様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２７　年　　６　月　１５　日現在

機関番号：　８２４０１
研究種目：　若手研究(B)
研究期間：　2011～2014
課題番号：　２３７１０２２２

研究課題名（和文）５’キャップ構造、アンチセンス、プロモーター近傍ＲＮＡの機能の解明

研究課題名（英文）Elucidation of the function of 5' capped antisense promoter-associated RNA

研究代表者
ＤＥ・ＨＯＯＮ　ＭＩＣＨＩＥＬ（De Hoon, Michiel）

独立行政法人理化学研究所・ライフサイエンス技術基盤研究センター・ユニットリーダー

研究者番号：７０５２５６１７

交付決定額（研究期間全体）：（直接経費）　　2,500,000円

研究成果の概要（和文）：私たちのプロジェクトは、CASPARsの生物学的役割、遺伝子の活性の調節に関与することができるRNA分子の新規ファミリーを理解することに焦点を当てています。白血病細胞株では、癌に関与する転写因子をコードするものを含め、数百の遺伝子でCASPARsを同定しました。特定CASPAR転写の阻害はCASPARsが機能的であり、遺伝子活性の調節において役割を果たすことを示唆し、これらの遺伝子の発現の変化を引き起こしました。次に、我々は、次世代シークエンシングを用いCASPAR発現の動態を測定します。本研究では、このように一般的に癌で調節不全されている癌遺伝子の調節の我々の理解に貢献しています。

研究成果の概要（英文）：Our project focuses on understanding the biological role of CASPARs, a novel family of RNA molecules that overlap genes near their promoter, which is the key genomic region for regulating the activity of genes. In the THP-1 leukemia cell line, we have identified CASPARs at hundreds of genes, including those coding for transcription factors involved in cancer. Upon inhibition of specific CASPAR transcripts, we found expression changes of these genes, suggesting that the CASPARs are functional and play a role in regulating gene activity. We are currently planning to measure the genome-wide response to CASPAR inhibition using next-generation sequencing. We also subjected the THP-1 cells to PMA stimulation, causing the cells to differentiate into monocytes, and extracted RNA at several time points to measure the dynamic behavior of CASPARs during differentiation. This study thus contributes to our understanding of the regulation of oncogenes, which are typically dysregulated in cancer.

研究分野：　Molecular Biology

キーワード：　Transcriptome　Non-coding RNA　Sequencing　Gene regulation

１．研究開始当初の背景 Background and motivation of the research at the time of application.

(1) Almost all cells in the human body have an identical DNA sequence. However, their morphology, biological function, and behaviors are strikingly different, which can be explained by the specific programs of gene expression and regulation that govern the biological character of each cell. The importance of gene regulation in human disease is particularly apparent in cancer cells, in which dysregulation of gene expression leads to excessive or inappropriate cell proliferation.

(2) Genes are primarily regulated by transcription factors, which are proteins binding to specific regulatory DNA sequences, as well by as non-coding RNAs such as micro-RNAs, small interfering RNAs, and long non-coding RNAs. Next-generation sequencing approaches have revealed a wide variety of novel small RNAs whose cellular role is currently not well understood. In particular, several families of small RNAs have been found near the promoters of protein-coding genes, such as promoter-associated long and short RNAs, transcription initiation RNAs, and promoter upstream transcripts (PROMPTs). While the biogenesis and biological function of these RNAs are largely unknown, they are suspected to play a role in gene regulation because of their close association with promoters.

(3) We recently developed a novel technique for cloning full-length transcripts from poly-A-minus RNA, and used this methodology to sequence long poly-A-minus transcripts obtained from human THP-1 myelomonocytic leukemia cells. A bioinformatics analysis of these sequences revealed a novel family of RNAs that have a 5' cap but lack a poly(A) tail, and originate near the promoters of coding genes but on the opposite strand of DNA. These promoter- associated antisense transcripts have been found at more than 1700 genes, indicating that they are a widespread phenomenon in transcription. In a gene ontology analysis, the categories biological regulation, developmental process, cell differentiation, and negative regulation of biological process were shown to be very strongly statistically overrepresented ($p \sim 10^{-20}$ or less) among genes with such promoter-associated anti-sense transcripts. Furthermore, the transcripts are preferentially associated with lowly expressed genes, suggesting that they have an inhibitory regulatory function. We therefore named this novel family of non-coding transcripts CASPAR RNAs, for capped anti-sense promoter-associated regulatory RNAs. CASPAR RNAs typically have a length of 200 nucleotides or more and are not spliced.

２．研究の目的 Purpose of the research

Our aim is to establish the existence, genomic start and end, subcellular localization, and expression dynamics of CASPAR transcripts at selected genes, as well as to elucidate their regulatory role by measuring the expression of the sense transcript as well as the genome-wide response of the transcriptome upon transfection of specific CASPAR transcripts.

３．研究の方法 Research method

(1) CASPAR selection
We focused our research project on CASPAR transcripts at four genes encoding transcription factors, and at which CASPAR transcripts were detected in the previous sequencing libraries. The four genes we selected encode the transcription factors MYB (a proto-oncogene), SPI1 (also known as PU.1, an oncogene that activates gene expression during myeloid development), STAT6 (important in signaling by cytokines and growth factors), and GFI1 (a transcriptional repressor independent of growth factors); these transcription factors are key regulators of proliferation and differentiation of THP-1 leukemia cells.

(2) RACE experiments
We determined the extent and subcellular localization of CASPAR transcripts at each of these four genes using RACE (Rapid Amplification of cDNA Ends). Given the limitations in the transcript length that can be read by high-throughput sequencers, in most cases only a partial (5' or 3') CASPAR transcript had been sequenced, rather than the full-length CASPAR sequence. The true length of CASPAR transcripts therefore remained obscure. In addition, since only a few CASPAR

transcripts were seen at each gene, it is unclear whether a single isoform of CASPAR transcripts is produced at each gene, or several isoforms with varying start and end positions. To address these questions, we performed RACE experiments followed by sequencing using the 454 Titanium sequencer. The RACE experiments were performed on RNA separated by their nuclear or cytoplasmic origin to determine the subcellular localization of CASPAR transcripts. RACE primers were chosen such that the 5' and 3' RACE products overlap in order to ascertain that the 5' and 3' ends of the cDNA found in the RACE experiments originate from the same CASPAR transcript.
Upon stimulation of THP-1 cells by phorbol myristate acetate (PMA), the cells cease proliferation, become adherent, and differentiate into a mature monocyte- and macrophage- like phenotype. The sequencing data in which the CASPARs were originally discovered were obtained only at a single time point after stimulation by PMA, and consequently the dynamics of CASPAR transcript expression is presently unknown. We therefore performed the RACE experiment on RNA obtained before stimulation, as well as after 96 hours of stimulation by PMA.

(3) Transfection of CASPAR transcripts into THP-1 cells
To establish the regulatory role of the CASPAR transcripts directly, we transfected specific CASPAR transcripts into unstimulated THP-1 cells. We produced selected CASPAR transcripts in-vitro using a transcription kit, as well as an m7G capping enzyme to add the 5' cap to the transcript. Upon transfection of these CASPAR transcripts into THP-1 cells, we measured the expression response of the sense transcript by qRT-PCR. We decided, as a first attempt, to transfect CASPAR transcripts instead knocking them down as the knock-down procedure may directly affect the sense gene overlapping the CASPAR.

(4) Inhibition of CASPAR expression
As described below, in the end the CASPAR transfection experiments did not give reliable results due to the interferon response of THP-1 cells to foreign DNA. We therefore attempted to knock down the CASPAR transcript using antisense oligo gapmers. For each CASPAR transcript, we designed two antisense oligo gapmers.

(5) Expression dynamics
As CASPAR transcripts have a 5' cap, we reasoned that CAGE (Cap Analysis of Gene Expression) is the ideal approach to measure the expression of CASPAR transcripts on a genome-wide scale. However, the standard CAGE protocol will predominantly detect transcripts longer than the size of CASPAR transcripts. We therefore concluded that it is necessary to enrich for shorter transcripts in the CAGE library. This is challenging as the use of a random primer for the reverse transcription reaction loses the information of the length of the original transcript, whereas oligo-dT priming will not work against CASPAR transcripts as they do not have a poly(A)-tail. We overcame this problem by using a ligation strategy.

４．研究成果 Research result

(1) RACE experiments
We performed 5' and 3' RACE (rapid amplification of cDNA ends) to elucidate the exact start and end of the CASPAR transcript we previously identified by deep sequencing. We extracted RNA in three independent experiments from THP-1 myelomonocytic leukemia cells before, as well as 96 hours after stimulation by phorbol myristate acetate (PMA). PMA causes the cells to cease proliferation, become adherent, and to differentiate into a mature monocyte- and macrophage-like phenotype. We performed RACE against 13 putative CASPAR transcripts that overlap the 5' end of the SPI1, MYB, GFI1, or STAT6 genes, which encode transcription factors that play a major role in differentiation of THP-1 myelomonocytic leukemia cells into monocytic cells. The RNA at both time points was subsequently divided into a nuclear and a cytoplasmic fraction to allow us to identify the subcellular localization of CASPAR transcripts. We reverse transcribed the CASPAR transcripts and amplified the cDNA products in three subsequent PCR amplification steps, using outer, inner, and replacement primers to maximize the sensitivity and specificity. The distribution of cDNA lengths in the library suggested that we successfully amplified the CASPAR transcripts of interest. We then performed high-throughput sequencing of these RACE libraries to determine the exact 5'

and 3' end of the CASPAR transcripts. All thirteen CASPAR transcripts as well as the positive control, a histone transcript, were confirmed in the RACE libraries at both time points and in both the nuclear and the cytoplasmic fractions. The RACE libraries revealed that the CASPAR transcripts have specific 5' and 3' ends, in agreement with our hypothesis that they are functional and not random or degradation products. The 3' RACE data also confirm that the CASPAR transcripts do not have a poly(A) tail.

(2) Transfection experiments
We have then created expression vectors to transfect CASPARs into THP-1 cells, which was challenging due to their lack of a poly(A) tail. As a test, we selected a CASPAR transcript that overlaps the SPI1 gene, and attempted transfecting it into THP-1 cells. We confirmed by qPCR expression measurements that the transfection was successful, and that the expression vector containing the CASPAR transcript sequence is being transcribed in the THP-1 cell.
We then performed the transfection into THP-1 cells of two CASPAR transcripts that overlap the gene encoding the STAT6 transcription factor. Indeed we saw a decrease of 20-30% in the expression of STAT6 upon transfection of CASPAR5. However, we also saw a decrease in the expression of STAT6 after the transfection of GFP (green fluorescent protein). This may be due to an interferon response in which transfection of foreign DNA into THP-1 cells is causing them to differentiate from a monoblast to a macrophase state. This was confirmed by measuring the expression of the MYB and GFI1 genes, which were similarly affected by the transfection.

(3) Knockdown experiments
Based on these experimental results, we concluded that the interferon response in THP-1 cells to transfections would mask any effect of the transfected CASPAR. We therefore changed our strategy to knock down the endogenously expressed CASPAR transcripts using antisense oligos (gapmers). We used 2 antisense oligos for each of the 13 CASPARs and measured the expression of the corresponding sense gene to discover any cis-regulatory effect. For most knockdowns, we indeed found an upregulation or downregulation of the sense gene (as measured by qPCR).
We then developed a strand-specific qPCR protocol to verify the knockdown efficiency of the CASPAR transcript itself. The strand-specific qPCR confirmed the knockdown efficiency of the three CASPARs. We will now proceed to perform CAGE expression profiling on the knockdown samples to assess the genome-wide transcriptome response to the CASPAR knockdown.

(4) Expression dynamics
We also subjected THP-1 cells to PMA stimulation, causing the THP-1 monoblast cells to differentiate into monocytes, and extracted RNA at several time points during this differentiation process. We produced CAGE libraries from these RNA samples, using a novel protocol under development in our laboratory to enrich the CAGE library for short transcripts. We performed paired-end sequencing of a test library on a MiSeq sequencer, confirming that this protocol can detect CASPAR transcripts. Next, we will apply single-end sequencing to these libraries to measure the CAGE expression dynamics of the CASPARs.
We aim to describe the genome-wide transcriptome response to the CASPAR knockdown, together with the dynamics of CASPAR expression, in a future publication.

５．主な発表論文等 Important papers
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　0　件）

〔学会発表〕（計　1　件）
The 5th Biennale RIKEN Joint Retreat, Kakegawa, Shizuoka, February 5, 2013; Oral presentation number SIV-1; title “Capped Anti-Sense Promoter-Associated Regulatory (CASPAR) Transcripts: A Novel Family of Regulatory Non-Coding RNAs”
〔図書〕（計　0　件）

〔産業財産権〕 Industrial property rights
○**出願状況（計　0　件）**

○**取得状況（計　0　件）**

６．研究組織
(1)**研究代表者**
DE HOON MICHIEL
独立行政法人理化学研究所・ライフサイエンス技術基盤研究センター・ユニットリー

ダー
研究者番号：70525617

⑵研究分担者
（　　）

研究者番号：

⑶連携研究者
（　　）

研究者番号：