様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月 １７ 日現在

機関番号：８２４０１
研究種目：若手研究(B)
研究期間：2017～2019
課題番号：１７Ｋ１８３６６

研究課題名（和文）Discovery and validation of pan cancer epigenetic biomarkers

研究課題名（英文）Discovery and validation of pan cancer epigenetic biomarkers

研究代表者
カチコフスキー　ボグミル（Kaczkowski, Bogumil）

国立研究開発法人理化学研究所・生命医科学研究センター・研究員

研究者番号：５０６４８１３６

交付決定額（研究期間全体）：（直接経費）　　2,600,000円

研究成果の概要（和文）：長年に亘る研究にも関わらず、診療機関で使用されている癌バイオマーカーは殆ど無い。DNAメチル化は安定で長期的な調節に関与している為、異常なDNAメチル化によって引き起こされ通常とは異なって発現する遺伝子がバイオマーカーの有力な候補である事を提案する。我々は公的に利用可能なトランスクリプトームとエピゲノムデータを統合する計算分析を行い、プロモーターの低メチル化の為に肺癌で上方制御されるコアセット遺伝子を発見した。更に詳しく研究する為、正常な細胞を脱メチル化剤とHDAC阻害剤で処理する攪乱実験を行った。攪乱実験後、遺伝子発現解析（CAGE）、DNAメチル化アレイ、シングルセルCAGE実験を行った。

研究成果の学術的意義や社会的意義
The results and data generated in the project help us to better understand the link between the aberrant epigenetic changes and the gene expression in cancer. This opens a way to uncover the theragnostic potential of epigenetically regulated genes in clinical cancer research.

研究成果の概要（英文）：Despite decades of research, few cancer biomarkers are being used in clinics. Based on the assumption that DNA methylation is involved in stable, long-term regulation, we propose that differentially expressed genes that are caused by aberrant DNA methylation are optimal candidates for biomarkers. We performed computation analyses integrating publicly available transcriptomic and epigenomic data. We discovered 49 coding genes and 10 noncoding RNAs, which are upregulated in NSCLC lung cancer due to promoter hypomethylation. We also observed that multiple copies of the REP522 DNA repeat family are activated in lung cancer by DNA hypomethylation and histone modification. To study the link between DNA methylation and transcription more closely, we performed perturbation experiments, where normal cells were treated with a demethylating agent and histone deacetylase inhibitor. The perturbations were followed by gene expression profiling (CAGE), DNA methylation array, and single-cell C1-CAGE.

研究分野：Computational Cancer Biology

キーワード：cancer　epi-genome　transcriptome　genome

１．研究開始当初の背景

Research Background

Despite decades of research, few cancer biomarkers are being used in clinics. In our previous pan-cancer work published in *Cancer Research* (Kaczkowski, B, et al.　2016), we identified de-regulated protein-coding genes and lncRNA across cancer cell lines by analyzing Cap Analysis of Gene Expression (CAGE) profiles from FANTOM5 project.
Here, based on the assumption that DNA methylation is involved in stable, long-term regulation, we propose that differentially expressed genes that are caused by aberrant DNA methylation are optimal candidates for biomarkers.

２．研究の目的

Purpose of Research

The purpose of this research is to find a novel approach to deeply integrate transcriptomics and epigenetics data and to find robust cancer biomarkers that are regulated at epigenetic level (epi-markers/drivers), including novel lncRNAs and transcribed repetitive elements. To that end, we performed computation analyses integrating publicly available transcriptomic and epigenomic data. These included computational analysis of the integrative analyses of gene expression and DNA methylation across cancer samples from TCGA consortium. To study the link between DNA methylation and transcription more closely, we performed a set of controlled, perturbation experiments, where normal epithelial cells (MCF10A) were treated with a demethylating agent and histone deacetylase inhibitor. The perturbations were followed by gene expression profiling (CAGE), DNA methylation array, and single-cell C1-CAGE.

３．研究の方法

Methods

(1) Identifying epigenetically regulated genes in NSCLC by integrative analyses of CAGE, RNA-seq and DNA methylation data

We analyzed and integrated publicly available data: **A)** CAGE expression data of 16 normal lung epithelial cells and 16 NSCLC cell lines (FANTOM5); **B)** RNA-seq of 515 lung adenocarcinomas with 59 normal tissue controls, 501 lung squamous cell carcinoma with 49 normal tissue controls (TCGA); and **C)** DNA methylation array data from TCGA project and GSE36216 dataset.

(2) Pan-cancer analysis of REP522 repeat activation in primary cancers.

I analyzed RNA-Seq data from 21 TCGA primary tumor types and normal tissue control data profiled by TCGA (The Cancer Genome Atlas, https://gdc-portal.nci.nih.gov)

(3) Epigenomic perturbation in the MCF10A cell line.

We performed a series of epigenetic perturbations in normal epithelial cells (MCF10A). We tested four conditions, where the cells were treated with: **1)** DAC (5-aza-2'-deoxycytidine, DNA demethylating agent, 500nM), **2)** TSA (Histone Deacetylase inhibitor, 500nM), **3)** DAC and TSA combination, and **4)** DMSO control.

(4) CAGE and methylation profiling after perturbations (bulk).

After the perturbations, the cells were profiled by CAGE 5'start RNA sequencing to allow for promoter level gene expression analysis. We also performed DNA methylation profiling using Illumina EPIC DNA methylation array that covers ~830 thousand methylation sites across the genome.

(5) Single-cell gene expression profiling after perturbations.

To study the heterogeneity of response to the demethylating drug we also performed C1-CAGE profiling after the perturbations. C1-CAGE is a single-cell implementation of Cap Analysis of Gene Expression (Kouno, T. et al. Nat. Commun. 2019)

４．研究成果

Research Results

(1) Identifying epigenetically regulated genes in NSCLC by integrative analysis CAGE and DNA methylation data

In collaboration with researchers from Tokyo University, we performed the integrative analyses of gene expression and DNA methylation in lung cancer cell lines and clinical tumors. We discovered a set of 49 coding genes and 10 long noncoding RNAs (lncRNA), which are upregulated in NSCLC cell lines due to promoter hypomethylation. We validated 22 epigenetically up-regulated genes in the adenocarcinoma and squamous cell cancer subtypes of lung cancer using RNA-seq data from The Cancer Genome Atlas.

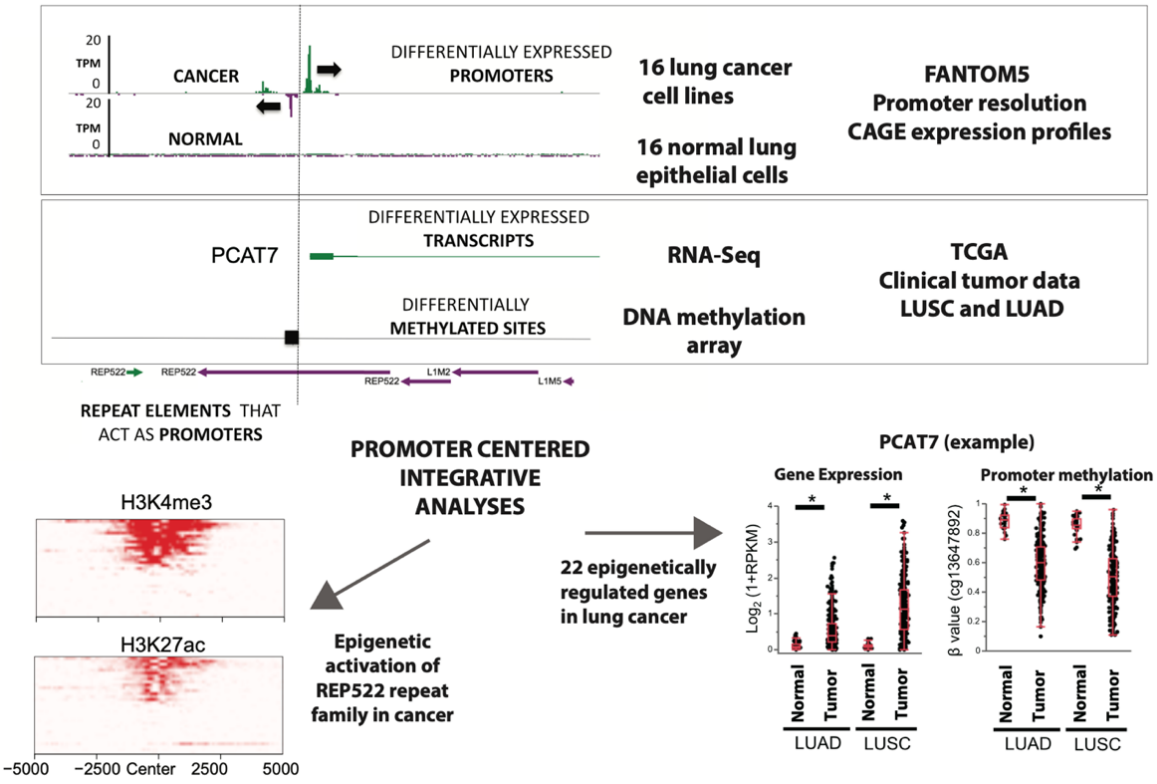

Figure 1 Visual Overview of the article "*Integrative CAGE and DNA Methylation Profiling Identify Epigenetically Regulated Genes in NSCLC*" as published in *Molecular Cancer Research* (Horie, M., Kaczkowski, B.*, et al., 2017, **co-corresponding author.*)

REP522 is an unclassified interspersed repeat of 1.8 Kb in size, with a large palindrome of ~600nt in the center. It is a small and elusive family of repeat elements with just 368 copies in the genome. Previously, we reported that many REP522 repeats harbor bidirectional promoters that are activated in multiple cancer types (Kaczkowski, B, et al. 2016). Such REP522 elements are silent in normal cells, but become active promoters in cancer cells and drive the expression of multiple long non-coding RNAs and pseudogenes.

Here, we observed that multiple copies of the REP522 DNA repeat family are, in fact, epigenetically activated in lung cancer by DNA hypomethylation and histone modification typical to active promoters (H3K4me3). The activated REP522 repeat elements act as bi-directional promoters for cancer-specific lncRNAs, e.g. RP1-90G24.10, AL022344.4, and PCAT7. (**See Figure 1 for Visual Overview**).

(2) Pan-cancer analysis of REP522 repeat activation in primary cancers.

Here, I performed the pan-cancer analyses using RNA-Seq data from 21 tumor types profiled by The Cancer Genome Atlas (TCGA). I calculated the frequency (%) of activation/expression of REP522 promoters across 21 cancer types (7916 primary tumors and 725 normal tissue controls) **(Figure 2)** The results were presented as a poster at the Human Genomics meeting (Kaczkowski B.*, et. al. 2018, *Human Genomics* 2018, 12(Suppl 1):9).

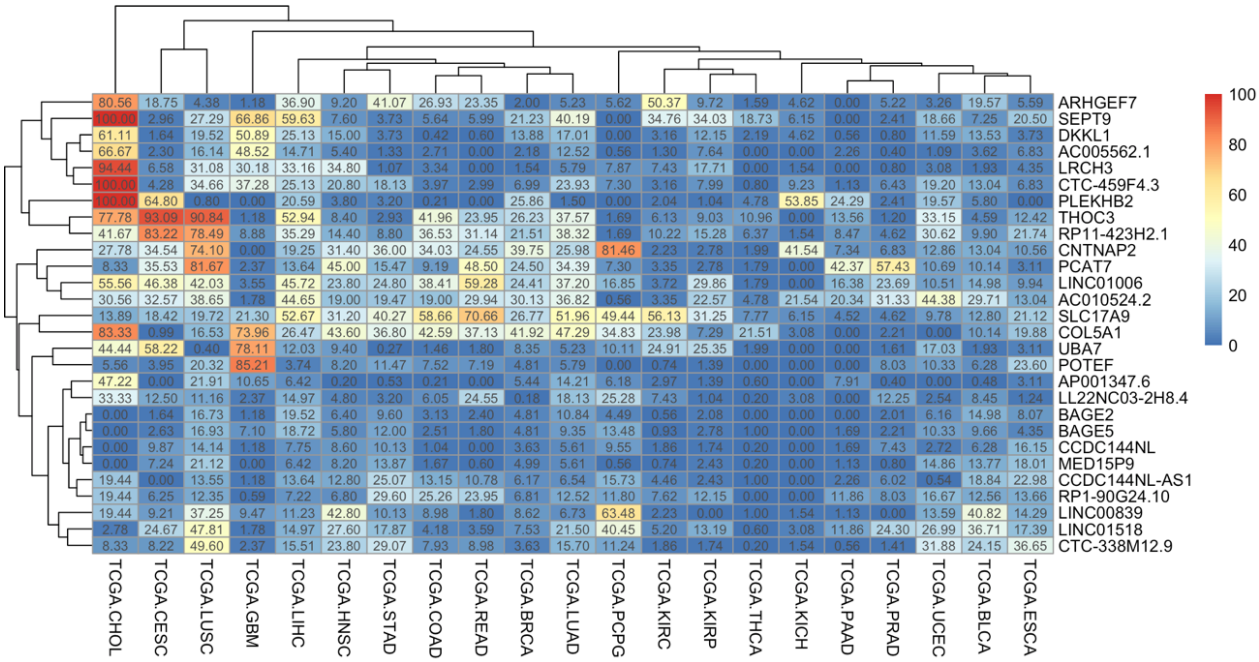

Figure 2 The heatmap shows the frequency (%) of activation/expression of REP522 promoters across 21 TCGA cancer types. Activation threshold was set to mean(x) + 3sd(x), where x is the expression in normal samples. Genes with a frequency of activation > 5% in 7 (out of 21, 1/3) TCGA cancer types are shown. (Kaczkowski B.*, et. al. 2018, *Human Genomics* 2018, 12(Suppl 1):9, poster abstract).

(3) Epigenomic perturbation in the MCF10A cell line.

To complement the computational analysis of publicly available data, we performed new experiments where we are perturbing the DNA methylation and histone acetylation in normal epithelial cells (MCF10A), which is followed by promoter level transcriptomic profiling using CAGE technology. This controlled, perturbation experiment enables us to understand the direct link between the epigenetic aberrations and transcription and if hypomethylation on its own is enough to activate the transcription of epigenetic cancer biomarkers and repeat elements including REP522. This work is done in collaboration with Dr. Kazuhide Watanabe from RIKEN IMS. Preliminary results of CAGE profiling show that DAC/TSA combination has the strongest effect on the transcriptome (**Figure3**). Further analyses are now ongoing.

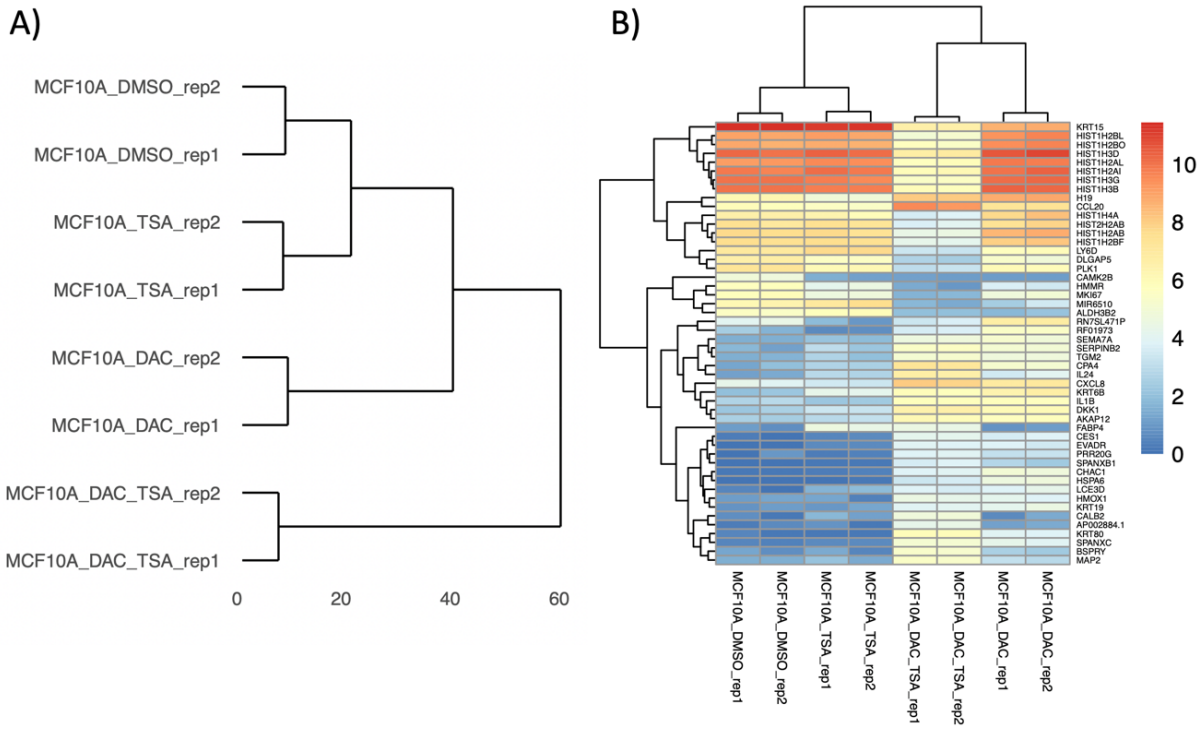

Figure 3 The effect of DAC (DNA demethylating agent) and TSA (Histone Deacetylase inhibitor) on gene expression in MCF10A cell line (+ DMSO control). A) Hierarchical clustering based on 500 most variant genes. B) Heatmap visualizing the expression (log2 tag per million) of 50 most variant genes.

(4) Single cell gene expression profiling after perturbations.

Single-cell sequencing will offer us an opportunity to study the heterogeneity of how individual cells respond to the demethylating drug. The single-cell experiments have been performed and RNA samples are awaiting C1 CAGE library preparation and deep sequencing. The single-cell sequencing data are estimated to be generated by September 2020 and will be analyzed soon afterwards.

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　1件）

１．著者名 Horie M, Kaczkowski B, Ohshima M, Matsuzaki H, Noguchi S, Mikami Y, Lizio M, Itoh M, Kawaji H, Lassmann T, Carninci P, Hayashizaki Y, Forrest ARR, Takai D, Yamaguchi Y, Micke P, Saito A, Nagase T.	４．巻 15
２．論文標題 Integrative CAGE and DNA Methylation Profiling Identify Epigenetically Regulated Genes in NSCLC.	５．発行年 2017年
３．雑誌名 Molecular Cancer Research	６．最初と最後の頁 1354-1365
掲載論文のDOI（デジタルオブジェクト識別子） 10.1158/1541-7786.MCR-17-0191	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計0件

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考