

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第6825768号
(P6825768)

(45) 発行日 令和3年2月3日 (2021. 2. 3)

(24) 登録日 令和3年1月18日 (2021.1.18)

(51) Int. Cl.

F I

C 1 2 N 15/10 (2006.01)

C 1 2 Q 1/6851 (2018.01)

C 1 2 Q 1/686 (2018.01)

C 1 2 N 15/10 Z N A Z

C 1 2 Q 1/6851 Z

C 1 2 Q 1/686 Z

請求項の数 18 (全 38 頁)

(21) 出願番号	特願2017-502400 (P2017-502400)	(73) 特許権者	504132272
(86) (22) 出願日	平成28年2月23日 (2016. 2. 23)		国立大学法人京都大学
(86) 国際出願番号	PCT/JP2016/055314		京都府京都市左京区吉田本町 3 6 番地 1
(87) 国際公開番号	W02016/136766	(74) 代理人	100080791
(87) 国際公開日	平成28年9月1日 (2016. 9. 1)		弁理士 高島 一
審査請求日	平成31年2月22日 (2019. 2. 22)	(74) 代理人	100125070
(31) 優先権主張番号	特願2015-33432 (P2015-33432)		弁理士 土井 京子
(32) 優先日	平成27年2月23日 (2015. 2. 23)	(74) 代理人	100136629
(33) 優先権主張国・地域又は機関	日本国 (JP)		弁理士 鎌田 光宣
		(74) 代理人	100121212
			弁理士 田村 弥栄子
		(74) 代理人	100163658
			弁理士 小池 順造
		(74) 代理人	100174296
			弁理士 當麻 博文

最終頁に続く

(54) 【発明の名称】 核酸配列増幅方法

(57) 【特許請求の範囲】

【請求項 1】

生物学的試料における遺伝子発現量の相対的な関係を保持している増幅産物を含む核酸集団を調製する方法であって、

- (a) センス鎖とアンチセンス鎖とからなる 2 重鎖 DNA であって、該センス鎖は、 (1) 任意の配列からなる付加核酸配列 X、 (2) ポリ T 配列、 (3) 生物学的試料から単離した mRNA 配列を鋳型として使用して調製された c DNA のセンス鎖配列、 (4) ポリ A 配列および (5) 付加核酸配列 X とは異なる任意の配列からなる付加核酸配列 Y を、 5 ' から 3 ' の方向にこの順序で含む、 2 重鎖 DNA を鋳型として、 5 ' 末端にアミンを付加した付加核酸配列 X を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 1 のプライマーと、付加核酸配列 Y を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 2 のプライマーとを用いて、該 2 重鎖 DNA を増幅する工程、
- (b) 前記工程 (a) により得られた 2 重鎖 DNA を断片化する工程、
- (c) 前記工程 (b) により得られた断片化 2 重鎖 DNA の 5 ' 末端をリン酸化して、 5 ' 末端がリン酸化された 2 重鎖 DNA を得る工程、
- (d) 前記工程 (c) により得られた 5 ' 末端がリン酸化された 2 重鎖 DNA を鋳型として、任意の配列からなる付加核酸配列 Z および前記付加核酸配列 Y をこの順序で含み、任意でその下流にポリ T 配列をさらに含んでもよい第 3 のプライマーを用いて c DNA を調製、該 c DNA の 3 ' 末端へアデニン (A) を付加する工程、
- (e) 前記工程 (d) により得られた 2 重鎖 DNA へ、 3 ' 末端にチミン (T) をオーバ

ーハングで有する任意の配列 V を含む 2 重鎖 DNA を連結させる工程、および
 (f) 前記工程 (e) により得られた 2 重鎖 DNA を鋳型として、前記配列 V を含む第 4
 のプライマーと、前記付加核酸配列 Z を含み、任意でその下流に前記付加核酸配列 Y をさ
 らに含んでもよい第 5 のプライマーとを用いて、該 2 重鎖 DNA を増幅する工程
 を含む方法。

【請求項 2】

前記工程 (a) で用いる 2 重鎖 DNA が、次の工程を含む方法によって調製される、請
 求項 1 に記載の方法：

(i) 生物学的試料から単離した mRNA を鋳型として、前記付加核酸配列 Y およびポリ
 T 配列とからなる第 6 のプライマーを用いて逆転写することにより一次鎖 c DNA を調製
 する工程、

10

(i i) 工程 (i) により得られた一次鎖 c DNA をポリ A テーリング反応に付し、次い
 でこれを鋳型として、前記付加核酸配列 X およびポリ T 配列とからなる第 7 のプライマー
 を用いて二次鎖である 2 重鎖 DNA を調製する工程、および

(i i i) 工程 (i i) により得られた 2 重鎖 DNA を、前記付加核酸配列 X を含み、任
 意でその下流にポリ T 配列をさらに含んでもよい第 8 のプライマーと、前記付加核酸配列
 Y を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 9 のプライマーとを用い
 て増幅する工程。

【請求項 3】

前記工程 (b) における断片化が、超音波処理で行われる、請求項 1 または 2 に記載の
 方法。

20

【請求項 4】

前記工程 (c) において、5' 末端のリン酸化と同時に末端の平滑化を行う、請求項 1
 から 3 のいずれか 1 項に記載の方法。

【請求項 5】

前記工程 (c) において、さらに、200 塩基から 250 塩基、または 300 塩基から
 350 塩基の大きさの断片化 2 重鎖 DNA を選択する工程を含む、請求項 1 から 4 のいづ
 れか 1 項に記載の方法。

【請求項 6】

前記工程 (a) において、増幅が、2 から 8 サイクルの PCR によって行われる、請求
 項 1 から 5 のいずれか 1 項に記載の方法。

30

【請求項 7】

前記工程 (f) において、増幅が、5 から 20 サイクルの PCR によって行われる、請
 求項 1 から 6 のいずれか 1 項に記載の方法。

【請求項 8】

前記工程 (i i i) において、増幅が、5 から 30 サイクルの PCR によって行われる
 、請求項 2 から 7 のいずれか 1 項に記載の方法。

【請求項 9】

前記工程 (f) において用いる第 5 のプライマーが、さらにバーコード配列を含む、請
 求項 1 から 8 のいずれか 1 項に記載の方法。

40

【請求項 10】

生物学的試料が 1 個から数個の細胞である、請求項 1 から 9 のいずれか 1 項に記載の方
 法。

【請求項 11】

生物学的試料が 1 個の細胞である、請求項 10 に記載の方法。

【請求項 12】

請求項 1 から 11 のいずれか 1 項に記載の方法で調製された核酸集団における前記増幅
 された 2 重鎖 DNA の量を、次世代シーケンサーを用いて測定することを含む、当該核酸
 集団を調製した細胞に含まれる mRNA 量の測定方法。

【請求項 13】

50

以下を含む、請求項 1 の方法において使用するためのキット。

(a) 5 ' 末端にアミンを付加した任意の配列からなる付加核酸配列 X を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 1 のプライマー

(b) 任意の配列からなる付加核酸配列 Y を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 2 のプライマー

(c) 任意の配列からなる付加核酸配列 Z および前記付加核酸配列 Y をこの順序で含み、任意でその下流にポリ T 配列をさらに含んでもよい第 3 のプライマー

(d) 3 ' 末端にチミン (T) をオーバーハングで有する任意の配列 V を含む 2 重鎖 DNA

(e) 前記配列 V を含む第 4 のプライマー

(f) 前記付加核酸配列 Z を含み、任意でその下流に前記付加核酸配列 Y をさらに含んでもよい第 5 のプライマー。

【請求項 1 4】

前記 (f) の第 5 のプライマーが、さらにバーコード配列を含む、請求項 1 3 に記載のキット。

【請求項 1 5】

(g) 前記付加核酸配列 Y およびポリ T 配列からなる第 6 のプライマー、

(h) 前記付加核酸配列 X およびポリ T 配列からなる第 7 のプライマー、

(i) 前記付加核酸配列 X を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 8 のプライマー、および

(j) 前記付加核酸配列 Y を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 9 のプライマー

をさらに含む、請求項 1 3 または 1 4 に記載のキット。

【請求項 1 6】

前記第 6 のプライマーと前記第 9 のプライマーとが同一、および / または、前記第 7 のプライマーと前記第 8 のプライマーとが同一である、請求項 1 5 に記載のキット。

【請求項 1 7】

前記第 2 のプライマーと前記第 9 のプライマーとが同一である、請求項 1 5 または 1 6 に記載のキット。

【請求項 1 8】

DNA 増幅に用いるポリメラーゼをさらに含む、請求項 1 3 から 1 7 のいずれか一項に記載のキット。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、次世代シーケンサーを用いて mRNA の定量を行うための試料を作製するための核酸配列増幅方法、詳細には少数の細胞、好ましくは単一細胞レベルでの、次世代シーケンサーを用いた mRNA の定量解析を可能にする核酸配列増幅方法に関する。

【背景技術】

【0002】

単一細胞での定量トランスクリプトーム解析は、発生学、幹細胞およびガン研究を行うための重要なツールである。この単一細胞による解析を行うためには、単一細胞中の mRNA を逆転写して製造する cDNA を増幅する必要がある、この増幅法として 2 つの方法が提示されている。一つは、PCR による増幅法であり、もう一つは、T7 RNA ポリメラーゼによる増幅方法である。PCR を用いる場合、増幅効率が高く、簡便で安定性の高い方法であるため、単一細胞のトランスクリプトーム解析には、とても有用である。

【0003】

単一細胞からの cDNA の定量的な増幅を確認するため、増幅した cDNA を網羅的に解析する方法として、長い第一の cDNA を合成し、増幅産物を RNA シークエンス (RNA-seq) により分析する方法が提案されている (非特許文献 1 ~ 3) 。他の方法として、単一細胞からの cD

10

20

30

40

50

NAの増幅方法は、テンプレートスイッチング法と呼ばれる方法で増幅されたcDNAの全長をRNA-seqに用いている（非特許文献4および5）。転写産物の絶対的定量を進めるため、第一のcDNAの5'側または3'側にタグを付けおこなう単一分子識別（UMI）などが行われている（非特許文献6～10）。この他にも、バーコード配列により個々の細胞を認識することで分析する方法やマイクロ流路により単一細胞を捕獲して行う方法などが提案されている（非特許文献11および12）。

【0004】

このように、RNA-seqの方法が提案されているが、定量の再現性・精度が問題視されている。すなわち、RNA-seqは、上述のとおりPCR増幅を伴う例が多いが、PCRは増幅率が100%ではないので、特に少数コピーから増幅された場合には、増幅後のコピー数の再現性が悪い。また単一細胞由来のcDNAを解析するには多数のサンプルを解析する必要がある。既存の方法では一細胞あたりの解析単価が高いため、多くの細胞を解析することが難しい。

【先行技術文献】

【非特許文献】

【0005】

【非特許文献1】Tang, F., et al., Nat Methods, 6, 377-382, 2009

【非特許文献2】Tang, F., et al., Nature protocols, 5, 516-535, 2010

【非特許文献3】Sasagawa, Y., et al., Genome Biol, 14, R31, 2013

【非特許文献4】Ramskold, D., et al., Nat Biotechnol, 30, 777-782, 2012

【非特許文献5】Picelli, S., et al., Nat Methods, 10, 1096-1098, 2013

【非特許文献6】Islam, S., et al., Genome Res, 21, 1160-1167, 2011

【非特許文献7】Kivioja, T., et al., Nat Methods, 9, 72-74, 2012

【非特許文献8】Islam, S., et al., Nat Methods, 11, 163-166, 2014

【非特許文献9】Hashimshony, T., et al., Cell reports, 2, 666-673., 2012

【非特許文献10】Grun, D., et al., Nat Methods, 11, 637-640, 2014

【非特許文献11】Streets, A.M., et al., Proc Natl Acad Sci U S A, 111, 7048-7053, 2014

【非特許文献12】Jaitin, D.A., et al., Science, 343, 776-779, 2014

【発明の概要】

【発明が解決しようとする課題】

【0006】

このように、これまでの核酸配列増幅方法により得られた試料は、cDNAの全長を対象としているため、cDNAが長くなると次世代シーケンサーを用いたmRNAの解析の定量性が低くなる。また多数のサンプルを同時に解析するため、より定量性の高い試料を低コストで調製するための核酸配列の増幅方法が求められる。

【課題を解決するための手段】

【0007】

次世代シーケンサーを用いたmRNAの解析の定量性を高め、かつ解析単価を低くする必要がある。そこで、本発明では以下の手法により、mRNAが有するポリA配列を利用して、cDNAを増幅し、さらに、断片化した後、選択的にプライマー配列を付加することで、3'末端側のみを含む試料を得ることに成功した。これにより、より定量性高くかつ多数のサンプルの同時解析を可能にするSC3-seq (Single cell mRNA 3' end sequence) 法を開発した。

【0008】

即ち、本発明は、以下に関する；

[1] 生物学的試料における遺伝子発現量の相対的な関係を保持している増幅産物を含む核酸集団を調製する方法であって、

(a) 任意の付加核酸配列X、ポリT配列、生物学的試料から単離したmRNA配列、ポリA配列および任意の付加核酸配列Yの順で構成される2重鎖DNAを鋳型として、5'末端にアミンを付加した任意の付加核酸配列Xを含み、任意でその下流にポリT配列をさ

10

20

30

40

50

らに含んでもよい第1のプライマーと、任意の付加核酸配列Yを含み、任意でその下流にポリT配列をさらに含んでもよい第2のプライマーとを用いて、該2重鎖DNAを増幅する工程、

(b) 前記工程(a)により得られた2重鎖DNAを断片化する工程、

(c) 前記工程(b)により得られた断片化2重鎖DNAの5'末端をリン酸化する工程、

(d) 前記工程(c)により得られた5'末端がリン酸化された2重鎖DNAを鋳型として、任意の付加核酸配列Zおよび前記付加核酸配列Yをこの順序で含み、任意でその下流にポリT配列をさらに含んでもよい第3のプライマーを用いてcDNAを調製、該cDNAの3'末端へアデニン(A)を付加する工程、

(e) 前記工程(d)により得られた2重鎖DNAへ、3'末端にチミン(T)をオーバーハングで有する任意の配列Vを含む2重鎖DNAを連結させる工程、および

(f) 前記工程(e)により得られた2重鎖DNAを鋳型として、前記配列Vを含む第4のプライマーと、前記付加核酸配列Zを含み、任意でその下流に前記付加核酸配列Yをさらに含んでもよい第5のプライマーとを用いて、該2重鎖DNAを増幅する工程を含む方法。

[2] 前記工程(a)で用いる任意の付加核酸配列X、ポリT配列、生物学的試料から単離したmRNA配列、ポリA配列および任意の付加核酸配列Yの順で構成される2重鎖DNAが、次の工程を含む方法によって調製される、[1]に記載の方法：

(i) 生物学的試料から単離したmRNAを鋳型として、前記付加核酸配列YおよびポリT配列とからなる第6のプライマーを用いて逆転写することにより一次鎖cDNAを調製する工程、

(ii) 工程(i)により得られた一次鎖cDNAをポリAテリング反応に付し、次いでこれを鋳型として、前記付加核酸配列XおよびポリT配列とからなる第7のプライマーを用いて二次鎖である2重鎖DNAを調製する工程、および

(iii) 工程(ii)により得られた2重鎖DNAを、前記付加核酸配列Xを含み、任意でその下流にポリT配列をさらに含んでもよい第8のプライマーと、前記付加核酸配列Yを含み、任意でその下流にポリT配列をさらに含んでもよい第9のプライマーとを用いて増幅する工程。

[3] 前記工程(b)における断片化が、超音波処理で行われる、[1]または[2]に記載の方法。

[4] 前記工程(c)において、5'末端のリン酸化と同時に末端の平滑化を行う、[1]から[3]のいずれか1項に記載の方法。

[5] 前記工程(c)において、さらに、200塩基から250塩基、または300塩基から350塩基の大きさの断片化2重鎖DNAを選択する工程を含む、[1]から[4]のいずれか1項に記載の方法。

[6] 前記工程(a)において、増幅が、2から8サイクルのPCRによって行われる、[1]から[5]のいずれか1項に記載の方法。

[7] 前記工程(f)において、増幅が、5から20サイクルのPCRによって行われる、[1]から[6]のいずれか1項に記載の方法。

[8] 前記工程(iii)において、増幅が、5から30サイクルのPCRによって行われる、[2]から[7]のいずれか1項に記載の方法。

[9] 前記工程(f)において用いる第5のプライマーが、さらにバーコード配列を含む、[1]から[8]のいずれか1項に記載の方法。

[10] 生物学的試料が1個から数個の細胞である、[1]から[9]のいずれか1項に記載の方法。

[11] 生物学的試料が1個の細胞である、[10]に記載の方法。

[12] [1]から[11]に記載の方法で調製された核酸集団における前記増幅された2重鎖DNAの量を、次世代シーケンサーを用いて測定することを含む、当該核酸集団を調製した細胞に含まれるmRNA量の測定方法。

10

20

30

40

50

[1 3] 以下を含む、次世代シーケンサーによる mRNA 量の測定に適用するための cDNA 集団を調製するためのキット。

(a) 5' 末端にアミンを付加した任意の付加核酸配列 X を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 1 のプライマー

(b) 任意の付加核酸配列 Y を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 2 のプライマー

(c) 任意の付加核酸配列 Z および前記付加核酸配列 Y をこの順序で含み、任意でその下流にポリ T 配列をさらに含んでもよい第 3 のプライマー

(d) 3' 末端にチミン (T) をオーバーハングで有する任意の配列 V を含む 2 重鎖 DNA

(e) 前記配列 V を含む第 4 のプライマー

(f) 前記付加核酸配列 Z を含み、任意でその下流に前記付加核酸配列 Y をさらに含んでもよい第 5 のプライマー

[1 4] 前記 (f) の第 5 のプライマーが、さらにバーコード配列を含む、[1 3] に記載のキット。

[1 5] (g) 前記付加核酸配列 Y およびポリ T 配列からなる第 6 のプライマー、

(h) 前記付加核酸配列 X およびポリ T 配列からなる第 7 のプライマー、

(i) 前記付加核酸配列 X を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 8 のプライマー、および

(j) 前記付加核酸配列 Y を含み、任意でその下流にポリ T 配列をさらに含んでもよい第 9 のプライマー

をさらに含む、[1 3] または [1 4] に記載のキット。

[1 6] 前記第 6 のプライマーと前記第 9 のプライマーとが同一、および / または、前記第 7 のプライマーと前記第 8 のプライマーとが同一である、[1 5] に記載のキット。

[1 7] 前記第 2 のプライマーと前記第 9 のプライマーとが同一である、[1 5] または [1 6] に記載のキット。

[1 8] DNA 増幅に用いるポリメラーゼをさらに含む、[1 3] から [1 7] のいずれか一項に記載のキット。

【発明の効果】

【 0 0 0 9 】

本発明は、簡便な PCR 法による、オリゴヌクレオチドマイクロアレイへ直接応用できる、信頼性の高い、定量的な極微量 cDNA 増幅技術を提供する。本発明方法では、一日の実験で単一細胞からマイクロアレイ実験に十分な量の鋳型 cDNA を合成・増幅することができる。従来法と本発明方法の比較を、いくつかの遺伝子産物をプローブにした、リアルタイム PCR 実験を用いて行い、システムティック誤差 (系統誤差)、ランダム誤差ともに疑う余地もなく著しく改善されていることが確認された。さらに、本発明方法を用いて行ったトランスクリプトーム解析実験では、従来法よりも遥かに改善された、再現性の良い定量的な単一細胞レベルでの解析が可能となったことが確認された。

【図面の簡単な説明】

【 0 0 1 0 】

【図 1】図 1 A は、SC3-seq の概念図を示す。SC3-seq は、図中破線枠で示した 3' 末端側のみを解析の対象とすることを意味する。図 1 B は、マウス mm10 データベースにアノテーションされた 21,254 個のタンパク質をコードする遺伝子を転写産物の長さごとに整列させたグラフを示す (左図)。右図は、全ての転写産物の長さの合計が、60 Mbp であるところ、全ての転写産物の 3' 末端から 200 bp の合計は、4 Mbp であることを示すグラフである。図 1 C は、SC3-seq のスキームを示す。左図は、cDNA の合成と増幅の工程を示し、右図は、ライブラリー構築の工程を示す。図 1 D は、mESC から抽出した RNA の 100ng において、平均 SC3-seq トラック (リード密度 (RPM, $\times 1,000$ リード) を、アノテーションされた TS (転写終結サイト) からのリードの位置に対してプロットしたグラフを示す。図中、赤線は、センス鎖でマッピングされたリードのトラックを示し、青線は、アンチセンス鎖で

10

20

30

40

50

マッピングされたリードのトラックを示す。図 1 Eは、Pou5f1およびNanog遺伝子座のSC3-seqリードの位置を示す。赤いピークは、センス鎖でマッピングされたリードを示し、青いピークは、アンチセンス鎖でマッピングされたリードを示す。図 1 Fは、3'末端延長の定義拡張に対して、遺伝子数をプロットしたグラフを示す。遺伝子数は、カラーコードごとにリード数が増加することを示している。図 1 Gは、横軸に示した長さに対して、正しい遺伝子数（黒バー）またはTTSの定義の拡張によるミスアノテーションした遺伝子数を示したグラフである。10 KbまでTTSを定義を拡張された2倍以上の遺伝子数（図 1 Fにおける $\times 2$, $\times 3$, $\times 4$ ）を提示する205個の遺伝子が、正しい遺伝子数またはミスアノテーションした遺伝子数に対して公開されたRNA-seqデータと比較することで検出された。

【図 2】図 2 Aは、SC3-seq [$\log_2(\text{RPM}+1)$]ごとに、Q-PCR (CT値) によって表された増幅されたcDNA (SC3-seqによるライブラリー構築前の1 ng (中央図) および10 pg (右図) の全RNA) の発現レベルと比較されたグラフを示す。左図は、比較の方法の概念図を示す。図 2 Bは、mESCの全RNA (MS01T01およびMS01T17、それぞれ100 ngおよび10 pgの全RNAを示す) の希釈におけるERCC RNAの量とSC3-seq [$\log_2(\text{RPM}+1)$]によるSC3-seq [$\log_2(\text{RPM}+1)$]のレベルの間の関係を示すグラフである。10 pgあたり10コピー以上を有するERCC spike-in RNAに対するSC3-seqデータが回帰直線のために用いられた。図 2 Cは、mESCの全RNAの100 ng, 10 ng, 1 ng, 100 pgおよび10 pgから2つの独立して増幅された複製物間の比較を示した散布図である。図中、白色および黄色の領域は、それぞれ、2倍および4倍の違いがある発現レベルを示す。図中、100 ngのRNAにおけるERCC spike-in RNAのSC3-seqリードにより検出された10 pgの全RNAあたりのコピー数を点線（縦線）で示した（右から、1,000コピー、100コピー、10コピーおよび1コピー）。図 2 Dは、mESCの全RNAの10 ng, 1 ng, 100 pgおよび10 pgからの複製物と100 ngの全RNAからの複製物間の比較を示した散布図である。図中、白色および黄色の領域は、それぞれ、2倍および4倍の違いがある発現レベルを示す。図中、100 ngのRNAにおけるERCC spike-in RNAのSC3-seqリードにより検出された10 pgの全RNAあたりのコピー数を点線（縦線）で示した（右から、1,000コピー、100コピー、10コピーおよび1コピー）。図 2 Eは、mESCの全RNAの100 ngの平均SC3-seqデータ ($\log_2(\text{RPM}+1)$) と10 pgの全RNAの平均SC3-seqデータを比較した散布図を示す。図中、白色および黄色の領域は、それぞれ、2倍および4倍の違いがある発現レベルを示す。図中、100 ngのRNAにおけるERCC spike-in RNAのSC3-seqリードにより検出された10 pgの全RNAあたりのコピー数を点線（縦線）で示した（右から、1,000コピー、100コピー、10コピーおよび1コピー）。図 2 Fは、8つの10 pgのRNAサンプルにおけるSC3-seqによる遺伝子発現レベルに対する遺伝子発現レベルの標準偏差をプロットしたグラフである。図 2 Gは、相関係数 (R^2) の最小値 (min) および最大値 (max) ならびに2倍および4倍の発現差を有する遺伝子（全ての遺伝子発現および10pgあたり20コピー以上発現した遺伝子）のパーセンテージを比較した関係を示す。図 2 Hは、100 ng, 10 ng, 1 ng, 100 pgおよび10 pgのmESCの全RNAにおけるERCC spike-in RNAのコピー数から算出された10 pg RNAあたりの全mRNA分子数を示すグラフである。

【図 3】図 3 Aは、100 ngの全RNAにおける発現レベル ($\log_2(\text{RPM}+1)$) の機能として、10 pgの全RNAからのSC3-seqのカバレッジを示すグラフである。黒線は、単一サンプル分生におけるカバレッジの平均値を示す。8つの増幅サンプルの 1-8において検出された転写産物の検出定義の下での多数サンプル解析の結果をそれぞれ示す。図中、100 ngのRNAにおけるERCC spike-in RNAのSC3-seqリードによって算出された10 pgの全RNAあたりのコピー数を点線（縦線）で示す（右から、1,000コピー、100コピー、10コピーおよび1コピー）。図 3 Bは、発現レベル ($\log_2(\text{RPM}+1)$) の機能として、10 pgの全RNAからのSC3-seqの正確さ (Accuracy) を示すグラフである。黒線は、単一サンプル分生における正確さの平均値を示す。8つの増幅サンプルの 1-8において検出された転写産物の検出定義の下での多数サンプル解析の結果をそれぞれ示す。図中、100 ngのRNAにおけるERCC spike-in RNAのSC3-seqリードによって算出された10 pgの全RNAあたりのコピー数を点線（縦線）で示す（右から、1,000コピー、100コピー、10コピーおよび1コピー）。図 3 Cは、リードの機能として、遺伝子数 ($\log_2(\text{RPM}+1)$) 4、全長リードにより決定された遺伝子発現

10

20

30

40

50

レベルと比較して 2倍)を100 ng, 10 ng, 1 ng, 100 pgおよび10 pgのmESCの全RNAからのSC3-seqごとにプロットしたグラフである。図3Dは、リードの機能として、パーセンテージ(全長リードにより決定された遺伝子発現レベルと比較して 2倍)を100 ngの全RNAにおける発現レベルの範囲によって分類された10 pgのmESCの全RNAからのSC3-seqごとにプロットしたグラフである。

【図4】図4Aは、SC3-seq(100 ng(複製1回、MS01T01)、10 ng(複製1回、MS01T05)および10 pg(複製1回、MS01T17))のESCの全RNAからの希釈サンプル、ならびに単一のmESC(MS04T18)および単一のヒトESC(MS04T66)、Smart-seq2(1 ng(Smart-seq2_1ng, HEK_rep1)および10 pg(Smart-seq2_10 pg, HEK_rep1)のHEK293の全RNAからの希釈サンプル、ならびに単一のmESCおよび単一のマウス胚性線維芽細胞(Smart-seq2_MEF、複製1回)およびYanらによる単一細胞のRNA-seq(単一のヒトESC(Yan_hESC_1)および全長RNA-seq(Ohta_mESCおよびOhta_MEF))によって検出された発現レベルと転写産物の長さをプロットしたグラフである。SC3-seqによる発現レベルは、 \log_2 RPMで表し、他の方法では、 \log_2 FPKMで表した。各散布図右側のヒストグラムは、異なる長さの転写産物の遺伝子発現レベルの分布を示している。図4Bは、3つの単一細胞によるRNA-seq法による転写産物の3'末端でマッピングされたリードの分布を示すグラフである。図4Cは、3つの単一細胞によるRNA-seq法による転写産物の長さごとのマッピングされたリードの分布を示すグラフである。図4Dは、3つの単一細胞によるRNA-seq法による全て(上図)および短い転写産物(1 Kbp未満、マウスおよびヒトでそれぞれ913および832個の遺伝子)(下図)に対する検出限界(マウスおよびヒトでそれぞれ、遺伝子発現レベルが、上位6555番目以上および6217番目以上(マウスおよびヒトに対してアノテーションされた全ての転写産物の1/4)、 $\sim \log_2$ RPM 3.69 ± 0.05 (SC3-seq), $\sim \log_2$ FPKM 2.21 ± 1.28 (Yan et al.), and $\sim \log_2$ FPKM 2.92 ± 0.27 (Picelli et al.)、全長リードによる遺伝子発現レベルと比較して 2倍である)を分析したグラフである。

【図5】図5AおよびBは、全ての発現遺伝子(全てのサンプルにおいて \log_2 (RPM+1) 4、12,010遺伝子)でのUnsupervised hierarchical clustering (UHC)(図5A)およびエピブラスト、原始内胚葉(PE)および栄養外胚葉(TE)に対するマーカー遺伝子の発現レベルのヒートマップ(図5B)を示す。アノテーションされた細胞型(エピブラスト、PE、極性TEおよび壁TE)は、分類、位置およびマーカー遺伝子の発現で定義された。図5Cは、全ての発現遺伝子による細胞のPrincipal component analysis (PCA)の結果を示す。PC1およびPC2(上図)またはPC1およびPC3(下図)で展開した図として示す。図5Dは、エピブラスト(9サンプル)およびPE(9サンプル)間(左図)、ならびにmTE(9サンプル)およびpTE(10サンプル)間(右図)における平均遺伝子発現の差をプロットしたグラフである。遺伝子発現の差は、平均 \log_2 (RPM+1) 4である一つの細胞種において4倍差以上のものを示す。PE(504遺伝子)、エピブラスト(309遺伝子)、pTE(231遺伝子)およびmTE(391遺伝子)で発現が上昇している遺伝子としてそれぞれ、青色、緑色、黄色および赤色で示す。図5Eは、4つの細胞種のうちmTE(左図)またはpTE(右図)で発現が上昇している遺伝子の発現を示すグラフである。図中、ボックスの中にあるバーは、平均発現レベルを示す。図5Fは、mTE(上図)またはpTE(下図)で発現が上昇している遺伝子のGene ontology (GO)分析の結果を示す。図5Gは、4つの細胞種におけるERCC spike-in RNAのコピー数に基づき算出された平均遺伝子発現レベルを示すグラフである。図中、ボックスの中にあるバーは、平均発現レベルを示す。

【図6】図6Aは、SNLフィーダー細胞上で培養したhiPSC colonies (585B1)(上図)およびフィーダーフリー条件で培養した同細胞の位相差顕微鏡像を示す。図6Bは、全ての発現遺伝子(全てのサンプルにおいて \log_2 (RPM+1) 4、12,406遺伝子)でのUHCの結果および遺伝子発現レベルのヒートマップを示す。図6Cは、発現する全ての遺伝子による全ての細胞のPCAの結果を示す。細胞をPC1およびPC2によって展開してプロットした。図6Dは、フィーダー細胞上(左図)およびフィーダーフリー条件(右図)でのhiPSCの遺伝子発現を各グループ(図6CにおけるMS04T72, MS04T67およびMS04T78)での最大発現レベルを標準偏差(SD)に対してプロットしたグラフを示す。最大遺伝子発現レベルが、6お

10

20

30

40

50

よびSDが、2である遺伝子を不均一発現遺伝子とした（フィーダー細胞上で培養したhiPSCでは699遺伝子、フィーダーフリー条件でのhiPSCでは61遺伝子）。図6Eは、図6Dでの不均一発現遺伝子の関係を表すベン図を示す。図6Fは、フィーダー細胞上で培養したhiPSCとフィーダーフリー条件でのhiPSCにおける遺伝子発現レベルのSD値をプロットしたグラフを示す。

【図7】図7Aは、SC3-seqにより3'末端でライブラリー構築が可能とするメカニズムの概略図を示す。断片化後、5'末端側でV3タグを有する断片、タグを持たない断片、およびV1タグを有する断片の3つの断片を生じること示す。これらは全てポリッシュされ平滑化末端においてリン酸化されていることを示す。内部アダプターの延長工程において、V1タグを付加された3'末端を有する断片のみInt配列を付加される。P1アダプターをライゲーションする工程において、ポリッシュされた部位において、V3タグ末端を有する断片および内部断片にもP1アダプターを付加されるが、IntV1タグを有する断片の3'末端はリン酸基を持たないため、P1アダプターが付加されない。最後に、P1およびIntV1タグを持つ断片のみ増幅することで、選択的に3'末端でライブラリー構築が可能となる。図7Bは、100 ng, 10 ng, 1 ng, 100 pgおよび10 pgのESCの全RNAから増幅されたERCC spike-in RNAの増幅レベルのQ-PCRの結果を示す。10 pgのRNAあたりのコピー数および一致するERCCコードを示す。図7Cは、増幅されたcDNA（1 ng（中央図、4サンプル）および10 pg（右図、16サンプル）の全RNAからのV1V3 cDNA）の発現レベル（Q-PCRのCT値）およびSC3-seqライブラリーの発現レベル（Q-PCRのCT値）を比較したグラフを示す（全RNAの1 ngおよび10 pgに対してそれぞれMS01T05およびMS01T17とした）。左図は、比較の方法の概念図を示す。

【図8】図8Aは、Let7a-7d遺伝子座でのリードの検出を示す。上段は、SC3-seqのセンス鎖でのマッピングを示し、中段は、SC3-seqのアンチセンス鎖でのマッピングを示し、下段は、Ohta et alによるマッピングを示す。図8Bは、Mir290-295の遺伝子座でのリードの検出を示す。ノンコーディングD7Ertd143eは、Mir290-295の前駆体を示す。上段は、SC3-seqのセンス鎖でのマッピングを示し、中段は、SC3-seqのアンチセンス鎖でのマッピングを示し、下段は、Ohta et alによるマッピングを示す。図8Cは、Mir684-1の遺伝子座でのリードの検出を示す。単一のmiRNAは、Dusp19をコードする遺伝子のイントロンにコードされている。上段は、SC3-seqのセンス鎖でのマッピングを示し、中段は、SC3-seqのアンチセンス鎖でのマッピングを示し、下段は、Ohta et alによるマッピングを示す。図8Dは、分類されていないノンコーディングRNA（Gm19693）のリードの検出を示す。H2afzの3'末端の逆鎖でアノテーションをされている。上段は、SC3-seqのセンス鎖でのマッピングを示し、中段は、SC3-seqのアンチセンス鎖でのマッピングを示し、下段は、Ohta et alによるマッピングを示す。

【図9】図9Aは、mESCの全RNA（100 ng: 2つの複製物、10 ng: 2つの複製物、1 ng: 4つの複製物、100 pg: 8つの複製物、10 pg: 16つの複製物）の希釈後のERCC RNAの量とSC3-seq（ $\log_2(\text{RPM}+1)$ ）によるERCC spike-in RNAの算出レベルとの関係を示す。10 pgあたりの10コピー以上を有するERCC spike-in RNAに対するSC3-seqのデータは、回帰直線に利用された。図9Bは、ESCの全RNAの希釈物からのSC3-seqによる測定された全てのサンプルと増幅された全てのサンプルとの相関係数（ R^2 ）のヒートマップを示す（左図は、全ての発現遺伝子でのデータを示し、右図は、10 pgあたり20コピー以上の発現遺伝子でのデータを示す）。図9Cは、図9Bで示されたグループ間のペアワイズでの相関係数の最大値(max)および最小値(min)を示す（左図は、全ての発現遺伝子でのデータを示し、右図は、10 pgあたり20コピー以上の発現遺伝子でのデータを示す）。

【図10】図10Aは、SC3-seq（左図）およびSmart-seq2（右図）により増幅した希釈サンプルと測定して希釈サンプル缶の相関係数のヒートマップを示す。SC3-seqおよびSmart-seq2で、それぞれ0.1 RPMおよび0.1 FPKM未満の発現量が0.1として設定された。図10Bは、図10Aで示されたグループ間のペアワイズでの相関係数の最大値(max)および最小値(min)を示す（左図は、SC3-seq、右図は、Smart-seq2）。図10Cは、3つのRNA-seqによる全転写産物（上段左図）、1 Kbp未満の転写産物（下段左図）、750 bp未満の転写産物

10

20

30

40

50

(上段右図)および500 bp未満の転写産物(下段右図)の検出限界分析の結果を示す(マウスおよびヒトでそれぞれ、遺伝子発現レベルが、上位6555番目以上および6217番目以上(マウスおよびヒトに対してアノテーションされた全ての転写産物の1/4)で、 $\sim \log_2$ RPKM 3.69 ± 0.05 (SC3-seq), $\sim \log_2$ FPKM 2.21 ± 1.28 (Yan et al.)および $\sim \log_2$ FPKM 2.92 ± 0.27 (Picelli et al.)、全長リードによる遺伝子発現レベルと比較して 2 倍である)。

【図 1 1】図 1 1 Aは、E4.5日目の着床前胚におけるマーカー遺伝子(NANOG (エピブラスト)、POU5F1 (エピブラストおよびPE)、GATA4 (PE)およびCDX2 (TE))の発現を免疫蛍光染色分析した結果を示す。スケールバーは、100 μ m。図 1 1 Bは、E4.5日目の着床前胚の単一細胞の増幅cDNA(品質確認した67のcDNA)におけるマーカー遺伝子(NANOG (エピブラスト)、GATA4 (PE)、CDX2 (TE)およびGapdh (housekeeping))の発現をQ-PCR分析した結果を示す。図 1 1 Cは、SC3-seq [\log_2 (RPM+1)]により測定した増幅サンプルのERCC spike-in RNAを元のコピー数と比較した散布図を示す。回帰曲線および相関係数は、コピー数が10以上であるプローブの平均から算出した。図 1 1 Dは、各細胞における遺伝子発現レベルの分布をボックスプロットした結果を示す。図 1 1 Eは、全ての胚細胞間での相関係数(R^2)のヒートマップを示す。図 1 1 Fは、PEと比較してエピブラストで高発現している遺伝子(左図)またはエピブラストに比較してPEで高発現している遺伝子(右図)の発現レベルをボックスプロットした結果を示す。図 1 1 Gは、PEと比較してエピブラストで高発現している遺伝子(上図)またはエピブラストに比較してPEで高発現している遺伝子(下図)のGO解析の結果を示す。

【図 1 2】図 1 2 Aはフィーダー細胞上またはフリーで培養したhiPSCs (585A1 and 585B1)から増幅したcDNA(品質確認した112のcDNA)多能性遺伝子(POU5F1, NANOG, SOX2)およびGAPDH (housekeeping)の発現をQ-PCR分析した結果を示す。図 1 2 Bは、SC3-seq (\log_2 (RPM+1))によって測定した増幅サンプルのERCC spike-in RNAを元のコピー数と比較した散布図を示す。回帰曲線および相関係数は、コピー数が10以上であるプローブの平均から算出した。図 1 2 Cは、各細胞における遺伝子発現レベルの分布をボックスプロットした結果を示す。図 1 2 Dは、全ての胚細胞間での相関係数(R^2)のヒートマップを示す。

【図 1 3】図 1 3 Aは、カニクイザル着床前後胚より、SC3-seqを用いて得られた発現遺伝子解析の結果である。全ての発現遺伝子(全てのサンプルにおいて \log_2 (RPM+1) 4、18,353遺伝子)でのUHCおよび、多能性細胞マーカー、原始内胚葉マーカー、原腸陥入に伴う分化マーカーの発現レベルをヒートマップによって示す。図 1 3 B、Cは、全ての発現遺伝子によるPCAの結果を示す。図 1 3 Dは、胚体内細胞発生において大きく発現変動する遺伝子の発現レベルをヒートマップで示した。右は各クラスターに含まれる代表的遺伝子と、Gene Ontology解析の結果を示す。

【図 1 4】図 1 4 Aは、illumina社の次世代シーケンサー(Miseq, Nextseq500, Hiseq2000/2500/3000/4000)に対応したライブラリ作成の概略図を示す。上述のSOLiD5500xl用ライブラリとの違いは、タグにillumina社指定のDNA配列を使用した部分(赤破線)である。図 1 4 Bは、mESCから抽出したRNAの1ngにおいて、平均SC3-seqトラック(リード密度(RPM, $\times 1,000$ リード))を、アノテーションされたTTS(転写終結サイト)からのリードの位置に対してプロットしたグラフを示す。図 1 4 Cは、mESCの全RNAの1 ngおよび10 pgから2つの独立して増幅された複製産物をillumina社Miseqで解析し、複製産物間での比較を示した散布図である。

【図 1 5】図 1 5 Aは、上述のV1V3をP1P2という新規DNA配列に変更し、illumina社の次世代シーケンサー(Miseq, Nextseq500, Hiseq2000/2500/3000/4000)を用いて解析するためのSC3-seq法の概略を示す。図 1 5 Bは、mESCの全RNAの1 ngおよび10 pgからP1P2タグを利用し、illumina社Miseqを用いて解析した、2つの独立して増幅された複製産物間の比較を示した散布図である。図 1 5 C、Dは、1ng および10pgRNAより、それぞれV1V3タグとP1P2タグを用いて増幅されたcDNAを、illumina社Miseqを用いて解析し、全遺伝子の発現レベルの分布をボックスプロットで示した結果(C)と検出された遺伝子の数を棒グラフ

10

20

30

40

50

で示す(D)。

【発明を実施するための形態】

【0011】

(1) 生物学的試料における遺伝子発現量の相対的な関係を保持している増幅産物からなる核酸集団を調製する方法

本発明は、生物学的試料における遺伝子発現量の相対的な関係を保持している増幅産物を含む核酸集団を調製する方法、およびその方法により得られた核酸集団を提供する。

本発明において、「生物学的試料における遺伝子発現量の相対的な関係を保持している増幅産物」とは、生物学的試料における遺伝子産物群全体の構成（各遺伝子産物間の量比）がほぼ保持されている増幅産物（群）であって、次世代シーケンサーによりmRNAを定量する標準プロトコルに適用できる水準が確保されている増幅産物（群）を意味する。

10

【0012】

本発明において「生物学的試料」とは、mRNAの3'末端にポリAを持つ生物種、例えばヒトやマウス、カニクイザルなどの哺乳類を含む動物、植物、菌類、原生生物などの真核生物の細胞を意味する。本発明は特に、生物学的試料として発生過程における胚に含まれる細胞、多様性を有する多能性幹細胞への応用が期待される。

【0013】

生物学的試料としての細胞の数は特に制限されないが、本発明が生物学的試料における遺伝子発現量の相対的な関係を保持したまま再現性良く増幅可能である点を考慮すると、細胞の数は100個以下、数十個、1個から数個、究極には1個の細胞レベルに応用可能である。

20

【0014】

「次世代シーケンサーによるmRNAの定量」とは、RNAシーケンシング（RNA-Seqとも言う。）を意味し、mRNAの配列決定と合わせて、当該配列を有するmRNAのカウンティング、すなわち定量を行うことである。このような目的に用いられる次世代シーケンサーとして、illumina社、Life Technologies社、Roche Diagnostic社から市販されているものを用いることができる。

【0015】

本発明方法は以下の工程を含む。

(a) 任意の付加核酸配列X、ポリT配列、生物学的試料から単離したmRNA配列（実際にはmRNA配列に対応するcDNA配列である（以下同じ））、ポリA配列および任意の付加核酸配列Yの順で構成される2重鎖DNAを鋳型として、5'末端にアミンを付加した任意の付加核酸配列Xを含む第1のプライマーと、任意の付加核酸配列Yを含む第2のプライマーとを用いて、該2重鎖DNAを増幅する工程、

30

(b) 前記工程(a)により得られた2重鎖DNAを断片化する工程、

(c) 前記工程(b)により得られた断片化2重鎖DNAの5'末端をリン酸化する工程、

(d) 前記工程(c)により得られた5'末端がリン酸化された2重鎖DNAを鋳型として、任意の付加核酸配列Zおよび前記付加核酸配列Yをこの順序で含む第3のプライマーを用いてcDNAを調製、該cDNAの3'末端へアデニン(A)を付加する工程、

40

(e) 前記工程(d)により得られた2重鎖DNAへ、3'末端にチミン(T)をオーバーハングで有する任意の配列Vを含む2重鎖DNAを連結させる工程、および

(f) 前記工程(e)により得られた2重鎖DNAを鋳型として、任意の配列Vを含む第4のプライマーと、前記付加核酸配列Zを含む第5のプライマーとを用いて、該2重鎖DNAを増幅する工程

【0016】

本発明の方法の概要を図1C、図14Aおよび図15Aに示す。これらの図はあくまでも本発明の方法の一例についての説明であり、当業者は適宜変更を加えて本発明を実施することができる。以下において、本発明方法の各工程について図1Cおよび図15Aを参照しつつ詳細に説明する。

50

【0017】

(a) 任意の付加核酸配列 X、ポリ T 配列、生物学的試料から単離した mRNA 配列、ポリ A 配列および任意の付加核酸配列 Y の順で構成される 2 重鎖 DNA を鋳型として、5' 末端にアミンを付加した任意の付加核酸配列 X を含む第 1 のプライマーおよび任意の付加核酸配列 Y を含む第 2 のプライマーを用いて増幅する工程

【0018】

本工程において用いられる任意の付加核酸配列 X、ポリ T 配列、生物学的試料から単離した mRNA 配列、ポリ A 配列および任意の付加核酸配列 Y の順で構成される 2 重鎖 DNA は、次の工程 (i) から (iii) を含む方法で準備することができる (WO 2006/085616 を参照のこと)。

10

【0019】

(i) 一次鎖 cDNA の調製

生物学的試料から単離した mRNA を鋳型として、前記付加核酸配列 Y およびポリ T 配列からなる第 6 のプライマーを用いて逆転写することにより一次鎖 cDNA を調製する。

【0020】

以後の PCR 反応における増幅効率が鋳型 cDNA の長さに依存しないように、好ましくは逆転写反応の時間を 5 - 10 分、より好ましくは約 5 分間にまで短縮すればよい。これにより、全長の長い mRNA について、長さのそろった一次鎖 cDNA が合成される。「一次鎖 cDNA の長さがほぼ均一」とは、このような全長の長い mRNA について長さのそろった一次鎖 cDNA が得られることを意味し、より短い cDNA の存在を排除するものではない。

20

【0021】

当該調製の後、残存する第 6 のプライマーを分解その他の方法により除去することが望ましい。典型的にはエキソヌクレアーゼ I またはエキソヌクレアーゼ T により残存プライマーを分解すればよい。あるいは、残存プライマーの 3' 側をアルカリホスファターゼ等により修飾することにより、失活させることもできる。

【0022】

(ii) ポリ A テーリング反応、および第 7 のプライマーを用いた二次鎖 (2 重鎖) cDNA の調製工程

(i) により得られた一次鎖 cDNA をポリ A テーリング反応に付し、次いでこれを鋳型として、前記付加核酸配列 X およびポリ T 配列からなる第 7 のプライマーを用いて二次鎖である 2 重鎖 DNA を調製する。

30

【0023】

ここに使用する第 1 のプライマーと工程 (i) にて用いる第 2 のプライマーとは、相互に核酸配列が異なるが、一定の同一性を有し、かつプロモーター配列を含まないことを特徴とする。

【0024】

以下、第 6 および第 7 のプライマーについて、より詳細に説明する。

工程 (i) にて用いる第 6 のプライマー：付加核酸配列 Y およびポリ T 配列、および工程 (ii) にて用いる第 7 のプライマー：付加核酸配列 X およびポリ T 配列、における付加核酸配列 X および Y は、互いに配列が異なる。cDNA の 3' 側と 5' 側に対して異なるプライマーを用いることによって、以後の PCR 増幅において 3' 側と 5' 側を区別できる方向性をつけることができる。

40

【0025】

付加核酸配列 X および Y はさらに、X および Y における共通配列の Tm 値が、第 6 および第 7 のプライマーにおける Tm 値それぞれよりも低くなるように共通配列を選択し、かつ可能な限りその値が離れるようにする。そうすることで、以後の PCR 反応におけるアニーリングの際に第 6 および第 7 のプライマーが互いに別の部位にアニールしてしまう望ましくないクロスアニーリングを防止できる。換言すれば、共通配列の Tm 値は、第 6 および第 7 のプライマーをアニーリングする以後のアニーリング温度を超えないように選択

50

る mRNA を鋳型として用いた場合、20 サイクルが例示される。

【0030】

工程 (iii) において、PCR 増幅におけるアニーリング温度を、用いるプライマーの T_m 値に近づけることにより、非特異的なアニーリングを抑制することができる。例えば、配列番号 1 に示す核酸配列を有する核酸分子および配列番号 2 に示す核酸配列を有する核酸分子のプライマー対を用いる場合、アニーリング温度は、60 以上 90 未満、好ましくは約 70、最も好ましくは 67 である。

【0031】

工程 (iii) では、同一の出発サンプルに由来する一次鎖 cDNA を複数の、例えば 3 - 10 個、好ましくは約 4 個のチューブに小分けし、それぞれ PCR 反応を行い、最後に再び混合するのが好ましい。そうすることにより、ランダム誤差が平均化され、それを著しく抑制することができる。

【0032】

工程 (iii) で増幅された 2 重鎖 DNA は、例えば、Life Technologies 社から市販されている ERCC spike-in RNA を鋳型として用いて共に増幅し、その量と想定されるコピー数を比較することで、上述した PCR による増幅が正常に行われていたかを確認することができる。

【0033】

本工程 (a) では、上述の通り得られた 2 重鎖 DNA を鋳型として、5' 末端にアミンを付加した前記付加核酸配列 X を含む第 1 のプライマーと、前記付加核酸配列 Y を含む第 2 のプライマーとを用いて、該 2 重鎖 DNA を増幅することで、当該 2 重鎖 DNA 中の付加核酸配列 X の 5' 末端にアミンを付加することができる。当該第 1 および第 2 のプライマーは、それぞれ付加核酸配列 X および Y の下流に、任意でポリ T 配列をさらに含んでもよい。従って、第 1 のプライマーとして、工程 (i) で用いられる第 7 のプライマーの 5' 末端にアミンを付加したものを、また、第 2 のプライマーとして、工程 (i) で用いられる第 6 のプライマーを、それぞれ用いることができる (図 1C)。あるいは、工程 (iii) において、第 8 および第 9 のプライマーとして、ポリ T 配列を含まないプライマーを用いる場合には、該第 8 のプライマーの 5' 末端にアミンを付加したものを第 1 のプライマーとして、該第 9 のプライマーを第 2 のプライマーとして、それぞれ用いることもできる (図 15A)。

当該アミンの付加により、続く工程 (c) において mRNA 配列の 5' 末端側のリン酸化を抑制することができ、mRNA 配列の 3' 末端側のみを有するライブラリーを構築することを可能とする。本工程 (a) での増幅は、2 重鎖 DNA 中の付加核酸配列 X の 5' 末端にアミンを付加することができれば、特に限定されないが、例えば、2 から 8 サイクルの PCR によって行われ、好適には、4 サイクルである。付加されるアミンは工程 (c) における 5' 末端のリン酸化を抑制し得る限り特に制限はないが、好ましくはアミノ基が付加される。

【0034】

(b) 前記工程 (a) により得られた 2 重鎖 DNA を断片化する工程

本工程 (b) は、工程 (a) で得られた 2 重鎖 DNA を断片化する工程である。DNA の断片化は、超音波を用いて分裂させる方法や DNA 断片化酵素を用いて行う方法が例示される、本発明では、超音波を用いて分裂させる方法が好適に用いられる。

【0035】

(c) 前記工程 (b) により得られた断片化 2 重鎖 DNA の 5' 末端をリン酸化する工程

本工程 (c) は、工程 (b) により得られた断片化 2 重鎖 DNA の 5' 末端をリン酸化する工程であり、当該リン酸化は、自体公知の核酸キナーゼを用いて行うことができる。ただし、上述のとおり、アミンの付加された 5' 末端は、リン酸化することができない。工程 (b) において、超音波を用いて断片化した場合、切断された末端が、平滑化されていない可能性があるため、DNA ポリメラーゼを用いて、末端の平滑化を行ったのち、本工程 (c) を行うことが望ましい。

本工程（c）の後、任意の塩基長の大きさの断片化2重鎖DNAを選択する工程を行っても良い。本塩基長は、シーケンスによりmRNAが認識できる長さであれば特に限定されないが、好ましくは、Life Technologies社SOLiD5500xlでは200塩基から250塩基長、illumina社Miseq/NextSeq500/Hiseq2000/2500/3000/4000では350塩基から500塩基長である。

【0036】

DNAを選択する工程は、DNA吸着法、ゲル濾過法、ゲル電気泳動法などの当該分野で公知のいずれの手段を用いて行ってもよい。Beckman Coulter社のAMPureXP beadsを用いて行うことで、簡便に行うことができる。

【0037】

（d）前記工程（c）により得られた5'末端をリン酸化された2重鎖DNAを鋳型として、任意の付加核酸配列Zおよび前記付加核酸配列Yをこの順序で含む第3のプライマーを用いてcDNAを調製、その3'末端へアデニン（A）を付加する工程

本工程（d）では、工程（c）により得られた5'末端がリン酸化された2重鎖DNAを鋳型として、前記付加核酸配列Y（および任意でさらにポリT配列）を含む第2のプライマーの5'側にさらに任意の付加核酸配列Zを含む、第3のプライマーを用いて、該5'末端がリン酸化された2重鎖DNAへ該付加核酸配列Zを付加する工程である。当該工程では、工程（c）により得られた5'末端がリン酸化された2重鎖DNAを鋳型として、当該第3のプライマーを添加して、DNAポリメラーゼを用いて伸長反応を行うことにより、該付加核酸配列Zを該2重鎖DNA中の付加核酸配列Yに付加することができる。DNAポリメラーゼとして3'末端にアデニン（A）を付加するTdT活性を有する酵素を用いることにより、該2重鎖DNAの各鎖3'末端にアデニン（A）が付加される。本工程（d）により、5'末端がリン酸化された生物学的試料から単離したmRNA配列、ポリA配列、任意の付加核酸配列Yおよび任意の付加核酸配列Zの順で構成される2重鎖DNAを得ることができる。任意の付加核酸配列Zは、使用する次世代シーケンサーに依存した配列であり、当該シーケンサーの製造業者が推奨する配列を用いて行うことができる。例えば、Life technologies社のSOLiD5500XLを用いる場合、任意の付加核酸配列Zとして、配列番号3（ctgctgtacggccaaggcgt）で表されるInt配列を用いることができる。また、illumina社のMiseq/NextSeq500/Hiseq2000/2500/3000/4000を用いる場合、該付加核酸配列Zとして、配列番号13（gtgactggaggttcagacgtgtgctcttccgatc）で表されるRd2SP配列を用いることができる。

【0038】

（e）前記工程（d）により得られた2重鎖DNAへ、3'末端にチミン（T）をオーバーハングで有する任意の配列Vを含む2重鎖DNAと連結させる工程

センス鎖（工程（d）により得られた2重鎖DNAのうちmRNA配列のセンス鎖を含む鎖に連結する鎖）の3'末端にチミン（T）をオーバーハングで有する任意の配列Vを含む2重鎖DNAとは、3'末端側に1塩基（T）のみ相補鎖がない状態の2重鎖DNAであり、当該オーバーハング部分が、連結される2重鎖DNAの5'末端において、同様に1塩基（A）オーバーハングである相補鎖を有する場合、突出末端でのライゲーションが可能となる。本発明において、配列Vは、使用する次世代シーケンサーに依存した配列であり、例えば、Life technologies社やillumina社において市販されている配列を用いて行うことができる。具体的には、例えば、図1CにおけるP1-T（配列番号11：ccac tacgcctccgcttttctctctatgt）や図15AにおけるRd1SP-T（配列番号12：tctttccctacacgacgtcttccgatct）（いずれもセンス鎖）を用いることができる。

【0039】

（f）前記工程（e）により得られた2重鎖DNAを鋳型として、前記配列Vを含む第4のプライマーならびに前記付加核酸配列Zを含む第5のプライマーを用いて増幅する工程

本工程（f）において、用いる第4のプライマーおよび第5のプライマーは、使用する次世代シーケンサーに依存した配列である。第5のプライマーは、付加核酸配列Zを少なくとも含有していればよく、任意でその下流に前記付加核酸配列Yさらに含んでもよい

10

20

30

40

50

。より好ましくは、第5のプライマーはバーコード配列をさらに含むことが望ましく、任意の配列を有するアダプター配列をさらに含んでいることが望ましい。例えば、当該第4のプライマーおよび第5のプライマーは、Life technologies社やillumina社において市販されている配列を用いて行うことができる。

本工程(f)において、増幅は、第4のプライマーおよび第5のプライマーで増幅されない2重鎖DNA(mRNA配列の3'末端側または内部配列を含む断片化2重鎖DNA)が相対的に減少すれば、特にサイクル数は限定されないが、例えば、5から30サイクルのPCRが例示され、好適には、9サイクルである。

【0040】

上記工程により得られる本発明の核酸集団は、次世代シーケンサーに適用し、配列決定および当該核酸数を測定するための試料として有用である。特に、3'側のみを特異的に抽出していることから、確実にmRNAの一部を含有したライブラリーを構築することができる。従って、増幅されたmRNAの定量に優れていることから、従来のRNA-Seqと比較して、より正確な網羅的mRNAの発現量を検出することができる。

【0041】

別の態様として本発明は、次世代シーケンサーによるmRNA量の測定に適用するためのcDNA集団を調製するためのキットを提供する。

本発明のキットは、上述した以下のプライマーによって構成されても良い；

(a) 5'末端にアミンを付加した任意の付加核酸配列X(および任意でさらにポリT配列)を含む第1のプライマー

(b) 任意の付加核酸配列Y(および任意でさらにポリT配列)を含む第2のプライマー

(c) 任意の付加核酸配列Zおよび前記付加核酸配列Y(および任意でさらにポリT配列)をこの順序で含む第3のプライマー

(d) 3'末端にチミン(T)をオーバーハングで有する任意の配列Vを含む2重鎖DNA

(e) 前記配列Vを含む第4のプライマー、および

(f) 前記付加核酸配列Z(および任意でさらに前記付加核酸配列Y)を含む第5のプライマー。

当該キットは、任意の付加核酸配列X、ポリT配列、生物学的試料から単離したmRNA配列、ポリA配列および任意の付加核酸配列Yの順で構成される2重鎖DNAを調製するための以下のプライマーセット；

(g) 前記付加核酸配列YおよびポリT配列からなる第6のプライマー、

(h) 前記付加核酸配列XおよびポリT配列からなる第7のプライマー、

(i) 前記付加核酸配列Xを含み、任意でその下流にポリT配列をさらに含んでもよい第8のプライマー、および

(j) 前記付加核酸配列Yを含み、任意でその下流にポリT配列をさらに含んでもよい第9のプライマー

をさらに含んでもよい。ここで、前記第6のプライマーと前記第9のプライマーとは同一であってよく、また、前記第7のプライマーと前記第8のプライマーとは同一であってよい。さらに、前記第2のプライマーと前記第9のプライマーとが同一であってもよい。

当該キットは、PCR反応に必要な他の試薬(例えば、DNAポリメラーゼ、dNTPミックス、緩衝液等)、ライゲーション反応、逆転写反応、末端リン酸化反応等に必要な他の試薬などをさらに含んでもよい。

【0042】

以下に本発明を実施例によりさらに詳細に説明するが、本発明はこれら実施例に制限されるものではない。

【実施例1】

【0043】

RNA抽出

マウスを用いた動物実験は、京都大学の倫理ガイダンスに従って行われた。マウス胚性

10

20

30

40

50

幹細胞 (mESC) 株であるBVSC R8は従来法に従って培養し (Hayashi, K. et al, Cell, 146, 519-532, 2011)、細胞株からの全RNAの抽出は、RNeasy mini kit (Qiagen (74104)) を用いて製造者マニュアルに従って行われた。抽出したRNAは、double-distilled water (DDW)により250 ng/μl, 25 ng/μl, 2.5 ng/μl, 250 pg/μlおよび25 pg/μlの濃度へと希釈し、single-cell mRNA 3-prime end sequencing (以下、SC3-seqという)の定量性評価に用いた。

【0044】

マウス胚盤胞を摘出するため、C57BL/6マウスを交配させ、交尾栓を確認した日の午後をE0.5日目とした。E4.5日目の着床前の胚盤胞をKSOM (Merck Millipore (MR-020P-5D)) を用いて子宮より摘出し、dissection microscope (Leica Microsystems (M80)) 下でガラス針により、内部細胞塊(ICM)および極性栄養外胚葉(pTE)を含む極性部位と壁栄養外胚葉(mTE)を含む壁部位へと両断した。各フラグメントを0.25% trypsin/PBS (Sigma-Aldrich (T4799)) 中で約10分間37℃でインキュベートし、ピペティングにより単一細胞へと分離し、0.1 mg/ml PVA/PBS (Sigma-Aldrich (P8136)) へ懸濁し、SC3-seq分析に用いた。

【0045】

ヒト試料の取扱いは、京都大学医学研究科の倫理委員会の承認を得て、各ドナーのインフォームドコンセントの下に行った。ヒトiPSC細胞 (hiPSC) である、2つのiPSC株585A1および585B1 (Okita, K. et al, Stem Cells, 31, 458-466, 2013) は、20% (vol/vol) Knockout Serum Replacement (KSR) (Life Technologies (10828-028))、1% (vol/vol) GlutaMax (Life Technologies (35050-061))、0.1 mM nonessential amino acids (Life Technologies (11140-050))、4 ng/ml recombinant human bFGF (Wako Pure Chemical Industries (064-04541)) および0.1 mM 2-mercaptoethanol (Sigma-Aldrich (M3148)) を添加したDMEM/F12 (Life Technologies (11330-32)) 中でSNLフィーダー細胞上で培養する従来の培養条件、またはフィーダーフリー条件 (Nakagawa, M, et al, Scientific reports, 4, 3594, 2014、またはMiyazaki, T. et al, Nat Commun, 3, 1236, 2012) で培養した。フィーダー細胞からhiPSCを単離するため、CTK solutions (0.25% Trypsin (Life Technologies (15090-046))、0.1 mg/mL Collagenase IV (Life Technologies (17104-019))、1 mM CaCl₂ (Nacalai Tesque (06729-55))) で処理し、フィーダー細胞を除去後、acutase (Innovative Cell Technologies) を用いて単一細胞へ分離した。フィーダーフリー条件での細胞培養物から単一細胞を調製するため、0.5 × TrypLE Select (TrypLE Select) (Life Technologies (12563011)) を1:1で0.5 mM EDTA/PBSで希釈して用いた。分離した単一細胞状態のhiPSCsは、10 μM of the ROCK inhibitor Y-27632 (Wako) を添加した1% KSR/PBSに移し、SC3-seq分析に用いた。

【0046】

カニクイザルを用いた動物実験は滋賀医科大学動物生命科学研究センターにおける倫理委員会の承認を得て行った。カニクイザルの過排卵や卵採取、人工授精、初期胚培養、仮親への移植、着床胚の摘出など、生殖工学的操作は既報の方法により行った。J Yamasaki and others, 'Vitrification and Transfer of Cynomolgus Monkey (Macaca Fascicularis) Embryos Fertilized by Intracytoplasmic Sperm Injection.', Theriogenology, 76.1 (2011), 33-38 <<http://dx.doi.org/10.1016/j.theriogenology.2011.01.010>>.

【0047】

cDNA合成およびSC3-seq分析のための増幅

単離した単一細胞のRNAからのV1V3-cDNA合成および増幅は、既存の方法 (Kurimoto, K. et al, Nucleic Acids Res, 34, e42, 2006またはKurimoto, K. et al, Nature protocols, 2, 739-752, 2007) に従って行った。ただし、Qiagen RNase inhibitor (0.4 U/sample) (Qiagen (129916))、Porcine Liver RNase inhibitor (0.4 U/sample) (Takara Bio (2311A)) およびExternal RNA Controls Consortium (ERCC) により開発されたspike-in RNA (ERCC spike-in RNA) (Life Technologies (4456740)) の使用、ならびに全RNA量ごとのPCRサイクル数 (100 ngの全RNAでは7 cycles、10 ngの全RNAでは11 cycles、1 ng

の全RNAでは14 cycles、100 pgの全RNAでは17 cycles、10 pgの全RNAでは20 cycles)は、従来法とは異なる。P1P2-cDNAの合成および増幅は、SuperScript4(Life Technologies (18090200))とKOD FX NEO (Toyobo (KFX-201))の使用がV1V3-cDNA合成および増幅と異なる。ERCC spike-in RNAの全62,316または12,463コピーを、10 pgの全RNAまたは単一細胞あたりの溶解液に添加した。SC3-seqライブラリーの構築前に、増幅したcDNAの品質をERCC spike-in RNAおよび内在性遺伝子の定量PCR (Q-PCR) におけるCt valuesを調べることで、およびLabChip GX (Perkin Elmer) またはBioanalyzer 2100 (Agilent Technologies) によってcDNAフラグメントの割合を調べることで評価した。なお、mESCの全RNA希釈分析のためには、ERCC spike-in RNAとして、ERCC-00074 (9030 copies)、ERCC-00004 (4515 copies)、ERCC-00113 (2257 copies)、ERCC-00136 (112.8 copies)、ERCC-00042 (282.2 copies)、ERCC-00095 (70.5 copies)、ERCC-00019 (17.6 copies) およびERCC-00154 (4.4 copies) を用い、マウスの着床前胚およびhiPSCsのためには、ERCC-00096 (1806 copies)、ERCC-00171 (451.5 copies)、ERCC-00111 (56.4 copies) を用い、内在性遺伝子については、表 1 に記載の遺伝子を用いた。

【 0 0 4 8 】

【表 1 - 1】

	Target	Primer set name	sense primer	anti-sense primer	
ERCC spike-in	ERCC-0007	ERCC-00074-qPC	AAGTGAGGCTCTCTCAT	CTGGGGTTATGAGTAG	10
	4	R-s1/as1	TGGTTT	GGATGAG	
	ERCC-0000	ERCC-00004-qPC	GACATCTTCATAAGGGG	GGCAAGCCGGTGATTT	
	4	R-s1/as1	TTGGGT	TATCTAC	
	ERCC-0011	ERCC-00113-qPC	TACCCTGTAGAACCCGA	TTAAGCTCTGCACCTGT	
	3	R-s1/as1	CTTTTG	TACACT	
	ERCC-0013	ERCC-00136-qPC	GATGTTTGGACTGATGG	GATTTTCAGCATGTTGAG	
	6	R-s1/as1	AAGCAC	CTTCGT	
	ERCC-0004	ERCC-00042-qPC	GTGGTCTGCATAAGGGT	GCTTTGTCTTTAAACGC	
	2	R-s1/as1	AGAGAG	TCACCT	
	ERCC-0009	ERCC-00095-qPC	ATCCGTCAATAAAACCT	GGCTTTGTGGGATGAG	20
	5	R-s1/as1	CTGGCT	GTAAAG	
	ERCC-0001	ERCC-00019-qPC	TCTGTCTGTTAGTGAGA	TTTTGACCTAGCCCCAT	
	9	R-s1/as1	GCCCTA	CTACAC	
	ERCC-0015	ERCC-00154-qPC	CACGCGCTATTGACACG	TGGATATCCTCGCTTGA	
	4	R-s1/as1	TTATTT	GTTGAC	
	ERCC-0009	ERCC-00096-qPC	GATCCCGGAAGATACGC	CGCAGGTTGATGCTTCC	
	6	R-s1/as1	TCTAAG	AATAAA	
	ERCC-0017	ERCC-00171-qPC	CAGGCAAGAGTTCAATC	TAGCCTTCAGTGACTGT	
	1	R-s1/as1	GCTTAG	GAGATG	
Mouse	ERCC-0011	ERCC-00111-qPC	CCAACCCACATTGTAA	GTCTTTACTTACGCGCT	30
	1	R-s1/as1	CTTCG	CCTCT	
	<i>Cdx2</i>	<i>Cdx2</i> (+)/(-)	ATTGTTTGCTGCTGTTC	CGACTTCCCTTCACCAT	
			GAGTC	ACAAC	
	<i>Dnmt3a</i>	<i>Dnmt3a</i> -2(+)/(-)	GACTCGCGTGCAATAAC	GGTCACTTTCCCTCACT	
			CTTAG	CTGG	
	<i>Dnmt3b</i>	<i>Dnmt3b</i> (+)/(-)	CTCGCAAGGTGTGGGCT	CTGGGCATCTGTCATCT	
			TTTGTAAC	TTGCACC	
	<i>Dppa2/ECA</i>	<i>musDppa2</i> -qPCR-	CATGTAGCAACTCCAGT	ACATCAGTACTGGCTAA	
	<i>T15-2</i>	s1/as1	GGACC	GTGGT	
	<i>Dppa3/Stella</i>	<i>Stella</i> -B(+)/(-)	AGGCTCGAAGGAAATG	TCCTAATTCTCCCGATT	40
	<i>a</i>		AGTTTG	TTCG	

【 0 0 4 9 】

続き

【 0 0 5 0 】

【表 1 - 2】

<i>Dppa4/ECA</i>	musDppa4-	GTTTTGCTGCAAGTAG	GA CTGCTGAACTGGTT	10
<i>T15-1</i>	qPCR-s1/as1	GTCCTG	ATGACG	
<i>Dppa5/Esg</i>	Qesg1-5/3	AAGGAGTGCTGAAGCT	CAGCTTAACCTGCATC	10
<i>1</i>		GGAGG	CAGGTC	
<i>Esrrb</i>	musEsrrb-qPCR-	GCCTTTACTATCTGTGC	TAGTGCTTCTCTTTGGT	10
	s1/as1	CTGGT	GCTGT	
<i>Gapdh</i>	Qgapdh5/3	ATGAATACGGCTACAG	CTCTTGCTCAGTGCCT	10
		CAACAGG	TGCTG	
<i>Gata2</i>	musGata2-qPCR-	AAAGTGTCCCAAGCTTC	GCACTTGTTGACTCA	10
	s1/as1	GAT	GCAC	
<i>Gata4</i>	Gata4b(+)/(-)	CCTAAACCTTACTGGCCGT	ACAATGTTAACGGGTTGT	10
		AGC	GGAG	
<i>Kit</i>	c-Kit(+)/(-)	CAGTTACCGCGCTCTGT	GCCCCCTAAGTACCTG	10
		TTG	ACATCC	
<i>Myc</i>	c-myc-1(+)/(-)	AAGGAGAACGGTTCCT	GCTGAAGCTTACAGTC	20
		TCTGAC	CCAAAG	
<i>Nanog</i>	Qnanog-5/3	CTTTCACCTATTAAGGTGC	TGGCATCGGTTTCATG	20
		TTGC	GTAC	
<i>Pou5f1/Oct</i>	Qoct4-5/3	GATGCTGTGAGCCAAG	GGCTCCTGATCAACAG	20
<i>4</i>		GCAAG	CATCAC	
<i>Ppia</i>	PPIA(+)/(-)	TTACCCATCAAACCATT	AACCCAAAGAACTTCA	20
		CCTTCTG	GTGAGAGC	
<i>Prdm14</i>	prdm14-B(+)/(-)	ACAGCCAAGCAATTTG	TTACCTGGCATTTCAT	30
		CACTAC	TGCTC	
<i>Rplp0/Arbp</i>	PO(+)/(-)	CAAAGCTGAAGCAAAG	AATTAAGCAGGCTGAC	30
		GAAGAG	TTGGTTG	
<i>Sox2</i>	Qsox2-5/3	CATGAGAGCAAGTACT	CCAACGATATCAACCT	30
		GGCAAG	GCATGG	
<i>POU5F1/O</i>	hPOU5F1_F/R_q	CTGTCTCCGTCACCACT	AAACCCTGGCACAAAC	40
<i>CT4</i>	RT	CTG	TCCA	
<i>GAPDH</i>	hsGAPDH-qPCR-	ACAAGAGGAAGAGAG	TCTACATGGCAACTGT	40
	s1/as1	AGACCCT	GAGGAG	
<i>NANOG</i>	hsNANOG-qPCR-	AGAGGTCTCGTATTTGC	AAACACTCGGTGAAAT	40
	s1	TGCAT	CAGGGT	
<i>SOX2</i>	hsSOX2-qPCR-	TGAATCAGTCTGCCGA	TCTCAAACCTGTGCATA	40
	s1/as1	GAATCC	ATGGAGT	

【 0 0 5 1】

定量PCR (Q-PCR)

Q-PCRは、Power SYBR Green PCR Master mix (Life Technologies (4367659)) を用いて、CFX384 real-time qPCR system (Bio-Rad) により製造者マニュアルに従って行った。プライマー配列は、表 1 に示す。

【 0 0 5 2】

SOLiD 5500XL systemのためのSC3-seq用のライブラリー構築

5 ngの増幅および品質確認済みのcDNAをpre-amplification buffer ((1 × ExTaq buffer) (Takara Bio (RR006))、0.2 μMの各dNTP (Takara Bio (RR006))、0.01 μg/μl N-V3 (dT)24 primer (HPLC-purified, 5' 側末端にアミン付加)、0.01 μg/μl V1(dT)24 primer (HPLC-purified)および0.025 U/μl ExTaqHS (Takara Bio (RR006))) に添加し、PCRを4サイクル回して増幅した。プライマーダイマーといった副産物を0.6 × volumeのAMPureXP beads (Beckman Coulter (A63881)) を用いてサイズ選択により取り除いた。精製したcDNAsは、double-distilled water (DDW)で希釈して130 μlとし、Covaris S2またはE210 (Covaris) を用いて断片化した。さらに、End-polish buffer (1 × NEBnext End Repair Reaction buffer (NEB (B6052S))、0.01 U/μlのT4 DNA polymerase (NEB (M0203)) および0.033 U/μlのT4 polynucleotide kinase (NEB (M0201))) 中で20、30分間処理することで末端ポリッシュを行い、0.8倍量のAMPureXPを加え、さらに20分間攪拌し、1.2倍量のAMPureXPに移し、cDNAを精製した。続いて、Int-adaptor配列と精製cDNAを提供するため、cDNAは30 μlのInternal adaptor extension buffer (1 × ExTaq Buffer、0.23 mMの各dNTP、0.67 μMのIntV1 (dT)24 primer (HPLC-purified)、0.033 U/μlのExTaqHS) 中で95 °Cを3 min、67 °Cを2 minおよび72 °Cを2 minのプログラムでサーマルサイクラーに供した。反応は、氷冷することで停止させ、20 μlのP1-adaptor ligation buffer (10 μlの5 × NEBNext Quick Ligation Reaction Buffer (NEB (B6058S))、0.6 μlの5 μM of the P1-T adaptor (Life Technologies (4464411)) および1 μlのT4 ligase (NEB (M0202M)) の混合物) を加えた後、20 °Cで15分間、72 °Cで20分間インキュベーションした。1.2倍量のAMPure XPを添加し、cDNAの精製を2度行った後、cDNAをFinal amplification buffer (1 × ExTaq buffer、0.2 mMの各dNTP、1 μMのP1 primer、1 μMのBarT0XX_IntV1 primer (HPLC-purified) (XXは2ケタの整数を示し、具体的なプライマーは表2に列挙した)、0.025 U/μlのExTaqHS) に加え、95 °Cで3 minのインキュベーション後、95 °Cで30 sec、67 °Cで1 minおよび72 °Cで1 minを9 cycles、その後72 °Cで3 minのプログラムでサーマルサイクラーに供し、PCRによって増幅した。最後に、cDNAライブラリーは、1.2倍量のAMPureXPを用いて精製し、20 μl TE bufferへ溶解した。構築したライブラリーの品質と容量は、a Qubit dsDNA HS assay kit (Life Technologies (Q32851)) およびSOLiD Library TaqMan Quantitation kit (Life Technologies (4449639)) を用いてLabChip GXまたはBioanalyzer 2100により評価した。emulsion PCRによるビーズ上でのライブラリーの増幅は、SOLiD™ EZ Bead™ System (Life Technologies (4449639)) を用いて、製造者マニュアルに従ってE120スケールで実施した。得られたビーズライブラリーは、flowchipに装填し、SOLiD 5500XL systemの50 bpおよび5 bp barcode plus Exact Call Chemistry (ECC)で測定した。

【0053】

10

20

30

【表 2 - 1】

Oligonucleotids for SC3-seq library construction

Oligonucleotids name	sequence
V1(dT)24	ATATGGATCCGGCGCGCCGTCGACTTTTTTTTTTTTTTTTTTTTTT
V3(dT)24	ATATCTCGAGGGCGCGCCGGATCCTTTTTTTTTTTTTTTTTTTTTT
N-V3(dT)24	(NH2)-ATATCTCGAGGGCGCGCCGGATCCTTTTTTTTTTTTTTTTTTTTTT
IntV1(dT)24	CTGCTGTACGGCCAAGGCGTATATGGATCCGGCGCGCCGTCGACTTTTTTTTTTTTTTTTTTTTTT
P1 primer	CCACTACGCTCCGCTTCCTCTCTATG
BarT001-IntV1	ctgccccgggttctcattctctGTGTAAGAGGctgctgtacggccaaggcgatatggatcc
BarT002-IntV1	ctgccccgggttctcattctctAGGGAGTGGTctgctgtacggccaaggcgatatggatcc
BarT003-IntV1	ctgccccgggttctcattctctATAGGTTATctgctgtacggccaaggcgatatggatcc
BarT004-IntV1	ctgccccgggttctcattctctGGATGCGGTCctgctgtacggccaaggcgatatggatcc
BarT005-IntV1	ctgccccgggttctcattctctGTGGTGTAAGctgctgtacggccaaggcgatatggatcc
BarT006-IntV1	ctgccccgggttctcattctctGCGAGGGACActgctgtacggccaaggcgatatggatcc
BarT007-IntV1	ctgccccgggttctcattctctGGGTTATGCCctgctgtacggccaaggcgatatggatcc
BarT008-IntV1	ctgccccgggttctcattctctGAGCGAGGATctgctgtacggccaaggcgatatggatcc
BarT009-IntV1	ctgccccgggttctcattctctAGGTTGCGACctgctgtacggccaaggcgatatggatcc
BarT010-IntV1	ctgccccgggttctcattctctGCGGTAAGCTctgctgtacggccaaggcgatatggatcc
BarT011-IntV1	ctgccccgggttctcattctctGTGCGACACGctgctgtacggccaaggcgatatggatcc
BarT012-IntV1	ctgccccgggttctcattctctAAGAGGAAActgctgtacggccaaggcgatatggatcc
BarT013-IntV1	ctgccccgggttctcattctctGCGGTAAGCctgctgtacggccaaggcgatatggatcc
BarT014-IntV1	ctgccccgggttctcattctctGTGCGGACAGctgctgtacggccaaggcgatatggatcc
BarT015-IntV1	ctgccccgggttctcattctctGAGTTGAATGctgctgtacggccaaggcgatatggatcc
BarT016-IntV1	ctgccccgggttctcattctctGGGAGACGTTctgctgtacggccaaggcgatatggatcc
BarT017-IntV1	ctgccccgggttctcattctctGGCTACCGCctgctgtacggccaaggcgatatggatcc
BarT018-IntV1	ctgccccgggttctcattctctAGGCGGATGActgctgtacggccaaggcgatatggatcc
BarT019-IntV1	ctgccccgggttctcattctctATGTTAACTGctgctgtacggccaaggcgatatggatcc
BarT020-IntV1	ctgccccgggttctcattctctGTCAAGCTTTctgctgtacggccaaggcgatatggatcc
BarT021-IntV1	ctgccccgggttctcattctctGTGCGGTTCCctgctgtacggccaaggcgatatggatcc
BarT022-IntV1	ctgccccgggttctcattctctGAGAAAGTGAActgctgtacggccaaggcgatatggatcc
BarT023-IntV1	ctgccccgggttctcattctctGCGGTGCTTGctgctgtacggccaaggcgatatggatcc
BarT024-IntV1	ctgccccgggttctcattctctGGGTGCGTATctgctgtacggccaaggcgatatggatcc
BarT025-IntV1	ctgccccgggttctcattctctAACATGATGActgctgtacggccaaggcgatatggatcc
BarT026-IntV1	ctgccccgggttctcattctctCGGGAGCCCGctgctgtacggccaaggcgatatggatcc
BarT027-IntV1	ctgccccgggttctcattctctCAGCAAACCTTctgctgtacggccaaggcgatatggatcc
BarT028-IntV1	ctgccccgggttctcattctctAGCTTACTACctgctgtacggccaaggcgatatggatcc
BarT029-IntV1	ctgccccgggttctcattctctGAATCTAGGGctgctgtacggccaaggcgatatggatcc
BarT030-IntV1	ctgccccgggttctcattctctGTAGCGAAGActgctgtacggccaaggcgatatggatcc
BarT031-IntV1	ctgccccgggttctcattctctGCTGGTGCGTctgctgtacggccaaggcgatatggatcc
BarT032-IntV1	ctgccccgggttctcattctctGTTGGGTGCTgctgtacggccaaggcgatatggatcc
BarT033-IntV1	ctgccccgggttctcattctctCGTTGGATACTgctgtacggccaaggcgatatggatcc
BarT034-IntV1	ctgccccgggttctcattctctTCGTTAAAGGctgctgtacggccaaggcgatatggatcc
BarT035-IntV1	ctgccccgggttctcattctctAAGCGTAGGActgctgtacggccaaggcgatatggatcc
BarT036-IntV1	ctgccccgggttctcattctctGTTCTACATctgctgtacggccaaggcgatatggatcc
BarT037-IntV1	ctgccccgggttctcattctctCTGTTATACCctgctgtacggccaaggcgatatggatcc
BarT038-IntV1	ctgccccgggttctcattctctGTCGTCTTAGctgctgtacggccaaggcgatatggatcc
BarT039-IntV1	ctgccccgggttctcattctctTATCGTAGTctgctgtacggccaaggcgatatggatcc
BarT040-IntV1	ctgccccgggttctcattctctAAAAGGGTTActgctgtacggccaaggcgatatggatcc
BarT041-IntV1	ctgccccgggttctcattctctGTGGGATTGctgctgtacggccaaggcgatatggatcc
BarT042-IntV1	ctgccccgggttctcattctctGAATGTACTActgctgtacggccaaggcgatatggatcc
BarT043-IntV1	ctgccccgggttctcattctctCGTAGGGTTctgctgtacggccaaggcgatatggatcc

10

20

30

【 0 0 5 4 】

【表 2 - 2】

BarT044-IntV1	ctgccccgggttctcattctctAAGGATGATCctgctgtacggccaaggcgatatggatcc	
BarT045-IntV1	ctgccccgggttctcattctctGTACTTGGCTctgctgtacggccaaggcgatatggatcc	
BarT046-IntV1	ctgccccgggttctcattctctGTCTCGTCAActgctgtacggccaaggcgatatggatcc	
BarT047-IntV1	ctgccccgggttctcattctctGAGGGATGGCctgctgtacggccaaggcgatatggatcc	
BarT048-IntV1	ctgccccgggttctcattctctGCCGTAAAGTctgctgtacggccaaggcgatatggatcc	
BarT049-IntV1	ctgccccgggttctcattctctATGTCATAAGctgctgtacggccaaggcgatatggatcc	
BarT050-IntV1	ctgccccgggttctcattctctGAAGGCTTGCctgctgtacggccaaggcgatatggatcc	
BarT051-IntV1	ctgccccgggttctcattctctAAGCAGGAGTctgctgtacggccaaggcgatatggatcc	
BarT052-IntV1	ctgccccgggttctcattctctGTAATTGTAActgctgtacggccaaggcgatatggatcc	
BarT053-IntV1	ctgccccgggttctcattctctGTATCAAGTctgctgtacggccaaggcgatatggatcc	
BarT054-IntV1	ctgccccgggttctcattctctAAAAGGCGGActgctgtacggccaaggcgatatggatcc	
BarT055-IntV1	ctgccccgggttctcattctctAGCTTAAGCctgctgtacggccaaggcgatatggatcc	
BarT056-IntV1	ctgccccgggttctcattctctGCATGTACCCctgctgtacggccaaggcgatatggatcc	10
BarT057-IntV1	ctgccccgggttctcattctctTAGTAAGAActgctgtacggccaaggcgatatggatcc	
BarT058-IntV1	ctgccccgggttctcattctctTAAAGTGGCGctgctgtacggccaaggcgatatggatcc	
BarT059-IntV1	ctgccccgggttctcattctctAAGTAATGTCctgctgtacggccaaggcgatatggatcc	
BarT060-IntV1	ctgccccgggttctcattctctGTGCCTCGGTctgctgtacggccaaggcgatatggatcc	
BarT061-IntV1	ctgccccgggttctcattctctAAGATTATCctgctgtacggccaaggcgatatggatcc	
BarT062-IntV1	ctgccccgggttctcattctctAGGTGAGGGTctgctgtacggccaaggcgatatggatcc	
BarT063-IntV1	ctgccccgggttctcattctctGCGGGTTCGActgctgtacggccaaggcgatatggatcc	
BarT064-IntV1	ctgccccgggttctcattctctGTCTACACctgctgtacggccaaggcgatatggatcc	
BarT065-IntV1	ctgccccgggttctcattctctGGGATCAAGCctgctgtacggccaaggcgatatggatcc	
BarT066-IntV1	ctgccccgggttctcattctctGATGTAATGctgctgtacggccaaggcgatatggatcc	
BarT067-IntV1	ctgccccgggttctcattctctGTCTTAGGctgctgtacggccaaggcgatatggatcc	
BarT068-IntV1	ctgccccgggttctcattctctGCATTGACGActgctgtacggccaaggcgatatggatcc	
BarT069-IntV1	ctgccccgggttctcattctctGATATGCTTctgctgtacggccaaggcgatatggatcc	
BarT070-IntV1	ctgccccgggttctcattctctGCCCTACAGActgctgtacggccaaggcgatatggatcc	
BarT071-IntV1	ctgccccgggttctcattctctACAGGGAACctgctgtacggccaaggcgatatggatcc	
BarT072-IntV1	ctgccccgggttctcattctctAAGTGAATACctgctgtacggccaaggcgatatggatcc	20
BarT073-IntV1	ctgccccgggttctcattctctGCAATGACGTctgctgtacggccaaggcgatatggatcc	
BarT074-IntV1	ctgccccgggttctcattctctAGGACGCTGActgctgtacggccaaggcgatatggatcc	
BarT075-IntV1	ctgccccgggttctcattctctGTATCTGGGctgctgtacggccaaggcgatatggatcc	
BarT076-IntV1	ctgccccgggttctcattctctAAGTTTAGGctgctgtacggccaaggcgatatggatcc	
BarT077-IntV1	ctgccccgggttctcattctctATCTGGTCTTctgctgtacggccaaggcgatatggatcc	
BarT078-IntV1	ctgccccgggttctcattctctGGCAATCATctgctgtacggccaaggcgatatggatcc	
BarT079-IntV1	ctgccccgggttctcattctctAGTAGAATTctgctgtacggccaaggcgatatggatcc	
BarT080-IntV1	ctgccccgggttctcattctctGTTACGGTctgctgtacggccaaggcgatatggatcc	
BarT081-IntV1	ctgccccgggttctcattctctGAACGTCACTctgctgtacggccaaggcgatatggatcc	
BarT082-IntV1	ctgccccgggttctcattctctGTGAAGGGAActgctgtacggccaaggcgatatggatcc	
BarT083-IntV1	ctgccccgggttctcattctctGGATGGCGTActgctgtacggccaaggcgatatggatcc	
BarT084-IntV1	ctgccccgggttctcattctctGCGGATGAACctgctgtacggccaaggcgatatggatcc	
BarT085-IntV1	ctgccccgggttctcattctctGAAAGCGTTctgctgtacggccaaggcgatatggatcc	
BarT086-IntV1	ctgccccgggttctcattctctAGTACCAGGActgctgtacggccaaggcgatatggatcc	
BarT087-IntV1	ctgccccgggttctcattctctATAGCAAAGCctgctgtacggccaaggcgatatggatcc	
BarT088-IntV1	ctgccccgggttctcattctctGTTGATCATGctgctgtacggccaaggcgatatggatcc	30
BarT089-IntV1	ctgccccgggttctcattctctAGGCTGTCTActgctgtacggccaaggcgatatggatcc	
BarT090-IntV1	ctgccccgggttctcattctctGTGACCTACTctgctgtacggccaaggcgatatggatcc	
BarT091-IntV1	ctgccccgggttctcattctctGCGTATTGGGctgctgtacggccaaggcgatatggatcc	
BarT092-IntV1	ctgccccgggttctcattctctAAGGGATTActgctgtacggccaaggcgatatggatcc	
BarT093-IntV1	ctgccccgggttctcattctctGTTACGATGctgctgtacggccaaggcgatatggatcc	
BarT094-IntV1	ctgccccgggttctcattctctATGGGTGTTTctgctgtacggccaaggcgatatggatcc	
BarT095-IntV1	ctgccccgggttctcattctctGAGTCCGGCActgctgtacggccaaggcgatatggatcc	
BarT096-IntV1	ctgccccgggttctcattctctAATCGAAGActgctgtacggccaaggcgatatggatcc	

【 0 0 5 5 】

Illumine Miseq systemのためのSC3-seq用のライブラリー構築

2 ngの増幅および品質確認済みのcDNAをpre-amplification buffer ((1 × KOD FX NEO buffer)、0.4 μMの各dNTP (Takara Bio (RR006))、0.3 μM N-P1 primer (HPLC-purified, 5' 側末端にアミン付加)、0.3 μM P2 primer (HPLC-purified)および0.02 U/μl KOD FX NEO) に添加し、PCRを4サイクル回して増幅した。その後PCR産物を0.6 × volumeのAMPureXP beads (Beckman Coulter (A63881)) を用いて精製した。精製したcDNAsは、double-distilled water (DDW)で希釈して130 μlとし、Covaris S2またはE210 (Covaris) を用いて断片化した。さらに、End-polish buffer (1 × NEBnext End Repair Reaction buffer (NEB (B6052S))、0.01 U/μlのT4 DNA polymerase (NEB (M0203)) および0.033 U/μlのT4 polynucleotide kinase (NEB (M0201))) 中で20、30分間処理することで末端ポリッシュを行い、0.7倍量のAMPureXPを加え、さらに20分間攪拌し、0.9倍量のAMPureXPに移し、cDNAを精製した。続いて、Rd2SP-adaptor配列と精製cDNAを提供するため、cDNAは30 μlのInternal adaptor extension buffer (1 × ExTaq Buffer、0.23

40

30

50

mMの各dNTP、0.67 μ MのRd2SP-P2 primer (HPLC-purified)、0.033 U/ μ lのExTaqHS) 中で95 $^{\circ}$ Cを3 min、60 $^{\circ}$ Cを2 minおよび72 $^{\circ}$ Cを2 minのプログラムでサーマルサイクラーに供した。反応は、氷冷することで停止させ、20 μ lのRd1SP-adaptor ligation buffer (10 μ lの5 \times NEBNext Quick Ligation Reaction Buffer (NEB (B6058S))、0.6 μ lの10 μ M of the Rd1SP adaptorおよび1 μ lのT4 ligase (NEB (M0202M)) の混合物) を加えた後、20 $^{\circ}$ Cで15分間、72 $^{\circ}$ Cで20分間インキュベーションした。0.8倍量のAMPure XPを添加し、cDNAの精製を2度行った後、cDNAをFinal amplification buffer (1 \times KOD FX NEB 0 buffer、0.4 mMの各dNTP、0.3125 μ MのS5XX primer (XXは2ケタの整数を示し、具体的なプライマーは表3に列挙した)、0.3125 μ MのN7XX primer (HPLC-purified) (XXは2ケタの整数を示し、具体的なプライマーは表3に列挙した)、0.02 U/ μ lのKOD FX NEO) に加え、95 $^{\circ}$ Cで3 minのインキュベーション後、95 $^{\circ}$ Cで10 sec、60 $^{\circ}$ Cで1 minおよび68 $^{\circ}$ Cで1 minを9 cycles、その後68 $^{\circ}$ Cで3 minのプログラムでサーマルサイクラーに供し、PCRによって増幅した。最後に、cDNAライブラリーは、0.9倍量のAMPureXPを用いて精製し、20 μ l TE bufferへ溶解した。構築したライブラリーの品質と容量は、a Qubit dsDNA HS assay kit (Life Technologies (Q32851))、KAPA Library Quantification Kits(KAPA (KK4828))、およびLabChip GXまたはBioanalyzer 2100により評価した。得られたLibrary DNAはMiseq Reagent kit v3, 150cycle (illumina (MS-102-3001))を用いて解析した。

【 0 0 5 6 】

【表 3】

Oligonucleotids for SC3-seq library construction

Oligonucleotids name	sequence
P1(dT)24	CCACTACGCTCCGCTTTCCTCTCTATGTTTTTTTTTTTTTTTTTTTTT
P2(dT)24	CTGCCCCGGGTTCTCATCTTTTTTTTTTTTTTTTTTTTTT
P1(28)	CCACTACGCTCCGCTTTCCTCTCTATG
P2(21)	CTGCCCCGGGTTCTCATCT
trd2SP-P2	GTGACTGGAGTTACAGAGTGTGCTCTCCGATCTGCCCCGGGTTCTCATCT
trd1SP-Ts	TCTTTCCCTACACGACGCTCTCCGATCT
trd1SP-Tas	GATCGGAAGAGCGTCGTGTAGGGAAGA
trd2SPV1(dT)20	GTGACTGGAGTTACAGAGTGTGCTCTCCGATCATATGGATCCGGCGCGCGTGCAGCTTTTTTTTTTTTTTTTTT
S502	AATGATACGGCGACCAACCGAGATCTACACTCTCTATACACTCTTTCCCTACACGACGCTCT
S503	AATGATACGGCGACCAACCGAGATCTACACTCTCTATACACTCTTTCCCTACACGACGCTCT
S505	AATGATACGGCGACCAACCGAGATCTACAGTAAGGAGACACTCTTTCCCTACACGACGCTCT
S506	AATGATACGGCGACCAACCGAGATCTACACTGCAACACTCTTTCCCTACACGACGCTCT
S507	AATGATACGGCGACCAACCGAGATCTACAGGAGTAACACTCTTTCCCTACACGACGCTCT
S508	AATGATACGGCGACCAACCGAGATCTACACTAAGCCTACACTCTTTCCCTACACGACGCTCT
S510	AATGATACGGCGACCAACCGAGATCTACACGCTTAATACACTCTTTCCCTACACGACGCTCT
S511	AATGATACGGCGACCAACCGAGATCTACACTCTCCGACACTCTTTCCCTACACGACGCTCT
S513	AATGATACGGCGACCAACCGAGATCTACACTGCACTAGACACTCTTTCCCTACACGACGCTCT
S515	AATGATACGGCGACCAACCGAGATCTACACTTAGTACACTCTTTCCCTACACGACGCTCT
S516	AATGATACGGCGACCAACCGAGATCTACACTTAGTACACTCTTTCCCTACACGACGCTCT
S517	AATGATACGGCGACCAACCGAGATCTACAGCGTAAGAACACTCTTTCCCTACACGACGCTCT
S518	AATGATACGGCGACCAACCGAGATCTACACTATTAAAGACTCTTTCCCTACACGACGCTCT
S520	AATGATACGGCGACCAACCGAGATCTACAGGCTATACACTCTTTCCCTACACGACGCTCT
S521	AATGATACGGCGACCAACCGAGATCTACAGGCTATACACTCTTTCCCTACACGACGCTCT
S522	AATGATACGGCGACCAACCGAGATCTACACTATGCGAACACTCTTTCCCTACACGACGCTCT
N701	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N702	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N703	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N704	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N705	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N706	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N707	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N710	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N711	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N712	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N714	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N715	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N716	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N718	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N719	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N720	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N721	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N722	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N723	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N724	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N726	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N727	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N728	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT
N729	CAAGCAGAAGACGGCATACGAGATTCGCTTATGTTGACTGGAGTTACAGCGTGT

【 0 0 5 7 】

参照ゲノムへのマッピング

アダプターまたはpoly-A配列は、cutadapt-1.3によってトリムし、全てのリードを検索した。ただし、30 bp以下のトリムされたリードは排除した。アダプターおよびpoly-A配列は、それぞれ総リードの約1-20%および5%で確認された。30 bp以上のトリムされなかったリードおよびトリムされたリードは、mouse genome mm10およびERCC spike-in RNAの “-no-coverage-search” オプションでtophat-1.4.1/bowtie1.0.1によりマッピングされた。ゲノムおよびERCC上マッピングされたリードは、分離され、ゲノム上のリードは、“-c

10

20

30

40

50

ompatible-hits-norm”、“-no-length-correction”および“-library-type fr-second strand”オプションおよびextended TTS（転写終結サイト）でアノテーションされたmm10参照遺伝子を用いてcufflinks-2.2.0により発現レベルに変換した。いくつかの遺伝子の発現レベルを評価した際に、cufflinksのオプション“-max-mle-iterations”をデフォルトである5,000でおこなったところ“FAILED”の結果であったため、cufflinksのオプション“-max-mle-iterations”は50,000にセットした。cufflinksを用いた参照遺伝子のアノテーションのため、3'方向へ参照より長い転写産物を有する遺伝子の発現レベルを正しく見積もるため、10kb下流以上まで参照遺伝子の転写終結サイト（TTS）を拡張した。細胞あたりの転写コピー数を評価するため、ERCC spike-in RNAのリードは、遺伝子発現解析に用いたゲノム上にマッピングされた総リードによって、million-mappedリードあたりのリード数（RPM）に標準化された。マッピングされたリードは、igv-2.3.34を用いて視覚化された。HTSeq-0.6.0によるマッピングされたリードの発現レベルへの変換は、上述したオプションを用いたcufflinks-2.2.0による結果と一致した。

10

【0058】

SC3-seqと従来法で得られたデータ分析

分析は、R software version 3.0.2およびExcel (Microsoft)を用いて行った。SC3-seqによる発現データは、図4ならびに図10Aおよび10B以外では、発現量として \log_2 (RPM+1)を用いて分析された。図4では、0.01未満のRPM/FPKM valuesは、相関係数の計算のため、0.01として行った。図10Aおよび10Bでは、0.1未満のRPM/FPKM valuesは、相関係数の計算のため、0.1として行った。

20

【0059】

転写産物の3'末端での最適定義の判断は、参照遺伝子アノテーションgff3 fileを修正して行った。転写産物の3'末端の定義が、1 kb単位で10 kbまで、10 kb単位で100 kbまで拡張した場合、全ての遺伝子の発現量を算出した。発現レベルが10%, 20%, 30%, 40%, 50%, 80%, 100%, 200%および300%まで増加した遺伝子がカウントされた。

【0060】

10 pgの全RNAあたりのコピー数の評価は、コピー数の機能として全32増幅サンプルにおいてERCC spike-in RNAの全ての \log_2 (RPM+1)値に対する単線形回帰直線を記載することによって実施された（ERCC-00116およびERCC-00004、ならびに10 pgあたり100コピー数以下であるERCC spike-in RNAを除いた）。10 pgの全RNAあたり1000コピー、100コピー、10コピーおよび1コピーに対する \log_2 (RPM+1)は、それぞれ、11.78、8.10、4.43および0.76であった。

30

【0061】

異なる発現レベルのため100 ngの全RNA [\log_2 (RPM+1) 1]から増幅したサンプルにおける遺伝子数のパーセンテージとして、10 pgの全RNA [\log_2 (RPM+1) 1]から増幅したサンプルにおいて検出された遺伝子数によってカバレッジを定義する。正確さ（accuracy）は、10 pgの全RNA [\log_2 (RPM+1) 1]から増幅したサンプルにおける遺伝子数のパーセンテージとして、100 ngの全RNAs [\log_2 (RPM+1) 1]から増幅したサンプルにおいて検出された遺伝子数に基づいて定義付けられる。発現した遺伝子は、100 ngのRNAからSC3-seqで準備したサンプルにおける [\log_2 (RPM+1) 1]で検出された遺伝子として定義された。カバレッジと正確さに対するマルチサンプル解析（8サンプル）は、8つの増幅したサンプルの1以上から8以上がリードを提示する検出定義のもとでカバレッジと正確さを算出することによって行われた。

40

【0062】

遺伝子発現の検出限界の解析のため、cufflinksによって遺伝子マッピングされたリードの変換は、10,000にまで減少させたマッピングされたリード数で繰り返した。有意な発現レベル [\log_2 (RPM+1) > 4]で検出された遺伝子数は、異なるマッピングされたすべてのリード数で計数された。他の方法で得られた公開データとSC3-seqを用いて得られたデータの比較において、SC3-seqに対する発現量として \log_2 RPMを用い、マウスおよびヒト遺伝子においてそれぞれ上位6,555番目および6,217番目までを用いた。

50

【 0 0 6 3 】

E4.5の胚盤胞、ヒトiPS細胞およびカニクイザル初期胚におけるSC3-seqによる単一細胞解析のデータ解析

Gplotsおよびqvalue packagesを備えたR software version 3.1.1およびEXCELを用いて解析をおこなった。すべての発現データ解析は、 \log_2 (RPM+1)値を用いて行われた。サンプルにおいて \log_2 (RPM+1)値は4以下である遺伝子（およそ20コピー/細胞）は、解析から除外した。Unsupervised hierarchical clustering (UHC)は、Euclidian distancesおよびWard distance functionsを備えたhclust functionを用いて行った。principal component analysis (PCA)は、スケーリングなしのprcomp functionを用いて行われた。

【 0 0 6 4 】

多群間での発現が異なる遺伝子群（DEG）を認識するため、Oneway_analysis of variance (ANOVA)およびqvalue functionが、それぞれ、p値およびfalse discovery ratioの算出のため用いられた。DEGは、サンプル間で4倍以上の違いを提示したもの(FDR < 0.01)および群内での平均発現レベルが \log_2 (RPM+1) = 4であるものとして定義した。遺伝子オントロジー（GO）解析は、DAVID web toolを用いて実施した。

【 0 0 6 5 】

マウスE4.5胚の免疫蛍光解析

ホールマウント免疫蛍光解析のため、単離した胚は、4%パラホルムアルデヒドのPBS中で20分間室温で固定し、2%BSA/PBSで洗浄し、permeabilization solution (0.5% Triton X / 1.0% BSA / PBS)中で20分間、室温でインキュベートした。2%BSA/PBSで2度洗浄後、胚は一次抗体を含有する2% BSA/PBS中で4 ー夜インキュベートし、2%BSA/PBSで3度洗浄後、二次抗体および4,6-diamidino-2-phenylindole (DAPI)を含有する2%BSA/PBS中で1時間、室温でインキュベートし、2%BSA/PBSで3度洗浄した。VECTASHIELD Mounting Medium (Vector Laboratories (H-1000))を用いてマウントした。一次抗体として、anti-mouse NANOG (rat monoclonal) (eBioscience (eBio14-5761))、anti-mouse POU5F1 (mouse monoclonal) (Santa Cruz (sc-5279))、anti-mouse GATA4 (goat polyclonal) (Santa Cruz (sc-1237))、anti-mouse CDX2 (rabbit monoclonal, clone EPR2764Y) (Abcam (ab76541))を用いた。二次抗体として、Alexa Fluor 488 anti-rat IgG (Life Technologies (A21208))、Alexa Fluor 555 anti-rabbit (Life Technologies (A31572))、Alexa Fluor 568 anti-mouse IgG (Life Technologies (A10037))およびAlexa Fluor 647 anti-goat IgG (Life Technologies (A21447)) (全てdonkey polyclonal)を用いた。蛍光像は、共焦点顕微鏡 (Olympus (FV1000))を用いて得られた。

【 0 0 6 6 】

Accession numbers

Accession numbersは、本分野で一般的に用いられている次のデータを用いた；SC3-seq data (GSE63266, GSE74767)、RNA-seq data for mESCsおよびmouse embryonic fibroblasts (MEFs)(GSE45916) (Ohta, S., et al., Cell reports, 5, 357-366, 2013), SMART-seq2 data for MEF (GSE49321) (Picelli, S., et al., Nat Methods, 10, 1096-1098, 2013)およびsingle-cell RNA-seq data for hESCs. (GSE36552) (Yan, L., et al., Nat Struct Mol Biol, 20, 1131-1139, 2013)。

【 実施例 2 】

【 0 0 6 7 】

SC3-seqの設計と構築

単一細胞cDNAの増幅は、高密度オリゴヌクレオチドマイクロアレイ様の3'末端側を濃縮する方法を用いた (Kurimoto, K., et al., Nucleic Acids Res, 34, e42, 2006およびKurimoto, K., et al., Nature protocols, 2, 739-752, 2007)。当該方法は、マウス胚盤胞など不均一な細胞種の解析、始原生殖細胞 (PGC) の発生におけるトランスクリプトームの解明、大脳皮質の発生における神経細胞種の解明など多様性のある細胞の発生解析に有用である。本方法は、単一細胞でのトランスクリプトーム解析だけでなく多くの細胞を対象とする場合においても有用であることが示されている。本方法は、cDNAの全長を含

む長いcDNAが合成され、RNA-seqにより解析されること狙って改良している。しかし、全長cDNAの合成効率の悪さおよび長いcDNAを増幅する際のバイアスの影響を考慮すると、cDNAの3'末端側の配列を増幅することが遺伝子発現レベルの適切な解析となると考えられた(図1A)。さらに、3'末端側の配列のみであれば、必要な配列長がより短くなり(図1B)、コスト面において有利である。

【0068】

単一細胞から合成されたcDNAの3'末端を増幅およびシークエンスする方法をSC3-seqと呼び、その方法を図1Cに示す。cDNA増幅のため、従来の方法で単一細胞レベルのRNAからcDNAを増幅した(図1C)。最初のcDNA鎖は、V1 (dT)24 primerを用いて合成し、過剰のV1 (dT)24 primerとアニールしたmRNAは、それぞれExonuclease IおよびRNaseHで消化され、poly (dA)テールは、最初のcDNA鎖の3'末端側に付加された。2番目のcDNA鎖は、V3 (dT)24 primerを用いて合成され、得られたcDNAは、V1 (dT)24 primerおよびV3(dT)24 primerを用いて、初期量に依存したPCRサイクル数(単一細胞または10 pgの全RNAあたり20 サイクル)で増幅した。シークエンサーとしてSOLiDを用いる場合、ライブラリーの構築のため、V1(dT)24 primerを有する3'末端側を増幅する方法を用いた(図1Cおよび図7A)。増幅したcDNAは、PCRによりNH2-V3(dT)24 primerによりタグを付し、プライマーダイマーをAMPure XPにより3度精製することで除去し、タグを付したcDNAは、超音波により断片化した。さらに、末端の平滑化(end-polish)およびAMPure XPによるサイズ選択した。得られた200-250 bpのcDNAは、変性させ、3'末端側を捕捉するため内部アダプター拡張配列を伴うIntV1(dT)24 primerでアニールした。P1アダプターを用いてライゲーションおよび配列拡張し、P1 primer、96長バーコードおよびIntV1(dT)24 primerを有するBarTOXX-IntV1 primerを用いて最終増幅した。最終増幅産物は、P1アダプター末端からシークエンスされ、遺伝子座におけるmRNAの3'末端側のマッピングを行った。SC3-seqは、mRNAの3'末端側のみの配列リードを提供することから、リード数は全長にかかわらず絶対的なmRNAの発現レベルの割合となり、単純でより正確な遺伝子発現量の定量性をもたらす。100万のマッピングされたリードあたりの配列リードにより標準化して正確に定量した。SC3-seqを評価するため、全RNA (100 ng (10,000細胞相当)を2複製、10 ng (1,000細胞相当)を2複製、1 ng (100細胞相当)を4複製、100 pg (10細胞相当)を8複製および10 pg (単一細胞相当)を16複製)をmESCから単離し、External RNA Controls Consortium (ERCC)により開発されたexternal spike-in RNA controlとともにRNAを採取し、増幅(100 ng, 10 ng, 1 ng, 100 pgおよび10 pgの全RNA に対して、それぞれ7, 11, 14, 17および20の初期PCRサイクル)して用い、SC3-seqによりこれらをシークエンスした。初期cDNA増幅する間に、ERCC spike-in RNAがspiked-in copy数へ増幅することを定量PCRで確認した(図7B)。mESCで発現している遺伝子数は、初期増幅したcDNA量と一致している。また、cDNAの増幅とSC3-seqのライブラリー構築工程の間に遺伝子の内容が保持されることが示された。

【0069】

図1Dは、SC3-seq(～40-50%のマッピング効率)による100 ngのRNAの増幅産物の一つの平均配列リード分布を示す。SC3-seqのデザインのとおりに、マッピングリードは、マッピングされた全てのRefSeq遺伝子の3'末端(TTSから150 bp上流)を多く含んでいた。エクソンのアンチセンス鎖でマッピングされたリードが少し見られた。これは、V1 (dT)24 primerが、増幅のためのcDNAの5'末端におけるミスアニールすることによる増幅産物の一部であることを意味している(Figure 1C)。一例として、Pou5f1遺伝子座周辺のSC3-seq trackは、エクソンのアンチセンス鎖における小さいピークと共に、Pou5f1のセンス鎖の3'末端に一致する単一のピークを提示した(Figure 1E)。遺伝子断片は、SC3-seqのリードのピークは、アノテートされたRefSeq transcriptの3'末端より下流に観察された。このことは、転写産物の3'末端の上流に一致する遺伝子座を提示している(図1Dおよび1E)。もう一例、Nanog遺伝子座において、SC3-seqピークは、Nanogの3'末端の1 Kb下流に観察された(図1E)。実際、公開されているmESCの網羅的なRNA-seqデータは、ピークがNanogの上流エクソンに連続的であることが確認されているが(Anders, S, et al., Bioinformatics, 31, 166-169, 2015)、SC3-seqによって検出された3'末端ピークは、Nanogの転写産

10

20

30

40

50

物の一部であることが示された。以上より、SC3-seqでは、RefSeqの転写産物の断片の3'末端側を再定義できることが示唆された。in silicoでのRefSeqの転写産物のTTSの拡張とSC3-seqによるマッピングされたリードにおいて増加を示した遺伝子数の関係を調べた。図1Fに示したように、マッピングされたリードの増加を提示する遺伝子数は、アノテーションされたTTSからTTSの定義の拡張により増加し、450遺伝子は、10 KbまでTTSの定義を拡張することによりマッピングされたリードが増加した。10 Kb以上までの拡張が、間違っ

てアノテーションすることを見出したため、直前の上流遺伝子の発現を示すアノテーションされたTTSから10 Kb下流以内で検出したピークを定義づけ、SC3-seqのデータを解析した。公開されたRNA-seqデータの解析によって、この定義付けによるmESCの網羅的RNA量の安定性を確認した。

10

【実施例3】

【0070】

SC3-seqの定量性評価

SC3-seqの定量性を評価するため、全RNAの1 ng (100細胞相当)および10 pg (1細胞相当)から増幅されたcDNAにおける遺伝子数としてQ-PCRにより測定されたthreshold-cycle (CT)値と同じcDNAから用意されたライブラリーにおける同じ遺伝子セットのSC3-seqリードに対する $[\log_2 (\text{RPM}+1)]$ 値の関係を分析した。図2Aに示されたように、CT値とSC3-seqのリードは、両cDNAの間に相関関係を示した(1ngおよび10 pg RNAからのcDNAに対してそれぞれ $R^2 = 0.9858$ および 0.9428 であった)。さらに、10 pgのRNAにおける30コピー以下であったspike RNAは、不十分な増幅を提示したが、希釈して得られた全ての増幅されたライブラリーにおけるERCC spike-in RNAのSC3-seqリードの $[\log_2 (\text{RPM}+1)]$ 値は、元のコピー数に相関した(図2B)(各希釈における10 pgの全RNAあたりのERCC spike-in RNAのコピー数を示した)。これにより、SC3-seqのリード数から10 pgのRNAあたりの遺伝子のコピー数を見積もることを可能とした。散布図分析により、100 ng(10,000細胞相当)、10 ng (1,000細胞相当)および1 ng (100細胞相当)のRNAから増幅したサンプルが、全ての遺伝子2倍差分直線内に全ての遺伝子が入り、非常に良い相関関係(それぞれ、 $R^2 = 0.994$, 0.992 および 0.988)を示すことが確認された(図2Cおよび2G)。100 pg (10細胞相当)のRNAから増幅したサンプルは、2倍および4倍差分直線内にそれぞれ89.6%および97.8%がプロットされ、良い相関関係を示した(図2Cおよび2G)。10 pg (1細胞相当)のRNAから増幅したサンプルは、2倍および4倍差分直線内にそれぞれ77.6%および86.1%がプロットされ、良い相関関係を示した(図2Cおよび2G)。増幅されたサンプルは、10 pgのRNAあたり20コピー以上発現していた遺伝子は良い相関関係を有しており、特に2倍および4倍差分直線内のそれぞれ85.4%および99.6%の遺伝子で $R^2 = 0.764$ であり、10 pgのRNAから増幅した遺伝子の相関性は高かった(図2Cおよび2G)。

20

30

【0071】

100 ng (10,000細胞相当)のRNAから増幅したサンプルの配列プロファイルと比較した場合、10 ng (1,000細胞相当)、1 ng (100細胞相当)および100 pg (10細胞相当)のRNAから増幅したサンプルは高い相関性を示し(それぞれ、 $R^2 = 0.991$, 0.989 および 0.963)、10 pg (1細胞相当)のRNAから増幅したサンプルは、2倍および4倍差分直線内のそれぞれ75%および87%の遺伝子を有し、高い相関性を示した($R^2 = 0.797$)(図2Dおよび2G)。10 pgのRNAあたり20コピー以上発現する遺伝子では、10 pg (1細胞相当)のRNAから増幅したサンプルは、2倍および4倍差分直線内のそれぞれ90.8%および99.9%の遺伝子を有し、高い相関性を示した($R^2 = 0.813$)(図2Dおよび2G)。

40

【0072】

単一細胞相当のRNAの場合のSC3-seqを評価するため、100 ngのRNAから増幅された2つのサンプルと10 pgのRNAから増幅された8つのサンプルにおける平均log発現量を比較した。散布図分析をしたところ、平均されたサンプルは、2倍および4倍差分直線内のそれぞれ79.8%および97%の遺伝子を有し、高い相関性を示した($R^2 = 0.939$)(図2Eから2G)。10 pgのRNAあたり20コピー以上発現する遺伝子では、2倍および4倍差分直線内のそれぞれ98.6%および99.9%以上の遺伝子を有し、高い相関性を示した($R^2 = 0.930$)(図2Eから

50

2G)。以上より、100 ng (10,000細胞相当)から10 pg (1細胞相当)の領域のRNAでは、SC3-seqの定量性はとても高いことが示された。これらの結果より、10 pg (1細胞相当)のmESCの全RNAに存在するmRNA分子数は、およそ300,000であると思われることができ、これは従来の値とよく一致している (図2H)。

【0073】

SC3-seqを評価するため、さらに、8増幅のデータを用いて10 pgのRNAのSC3-seqのカバレッジ (10 pgのRNA (\log_2 (RPM+1) - 1)で検出される遺伝子数/100 ngのRNA (\log_2 (RPM+1) - 1)で検出される遺伝子数) および正確さ (100 ngのRNAs (\log_2 (RPM+1) - 1)で検出され、さらに10 pgのRNA (\log_2 (RPM+1) - 1)で検出される遺伝子数) を評価した。

【0074】

発現遺伝子は、100 ngのRNAからSC3-seqにより調製されたサンプルにおける [\log_2 (RPM+1) - 1]で検出される遺伝子として定義付けた。発現レベルの機能として、単回増幅サンプルのカバレッジが、プロットされた (図3Aの黒線)。これまでの報告および上記のデータ (図2) から、カバレッジは発現レベルに依存しているが、10 pgのRNAあたり10コピー以上発現している発現遺伝子の大部分 (累積パーセンテージ、94.1%) は検出できている (図3A)。単回増幅サンプルの正確さも同様にプロットし、検出された遺伝子の99.7% (累積パーセンテージ) が10 pgのRNAあたり10コピー以上の発現レベル領域で発現していることが確認された (図3B)。多数サンプル解析 (8サンプル) を行ったところ、カバレッジは、8増幅サンプルの1以上から5以上が、リード (10 pgのRNAあたり10コピー以上) を提示している検出定義の下で改良された。このとき、正確さは、全ての検出定義においてほぼ100%であった (図3Aおよび3B)。以上より、単一細胞相当のRNAからSC3-seqで調製された単回サンプルは、優れたカバレッジと正確さを有していること、さらに、多数サンプル解析でもカバレッジおよび正確さ共に改善することが示された。

【0075】

SC3-seqの利点として、比較的少数の配列リードでも遺伝子発現の定量性が高いことが挙げられる。続いて、100 ng (10,000細胞相当)から10 pg (1細胞相当)のRNAで遺伝子発現量を測定することができるリード数を調べた。 \log_2 (RPM+1) - 4 (10 pgのRNAあたり5コピー以上) の条件での遺伝子数をプロットし、リード数を検出した。図3Cに示したように、SC3-seqで調整した100 ng (10,000細胞相当)から10 pg (1細胞相当)のRNAで \log_2 (RPM+1) - 4の条件で検出された遺伝子数は、上限がおよそ0.2メガのマッピングされたリードであり、検出された遺伝子数は、100 ng (10,000細胞相当)から100 pg (10細胞相当)のRNAでほぼ一定であり (~7,000)、10 pgのRNA (1細胞相当)でも同様であった (~6,000)。続いて、配列リードに対して100 ngのRNAから検出された遺伝子中、10 pgのRNAからSC3-seqで検出された遺伝子を発現量の領域ごとにパーセンテージをプロットした (図3D)。この分析から、 \log_2 (RPM+1) - 6 (10 pgのRNAあたり、20コピー以上) の条件での100%の遺伝子および \log_2 (RPM+1) - 4 (10 pgのRNAあたり5コピー以上) の条件でのおよそ60%の遺伝子は、およそ0.2メガのマッピングされたリードで上限として検出されることが確認された (図3D)。SC3-seqによる実質的な量 (\log_2 (RPM+1) - 4、10 pgのRNAあたり5コピー以上) で発現した遺伝子を認識するため、サンプルあたり0.2メガマッピングされたリードを実施することで十分であり、多数のサンプルを同時進行でシーケンシングできる少数で十分であることが見出された。

【実施例4】

【0076】

他の方法によるSC3-seqの機能比較

SC3-seqの定量的な機能を比較するため、単一細胞のRNA-seqの他の方法と比較した。まず、SC3-seqおよび対照として他の方法により調製されたサンプルにおける発現レベルおよび転写産物の長さの関係を調べた。さらに、mESCおよびMEFの網羅的なRNA定量として公開されているRNA-seqとも比較した。図4Aに示されたように、対照サンプルは転写産物の長さおよび細胞種にかかわらず発現レベル領域の多様な分布を示した (Ohta_mESCsおよびOhta_MEFs)。平均の発現レベルが全ての転写産物の長さにおいて同様であった (発現レ

10

20

30

40

50

ベルのモードは、 \log_2 FPKM (fragment per kilobase per million mapped reads) = 2である)。おどろくべきことに、100 ngおよび10 pgのRNAのSC3-seqにより検出されたmESCにおける転写産物の長さの機能として、発現レベル領域の分布は、対照RNA-seqによるものと同様であった(図4A)。例えば、発現レベルのモードは全ての転写産物の長さにおいて同様であるように(およそ \log_2 RPM = 4)、転写産物の長さにかかわらず遺伝子の発現レベルは一定であった。この結果は、SC3-seqは、単一細胞レベルを初期マテリアルとして用いても全ての転写産物の長さの遺伝子発現を代替していることが示された。

【0077】

一方、単一のヒトESCに対するYanらの結果(全長RNA-seq増幅法、Tang, F., et al, Nat Methods, 6, 377-382, 2009、Tang, F., et al, Nature protocols, 5, 516-535, 2010 およびOhta, S. et al., Cell reports, 5, 357-366, 2013)およびPicelliらの単一のMEFに対する報告(SMART-seq 2、Picelli, S., et al, Nat Methods, 10, 1096-1098, 2013)は、同一の傾向を示しているが、発現レベルのモードは、長い転写産物における合成および増幅の低い効率のため、転写産物の長さ依存することから(例えば、より長い遺伝子の発現レベルのモードは、低く、より短い遺伝子では高い)、SC3-seqおよび対照RNA-seqによる結果と異なっている(図4A)。

【0078】

続いて、SC3-seqおよび他の単一細胞RNA-seq法の転写産物におけるリードの位置の違いを調べた。図4Bおよび4Cに示されたとおり、SC3-seqは、Figure 1に示すとおり、全ての転写産物の長さにおいて、3'末端に限定した明瞭なピークを提示した。一方、他の方法では、転写産物の長さにより不均一さを示し、特に長い転写産物において3'末端で歪んだリードとなった(図4C)。

【0079】

これらの結果から、SC3-seqは、短い転写産物(< 500 bp)で、0.2メガのマッピングされたリードが必要であり(遺伝子発現レベルは、上位6,555番目以上の遺伝子(マウスでアノテーションされた転写産物の1/4)、 \log_2 RPM 3.69 ± 0.05)、他の方法では、短い転写産物(< 500 bp)で、1メガ以上のマッピングされたリードが必要である(500bp未満、遺伝子発現レベルは、上位6,555番目および6,217番目以上の遺伝子(それぞれ、マウスおよびヒトでアノテーションされた転写産物の1/4)、 \log_2 FPKM 2.21 ± 0.05 (Yan et al.) または2.92 ± 0.27 (Picelli et al.)) (図4D)。以上より、SC3-seqは従来の方法に比べて単一細胞のトランスクリプトーム解析においてより定量的で有効な方法であることが示唆された。

【実施例5】

【0080】

着床前のマウス胚における細胞種の差異の遺伝子解析

E4.5日目の着床前のマウス胚は、少なくとも3つの細胞種、エピブラスト、始原内胚葉(PE)および栄養外胚葉(TE)から成る。前2つの細胞は、内部細胞塊(ICM)を形成している。解剖学的には、TEは2つの細胞種、ICMに直接接触して胚外胚葉(ExE)を形成するpTE(極性TE)と胚盤胞の胚体外組織の一部と後の初期の栄養芽層巨大細胞を形成するmTE(壁TE)に分類される。これまで、E4.5日目のpTEおよびmTE間の遺伝子発現の違いは報告されていない。SC3-seqを用いて、胚盤胞における細胞種の区別ができるか否かについて調べた。

【0081】

E4.5日目の着床前の胚盤胞(図11A)を単離し、胚と胚体外に分け、それぞれを単一細胞へと分散させ、mTEまたはpTE/ICMを含むと考えられる単一細胞を単離し、cDNAを増幅した。GapdhおよびERCC spike-in RNAの発現レベルを解析することでcDNAの増幅効率を調べ(図11Bおよび11C)、主なマーカーであるNanog (epiblast), Gata4 (PE), Cdx2 (TE)およびGata2 (TE)の発現を解析することで、cDNAを大きく分類した(図11B)。67の単一細胞cDNAのうち、SC3-seq解析により37の代表的cDNAを解析した(解剖学的部位およびマーカーの発現により、12のmTE, 8のm/pTE, 9のエピブラスト, 9のPEとした。全てのTEは、

免疫蛍光染色によりCDX2を提示したが、このステージでは、Gata2陽性TEにおいてCdx2のmRNAが欠失していることを確認した(図11Aおよび11B)。Unsupervised hierarchical clustering (UHC)は、これらの細胞が大きく2つのクラスターに分類でき、これらは、さらに2つのサブクラスターへ分けられることが示された(図5A)。一つのクラスターは、主要なマーカー遺伝子の発現に基づいて(エピブラストではFgf4, Nanog, Sox2およびKlf2、ならびにPEではGata4, Pdgfra, Sox17, Sox7およびGata6)、エピブラストおよびPEの2つのサブクラスターから成っていることが見出された(図5A)。もう一方のクラスターの細胞では、12のmTE(胚体外組織から単離した細胞)のうち9が一つのサブクラスターに分類され、残りの3つが、マーカーと解剖学的部位により確認されたpTEを有する隣のサブクラスターに分類された。UHC解析によるprinciple component analysis (PCA)では、4つのグループに分類された(図5C)。これらの知見は、SC3-seqが胚発生における細胞種を区別できること、およびmTEおよびpTEが、E4.5日目で全体的に遺伝子発現が異なることを提示することに成功した。

10

【0082】

さらに、SC3-seqを評価するため、エピブラストとPE間で異なる発現(one-way analysis of variance (ANOVA))に基づいて、4倍以上、 \log_2 (RPM+1)が4およびfalse discovery ratios (FDR) が< 0.01である平均発現レベルにおいて発現の異なる遺伝子として定義した)をする遺伝子を見出した。これまでの知見と一致して、エピブラストで発現が高い遺伝子(313遺伝子)は、Tdgf1, Utf1, Sox2およびNanogを含み、“転写制御”、“遺伝子発現の負の制御”および“巨大分子代謝プロセスの負の制御”などの遺伝子オントロジー(GO)用語が多く見られた。一方、PEで発現が高い遺伝子(502遺伝子)は、Sparc, Lama1, Lamb1, Col4a1, Gata4, Gata6, PdgfraおよびSox17を含み、GO用語として、“脂質生合成プロセス”、“糖脂質代謝プロセス”および“出生または孵化における胚発生”が見られた(図5Dおよび図11G)。

20

【0083】

続いて、pTEおよびmTE間で発現の異なる遺伝子を探索した。Hspd1, Ddah1, Gsto1, and Dnmt3bを含むpTEで高く発現している218遺伝子を見出し、“細胞周期”、“M期”および“有糸分裂終期”と言った細胞周期に係るGO用語が多くみられた(図5Dおよび5F)。mTEと比較してpTEで多く発現している遺伝子は、pTEでの発現量と同じで、エピブラストおよびPEで発現しており(図5E)、これらの遺伝子は特に胚盤胞のmTEで下方制御されていることが確認された。一方、Slc2a3, Basp1, Klf5, Gata2およびDppa1を含むmTEで上方制御されている392遺伝子を見出し、“ビタミン輸送”、“PDGFRシグナル経路”、“脂質蓄積”、“胚胎盤発生”および“膜組織化”と言ったGO用語が多く見られた(図5Dおよび5F)。これらのデータは、E4.5日目で、mTEは有糸分裂が止まるまたは遅くなるといった、初期の栄養芽層巨大細胞への分化による複製末期における知見によく一致している。

30

【0084】

各細胞種あたり100コピー以上発現する遺伝子の平均コピー数を計算し、単一のmTEが、単一のエピブラストやpTEよりも転写産物を多く(2倍程度)有することが見出された(図5G)。さらに、このことは、mTEがこの胚ステージで典型的な単一細胞と比較して多くの転写産物を含む複製末期であるという知見をサポートしている。このように、ERCC spike-in RNAと共に用いた場合、SC3-seqは、単一細胞での転写産物の定量が可能となる。

40

【実施例6】

【0085】

ヒトiPS細胞の不均一性の検出

SC3-seqにより、均一な細胞集団における遺伝子発現の不均一性を検出できるかを調べた。フィーダー細胞上で培養したhiPSC(on-feeder hiPSC)およびフィーダーフリー条件で培養したhiPSC(feeder-free hiPSC)での遺伝子発現を測定した。このため、2つのhiPSC株(585A1および585B1)をSNLフィーダー細胞上およびフィーダーフリー条件で培養し(全部で112個分の単一細胞cDNAを得た)(図6A、図12Aおよび12B)、細胞(on-feeder 585A1, on-feeder 585B1, feeder-free 585A1およびfeeder-free 585B1を、それぞれ7個、

50

7個, 8個, and 7個の単一細胞分) に対してSC3-seqを用いて分析を行った(図12C)。UHC分析により、on-feeder hiPSCおよびfeeder-free hiPSCが、細胞株の違いにかかわらず、2つの異なるクラスターに分類できることが見出された。ただし、一つのon-feeder hiPSCは、feeder-freeクラスターに分類され、一つのfeeder-free hiPSCは、on-feeder クラスターに分類されたことを除いた(図6B)。このことから、on-feeder hiPSCおよびfeeder-free hiPSCは、とても似ているが区別できることを示唆している。続いて、PCA分析の結果、二つの異常値を除いて、feeder-free hiPSCが、互いに近くに分類され、on-feeder hiPSCは、PC2軸に沿って分散されることが見出された(図6C)。このことは、on-feeder hiPSCの遺伝子発現は、feeder-free hiPSCと比べてより不均一であることを示している。さらに、on-feeder hiPSCおよびfeeder-free hiPSCにおける遺伝子発現レベルに対する遺伝子発現レベルの標準偏差(SD)をプロットした(図6D)。on-feeder hiPSCにおいて、最大発現レベルである $\log_2(\text{RPM}+1)$ 6であり、SD 2である630遺伝子を見出したが、feeder-free hiPSCでは、109個であった。すなわち、PCA分析の結果と一致して、全ての遺伝子発現領域において、feeder-free hiPSCのSD値と比較して、on-feeder hiPSCのSD値が高いことが示された。feeder-free hiPSCの遺伝子発現レベルのSD値を、on-feeder hiPSCのものに対してプロットしたところ、PK1B, UNC5D, SCGB3A2, HERC1, RPS4Y1およびRBM14/RBM4を含むfeeder-free hiPSCでSD値が高い75の遺伝子、ZFP42, ANXA3, LEFTY1, PTCD1およびLDOC1を含むfeeder-free hiPSCおよびon-feeder hiPSC共にSD値が高い(SD 2)遺伝子、ならびにFGF19, CAV1, NODAL, SGK1, CTGFおよびSFRP1を含むon-feeder hiPSCでより高いSD値を示す596個の遺伝子を見出した(図6B、6Eおよび6F)。以上の知見は、feeder-free hiPSCが、on-feeder hiPSCよりも遺伝子発現において均一性が高いことを示している。SC3-seqは、均一細胞集団における遺伝子発現の不均一性を見出すための強力な方法であることが示唆された。

【実施例7】

【0086】

カニクイザル着床前後胚における遺伝子解析

全ての関連する系列及びそれらの前駆体を含む、着床前(pre)及び着床後(post)の胚からの390個の代表的な細胞(pre:193細胞; post:197細胞)のトランスクリプトームを、SC3-seq法により調べた。UHCは、全ての細胞を2つの大きなクラスターに分類し、1つは少なくとも6つの異なるクラスターを有する着床前の細胞を主として構成され、もう1つは少なくとも7つの異なるクラスターを有する着床後の細胞のみから構成されていた(図13a)。それぞれの細胞は、UHCによるクラスターと、多能性細胞マーカー、原始内胚葉マーカー、原腸陥入に伴う分化マーカー遺伝子の発現パターンより定義し、分類した(post_paTE; 着床後胚由来壁側栄養外胚葉、preL_TE; 着床前胚由来後期栄養外胚葉、HYP; 着床前胚由来胚盤膜下層、preE_TE; 着床前胚由来早期栄養外胚葉、ICM; 着床前胚由来内部細胞塊、pre_EPI; 着床前胚由来エピブラスト、postE_EPI; 着床後胚由来早期エピブラスト、postL_EPI; 着床後胚由来後期エピブラスト、Gast1, 2a, 2b; 着床後胚由来原腸陥入細胞、YE; 着床後胚由来卵黄内胚葉、ExMchy; 着床後胚由来胚体外間質細胞)。図13B、Cは、全ての発現遺伝子によるPCAの結果を示し、図13Dは、胚体内細胞発生において大きく発現変動する遺伝子の発現レベルをヒートマップで示した。以上より、同一グループの細胞は、類似の遺伝子発現パターンを示すことが認められたが、これは、SC3-seqが、高い定量性・再現性を有することを反映している。また、これらの知見は、SC3-seqが、カニクイザルのエピブラストの発生において、様々な細胞集団や、それらの発現パターンを区別することに成功した。

【実施例8】

【0087】

SC3-seq法の他の次世代シーケンサーへの適用性の検討

図14Aは、図1CのIntV1(dT)₂₄配列(配列番号4: ctgctgtacggccaaggcgatatggatc cggcgcgccgtcgacttttttttttttttttttttttt)をRd2SPV1(dT)₂₀配列(配列番号14: gtga ctggagttcagacgtgtgtcttccgatcatatggatccggcgcgccgtcgacttttttttttttttttttttt)に、

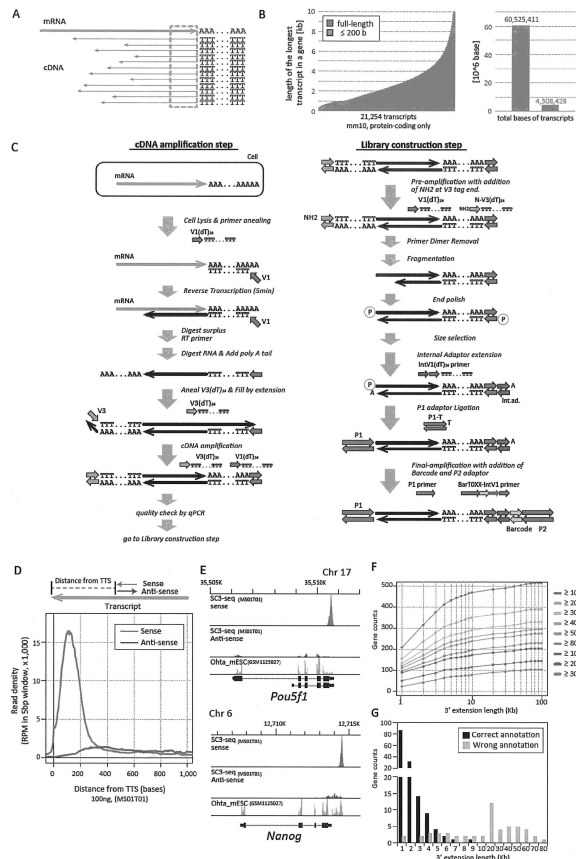
図 1 C のP1-T配列をRd1SP-T配列に変更することで、illumina社の次世代シーケンサー (Miseq, Nextseq500, Hiseq2000/2500/3000/4000) に対応したライブラリー作成の概略図を示す。mESCから抽出したRNAの1ngにおいて、平均SC3-seqトラック (リード密度 (RPM, $\times 1,000$ リード) を、アノテーションされたTTS (転写終結サイト) からのリードの位置に対してプロットした (図 1 4 B)。mESCの全RNAの1 ngおよび10 pgから、2つの独立して増幅された複製産物を、上記illumina社Miseqを用いて解析した (図 1 4 C)。全RNAの1 ngから増幅したサンプルは、非常に良い相関関係 ($R^2=0.972$) を示した。尚、Miseqで解析できれば他の次世代シーケンサー (Nextseq500, Hiseq2000/2500/3000/4000) でも解析可能であることがillumina社から公表されている。実際、本発明者らもHiseq2500でも解析可能であることを確認している (データは示さず)。

10

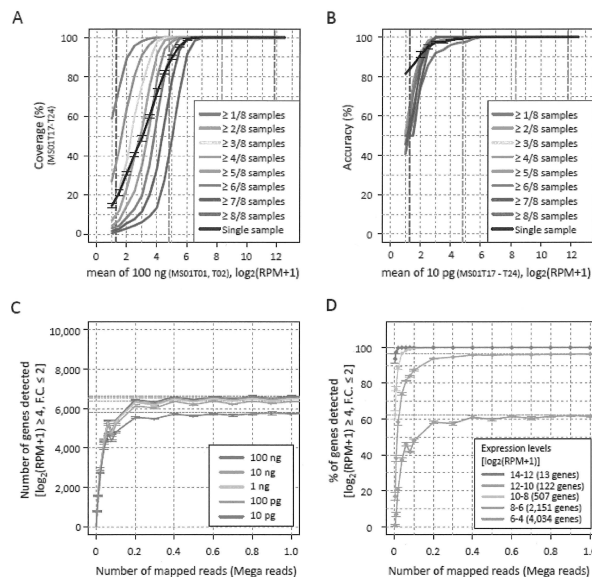
図 1 5 A は、上記illumina社Miseqを用いた解析において、cDNA増幅の際に用いたV1 (dT)₂₄ 配列をP2(dT)₂₄ 配列に、V3 (dT)₂₄ 配列をP1(dT)₂₄ 配列に変更し、さらにライブラリー作成の際に用いたV1(dT)₂₄ 配列をP2配列に、N-V3 (dT)₂₄ 配列 (配列番号 5 : (NH₂)-ata tctcgagggcgcgccggtatccttttttttttttttttttttttt) をN-P1配列 (配列番号 9 : (NH₂)-cca ctacgcctccgctttcctctctatg) に、Rd2SPV1(dT)₂₄ 配列をRd2SP-P2配列 (配列番号 15 : g tgactggagttcagacgtgtgctcttccggtcctgccccgggttcctcattct) に変更した、SC3-seqの概略図を示す。mESCの全RNAの1 ngおよび10 pgから、P1(dT)₂₄ 配列及びP2(dT)₂₄ 配列 (P1P2タグ) を利用し、2つの独立して増幅された複製産物をillumina社のMiseqを用いて解析した (図 1 5 B)。全RNAの1 ngから増幅したサンプルは、非常に良い相関関係 ($R^2=0.974$) を示した。全RNAの1ng及び10pgRNAより、それぞれV1 (dT)₂₄ 配列及びV3 (dT)₂₄ 配列 (V1V3タグ) を用いて増幅されたcDNAと、P1P2タグを用いて増幅されたcDNAを、illumina社Miseqを用いて解析し、全遺伝子の発現レベルの分布をボックスプロット (図 1 5 C) 及び検出された遺伝子の数を棒グラフ (図 1 5 D) で示した。V1V3タグとP1P2タグを用いて増幅されたサンプルは、同様のパターンを示したが、PIP2タグを用いた場合には、VIV3タグを用いた場合と比較してやや良好な結果を示した。以上より、SC3-seqが、Life Technologies社SOLiD5500xl以外にも、より汎用性の高いillumina社Miseqを用いた解析にも適用可能であることが示された。

20

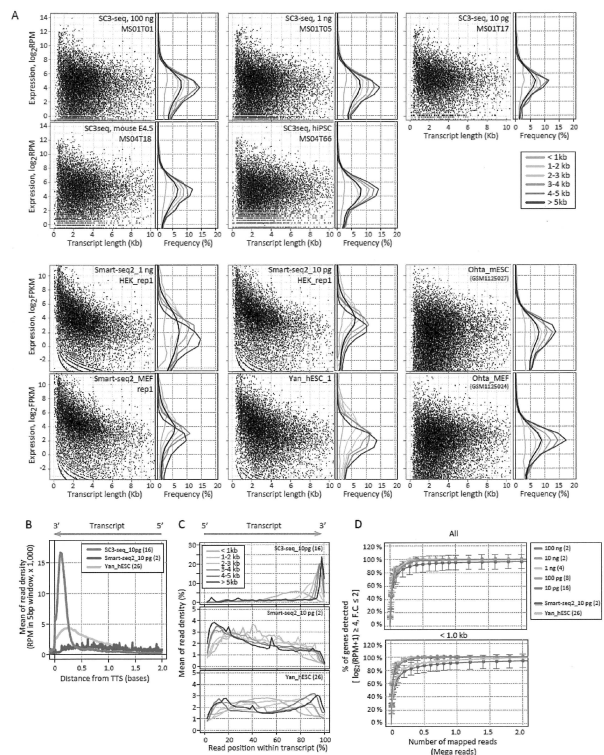
【 図 2 】



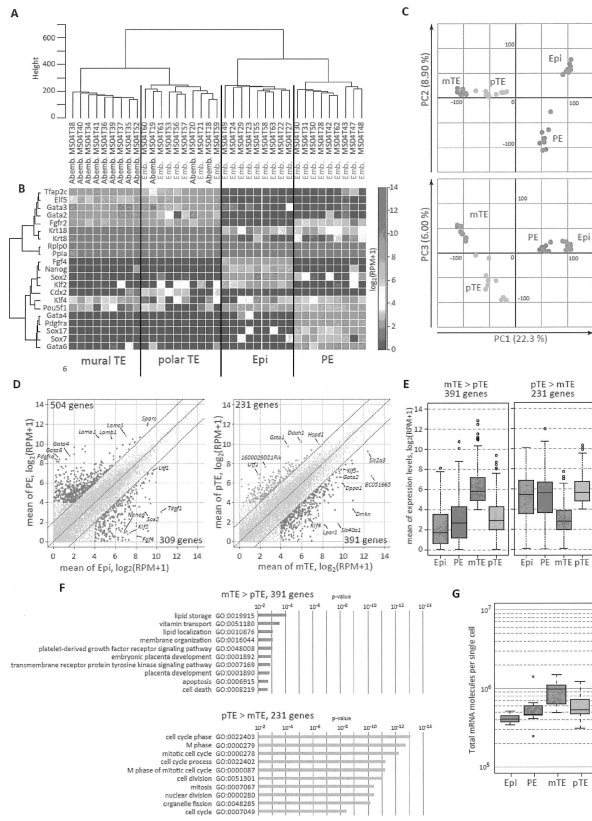
【圖 3】



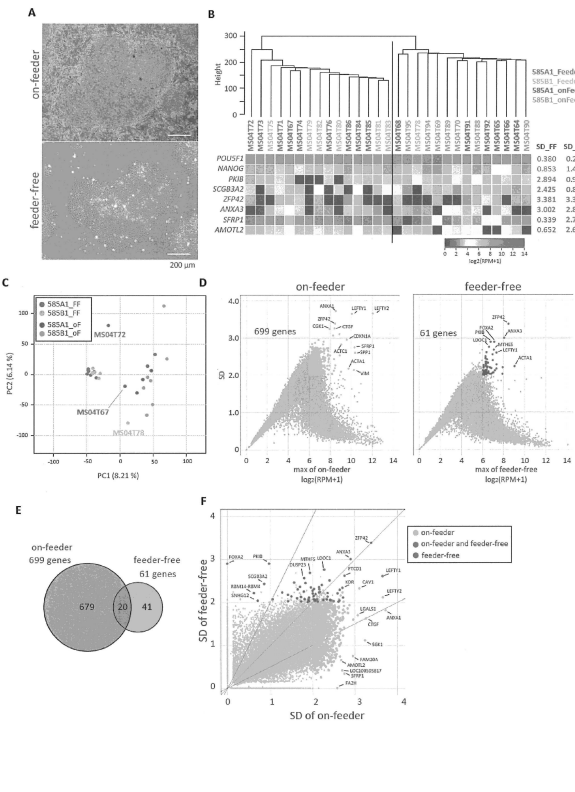
【 図 4 】



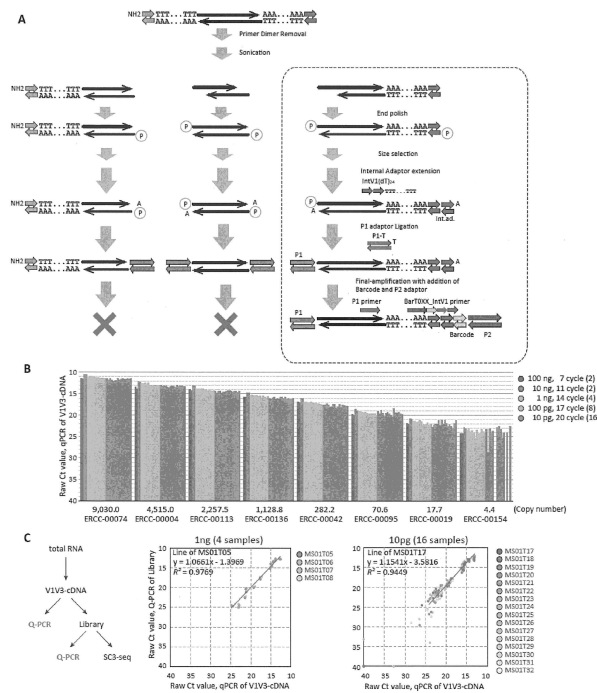
【図 5】



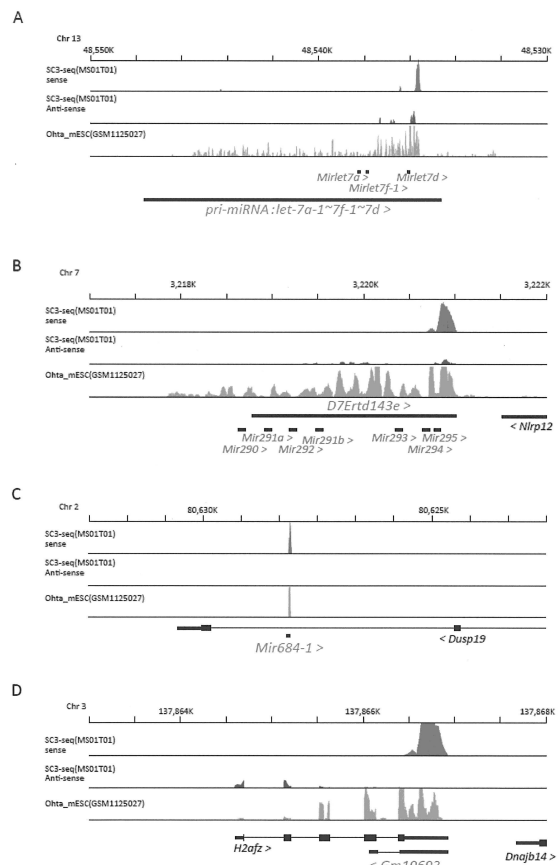
【図 6】



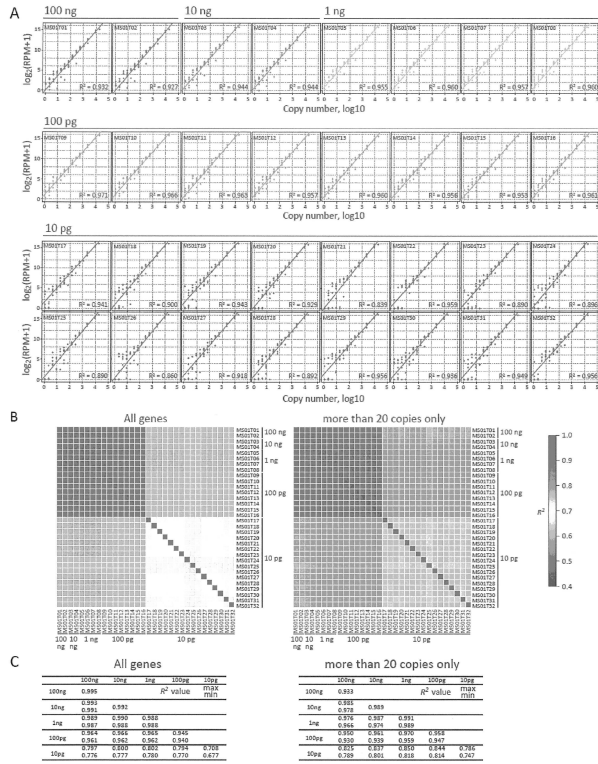
【図 7】



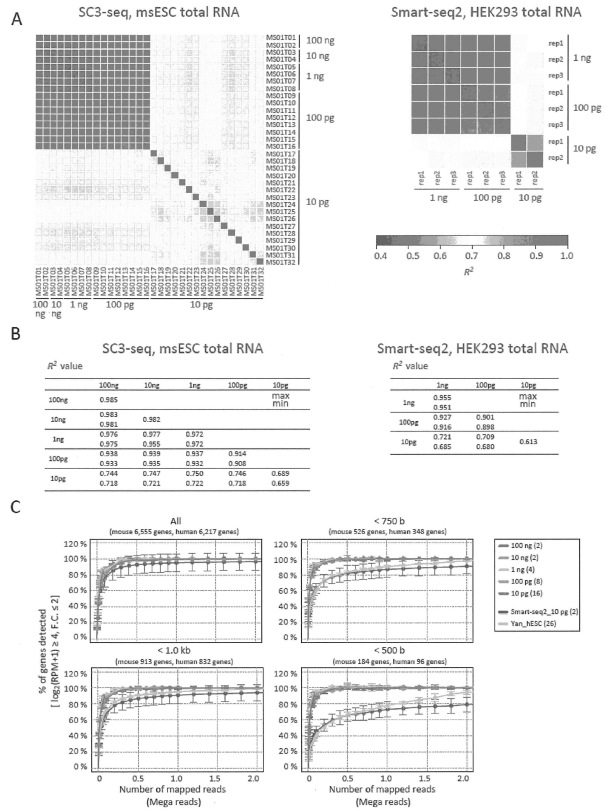
【図 8】



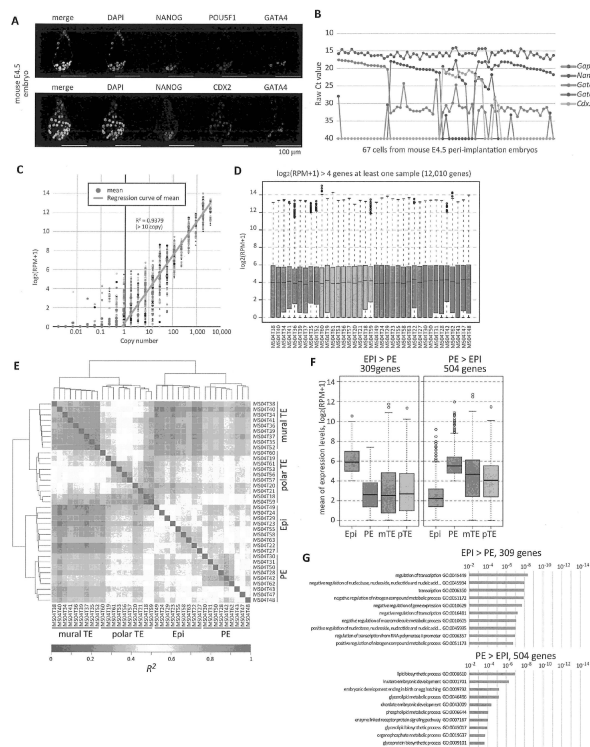
【図 9】



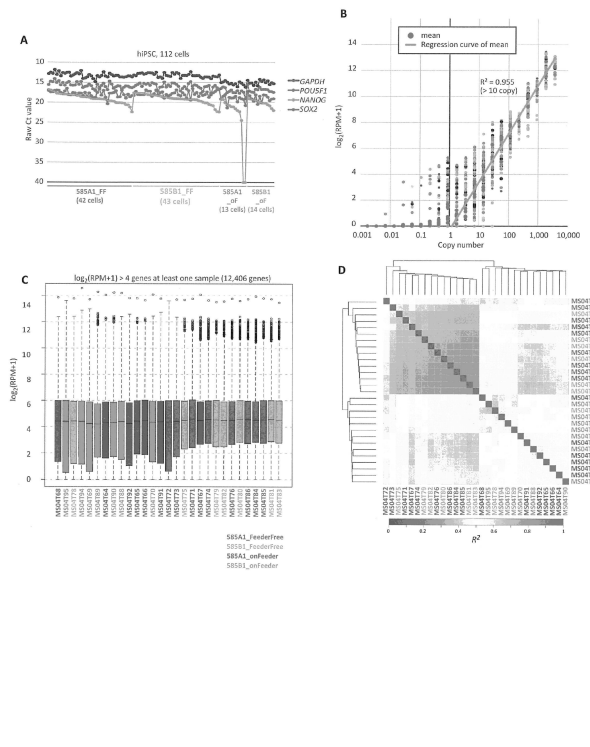
【図 10】



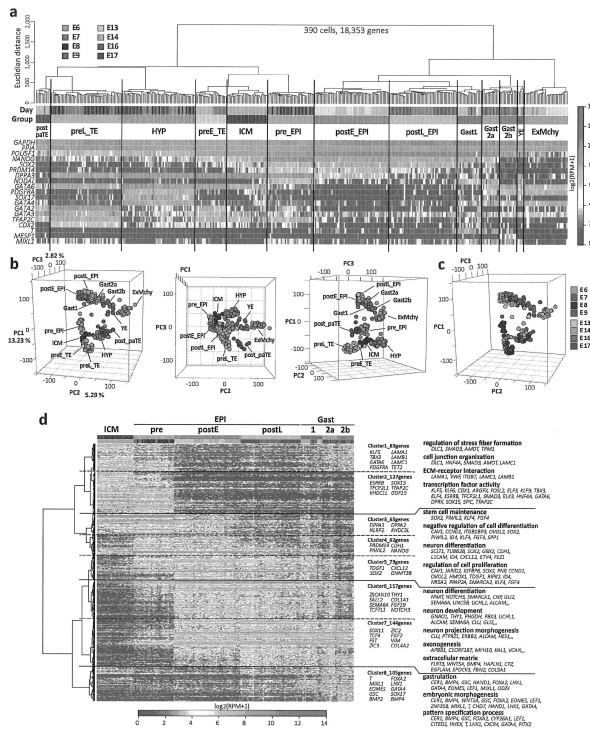
【図 11】



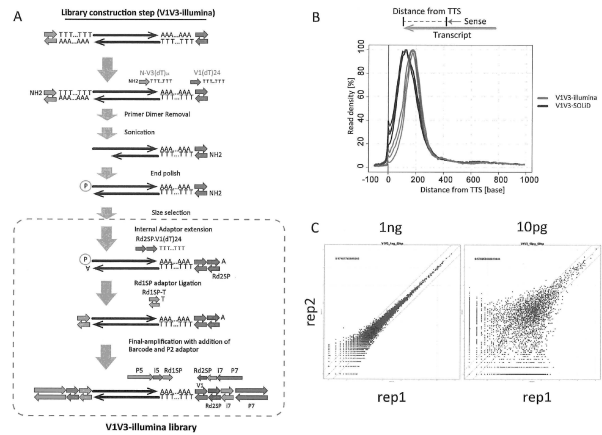
【図 12】



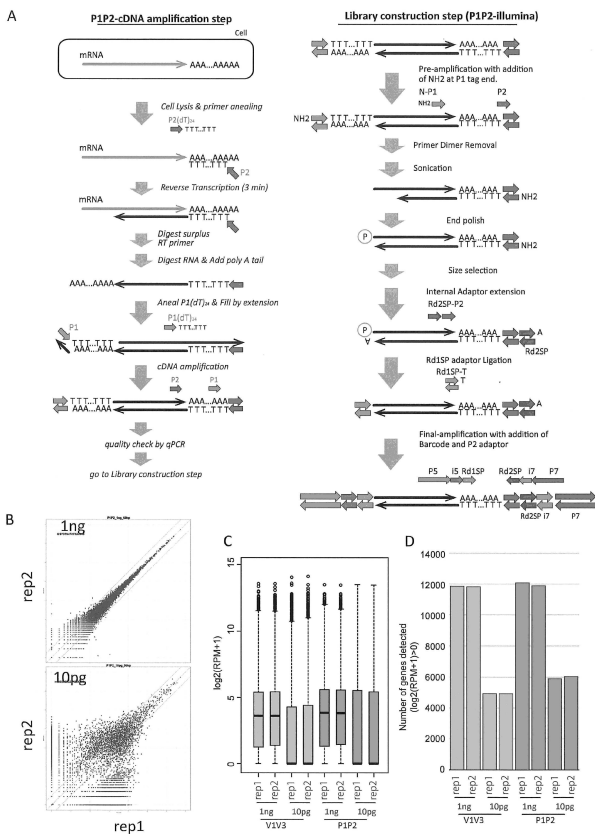
【図 13】



【図 14】



【図 15】



フロントページの続き

- (74)代理人 100137729
弁理士 赤井 厚子
- (74)代理人 100151301
弁理士 戸崎 富哉
- (74)代理人 100203253
弁理士 村岡 皓一郎
- (72)発明者 斎藤 通紀
京都府京都市左京区吉田本町3番地1 国立大学法人京都大学内
- (72)発明者 中村 友紀
京都府京都市左京区吉田本町3番地1 国立大学法人京都大学内
- (72)発明者 薮田 幸宏
京都府京都市左京区吉田本町3番地1 国立大学法人京都大学内

審査官 鈴木 崇之

- (56)参考文献 BECK AH et al., 3'-End Sequencing for Expression Quantification (3SEQ) from Archival Tumor Samples, PLoS One, 2010, 5(1), e8768, ISSN: 1932-6203
- LIANG J et al., Single-cell sequencing technologies: current and future, J. Genet. Genomics, 2014, 41(10), pp.513-528, ISSN: 1673-8527
- Genome Biol., 2013, 14(4):R31

(58)調査した分野(Int.Cl., DB名)

C12N 15/00 - 15/90
C12Q 1/68 - 1/6897
JSTPlus/JMEDPlus/JST7580(JDreamIII)
CPlus/MEDLINE/BIOSIS/WPIDS(STN)