様式19 別紙1

課題番号	LS127

先端研究助成基金助成金(最先端・次世代研究開発支援プログラム)
実施状況報告書(平成 23 年度)

本様式の内容は一般に公表されます

研究課題名	Genome-wide identification of non-coding RNA function for cell differentiation
研究機関・ 部局・職名	RIKEN、The Institute of Physical and Chemical and Research、Omics Science Center, Functional Genomics Technology Team, Team Leader
氏名	Piero Carninci

1. 当該年度の研究目的

In the second year of the project we have essentially accomplished the AIM1, which consists of characterization of the transcriptome of human and mouse ES, iPS and their corresponding differentiated cell. The AIM 1 was to use CAGE, CAGEscan, sRNAseq and RNAseq libraries from the above mentioned samples. We have also moved towards the AIM2, which consists of perturbation of expression of ncRNAs to understand the phenotype caused by these ncRNAs. The purpose was to test overexpression and knockdown strategies for selected ncRNAs and REs (detected at the AIM1) by using siRNA, shRNA, PNA, LNA, or promoter targeting RNAs and monitor their expression by qRT-PCR as well as global expression. Also, the purpose was to achieve cloning ncDNAs in lentivirus expression system. We also wanted to start the AIM3, which consists of induction of differentiation/dedifferentiation by ncRNAs. One of the purposes of AIM3 to identify the iPS-specific ncRNAs that are capable of inducing differentiation of iPS upon repression using knockdown. The other purpose was to over express them in the differentiated cell (MEF or other) to measure their potential for dedifferentiation.

2. 研究の実施状況

Due to very limited time (Feb-Mar 2010) for the first year of the project, only a very limited part of research was achieved. In the second year of the project (fiscal year 2011), we have comprehensively proceeded towards the goals of the project.

For the AIM1, we have created a massive resource of transcriptome data from nuclear and cytoplasmic RNA fractions of 12 different embryonic and iPS cell lines as well as from 5 different differentiated cell lines. The amount of data is beyond the expectations at the beginning of the project. Among the cell lines, nine are from murine and 8 are from human origin. All the libraries (Total 34/library type) produced are sequenced with Illumuna HiSeq200 v3 and Illumina MiSeq. We have produced 989 million tags for CAGE libraries, 1.7 billion tags for CAGEscan, 2 billion tags for sRNAseq, and 1 million tags from a new method developed here at OSC to detect chromatin associated RNAs (CAR). This differs from the general methods because

the data identifies the specific binding of such RNAs on the chromatin. Moreover, we are expecting ~600 million tags from RNAseq libraries and ~1 billion tags from libraries prepared for CARs. This is the largest and the most diverse dataset on stem cells if this nature, which provides the most complete collection of transcription starting site and RNA transcript expression.

To date, with the sequencing dataset, we have analyzed all of the 18 mouse libraries. Human dataset analysis is ongoing currently. Genome wide promoter activity was measure with CAGE for mouse cell lines. A total of 102,495 tag clusters were detected in the dataset using Paraclu clustering of CAGE tags. Hierarchical clustering of the murine CAGE samples shows distinct clustering of nuclear and cytoplasmic fractions, with sub-clusters corresponding to differentiation stages.

Differential expression analyzes were performed separately for the nuclear and cytoplasmic samples between stem and differentiated cells using DESeq R package on single mapped tag clusters with expression value over 1 tag/million and removing cluster expressed in a single sample. Interestingly, the nuclear stem cell specific group display the greater amount of differentially expressed clusters (pAdj<0.05), with 31% of them being not annotated by current genomic databases (Figure 1A). Among nuclear stem cell specific clusters, enrichment is observed for clusters mapping in intergenic region or having antisense transcription pattern (Figure 1B). Finally, we identified a large number of nuclear stem cell specific differentially expressed clusters mapping to retrotransposon elements.

	Cytoplasmic Stem cell specific	*Cytoplasmic differentited cell specific*	*Nuclear Stem cell specific*	*Nuclear Differentiated cell specific*
Protein coding	1663	1626	2518	2318
Processed transcript	115	84	212	113
Retained intron	75	65	86	76
Antisense	19	20	40	27
Transcribed pseudogene	19	9	31	4
Pseudogene	7	5	10	3
Nonsense mediated decay	20	19	38	25
Retrotransposed	2	-	3	-
lincRNA	46	30	136	50
miRNA	3	-	8	1
snoRNA	-	1	-	1
misc RNA	1	1	3	1
Non coding	-	2	4	1
IG gene	-	93	-	78
Mt rRNA	-	12	-	2
No annotations	**317 (14%)**	**125 (6%)**	**1356 (31%)**	**341 (11%)**
Total	*2287*	*2092*	*4445*	*3041*

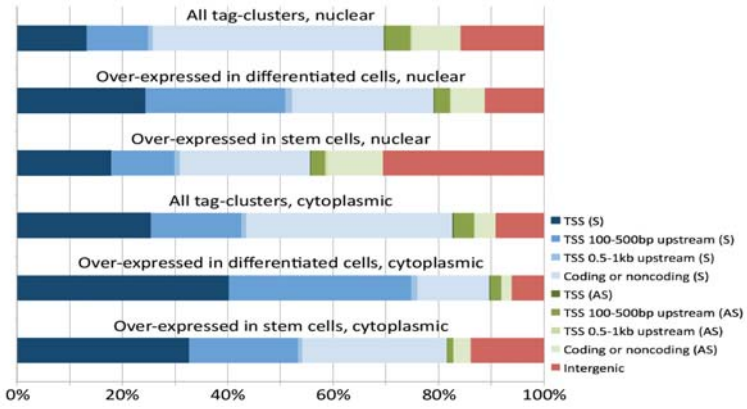

Figure1 (A) Total differentially expressed Tag clusters per *Ensembl* biotype. (B) Fractions of tag clusters relative to their genomic locations.

Among tag clusters overlapping repeat elements, we found enrichment for stem cell over expressed clusters in SINE.B4, LINE.L1 and LTR (ERVK, MaLR, ERV1, ERVL) (Table 1). We look at expression of LINE.L1 relative to their sizes in mouse stem cells. Full length elements (size >4kb) appear to be overexpressed in ES and iPS nuclear samples.

Table 1: Enrichement of repeat elements among over-expressed nuclear tag-clusters (pAdj<0.05)

	Alu	B2	B4	L1	L2	ERVK	MaLR	ERV1	ERVL	Low complexity	Simple repeat
Nucl. over-expressed	93	26	63	99	13	665	250	66	57	137	124
Total expressed	1731	816	1020	1745	343	3819	1883	499	518	2923	3005
Fisher'exact test (p)	0.08845	0.9874	0.0154	0.02697	0.8124	2.20E-16	2.20E-16	3.92E-14	2.62E-09	0.4839	0.9287

Among the nuclear mouse CAGE tag clusters, we identified 8,958 pairs (distant by <500bp) showing a typical divergent transcription described, which resemble enhancer transcription. 172 of these pairs appears to be significantly upregulated in stem cell samples, while 73 are down regulated (pAdj<0.05).

We have also analyzed mouse short RNA seq data. The clustering of shortRNA libraries based on known and novel miRNA expression levels showed distinct clustering of differentiated and stem cells as well as distinct clustering of MEFs and lymphocytes. The number of novel miRNAs detected in each sample ids shown in Figure 2. After filtering out miRNAs that overlapped known miRNA loci, and those expressed in only two libraries, we identifies 44 novel robust miRNAs.

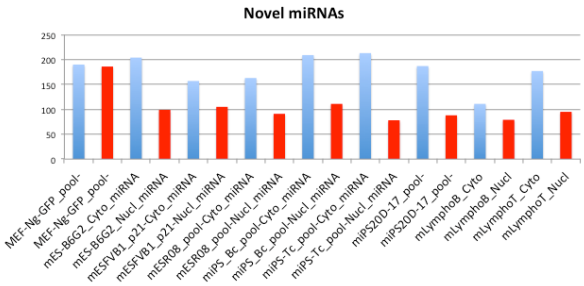

Figure 2. Total number of miRNAs identified in short RNA samples

AIM2: For preliminary analysis, we have selected 123 ncRNA candidates *in silico* for siRNA screening based on our earlier CAGE data. Among them, expression of 73% CAGE tags could be confirmed by qRT-PCR. Then we have screened those candidates with siRNA accompanied with a scrambled control for knock-down efficiency of the siRNA used at 48 hour time point. 12 of the siRNAs were used for next round of screening on knock-down efficiency on stemness. One out of 12 was found efficiently knocked down Nanog, Oct3/4 and Sox2. Later, from the CAGEscan data, we have further selected 48 ncRNA candidates for siRNA screening. After first screening with the siRNA knockdown efficiency, we have selected 32 ncRNA candidates. Second screening with 32 stem specific ncRNA candidates shows that 3 can efficiently knockdown Nanog, oct3/4 and Sox2 factors, confirming that the initial idea behind the project was correct. The knockdown efficiency of two ncRNA out of three was further validated by Cellomics. We are now testing

whether shRNA transduction may KD candidates for which siRNA transfections were inefficient.

We have also start to work on AIM3 by preparing the pipeline for the cloning of ncRNA detected with the genomics data (CAGEscan, CAGE, RNAseq). This pipeline will be fully expanded in FY2012, including the perturbation (by expression of long ncRNAs to test their function in cell stemness).

In addition, we are developing a new methodology to identify chromatin associated RNAs (CARs). The method will simultaneously identify the RNAs interacting with chromatin and DNA sequences where they interact. With this method we are planning to locate the interaction site of potential ncRNAs (capable of differentiation/dedifferentiation) on the chromatin and make model. For this, we have prepared and sequenced CAR libraries with Illumina MiSeq. With preliminary analysis, we found tags for both DNA and RNA sequences and genomic feature analysis shows they can be mapped on protein coding, non coding and repeat elements. We are now deeply sequencing the library on Illumina HiSeq2000. This set will provide strong non-coding RNA candidates for further analysis in the 3rd and 4th year of the project.

3. 研究発表等

雑誌論文 計 3 件	（掲載済み－査読有り） 計 3 件 **Metazoan promoters: emerging** characteristics and insights into transcriptional regulation. Lenhard B, Sandelin A, Carninci P. Nat Rev Genet. 2012 Mar 6;13(4):233-45. doi: 10.1038/nrg3163. PMID: 22392219 Chromatin-associated **RNA** interference components contribute to transcriptional regulation in Drosophila. Cernilogar FM, Onorati MC, Kothe GO, Burroughs AM, Parsi KM, Breiling A, Lo Sardo F, Saxena A, Miyoshi K, Siomi H, Siomi MC, **Carninci** P, Gilmour DS, Corona DF, Orlando V. Nature. 2011 Nov 6;480(7377):391-5. doi: 10.1038/nature10492. PMID: 22056986 Long **non-coding RNA** modifies chromatin: epigenetic silencing by long **non-coding** RNAs. Saxena A, **Carninci** P. Bioessays. 2011 Nov;33(11):830-9. doi: 10.1002/bies.201100084. Epub 2011 Sep 14. Review. PMID: 21915889

様式19 別紙1

会議発表 計 5 件	専門家向け 計 5 件 ・ The Biology of Genomes, Cold Spring Harbor, USA, "The Diversity of Human Small RNAs" Piero Carninci, May 10 – May 14, 2011 ・ Theo Murphy International Scientific Meeting, Buckinghamshire, UK," The complexity of the mammalian transcriptome", Piero Carninci, Sept. 18 – Sept. 21,2011 ・ Cold Spring Harbor Conferences Asia: Bioinformatics of Human and Animal Genomics, Suzhou, China, "COMPLEXITY OF MAMMALIAN TRANSCRIPTION ANALYZED BY DEEPCAGE", Piero Carninci, Nov. 14 – Nov. 16, 2011 ・ Stazoine Zoologica 2011 Seminar Series, Naples, Italy, "The complexity of mammalian transcription" Piero Carninci, Nov. 28, 2011 ・ Functional Genomics & Systems Biology 2011, Cambridge, UK, "COMPLEXITY OF MAMMALIAN TRANSCRIPTION ANALYZED BY DEEPCAGE", Piero Carninci, Nov. 29 – Dec. 1, 2011 一般向け 計 0 件
図 書 計 0 件	
産業財産権 出願・取得状況 計 1 件	（取得済み）計 0 件 （出願中） 計 1 件 Title：FUNCTIONAL NUCLEIC ACID MOLECULE AND USE THEREOF Date of application：March 30,201 Application no：PCT/JP2012/059430
Webページ （URL）	Genome.gsc.riken.jp/osc/English/members/Piero_Carninci.html
国民との科学・技術対話の実施状況	RIKEN Yokohama Science Café、Yokohama City Central Public Library, Yokohama Japan "ゲノムってどうやって調べるの？どう役に立つの？", Piero Carninci March 4,2012.
新聞・一般雑誌等掲載 計 0 件	
その他	

4. その他特記事項

様式19 別紙2

課題番号	LS127

実施状況報告書(平成23年度) 助成金の執行状況

本様式の内容は一般に公表されます

1. 助成金の受領状況(累計) (単位:円)

	①交付決定額	②既受領額(前年度迄の累計)	③当該年度受領額	④(=①−②−③)未受領額	既返還額(前年度迄の累計)
直接経費	138,000,000	47,338,000	0	90,662,000	
間接経費	41,400,000	14,201,400	0	27,198,600	
合計	179,400,000	61,539,400	0	117,860,600	0

2. 当該年度の収支状況 (単位:円)

	①前年度未執行額	②当該年度受領額	③当該年度受取利息等額(未収利息を除く)	④(=①+②+③)当該年度合計収入	⑤当該年度執行額	⑥(=④−⑤)当該年度未執行額	当該年度返還額
直接経費	47,098,012	0	0	47,098,012	46,240,422	857,590	
間接経費	14,201,400	0	0	14,201,400	14,201,400	0	
合計	61,299,412	0	0	61,299,412	60,441,822	857,590	0

3. 当該年度の執行額内訳 (単位:円)

	金額	備考
物品費	24,630,452	シーケンス用キット、実験試薬、実験用消耗品、等
旅費	2,949,048	任期制職員赴任手当等、実習生滞在費、The Bi
謝金・人件費等	17,957,576	職員人件費
その他	703,346	論文投稿料、印刷料
直接経費計	46,240,422	
間接経費計	14,201,400	
合計	60,441,822	

4. 当該年度の主な購入物品(1品又は1組若しくは1式の価格が50万円以上のもの)

物品名	仕様・型・性能等	数量	単価(単位:円)	金額(単位:円)	納入年月日	設置研究機関名
				0		
				0		
				0		