs (*CARLo*s) in the 8q24.21 region. Interestingly, we found that the cancer-associated variant rs6983267 regulating the enhancer activity is significantly associated with the expression of one of the lncRNAs *CARLo-5* and that *CARLo-5* has an oncogenic function. By showing direct interaction between the enhancer region and active regulatory region of the *CARLo-5* promoter, we provide a regulatory mechanism of cancer susceptibility caused by the cancer-associated variants.

Author contributions: T.K. and C.M.C. designed research; T.K., R.C., Y.-J.J., H.S., P.F., and H.A. performed research; T.K., J.-H.L., E.T., H.N., Y.P., T.J.L., Z.L., H.-L.S., H.W., J.S.O., K.S.S., and S.-B.K. contributed new reagents/analytic tools; T.K., R.C., Y.-J.J., J.-H.L., J.H.L., J.K.P., M.-I.H., S.-H.K., J.H.C., and B.H.S. analyzed data; and T.K. and C.M.C. wrote the paper.

The authors declare no conflict of interest.

Data deposition: Sequence information of *CARLo*s is available in GenBank (accession nos. JX003864, JX003865, JX003866, JX003867, JX003868, JX003869, JX003870, JX003871, JX003872, JX003873, and JX003874). The raw data of NanoString analysis has been deposited in the Gene Expression Omnibus (GEO), www.ncbi.nlm.nih.gov/geo (accession no. GSE48877).

¹Present address: Epithelial Biology Center, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232.

²To whom correspondence should be addressed. E-mail: carlo.croce@osumc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1400350111/-DCSupplemental.

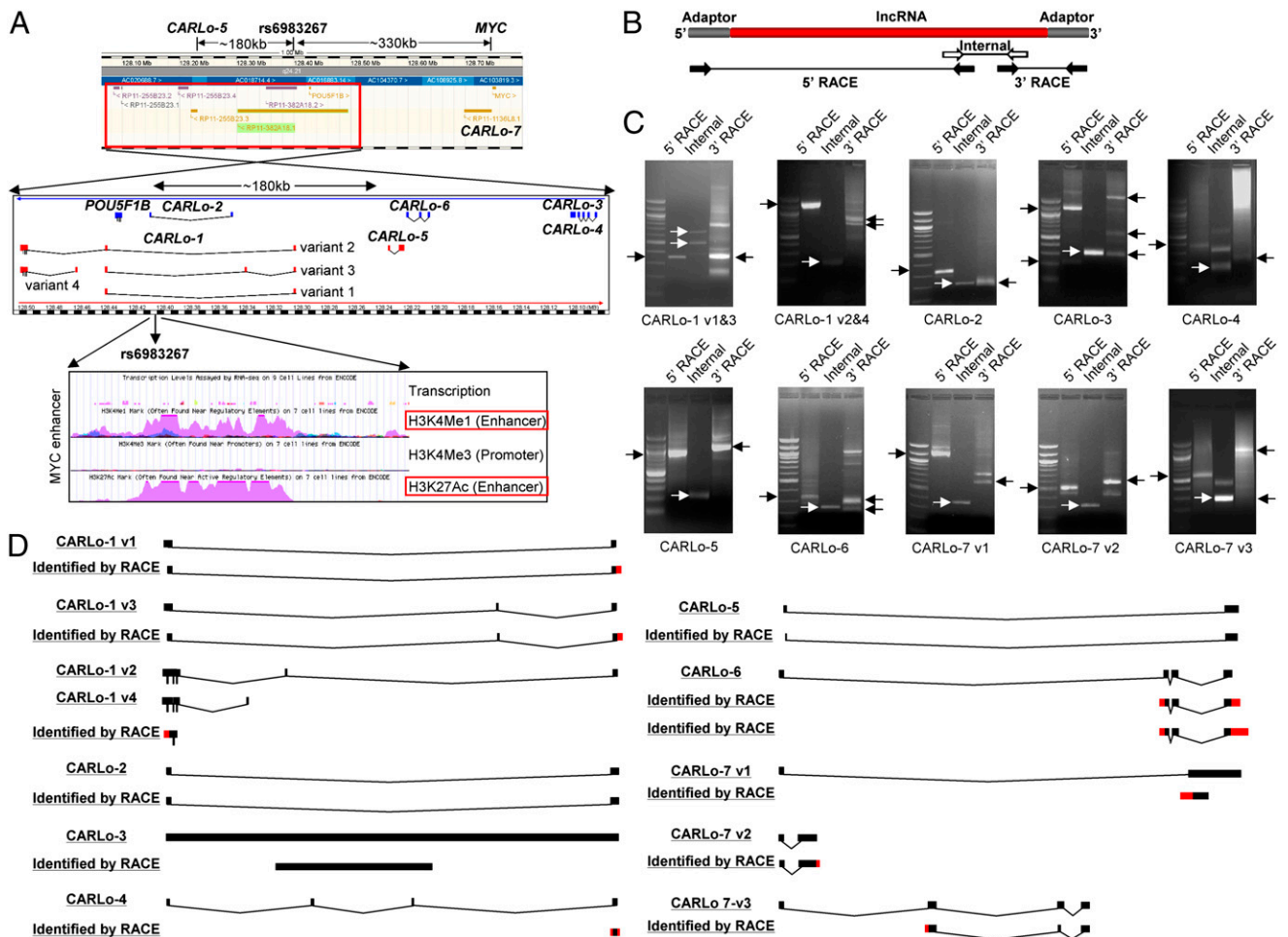


Fig. 1. Identification of lncRNAs in the 8q24 gene desert. (A) Schematic view of the 8q24.21 region showing lncRNAs (*CARLos*) manually annotated by HAVANA. The region including *CARLo-1~6* in the *Top* is symmetrically rotated in the *Middle*. (*Bottom*) ENCODE regulation data in the region including rs6983267. The markers representing regulatory (enhancer) activity are highlighted with red-colored boxes. (B) Schematic view of RACE strategy. The internal region covers both starting points of 5' and 3' RACE. The 5' (*Left* lane) and 3' (*Right* lane) RACE and amplification of the internal region (*Middle* lane) covering both starting points of 5' and 3' RACE. Black arrows indicate appropriate 5' or 3' RACE products, and white arrows indicate PCR products of the internal regions. The appropriate PCR products were identified by direct sequencing. (D) Schematic view of *CARLos* predicted by HAVANA (*Upper*) and identified by RACE and direct sequencing (*Lower*). The regions in red indicate additional sequence identified by RACE.

and their actual sequences were determined by direct sequencing (Fig. 1 C and D). Our findings of lncRNAs demonstrate that the 8q24 gene desert is a gene desert for protein-coding genes, but not for loci-encoding noncoding RNAs including lncRNAs.

It was known that the region including rs6983267 has enhancer activity (5). Using the Encyclopedia of DNA Elements (ENCODE) regulation database available in the University of California Santa Cruz genome browser, we found that markers of regulatory elements such as H3K4Me1 (20) and H3K27Ac (21) are detected at high levels in the region including rs6983267 (Fig. 1A, *Bottom*). On the other hand, promoter marker H3K4Me3 (22) and transcripts are not significantly detected in the region, supporting the previous findings (5). Taken together, these data indicate that the region including rs6983267 has enhancer activity rather than gene-coding activity. Based on the evidence, we hypothesized that the rs6983267 region with enhancer activity regulates expression of adjacent noncoding RNAs including *CARLos*.

Correlation Between *CARLo-5* Expression and rs6983267 in the *MYC* Enhancer Region. To determine whether the *MYC* enhancer region including rs6983267 is involved in the regulation of

CARLos, we tested expression levels of *CARLos* in normal colon tissues classified by alleles of rs6983267. Interestingly, of seven *CARLos*, the expression level of *CARLo-5* is significantly induced in the tissues with heterozygous [guanine (G)/thymine (T)] or homozygous (G) cancer risk allele of rs6983267 (Fig. 2 A and B). These results strongly suggest that the rs6983267 region with enhancer activity is implicated in *CARLo-5* regulation. The ENCODE transcription factor binding database shows that TCF4 (TCF7L2) significantly binds to the promoter region of *CARLo-5* in various cancer-derived cells such as HCT116, HepG2, HeLa, and MCF7 (Fig. 3A, *Upper*). Using various prediction programs of transcription factor binding, however, we found that there is no predicted sequence for TCF4 binding in the *CARLo-5* promoter region (data not shown). Interestingly, it was reported that the enhancer region including rs6983267 binds with TCF4 (TCF7L2), and the region has a different affinity to the transcription factor TCF4 (TCF7L2) by the different alleles of rs6983267 (8). These results suggest that the 8q24 enhancer region including rs6983267 regulates the expression of *CARLo-5* through long-range interaction with the promoter of *CARLo-5*.

instructions (Life Technologies). All RT reactions, including no-template controls and RT minus controls, were run in a GeneAmp PCR 9700 Thermocycler (Life Technologies). RNA concentrations were determined with a NanoDrop 20000 (Thermo Fisher Scientific, Inc.). Samples were normalized to GAPDH or OAZ1 for mRNAs and lncRNAs (Life Technologies). Gene expression levels were quantified using the ABI Prism 7900HT Sequence detection system (Life Technologies). Comparative real-time PCR was performed in triplicate, including no-template controls. Relative expression was calculated using the comparative Ct method. Taqman assay for CARLo-5 (Taqman assay ID: AJHSNV9) was designed using a custom assay design tool (Life Technologies), based on sequence information identified by RACE.

Cells, Oligonucleotides, Plasmids, and Transfection. All cell lines were cultured as recommended by ATCC. Custom siRNAs for CARLo-5 were designed by the Dharmacon custom siRNA design tool based on the sequence information identified by RACE. The sequence of the custom siRNAs is described below. As a negative control, smartpool ON-TARGETplus negative control (CTRL) siRNA (mixture of four negative control siRNAs, siCTRL-pool) and ON-TARGETplus Nontargeting siRNA #1 (siCTRL-1) were used. As a mock test, transfection mixture without siRNA was used. Transfection was performed with Lipofectamine RNAiMAX (Invitrogen) in accordance with manufacturer's instructions, and siRNAs were used at 50 nM final concentration.

CARLo-5 siRNA-1 (siCARLo5-1): GGAGGGUGUCUAGCAAUAUU

CARLo-5 siRNA-2 (siCARLo5-2): GAGAAGACCAUAAGAAGAU

CARLo-5 siRNA-3 (siCARLo5-3): UGGAAAGGUGCCGAGACA

Cell Proliferation Assay. For cell proliferation assay, MTS assay from Promega (CellTiter 96 AQueous One Solution Cell Proliferation Assay) was used following manufacturer's instructions. Briefly, cells in 96-well plates were incubated for 72 or 96 h in humidified 5% (vol/vol) CO₂ atmosphere after transfection with indicated siRNAs, followed by addition of 20 μ L CellTiter 96 AQueous One Solution and 1–4 h incubation in humidified 5% (vol/vol) CO₂ atmosphere.

Soft Agar Colony Formation Assay. Soft agar colony formation assay was performed with CytoSelect cell transformation assay (Cell Biolab, Inc.) in accordance with manufacturer's instructions. The cells transfected with the indicated siRNA were incubated for 24 h followed by soft agar colony formation assay.

Xenograft and Tumor-Free Survival Analysis. Animal experiments were approved by The Ohio State University animal care and use committee and conducted following The Ohio State University animal policy in accordance

with National Institutes of Health guidelines. The 0.5 million cells (HCT116 and RKO) transfected with indicated siRNAs 24 h before injection were s.c. injected into the right flanks of a 5-wk-old female athymic nude mouse (Jackson Laboratory). Tumor initiation was observed and counted when a palpable tumor (\sim 50–60 mm³) formed. Tumor volume was determined by measuring the length and the width of the tumor mass and calculating the volume [volume = (width)²(length)/2]. The palpable mass was observed until the volume reaches 100 mm³ to verify that the palpable mass is a tumor mass. The parental RKO and HCT116 cells that were not transfected with siRNA were also injected, and the results of tumor-free survival with parental cells showed similar results with control siRNA-transfected cells. All athymic nude mice developed palpable tumors 5 wk after injection with the parental cells.

Flow Cytometry Analysis. For DNA content analysis, cells were fixed in methanol at -20 °C, washed again, rehydrated, resuspended in PBS containing 2 μ g/mL propidium iodide (PI) and 5 μ g/mL RNase A, and analyzed by BD FACS Calibur Flow Cytometer.

NanoString Gene Expression Assay and Data Analysis. For NanoString Gene expression assay analyses, the nCounter Virtual Cell Cycle Gene Set was used, following manufacturer's instructions (NanoString Technologies). Briefly, total RNA (100 ng) were used as input for nCounter mRNA sample preparation reactions. All sample preparation was performed according to manufacturer's instructions (NanoString Technologies). Hybridization reactions were performed according to manufacturer's instructions with 5 μ L of the fivefold diluted sample preparation reaction. All hybridization reactions were incubated at 64 °C for a minimum of 18 h. Hybridized probes were purified and counted on the nCounter Prep Station and Digital Analyzer (NanoString Technologies) following manufacturer's instructions. For each assay, a high-density scan (600 fields of view) was performed. Data analysis was performed using the nSolver analysis software (NanoString Technologies) and dChip software.

Statistics. All graph values represent means \pm SD from three independent experiments with each measured in triplicate. The differences between two groups were analyzed with unpaired two-tailed Student *t* test. *P* < 0.05 was considered statistically significant and indicated with asterisks as described in the figure legends.

ACKNOWLEDGMENTS. We thank Dr. Thomas D. Schmittgen (The Ohio State University) for the revision of this manuscript. We also thank the human and vertebrate analysis and annotation team for use of the genome annotation database and the Encyclopedia of DNA Elements (ENCODE) consortium for use of the ENCODE regulation database.

- Yeager M, et al. (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39(5):645–649.
- Haiman CA, et al. (2007) A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* 39(8):954–956.
- Tomlinson I, et al. (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39(8):984–988.
- Sur IK, et al. (2012) Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* 338(6112):1360–1363.
- Pomerantz MM, et al. (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 41(8):882–884.
- Ahmadiyeh N, et al. (2010) 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc Natl Acad Sci USA* 107(21):9742–9746.
- Sansom OJ, et al. (2007) Myc deletion rescues Apc deficiency in the small intestine. *Nature* 446(7136):676–679.
- Tuupainen S, et al. (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 41(8):885–890.
- Zanke BW, et al. (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39(8):989–994.
- Prokunina-Olsson L, Hall JL (2009) No effect of cancer-associated SNP rs6983267 in the 8q24 region on co-expression of MYC and TCF7L2 in normal colon tissue. *Mol Cancer* 8:96–100.
- Pomerantz MM, et al. (2009) Evaluation of the 8q24 prostate cancer risk locus and MYC expression. *Cancer Res* 69(13):5568–5574.
- Wapinski O, Chang HY (2011) Long noncoding RNAs and human disease. *Trends Cell Biol* 21(6):354–361.
- Guttman M, et al. (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477(7364):295–300.
- Klattenhoff CA, et al. (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152(3):570–583.
- Hung T, et al. (2011) Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 43(7):621–629.
- Huarte M, et al. (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142(3):409–419.
- Gupta RA, et al. (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464(7291):1071–1076.
- Yang F, et al. (2011) Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology* 54(5):1679–1689.
- Xu D, et al. (2013) Long noncoding RNAs associated with liver regeneration 1 accelerates hepatocyte proliferation during liver regeneration by activating Wnt/ β -catenin signaling. *Hepatology* 58(2):739–751.
- Heintzman ND, et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459(7243):108–112.
- Creyghton MP, et al. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* 107(50):21931–21936.
- Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823–837.
- Sotelo J, et al. (2010) Long-range enhancers on 8q24 regulate c-Myc. *Proc Natl Acad Sci USA* 107(7):3001–3005.
- Ling H, et al. (2013) CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res* 23(9):1446–1461.
- Hagège H, et al. (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2(7):1722–1733.