

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7689737号
(P7689737)

(45)発行日 令和7年6月9日(2025.6.9)

(24)登録日 令和7年5月30日(2025.5.30)

(51)国際特許分類		F I		
C 1 2 Q	1/6809(2018.01)	C 1 2 Q	1/6809	Z
C 1 2 M	1/00 (2006.01)	C 1 2 M	1/00	A
C 1 2 Q	1/6869(2018.01)	C 1 2 Q	1/6869	Z

請求項の数 10 (全50頁)

(21)出願番号	特願2021-576209(P2021-576209)	(73)特許権者	518269418 Karydo Therapeutix 株式会社 東京都千代田区一番町6番地
(86)(22)出願日	令和3年2月6日(2021.2.6)	(74)代理人	110000796 弁理士法人三枝国際特許事務所
(86)国際出願番号	PCT/JP2021/004470	(72)発明者	佐藤 匠徳 京都府相楽郡精華町光台二丁目2番地2 株式会社国際電気通信基礎技術研究所内
(87)国際公開番号	WO2021/157739	審査官	川野 汐音
(87)国際公開日	令和3年8月12日(2021.8.12)		
審査請求日	令和6年1月17日(2024.1.17)		
(31)優先権主張番号	特願2020-18989(P2020-18989)		
(32)優先日	令和2年2月6日(2020.2.6)		
(33)優先権主張国・地域又は機関	日本国(JP)		

最終頁に続く

(54)【発明の名称】 シングルセルRNA-Seq解析のカウントデータセットの補正方法、シングルセルRNA-Seq解析の方法、細胞種の構成比率の解析方法、並びにこれらの方法を実行するた

(57)【特許請求の範囲】

【請求項1】

解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けすること、を含むシングルセルRNA-Seq解析のカウントデータセットの補正方法であって、

前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、

前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、

前記補正方法。

【請求項2】

前記シグネチャ遺伝子セットは、所定数の遺伝子数を含む、請求項1に記載の補正方法。

【請求項3】

解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けすることと、

重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対

象細胞を含む解析対象器官を構成する各細胞種におけるRNAの発現パターンを解析すること、

を含む、シングルセルRNA-Seqの解析方法であって、

前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、

前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、

前記解析方法。

10

【請求項4】

解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けすることと、

重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する細胞種の構成比率を解析すること、

を含む、

解析対象器官を構成する細胞種の構成比率の解析方法であって、

前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、

20

前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、

前記解析方法。

【請求項5】

シングルセルRNA-Seq解析のカウントデータセットの補正装置であって、

前記補正装置は、制御部を備え、

前記制御部は、

解析対象細胞から取得したシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けし、

30

前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、

前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、

前記補正装置。

40

【請求項6】

シングルセルRNA-Seqの解析装置であって、

前記解析装置は、制御部を備え、

前記制御部は、

解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けし、

重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する各細胞種におけるRNAの発現パターンを解析し、

前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝

50

子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、
前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、
 前記解析装置。

【請求項 7】

解析対象器官を構成する細胞種の構成比率の解析装置であって、
 前記解析装置は、制御部を備え、
 前記制御部は、
 解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けし、
 重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する細胞種の構成比率を解析し、
前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、
前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、
 前記解析装置。

10

【請求項 8】

コンピュータに実行させた時に、
 コンピュータに、
 解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けするステップを備える処理を実行させる、
 シングルセルRNA-Seq解析のカウントデータセットの補正プログラムであって、
前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、
前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、
 前記補正プログラム。

20

30

【請求項 9】

コンピュータに実行させた時に、
 コンピュータに、
 解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けするステップと、
 重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する各細胞種におけるRNAの発現パターンを解析するステップと、
 を備える処理を実行させる、シングルセルRNA-Seqの解析プログラムであって、
前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、
前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析

40

50

対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、
前記解析プログラム。

【請求項 10】

コンピュータに実行させた時に、

コンピュータに、

解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて、重み係数によって重み付けするステップと、

重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する細胞種の構成比率を解析するステップと、

を備える処理を実行させる、解析対象器官を構成する細胞種の構成比率の解析プログラムであって、

前記重み係数は、参照細胞種の各細胞種における各細胞種を特徴付けるシグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと、解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータから算出され、

前記シグネチャ遺伝子セットのシングルセルRNA-Seq解析のカウントデータと前記解析対象となる各器官全体に含まれる全RNAをRNA-Seq解析することによって得られたカウントデータは共に正規化されている、

前記解析プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本明細書には、シングルセルRNA-Seq解析のカウントデータセットの補正方法、シングルセルRNA-Seqの解析方法、細胞種の構成比率の解析方法、並びにこれらの方法を実行するための装置及びコンピュータプログラムが開示される。

【背景技術】

【0002】

ヒトの器官は約 1×10^8 - 3×10^{12} 細胞で構成されている。器官の細胞構成および/または細胞のフェノタイプの変化は、その機能不全、リモデリングおよび再生と密接に関連している。個々の器官は細胞の混合集団である。このため、シングルセルRNA-Seq (Single-cell RNA-SeqまたはscRNA-Seq) は、器官の細胞構成および/または細胞のフェノタイプの変化を捉えるために、それぞれの器官の細胞の集団に対して包括的な遺伝子発現プロファイルを解析し、その解析データを単一細胞の発現レベルに分解して、単一細胞の変化の情報を導き出している (非特許文献 1 から非特許文献 5)。したがって、scRNA-Seqは、詳細な正常および病的器官の分子細胞アトラスを生成する強力な方法といわれている。

【0003】

しかし、scRNA-Seqには、限界がある。まず、scRNA-Seqに供するためには、器官から採取した組織を、消化酵素を使って、又は物理的な破碎により個々細胞を回収しなければならない。この前提として、このような細胞を回収するためには組織が新鮮である必要がある。言い換えると、一般的に手術等で採取された組織は、数ヶ月から数年凍結保存されている場合が多く、このような保存組織は、scRNA-Seqに用いることができない。手術の後の病理組織診断において、希少な疾患であると判明しても、その希少な病理学的サンプルを新たに入手することは困難であり、RNAの発現解析に使用できるサンプルは、凍結保存されたものとなるのが一般的である。また、ヒトの組織の採取は、一般的に生検であることが多く、サンプル量が少ないといった問題がある。仮に剖検等を行って器官全体を採取できたとしても、心臓や脳のように大きな器官の場合、scRNA-Seqの目的でそれらの器官全体から個々の細胞を分離することは、不可能ではないが、現実的ではない。

【0004】

10

20

30

40

50

また、これは多くの場合、薬物効果および/または病因の研究では、同じ被験者の異なる複数の器官において薬物による影響、及び/又は病態を解析することが必要となるが、ヒトの場合、同一者から複数種の器官を採取した解析が難しいという問題もある。

【0005】

さらに、scRNA-Seqには、遺伝子発現における実験手法上のアーチファクトの問題もある。このような例として、細胞の分離工程で、細胞内で異常な遺伝子発現を誘発することが報告されている。

【0006】

上記課題を解決するために、コンピュータによる器官全体RNAデータベースデコンボリューション (computational deconvolution of whole-organ RNA datasets) が、提唱されている。器官全体RNAデータベースデコンボリューションは、細胞種ごとの細胞分離を行わずに、採取した被験組織からRNAの抽出を行い、RNA-Seqにより発現しているRNAの配列情報を得た後に、コンピュータによって計算された被験組織に含まれる細胞種の割合に基づいて、細胞種ごとの各RNAの発現量を推定する方法である。この方法は、新鮮な組織だけでなく、凍結保存した組織を用いたRNA発現解析が可能である。また、この方法は、多数の器官からのRNAの同時精製も可能である。

【0007】

器官全体RNA-Seqデータをデコンボリューションするためのいくつかのコンピュータ解析方法がこれまで提案されている (非特許文献6 - 19)。これらの方法は、対応する器官のほぼ全体のRNA-Seqデータを使用して、解析対象の器官の細胞種の組成を計算する。

【0008】

最近、Multi-Subject Single Cell deconvolution (MuSiC) (非特許文献17)、Damped Weighted Least Squares (DWLS) (非特許文献18)、及びComplete Deconvolution for Sequencing data (CDSeq) 法 (非特許文献19) が報告された。これら3つの方法は、以前に報告された非特許文献6から非特許文献16に記載の方法よりも優れているといわれている。

【先行技術文献】

【非特許文献】

【0009】

【文献】Deng, Q., Ramskold, D., Reinius, B. & Sandberg, R. *Science* 343, 193-196, doi:10.1126/science.1245316 (2014).

【文献】Han, X. et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* 172, 1091-1107 e1017, doi:10.1016/j.cell.2018.02.001 (2018).

【文献】Regev, A. et al. Science Forum: The Human Cell Atlas. *Elife* 6, doi:10.7554/eLife.27041 (2017).

【文献】Sandberg, R. *Nature methods* 11, 22-24, doi:10.1038/nmeth.2764 (2014).

【文献】Tabula Muris, C. et al., *Nature* 562, 367-372, doi:10.1038/s41586-018-0590-4 (2018).

【文献】Abbas, A. R. et al., *PloS one* 4, e6098, doi:10.1371/journal.pone.0006098 (2009).

【文献】Avila Cobos, F. et al., *Bioinformatics* 34, 1969-1979, doi:10.1093/bioinformatics/bty019 (2018).

【文献】Gaujoux, R. & Seoighe, C., *Infect Genet Evol* 12, 913-921, doi:10.1016/j.meegid.2011.08.014 (2012).

【文献】Gong, T. et al. *PloS one* 6, e27156, doi:10.1371/journal.pone.0027156 (2011).

【文献】Gong, T. & Szustakowski, J. D., *Bioinformatics* 29, 1083-1085, doi:10.1093/bioinformatics/btt090 (2013).

【文献】Li, B. et al., *Genome biology* 17, 174, doi:10.1186/s13059-016-1028-

10

20

30

40

50

7 (2016).

【文献】Newman, A. M. et al., Nature methods 12, 453-457, doi:10.1038/nmeth.3337 (2015).

【文献】Repsilber, D. et al., BMC bioinformatics 11, 27, doi:10.1186/1471-2105-11-27 (2010).

【文献】Shen-Orr, S. S. & Gaujoux, R., Curr Opin Immunol 25, 571-578, doi:10.1016/j.coi.2013.09.015 (2013).

【文献】Wang, N. et al., Bioinformatics 31, 137-139, doi:10.1093/bioinformatics/btu607 (2015).

【文献】Zhong, Y. et al., BMC bioinformatics 14, 89, doi:10.1186/1471-2105-14-89 (2013).

【文献】Tsoucas, D. et al., Nat Commun 10, 2975, doi:10.1038/s41467-019-10802-z (2019).

【文献】Wang, X. et al., Nat Commun 10, 380, doi:10.1038/s41467-018-08023-x (2019).

【文献】Kang, K. et al., PLoS computational biology 15, e1007510, doi:10.1371/journal.pcbi.1007510 (2019).

【発明の概要】

【発明が解決しようとする課題】

【0010】

しかし、非特許文献17から非特許文献19に記載方法、合成データセット、培養細胞、いくつかの組織の混合物、及び/又は1個から4個の実際の器官に由来するRNA-Seqデータにおいて有用性が検証されたに過ぎない。言い換えると、より多様な実際の器官への適用可能性は検討されていない。本発明者は、MuSiC（非特許文献17）およびDWLS法（非特許文献19）について性能を評価した。これらは、1個から4個の実際の器官についてデコンボリューションを行う、他の以前の方法と比較され、他の方法よりも優れていることが示された最新の2つの方法である。しかし、後述する効果の検証に示すように、MuSiC又はDWLS法によって、コンピュータにより算出された細胞種の比率が、実際のscRNA-Seqの研究によって実験的に推定されたものから乖離しており、また乖離の程度も様々であった。特に、骨格筋及び心臓では乖離が顕著であった。

【0011】

そこで、このよう乖離を解消するために、本発明は、実際の組織における各種細胞の割合により近い、各細胞種の割合を推定するためのRNA-Seqデータのデコンボリューション法を提供することを課題とする。また、より多種の組織に対応可能なRNA-Seqデータのデコンボリューション法を提供することを課題とする。

【課題を解決するための手段】

【0012】

本発明のある実施形態は、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けすること、を含むシングルセルRNA-Seq解析のカウントデータセットの補正方法に関する。

【0013】

好ましくは、前記重み付けが、各細胞種を特徴付けるシグネチャ遺伝子セットの発現に基づいて行われ、前記シグネチャ遺伝子セットは、所定数の遺伝子数を含む。

【0014】

本発明のある実施形態は、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けすることと、重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する各細胞種におけるRNAの発現パターンを解析すること、と含む、シングルセル

RNA-Seqの解析方法に関する。

【0015】

本発明のある実施形態は、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けすることと、重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する細胞種の構成比率を解析することを含む、解析対象器官を構成する細胞種の構成比率の解析方法に関する。

【0016】

本発明のある実施形態は、シングルセルRNA-Seq解析のカウントデータセットの補正装置(10)に関する。補正装置(10)は、制御部(101)を備える。制御部(101)は、解析対象細胞から取得したシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けする。

10

【0017】

本発明のある実施形態は、シングルセルRNA-Seqの解析装置に関する。解析装置(20)は、制御部(201)を備える。制御部(201)は、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けし、重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する各細胞種におけるRNAの発現パターンを解析する。

20

【0018】

本発明のある実施形態は、解析対象器官を構成する細胞種の構成比率の解析装置に関する。解析装置(20)は、制御部(201)を備える。制御部(201)は、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けし、重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する細胞種の構成比率を解析する。

【0019】

本発明のある実施形態は、コンピュータに実行させた時に、コンピュータに、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けするステップを備える処理を実行させる、シングルセルRNA-Seq解析のカウントデータセットの補正プログラムに関する。

30

【0020】

本発明のある実施形態は、コンピュータに実行させた時に、コンピュータに、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けするステップと、重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する各細胞種におけるRNAの発現パターンを解析するステップと、を備える処理を実行させる、シングルセルRNA-Seqの解析プログラムに関する。

40

【0021】

本発明のある実施形態は、コンピュータに実行させた時に、コンピュータに、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、前記解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けするステップと、重み付けされたシングルセルRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する細胞種の構成比率を解析するステップと、を備える処理を実行させる、解析対象器官を構成する細胞種の構成比率の解析プログラム。

【発明の効果】

50

【 0 0 2 2 】

本発明は、RNA配列データベースから、実際の組織における各種細胞の割合により近い、各細胞種の割合を推定することができる。また、本発明によれば、より多種の組織において、各細胞種の割合を推定することができる。

【図面の簡単な説明】

【 0 0 2 3 】

【図 1】補正装置 10 のハードウェア構成の例を示す。

【図 2】補正プログラム 1042 の処理の流れを示す。

【図 3】解析装置 10 のハードウェア構成の例を示す。

【図 4】解析プログラム 2042 の処理の流れを示す。

【図 5】大動脈、脳、脂肪、心臓、腎臓、大腸、肝臓、及び肺における各器官に存在する各細胞種の参照細胞種の構成比率と、MuSiCで予測された細胞種の構成比率と、DWLS法で予測された細胞種の構成比率を示す。

【図 6】骨髄、膵臓、皮膚、骨格筋、脾臓、及び胸腺における各器官に存在する各細胞種の参照細胞種の構成比率と、MuSiCで予測された細胞種の構成比率と、DWLS法で予測された細胞種の構成比率を示す。

【図 7】参照細胞種の構成比率と各器官の実際のscRNA-Seqデータから得られた推定器官全体RNA-Seqデータセットと、実際の器官全体RNA-Seqデータセットとの比較を示す。

【図 8】各器官に存在する各細胞種の重み係数とその分布範囲を示す。

【図 9】本発明で求められた細胞種特異的な重み係数を用いて推定した推定器官全体RNA-Seqデータセットと実際の器官全体RNA-Seqデータセットとの比較を示す。

【図 10】本発明による器官全体RNA-Seqデータのデコンボリューション法の概要を示す。図中 w は重みを示し、 m は各遺伝子のRNAのカウントを示し、 n は各細胞種の比率を示す。

【図 11】大動脈、脂肪、心臓、腎臓、肝臓、肺、大腸、骨髄、骨格筋、及び脾臓における各器官に存在する各細胞種の参照細胞種の構成比率と、本発明により推定した各細胞の構成比率と、MuSiCで予測された細胞種の構成比率と、DWLS法で予測された細胞種の構成比率を示す。

【図 12】参照細胞種の構成比率に対する、本発明により推定した各細胞の構成比率と、MuSiCで予測された細胞種の構成比率と、DWLS法で予測された細胞種の構成比率の平均二乗誤差 (MSE) を示す。

【図 13】大動脈、脂肪、心臓、腎臓、肝臓、肺、大腸、骨髄、骨格筋、及び脾臓における推定トランスクリプトカウントと、実際の器官内の各細胞種における遺伝子発現との比較を示す。

【図 14】推定scRNA-Seqカウントデータのt-Distributed Stochastic Neighbor Embedding (t-SNE) 解析結果を示す。

【図 15】心筋梗塞 (MI) のマウスモデルを用いて、本発明により心臓の細胞種の構成比率と、各細胞種における遺伝子発現プロファイルを推定した時の結果を示す。 a は、推定した細胞種の構成比率のShamに対する変化率を示す。 b は、推定した遺伝子発現プロファイルの変動解析結果を示す。

【図 16】マウスのデータを使用して算出した重み係数と推定scRNA-Seqカウントデータを用いて、ヒト器官全体RNA-Seqデータセットのデコンボリューションを行った時の結果を示す。 a は、ヒトの心臓及び腎臓について推定した細胞種の構成比率を示す。 b は、ヒトの心臓及び腎臓について推定した遺伝子発現プロファイルのt-Distributed Stochastic Neighbor Embedding (t-SNE) 解析結果を示す。

【発明を実施するための形態】

【 0 0 2 4 】

1. 補正方法、解析方法、及び解析対象器官を構成する細胞種の構成比率の解析方法

本発明のある実施形態は、シングルセルRNA-Seq解析のカウントデータセットの補正方法、補正装置及び補正プログラムに関する。

10

20

30

40

50

1 - 1 . シングルセルRNA-Seq解析のカウントデータセットの補正方法

【 0 0 2 5 】

シングルセルRNA-Seq (scRNA-Seq)解析のカウントデータセットの補正方法(以下、単に「補正方法」とも称する)は、解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、細胞種ごとの総RNAの含有量に基づいて重み付けすること、を含む。

1 - 1 - 1 . RNA-Seq解析

【 0 0 2 6 】

本明細書において、RNAは、RNA-Seq解析により解析できるRNAである限り制限されない。RNAには、mRNA、非翻訳RNA、microRNA等を含み得る。

【 0 0 2 7 】

RNAは、生物に存在する限り制限されない。生物は、器官を備える多細胞生物である限り制限されない。生物は、動物であっても植物であってもよいが、好ましくは動物である。動物として好ましくは、ヒト、マウス、ラット、イヌ、ネコ、ウサギ、ウシ、ウマ、ヤギ、ヒツジ、ブタ等の哺乳動物、ニワトリ等の鳥類等である。より好ましくはヒト、マウス、イヌ、ネコ、ウシ、ウマ、ブタ等の哺乳動物であり、さらに好ましくはヒト、マウス、イヌ、又はネコ等であり、さらにより好ましくはヒト又はマウスであり、最も好ましくはヒトである。また、生物には疾患を有する生物と有さない生物の両方が含まれる。

解析対象細胞は、前記生物の器官に存在する限り制限されない。器官は、好ましくは器官内の細胞構成がわかっている細胞である。

【 0 0 2 8 】

器官とは、生物に存在する、いくつかの組織の集まりで、一定の独立した形態および特定の機能を有するものを意味する。例えば、生物が哺乳類である場合、器官には、循環器系器官(心臓、動脈、静脈、リンパ管等)、呼吸器系器官(鼻腔、副鼻腔、喉頭、気管、気管支、肺等)、消化器系器官(口唇、頬部、口蓋、歯、歯肉、舌、唾液腺、咽頭、食道、胃、十二指腸、空腸、回腸、盲腸、虫垂、上行結腸、横行結腸、S状結腸、直腸、肛門、肝臓、胆嚢、胆管、胆道、膵臓、膵管等)、泌尿器系器官(尿道、膀胱、尿管、腎臓)、神経系器官(大脳、小脳、中脳、脳幹、脊髄、末梢神経、自律神経等)、女性生殖器系器官(卵巣、卵管、子宮、膣等)、乳房、男性生殖器系器官(陰茎、前立腺、精巣、精巣上体、精管)、内分泌系器官(視床下部、下垂体、松果体、甲状腺、副甲状腺、副腎等)、外皮系器官(皮膚、毛、爪等)、造血器系器官(血液、骨髄、脾臓等)、免疫系器官(リンパ節、扁桃、胸腺等)、骨軟部器官(骨、軟骨、骨格筋、結合組織、靭帯、腱、横隔膜、腹膜、胸膜、脂肪組織(褐色脂肪、白色脂肪)等)、感覚器系器官(眼球、眼瞼、涙腺、外耳、中耳、内耳、蝸牛等)等を含み得る。本発明において、対象とする組織としては、好ましくは心臓、大脳、肺、腎臓、脂肪組織、肝臓、骨格筋、精巣、脾臓、胸腺、骨髄、膵臓、皮膚(例えば、皮下組織より上層の表皮、乳頭層、網状層を含む。器官として、大動脈、脳、脂肪、心臓、腎臓、大腸、肝臓、肺、骨髄、膵臓、皮膚、骨格筋、脾臓、及び胸腺が好ましい。

【 0 0 2 9 】

RNA-Seq解析は、いわゆるトランスクリプトーム解析と呼ばれる解析であり、対象とするサンプル内に存在するRNAから、網羅的に配列情報を含むリードを取得し、そのリードを参照配列にマッピングし、発現している遺伝子や、そのカウント数(リード数ともいう)を解析する方法である。カウント数が遺伝子発現量に対応する。RNA-Seq解析のカウントデータには、発現している遺伝子の遺伝子名及び/又は遺伝子データベースへの登録番号と、各遺伝子のリードのカウント数を含み得る。

【 0 0 3 0 】

RNA-Seq解析は、次世代シーケンサー又は第3世代シーケンサーと呼ばれるDNAシーケンサーを用いて行うことができる。次世代シーケンサーとしては、例えば、イルミナ社(サンディエゴ、CA)のMiSeq9(登録商標)、HiSeq(登録商標)、NextSeq(登録商標)、MiSeq(登録商標);サーモ・フィッシャー社(ウォルサム、MA)のIon Proton

10

20

30

40

50

(登録商標)、Ion PGM (登録商標) ; ロシュ社 (バーゼル、スイス) のGS FLX + (登録商標)、GS Junior (登録商標) 等が挙げられる。第三世代シーケンサーとしては、Pa cBio Sequel (商標) 等が挙げられる。

【0031】

scRNA-Seq解析のカウントデータセットは、生物の個々の細胞に発現する遺伝子発現解析、及び/又はコンピュータ解析方法によって予測された遺伝子発現に基づいて生成されたカウントデータのセットである。例えば、scRNA-Seq解析のカウントデータセットは、RNA-Seq解析によって、実際の個々の細胞から取得されたカウントデータであり得る。また、scRNA-Seq解析のカウントデータセットは、例えば、非特許文献6 - 19に記載の方法により器官全体からRNA-Seq解析によって取得したカウントデータを、コンピュータ解析方法によって参照細胞構成比率に基づいてデコンボリューション等することにより予測されたカウントデータセットであってもよい。scRNA-Seq解析のカウントデータセットを予測する方法としては、例えば、Complete Deconvolution for Sequencing data (CDSeq) 法 (非特許文献19) が好ましい。

10

【0032】

1 - 1 - 2 . 各器官に存在する異なる細胞種におけるRNA含有量に基づく重み係数の算出
解析対象細胞から取得した、又は解析対象細胞について予測されたシングルセルRNA-Seq解析のカウントデータセットに、細胞種ごとの総RNAの含有量に基づいて重み付けするための重み係数の算出方法について説明する。

【0033】

はじめに、重み付けをするにあたり、各器官がどのような細胞種から構成されているかの情報を取得する必要がある。各器官の細胞構成は、非特許文献5又は非特許文献2に記載のscRNA-Seqデータや、NIH等に登録されているデータベースから取得することができる。これらの細胞種の構成は、実際に各器官の組織の細胞種の構成を解析して得られた情報である。このような各器官の細胞構成を「参照細胞種」とも呼ぶ。参照細胞種には、通常各細胞種において発現する遺伝子のscRNA-Seqのカウントデータセットが含まれている。また、参照細胞種には、各器官における参照細胞種の構成比率 (レファレンスとも呼ぶ) が含まれており、各細胞種の名称、又は略称を示すラベルと紐付いている。

20

【0034】

重み係数の算出において、大動脈、脂肪、腎臓、大腸、肝臓、肺、骨髄、膵臓、皮膚、脾臓、および胸腺については、非特許文献5に記載の各器官における細胞種の構成を参照細胞種及びその構成比率として使用することが好ましい。

30

また、骨格筋については、非特許文献2に記載されている細胞種の比率を参照細胞種とその構成比率を使用することが好ましい。

【0035】

心臓については、心筋細胞と非筋肉細胞の分離分析に合わせて、非特許文献5に記載されている参照細胞種の構成比率を修正し参照細胞種の構成比率として用いることが好ましい。具体的には、非特許文献5において採用されている心筋細胞の構成比率 (3.1%) は、さまざまな過去の研究に基づいて、組織解剖学の分野で一般的にコンセンサスが得られている比率 (30% から40%) と比較して、極端に少ない。したがって、本実施例では、心筋細胞の構成比率を30% に設定し、残りの70% を非筋肉細胞種の構成比率で割って、参照細胞種の構成比率とすることが好ましい。

40

【0036】

脳については、NIH (<http://www.nervenet.org/papers/BrainRev99.html#Numbers>) における報告に基づいて、参照細胞種及びその構成比率を決定することが好ましい。脳内の各細胞種を示すラベルとして、非特許文献5に記載されているscRNA-Seqデータの対応する細胞種ラベルを使用した。はじめに、まず、脳内の細胞種のクラスを「ニューロン」、「グリア細胞」、「内皮細胞」、「その他」の4つのクラスに分類し、それぞれのクラスの比率をそれぞれ75 : 23 : 7 : 4に設定する。この比率は、マウスの脳内の推定比率 (<http://www.nervenet.org/papers/BrainRev99.html#Numbers>) にしたがって

50

いる。次に、非特許文献5にしたがって、「ニューロン」、「グリア細胞」、および「その他」をより詳細な細胞種クラスにさらに分類する。具体的には、「ニューロン」のクラスをさらに「神経細胞 - 興奮性ニューロン、及びいくつかの神経幹細胞」と「神経細胞 - 抑制性ニューロン」に分類する。「その他」のクラスは、「脳周皮細胞-NA」と「乏突起膠細胞前駆細胞-NA」に分類する。「グリア細胞」のクラスに関しては、次の3つの前提に基づいて分類する。i) 「グリア細胞」は、非特許文献5にしたがって、「ミクログリア細胞-NA」、「アストロサイト-NA」、「バグマングリア細胞-NA」、及び「乏突起膠細胞NA」の4つの細胞種に分類する。ii) これらの4つのグリア細胞種の構成比率は、非特許文献5の記載にしたがう。iii) 脳全体の細胞において「ミクログリア細胞」が10 - 15%を占めると報告されているため、脳全体の「ミクログリア細胞-NA」の比率は0.1とする。これらの前提に基づいて、各脳細胞種の比率は、「マクロファージ - NA」（約0.2%）、「ミクログリア細胞 - NA」（10.0%）、「アストロサイト - NA」（約2.2%）、「バグマングリア細胞 - NA」（約2.1%）、「脳周皮細胞 - NA」（約1.5%）、「内皮細胞 - NA」（約6.4%）、「神経細胞 - 興奮性ニューロン、及びいくつかの神経幹細胞」（約47.5%）、「神経細胞 - 抑制性ニューロン」（約21.3%）、「乏突起膠細胞 - NA」（約8.7%）、「乏突起膠細胞前駆細胞 - NA」（約1.9%）と設定し、これらを脳の参照細胞種及びその構成比率として用いることができる。

10

本明細書において使用する参照細胞種、及びその細胞種の構成比率を、本書末尾の参照細胞種の構成比率リストに示す。

【0037】

20

さらに、重み係数を算出するにあたり、上記、参照細胞種の構成比率に加え、各細胞種における遺伝子発現、すなわち各細胞種における、scRNA-Seq解析のカウントデータが必要になる。しかし、一般的にscRNA-Seq解析の対象となっている遺伝子数は2万から3万ある。

【0038】

これら全ての遺伝子のカウントデータを使用してもよいが、各細胞種を特徴付けることができる遺伝子（シグネチャ遺伝子）を選択し、これらの遺伝子セットのカウントデータを使って重み係数を算出することが、効率的である。このような各細胞種を特徴付けるシグネチャ遺伝子セットは、例えば次に述べる方法により、算出することができる。

【0039】

30

はじめに、シグネチャ遺伝子を選択するにあたり、scRNA-Seq解析のカウントデータには、ERCCラベルが付されるスパイクイン遺伝子のカウントや、合計カウントに大きく影響するものの、非mRNAアーチファクトとして報告されているRn45s、Akap5、Lrrc17の3つの遺伝子に由来するカウントは削除することが好ましい。また、各遺伝子に由来するRNAカウントは、scRNA-Seqデータセット内の各細胞の総カウントが 10^0 、 10^3 、 10^4 、 10^5 、 10^6 である場合等に換算し、正規化しておくことが好ましい。

【0040】

シグネチャ遺伝子の選択には、例えば、ランダムフォレスト等の人工知能を訓練して生成した分類器を使用することができる。各器官の参照細胞種の構成比率と、各参照細胞種について報告されているscRNA-Seq解析のカウントデータセットを用いて、人工知能を訓練し、分類器を生成する。例えば、人工知能としてランダムフォレストを使用する場合、分類器の重要な特徴量を各細胞種のシグネチャ遺伝子名として抽出し、“Mean Decrease Gini”値を各遺伝子の重要度指標として、“Mean Decrease Gini”値の高い遺伝子を、シグネチャ遺伝子として抽出した。シグネチャ遺伝子は、“Mean Decrease Gini”値が高い方から、 10^0 遺伝子から 2000 遺伝子程度を抽出し、シグネチャ遺伝子セットとして使用することができる。

40

次に、各器官に存在する各細胞種について、scRNA-Seq解析のカウントデータセットをRNA含有量で補正するための重み係数を算出する。

【0041】

重み係数の算出には、各器官について、参照細胞種の各細胞種におけるシグネチャ遺伝

50

子セットのscRNA-Seq解析のカウントデータ（シグネチャ遺伝子scRNA-Seqデータともいう）と、各器官の全体に含まれる全RNAをRNA-Seq解析したカウントデータ（器官全体RNA-Seqデータともいう）を用いることができる。シグネチャ遺伝子scRNA-Seqカウントデータ及び器官全体RNA-Seqカウントデータはとも正規化して用いる。

【0042】

器官全体RNA-Seqデータとして、公開されているRNA-Seq解析のカウントデータセットを使用することができる。マウスの器官全体RNA-Seqデータは「i-organs.atr.jp」から取得することができる。ヒト器官全体RNA-Seqデータは「The Human Protein Atlas」（<https://www.proteinatlas.org/>;心臓（ERR315328）および腎臓（ERR315494））から取得することができる。

10

重み係数は、以下の方法にしたがって、算出することができる。

【0043】

【数1】

[math. 1]

$$\mathbf{y} \in \mathbb{R}^n, \quad w_j > 0, \quad \text{及び} \quad \mathbf{x}_j \in \mathbb{R}^n$$

は、それぞれ正規化された器官全体RNA-Seq解析のカウントのベクトル、各解析対象細胞 j 重み係数、及び正規化されたscRNA-Seq解析のカウントの行列を示す。ここで、 n はそれぞれの器官のシグネチャ遺伝子の遺伝子数を示す。加えて、解析対象細胞の組み合わせ C^m は、参照細胞種の構成比率が合計セットサイズ m 内に保たれるという制限下でランダムに選択する。また、解析対象細胞について予測されたscRNA-Seqのカウントデータセットを用いる場合には、正規化されたscRNA-Seq解析のカウントの行列にかえて、解析対象細胞について予測されたscRNA-Seqのカウントデータの行列を使用する。

20

次に下式（1）について説明する。

【0044】

【数2】

[math. 2]

$$\mathbf{m}\mathbf{y} = \sum_j w_j \mathbf{x}_j \quad (j \in C^m) \quad (1)$$

30

式（1）において m は倍率を示し、 n に依存して決定される。 m は、以下の計算のそれぞれにおいて、 n 未満に設定される。ここで w_j は、結果の値が0.01以上であるという制限の下で下式（2）にしたがって、二次計画問題を解くことにより算出する。

【0045】

【数3】

[math. 3]

$$\tilde{w}_j = \operatorname{argmin}_{w_j} \sum_i^S \left| \mathbf{m}\mathbf{y}_i - \sum_j w_j \mathbf{x}_j \right|^2 \quad s.t. \quad w_j \geq 0.01. \quad (2)$$

40

式（2）において、 S は、器官全体を対象とする各遺伝子のRNA-Seqのカウントデータセットの数を示す。例えば、異なる2個体から取得した対応する器官全体RNA-Seqのカウントデータセットを使用する場合、 S は2である。この二次計画問題は、Rで「quadprog

50

」パッケージを使用して解くことができる。解析対象細胞の組み合わせのランダム選択と
【数 4】

$$\begin{aligned} &[\text{math. 4}] \\ &\tilde{w}_j \end{aligned}$$

の算出の両方の工程は w_j の数が100以上になるまで再帰的に行い選択した解析対象細胞すべてについて行う。

【0046】

1 - 1 - 3 . scRNA-Seq解析のカウントデータセットの補正

算出された重み係数を使用し、解析対象細胞から取得した、又は解析対象細胞について予測されたscRNA-Seqのカウントデータセットに、解析対象細胞に対応する細胞種ごとの総RNAの含有量に基づいて重み付けする。

【0047】

重み付けは重み係数の分布がガウス分布にしたがうと仮定して、各解析対象細胞の遺伝子の重み付きカウントの平均と分散を下式(3)にしたがって算出する。

【数 5】

$$[\text{math. 5}]$$

$$\tilde{x}_j \sim \mathcal{N}\left(\mu_{w_j} x_j, \sigma_{w_j}^2 \text{diag}(x_j \odot x_j)\right) \quad (3)$$

式(3)において、

【数 6】

$$[\text{math. 6}]$$

$$\tilde{x}_j, \mu_{w_j}, \text{及び} \sigma_{w_j}^2$$

は、それぞれ解析対象細胞jの遺伝子の重み付きカウントベクトル、解析対象細胞jの重み係数の平均値、及び解析対象細胞jの重み係数の分散を示す。

【数 7】

$$[\text{math. 7}]$$

$$\text{作用素 } \odot$$

は、2つのベクトル間の要素ごとの積を表す。

【0048】

算出された各解析対象細胞の重み係数の平均と分散に基づいて、対応する細胞種における重み係数の平均と分散がガウス分布にしたがうと仮定して、対応する細胞種における重み係数の平均と分散を下式(4)にしたがって算出する。

【数 8】

10

20

30

40

50

[math. 8]

$$\bar{x}_k \sim \mathcal{N}\left(\frac{1}{N_k} \sum_j \mu_{w_j} x_j, \frac{1}{N_k} \sum_j \sigma_{w_j}^2 \text{diag}(x_j \odot x_j)\right) \quad (j \in C^k). \quad (4)$$

【0049】

式(4)において、 k 、 C^k 、及び N_k は、それぞれ細胞種、細胞種 k にラベルされた解析対象細胞の群、及び C^k における解析対象細胞の数を示す。

10

【0050】

上記式(2)にしたがって、算出された各器官に存在する細胞種ごとの重み係数の平均値、分散、四分位数を後述する重み係数リストに示す。

【0051】

本法によって重み付けされたscRNA-Seq解析のカウントデータセットは、推定scRNA-Seqカウントデータセットともいう。

【0052】

2. 各器官の細胞種の構成比率の解析と前記細胞種における全RNA発現パターンの解析

上記1-1-2. で算出された重み係数を使用して、解析対象器官を構成する細胞種の構成比率を解析することができる。解析対象器官を構成する細胞種の構成比率の解析は、上記1-1-3. で重み付けされたscRNA-Seq解析のカウントデータセットに基づいて、解析対象細胞を含む解析対象器官を構成する細胞種の構成比率を算出することを含む。言い換えると、この方法によって取得される細胞種の構成比率は、推定された構成比率である。

20

【0053】

また、上記1-1-2. で算出された重み係数を使用して、解析対象器官を構成する細胞種における全RNA発現パターンの解析を行うことができる。全RNA発現パターンの解析は、推定されたscRNA-Seq解析のカウントデータを取得することである。ここで、全RNAとは、シグネチャ遺伝子セット及びそれ以外の遺伝子から発現されるRNAを含むことを意図する。

30

【0054】

例えば、細胞種の構成比率と、各細胞種における全RNAの発現パターンとの解析は、ベイズの定理に基づいたアルゴリズムを設計し、同時に計算することができる。

計算は下式(5)にしたがうことができる。

【数9】

[math. 9]

$$y = \mathbf{Xr} \quad (5)$$

40

式(5)において、

【数10】

[math. 10]

$$y, \mathbf{X} = (\bar{x}_1, \dots, \bar{x}_k, \dots, \bar{x}_K), \text{ 及び } \mathbf{r}$$

は、それぞれ、器官全体RNA-Seq データベクトル、列を各細胞種の重み係数を用いて上

50

記式(4)にしたがって計算した重み付けしたカウントである推定scRNA-Seqデータとする行列、及び細胞種の構成比率に対応する係数ベクトルを示す。上記式(4)にしたがって計算した各細胞種の重み係数で重み付けしたカウントは、初期値であり、後述する式(7)の計算により新たな値に更新される。Xとrの算出には、ベイズの定理を用いる。ベイズの定理を適用するために、

【数11】

$$\begin{aligned} &[\text{math. 11}] \\ &\text{ノイズ } \mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I}) \end{aligned}$$

10

を式(5)に加え、下式(6)に示す確率モデルを採用する。

【数12】

$$\begin{aligned} &[\text{math. 12}] \\ &\mathbf{P}(\mathbf{y}|\mathbf{X}, \mathbf{r}) = \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{r}, \beta^{-1}\mathbf{I}) \end{aligned} \quad (6)$$

20

式(6)において、 β は各細胞種の遺伝子発現パターンを推定する際の分布のばらつきの度合いを制御するためのハイパーパラメータを示す。ベイズの定理にしたがって、X及びrの事後分布を下式(7)として得る。

【数13】

$$\begin{aligned} &[\text{math. 13}] \\ &\mathbf{P}(\mathbf{X}, \mathbf{r}|\mathbf{y}) \propto \mathbf{P}(\mathbf{y}|\mathbf{X}, \mathbf{r})\mathbf{P}(\mathbf{X})\mathbf{P}(\mathbf{r}) \end{aligned} \quad (7)$$

30

式(7)において、 $\mathbf{P}(\mathbf{X})$ 及び $\mathbf{P}(\mathbf{r})$ は、それぞれXとrの事前分布を示す。そして、 $\mathbf{P}(\mathbf{X})$ 及び $\mathbf{P}(\mathbf{r})$ は、下式(8)及び(9)として与えられる。

【数14】

$$\begin{aligned} &[\text{math. 14}] \\ &\mathbf{P}(\mathbf{X}) = \prod_{k=1}^K \mathbf{P}(\bar{\mathbf{x}}_k) = \prod_{k=1}^K \mathcal{N}(\bar{\mathbf{x}}_k | \boldsymbol{\mu}'_k, \boldsymbol{\Sigma}_k^{-1}) \end{aligned} \quad (8)$$

$$\mathbf{P}(\mathbf{r}) = \mathcal{N}(\mathbf{r} | \mathbf{0}, \alpha^{-1}\mathbf{I}) \quad (9)$$

40

式(9)において、 α は細胞種の比率を推定する際の分布のばらつきの度合いを制御するハイパーパラメータである。

【数15】

50

[math. 15]

事前分布 $P(\mathbf{X})$ において、 $\mathcal{N}(\bar{\mathbf{x}}_k | \boldsymbol{\mu}'_k, \boldsymbol{\Sigma}_k^{-1})$ は式 (4) と同じである。

$\boldsymbol{\mu}'_k = \frac{1}{N_k} \sum_j \mu_{wj} \mathbf{x}_j$ であり、 $\boldsymbol{\Sigma}_k^{-1} = \frac{1}{N_k} \sum_j \sigma_{wj}^2 \text{diag}(\mathbf{x}_j \odot \mathbf{x}_j)$ である。

事後分布を直接最大化する代わりに、

\mathbf{r} と、 $\bar{\mathbf{x}}_k$ のそれぞれについて、 $P(\mathbf{r} | \mathbf{y}, \mathbf{X})$ と $P(\bar{\mathbf{x}}_k | \mathbf{y}, \mathbf{r}, \{\bar{\mathbf{x}}_l\}_{l \neq k})$ を最大化した $\bar{\mathbf{x}}_k$ を推定する反復法を採用する。

具体的には、 $\bar{\mathbf{x}}_k$ の推定のために、下式 (10) で表される確率の式を採用する。

$$P(\bar{\mathbf{x}}_k | \mathbf{y}, \mathbf{r}, \{\bar{\mathbf{x}}_l\}_{l \neq k}) \propto P(\mathbf{y} | \mathbf{X}, \mathbf{r}) P(\bar{\mathbf{x}}_k) . \quad (10)$$

$P(\bar{\mathbf{x}}_k | \mathbf{y}, \mathbf{r}, \{\bar{\mathbf{x}}_l\}_{l \neq k})$ はガウス分布にしたがい、その平均と分散を下式 (11) 及び式 (12) したがって算出する。

$$\mathbb{E}[\bar{\mathbf{x}}_k] = (\boldsymbol{\beta} r_k^2 \mathbf{I} + \boldsymbol{\Sigma}_k^{-1})^{-1} \left(\boldsymbol{\beta} r_k \mathbf{y} + \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}'_k - \boldsymbol{\beta} \sum_{l \neq k} r_l \bar{\mathbf{x}}_l \right) \quad (11)$$

$$\text{Var}[\bar{\mathbf{x}}_k] = \boldsymbol{\beta} r_k^2 \mathbf{I} + \boldsymbol{\Sigma}_k^{-1} . \quad (12)$$

$\mathbb{E}[\bar{\mathbf{x}}_k]$ の結果において、すべての負の値を「0」に設定する。

\mathbf{r} の推定のために、下式 (13) で表される確率の式を採用する。

$$P(\mathbf{r} | \mathbf{y}, \mathbf{X}) \propto P(\mathbf{y} | \mathbf{X}, \mathbf{r}) P(\mathbf{r}) . \quad (13)$$

そして、 $P(\mathbf{r} | \mathbf{y}, \mathbf{X})$ はガウス分布にしたがい、その平均と分散を下式 (14) したがって算出する。

$$\mathbb{E}[\mathbf{r}] = (\boldsymbol{\alpha} \mathbf{I} + \boldsymbol{\beta} \mathbf{X}^t \mathbf{X})^{-1} \boldsymbol{\beta} \mathbf{X}^t \mathbf{y} . \quad (14)$$

結果の $\mathbb{E}[\mathbf{r}]$ と $\mathbb{E}[\bar{\mathbf{x}}_k]$ の両方について、すべての負の値を「0」に設定する。

遺伝子発現パターン ($\{\bar{\mathbf{x}}_k^*\}_{k=1}^K$) と細胞種の構成比率 (\mathbf{r}^*) を推定するため、 \mathbf{X} と \mathbf{r} の両方が収束するか、そうでない場合には、1001 回の反復が行われるまで交互に繰り返し計算を行う。

【0055】

最初の \mathbf{r} と \mathbf{X} として、細胞種の構成比率と式 (4) の計算式により重み付けされた参照データセットのカウントを使用する。ハイパーパラメータ と は、便宜上 10^{-3} 、 10^{-2} 、...、 10^3 に設定することができる。実際の器官全体 RNA-Seq と高い類似性 (高いピアソン相関係数及びスピアマン相関係数を示し、かつ低い平均二乗誤差に基づいて類似性を決定) を生じたシグネチャ遺伝子セット (100 から 2000 遺伝子) ハイパーパラメータ (および) の数の組み合わせの結果を最良の推定結果として選択できる。

【0056】

3 . scRNA-Seq 解析のカウントデータセットの補正装置、scRNA-Seq の解析装置、及び解析対象器官を構成する細胞種の構成比率の解析装置

3 - 1 . scRNA-Seq 解析のカウントデータセットの補正装置

図 1 に、scRNA-Seq 解析のカウントデータセットの補正装置 10 のハードウェアの構成を示す。

【0057】

補正装置 10 は、汎用コンピュータであり得る。補正装置 10 は、入力デバイス 111 と、出力デバイス 112 と、メディアドライブ 113 と通信可能に接続されている。補正装置 10 は、CPU 101 と、メモリ 102 と、ROM (read only memory) 103 と、記憶デバイス 104 と、通信インターフェース (I/F) 105 と、入力インターフェース (I/F) 106 と、出力インターフェース (I/F) 107 と、メディアインターフェース (I/F) 108 とを備える。補正装置 10 内の各構成はバス 109 によって互いにデータ通信可能に接続されている。

【0058】

10

20

30

40

50

記憶デバイス104は、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等によって構成される。記憶デバイス104には、オペレーティングシステム(OS)1041と、後述する補正プログラム1042と、アルゴリズムデータベース(DB)DB1と、参照細胞種データベース(DB)DB2と、器官全体RNA-Seqデータベース(DB)DB3が格納されている。補正プログラム1042は、オペレーティングシステム1041と協働して、コンピュータを補正装置10として機能させる。

【0059】

CPU101は、本実施形態において制御部101とも呼ばれる。

アルゴリズムデータベースDB1は、上記1-1-3.で述べた補正を行うための数式を格納している。参照細胞種データベースDB2には、各器官に含まれる細胞種を示すラベルと、その構成比率と、各細胞種のscRNA-Seq解析のデータカウントが紐付けられて格納されている。また、参照細胞種データベースDB2には、補正された各細胞種のscRNA-Seq解析のデータカウントが、器官名を示すラベルと、細胞種名を示すラベルと紐付けられて格納される。器官全体RNA-SeqデータベースDB3には、マウス、またはヒトの各器官全体RNA-Seq解析のカウントデータが器官ごとに登録されている。これらのデータは、1-1-2.で述べた公知のデータから生成され格納されている。

10

【0060】

入力デバイス111は、タッチパネル、キーボード、マウス、ペンタブレット、マイク等から構成され、補正装置10に文字入力又は音声入力を行う。入力デバイス111は、制御部101の外部から接続されても、補正装置10と一体となってもよい。

20

【0061】

出力デバイス112は、例えばディスプレイ等の表示デバイス、プリンタ等で構成され、各種操作ウィンドウ、解析結果等を入力する。

【0062】

メディアドライブ113は、USBドライブ、フレキシブルディスクドライブ、CD-ROMドライブ、又はDVD-ROMドライブ等であり得る。

【0063】

通信I/F105は、外部のデータベースや他のコンピュータとの通信を行う。出力I/F107は、出力デバイス112へ情報の送信を行う。

【0064】

3-2. 補正プログラムの処理

図2に補正プログラム1042の処理の流れを示す。

補正装置10の制御部101は、はじめにオペレータが入力デバイス111から入力した処理開始指令を受け付け、処理を開始する。制御部101は、ステップS1において、上記1-1-2.に記載した方法にしたがって、解析対象器官の各細胞種を特徴付けるシグネチャ遺伝子を選択する。

30

【0065】

次に、制御部101は、ステップS2において、参照細胞種データベースDB2からステップS1において取得したいシグネチャ遺伝子セットのscRNA-Seqカウントデータを取得する。

40

【0066】

次に、制御部101は、ステップS3において、器官全体RNA-SeqデータベースDB3から、器官全体RNA-Seqカウントデータを取得する。なお、ステップS3は、ステップS2の前であってもよい。

【0067】

次に、制御部101は、ステップS4において、アルゴリズムデータベースDB1から、上記1-1-2.で述べた式(1)から式(4)を読み出す。読み出した各式に、ステップS2で取得したシグネチャ遺伝子セットのscRNA-SeqカウントデータとステップS4で取得した器官全体RNA-Seqカウントデータを使用し、上記1-1-2.で説明した式に基づいて各器官に存在する細胞種ごとの重み係数を算出する。制御部101は、算出した

50

重み係数をアルゴリズムデータベースDB1に格納する。

【0068】

最後に、制御部101は、ステップS5において、1-1-3.にしたがって細胞種ごとに重み付けしたscRNA-Seq解析のカウントデータセットを取得し、参照細胞種データベースDB2に格納する。

【0069】

さらに、制御部101は、オペレータが入力デバイス111から入力した出力処理開始指令を受信し、出力デバイス112から、重み付けしたscRNA-Seq解析のカウントデータセットを出力してもよい。

【0070】

この実施形態では、ステップS1からS5を一台のコンピュータで行う例を示しているが、例えば、ステップS1と、ステップS2からステップS4と、ステップS5は、別のコンピュータが行ってもよい。すなわち、第1のコンピュータがステップS1にしたがってシグネチャ遺伝子を選択し、第2のコンピュータが第1のコンピュータから各器官に存在する各細胞種のシグネチャ遺伝子セットの情報を取得し、ステップS2からステップS4の処理を行い、重み係数を算出する。さらに、第3のコンピュータが重み付けされたscRNA-Seq解析のカウントデータセットを取得してもよい。

【0071】

さらに、ステップS1からステップS4を第1のコンピュータが行い、第2のコンピュータがステップS5を行ってもよい。

【0072】

また、ステップS1を第1のコンピュータが行い、第2のコンピュータがステップS2からステップS5を行ってもよい。

【0073】

3-3. scRNA-Seqの解析装置、及び解析対象器官を構成する細胞種の構成比率の解析装置

上記2.において説明したように、scRNA-Seqの解析と解析対象器官を構成する細胞種の構成比率の解析は同時に行うことができる。したがって、解析装置20は、両方の処理を行う。

【0074】

図3に、解析装置20のハードウェア構成を示す。解析装置20の構成は、記憶デバイス204を除き、基本的には補正装置10の構成と同じである。記憶デバイス204は、補正プログラム1042にかえて、後述する解析プログラム2042を格納している。さらに、記憶デバイス204は、記憶デバイス104と同様に、アルゴリズムデータベース(DB)DB1と、参照細胞種データベース(DB)DB2と、器官全体RNA-Seqデータベース(DB)DB3を格納している。

【0075】

3-4. 解析プログラム処理

図4に、解析プログラム2042の処理の流れを示す。

解析装置20の制御部201は、はじめにオペレータが入力デバイス211から入力した処理開始指令を受け付け、処理を開始する。制御部201は、ステップS11において、上記2.において、説明したアルゴリズムを、アルゴリズムデータベースDB1から読み出す。

【0076】

次に、制御部201は、ステップS13において、器官全体RNA-SeqデータベースDB3から、器官全体RNA-Seqカウントデータを取得する。

【0077】

続いて、制御部201は、ステップS13において上記3-2.で取得された重み付けしたscRNA-Seq解析のカウントデータセットを参照細胞種データベースDB2から読み出し、アルゴリズムに適用する。

10

20

30

40

50

【0078】

次に、制御部201は、アルゴリズムによって推定された解析対象器官を構成する細胞種の構成比率と、推定されたscRNA-Seq解析のカウントデータを、推定結果として記憶デバイス204に記録する。

【0079】

制御部201は、推定結果について、解析対象器官を構成する細胞種の構成比率のみを出力デバイス212から出力してもよく、推定されたscRNA-Seq解析のカウントデータのみを出力デバイス212から出力してもよい。また、制御部201は、両方の結果を出力デバイス212から出力してもよい。

【0080】

4. コンピュータプログラムを記録した記録媒体

補正プログラム1042及び解析プログラム2042は、記録媒体に記録されていてもよい。

【0081】

すなわち、各プログラムは、ハードディスク、フラッシュメモリ等の半導体メモリ素子、光ディスク等の記録媒体に記憶される。また各プログラムは、クラウドサーバ等のネットワークで接続可能な記録媒体に記憶されていてもよい。各プログラムは、ダウンロード形式の、又は記録媒体に記録されたプログラム製品として提供されてもよい。

【0082】

前記記録媒体へのプログラムの記憶形式は、前記各装置が前記プログラムを読み取り可能である限り制限されない。前記記録媒体への記憶は、不揮発性であることが好ましい。

【実施例】

【0083】

以下に実施例を示して、本発明についてより詳細に説明する。しかし、本発明は実施例に限定して解釈されるものではない。

【0084】

I. 方法

1. 参照細胞種の構成比率の計算

大動脈、脳、脂肪、心臓、腎臓、大腸、肝臓、肺、骨髄、膵臓、皮膚、骨格筋、脾臓、及び胸腺の14種の器官について、非特許文献5に記載のscRNA-Seqデータや、NIH等に

登録されているデータベースに基づいて参照細胞種の構成比率を算出した。大動脈、脂肪、腎臓、大腸、肝臓、肺、骨髄、膵臓、皮膚、脾臓、および胸腺については、非特許文献5に記載の各器官における細胞種の比率を参照細胞種の構成比率として使用した。

【0085】

骨格筋については、非特許文献2に記載されている細胞種の比率を参照細胞種の構成比率として使用した。

【0086】

心臓については、心筋と非筋肉細胞の分離分析に合わせて、非特許文献5に記載されている細胞種の構成比率を修正し参照細胞種の構成比率(Referene)とした。具体的には、非特許文献5において採用されている心筋細胞の比率(3.1%)は、さまざまな過去の研究に基づいて、組織解剖学の分野で一般的にコンセンサスが得られている比率(30%から40%)と比較して、極端に少ない。したがって、本実施例では、心筋細胞の比率を30%に設定し、残りの70%を非筋肉細胞種の割合で割って、参照細胞種の構成比率とした。

【0087】

脳については、NIH(<http://www.nervenet.org/papers/BrainRev99.html#Numbers>)における報告に基づいて、参照細胞種の構成比率を決定した。脳内の各細胞種を示すラベルとして、非特許文献5に記載されているscRNA-Seqデータの対応する細胞種ラベルを使用した。はじめに、まず、脳内の細胞種のクラスを「ニューロン」、「グリア細胞」、「内皮細胞」、「その他」の4つのクラスに分類し、それぞれのクラスの比率をそれぞれ

10

20

30

40

50

れ75:23:7:4に設定した。この比率は、マウスの脳内の推定比率 (<http://www.nervenet.org/papers/BrainRev99.html#Numbers>) にしたがっている。次に、非特許文献5にしたがって、「ニューロン」、「グリア細胞」、および「その他」をより詳細な細胞種クラスにさらに分類した。具体的には、「ニューロン」のクラスをさらに「神経細胞 - 興奮性ニューロン、及びいくつかの神経幹細胞」と「神経細胞 - 抑制性ニューロン」に分類した。「その他」のクラスは、「脳周皮細胞-NA」と「乏突起膠細胞前駆細胞-NA」に分類した。「グリア細胞」のクラスに関しては、次の3つの前提に基づいて分類した。i) 「グリア細胞」は、非特許文献5にしたがって、「ミクログリア細胞-NA」、「アストロサイト-NA」、「バークマングリア細胞-NA」、及び「乏突起膠細胞NA」の4つの細胞種に分類できる。ii) これらの4つのグリア細胞種の比率は、非特許文献5の記載にしたがう。iii) 脳全体の細胞において「ミクログリア細胞」が10-15%を占めると報告されているため、脳全体の「ミクログリア細胞-NA」の比率は0.1とする。これらの前提に基づいて、各脳細胞種の比率は、「マクロファージ-NA」(約0.2%)、「ミクログリア細胞-NA」(10.0%)、「アストロサイト-NA」(約2.2%)、「バークマングリア細胞-NA」(約2.1%)、「脳周皮細胞-NA」(約1.5%)、「内皮細胞-NA」(約6.4%)、「神経細胞 - 興奮性ニューロン、及びいくつかの神経幹細胞」(約47.5%)、「神経細胞 - 抑制性ニューロン」(約21.3%)、「乏突起膠細胞-NA」(約8.7%)、「乏突起膠細胞前駆細胞-NA」(約1.9%)と設定し、これらを脳の参照細胞種の構成比率として用いた。ヒトの心臓と腎臓については、参照細胞種の構成比率として、マウスにおける心臓の細胞種の構成比率と腎臓の細胞種の構成比率を使用した。

【0088】

各器官の参照細胞種の構成比率は、後述する参照細胞種の構成比率リストに示す。

また、各参照細胞種の構成比率リストに示される各細胞種については、公知データベースに細胞種ごとのscRNA-Seqのカウントデータが登録されている。

【0089】

2. データの前処理とRNAカウントの正規化

すべてのデータ処理と分析は、ソフトウェア「R」バージョン3.6.1を使用して実行した。すべての細胞種ラベルは、既報のscRNA-Seq研究において付されたラベルと同様とした。scRNA-Seqデータに付されている遺伝子シンボルは、「org.Mm.eg.db」のRパッケージ内の「org.Mm.egALIAS2EG」に由来するentrez遺伝子IDによって各器官全体RNA-Seqデータに対応付け変換された。ERCCラベルが付された遺伝子はスパイクイン遺伝子であるため削除した。さらに、Rn45s、Akap5、Lrrc17の3つの遺伝子に由来するRNAのカウントも、合計カウントに大きく影響する非mRNAアーチファクトであるため削除した。次に、各遺伝子に由来するRNAカウントを、scRNA-Seqデータセット内の各細胞の総カウントが100である場合に換算し、正規化した。この正規化の工程は、器官全体RNA-Seqデータセットに含まれる各RNAについても同様に行った。

【0090】

3. 細胞種の識別のためのシグネチャ遺伝子セットの選択

ランダムフォレスト(RF)を使用し、前のセッションで説明した参照細胞種の構成比率データセットとscRNA-Seqデータを用いて、各細胞種のシグネチャ遺伝子をコンピュータで選択した。この選択ではRFによる分類器のチューニングと作成にRの「randomForest」パッケージを使用した。scRNA-Seqデータを最初に2分割し、一方をRFによる分類器を作成するためのトレーニングデータとし、もう一方を分類器の精度を検証するためのF1スコア計算用のテストデータとした。前のセッションで説明した細胞種の構成比率を維持したデータセットでRF分析を行った。分類器の生成に続いて、分類器の重要な特徴量を各細胞種のシグネチャ遺伝子名として抽出し、“Mean Decrease Gini”値を各遺伝子の重要度指標として使用した。

【0091】

4. データベース

本実施例では、すべて公開データセットを使用した。マウスの器官全体RNA-Seqデータ

、及び心筋梗塞モデルマウスの器官全体RNA-Seqは「i-organs.atr.jp」から取得した。ヒト器官全体RNA-Seqデータは「The Human Protein Atlas」(<https://www.proteinatlas.org/>;心臓 (ERR315328) および腎臓 (ERR315494)) から取得した。scRNA-Seqデータは、非特許文献5 (大動脈、脳、脂肪、心臓、腎臓、大腸、肝臓、肺、骨髄、膵臓、皮膚、脾臓、胸腺)、「マウス細胞アトラス」の骨格筋から取得した。

【0092】

5. 推定遺伝子発現変動解析

予測計算したすべての遺伝子のRNAカウントの合計を100万コピーに正規化した。各遺伝子の正規化されたカウントを整数に四捨五入し、Rパッケージ「DESeq2 (バージョン1.24.0)」を使用して解析した。

【0093】

II. 既報の器官全体RNA-Seqデータのデコンボリューション法の性能検証

はじめに、既報の方法であるMuSiC (非特許文献17) およびDWLS法 (非特許文献19) について、それぞれの方法によって算出された細胞種の構成比率と、scRNA-Seqデータ及び過去の報告から得られた参照細胞種の構成比率とのサイド・バイ・サイド比較を行い、各デコンボリューション法の性能を検証した。

【0094】

1. 既報の方法による細胞種の構成比率の算出

MuSiC (非特許文献17) およびDWLS法 (非特許文献19) は、それぞれの文献にしたがった。DWLS法でquadratic problem solver を行う際にはsolve.QP (R package: quadprog) をsolve_osqp (R package: osqp)に置き換えた。

【0095】

2. 結果

MuSiC又はDWLS法によって、コンピュータにより算出された各器官の推定した細胞種の構成比率と上記2.において準備した各器官の参照細胞種の構成比率との比較結果を図5及び図6に示す。MuSiC又はDWLS法によって推定された各器官における推定した細胞種の構成比率と参照細胞種の構成比率は乖離しており、また乖離の程度も様々であった。特に、骨格筋、心臓、膵臓、及び肝臓では乖離が顕著であった。

【0096】

心臓は、心筋細胞と非筋肉細胞で構成されている。心筋細胞は心臓の最も広い体積を占める。しかし、細胞数を比較すると、心筋細胞よりも非筋肉細胞の方が多い。この事実に対して、MuSiC又はDWLS法によって計算された心臓の細胞種の組成は、心筋細胞が90%を占めると算出されている。同じ傾向が、骨格筋にも見られた。

【0097】

このように、参照細胞種の構成比率と推定した細胞種の構成比率が乖離する原因の1つとして、異なる細胞種間でのトータルRNAの含有量の差が考えられた。トータルRNAの含有量は、50,000 transcripts/cellから300,000 transcripts/cellの範囲で、細胞ごとに異なることが報告されている。心臓では、心筋細胞の体積は、内皮細胞や線維芽細胞などの非筋肉細胞の体積の20倍から25倍といわれている。したがって、細胞あたりのトータルRNAの含有量は、筋肉細胞と非筋肉細胞で大きく異なる可能性がある。実際、この可能性は、MuSiC及びDWLS法では考慮されていない。このような点が、参照と推定した細胞種の構成比率の乖離につながったものと考えられる。

【0098】

III. 推定器官全体RNA-Seqデータセットと実際の器官全体RNA-Seqデータセットの乖離原因の検証

参照細胞種の構成比率と推定した細胞種の構成比率の乖離は、器官の全RNAサンプルを抽出した際に器官から採取した組織に含まれている各細胞種間でのトータルRNAの含有量の差が原因であると仮説を立て、実際の遺伝子発現プロファイルと推定遺伝子発現プロファイルを比較することにより、この仮説を検証した。推定器官全体RNA-Seqデータは、1.の1で取得した参照細胞種の構成比率に1.の2.で取得したカウントデータをかけ合わせた

10

20

30

40

50

結果である。

【 0 0 9 9 】

図 7 に推定器官全体RNA-Seqデータを示す。推定器官全体RNA-Seqデータは、複数の細胞種から構成される組織について、既知の参照細胞種の構成比率に基づいて重み付けすることにより正規化した遺伝子ごとのトランスクリプトのカウントの合計として算出した。

【 0 1 0 0 】

図 7 に示される結果は、RFによって算出された各器官の細胞種を識別するために使用した各器官の各細胞種における上位ランク数に応じたシグネチャ遺伝子のindicated number (遺伝子数) である。上位ランク数は、シグネチャ遺伝子の中の100遺伝子、300遺伝子、2000遺伝子とした。但し、大動脈及び腎臓ではシグネチャ遺伝子の総数が2000に満たないため、大動脈は2000遺伝子ではなく1577遺伝子で、腎臓は1461遺伝子と比較した。14の器官の実際の遺伝子発現プロファイルと推定遺伝子発現プロファイル間の類似性 / 非類似性は、ピアソン相関係数によって示す。

10

【 0 1 0 1 】

図 7 に示すように、10個の器官(大動脈、脳、心臓、大腸、肝臓、肺、膵臓、皮膚、骨格筋、胸腺)でピアソン相関係数が0.75未満であった。これらの器官では、器官全体RNA-Seqデータを再構成するには、1.の1で取得した参照細胞種の構成比率に1.の2.で取得したカウントデータをかけ合わせただけでは不十分であることを示している。

【 0 1 0 2 】

IV. データセット間の乖離を解消するための細胞種特異的な重み係数の設定と検証

20

各組織に存在する異なる細胞種におけるRNA含有量を補正するための重み係数を算出し、その精度の検証を行った。

【 0 1 0 3 】

1. 細胞種特異的係数の算出

各器官に存在する細胞種ごとの重み係数は、以下の手法にしたがって算出した。

【数 1 6】

[math. 16]

$$\mathbf{y} \in \mathbb{R}^n, \quad w_j > 0, \quad \text{及び} \quad \mathbf{x}_j \in \mathbb{R}^n$$

30

は、それぞれ正規化された器官全体RNA-Seqカウントのベクトル、各解析対象細胞j重み係数、及び正規化されたscRNA-Seqカウントの行列を示す。ここで、nはそれぞれの器官のシグネチャ遺伝子の遺伝子数を示す。RF解析によって得られた“Mean Decrease Gini”に基づくランキングによって、上位100、300、又は2,000の遺伝子をシグネチャ遺伝子として選択した。RF解析において最大シグネチャ遺伝子数が2000未満の器官については、すべての遺伝子を使用した。加えて、解析対象細胞の組み合わせ C^m は、参照細胞種の構成比率が合計セットサイズm内に保たれるという制限下でランダムに選択した。

【 0 1 0 4 】

40

次に下式(1)について説明する。

【数 1 7】

[math. 17]

$$m\mathbf{y} = \sum_j w_j \mathbf{x}_j \quad (j \in C^m) \quad (1)$$

式(1)においてmは倍率を示し、nに依存して決定される。mは、以下の計算のそれぞれにおいて、n未満に設定される。ここで w_j は、結果の値が0.01以上であるという制限の

50

下で下式(2)にしたがって、二次計画問題を解くことにより算出した。

【数18】

[math. 18]

$$\tilde{w}_j = \operatorname{argmin}_{w_j} \sum_i^S \left| m y_i - \sum_j w_j x_j \right|^2 \quad s.t. \quad w_j \geq 0.01. \quad (2)$$

式(2)において、Sは、器官全体を対象とする各遺伝子のRNA-Seqのカウントデータセットの数を示す。今回の検討では、異なる2個体から取得した対応する器官全体RNA-Seqのカウントデータセットを使用した。このため、Sは2である。この二次計画問題は、Rで「quadprog」パッケージを使用して解いた。解析対象細胞の組み合わせのランダム選択と

10

【数19】

[math. 19]

\tilde{w}_j

20

の算出の両方の工程は w_j の数が100以上になるまで再帰的に行い選択した解析対象細胞すべてについて行った。

次に、重み係数の分布がガウス分布にしたがうと仮定して、各解析対象細胞の遺伝子の重み付きカウントの平均と分散を下式(3)にしたがって算出した。

【数20】

[math. 20]

$$\tilde{x}_j \sim \mathcal{N}(\mu_{w_j} x_j, \sigma_{w_j}^2 \operatorname{diag}(x_j \odot x_j)) \quad (3)$$

30

式(3)において、

【数21】

[math. 21]

\tilde{x}_j 、 μ_{w_j} 、及び $\sigma_{w_j}^2$

40

は、それぞれ解析対象細胞jの遺伝子の重み付きカウントベクトル、解析対象細胞jの重み係数の平均値、及び解析対象細胞jの重み係数の分散を示す。

【数22】

[math. 22]

作用素 \odot

は、2つのベクトル間の要素ごとの積を表す。

50

【 0 1 0 5 】

算出された各解析対象細胞の重み係数の平均と分散に基づいて、対応する細胞種における重み係数の平均と分散がガウス分布にしたがうと仮定して、対応する細胞種における重み係数の平均と分散を下式（ 4 ）にしたがって算出した。

【数 2 3】

[math. 23]

$$\bar{x}_k \sim \mathcal{N}\left(\frac{1}{N_k} \sum_j \mu_{w_j} x_j, \frac{1}{N_k} \sum_j \sigma_{w_j}^2 \text{diag}(x_j \odot x_j)\right) \quad (j \in C^k). \quad (4)$$

10

式（ 4 ）において、 k 、 C^k 、及び N_k は、それぞれ細胞種、細胞種 k にラベルされた解析対象細胞の群、及び C^k における解析対象細胞の数を示す。

上記式（ 2 ）にしたがって、算出された各器官に存在する細胞種ごとの重み係数、その平均値、分散、四分位数を後述する重み係数リストに示す。

【 0 1 0 6 】

2 . 結果

上記計算式（ 2 ）により、各細胞種の重み係数及びその範囲が作成された（図 8）。筋細胞の重み係数は、心臓と骨格筋の両方で実際に非筋細胞種よりも高くなった（図 8）。これらの細胞種特異的な重み係数を使用して、各細胞種のトランスクリプトカウントに重み付けをした。次に、各器官に含まれる各細胞種の参照細胞種の構成比率にしたがって、重み係数により重み付けされたトランスクリプトカウントにさらに各器官における各細胞種の参照細胞種の構成比率を適用し、RNA-Seqデータセットを生成した。この算出方法を推定器官全体RNA-Seq（v-RNA-Seq）と呼び、推定器官全体RNA-Seqで得られたRNA-Seqデータセットを推定器官全体RNA-Seqデータセットと呼ぶ。

20

【 0 1 0 7 】

次に、推定器官全体RNA-Seqデータセットと、対応する実際の器官全体RNA-Seqデータセットとを比較した。その結果を図 9 に示す。図 7 と比較して、ほとんどの器官において各データセットが示す遺伝子発現プロファイルの乖離が改善された（ピアソン相関係数 = 0.8-1.0）。

30

【 0 1 0 8 】

V . 各器官の細胞種の構成比率の計算と各器官に含まれる全細胞種の全RNA発現パターンの推定

上記IV . において算出された細胞種ごとのRNA含有量に基づく特異的重み係数を使用して、ベイズの定理に基づいたアルゴリズムを設計し、各器官に含まれる各細胞種の比率とその各細胞種における遺伝子発現パターンの両方を同時に計算した。

【 0 1 0 9 】

1 . 細胞種の構成比率と遺伝子発現パターンの計算

細胞種の構成比率と遺伝子発現パターンの計算は、下式（ 5 ）にしたがった。各細胞種において重み係数によって重み付けされたトランスクリプトカウントの平均と分散を上記式（ 4 ）にしたがって算出した。

40

【数 2 4】

[math. 24]

$$y = Xr \quad (5)$$

式（ 5 ）において、

50

【数 2 5】

[math. 25]

\mathbf{y} 、 $\mathbf{X} = (\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_k, \dots, \bar{\mathbf{x}}_K)$ 、及び \mathbf{r}

は、それぞれ、器官全体RNA-Seq データベクトル、列を上記式にしたがって計算された各細胞種の重み係数で重み付けしたカウントとする推定scRNA-Seqデータとする行列、及び細胞種の構成比率に対応する係数ベクトルを示す。Xとrの算出には、ベイズの定理を用いた。ベイズの定理を適用するために、

10

【数 2 6】

[math. 26]

ノイズ $\mathcal{N}(\mathbf{0}, \beta^{-1}\mathbf{I})$

を式 (5) に加え、下式 (6) に示す確率モデルを採用した。

【数 2 7】

20

[math. 27]

$$P(\mathbf{y}|\mathbf{X}, \mathbf{r}) = \mathcal{N}(\mathbf{y}|\mathbf{X}\mathbf{r}, \beta^{-1}\mathbf{I}) \quad (6)$$

式 (6) において、 β はハイパーパラメータを示す。ベイズの定理にしたがって、X 及び r の事後分布を下式 (7) として得た。

【数 2 8】

30

[math. 28]

$$P(\mathbf{X}, \mathbf{r}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{X}, \mathbf{r})P(\mathbf{X})P(\mathbf{r}) \quad (7)$$

式 (7) において、 $P(\mathbf{X})$ 及び $P(\mathbf{r})$ は、それぞれ X と r の事前分布を示す。そして、 $P(\mathbf{X})$ 及び $P(\mathbf{r})$ は、下式 (8) 及び (9) として与えられる。

【数 2 9】

40

50

[math. 29]

$$P(\mathbf{X}) = \prod_{k=1}^K P(\bar{\mathbf{x}}_k) = \prod_{k=1}^K \mathcal{N}(\bar{\mathbf{x}}_k | \boldsymbol{\mu}'_k, \boldsymbol{\Sigma}_k^{-1}) \quad (8)$$

$$P(\mathbf{r}) = \mathcal{N}(\mathbf{r} | \mathbf{0}, \alpha^{-1} \mathbf{I}) \quad (9)$$

式 (9) において、 α はハイパーパラメータである。事前分布 $P(\mathbf{X})$ において、 $\mathcal{N}(\bar{\mathbf{x}}_k | \boldsymbol{\mu}'_k, \boldsymbol{\Sigma}_k^{-1})$ は式 (4) と同じである。

$\boldsymbol{\mu}'_k = \frac{1}{N_k} \sum_j \mu_{wj} \mathbf{x}_j$ であり、 $\boldsymbol{\Sigma}_k^{-1} = \frac{1}{N_k} \sum_j \sigma_{wj}^2 \text{diag}(\mathbf{x}_j \odot \mathbf{x}_j)$ である。

事後分布を直接最大化する代わりに、

\mathbf{r} と、 $\bar{\mathbf{x}}_k$ のそれぞれについて、 $P(\mathbf{r} | \mathbf{y}, \mathbf{X})$ と $P(\bar{\mathbf{x}}_k | \mathbf{y}, \mathbf{r}, \{\bar{\mathbf{x}}_l\}_{l \neq k})$ を最大化した $\bar{\mathbf{x}}_k$ を推定する反復法を採用した。

具体的には、 $\bar{\mathbf{x}}_k$ の推定のために、下式 (10) で表される確率の式を採用した。

$$P(\bar{\mathbf{x}}_k | \mathbf{y}, \mathbf{r}, \{\bar{\mathbf{x}}_l\}_{l \neq k}) \propto P(\mathbf{y} | \mathbf{X}, \mathbf{r}) P(\bar{\mathbf{x}}_k) . \quad (10)$$

$P(\bar{\mathbf{x}}_k | \mathbf{y}, \mathbf{r}, \{\bar{\mathbf{x}}_l\}_{l \neq k})$ はガウス分布にしたがい、その平均と分散を下式 (11) 及び式 (12) したがって算出した。

$$\mathbb{E}[\bar{\mathbf{x}}_k] = (\boldsymbol{\beta} \mathbf{r}_k^2 \mathbf{I} + \boldsymbol{\Sigma}_k^{-1})^{-1} \left(\boldsymbol{\beta} \mathbf{r}_k \mathbf{y} + \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}'_k - \boldsymbol{\beta} \sum_{l \neq k} r_k r_l \bar{\mathbf{x}}_l \right) \quad (11)$$

$$\text{Var}[\bar{\mathbf{x}}_k] = \boldsymbol{\beta} \mathbf{r}_k^2 \mathbf{I} + \boldsymbol{\Sigma}_k^{-1} . \quad (12)$$

$\mathbb{E}[\bar{\mathbf{x}}_k]$ の結果において、すべての負の値を「0」に設定した。

\mathbf{r} の推定のために、下式 (13) で表される確率の式を採用した。

$$P(\mathbf{r} | \mathbf{y}, \mathbf{X}) \propto P(\mathbf{y} | \mathbf{X}, \mathbf{r}) P(\mathbf{r}) . \quad (13)$$

そして、 $P(\mathbf{r} | \mathbf{y}, \mathbf{X})$ はガウス分布にしたがい、その平均と分散を下式 (14) したがって算出した。

$$\mathbb{E}[\mathbf{r}] = (\alpha \mathbf{I} + \boldsymbol{\beta} \mathbf{X}^t \mathbf{X})^{-1} \boldsymbol{\beta} \mathbf{X}^t \mathbf{y} . \quad (14)$$

結果の $\mathbb{E}[\mathbf{r}]$ と $\mathbb{E}[\bar{\mathbf{x}}_k]$ の両方について、すべての負の値を「0」に設定した。

遺伝子発現パターン ($\{\bar{\mathbf{x}}_k^*\}_{k=1}^K$) と細胞種の構成比率 (\mathbf{r}^*) を推定するため、 \mathbf{X} と \mathbf{r} の両方が収束するか、そうでない場合には、1001 回の反復が行われるまで交互に繰り返し計算を行った。

【 0 1 1 0 】

最初の \mathbf{r} と \mathbf{X} として、細胞種の構成比率と式 (4) の計算式により重み付けされた参照データセットのカウントを使用した。ハイパーパラメータ α と $\boldsymbol{\beta}$ は、便宜上 $10^{-3}, 10^{-2}, \dots, 10^3$ に設定した。実際の器官全体 RNA-Seq と高い類似性 (高いピアソン相関係数及びスピアマン相関係数を示し、かつ低い平均二乗誤差に基づいて類似性を決定) を生じたシグネチャ遺伝子セット (100, 300, 2,000 / 1,577 / 1,461) ハイパーパラメーター (および) の数の組み合わせの結果を最良の推定結果として選択した。この計算の概要を図 10 に示す。本発明の方法により推定した細胞種の構成比率及び参照細胞種比率との比較結果を図 11 及び図 12 に示す。また、本発明の方法により推定した scRNA-Seq カウントデータの実際の scRNA-Seq との比較結果及び t-Distributed Stochastic Neighbor Embedding (t-SNE) の解析結果を図 13 及び図 14 に示す。

【 0 1 1 1 】

2 . 細胞種識別の検証

上記 V. 1. において算出した推定 scRNA-Seq カウントデータが、各器官に存在する細胞種を識別できているか、t-Distributed Stochastic Neighbor Embedding (t-SNE) を使用して検証した。各器官に存在するそれぞれの細胞種に属する細胞において、合計サンプリングサイズを 3,000 に設定し、各細胞種のサンプリングされたセルの数と、細胞種 k の

推定scRNA-Seqカウントデータを、
【数 3 0】

[math. 30]

それぞれ $P^*(\mathbf{r})$ と $P^*(\bar{\mathbf{x}}_k)$ とした。

合計サンプリングサイズが3,000に達するまで、このサンプリングプロセスを繰り返した。次に、Rパッケージ「Rtsne」を使用して、parameter perplexity = 50として、サンプリングされた推定scRNA-Seqカウントデータにt-SNEを適用した。

10

【 0 1 1 2】

3 . 結果

本発明においては、2つのハイパーパラメータ α と β を定義して、細胞種の比の組み合わせの影響を考慮した。異なる器官レベルの遺伝子発現パターン、例えば正常な器官と病理器官の遺伝子発現パターンは、異なる場合がある。しかし、この違いには、i)各細胞種における遺伝子発現パターンは一見同じままであるにもかかわらず、各細胞種の比率が異なっている場合、ii)細胞種の比率はおなじであるものの、同じ細胞種間で遺伝子発現パターンの違いが生じている場合、が考えられる。また、i)とii)が組み合わせられている可能性もある。そこで、 α と β の広範な範囲の包括的な組み合わせを評価して、器官レベルのトランスクリプトームの挙動を説明するために、細胞種の組成と各細胞タイプの重み付けトランスクリプトームカウントの最適な組み合わせを算出した。

20

【 0 1 1 3】

この方法により、10の器官（大動脈、脂肪、心臓、腎臓、肝臓、肺、大腸、骨髄、骨格筋、脾臓）の細胞種の構成比率を計算した。この結果を図 1 1 に示す。図 5 及び図 6 で使用した14の器官から脳、膵臓、皮膚、胸腺は、次の理由により検討から除外した。1) 実際の細胞種の比率が利用可能でない。2) 膵臓は、実際に膵島に由来する。膵島では実際の細胞種の比率を利用できるが、実際の膵臓全体の比率を示すものではない。3) 皮膚または胸腺の場合、細胞種と特異的重み係数を利用してピアソン相関係数が0.8を超えなかった。

30

【 0 1 1 4】

上記10の器官の計算された細胞種の構成比率は、実際のscRNA-Seq研究によって実験的に決定された参照細胞種の構成比率と類似していた（図 1 1）。特に、MuSiCとDWLS法によって推定される心筋細胞と骨格筋細胞の異常に大きい比率は、それぞれV-scRNA-Seqによって改善された。この結果を図 1 1 に示す。また、図 1 2 に示すように、参照細胞種の構成比率に対する平均二乗誤差（MSE）は、V-scRNA-Seqが5つの実際の臓器（脂肪、心臓、大腸、肝臓、骨格筋）で他の手法を上回った。

【 0 1 1 5】

また、推定器官全体RNA-Seqデータセットに含まれる、骨格筋を除く検討した器官のいずれかで発現していた23,131の遺伝子、及び骨格筋の14,323（骨格筋）の遺伝子について本発明の方法に従って、細胞種特異的重み係数及び参照細胞種の構成比率で補正した推定トランスクリプトカウントを計算し、補正した推定トランスクリプトカウントを実際の10の器官内の各細胞種における遺伝子発現と比較した。

40

【 0 1 1 6】

ピアソン相関係数は、推定トランスクリプトカウントは、すべての細胞種と器官について実際のカウントに匹敵することが示された（図 1 3）。また、異なる器官間で同じもしくは関連する細胞種のアノテーションの類似性と関連性も示された（図 1 3）。

【 0 1 1 7】

10種の器官のすべてのV-scRNASeqデータを用いたt-SNE解析により、各器官のすべてでそれぞれの細胞種が遺伝子発現プロファイルにしたがって分類できることを示していた

50

(図 1 4)。

【 0 1 1 8 】

VI. 疾患における細胞種比と遺伝子発現の変化の計算

次に、我々の方法が疾患プロセスに伴う各細胞種の細胞種比と遺伝子発現の変化を検出できるかどうかを評価した。心血管疾患は、世界最大の死因である ([https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)))。心臓病の間の細胞種構成が経時的に変化することを示す報告が存在する。さらに、心臓は、上述したように、以前に公開されたデコンボリューション法ではなく、本発明に係る方法が細胞種の構成比率とその遺伝子発現パターンの両方を効果的に計算できる器官である。したがって、心筋梗塞 (MI) のマウスモデルに本発明の方法を適用し、本発明に係る方法が、心臓における細胞種の構成比率と、既に知られているMI時の細胞種依存的な遺伝子発現の経時的变化の両方を検出できるか検討した。

10

疾患モデル計算は、はじめに、各ステージ (E、M、L) の正常マウスの同じ参照細胞種の構成比率を使用して、sham心臓の器官全体RNA-Seqデータを使用して重み係数を計算した。次に、sham/MIモデルの心臓全RNA-Seqデータを使用して、上記のように各段階の細胞種の構成比率と遺伝子発現プロファイルを計算した。

【 0 1 1 9 】

図 1 5 にその結果を示す。

心筋梗塞の動物モデルの作製方法は、公知である。心筋梗塞の3つの段階は、以下の通りである。1) 冠動脈結紮後1日 (E-MI、早期心筋梗塞段階)、2) 冠動脈結紮7日後 (M-MI、早期線維症段階)、3) 冠動脈結紮後8週間 (L-MI、心臓リモデリング段階)。この解析では、shamコントロール (E-sham、M-sham、L-sham) のRNA-Seqデータと正常なマウス心臓の参照細胞種の構成比率を使用して、各細胞種の重み係数を計算した。

20

【 0 1 2 0 】

演算により、予想される2つのMIに関連する主要な細胞種構成の変化、具体的には心筋細胞の減少、線維芽細胞の増加が検出された (図 1 5 a)。本方法により、M-MI期に特徴的である筋線維芽細胞の増加が検出された (図 1 5 a)。これは、報告されている実験結果とも一致していた。

【 0 1 2 1 】

本発明によって、算出された心筋梗塞時の各細胞種における遺伝子発現変化は、以前の実験研究から予想される複数の特徴も検出された (図 1 5 b)。

30

【 0 1 2 2 】

心筋細胞では、Nppb遺伝子、Sparc遺伝子、Col4a1遺伝子の統計的に有意な発現上昇 (\log_2 倍の変化 0.7)、及びMyh6遺伝子の発現低下 (\log_2 倍の変化 0.7) が検出された (図 1 5 b)。線維芽細胞では、Col4a1遺伝子、Col1a1遺伝子、及びSparc遺伝子の統計的に有意な発現上昇 (\log_2 倍の変化 0.7) が検出された (図 1 5 b)。ここで、統計学的に有意とは、adjusted p値 0.001を意味する。これらの既知のMI病態におけるランドマーク遺伝子に加えて、本方法により、各細胞種で病態に依存して発現が変動する他の多くの遺伝子が見つかった (図 1 5 b)。

【 0 1 2 3 】

40

VII. 推定ヒトscRNA-Seqの検証

マウスの重み係数とV-scRNA-Seqをヒト器官全体RNA-Seqデータセットのデコンボリューションへ応用できることを検証した。心臓および腎臓の公的に入手可能なヒト器官全体RNA-Seqデータを使用して、それらの細胞種の構成比率とトランスクリプトームプロファイルを計算した。

【 0 1 2 4 】

ヒト器官全体RNA-Seqデータセットに格納されている各遺伝子から発現されているRNAの総カウントのそれぞれを、最初に100に正規化した。次に、マウスとヒトの間で共通な名前の遺伝子を抽出することにより、マウスの遺伝子シンボルをヒトのものに一致させた。これらのマウスとヒトの共通の遺伝子セットを使用して、マウスデータセットについて

50

説明した計算手法を適用した。ヒトの心臓および腎臓の器官全体RNA-Seqデータは、「The Human Protein Atlas」(<https://www.proteinatlas.org/>)から取得した。

【0125】

結果を図8に示す。ヒトの心臓および腎臓について計算された細胞種の構成比率が、対応する正常なマウス器官の細胞種の構成比率と同様であることが示された(図16a)。さらに、ヒトの心臓および腎臓の推定scRNA-Seqデータのt-SNEの解析結果は、各器官における既知の細胞種の遺伝子発現プロファイルに基づく分類が可能であることを示した(図16b)。これらの結果は、細胞種特有の重み係数とV-scRNASeqフレームワークの異種間での適用可能性を示している。

【0126】

参照細胞種の構成比率リスト

下記リストは、Organ:Cell type:Abbreviation:Referenceの順に並んでいる。「;」は各細胞種のデータの区切りを意図している。細胞の構成比率は器官全体が「1」となるように正規化している。ここには、代表的な細胞種を示しているため、各器官における各細胞種の構成比率の合計は必ずしも1にはならない。

Aorta:Aorta-endothelial cell-NA:EC :0.40 ;
 Aorta:Aorta-erythrocyte-NA:ERC:0.21 ;
 Aorta:Aorta-fibroblast-NA:FC :0.22 ;
 Aorta:Aorta-professional antigen presenting cell-NA:PAP:0.16 ;
 Brain:Brain_Myeloid-macrophage-NA:MAC:0.00 ;
 Brain:Brain_Myeloid-microglial cell-NA:MI :0.10 ;
 Brain:Brain_Non-Myeloid-astrocyte-NA:AS :0.02 ;
 Brain:Brain_Non-Myeloid-Bergmann glial cell-NA:BGC:0.00 ;
 Brain:Brain_Non-Myeloid-brain pericyte-NA:BP :0.02 ;
 Brain:Brain_Non-Myeloid-endothelial cell-NA:EC :0.06 ;
 Brain:Brain_Non-Myeloid-neuron-excitatory neurons and some neuronal stem cells:NEUR2 :0.47 ;
 Brain:Brain_Non-Myeloid-neuron-inhibitory neurons:NEUR1 :0.21 ;
 Brain:Brain_Non-Myeloid-oligodendrocyte-NA:OLC:0.09 ;
 Brain:Brain_Non-Myeloid-oligodendrocyte precursor cell-NA:OPC:0.02 ;
 Fat:Fat-B cell-NA:B:0.10 ;
 Fat:Fat-endothelial cell-NA:EC :0.16 ;
 Fat:Fat-mesenchymal stem cell of adipose-mesenchymal progenitor:MSA:0.43 ;
 Fat:Fat-myeloid cell-NA:MYE:0.20 ;
 Fat:Fat-NA-NA:NA :0.01 ;
 Fat:Fat-natural killer cell-NA:NK :0.01 ;
 Fat:Fat-T cell-NA:T:0.08 ;
 Heart:Heart-cardiac muscle cell-NA:CM :0.30 ;
 Heart:Heart-endocardial cell-NA:ECC:0.02 ;
 Heart:Heart-endothelial cell-NA:EC :0.20 ;
 Heart:Heart-fibroblast-NA:FC :0.29 ;
 Heart:Heart-leukocyte-NA:LEU:0.13 ;
 Heart:Heart-myofibroblast cell-NA:MYF:0.05 ;
 Heart:Heart-NA-conduction cells:CC :0.01 ;
 Heart:Heart-smooth muscle cell-NA:SM :0.01 ;
 Kidney:Kidney-endothelial cell-NA:EC :0.19 ;
 Kidney:Kidney-epithelial cell of proximal tubule-NA:PT :0.48 ;
 Kidney:Kidney-kidney collecting duct epithelial cell-NA:CD :0.22 ;
 Kidney:Kidney-leukocyte-NA:LEU:0.02 ;
 Kidney:Kidney-macrophage-NA:MAC:0.09 ;

10

20

30

40

50

Large Intestine:Large Intestine-Brush cell of epithelium proper of large intestine-Tuft cell:TUF:0.01 ;
 Large Intestine:Large Intestine-enterocyte of epithelium of large intestine-Enterocyte (Distal):EN-D :0.06 ;
 Large Intestine:Large Intestine-enterocyte of epithelium of large intestine-Enterocyte (Proximal):EN-P :0.21 ;
 Large Intestine:Large Intestine-enteroendocrine cell-Chromaffin Cell:CHR:0.01 ;
 Large Intestine:Large Intestine-epithelial cell of large intestine-Lgr5- amplifying undifferentiated cell:EP1:0.16 ;
 Large Intestine:Large Intestine-epithelial cell of large intestine-Lgr5- undifferentiated cell:EP2:0.10 ; 10
 Large Intestine:Large Intestine-epithelial cell of large intestine-Lgr5+ amplifying undifferentiated cell (Distal):EP3-D:0.03 ;
 Large Intestine:Large Intestine-epithelial cell of large intestine-Lgr5+ amplifying undifferentiated cell (Proximal):EP3-P:0.05 ;
 Large Intestine:Large Intestine-epithelial cell of large intestine-Lgr5+ undifferentiated cell (Distal):EP4-D:0.08 ;
 Large Intestine:Large Intestine-epithelial cell of large intestine-Lgr5+ undifferentiated cell (Proximal):EP4-P:0.12 ;
 Large Intestine:Large Intestine-large intestine goblet cell-Goblet cell (Distal):GB1-D:0.09 ; 20
 Large Intestine:Large Intestine-large intestine goblet cell-Goblet cell (Proximal):GB1-P:0.05 ;
 Large Intestine:Large Intestine-large intestine goblet cell-Goblet cell, top of crypt (Distal):GB2-D:0.02 ;
 Liver:Liver-B cell-NA:B:0.07 ;
 Liver:Liver-endothelial cell of hepatic sinusoid-NA:EC :0.33 ;
 Liver:Liver-hepatocyte-NA:HE :0.42 ;
 Liver:Liver-Kupffer cell-NA:KUP:0.11 ;
 Liver:Liver-natural killer cell-NK/NKT cells:NK2:0.07 ; 30
 Lung:Lung-B cell-NA:B:0.02 ;
 Lung:Lung-ciliated columnar cell of tracheobronchial tree-multiciliated cells:CCC:0.01 ;
 Lung:Lung-classical monocyte-invading monocytes:CMN:0.07 ;
 Lung:Lung-epithelial cell of lung-alveolar epithelial type 1 cells, alveolar epithelial type 2 cells, club cells, and basal cells:EP5:0.06 ;
 Lung:Lung-leukocyte-mast cells and unknown immune cells:LEU2 :0.02 ;
 Lung:Lung-lung endothelial cell-NA:EC :0.34 ;
 Lung:Lung-monocyte-circulating monocytes:MN2:0.07 ;
 Lung:Lung-myeloid cell-dendritic cells, alveolar macrophages, and interstitial macrophages:MYE2 :0.01 ; 40
 Lung:Lung-NA-lung neuroendocrine cells and unknown cells:NC :0.03 ;
 Lung:Lung-natural killer cell-NA:NK :0.02 ;
 Lung:Lung-stromal cell-NA:SC :0.33 ;
 Lung:Lung-T cell-NA:T:0.03 ;
 Marrow:Marrow-B cell-Cd3e+ Klrb1+ B cell:B2 :0.01 ;
 Marrow:Marrow-basophil-NA:BAS:0.00 ;
 Marrow:Marrow-common lymphoid progenitor-NA:CLP:0.04 ;
 Marrow:Marrow-granulocyte-NA:GRA:0.16 ;
 Marrow:Marrow-granulocyte monocyte progenitor cell-NA:GMP:0.02 ; 50

Marrow:Marrow-granulocytopoietic cell-NA:GC :0.05 ;
 Marrow:Marrow-hematopoietic precursor cell-NA:HPC:0.08 ;
 Marrow:Marrow-immature B cell-NA:IB :0.06 ;
 Marrow:Marrow-immature natural killer cell-NA:INK:0.01 ;
 Marrow:Marrow-immature NK T cell-NA:INKT :0.01 ;
 Marrow:Marrow-immature T cell-NA:IT :0.02 ;
 Marrow:Marrow-late pro-B cell-Dntt- late pro-B cell:LPB1 :0.04 ;
 Marrow:Marrow-late pro-B cell-Dntt+ late pro-B cell:LPB2 :0.03 ;
 Marrow:Marrow-macrophage-NA:MAC:0.03 ;
 Marrow:Marrow-mature natural killer cell-NA:MNT:0.01 ; 10
 Marrow:Marrow-megakaryocyte-erythroid progenitor cell-NA:EPC:0.01 ;
 Marrow:Marrow-monocyte-NA:MN :0.04 ;
 Marrow:Marrow-naive B cell-NA:NBC:0.12 ;
 Marrow:Marrow-pre-natural killer cell-NA:PNK:0.00 ;
 Marrow:Marrow-precursor B cell-pre-B cell (Philadelphia nomenclature):PB :0.11 ;
 Marrow:Marrow-regulatory T cell-NA:RT :0.00 ;
 Marrow:Marrow-Slamf1-negative multipotent progenitor cell-NA:MPC1 :0.10 ;
 Marrow:Marrow-Slamf1-positive multipotent progenitor cell-NA:MPC2 :0.04 ;
 SkMuscle:B cell_Jchain high(Muscle):B3 :0.02 ; 20
 SkMuscle:B cell_Vpreb3 high(Muscle):B4 :0.09 ;
 SkMuscle:Dendritic cell(Muscle):DEN:0.01 ;
 SkMuscle:Endothelial cell(Muscle):EC :0.02 ;
 SkMuscle:Erythroblast_Car1 high(Muscle):ERB1 :0.03 ;
 SkMuscle:Erythroblast_Car2 high(Muscle):ERB2 :0.16 ;
 SkMuscle:Granulocyte monocyte progenitor cell(Muscle):GMP:0.08 ;
 SkMuscle:Macrophage_Ms4a6c high(Muscle):MAC2 :0.13 ;
 SkMuscle:Macrophage_Retnla high(Muscle):MAC3 :0.02 ;
 SkMuscle:Muscle cell_Tnnc1 high(Muscle):MC1:0.01 ;
 SkMuscle:Muscle cell_Tnnc2 high(Muscle):MC2:0.03 ; 30
 SkMuscle:Muscle progenitor cell(Muscle):MPC:0.08 ;
 SkMuscle:Neutrophil_Camp high(Muscle):NEUT1 :0.16 ;
 SkMuscle:Neutrophil_Prg2 high(Muscle):NEUT2 :0.01 ;
 SkMuscle:Neutrophil_Retnlg high(Muscle):NEUT3 :0.12 ;
 SkMuscle:Stromal cell(Muscle):SC :0.02 ;
 SkMuscle:T cell(Muscle):T:0.01 ;
 Pancreas:Pancreas-endothelial cell-NA:EC :0.06 ;
 Pancreas:Pancreas-leukocyte-NA:LEU:0.04 ;
 Pancreas:Pancreas-pancreatic A cell-pancreatic A cell:A:0.24 ;
 Pancreas:Pancreas-pancreatic acinar cell-acinar cell:ACI:0.10 ; 40
 Pancreas:Pancreas-pancreatic D cell-pancreatic D cell:D:0.11 ;
 Pancreas:Pancreas-pancreatic ductal cell-ductal cell:DUC:0.12 ;
 Pancreas:Pancreas-pancreatic PP cell-pancreatic PP cell:PP :0.05 ;
 Pancreas:Pancreas-pancreatic stellate cell-stellate cell:PSC:0.04 ;
 Pancreas:Pancreas-type B pancreatic cell-beta cell:BC :0.22 ;
 Skin:Skin-basal cell of epidermis-Basal IFE:BE:0.22 ;
 Skin:Skin-epidermal cell-Intermediate IFE:EPI :0.12 ;
 Skin:Skin-keratinocyte stem cell-Inner Bulge:KSC:0.26 ;
 Skin:Skin-keratinocyte stem cell-Outer Bulge:KSC2:0.37 ;
 Skin:Skin-leukocyte-NA:LEU:0.01 ; 50

Skin:Skin-stem cell of epidermis-Replicating Basal IFE:SCE:0.02 ;
 Spleen:Spleen-B cell-NA:B:0.77 ;
 Spleen:Spleen-macrophage-NA:MAC:0.03 ;
 Spleen:Spleen-T cell-NA:T:0.20 ;
 Thymus:Thymus-DN1 thymic pro-T cell-DN1 thymocytes:TPT:0.01 ;
 Thymus:Thymus-immature T cell-DN4-DP in transition Cd69 negative rapidly
 dividing thymocytes:IT3:0.15 ;
 Thymus:Thymus-immature T cell-DN4-DP in transition Cd69 negative thymocy
 tes:IT2:0.44 ;
 Thymus:Thymus-immature T cell-DN4-DP in transition Cd69 positive thymocy
 tes:IT4:0.37 ; 10
 Thymus:Thymus-leukocyte-antigen presenting cell:LEU3:0.02

【 0 1 2 7 】

参照細胞種の構成比率リスト

下記リストは、Organ:Singnature.gene.set.number:Cell.type:mean:var:min:first_
 quantile:Median:third_quantile:maxの順に並んでいる。「 ; 」は各細胞種のデータの
 区切りを意図している。

Aorta:100:EC :0.151788089:0.320824524:0.01:0.01:0.01:0.020643025:4.333
 799883;
 Aorta:100:ERC:24.67386955:27569.49096:0.01:0.268057658:0.947630617:4 20
 .647035248:1361.854647;
 Aorta:100:FC :1.120387302:7.507603394:0.01:0.014564004:0.061908957:0.
 516121627:12.41886869;
 Aorta:100:PAP:0.124841086:0.130858196:0.01:0.01:0.01211706:0.0326243
 13:1.661400716;
 Aorta:300:EC :0.335653916:0.888982181:0.01:0.01:0.01:0.10214728:5.7845
 60681;
 Aorta:300:ERC:20.7856725:15487.87723:0.01:0.132647672:0.943529224:3.
 121387266:1008.048707;
 Aorta:300:FC :1.122992247:3.637888239:0.01:0.103255397:0.318573613:1. 30
 121704441:9.938671176;
 Aorta:300:PAP:0.16699144:0.283733571:0.01:0.01:0.014927526:0.0856291
 13:3.587288911;
 Aorta:1577:EC :0.328052831:2.980101375:0.01:0.01:0.01:0.01:12.65695793;
 Aorta:1577:ERC:0.861381942:10.22004111:0.01:0.01:0.01:0.189752138:24.
 51986069;
 Aorta:1577:FC :1.157779979:11.62919843:0.01:0.01:0.011244537:0.227616
 723:17.14421124;
 Aorta:1577:PAP:0.236296791:0.806728782:0.01:0.01:0.01:0.01:4.64459890 40
 8;
 Brain:100:AS :0.434856565:2.615451954:0.01:0.022455987:0.046046996:0.
 12257169:18.38001032;
 Brain:100:BGC:1.072299836:4.247525505:0.052380374:0.193697173:0.441
 824763:0.815423538:9.813487019;
 Brain:100:BP :1.286448516:7.416001166:0.01:0.149224444:0.442108329:1.
 349882731:20.55492602;
 Brain:100:EC :0.155829289:2.482545762:0.01:0.011392245:0.018610335:0.
 058316551:33.77084044;
 Brain:100:MAC:0.048124031:0.004108694:0.01:0.013496606:0.023131426:
 0.063089991:0.377400612; 50

Brain:100:MI :0.012961869:2.3118E-05:0.01:0.010033301:0.011173433:0.013679577:0.064623845;

Brain:100:NEUR1 :2.653825355:114.4922959:0.01:0.014365065:0.065192324:0.488780763:74.1873715;

Brain:100:NEUR2 :1.516349011:26.86180756:0.01:0.017777754:0.088547977:0.515962167:38.7956663;

Brain:100:OLC:1.47033239:789.3444746:0.01:0.043020985:0.137535878:0.575570849:1014.927739;

Brain:100:OPC:1.384613588:14.07919044:0.01:0.012646413:0.040728358:0.702904737:25.50241672;

10

Brain:300:AS :1.74038249:13.24520375:0.01:0.015261047:0.195577225:1.884738702:41.41426285;

Brain:300:BGC:1.341778577:3.944690072:0.045824361:0.224444365:0.549709956:1.458466949:8.815660456;

Brain:300:BP :1.749492174:12.04296246:0.01:0.014499893:0.199956059:1.774169825:17.37539007;

Brain:300:EC :0.205803506:0.364383999:0.01:0.01:0.010304075:0.054289482:7.530690763;

Brain:300:MAC:0.010347881:4.91595E-06:0.01:0.01:0.01:0.01:0.024372891;

Brain:300:MI :0.010016051:1.25293E-07:0.01:0.01:0.01:0.01:0.02439228;

20

Brain:300:NEUR1 :0.091026428:0.378856142:0.01:0.01:0.01:0.01:5.354404503;

Brain:300:NEUR2 :1.881598235:117.9200715:0.01:0.01:0.01:0.010376092:122.6759504;

Brain:300:OLC:0.683957014:4.665791152:0.01:0.01:0.031833092:0.255883967:30.43686688;

Brain:300:OPC:0.371699481:4.009370499:0.01:0.01:0.01:0.022522107:22.58268291;

Brain:2000:AS :1.591406611:8.974491561:0.01:0.01:0.070425016:1.748463704:15.85464919;

30

Brain:2000:BGC:1.125268038:5.864155062:0.01:0.01:0.013973665:0.729619368:9.858942038;

Brain:2000:BP :0.010231754:6.39702E-06:0.01:0.01:0.01:0.01:0.03860368;

Brain:2000:EC :0.046223928:0.054173766:0.01:0.01:0.01:0.01:2.487765951;

Brain:2000:MAC:0.010208662:4.75753E-07:0.01:0.01:0.01:0.01:0.013557299;

Brain:2000:MI :0.01008294:2.25461E-06:0.01:0.01:0.01:0.01:0.062541367;

Brain:2000:NEUR1 :0.044018483:0.043085508:0.01:0.01:0.01:0.01:1.347998842;

Brain:2000:NEUR2 :1.008760807:29.81632686:0.01:0.01:0.01:0.01:51.96076429;

40

Brain:2000:OLC:0.55101755:2.370980823:0.01:0.01:0.01:0.153605575:17.776641;

Brain:2000:OPC:0.014554191:0.001657313:0.01:0.01:0.01:0.01:0.498116609;

Fat:100:B:0.046754003:0.011173744:0.01:0.010216413:0.015136787:0.037396414:1.302102477;

Fat:100:EC :2.414604433:19.41475516:0.01:0.010548113:0.280908337:2.789098145:41.78589451;

Fat:100:MSA:0.320323969:1.112225282:0.01:0.01:0.01:0.04529986:10.4911

50

1527;
 Fat:100:MYE:0.071228559:0.045200484:0.01:0.01:0.014329697:0.0436358
 57:3.146909307;
 Fat:100:NA :2.710777335:28.70862922:0.01:0.01:0.285071051:3.16731646
 8:27.15404948;
 Fat:100:NK :0.318498521:0.146633229:0.013489934:0.060943058:0.20101
 4136:0.368195143:1.823855366;
 Fat:100:T:0.45552476:2.077988331:0.01:0.01:0.029935117:0.292900664:1
 6.31685769;
 Fat:300:B:0.085593432:0.063367131:0.01:0.01:0.01:0.010472349:1.660266 10
 129;
 Fat:300:EC :1.951699824:32.94181255:0.01:0.01:0.01:0.449707234:46.603
 58358;
 Fat:300:MSA:0.301040647:2.273597383:0.01:0.01:0.01:0.01:20.45002588;
 Fat:300:MYE:0.145061513:0.432465824:0.01:0.01:0.01:0.011508338:8.063
 61426;
 Fat:300:NA :0.097562096:0.200890568:0.01:0.01:0.01:0.01:2.807266835;
 Fat:300:NK :0.01:8.01339E-29:0.01:0.01:0.01:0.01:0.01;
 Fat:300:T:0.010988287:0.000224172:0.01:0.01:0.01:0.01:0.237562315;
 Fat:2000:B:0.267000961:1.17913974:0.01:0.01:0.01:0.012349464:10.30491 20
 319;
 Fat:2000:EC :1.52498714:15.89721155:0.01:0.01:0.01:0.367139836:29.441
 72018;
 Fat:2000:MSA:0.352538559:3.267249585:0.01:0.01:0.01:0.01:26.45242703;
 Fat:2000:MYE:0.154984661:0.568213675:0.01:0.01:0.01:0.010339534:12.6
 5610984;
 Fat:2000:NA :0.038760219:0.01129055:0.01:0.01:0.01:0.01:0.520072065;
 Fat:2000:NK :0.011812948:8.05709E-05:0.01:0.01:0.01:0.01:0.062738677;
 Fat:2000:T:0.050755542:0.090293712:0.01:0.01:0.01:0.01:4.357269635;
 Heart:100:CC :0.01:4.19568E-32:0.01:0.01:0.01:0.01:0.01; 30
 Heart:100:CM :2.823889258:54.45296231:0.01:0.01:0.077634074:0.644519
 204:37.63426786;
 Heart:100:EC :0.383886505:8.95708179:0.01:0.01:0.0145107:0.127136156:
 60.17501651;
 Heart:100:ECC:0.142329405:0.121205364:0.01:0.012507964:0.028868296:0
 .164640401:2.344871042;
 Heart:100:FC :0.089502111:0.035613316:0.01:0.01:0.013279019:0.057356
 411:1.287831061;
 Heart:100:LEU:0.049459847:0.008453581:0.01:0.011171809:0.016984044:
 0.053998062:1.156060563; 40
 Heart:100:MYF:0.197874897:0.106739298:0.01:0.01282829:0.041756576:0.
 227034214:1.680778788;
 Heart:100:SM :1.348022055:1.282516297:0.297989562:0.449770835:0.871
 973921:1.854237253:4.078581375;
 Heart:300:CC :0.01:2.41086E-32:0.01:0.01:0.01:0.01:0.01;
 Heart:300:CM :1.706492592:22.04625552:0.01:0.01:0.014028892:0.385791
 139:22.6212958;
 Heart:300:EC :0.308624254:0.996382718:0.01:0.01:0.01441405:0.2280464
 96:14.13284018;
 Heart:300:ECC:0.075306832:0.026880368:0.01:0.01:0.018232343:0.066596 50

459:1.01598398;
Heart:300:FC :0.080447546:0.034630326:0.01:0.01:0.012546049:0.047817
12:1.673015079;
Heart:300:LEU:0.024387404:0.006626406:0.01:0.01:0.011895671:0.021707
587:1.457680288;
Heart:300:MYF:0.292706873:0.433907162:0.01:0.011944381:0.075827223:
0.295975505:6.11236449;
Heart:300:SM :0.123010953:0.023413091:0.01:0.01:0.056597312:0.185566
209:0.509664676;
Heart:2000:CC :0.010313811:7.39796E-07:0.01:0.01:0.01:0.01:0.01312443;
Heart:2000:CM :3.112064232:77.80962808:0.01:0.01:0.01:0.390638194:51.
1228028;
Heart:2000:EC :0.13368653:0.143577296:0.01:0.01:0.01:0.040965266:2.82
9282942;
Heart:2000:ECC:0.310990008:4.392293003:0.01:0.01:0.01:0.01:14.6819710
8;
Heart:2000:FC :0.053694616:0.018926396:0.01:0.01:0.01:0.02231088:1.64
8764347;
Heart:2000:LEU:0.039521164:0.287158212:0.01:0.01:0.01:0.01:9.75938906
8;
Heart:2000:MYF:0.017612063:0.002417966:0.01:0.01:0.01:0.01:0.42226639
1;
Heart:2000:SM :0.077480204:0.044825118:0.01:0.01:0.01:0.01:0.817152076;
Kidney:100:CD :0.861873721:5.325131329:0.01:0.01:0.027677075:0.42964
8536:15.29691009;
Kidney:100:EC :0.087685384:0.053268255:0.01:0.01:0.010467171:0.07961
2624:1.530825402;
Kidney:100:LEU:0.019678869:0.000210734:0.01:0.01:0.01:0.029438172:0.0
48074249;
Kidney:100:MAC:0.014345637:0.00027765:0.01:0.01:0.01:0.01:0.096159876;
Kidney:100:PT :1.066404822:6.559136387:0.01:0.01:0.042416515:0.65650
8148:14.31773311;
Kidney:300:CD :1.02990839:31.96503195:0.01:0.01:0.01:0.01:46.00176013;
Kidney:300:EC :0.148567471:0.401009826:0.01:0.01:0.01:0.01:3.505286789;
Kidney:300:LEU:0.01:3.0112E-29:0.01:0.01:0.01:0.01:0.01;
Kidney:300:MAC:0.010736962:1.52072E-05:0.01:0.01:0.01:0.01:0.03063493
1;
Kidney:300:PT :1.078855015:17.30970861:0.01:0.01:0.01:0.01:34.86889454;
Kidney:1461:CD :0.571606592:3.043433891:0.01:0.01:0.01:0.082063639:10
.68411549;
Kidney:1461:EC :0.262442826:0.506765362:0.01:0.01:0.01:0.040553741:4.
187360011;
Kidney:1461:LEU:0.448563929:1.140902526:0.01:0.01:0.025451239:0.1317
69898:3.072526545;
Kidney:1461:MAC:0.015812234:0.000494913:0.01:0.01:0.01:0.01:0.1261889
15;
Kidney:1461:PT :1.009683719:9.1880931:0.01:0.01:0.010200921:0.349203
618:18.39498163;
Large Intestine:100:CHR:0.012506713:9.45398E-05:0.01:0.01:0.01:0.01:0.05
9484444;

10

20

30

40

50

Large Intestine:100:EN-D :0.02890134:0.011490302:0.01:0.01:0.01:0.01:0.9
14454533;

Large Intestine:100:EN-P :0.245917773:0.409541469:0.01:0.01:0.01:0.0992
77919:5.442989504;

Large Intestine:100:EP1:0.010344637:3.46044E-05:0.01:0.01:0.01:0.01:0.11
3368671;

Large Intestine:100:EP2:0.108748193:1.93259855:0.01:0.01:0.01:0.01:19.76
858819;

Large Intestine:100:EP3-D:0.589057488:2.322023876:0.01:0.01:0.03608866
5:0.212778459:7.753674897;

Large Intestine:100:EP3-P:1.359012809:13.83295307:0.01:0.01:0.03810676
9:0.79029694:30.18008581;

Large Intestine:100:EP4-D:1.334742196:8.776057103:0.01:0.01:0.05175417
7:0.765595964:15.7202124;

Large Intestine:100:EP4-P:1.936669446:21.3101594:0.01:0.01:0.048946915
:0.996926491:29.43809417;

Large Intestine:100:GB1-D:0.567448195:1.714325103:0.01:0.01:0.04047053
4:0.529872099:10.85628586;

Large Intestine:100:GB1-P:0.206993092:0.398629305:0.01:0.01:0.01:0.0359
04714:4.782087384;

Large Intestine:100:GB2-D:0.020528901:0.001429011:0.01:0.01:0.01:0.01:0.
171902125;

Large Intestine:100:TUF:0.01:4.6781E-30:0.01:0.01:0.01:0.01:0.01;

Large Intestine:300:CHR:0.01:8.55585E-28:0.01:0.01:0.01:0.01:0.01;

Large Intestine:300:EN-D :0.028139446:0.0207549:0.01:0.01:0.01:0.01:1.42
3039804;

Large Intestine:300:EN-P :0.225660499:0.725122737:0.01:0.01:0.01:0.01:9.
328144914;

Large Intestine:300:EP1:0.124803964:1.666430783:0.01:0.01:0.01:0.01:21.9
8275274;

Large Intestine:300:EP2:0.136387448:1.286467877:0.01:0.01:0.01:0.01:12.0
6235413;

Large Intestine:300:EP3-D:0.100450079:0.080406616:0.01:0.01:0.01:0.0312
50293:1.703547206;

Large Intestine:300:EP3-P:2.013779797:104.0491511:0.01:0.01:0.01:0.0102
03473:93.19098724;

Large Intestine:300:EP4-D:0.618244402:4.172513336:0.01:0.01:0.01:0.0698
21296:14.62455624;

Large Intestine:300:EP4-P:2.216779474:81.90841528:0.01:0.01:0.01:0.0894
17422:72.73878781;

Large Intestine:300:GB1-D:0.625078047:4.444453549:0.01:0.01:0.01:0.1439
32698:20.8490123;

Large Intestine:300:GB1-P:0.136413664:0.845071847:0.01:0.01:0.01:0.01:9.
557555803;

Large Intestine:300:GB2-D:0.584599971:1.070237117:0.01:0.01:0.08999655
4:0.651617884:4.481089012;

Large Intestine:300:TUF:0.01:3.04504E-29:0.01:0.01:0.01:0.01:0.01;

Large Intestine:2000:CHR:0.015394585:0.000704757:0.01:0.01:0.01:0.01:0.
153127464;

Large Intestine:2000:EN-D :0.058661679:0.087600473:0.01:0.01:0.01:0.01:2

10

20

30

40

50

.68276365;
 Large Intestine:2000:EN-P :0.186182828:0.342939218:0.01:0.01:0.01:0.018
 775138:4.579226957;
 Large Intestine:2000:EP1:0.242031163:1.52035394:0.01:0.01:0.01:0.01:13.5
 4761958;
 Large Intestine:2000:EP2:0.137616207:0.684872242:0.01:0.01:0.01:0.01:7.3
 42157204;
 Large Intestine:2000:EP3-D:0.248573596:1.254007981:0.01:0.01:0.01:0.043
 345374:8.244720325;
 Large Intestine:2000:EP3-P:1.160239961:8.681215788:0.01:0.01:0.01:0.196 10
 549515:14.92215724;
 Large Intestine:2000:EP4-D:1.024156717:12.94504348:0.01:0.01:0.01:0.153
 132844:28.16052677;
 Large Intestine:2000:EP4-P:1.870495516:32.95764143:0.01:0.01:0.01:0.241
 904836:37.58578581;
 Large Intestine:2000:GB1-D:0.703346962:4.343873838:0.01:0.01:0.0172591
 45:0.246563149:15.28585118;
 Large Intestine:2000:GB1-P:0.267045238:3.849090953:0.01:0.01:0.01:0.01:
 20.24231902;
 Large Intestine:2000:GB2-D:0.602863068:0.645603881:0.01:0.02646894:0.2 20
 32151648:0.877725989:3.034948802;
 Large Intestine:2000:TUF:0.01:6.86261E-32:0.01:0.01:0.01:0.01:0.01;
 Liver:100:B:0.213797267:0.325687856:0.01:0.01:0.01:0.056571357:2.6133
 70426;
 Liver:100:EC :0.037007003:0.043483818:0.01:0.01:0.01:0.01:2.711247341;
 Liver:100:HE :1.577528039:35.04314183:0.01:0.01:0.01:0.183535291:43.73
 524356;
 Liver:100:KUP:0.509042737:6.621034858:0.01:0.01:0.01:0.012229472:18.8
 2181524;
 Liver:100:NK2:0.539076723:10.69009305:0.01:0.01:0.01:0.01:20.43361731; 30
 Liver:300:B:0.011314986:7.08967E-05:0.01:0.01:0.01:0.01:0.063914422;
 Liver:300:EC :0.113993526:0.733401923:0.01:0.01:0.01:0.01:10.81729271;
 Liver:300:HE :1.561492365:98.12025078:0.01:0.01:0.01:0.01:105.9140195;
 Liver:300:KUP:0.260603997:2.328280291:0.01:0.01:0.01:0.01:11.82954434;
 Liver:300:NK2:0.125984933:0.524647684:0.01:0.01:0.01:0.01:4.533412393;
 Liver:2000:B:0.046524022:0.014063618:0.01:0.01:0.01:0.01:0.629735622;
 Liver:2000:EC :0.155958403:0.528364338:0.01:0.01:0.01:0.01:8.591024824;
 Liver:2000:HE :1.498511944:62.89746473:0.01:0.01:0.01:0.01:83.09827976;
 Liver:2000:KUP:0.255600958:1.055424386:0.01:0.01:0.01:0.04115651:7.78
 9532314; 40
 Liver:2000:NK2:0.367333689:2.537527345:0.01:0.01:0.01:0.01:9.73738342
 9;
 Lung:100:B:0.01857141:0.001028567:0.01:0.01:0.01:0.01:0.129999746;
 Lung:100:CCC:1.456821314:8.825647444:0.01:0.01:0.01:1.547008977:9.74
 6616082;
 Lung:100:CMN:0.028888155:0.013039084:0.01:0.01:0.01:0.010769856:0.83
 5026957;
 Lung:100:EC :0.592266493:3.130564936:0.01:0.01:0.01:0.14637682:17.394
 6568;
 Lung:100:EP5:2.184258795:31.36444499:0.01:0.142788:0.468145799:1.199 50

237162:33.16037781;
 Lung:100:LEU2 :0.01:2.14937E-25:0.01:0.01:0.01:0.01:0.01;
 Lung:100:MN2:0.015465605:0.000828191:0.01:0.01:0.01:0.01:0.194112768;
 Lung:100:MYE2 :0.01115847:5.79673E-06:0.01:0.01:0.01:0.010703344:0.01
 6013027;
 Lung:100:NC :5.849657994:83.74798893:0.01:0.01:0.391509729:8.6635406
 96:29.4471241;
 Lung:100:NK :0.01:6.00811E-27:0.01:0.01:0.01:0.01:0.01;
 Lung:100:SC :0.596743132:2.135538046:0.01:0.01:0.025279551:0.3259755
 57:10.78671528;
 Lung:100:T:0.01:5.256E-27:0.01:0.01:0.01:0.01:0.01;
 Lung:300:B:0.044904601:0.017056636:0.01:0.01:0.01:0.01:0.498664407;
 Lung:300:CCC:1.891546254:5.51406788:0.01:0.056520696:0.298394181:3.
 445486709:6.056462974;
 Lung:300:CMN:0.01072913:1.36084E-05:0.01:0.01:0.01:0.01:0.032120838;
 Lung:300:EC :0.716551065:10.93183795:0.01:0.01:0.01:0.025238049:41.15
 891979;
 Lung:300:EP5:2.367899563:40.49001777:0.01:0.01:0.075846939:1.7243697
 1:34.37901999;
 Lung:300:LEU2 :0.01:3.0477E-29:0.01:0.01:0.01:0.01:0.01;
 Lung:300:MN2:0.011887933:9.99892E-05:0.01:0.01:0.01:0.01:0.080417127;
 Lung:300:MYE2 :0.01:4.19293E-28:0.01:0.01:0.01:0.01:0.01;
 Lung:300:NC :3.512526692:52.14415724:0.01:0.01:0.166310027:3.2610634
 33:26.50611444;
 Lung:300:NK :0.01:8.33236E-28:0.01:0.01:0.01:0.01:0.01;
 Lung:300:SC :0.453653037:1.848672295:0.01:0.01:0.01:0.044123625:8.947
 055684;
 Lung:300:T:0.011154861:3.06752E-05:0.01:0.01:0.01:0.01:0.036561792;
 Lung:2000:B:0.257379098:0.134236035:0.01:0.011041603:0.089877832:0.
 287607349:1.094854368;
 Lung:2000:CCC:3.063528155:12.0082984:0.01:0.061665312:1.300096759:5
 .627807116:9.698705171;
 Lung:2000:CMN:0.019441605:0.000835036:0.01:0.01:0.01:0.011049543:0.1
 85526719;
 Lung:2000:EC :0.690009542:4.988718328:0.01:0.01:0.01:0.165162893:19.5
 1617638;
 Lung:2000:EP5:1.706230301:7.732375989:0.01:0.065375787:0.461915327:
 1.751011363:10.1566163;
 Lung:2000:LEU2 :0.010031388:1.37926E-08:0.01:0.01:0.01:0.01:0.0104394
 27;
 Lung:2000:MN2:0.010778167:1.48816E-05:0.01:0.01:0.01:0.01:0.033806873;
 Lung:2000:MYE2 :0.154800639:0.063766588:0.01:0.014829632:0.05627163
 3:0.122672891:0.660438295;
 Lung:2000:NC :2.838071299:26.0408083:0.01:0.01:0.011684946:2.7094524
 58:15.53664066;
 Lung:2000:NK :0.011379011:1.8491E-05:0.01:0.01:0.01:0.01:0.026888432;
 Lung:2000:SC :0.325228195:1.505962612:0.01:0.01:0.01:0.040614985:13.0
 8108162;
 Lung:2000:T:0.435842093:0.882885258:0.01:0.01:0.037300932:0.3132657
 77:4.183441938;

10

20

30

40

50

Marrow:100:B2 :0.185902275:0.434241737:0.01:0.01:0.01:0.01:2.71241887
2;
Marrow:100:BAS:0.014519176:0.000265498:0.01:0.01:0.01:0.01:0.06874928
4;
Marrow:100:CLP:2.420824383:24.86621029:0.01:0.01:0.214101778:2.4309
00549:26.01299376;
Marrow:100:EPC:1.904845804:16.12612784:0.01:0.01:0.210311004:1.7079
37823:19.04839729;
Marrow:100:GC :1.621135952:4.828924043:0.01:0.117158592:0.707247043
:2.393001514:11.24602065;
Marrow:100:GMP:0.181084318:0.236147535:0.01:0.01:0.016960511:0.0926
67507:3.474723251;
Marrow:100:GRA:0.374090145:0.702833462:0.01:0.012786779:0.04777679
6:0.259564272:7.914217848;
Marrow:100:HPC:1.13030154:12.97866174:0.01:0.01:0.01:0.195455929:27.
27162771;
Marrow:100:IB :0.31131546:5.521495558:0.01:0.01:0.01:0.01:32.72827775;
Marrow:100:INK:0.01:3.00023E-30:0.01:0.01:0.01:0.01:0.01;
Marrow:100:INKT :0.047342165:0.025679816:0.01:0.01:0.01:0.01:0.709047
616;
Marrow:100:IT :0.053571303:0.039218934:0.01:0.01:0.01:0.01:1.336584538;
Marrow:100:LPB1 :1.608961936:8.767079358:0.01:0.01:0.177612782:1.660
130475:16.04030105;
Marrow:100:LPB2 :2.376398871:16.3857241:0.01:0.01:0.239978477:2.9364
07898:16.65895823;
Marrow:100:MAC:0.0259025:0.006520404:0.01:0.01:0.01:0.01:0.564218133;
Marrow:100:MN :0.038996541:0.026703762:0.01:0.01:0.01091448:0.01931
8394:1.81463237;
Marrow:100:MNT:0.056297223:0.082362769:0.01:0.01:0.01:0.01:1.8258946
31;
Marrow:100:MPC1 :0.367024786:1.315133774:0.01:0.01:0.01:0.090052996:
9.312120992;
Marrow:100:MPC2 :0.175320535:0.418543065:0.01:0.01:0.01:0.023679329:
5.238568571;
Marrow:100:NBC:0.095766964:0.126490199:0.01:0.01:0.01:0.036544526:5.
940112975;
Marrow:100:PB :0.010675228:0.000122283:0.01:0.01:0.01:0.01:0.21806308
1;
Marrow:100:PNK:2.482467108:17.15464909:0.01:0.022599142:0.03668095
6:3.944061666:14.74284385;
Marrow:100:RT :0.012112594:6.19298E-05:0.01:0.01:0.01:0.01:0.040539169;
Marrow:300:B2 :0.072463781:0.080445354:0.01:0.01:0.01:0.01:1.51778171
3;
Marrow:300:BAS:0.518609643:3.362888996:0.01:0.01:0.01:0.01:6.62192535
9;
Marrow:300:CLP:1.284579894:25.89444494:0.01:0.01:0.01:0.047828455:31
.57063454;
Marrow:300:EPC:0.975661564:17.3377232:0.01:0.01:0.01:0.017342857:20.
88881381;
Marrow:300:GC :0.95647576:6.392217988:0.01:0.01:0.021089981:0.538024

10

20

30

40

50

117:20.6035697;
Marrow:300:GMP:1.437037515:20.63563736:0.01:0.01:0.01:0.149559748:24
.18271454;
Marrow:300:GRA:0.626599281:4.597668957:0.01:0.01:0.01:0.147438321:23
.3880025;
Marrow:300:HPC:1.304453698:29.11832644:0.01:0.01:0.01:0.031238635:46
.69346883;
Marrow:300:IB :0.581646271:13.28705952:0.01:0.01:0.01:0.01:40.52106243;
Marrow:300:INK:0.010677684:7.80735E-06:0.01:0.01:0.01:0.01:0.02152063
4;
Marrow:300:INKT :0.011091919:2.26534E-05:0.01:0.01:0.01:0.01:0.0307464
58;
Marrow:300:IT :0.187626221:0.516269114:0.01:0.01:0.01:0.01:4.414912871;
Marrow:300:LPB1 :2.327177218:39.45478944:0.01:0.01:0.01:0.669150502:
47.664881;
Marrow:300:LPB2 :3.027541169:44.75653921:0.01:0.01:0.028413594:1.591
414395:32.7034205;
Marrow:300:MAC:0.533265095:3.571316901:0.01:0.01:0.01:0.038816232:1
3.60439309;
Marrow:300:MN :0.113166634:1.302606568:0.01:0.01:0.01:0.01:14.073386
51;
Marrow:300:MNT:0.187233959:0.61635283:0.01:0.01:0.01:0.01:4.78664248
1;
Marrow:300:MPC1 :0.201187132:1.513216898:0.01:0.01:0.01:0.01:14.1286
1891;
Marrow:300:MPC2 :0.049832787:0.104629922:0.01:0.01:0.01:0.01:3.56267
2271;
Marrow:300:NBC:0.186056236:4.138595569:0.01:0.01:0.01:0.01:37.747659
59;
Marrow:300:PB :0.011186099:0.00027564:0.01:0.01:0.01:0.01:0.317098022;
Marrow:300:PNK:2.032067255:22.65497581:0.01:0.01:0.063921568:1.5014
03207:18.24728518;
Marrow:300:RT :0.01:1.13792E-28:0.01:0.01:0.01:0.01:0.01;
Marrow:2000:B2 :0.167172108:0.259600952:0.01:0.01:0.01:0.01:2.4877431
47;
Marrow:2000:BAS:2.130613564:8.466073636:0.01:0.056819654:0.8253349
25:3.54974934:9.299718758;
Marrow:2000:CLP:1.098629192:10.50141526:0.01:0.01:0.01:0.265548246:2
8.26777702;
Marrow:2000:EPC:0.420817747:2.493197877:0.01:0.01:0.01:0.093986726:1
0.33754185;
Marrow:2000:GC :0.969074716:3.773959377:0.01:0.01:0.069629699:1.1076
61712:13.42674231;
Marrow:2000:GMP:1.71162988:8.833804552:0.01:0.051021256:0.49440933
8:1.397984403:12.66009452;
Marrow:2000:GRA:0.675219767:3.340806905:0.01:0.01:0.014277322:0.347
261219:14.96173219;
Marrow:2000:HPC:0.894207305:5.34592313:0.01:0.01:0.01:0.400270601:20
.58178485;
Marrow:2000:IB :0.369499113:2.05090549:0.01:0.01:0.01:0.010907813:14.

10

20

30

40

50

24513112;
Marrow:2000:INK:0.056216304:0.027923716:0.01:0.01:0.01:0.01:0.6999968
17;
Marrow:2000:INKT :0.111411572:0.167818618:0.01:0.01:0.01:0.01:1.79797
6351;
Marrow:2000:IT :0.11687255:0.091009648:0.01:0.01:0.01:0.01347243:1.73
9893257;
Marrow:2000:LPB1 :1.893858092:16.46027996:0.01:0.01:0.050292558:1.49
6103716:27.37277965;
Marrow:2000:LPB2 :3.302615802:36.83218789:0.01:0.01:0.210685719:3.78 10
4145256:27.26089787;
Marrow:2000:MAC:0.423312503:0.893001499:0.01:0.01:0.01:0.220012709:
4.183752057;
Marrow:2000:MN :0.263946748:2.930026775:0.01:0.01:0.01:0.015794453:1
6.7846469;
Marrow:2000:MNT:0.181549588:0.183314907:0.01:0.01:0.01:0.029673303:
1.778545776;
Marrow:2000:MPC1 :0.362290945:1.790559251:0.01:0.01:0.01:0.03589056
6:14.05147694;
Marrow:2000:MPC2 :0.147777286:0.464731193:0.01:0.01:0.01:0.01:6.2239 20
1615;
Marrow:2000:NBC:0.137362883:0.348696493:0.01:0.01:0.01:0.01:7.142698
962;
Marrow:2000:PB :0.058905676:0.294446787:0.01:0.01:0.01:0.01:9.9288199
57;
Marrow:2000:PNK:1.833011369:8.638914818:0.01:0.011389807:0.2268013
09:2.007042567:9.85192995;
Marrow:2000:RT :0.010730026:7.99407E-06:0.01:0.01:0.01:0.01:0.02095039;
Pancreas:100:A:0.01:2.95436E-27:0.01:0.01:0.01:0.01:0.01;
Pancreas:100:ACI:11.89919455:1239.078518:0.01:0.027755043:1.99274542 30
4:10.39200207:299.3981303;
Pancreas:100:BC :0.014382556:0.001847759:0.01:0.01:0.01:0.01:0.5315780
68;
Pancreas:100:D:0.01:3.72485E-28:0.01:0.01:0.01:0.01:0.01;
Pancreas:100:DUC:0.085962055:0.033775105:0.01:0.01:0.012791323:0.057
646561:1.321106184;
Pancreas:100:EC :9.397344192:1396.05103:0.01:0.01:0.188111294:0.97974
1716:253.3114757;
Pancreas:100:LEU:0.961108391:10.44635577:0.01:0.014391258:0.0412671
31:0.085491737:15.68452118; 40
Pancreas:100:PP :0.025692694:0.002933414:0.01:0.01:0.01:0.011266917:0.
308624983;
Pancreas:100:PSC:2.221217553:63.14401388:0.01:0.01:0.054436934:0.626
533357:46.16280282;
Pancreas:300:A:0.015787667:0.006866904:0.01:0.01:0.01:0.01:1.19647179;
Pancreas:300:ACI:13.64002226:1753.848566:0.01:0.01:0.01:2.740798527:3
12.48915;
Pancreas:300:BC :0.023946869:0.035699395:0.01:0.01:0.01:0.01:2.6076000
67;
Pancreas:300:D:0.01:1.69592E-25:0.01:0.01:0.01:0.01:0.01; 50

Pancreas:300:DUC:0.14580081:0.480585724:0.01:0.01:0.01:0.01:6.811039251;

Pancreas:300:EC :1.476887113:58.74553799:0.01:0.01:0.01:0.01:53.07849045;

Pancreas:300:LEU:2.617639833:230.4585141:0.01:0.01:0.01:0.01:93.58914858;

Pancreas:300:PP :0.01:3.8384E-25:0.01:0.01:0.01:0.01:0.01;

Pancreas:300:PSC:0.158792338:0.445999812:0.01:0.01:0.01:0.01:4.047027631;

Pancreas:2000:A:0.013012481:0.001619796:0.01:0.01:0.01:0.01:0.584870878; 10

Pancreas:2000:ACI:10.86705529:601.7602352:0.01:0.01:0.815365852:8.845002453:155.7998631;

Pancreas:2000:BC :0.012869098:0.000777185:0.01:0.01:0.01:0.01:0.381544425;

Pancreas:2000:D:0.010021663:1.6756E-08:0.01:0.01:0.01:0.01:0.01106179;

Pancreas:2000:DUC:0.206309149:0.698594703:0.01:0.01:0.01:0.020448076:6.846125984;

Pancreas:2000:EC :0.624484527:6.177609328:0.01:0.01:0.01:0.022861064:16.26364263; 20

Pancreas:2000:LEU:0.224386478:1.390589728:0.01:0.01:0.01:0.016633662:7.289777506;

Pancreas:2000:PP :0.01169833:5.85563E-05:0.01:0.01:0.01:0.01:0.049072198;

Pancreas:2000:PSC:0.22808577:0.454643165:0.01:0.01:0.01:0.024362231:2.982198696;

Skin:100:BE:0.088525095:0.433917818:0.01:0.01:0.01:0.01:9.105400656;

Skin:100:EPI :1.497255989:8.07664797:0.01:0.023857561:0.156802628:1.353358422:15.12733217;

Skin:100:KSC:0.025219498:0.016150493:0.01:0.01:0.01:0.01:1.945234359; 30

Skin:100:KSC2:0.317196037:3.681240333:0.01:0.01:0.01:0.03712216:40.56973381;

Skin:100:LEU:6.537198302:190.7648029:0.01:0.01:0.076820057:6.008499214:46.96736574;

Skin:100:SCE:0.098152807:0.13388837:0.01:0.01:0.01:0.036533711:2.313070393;

Skin:300:BE:0.108357985:1.13393988:0.01:0.01:0.01:0.01:18.55862333;

Skin:300:EPI :1.305163535:21.44147441:0.01:0.01:0.01:0.098746599:36.14180497;

Skin:300:KSC:0.019268223:0.015211094:0.01:0.01:0.01:0.01:2.485425355; 40

Skin:300:KSC2:0.419843734:9.493485969:0.01:0.01:0.01:0.01:63.04223322;

Skin:300:LEU:7.856353791:227.1964209:0.01:0.01:0.01:10.01127642:50.34195125;

Skin:300:SCE:0.041529126:0.025377722:0.01:0.01:0.01:0.01:1.015417564;

Skin:2000:BE:0.054235631:0.107612828:0.01:0.01:0.01:0.01:4.894140689;

Skin:2000:EPI :1.463389585:13.46542866:0.01:0.01:0.041508019:0.580429537:24.10158418;

Skin:2000:KSC:0.057082183:0.283160697:0.01:0.01:0.01:0.01:8.947880368;

Skin:2000:KSC2:0.358296776:2.180414803:0.01:0.01:0.01:0.01:17.56111049; 50

Skin:2000:LEU:3.062859377:42.92355399:0.01:0.01:0.039353867:2.372801805:22.81755758;

Skin:2000:SCE:0.418997541:4.3441421:0.01:0.01:0.01:0.01:13.10047803;

SkMuscle:100:B3 :0.576226174:6.536569269:0.01:0.01:0.01:0.01:0.010556533:12.0156782;

SkMuscle:100:B4 :0.126395942:0.228262628:0.01:0.01:0.01:0.01:3.576026664;

SkMuscle:100:DEN:0.01:7.18298E-31:0.01:0.01:0.01:0.01:0.01;

SkMuscle:100:EC :0.675764223:2.800705161:0.01:0.01:0.030371209:0.285263553:7.310227407;

SkMuscle:100:ERB1 :0.01:4.75706E-24:0.01:0.01:0.01:0.01:0.01;

SkMuscle:100:ERB2 :0.010000162:4.76075E-12:0.01:0.01:0.01:0.01:0.010029355;

SkMuscle:100:GMP:0.087430963:0.486566491:0.01:0.01:0.01:0.01:6.365643955;

SkMuscle:100:MAC2 :0.029195195:0.03837845:0.01:0.01:0.01:0.01:2.308267274;

SkMuscle:100:MAC3 :0.013829597:0.000133662:0.01:0.01:0.01:0.01:0.06064265;

SkMuscle:100:MC1:28.93245718:212.9191713:0.101008:28.7416311:35.43885636:36.5236684:44.08999214;

SkMuscle:100:MC2:15.04398917:182.2259111:0.01:5.74097169:11.36666489:21.01674324:54.58031214;

SkMuscle:100:MPC:0.02286293:0.004120317:0.01:0.01:0.01:0.01:0.523549078;

SkMuscle:100:NEUT1 :0.010017958:4.95878E-08:0.01:0.01:0.01:0.01:0.012908385;

SkMuscle:100:NEUT2 :0.011664165:9.56584E-06:0.01:0.01:0.01:0.01:0.011539886:0.017153774;

SkMuscle:100:NEUT3 :0.018413406:0.009697599:0.01:0.01:0.01:0.01:1.162636555;

SkMuscle:100:SC :0.087895601:0.009532824:0.01:0.01:0.05575928:0.105801958:0.345386647;

SkMuscle:100:T:0.109161343:0.059970569:0.01:0.01:0.01:0.028898199:0.752215042;

SkMuscle:300:B3 :0.010361993:2.88286E-06:0.01:0.01:0.01:0.01:0.017963856;

SkMuscle:300:B4 :0.029963223:0.03376483:0.01:0.01:0.01:0.01:1.858624152;

SkMuscle:300:DEN:0.16913301:0.253233149:0.01:0.01:0.01:0.01:1.601330102;

SkMuscle:300:EC :0.895315862:11.40577034:0.01:0.01:0.01:0.01:16.86925061;

SkMuscle:300:ERB1 :0.01:9.0575E-24:0.01:0.01:0.01:0.01:0.01;

SkMuscle:300:ERB2 :0.010088532:1.41868E-06:0.01:0.01:0.01:0.01:0.026024377;

SkMuscle:300:GMP:0.037725117:0.063800617:0.01:0.01:0.01:0.01:2.311184743;

SkMuscle:300:MAC2 :0.014416578:0.002691851:0.01:0.01:0.01:0.01:0.619487802;

10

20

30

40

50

SkMuscle:300:MAC3 :0.010123343:3.80338E-07:0.01:0.01:0.01:0.01:0.013083577;

SkMuscle:300:MC1:7.592685729:87.58666453:0.01:1.765507418:4.031314363:10.95550861:30.46760561;

SkMuscle:300:MC2:15.5601171:833.7239616:0.01:0.01:1.157277635:12.19090123:106.3308553;

SkMuscle:300:MPC:0.030102624:0.035158047:0.01:0.01:0.01:0.01:1.75892828;

SkMuscle:300:NEUT1 :0.01:5.49289E-23:0.01:0.01:0.01:0.01:0.01;

SkMuscle:300:NEUT2 :0.032082425:0.003901068:0.01:0.01:0.01:0.01:0.186659401; 10

SkMuscle:300:NEUT3 :0.01087057:0.000103831:0.01:0.01:0.01:0.01:0.129268097;

SkMuscle:300:SC :0.033081414:0.009127034:0.01:0.01:0.01:0.01:0.465886089;

SkMuscle:300:T:0.010433177:1.68878E-06:0.01:0.01:0.01:0.01:0.013898597;

SkMuscle:2000:B3 :2.27902329:55.61744578:0.01:0.01:0.01:0.01:29.79690193;

SkMuscle:2000:B4 :0.081475145:0.454831686:0.01:0.01:0.01:0.01:6.820171116; 20

SkMuscle:2000:DEN:0.023446865:0.001476213:0.01:0.01:0.01:0.01:0.132248841;

SkMuscle:2000:EC :1.148049772:26.0796183:0.01:0.01:0.01:0.01:26.51870209;

SkMuscle:2000:ERB1 :0.01:2.62163E-29:0.01:0.01:0.01:0.01:0.01;

SkMuscle:2000:ERB2 :0.012219673:0.000496529:0.01:0.01:0.01:0.01:0.28302578;

SkMuscle:2000:GMP:0.010542119:2.43931E-05:0.01:0.01:0.01:0.01:0.054995847;

SkMuscle:2000:MAC2 :0.09516298:0.982113219:0.01:0.01:0.01:0.01:11.65243136; 30

SkMuscle:2000:MAC3 :0.01:2.2775E-30:0.01:0.01:0.01:0.01:0.01;

SkMuscle:2000:MC1:7.164084475:71.86606507:0.01:0.787027995:4.186646381:11.04237596:27.32437377;

SkMuscle:2000:MC2:14.32686526:568.351699:0.01:0.156707198:2.07884161:13.74566834:82.41595513;

SkMuscle:2000:MPC:0.274615674:3.129079571:0.01:0.01:0.01:0.01:14.27447261;

SkMuscle:2000:NEUT1 :0.012227059:0.00053056:0.01:0.01:0.01:0.01:0.306938325; 40

SkMuscle:2000:NEUT2 :0.01:2.46619E-30:0.01:0.01:0.01:0.01:0.01;

SkMuscle:2000:NEUT3 :0.215387785:5.34538476:0.01:0.01:0.01:0.01:27.06598056;

SkMuscle:2000:SC :0.384236497:3.218447212:0.01:0.01:0.01:0.01:8.613897029;

SkMuscle:2000:T:0.01:4.29784E-30:0.01:0.01:0.01:0.01:0.01;

Spleen:100:B:0.789739043:13.27629372:0.009999999:0.01:0.014974386:0.175114572:76.01353696;

Spleen:100:MAC:0.033677171:0.000928003:0.01:0.01052609:0.019320469:0.044604965:0.119540695; 50

Spleen:100:T:0.889478679:2.626844949:0.01:0.028523727:0.229516685:1.134817519:11.21089027;
 Spleen:300:B:0.6007074:5.718042953:0.009999997:0.01:0.01:0.014069245:24.35418936;
 Spleen:300:MAC:0.110306986:0.104400051:0.01:0.01:0.01:0.017623009:1.695440132;
 Spleen:300:T:1.193135349:9.943389873:0.01:0.01:0.01:0.306849232:26.93225583;
 Spleen:2000:B:0.527834996:2.847863632:0.01:0.01:0.01:0.026519544:14.2891133;
 Spleen:2000:MAC:0.028947532:0.008489124:0.01:0.01:0.01:0.01:0.526640415;
 Spleen:2000:T:0.902514553:4.96652535:0.01:0.01:0.01:0.263963255:13.63243794;
 Thymus:100:IT2:0.892838408:8.146903544:0.01:0.01:0.01:0.187870672:27.72681703;
 Thymus:100:IT3:1.517353985:8.980022847:0.01:0.01:0.087409456:1.088970761:12.81068485;
 Thymus:100:IT4:0.509593999:2.415552664:0.01:0.01:0.019557589:0.222231944:12.92113673;
 Thymus:100:LEU3:0.075703947:0.050678587:0.01:0.01:0.01:0.010156311:0.855605674;
 Thymus:100:TPT:0.603227667:0.873439087:0.01:0.02145553:0.26503743:0.596217908:2.432418181;
 Thymus:300:IT2:0.569771808:9.930095561:0.01:0.01:0.01:0.01:31.00492796;
 Thymus:300:IT3:1.539814872:17.40773119:0.01:0.01:0.01:0.178454647:21.665993;
 Thymus:300:IT4:0.293142261:3.327556195:0.01:0.01:0.01:0.01:19.65626539;
 Thymus:300:LEU3:0.100077789:0.113494376:0.01:0.01:0.01:0.01:1.270564448;
 Thymus:300:TPT:0.01:5.05421E-28:0.01:0.01:0.01:0.01:0.01;
 Thymus:2000:IT2:0.461222048:2.278633079:0.01:0.01:0.01:0.015257748:1.52396779;
 Thymus:2000:IT3:1.632787561:11.73465067:0.01:0.01:0.032282811:1.271299967:19.54368881;
 Thymus:2000:IT4:0.249174221:1.215448065:0.01:0.01:0.01:0.01:7.965207557;
 Thymus:2000:LEU3:1.037253207:2.514034235:0.01:0.04187717:0.257041468:0.893381107:4.991732332;
 Thymus:2000:TPT:0.011241477:9.2476E-06:0.01:0.01:0.01:0.01:0.017448864;

10

20

30

40

【符号の説明】

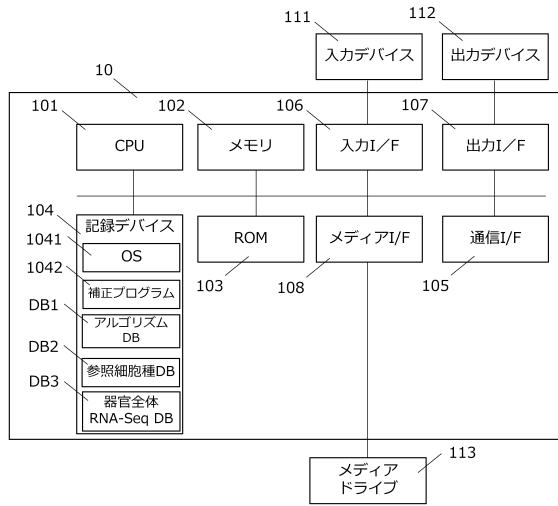
【 0 1 2 8 】

- 1 0 補正装置
- 1 0 1 制御部
- 2 0 解析装置
- 2 0 1 制御部

【図面】

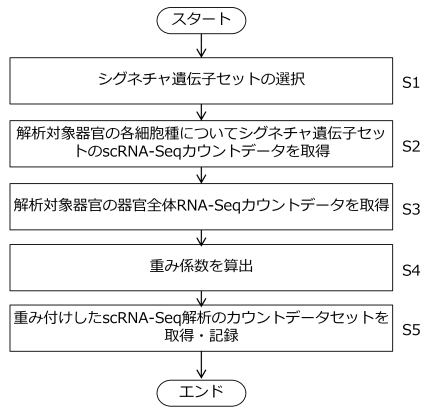
【図 1】

Fig. 1



【図 2】

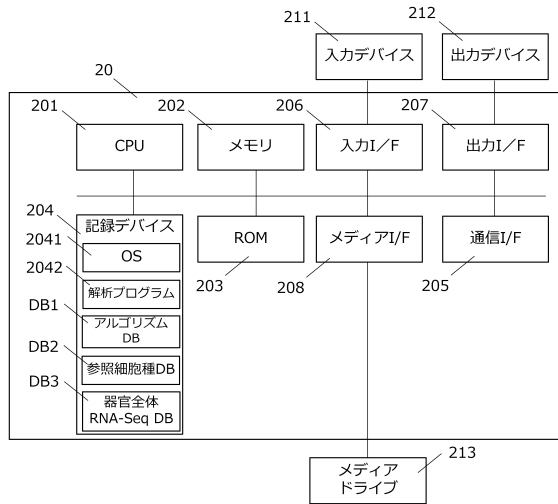
Fig. 2



10

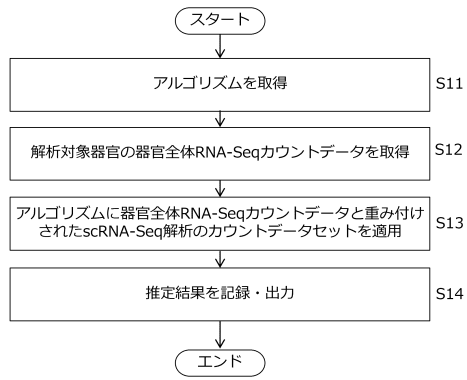
【図 3】

Fig. 3



【図 4】

Fig. 4



20

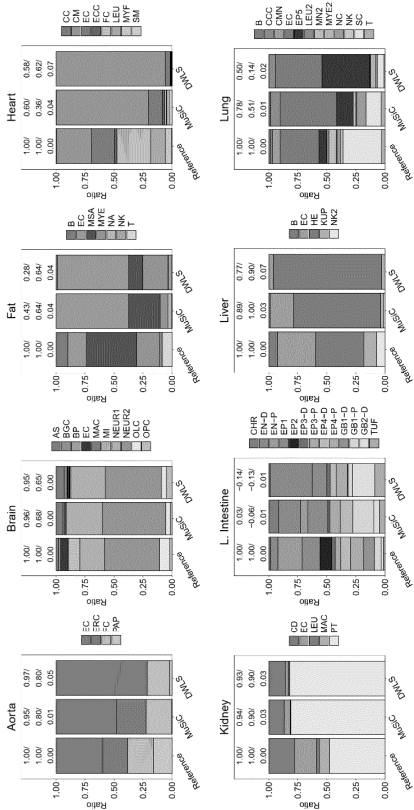
30

40

50

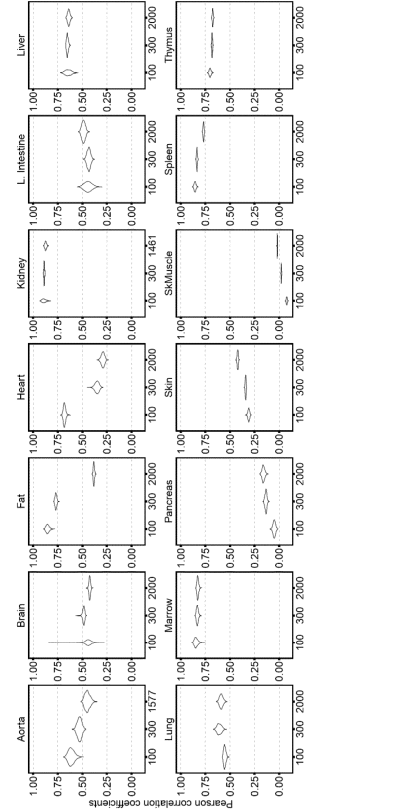
【 5 】

Fig. 5



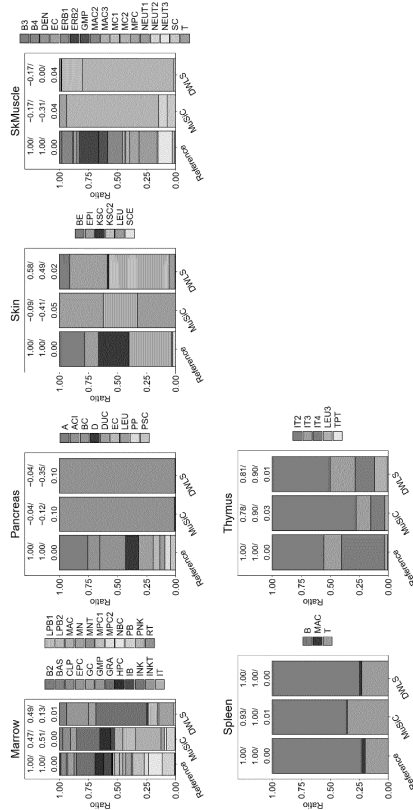
【 7 】

Fig. 7



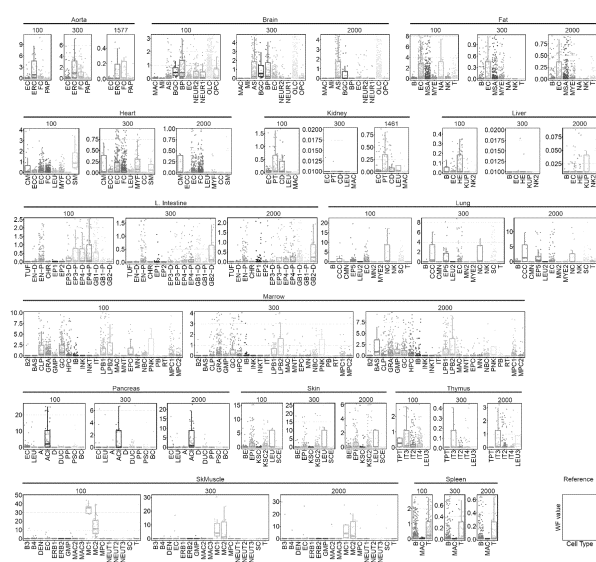
【 6 】

Fig. 6



【 8 】

Fig. 8



10

20

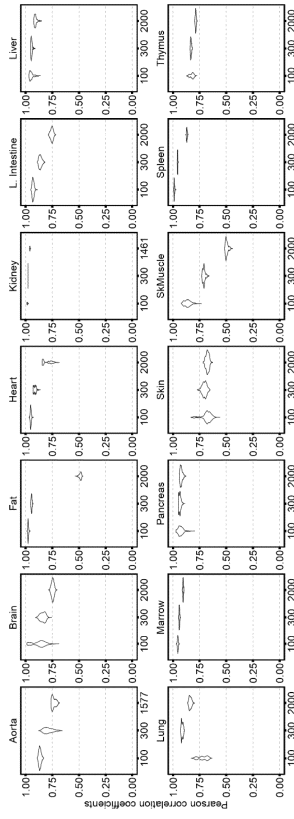
30

40

50

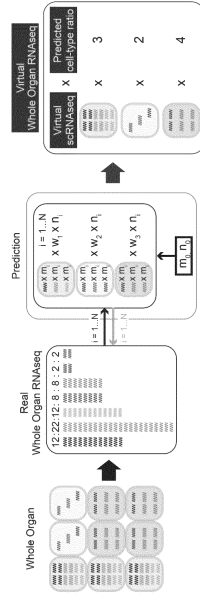
【 9 】

Fig. 9



【 1 0 】

Fig. 10

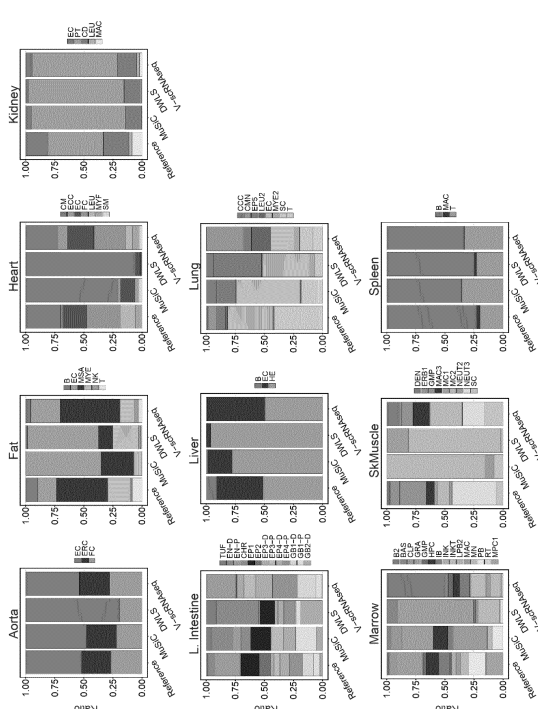


10

20

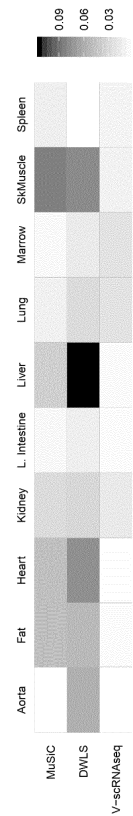
【 1 1 】

Fig. 11



【 1 2 】

Fig. 12



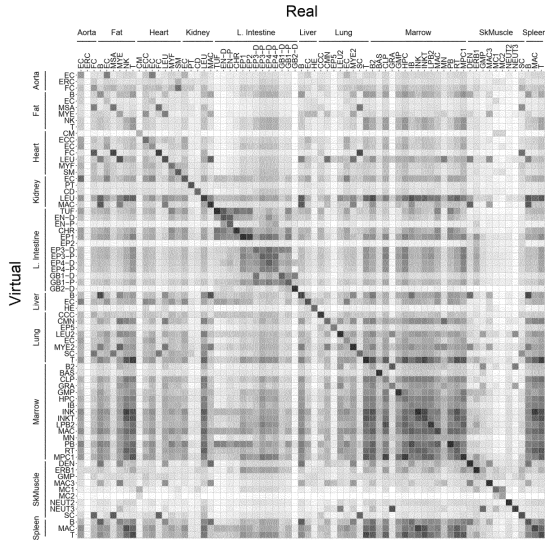
30

40

50

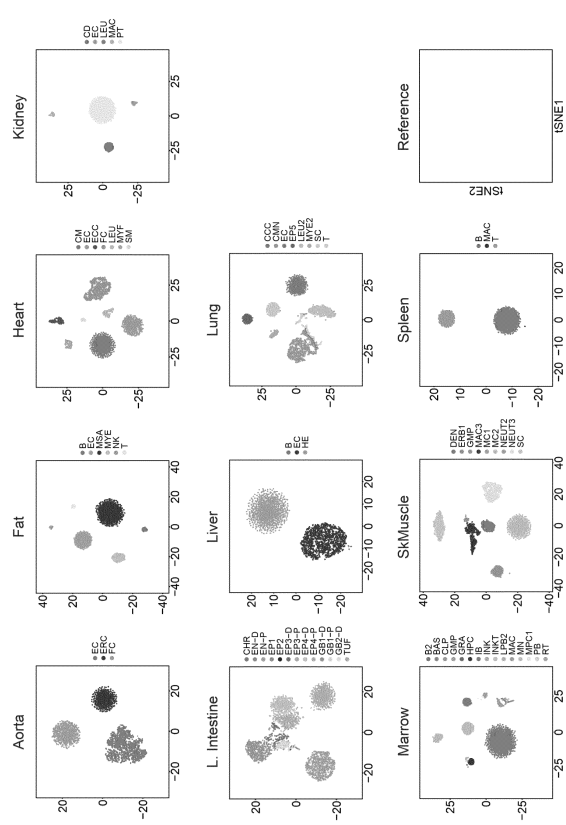
【 1 3 】

Fig. 13



【 1 4 】

Fig. 14

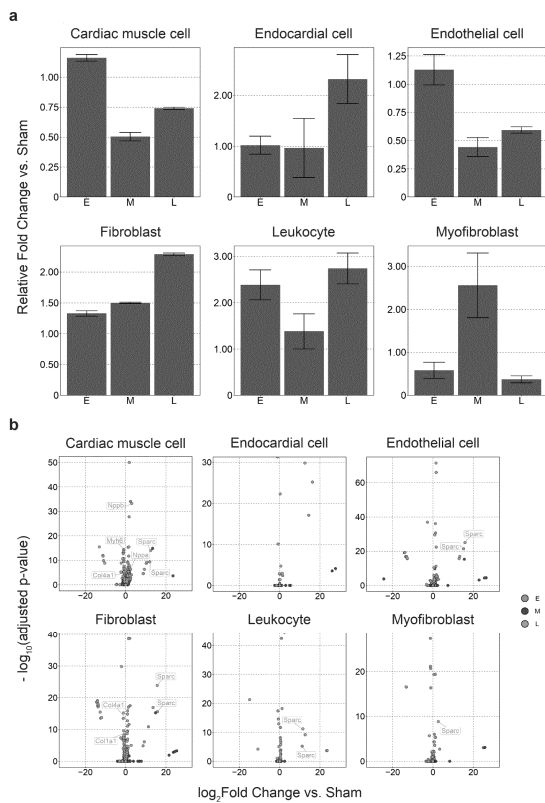


10

20

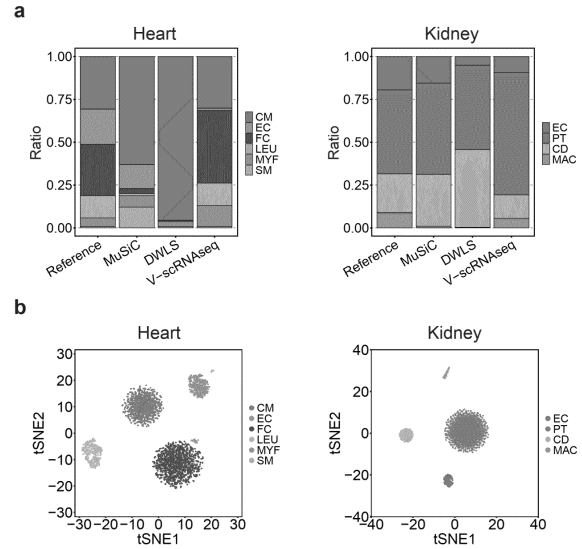
【 1 5 】

Fig. 15



【 1 6 】

Fig. 16



30

40

50

フロントページの続き

(54)【発明の名称】 めの装置及びコンピュータプログラム

(56)参考文献 国際公開第2019/018684(WO, A1)

KANG K. et al. , CDSeq: A novel complete deconvolution method for dissecting heterogeneous samples using gene expression , PLOS Computational Biology , 2019年 , Vol.15, No.12, e1007510

MONACO G. et al. , RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Type , Cell Reports , 2019年 , Vol.26, p.1627-1640

TSOUKAS D. et al. , Accurate estimation of cell-type composition from gene expression data , NATURE COMMUNICATIONS , 2019年 , Vol.10, 2975

(58)調査した分野 (Int.Cl. , DB名)

C12Q 1/6809

C12Q 1/6869

JSTPlus/JMEDPlus/JST7580(JDreamIII)

CAPLUS/MEDLINE/EMBASE/BIOSIS(STN)