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 ２８ 年　 ６ 月　 １ 日現在

機関番号：８２４０１
研究種目：若手研究(A)
研究期間：2013～2015
課題番号：２５７１００１８

研究課題名（和文）Single cell transcriptome analysis

研究課題名（英文）Single cell transcriptome analysis

研究代表者
ＰＬＥＳＳＹ　Ｃｈａｒｌｅｓ（Plessy, Charles）

国立研究開発法人理化学研究所・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：６０３９１９８４

交付決定額（研究期間全体）：（直接経費）　　19,400,000円

研究成果の概要（和文）：mRNAの定量シーケンスにより、単一細胞において遺伝子の働きを研究する手法を開発した。まず、DNA複製の前後において異なった蛍光色を持つ細胞を解析し、単一細胞の分裂周期の状態を推測することのできる働きを持つ遺伝子を同定した。そして、6000以上の単一細胞をマイクロウェルプレートに単離するためにフローサイトメーターを用い、Labcyteと呼ばれるナノリッターの溶液を分注できる装置を使い、各ウェルでの反応量を減らした。予備実験では遺伝子の働きの違いを測定し、手法の精密さを評価するために、複数の参照細胞株を比較した。
これらの結果は再生医療やがん研究に潜在的効果を持つ細胞分化研究の道を開くであろう。

研究成果の概要（英文）：We have developed methods to study the activity of genes in isolated single cells by quantitative sequencing of messenger RNA copies. First, we analysed cells that have a different fluorescent colour before and after the replication of their DNA, and found genes whose activity can predict the stage of single cells their division cycle. Then, we prepared a high-throughput method using a flow cytometer to isolate more than 6,000 single cells in microwell plates. To keep a reasonable cost, we reduced the reaction volumes in each well, using a high-performance system for transferring nanoliters of liquid reagents, called "Labcyte Echo". In our pilot experiment, we compared reference cell lines, to estimate our method's precision to measure differences of gene activity. Altogether, these results open the way to studies of cell differentiation, which have potential impact on regenerative medicine and cancer research.

研究分野：ゲノム生物学

キーワード：生物学　ハイスループット　トランスクリプトーム　単一細胞

１．研究開始当初の背景

When this project was proposed, the field of single-cell analysis was in its early phase of technological development. Most publications in this field involved manual handling of the cells and were exploring how to analyze this complex data. The methods of choice for gene expression measurements were qPCR and RNA-seq. In bulk samples, we had developed nanoCAGE (nano Cap Analysis Gene Expression, Plessy et al., 2010, Nature Methods), which was more powerful than the usual qRT-PCR and RNA-seq approaches for the detection of gene promoters and the quantitation of their activity. Such data is especially useful for the study and modeling of gene regulation. Our goal was to bring these studies to the level of single cells.

２．研究の目的

The purpose of this project was to develop a CAGE method for single-cell transcriptome analysis and apply it to the study of self-renewal, heterogeneity and differentiation using embryonic stem cells as a model. Initially, we proposed embryonic stem cells as the biological model. However, due to changes in availability of biological material, and appearance in symposiums and in the literature of results from other groups that would have made our planned effort redundant, we decided to switch to a different biological model, the virus-infected epithelial W12E cells. Nevertheless, we expect that our technical developments will be useful beyond these biological models.

３．研究の方法

The core method used in this work is transcriptome analysis with the “nanoCAGE” approach (see point 1 above). Practically, this is a reverse transcription followed by two PCRs, where each of these reactions use specially crafted oligonucleotides in order to amplify the 5′ ends of the whole capped transcriptome, which include coding RNAs and long non-coding RNAs. The resulting transcriptome library is quantitatively sequenced on “next-generation” sequencers, currently using the Illumina technology. The final results are obtained by bioinformatics analysis using a mixture of standard tools and tools specially designed for CAGE analysis (see for example result (3) below).

４．研究成果

(1) We have developed a “nanoCAGE” method for single cells isolated in microwell plates with a flow cytometer. One of the critical limitations of the original protocol for bulk samples was its final amplification step, which was a low-yield PCR step requiring 80 ng of input cDNA. To achieve the efficiency needed for single-cell resolution, we replaced that step with a “tagmentation” procedure requiring only 0.25 ng of the same starting material. Tagmentation is the amplification of DNA molecules fragmented with transposases, and is implemented as standard commercial product. The new optimized nanoCAGE procedure is being published as a book chapter (Poulain et al., 2016, in press).

(2) In parallel, we implemented “C1 CAGE”, a similar single-cell nanoCAGE protocol for the commercial “Fludigm C1” automated platform that allows for the capture, imaging and processing of single-cells in microfluidic chambers (Kouno et al., in preparation). C1 CAGE is already available prior publication for all Fluidigm users.

(3) These new nanoCAGE protocols include a Unique Molecular Identifier (UMI) in each cDNA, which allows us to express the expression levels as molecule counts. We implemented a software workflow to assemble all the sequence reads originating from the same molecule in “CAGEscan fragments”, which are a single-molecule transcript models. The software workflow is available on GitHub (Population-Transcriptomics/C1-CAGE-preview)

(4) Since it was known (for instance through microarray analysis of synchronized culture cells) that the cell cycle has a footprint on transcriptome expression profiles, we searched for cell cycle markers that have predictive power at the single cell level. As a training set, we prepared single-cell RNA-seq libraries using a modified HeLa cell lines carrying “Fucci” transgenes, which are fluorescent markers of the cell cycle stage. We used the Fluidigm C1 platform to capture, image, and further process the cells and therefore obtained matched data of single-cell expression profiles and cell cycle stages (Böttcher et al, submitted). Using publicly available single-cell transcriptomes from other

cell types where the cell cycle stage was also recorded, we confirmed that our markers have predictive powers on other human cell types. (Motakis et al., in preparation). Nevertheless, we could not identify universal markers that have enough power on human and mouse data at the same time. Further work with extended training sets (Fucci mouse cells, …) may be needed to reach that goal.

(4) In preparation for our first large-scale single-cell nanoCAGE experiment, we developed a new method for reducing the fraction of sequence reads originating from rRNA sequences and oligonucleotide artifacts. This method, "pseudo-random primers" is based on the observation that reverse-transcription can be initiated even if there are mismatches at the 3′ end of the primers. Thus, we could design a set of 40 oligonucleotides that cover broadly the transcriptome, while at the same time having a significantly reduced affinity for rRNA and linker sequences (Arnaud et al., 2016, BioTechniques).

(5) We then optimized FACS-sorting conditions to isolate thousands of single cells in 384-well plates. We chose to sort the cells in dry wells and freeze the plates, in order to uncouple the FACS sessions and the reverse-transcription steps.

(6) Our last challenge was to conduce thousands of reverse-transcriptions in a small volume in order to keep the experiment under budget. In particular, in one 384-well plate, each reverse-transcription mixture must contain a different sample identifier. This ruled out the use of standard micropipettes since our goal was to have a total volume of maximum 1 µL. Instead, we used the "Labcyte Echo" platform, which transfers 2.5 nL droplets from source reagent containers to destination plates using acoustic waves.

(7) Using this technique, we produced a data set of thousand of single cells from the HeLa, SiHa, CaSki, and W12E cell lines. High-throughput sequencing of the nanoCAGE libraries is under way as this report is being written. All of these cell lines are related to the biology of virus-induced cervix cancers. The W12E cell line is a model of cervix epithelium differentiation in the context of viral infection and we plan further studies including a differentiation time course of these W12E cells.

５．主な発表論文等

（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 2 件）

① Targeted reduction of highly abundant transcripts using pseudo-random primers. Ophélie Arnaud, Sachi Kato, Stéphane Poulain, and Charles Plessy. BioTechniques, Vol. 60, No. 4, April 2016, pp. 169–174

② nanoCAGE: a method for the analysis of coding and non-coding 5'-capped transcriptomes. Stéphane Poulain, Sachi Kato, Ophélie Arnaud, Jean-Étienne Morlighem, Makoto Suzuki, Charles Plessy, Matthias Harbers. Methods in Molecular Biology (Walker, J.M., ed.). In press.

〔学会発表〕（計３件）

① The 29th International Mammalian Genome Conference, Yokohama Memorial Hall, Japan, 2015. 1 talk on "Single-molecule RNA sequencing in single cells" (Charles Plessy) and 2 posters on "Transcriptome analysis of FACS-sorted single cells with nanoCAGE" (Stéphane Poulain) and "Targeted reduction of highly abundant transcripts using pseudo-random primers" (Ophélie Arnaud).

② The 11th International Workshop on Advanced Genomics, Hitotsubashi Hall, Tokyo, Japan, 2015. 2 posters on "C1 CAGE, single-molecule reconstruction and pseudo-random primers for single-cell population transcriptomics" (Charles Plessy), and "Transcriptome analysis of FACS-sorted single cells with nanoCAGE" (Stéphane Poulain).

③ 2nd UK/Japan workshop on epigenetics and transcription. 1 talk on "Single-cell markers of the cell cycle" (Charles Plessy).

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計０件）

名称：

発明者：

権利者：

種類：

番号：

出願年月日：

国内外の別：

○取得状況（計０件）

名称：

発明者：

権利者：

種類：

番号：

取得年月日：

国内外の別：

〔その他〕

ホ　ー　ム　ペ　ー　ジ　等　：
http://population-transcriptomics.org/

６．研究組織

(1)研究代表者

ＰＬＥＳＳＹ　Ｃｈａｒｌｅｓ(PLESSY, Charles)

国立研究開発法人理化学研究所・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：60391984

(2)研究分担者

なし

(3)連携研究者

なし