

1. 测量含有多个亚群的细胞混合物样品的第一亚群细胞中基因组表达的方法,所述方法包括:

确定多个亚群特征性基因,包括:

至少一个亚群靶基因;和

至少一个亚群参照基因,其相对于所述亚群靶基因具有更低的生物学变异,

所述多个亚群特征性基因中的每一个都被确定为这样的基因,其具有由所述第一亚群细胞提供的其转录本在一个或多个校准细胞混合物中占至少预定的百分比,所述预定的百分比等于或大于50%;

在计算机系统中接收获自测量细胞混合物样品的数据,所述数据获自涉及所述细胞混合物样品的一个或多个反应,无需将所述细胞混合物样品中的细胞分离成多个亚群的细胞成分;

用计算机系统辅助,利用所述数据来确定所述细胞混合物样品中所述至少一个亚群靶基因的转录本的第一量和所述细胞混合物样品中所述至少一个亚群参照基因的转录本的第二量;和

计算参数,所述参数为所述第一量对所述第二量的相对值,所述参数代表了所述细胞混合物样品的所述第一亚群细胞中所述亚群靶基因的基因组表达量。

2. 根据权利要求1所述的方法,其还包括:

测量来自涉及所述细胞混合物样品的一个或多个反应的信号以获得数据。

3. 根据权利要求2所述的方法,其中所述一个或多个反应包括聚合酶链式反应(PCR)、微阵列或测序。

4. 根据权利要求1所述的方法,还包括:

将所述参数与一个或多个临界值比较,以确定所述细胞混合物样品的分类。

5. 根据权利要求1所述的方法,其中确定所述第一亚群的多个亚群特征性基因包括:

接收关于所述第一亚群细胞中多个基因的表达的数据,以及关于一个或多个校准细胞混合物中多个基因的表达的数据;

基于所述数据,确定在所述第一亚群细胞中表达的第一组基因;

接收一个或多个校准细胞混合物中所述第一亚群细胞的比例细胞计数;和

分析所述数据以确定:

所述一个或多个校准细胞混合物中所述第一组基因中的每个基因的第一表达水平;和

所述第一亚群中所述第一组基因中的每个基因各自的表达水平;和

对于所述第一组基因中的每个基因:

使用所述比例细胞计数和所述各自的表达水平相对于所述第一表达水平,确定所述第一亚群是否提供了在所述一个或多个校准细胞混合物中各自基因的转录本的至少的所述预定百分比。

6. 根据权利要求1所述的方法,其中确定所述第一亚群的所述多个亚群特征性基因包括:

接收关于多个基因中的哪个基因由在一个或多个校准细胞混合物中的不同亚群细胞表达的数据;

接收一个或多个校准细胞混合物中不同亚群的每个亚群各自的比例细胞计数,所述各

自的比例细胞计数包括所述第一亚群的第一比例细胞计数P(1)；

基于所述数据,确定在所述第一亚群细胞中表达的第一组基因;

对于所述第一组基因的每个基因:

分析所述数据,以确定表达各个基因的不同亚群的每个亚群中各个基因各自的表达水平,所述各自的表达水平包括所述第一亚群的第一表达水平Ep(1);

使用下述方程式

$$Ep(1) * P(1) \geq \frac{C\%}{(1 - C\%)} \sum_{i=2}^N Ep(i) * P(i)$$

确定所述第一亚群是否提供了在所述一个或多个细胞混合物中各个基因的转录本的至少的所述预定百分比,其中,N为对应于表达各个基因的亚群数目的整数,并且C%为所述预定百分比。

7.根据权利要求1所述的方法,其中所述第一亚群选自中性粒细胞、淋巴细胞、单核细胞、嗜酸性粒细胞、嗜碱性粒细胞、自然杀伤细胞、CD3总T细胞、CD4 T细胞、CD8 T细胞和B细胞淋巴细胞。

8.根据权利要求1所述的方法,其中所述预定百分比为66%或75%。

9.根据权利要求1所述的方法,其中所述第一亚群为B淋巴细胞,且所述细胞混合物样品为血液样品,其中所述至少一个亚群参照基因包括TNFRSF13C。

10.根据权利要求1所述的方法,其中所述第一亚群为T淋巴细胞,且所述细胞混合物样品为血液样品,其中所述至少一个亚群参照基因包括GIMAP7和/或LRRN3。

11.根据权利要求1所述的方法,其中所述第一亚群为粒细胞,且所述细胞混合物样品为血液样品,其中所述至少一个亚群参照基因包括AQP1和/或SOD2。

12.用于测量含有多个亚群的细胞混合物样品的第一亚群细胞中基因组表达的系统,其包括:

确定多个亚群特征性基因的装置,所述基因包括:

至少一个亚群靶基因;和

至少一个亚群参照基因,其相对于所述亚群靶基因具有更低的生物学变异,

所述多个亚群特征性基因中的每一个都被确定为这样的基因,其具有由所述第一亚群细胞提供的其转录本在一个或多个校准细胞混合物中占至少预定的百分比,所述预定的百分比等于或大于50%;

在计算机系统中接收获自测量细胞混合物样品的数据的装置,所述数据获自涉及所述细胞混合物样品的一个或多个反应,无需将所述细胞混合物样品中的细胞分离成多个亚群的细胞成分;

利用所述数据来确定所述细胞混合物样品中所述至少一个亚群靶基因的转录本的第一量和所述细胞混合物样品中所述至少一个亚群参照基因的转录本的第二量的装置;和

计算参数的装置,所述参数为所述第一量对所述第二量的相对值,所述参数代表了所述细胞混合物样品的所述第一亚群细胞中所述亚群靶基因的基因组表达量。

13.根据权利要求12所述的系统,还包括:

测量来自涉及所述细胞混合物样品的一个或多个反应的信号的装置。

14. 根据权利要求13所述的系统,其中所述一个或多个反应包括聚合酶链式反应(PCR)、微阵列或测序。

15. 根据权利要求12所述的系统,还包括:

将所述参数与一个或多个临界值比较,以确定所述细胞混合物样品的分类的装置。

16. 根据权利要求12所述的系统,其中确定所述第一亚群的多个亚群特征性基因的装置包括:

接收关于所述第一亚群细胞中多个基因的表达的数据,以及关于一个或多个校准细胞混合物中多个基因的表达的数据的装置;

基于所述数据,确定在所述第一亚群细胞中表达的第一组基因的装置;

接收一个或多个校准细胞混合物中所述第一亚群细胞的比例细胞计数的装置;和

分析所述数据以确定以下水平的装置:

所述一个或多个校准细胞混合物中所述第一组基因中的每个基因的第一表达水平;和

所述第一亚群中所述第一组基因中的每个基因各自的表达水平;和

对于所述第一组基因中的每个基因:

使用所述比例细胞计数和所述各自的表达水平相对于所述第一表达水平,确定所述第一亚群是否提供了在所述一个或多个校准细胞混合物中各自基因的转录本的至少的所述预定百分比的装置。

17. 根据权利要求12所述的系统,其中确定所述第一亚群的所述多个亚群特征性基因的装置包括:

接收关于多个基因中的哪个基因由在一个或多个校准细胞混合物中的不同亚群细胞表达的数据的装置;

接收一个或多个校准细胞混合物中不同亚群的每个亚群各自的比例细胞计数的装置,所述各自的比例细胞计数包括所述第一亚群的第一比例细胞计数P(1);

基于所述数据,确定在所述第一亚群细胞中表达的第一组基因的装置;

对于所述第一组基因的每个基因:

分析所述数据,以确定表达各个基因的不同亚群的每个亚群中各个基因各自的表达水平的装置,所述各自的表达水平包括所述第一亚群的第一表达水平Ep(1);

使用下述方程式

$$Ep(1) * P(1) \geq \frac{C\%}{(1 - C\%)} \sum_{i=2}^N Ep(i) * P(i)$$

确定所述第一亚群是否提供了在所述一个或多个校准细胞混合物中各个基因的转录本的至少的所述预定百分比的装置,其中,N为对应于表达各个基因的亚群数目的整数,并且C%为所述预定百分比。

18. 根据权利要求12所述的系统,其中所述第一亚群选自中性粒细胞、淋巴细胞、单核细胞、嗜酸性粒细胞、嗜碱性粒细胞、自然杀伤细胞、CD3总T细胞、CD4 T细胞、CD8 T细胞和B细胞淋巴细胞。

19. 根据权利要求12所述的系统,其中所述预定百分比为66%或75%。

20. 根据权利要求12所述的系统,其中所述第一亚群为B淋巴细胞,且所述细胞混合物

样品为血液样品，其中所述至少一个亚群参照基因包括TNFRSF13C。

21. 根据权利要求12所述的系统，其中所述第一亚群为T淋巴细胞，且所述细胞混合物样品为血液样品，其中所述至少一个亚群参照基因包括GIMAP7和/或LRRN3。

22. 根据权利要求12所述的系统，其中所述第一亚群为粒细胞，且所述细胞混合物样品为血液样品，其中所述至少一个亚群参照基因包括AQP1和/或SOD2。

23. 一种试剂盒，其用于测量含有多个亚群的细胞混合物的第一亚群细胞中的基因组表达，所述试剂盒包含：

多个亚群特征性基因的引物，包括：

至少一个第一寡核苷酸，其与亚群靶基因的转录本的至少一部分特异性杂交；和

至少一个第二寡核苷酸，其与至少一个亚群参照基因的转录本的至少一部分特异性杂交，所述亚群参照基因相对于所述亚群靶基因具有更低的生物学变异，所述多个亚群特征性基因中的每个基因都被确定为这样的基因，其具有由所述第一亚群细胞提供的其转录本在所述细胞混合物中占至少预定的百分比，所述预定的百分比等于或大于50%，所述引物适用于涉及含有多个亚群的细胞混合物样品的一个或多个反应，并且适合提供关于所述细胞混合物样品中至少一个亚群靶基因的转录本的第一量和所述细胞混合物样品中至少一个亚群参照基因的转录本的第二量的数据，所述数据用于确定代表了所述细胞混合物样品的所述第一亚群细胞中所述亚群靶基因的基因组表达量的参数，无需将所述细胞混合物样品中的细胞分离成多个亚群的细胞成分。

24. 根据权利要求23所述试剂盒，其中所述至少一个第一寡核苷酸是用于特异性扩增所述亚群靶基因的转录本的至少一部分的第一引物，并且其中所述至少一个第二寡核苷酸是用于特异性扩增所述亚群参照基因的转录本的至少一部分的第二引物。

25. 根据权利要求23所述试剂盒，其中所述至少一个第一寡核苷酸是用于检测所述亚群靶基因的转录本的第一标记探针，并且其中所述至少一个第二寡核苷酸是用于检测所述亚群参照基因的转录本的至少一部分的第二标记探针。

26. 根据权利要求23所述试剂盒，其中所述第一亚群为B淋巴细胞，且所述细胞混合物为血液样品，其中所述至少一个亚群参照基因包括TNFRSF13C。

27. 根据权利要求23所述试剂盒，其中所述第一亚群为T淋巴细胞，且所述细胞混合物为血液样品，其中所述至少一个亚群参照基因包括GIMAP7和LRRN3。

28. 根据权利要求23所述试剂盒，其中所述第一亚群为粒细胞，且所述细胞混合物为血液样品，其中所述至少一个亚群参照基因包括AQP1和SOD2。

29. 根据权利要求23所述试剂盒，其中所述第一亚群为B淋巴细胞，且所述细胞混合物为血液样品，其中所述至少一个亚群靶基因包括表II中所列的基因。

30. 根据权利要求23所述试剂盒，其中所述第一亚群为CD4 T淋巴细胞，且所述细胞混合物为血液样品，其中所述至少一个亚群靶基因包括表III中所列的基因。

31. 根据权利要求23所述试剂盒，其中所述第一亚群为CD8 T淋巴细胞，且所述细胞混合物为血液样品，其中所述至少一个亚群靶基因包括表IV中所列的基因。

32. 根据权利要求23所述试剂盒，其中所述第一亚群为粒细胞，且所述细胞混合物为血液样品，其中所述至少一个亚群靶基因包括表V中所列的基因。

33. 根据权利要求23所述试剂盒，其中所述第一亚群为CD3+ve 总T细胞，且所述细胞混

合物为血液样品,其中所述至少一个亚群靶基因包括表VI中所列的基因。

一种细胞类型的基因表达水平的测定

[0001] 相关申请的交叉引用

[0002] 本申请为2011年7月21日递交的美国专利申请第61/510,445号的PCT申请，并要求其权益，该申请名称为“用于测定一种细胞类型的基因表达水平或转录丰度的方法、程序和组合物(METHODS, PROCESS AND COMPOSITIONS FOR DETERMINATION OF GENE EXPRESSION LEVELS OR TRANSCRIPT ABUNDANCE OF A CELL TYPE)”，其通过引用全文并入本文用于所有目的。

发明领域

[0003] 本公开涉及分析血液样品中不同造血亚群中的基因表达水平(转录丰度)。更具体地，本发明实施方案可测定在多种不同细胞类型混合物中的特定造血细胞亚群的基因转录丰度，而无需分离细胞以获得该特定成分亚群(fraction)的细胞。

[0004] 发明背景

[0005] 外周血细胞中基因的表达水平(转录丰度, TA)是重要的生物标志。当前使用外周血液样品中的基因TA作为生物标志的方法是以这样的方式在外周血的细胞混合物样品中进行，其使得无法获得有关特定造血亚群，如B细胞淋巴细胞的TA的信息。例如，这一些方法获得了细胞混合物中所有细胞类型的表达水平结果，而非特定造血亚群的表达水平结果。

[0006] 具体地，在去卷积的数学方法中(Lu et al.)，确定了细胞混合物样品中每种成分细胞类型的比例计数。同一组别内的样品中特定细胞类型的基因表达水平被假定是恒定的。仅基于所述样品中特定成分的亚群的比例来确定在细胞混合物样品中的表达，然而不能测量各个体中细胞类型亚群的真实表达，因为其被假定为是恒定的。

[0007] 另一方面，特定造血亚群(例如，白细胞亚群)的基因TA是优选的生物标志。为了获得特定造血亚群的TA，本领域的当前技术状态需要预先进行细胞分离或细胞分选，以从外周全血样品中分离所述特定造血亚群，然后进行基因表达的定量分析。细胞亚群分离是费力而单调的工作。该过程在建立临床服务实验室方面实施很困难。因而，当前的方法在分析外周血液样品中基因表达上存在局限性。

[0008] 对纯化的造血(尤其是白细胞)亚群中的基因表达水平已经有了广泛的研究。为了得到供研究的特定白细胞亚群的样品，这些研究都在细胞分离或细胞培养步骤后进行。另外还存在检测不同类型的外周血样品的研究方向，包括全血和外周血单核细胞样品中的基因表达水平作为疾病、治疗响应和预后指标的生物标志。

[0009] 因此，提供新的方法、设备、系统和组合物是可取的，其用以免去在测定特定造血亚群(例如白细胞细胞类型)中靶基因的基因表达水平或TA过程中所需的细胞纯化或分离，并且能提供其他优势。

[0010] 发明概述

[0011] 本发明实施方案可提供用来测定特定细胞类型亚群的基因表达水平(转录丰度)的方法、系统和组合物，其通过分析由各种细胞类型的多个亚群组成的细胞混合物样品，而无需提前分离细胞类型亚群成分(component)。例如，可确定特定细胞类型亚群的特征性靶

基因和参照基因。当细胞混合物中某基因的至少50%的转录本(或大于50%的其他百分比)是来自该亚群时,该基因可定义为该亚群的特征性基因。特征性基因可由可能涉及分离亚群的校准实验鉴定,但这类实验会由分析的创建者进行,而非在生产(production)运行期间进行。细胞混合物中特征性靶基因和参照基因之间的相对丰度可用作生物标志,例如,等同于亚群中分离的细胞样品的相对丰度。如本文所述,细胞混合物中特征性靶基因和参照基因的相对表达水平可与在分离的亚群中测得的相对表达相关联。因而,可在无分离亚群细胞的困难步骤下获得类似的生物标志。

[0012] 根据一个实施方案,一种方法测量含有多个亚群的细胞混合物的第一细胞亚群的基因组表达。多个亚群特征性基因被确定,并且包括至少一个亚群靶基因和至少一个相对于亚群靶基因具有更低的生物学变异的亚群参照基因。在细胞混合物中,所述多个亚群特征性基因中的每一个都被确定为这样的基因,其具有由所述第一亚群细胞提供的其转录本的至少的预定百分比,其中所述预定百分比等于或大于50%。一种计算机系统测定细胞混合物中至少一个亚群靶基因的转录本的第一量和细胞混合物中至少一个亚群参照基因的转录本的第二量。计算了所述第一量对所述第二量的相对值参数。该参数表示了所述第一亚群中亚群靶基因的基因组表达量。

[0013] 根据另一个实施方案,一种方法确定第一亚群的亚群参照基因。接收有关多个基因中的哪些基因在多个细胞混合物样品中的所述第一细胞亚群中表达的数据。基于该数据确定所述第一细胞亚群中表达的第一组基因。一种计算机系统分析所述数据以确定多个第一亚群样品的每个样品中第一组基因中的每个基因各自的表达水平。每个第一亚群样品都来自不同对象的不同细胞混合物样品。使用各自的表达水平确定亚群基因子集,在多个细胞混合物样品中,其每个各自具有由所述第一亚群细胞提供的其各自转录本的至少的预定百分比。所述预定百分比等于或大于50%。分析所述亚群基因子集各自的表达水平以确定至少一个亚群参照基因,其在所述第一亚群样品中具有小于临界值的表达水平变异。

[0014] 根据另一个实施方案,提供了一种用于测量含有多个亚群的细胞混合物的第一细胞亚群中基因组表达的试剂盒。所述试剂盒包含多个亚群特征性基因的引物。所述引物包括至少一个第一寡核苷酸,其与亚群靶基因的转录本的至少一部分特异性杂交;和至少一个第二寡核苷酸,其与至少一个亚群参照基因的转录本的至少一部分特异性杂交,所述亚群参照基因相对于所述亚群靶基因具有更低的生物学变异。在所述细胞混合物中,多个亚群特征性基因中的每个基因都被确定为这样的基因,其具有由所述第一亚群细胞提供的其转录本的至少的预定百分比。所述预定百分比等于或大于50%。

[0015] 其他实施方案涉及与本文描述的方法相关的系统和计算机可读介质。

[0016] 参考下述详细说明和附图可以更好地理解本发明的特性和优点。

附图说明

[0017] 图1A所示为外周全血(WB)样品的有核细胞群中特定白细胞亚群比例的实例。图1B所示为外周血单核细胞(PBMC)样品的有核细胞群中特定白细胞亚群比例的实例。

[0018] 图2为说明用于根据本发明的实施方案测量含有多个亚群的细胞混合物的第一亚群中基因组表达的方法200的流程图。

[0019] 图3所示为根据本发明实施方案的关于细胞混合物样品的假想实例,当可在特定

亚群和目标细胞混合物样品中进行基因表达测量时,其用来确立用于选择细胞类型亚群特征性基因的属性。

[0020] 图4A所示为显示根据本发明的实施方案的外周全血样品中造血亚群的结果的表400。图4B所示为显示外周血单核细胞样品中造血亚群的结果的表450。

[0021] 图5所示为根据本发明的实施方案的图表500,其显示了细胞混合样品中来自特定细胞类型亚群的转录本百分比与特定细胞类型亚群成分的TA对比细胞混合样本的TA的倍数差别(X)的关系,所述细胞混合物样本为含有各种比例细胞计数0.05(5%)至0.5(50%)的亚群的细胞混合物样本。

[0022] 图6所示为根据本发明的实施方案的关于细胞混合物样品的假想实例,当仅可进行两个亚群而非所述细胞混合物样品的基因表达测量时,其用来确立用于选择细胞类型亚群特征性基因的特性。

[0023] 图7所示为根据本发明的实施方案的另一个关于细胞混合物样品的示意实施例,当仅可进行亚群而非所述细胞混合物样品的基因表达测量时,其用来确立用于选择细胞类型亚群特征性基因的特性。

[0024] 图8A所示为根据本发明的实施方案的图表800,其为观察到的B细胞亚群成分样品和平均全血样品之间TA的差别表达(X,倍数差别);图8B所示为显示根据本发明的实施方案的满足X50的标准的基因(由图8A的表800确定)的图表850。

[0025] 图9的图表显示了细胞混合物样品中使用亚群特征性基因的基因定量新方法(y轴,外周全血中KCNG1相对于TNFRSF13C的相对表达)和经纯化的亚群中通过金标法使用常规管家基因定量的相同靶基因(x轴,在纯化的B细胞成分中KCNG1相对于GAPDH的相对表达)之间的关联。

[0026] 图10A所示的表1000为关于采集自各个个体的CD3+ve总T细胞和全血样品之间的基因表达水平比较。数值(X)表示以所述细胞亚群表达水平高出全血样品表达水平的倍数表示的差别表达水平。所述数据用于评估亚群特征性基因的X50标准。图10B所示的表1050为用于确定亚群参照基因的TA变异量(variance)。图10C所示的表1060为富集后的CD3+ve总T细胞(金标法)中靶基因的表达水平与直接LS-TA检测的关联,以及使用两个亚群参照基因的效果。关联程度表示为决定系数R²。

[0027] 图11A的图表显示了MAF的直接LS-TA检测(y轴)和通过金标法使用富集的总T细胞样品定量的相同基因(x轴)之间的关联。图11B的图表显示使用两个亚群参照基因进行MAF的直接LS-TA检测。相关系数具有统计学意义(p<0.01)。

[0028] 图12A所示的表1200为关于采集自各个个体的粒细胞和全血液样品之间的基因表达水平比较。值(X)表示依据所述细胞亚群中比全血样品中表达水平高出的倍数的差别表达水平。所述数据用于评估亚群特征性基因的X50标准,图12B所示的表1250为用于确定亚群参照基因的分离的粒细胞样品中基因的TA变异量。图12C所示的表1260为使用不同亚群参照基因,通过其与通过金标法得到的结果(在分离的粒细胞样品中的TA)的关联进行直接LS-TA检测。关联程度表示为决定系数R²。

[0029] 图13的图表显示了RNASE3的直接LS-TA检测(y轴)和通过金标法使用富集的粒细胞样品定量的相同基因(x轴)之间的关联。相关系数具有统计学意义(p<0.01)。

[0030] 图14A的图表显示了使用微阵列数据对GZMK基因进行直接LS-TA检测。图14B的图

表显示了使用微阵列数据对MAF基因进行直接LS-TA检测。图14C的图表显示了使用微阵列数据对NELL2基因进行直接LS-TA检测。相关系数具有统计学意义 ($p < 0.01$)。

[0031] 图15所示为根据本发明实施方案的系统和方法可用的示例性计算机系统1500的方框图。

[0032] 定义

[0033] 造血亚群是指血液的细胞组合物，并且由大致分为红细胞、白血细胞(白细胞)和血小板的不同细胞类型的细胞群组成。白细胞占血液中有核细胞群的大多数。白细胞包括具有特定功能和特性的不同细胞类型，如中性粒细胞、淋巴细胞、单核细胞、嗜酸性粒细胞、嗜碱性粒细胞和自然杀伤细胞。

[0034] 下述白细胞亚群为实例。粒细胞(多形核白细胞)的特征在于其细胞质中存在细胞质颗粒。中性粒细胞、嗜碱性粒细胞和嗜酸性粒细胞为3种粒细胞。中性粒细胞占所有粒细胞的~90%。淋巴细胞与免疫应答有关。它们分为T细胞和B细胞，其负责不同的免疫机制、细胞介导免疫和抗体产生。T细胞还分为CD4和CD8T细胞，也称为辅助T细胞和T毒性/抑制细胞。自然杀伤细胞为与先天免疫有关的特定形式淋巴细胞。单核细胞为在循环中运输的巨噬细胞，并且涉及抗体呈递和吞噬作用。

[0035] 如本文所述，直接白细胞亚群特异性转录丰度分析(直接LS-TA)可测定一个白细胞亚群中靶基因的转录丰度(TA)，而无需预先进行细胞分离以从血液样品中分离亚群细胞。例如，如果该亚群为B淋巴细胞，则该分析可称为B细胞的直接LS-TA。B细胞的直接LS-TA可从细胞混合物样品如外周全血样品直接测定丰度，无需进行细胞分离以分离所述B细胞。本文公开的这种新的直接LS-TA方法可基于两种亚群特征性基因的定量进行。

[0036] 在细胞混合物中，亚群特征性基因(也称为亚群基因)是指任何这样的基因，其至少50%(或其他大于50%的百分比)的转录本由单个白细胞亚群或细胞类型提供。两种类型的亚群基因为亚群特征性靶基因和亚群特征性参照基因。

[0037] 亚群特征性靶基因(也称为亚群靶基因)选自确定的特定亚群的亚群特征性基因。这些基因可能涉及一些目标途径，可能在健康对象和患者之间存在差别化的表达，或者可能与其他目标基因共表达。

[0038] 亚群特征性参照基因(也称为亚群参照基因)选自确定的特定亚群的亚群特征性基因。在确定的特定亚群的亚群基因中，它们显示低水平或最低程度的生物学变异。

[0039] 发明详述

[0040] 外周血由不同血液细胞类型组成，包括红细胞(RBC)和各种白细胞亚群，它们以不同的比例存在。血液循环中单个白细胞亚群(例如粒细胞、淋巴细胞)的基因表达水平非常有前景用作特征性生物标志。目前测量血液样品中的转录丰度(TA)作为疾病生物标志没有对特定造血亚群中的TA进行定量，或者必须分离特定造血亚群的细胞作为生产运行的部分来定量TA。本发明实施方案避免了细胞分离步骤，且通过确定和使用特定造血亚群中表达大于50%(或其他大于50%的百分比)的亚群基因，能直接测定细胞混合物样品中特定细胞类型的基因表达水平。一个实例即为本文描述的直接LS-TA检测。

[0041] 本发明实施方案可使用一个或多个亚群特征性靶基因相对于一个或多个亚群特征性参照基因的相对转录丰度(例如，仅一个基因的或多个基因的总结性指标)作为特定细胞类型亚群中靶基因的基因表达水平的定量量度。实施例证明了，在外周血液样品中测量

的特征性靶基因(即对特定亚群特征性的)的该相对丰度与通过细胞分离如磁珠细胞分选获得的特定亚群部分中靶基因的基因表达水平相关。因此,已证明,特定白细胞亚群的转录丰度可在无需预先分离细胞下测定。

[0042] 本发明实施方案的潜在应用非常广泛,因为其使得能够测定外周血液样品中特定造血亚群的基因表达。例如,该测定的亚群基因(如本文所述)的相对丰度可用作生物标志,以用于诊断和预后自身免疫疾病、感染性疾病、慢性疾病、急性损伤、败血性休克、急性器官衰竭或与造血亚群细胞中改变的基因表达有关的病理性病症。此外,本发明实施方案可用于确定药物遗传学生物标志,以用于监测免疫抑制药物或其他与造血亚群细胞中改变的基因表达有关的药物的治疗响应。在实体器官的组织活检应用中,实例包括但不限于实体器官如肾脏、肝脏、皮肤等的组织中恶性细胞或各种浸润性淋巴细胞亚群的基因表达。

[0043] I. 不同血液样品中的亚群

[0044] 外周血由不同血液亚群(细胞类型)组成,其在全血的有核细胞群中以不同的比例存在。红细胞和白细胞为主要谱系。白细胞又由不同亚群组成。例如,在外周全血中,B细胞淋巴细胞占有核细胞的约5%,而粒细胞占有核细胞的多达65%。

[0045] 造血亚群的实例包括但不限于,各种血液细胞类型或白细胞亚群如淋巴细胞、粒细胞、T淋巴细胞和B淋巴细胞,它们以各种比例构成了外周血的细胞组分,其中外周血为细胞混合物样品的一个实例。外周血的细胞混合物样品包括但不限于全血、处理的全血或外周血单核细胞样品。除了血液样品,组织活检样品是本发明实施方案也可应用的细胞混合物样品的另一实例。

[0046] 不同白细胞亚群细胞以不同比例存在于外周血中,并且其中比例细胞计数高于5%。在标准的血液学教科书中已有确定的该比例的参考范围。图1A所示为表100,其为外周全血(WB)样品的有核细胞群中特定白细胞亚群的比例的实例。图1B所示为表150,其为外周血单核细胞(PBMC)样品的有核细胞群中特定白细胞亚群的比例的实例。

[0047] 从这两个表明显可知,所列的造血亚群代表外周血样品中5%至65%的总有核细胞。因此,使用代表外周血液样品中5%的造血细胞亚群的实例证明了实验的原理和效用。下述结果证明,对于占5%的细胞混合物样品的亚群,本发明实施方案能够测定造血亚群的特定基因表达。因此,其他占更高比例的细胞类型或亚群可作为本文描述的实施方案的目标。另外,由于5%是相对低的百分比,所占比例低于5%的造血亚群也可按照本发明实施方案处理。目标白细胞亚群可为,例如,淋巴细胞、B细胞、T细胞、中性粒细胞、粒细胞、单核细胞、嗜酸性粒细胞等。其他实例为活组织检查样品中的细胞类型亚群,例如各种器官中的浸润性淋巴细胞。

[0048] II. 测量表达水平

[0049] 特定基因的表达水平可提供有用的生物标志。例如,可将表达水平(也称为转录丰度)与临界值进行比较,以确定用于诊断和预后的类别。一些当前的方法测定细胞混合物中所有亚群的特定基因的表达水平(如下述A部分所述)。然而,该生物标志可能会不准确,因为表达水平的变化可源于各种因素,其中一些因素不是疾病或其他症状的指标。其他当前的方法(B部分)分离特定亚群的细胞以获得更准确的生物标志;然而,进行分离困难且费钱。C部分介绍了本发明的技术。

[0050] A. 所有细胞类型(亚群)

[0051] 典型的方法测定细胞混合物样品中由所有不同细胞类型提供的转录丰度(TA)的量。例如,通用管家基因可在相对定量技术如定量PCR (QPCR) 中用作内部参照基因。作为另一个实例,微阵列可用于定量细胞混合物样品中的基因表达,其中将数据以对细胞混合物样品中细胞类型或白细胞亚群不具特异性的整体方式归一化。作为另一个实例,测序技术可用于定量细胞混合物中的基因表达,其中也将数据以对细胞混合物中细胞类型或白细胞亚群不具特异性的整体方式归一化。

[0052] 用外周血如外周全血(WB) 样品或外周血单核细胞(PBMC) 样品中不同细胞亚群的混合物样品实施这类方法。通过这些现有方法获得的结果受限于:(a) 它们不对任何造血亚群具有特异性,和(b) 结果受到外周血的这类细胞混合物样品中各种造血亚群所占比例的影响。因而,所有这类方法没有提供细胞混合物样品中特定造血亚群的TA的评估或定量。这些现有技术的生物标志仅能定量由细胞混合物样品中所有不同细胞类型提供的总TA,并且不能表示特定造血或白细胞亚群的TA。

[0053] 全血样品或细胞混合物样品中的转录丰度受到至少两个因素的影响:(1) 样品中各种细胞亚群的比例,和(2) 每个单独细胞亚群中的转录丰度(Fan and Hegde 2005)。在大多数情况下,有关转录丰度变化的生物学和临幊上相关的问题与因素(2)有关,即,特定造血亚群中转录丰度的变化。因此,细胞混合物样品中转录丰度的任何变化都不能直接归因于目标造血亚群中转录丰度的潜在变化。

[0054] 测定特定白细胞亚群而非全血样品或其他细胞混合物中的转录丰度可消除由于各种白细胞亚群组成差异导致的细胞混合物中测量的转录丰度的混杂因素。本发明的实施方案易于转化为分析试剂盒,以用于这类目的并应用于定量在各种临幊情况中的这些亚群特异性生物标志。

[0055] B. 分离的特定亚群

[0056] 如上所述,现有方法使用通用管家基因作为参照基因(例如 β -肌动蛋白、GAPDH、核糖体蛋白基因) 测定外周血液品中的基因表达。类似地,各种对数据进行整体非特异性归一化的方法被用于分析大规模的表达数据如微阵列和测序。然而,这些参照基因和非特异性归一化方法对细胞混合物样品中的任何造血或细胞类型亚群都不具特异性。因此,通过现有方法从外周全血样品或其他细胞混合物样品确定的基因表达的相对丰度仅表示外周血混合物样品中存在的所有不同细胞类型的靶基因的细胞计数加权平均表达。

[0057] 因而,细胞混合物样品中的此类基因表达分析产生了混杂结果,因为当不同细胞亚群具有不同的基因表达水平时,不确定差别性基因表达是否由以下因素引起:单个细胞类型中的病理、多个细胞类型中同时发生的变化,或各种细胞类型计数的变化。因此,当前的方法使用细胞分选、隔离或分离以获得富集的特定细胞亚群样品,从而测定细胞特异性的基因表达变化。然而,分离细胞困难且费钱。

[0058] 当前的标准方法和金标法由以下组成:(1) 细胞分离过程以获得富集的特定目标细胞类型的样品,和(2) 测量所富集的目标细胞类型样品中的基因表达水平。分离目标细胞类型后,提取RNA,然后可进行逆转录步骤。此后,可通过通用的方法之一来测量基因表达(转录丰度, TA),如通过实时定量PCR、微阵列、测序或其他方法。

[0059] 可通过各种方法进行细胞分选或分离,包括:(1) 密度梯度法,例如将聚蔗糖用于分离PBMC和粒细胞;(2) 荧光激活细胞分选术(FACS),其中细胞类型特异性表面标记结合至

荧光标记的抗体，并通过流式细胞仪以电磁机理进行分选；(3) 磁激活细胞分选术(MACS)，其中细胞类型特异性表面标记结合至连接有磁珠的抗体，并可通过使样品通过磁场来分离目标细胞。最近，还开发出了微流体方法来提供细胞分离功能，以从细胞混合物样品中获得细胞类型富集的样品。

[0060] C. 不经分离的特定亚群

[0061] 如上所述，直接在细胞混合物样品中评估细胞类型特异性基因表达而无需预先进行细胞分选是可取的。本发明实施方案免于预先进行细胞分离的要求，并能直接在细胞混合物样品中测定细胞类型亚群特异性基因表达。本发明实施方案使用选定的细胞类型特征性参照基因和细胞类型特征性靶基因，以测定外周血的细胞混合物中特定造血或细胞类型亚群的基因表达水平(例如TA)。本发明实施方案也可应用于实体组织细胞混合物样品，如组织活检样品。

[0062] 以下的实施例使用B淋巴细胞为目标亚群。B淋巴细胞占外周全血中总白细胞群的5%。因而，该实施例证明了本发明的实施方案也可用于细胞类型混合物样品中占比例细胞计数的5%或更多的其他白细胞亚群。这类情况包括但不限于：(1) 外周全血样品中的自然杀伤细胞；(2) 外周全血样品中的CD4T细胞；(2) 外周全血样品中的CD8T细胞；(3) 外周全血样品中的单核细胞；(4) 外周全血样品中的中性粒细胞/粒细胞；(5) 外周全血样品中的嗜酸性粒细胞；(6) 外周血单核细胞制备物中的自然杀伤细胞；(7) 外周血单核细胞制备物中的CD4T细胞；(8) 外周血单核细胞制备物中的CD8T细胞；(9) 外周血单核细胞制备物中的单核细胞；和(10) 外周血单核细胞制备物中的B细胞。

[0063] 本发明各种实施方案的优点可包括，但不限于：(1) 无需预先进行细胞分离、富集或分选；(2) 无需获得每单位容积的每个造血亚群或白细胞亚群的细胞计数；(3) 可用于定量PCR情况，其中在每种样品中仅定量少数转录本；(4) 可用于通过各种方法定量多个亚群特征性基因(靶基因、参照基因或二者)的情况，包括但不限于，定量PCR、微阵列、测序或其他分子计数方法如数字PCR；和(5) 可用于从微阵列产生的大规模转录丰度数据，其中在每个样品中定量了数百至上万的转录本。

[0064] II. 方法

[0065] 在细胞混合物样品中，亚群特征性基因被定义为这样的基因，其至少50%(或其他更高的百分比)的转录本总量由特定的细胞类型亚群提供。本发明实施方案可基于不同情况使用标准来选择这类亚群特征性基因，例如，当可得到特定的细胞类型成分和细胞混合物样品的表达水平数据时(如图3所示)，或者当仅可得到特定的细胞类型成分和一些其他细胞成分但不能得到细胞混合物样品的表达水平数据时(如图4和图5所示)。

[0066] 只要选定了亚群特征性靶基因，就可定量细胞混合物样品中其相对于亚群特征性参照基因的表达水平(例如，TA)。由于这两种基因的转录本形成了细胞混合物中的大多数，混合物中的该表达水平可以如分离的亚群中表达水平的生物标志使用。定量TA的方法包括但不限于，定量PCR、微阵列、测序或杂交或其他适于定量基因转录本的方法。

[0067] 图2为说明根据本发明实施方案的方法200的流程图，该方法用于测量含有多个亚群的细胞混合物的第一亚群的基因组表达。所述第一亚群的细胞在测量前没有进行分离。相反，特定基因被确定为所述第一亚群特征性的基因，因此不是任何基因都可用于方法200。可以相对于也为所述第一亚群特征性的参照基因，将特征性靶基因的基因组表达水

平归一化。所述表达水平(其可表式为数字参数)可用作诊断的生物标志。下述一些步骤是在化验方法开发中部分实施的,而该化验方法的用户只需进行其余步骤。并非所有步骤在每个与方法200相关的实施方案中都被执行。其中一些或所有步骤可通过计算机系统执行。

[0068] 在步骤210中,接受了一个或多个样品的不同类型细胞(亚群)中哪个基因表达的数据。所述数据可获自采自多个生物体(例如,来自多个人或同样类型的动物或植物)的样品。所述数据可提供每个亚群的基因转录丰度(TA)。所述数据可通过分离亚群细胞并在随后定量所述转录本产生,例如通过本领域技术人员已知的方法,如定量聚合酶链式反应(QPCR)。计算机系统可使用所述数据来确定每个亚群转录本的相对和/或绝对定量。所述数据也可指定混合物中来自每个亚群的细胞成分。

[0069] 在一个实施方案中,步骤210测量基因组表达是由化验方法的生产者执行,而使用该化验方法的技术人员是不需要执行的。因而,仅进行一次分离以确定特征性基因,而在确定特征性基因后无需重复进行分离。在另一个实施方案中,数据可获自另外的来源,例如,科学论文公开的数据或作为试剂盒的部分的小册子中的数据,或通过由软件产生的显示。

[0070] 在一些实施方案中,仅获得了细胞混合物样品中第一亚群的基因表达数据。例如,可获得表示所述第一亚群中多个基因的表达的基因表达数据。另外,也可获得整个细胞混合物样品中这些基因的表达。在其他实施方案中,获得了细胞混合物样品中多个亚群的多个基因表达,但没有获得整个细胞混合物样品中的表达数据。

[0071] 在步骤220中,例如使用来自步骤210的表达数据,确定了多个亚群特征性基因。在细胞混合物中,亚群特征性基因被确定为这样的基因,其具有由所述第一亚群细胞提供的其转录本的至少的预定百分比。所述预定百分比等于或大于50%(例如,50%、60%、70%、80%、90%或小于100%的任何百分比)。例如,亚群基因的至少50%的转录本(例如mRNA)可由所述第一亚群的细胞提供。

[0072] 在一个实施方案中,确定可简单地为从列表中获得这些基因。例如,用于特定诊断检测的试剂盒或产品说明书可指定特定亚群的基因列表。因而,测定可由临床实验室技术人员和/或接收亚群特征性基因列表的软件执行。测定的基因可为一列基因中的所有基因或仅为其中一些基因(例如通过可基于待进行的诊断检测发生的一些选择机制),例如,下述步骤230和240中提到的基因。

[0073] 在另一个实施方案中,所述测定可能更加复杂。例如,所述测定可由设计检测的研究实验室执行,而一些后续步骤可由临床实验室执行(例如,如上所述,简单地接收列表)。在一个实施操作中,数据通过计算机系统进行分析以确定特定亚群细胞中不同基因相对于管家基因的表达水平,所述管家基因在许多细胞类型(例如细胞混合物中的所有细胞类型)中具有一致的表达。

[0074] 通过比较纯化的亚群细胞成分和细胞混合物样品,可将细胞混合物样品中特定细胞类型(例如所述第一亚群)所占的基因转录本百分比用来确定该基因的必要差别表达比值X。X为特定细胞亚群相对于细胞混合物样品的基因表达水平的倍数,例如,数值10是指由特定细胞类型表达的基因比细胞混合物样品中的表达高出10倍,图4的图表所示为特定细胞类型的各种百分比;图4在下文有更详细的描述。

[0075] 在步骤230中,例如,从亚群特征性基因列表中确定了亚群靶基因。通常,所述靶基因是已知或期望与特定状况例如疾病状态相关联的基因。在分析的开始,该关联性可能是

未知的。但一旦知道了该关联性，确定的靶基因即可用于检测中。在一个实施方案中，计算机系统可通过接收来自用户或输入文件中指定的数据选择靶基因，确定亚群靶基因。

[0076] 在步骤240中，确定了亚群参照基因。计算机系统可以与确定靶基因类似的方式确定参照基因。在一个实施方案中，亚群参照基因相对于所述亚群靶基因具有更低的生物学变异。该变异可仅存在于所述第一亚群或细胞混合物中。来自不同个体的各样品间的测量可用于测定变异。计算机系统可分析来自各种样品的数据，以确定参照基因和靶基因的表达水平的生物学变异，其可相对于管家基因进行确定。在各种实施操作中，所述变异可评估为变异系数 (coefficient of variation)，其可以小于200%、小于100% (或90%、80%、70%、60%、50%、40%、30%、20%、15%、10%或5%)。

[0077] 在一个实施方案中，可选择亚群靶基因和亚群参照基因，以便它们在特定亚群例如红细胞或血小板中不具有可观的表达水平。在另一个实施方案中，所述红细胞和血小板可假定为被从细胞混合物中过滤出来，因而在确定特征性基因时可将它们所提供的TA去除。

[0078] 在步骤250中，计算机系统确定细胞混合物中亚群靶基因的转录本的第一量和细胞混合物中亚群参照基因的转录本的第二量。所述第一量和第二量可提供这两个基因的转录丰度的定量。所述细胞混合物可以为采自患者的用于实验室分析的样品。

[0079] 计算机系统可基于计算机系统所接收的数据(例如，信号)确定所述量。例如，计算机系统可接收荧光信号(或从涉及细胞混合物的一个或多个反应检测到的其他信号)，并将这些信号转换为所述量(例如，使用CT技术)。所述荧光信号能被转换为数字信息，然后由计算机系统接收，计算机系统随后使用这些数字信息确定所述量。其他实例包括接收测序数据或微阵列数据的计算机。

[0080] 可通过本领域技术人员所知的许多机制测量基因转录本的量。可从反应(例如，PCR或测序反应)测量信号(例如，荧光信号)，且这些信号可用于确定转录本的量。例如，定量PCR可用于获得扩增曲线，其中CT法被用来从扩增曲线确定所述量。除了来自QPCR的荧光性，也可使用其他测量方法确定所述量，如微阵列、测序和杂交。测量可通过作为计算机系统一部分或沟通式连接至计算机系统的机器来执行。

[0081] 在步骤260中，从所述第一量和第二量计算参数。在一个方面，所述参数是所述第一量对所述第二量的相对值(例如，相对丰度)。例如，所述参数可仅为这两个量的比率。其他实例为复杂函数，以所述量为输入(例如作为各自函数的输入)，且函数的比率可用于确定该参数。该参数代表所述第一亚群中亚群靶基因的基因组表达量。该参数可能不是最准确的测量(例如，相对于来自分离的细胞成分的测量)，但其仍然能提供一个量度(如下述相关分析中所示)，且其测定更具效率。

[0082] 在步骤270中，所述参数被用作生物标志。例如，可将该参数与一个或多个临界值比较，以确定细胞混合物的分类。所述分类可包括细胞混合物(且因此，患者)是处于病变状态还是非病变状态。一个范围区间值可与非病变状态相关。另一个分类可为未分类的或不确定的。例如，特定范围可为不确定的；可针对未分类的样品进行另外检测。对于少数不确定的样品，接着可进行利用亚群的细胞分离的更为昂贵的技术，以便获得更为准确的测量。

[0083] 如上所述，外周血中靶基因的过表达增高，可能是由于：(A)一些特定细胞类型组分中该基因的平均表达得到增加，和/或(B)那些细胞类型组分的数目或比例得到增加。特

定细胞类型中该基因的平均表达可用作生物标志的生物学指数。然而，其被混合物样品中存在的特定白细胞亚群数目或比例混杂。尽管在没有预先进行细胞分离下，特定细胞类型中该基因表达值不能确切测定，如下述所示，本文公开的新参数与所述确切值相关联。因而，使用该参数获得的测量值是确切值的估计，且其测定更具效率。

[0084] 因而，本发明实施方案可免于预先进行细胞分离的要求，并允许直接测定外周全血中白细胞亚群特异性的基因表达。其潜在应用很广泛。例如，诊断自体免疫疾病、测定免疫抑制药物的效果、诊断感染性疾病。具体实例为区分发烧患者体内的感染病原体或者作为发烧患者体内的预后生物标志。当前的方法使用血液中的蛋白生物标志（如C反应性蛋白，CRP）区分具有严重感染的患者。然而，CRP并非特异性的。因此，全血中粒细胞的基因表达分析将是非常特征性的。其他具体实例包括用作评估移植后患者体内的免疫抑制剂治疗的药效动力学的生物标志，以及监测自体免疫疾病中的疾病活动。

[0085] 分析试剂盒可以开发用于定量细胞混合物样品中的细胞类型特征性靶基因和细胞类型特征性参照基因，从而确定细胞类型特异性基因表达水平。例如，本发明实施方案可提供直接白细胞亚群特异性转录丰度（直接LS-TA）检测试剂盒，其用于测定特定造血亚群中特定靶基因的转录丰度。该细胞类型特征性靶基因可为通路中的基因、响应外部刺激的基因或纯化的细胞群中发现的在基线情况下或经刺激后差别表达的基因。

[0086] 尽管本文中的主要实施例应用于人样品，本发明实施方案可用于其他物种。当存在多个细胞类型时，本发明实施方案也可应用于其他细胞混合物样品如组织活检样品。同样，具体转录本可能有所不同，例如，仅为小RNA或可选的拼接转录本。样品可为离体刺激样品。在这种情况下，将外周血样品与一些刺激物一起孵育指定长度的时间。刺激后进行基因表达分析以确定反应。

[0087] III. 确定特征性基因

[0088] 如上所述，本发明实施方案使用特征性基因获得特定细胞亚群特异性的表达水平的测量。特定亚群特征性的基因使得可省略对这些细胞的分离，以及直接在细胞混合物中进行测量以提供与分离的细胞中的测量值相关联的测量值。本部分描述确定特征性基因（例如，方法200的步骤220）的过程，其中所述过程可由检测试剂盒的生产商执行，从而实验室技术人员无需进行该分析。

[0089] 在一个实施方案中，特定造血亚群或目标细胞类型表达的基因转录本占所述基因在外周血样品或细胞混合物样品中转录本的总量的至少Y%（其中Y为等于或大于50%的预定百分比），这些基因在使用细胞混合物样品进行的特定造血亚群的转录丰度的定量是有用和特征性的。这些转录本被称为细胞类型特征性转录本，且来自亚群特征性基因。预定百分比Y的实例为60%、70%、80%、90%或100%，或50%至100%之间的任何百分比，可能包括50%。

[0090] 在各种实施方案中，确定特定亚群的特征性基因可使用细胞混合物样品（例如全血）中细胞类型成分的平均百分比，以及所有或大多数细胞类型亚群（例如粒细胞、B淋巴细胞等）或者仅一个细胞亚群和整个细胞混合物中基因的平均基因表达水平的了解。

[0091] A. 可取得亚群和细胞混合物的表达水平数据

[0092] 图3所示为根据本发明实施方案的关于细胞混合物样品的假想实例，当可进行特定亚群和目标细胞混合物样品的基因表达测量时，其用以确立用于选择细胞类型亚群特征性基因的特性。该实施例表示了可取得细胞混合物样品10（一个校准细胞混合物的实例）和

细胞分离后的一些细胞成分12、14或16的TA数据时的情况。细胞成分12–16为通过细胞分离过程从细胞混合物样品10获得的单个细胞类型(亚群)的实例。

[0093] 在所述细胞混合物样品10(例如,外周全血或全血(WB))中,假想存在3种类型细胞,其以不同形状包括“方形”18、“菱形”19和“圆形”17表示。如图所示,细胞混合物样品有20个细胞,其中每个形状实体代表一个细胞。为了简单起见仅描绘了3种细胞类型,但同样的原理适用于更多的细胞类型。

[0094] 在细胞混合物样品10的所有细胞组分中,每个亚群以一定范围内的已知比例的细胞计数存在。具体地,“方形”细胞18占样品中细胞的5%(20个细胞中的1个),“菱形”细胞19占样品中细胞的25%(20个细胞中的5个),并且“圆形”细胞17占样品中细胞的剩下70%(20个细胞中的14个)。使用这些比例是因为它们代表了人血液样品中各种造血亚群的典型比例。

[0095] “方形”细胞18为特定的目标细胞类型(亚群),其细胞类型特征性TA分析正在开发中。每个细胞成分样品12–16中显示的比例11a–11c表示转录本A(对应于基因A)对各个细胞类型中常规管家基因(HK)转录本的相对表达。这是定量基因表达水平(具体为TA)的金标法。在细胞混合物样品10的顶部显示了比例11d,其表示细胞混合物样品10中转录本A对转录本HK的整体相对表达。比例11d通过计算基因A的所有转录本(其为38)和管家基因的转录本(其为20—每个细胞为1)获得,从而提供比例38:20,其与19:10一样。这些比例11为实验中观察到的定量或表达数据。这些比例11也显示于每个细胞中。

[0096] 为了获得比例11a–11c,可分离细胞混合物样品10以产生3种细胞类型12、14和16中每种的3种纯化细胞成分的样品。对细胞成分样品进行转录本A相比管家基因的相对定量11。由于仅需要比例11a来测定亚群18的特征性基因,仅需分离方形细胞。

[0097] 在一个实施方案中,细胞混合物样品中的亚群特征性基因转录本总量的至少50%由特定的细胞类型亚群提供。可基于在细胞混合物样品10(比例11d)和特定细胞类型亚群成分样品12(比例11a)中观察到的两个表达数据提供标准,来选择这类亚群特征性基因。在全血用作细胞混合物样品的实例时,该临界值比例作为选择细胞类型特征性基因的标准被表示为X50(“方形”细胞:WB)。该临界值比例X50(“方形”细胞:WB)也适用于通过其他方法如微阵列或测序数据库获得的基因表达数据,其中进行强度数据的整体归一化。

[0098] 基于图3中的信息,19个转录本A可源自亚群18,且细胞混合物10中总共有38个转录本A。因而,细胞混合物中50%的转录本A由“方形”细胞类型18提供。因此,对于“方形”细胞类型18,转录本A满足细胞类型特征性转录本的标准,并且对应的基因为该亚群的特征性基因。

[0099] 该分析还显示,纯化的“方形”细胞样品12中转录本A对转录本HK的相对表达高出细胞混合物样品10的相对表达为10倍。其还表明,当细胞混合物中所述亚群包含至少5%的所述细胞时,纯化的目标细胞成分中具有高出10倍的相对表达的基因满足细胞类型特征性基因的要求。

[0100] B. 可取得亚群和混合物的表达时的通用计算方式(formulation)

[0101] 基于细胞混合物和分离的特定亚群样品中基因A和基因HK的相对表达,图3中构思的通用计算方式可确定某基因是否符合X50标准。可用于细胞混合物样品中细胞类型特征性表达分析的候选细胞类型特征性基因的选择,是基于特定细胞类型的细胞成分样品和细胞混合物样品之间的表达水平差别。提供了用于确定具有一定比例细胞计数P的亚群的临

界值表达水平的公式,以及显示不同比例细胞计数P的情况,细胞混合物中表达水平的X-倍差别相对于转录本百分比的关系的图表。

[0102] 变量Etotal是细胞混合物中转录本A对于转录本HK的相对表达水平(例如,转录丰度)。可使用任何合适的方法(例如,QPCR)从细胞混合物中测量Etotal。变量Ep是这基因相对于HK基因在特定细胞亚群(例如B淋巴细胞、粒细胞等)的表达水平。可从分离的特定亚群的细胞成分测量Ep。变量P是细胞混合物中该特定细胞亚群的比例细胞计数,变量X是特定亚群中选定基因的TA相比细胞混合物样品(例如全血)中TA的倍数差别,因此 $X=Ep/Etotal$ 。

[0103] 变量Eother是混合物样品(例如全血样品、外周血单核细胞样品中)存在的所有其他细胞的平均表达水平。Eother是基于其他亚群中的每个亚群的比例细胞计数的加权平均数,因而当各种其他细胞类型的表达水平不同时也适用。在下述方程式中Eother实际上并不需要测定,而是仅用于计算特定P的X50标准所需的X。

[0104] 所有表达术语(E)均提供为每个细胞的转录本的表达水平,且对应于转录本相对于常规通用管家基因的相对表达量。这些变量之间的相互关系如以下方程式:

[0105] $Ep = X * Etotal \quad (1)$

[0106] $Etotal = P * Ep + (1 - P) * Eother \quad (2)$

[0107] $Ep/X = P * Ep + (1 - P) * Eother \quad (3)$

[0108] $Eother = \frac{Ep \left(\frac{1}{X} - P\right)}{(1 - P)} \quad (4)$

[0109] 对于不同的P值,细胞混合物样品中X和占转录本总量的百分比之间的关系如图5所示,以下将进一步描述。在50%的靶转录本由特定亚群提供的特殊情况下, $P * Ep = (1 - P) * Eother$ (5),其中P*Ep是来自特定亚群的转录本量,(1-P)*Eother是来自其他亚群的转录本量,并且当二者均为50%时它们彼此相等。将等式(4)代入等式(5)可得:

[0110] $P * Ep = (1 - P) \frac{Ep \left(\frac{1}{X} - P\right)}{(1 - P)} \quad (6)$ 其提供了

[0111] $X = \frac{1}{2P} \quad (7)$

[0112] 因而,当指定细胞类型提供了细胞混合样本中转录本总量50%(X50标准)的时候,这两个样品中基因表达或TA的倍数差别X(即分离的亚群相对于细胞混合物)必然等于或大于 $(1/2P)$ 。例如,如果所述亚群占细胞混合物的5%,那么X必然至少为10,其如图3所示。这倍数差别的X50值被用作确定亚群特征性基因的标准。从图5中可读出由目标亚群提供的任何期望百分比的X值,其在下文将详细描述。

[0113] 图4A所示为表400,其为根据本发明实施方案的外周全血样品中造血亚群的结果。410列中列出了外周全血的不同亚群。420列所示为每个特定细胞亚群的比例细胞计数值(P),其获自标准图书和实验室参考范围。430列所示为满足来自亚群的基因转录本的50%或更多标准所需的倍数差别。正如所见到的,X的值等于 $1/2P$ 。

[0114] 图4B所示为表450,其为根据本发明实施方案的外周血单核细胞样品中造血亚群的结果。460列列出了外周血单核细胞的不同亚群。470列所示为每个特定细胞亚群的比例

细胞计数值(P),其获自标准图书和实验室参考范围。480列所示为满足来自亚群的基因转录本的50%或更多标准所需的倍数差别X。同样如同所见到的,X的值等于 $1/2P$ 。

[0115] 表格中的430列和480列可被用作确定细胞类型特征性基因的标准的临界值。例如,B淋巴细胞特征性转录本选自这样的转录本,其在B细胞成分样品中的表达高出外周全血平均全血样品中的表达至少10倍,如图4A中所示。这些值表示为X50,50%仅由单个亚群提供的X。换句话说,X50是当50%存在于细胞混合物样品中的TA是由特定的细胞类型提供时,该特定细胞成分的TA高于细胞混合物样品中TA的倍数差别。可用较高的百分比值可来提高在细胞混合物样品中进行细胞类型特异性表达测定。因此,60%、70%、80%、90%或更高可作为用于选择细胞类型特征性基因的标准的临界值。

[0116] 图5为图表500,其显示了细胞混合样品中来自特定细胞类型亚群的转录本百分比与特定细胞类型亚群成分的TA对比细胞混合样本的TA的倍数差别(X)的关系,所述细胞混合物样本为含有各种比例细胞计数0.05(5%)至0.5(50%)的亚群的细胞混合物样本。水平轴510是特定亚群中基因表达水平相对于细胞混合物中基因表达水平的X倍差别。垂直轴520是由特定亚群提供的转录本百分比。值0.5对应于50%的细胞混合物。用不同符号绘制的6条不同的线对应于具有不同比例细胞计数的不同亚群的关联性特性,尽管不是来自相同的样品。每个亚群具有不同的比例细胞计数P,如用相应符号表示的图例530所示。6条不同的线以相应符号表示0.05、0.1、0.15、0.2、0.3和0.5的比例细胞计数(P)。例如,线502表示的细胞类型特性为其比例细胞计数为0.05(5%),并且线501表示的另一细胞类型特性为其比例细胞计数为0.1(10%)。通常每条线均符合 $C\% (Y轴) = X * P$ 的形式。

[0117] 虚线说明了图5在确定细胞混合物中比例细胞计数为0.1(10%)和0.05(5%)的亚群的标准值例如X50中的应用,且其对应的X50值分别为5和10。例如,虚线的水平部分540从垂直轴上的0.5值开始,并与线501在X(显示于x-轴上)等于5的点处相交。然后水平虚线540与线502在X等于10的点处相交。由于线501和502表示以比例细胞计数0.1和0.05存在的细胞类型的关系特性,这两个值可对应于所需的X50临界值。当期望占更高的百分比时,相应的倍数差别X的值也可从图表500中获得。对于各种比例细胞计数值的不同亚群,当期望占其他百分比例如60%时,可使用该图表500通过在垂直轴的不同水平上绘制另一水平虚线来计算对应的X60。例如,对于比例细胞计数为0.05的亚群,X60为12,并且对于比例细胞计数为0.1的另一亚群,X60为6。由于所述线的通式为 $C\% = X * P$,可使用 $X = C\% / P$,其中C%表示为小数形式。C%=60%和P=0.1(线501)时,这提供 $0.6 / 0.1$,其得数为6,如图6所给定的。

[0118] C.仅两个亚群的表达

[0119] 在一些实施方案中,仅可获得特定细胞类型成分和一些其他细胞成分的表达水平,但不能获得细胞混合物样品的表达水平。以下讨论提出了使用亚群的表达水平确定基因是否满足成为亚群的特征性基因的标准的方法。

[0120] 图6所示为,根据本发明实施方案,当仅可获得两个或更多个亚群但不能获得细胞混合物的基因表达测量的情况。例如,仅可获得细胞成分样品但不能获得细胞混合物样品的TA数据。该数据仍然可用于确定亚群特征性基因。图6与图2的不同之处在于,对于细胞混合物样品20,没有进行基因表达的相对定量。而是,确定了经细胞分离过程后两个或更多个细胞类型成分中的转录丰度,且仅其中的两个成分具有任何可观的转录丰度。

[0121] 细胞混合物样品20(例如,外周全血)具有3种类型细胞,其以不同形状包括“方形”

28、“菱形”29和“圆形”27表示。同样，“方形”细胞28为目标细胞类型亚群，正在为其建立细胞类型特征性TA分析。使用同样的用于选择细胞类型特征性转录本的临界值标准，即目标细胞类型表达的基因转录本或RNA占细胞混合物样品20中该基因转录本的至少50%。细胞混合物样品20中，50%的转录本A是由“方形”细胞类型28提供。因此，转录本A满足“方形”细胞类型28的细胞类型特征性转录本的临界值标准。

[0122] 对细胞类型混合物样品20进行分选，以产生3种细胞类型(亚群)27、28和29的3种纯化的细胞成分样品22、24和26。可在这些细胞成分样品的至少一些中进行转录本A的相对定量，其包括纯化的方形细胞样品22、纯化的菱形细胞样品24、纯化的圆形细胞样品26。每个细胞内显示的比例21a-c表示特定亚群中转录本A对转录本HK的相对表达。这些比例21a-c是通过分离的细胞成分22-26的实验观察和获得的定量或表达数据。如同所见到的，亚群28的比例为20:1，亚群29的比例为4:1，且亚群27无可观的转录本A表达。

[0123] 这些比例表明，纯化的“方形”细胞样品22中的转录本A对转录本HK的相对表达高出“菱形”细胞样品24中的相对表达5倍。在这样的条件下，鉴于两个细胞成分的该比例细胞计数(即，亚群28对亚群29的1:5的比例细胞计数)，以及其他细胞类型(即，亚群27)中可忽略的表达，两种纯化的细胞成分之间具有高出至少5倍的相对表达差别的基因可用作亚群特征性基因。

[0124] 细胞类型特异性相对表达乘以目标细胞类型的比例细胞计数的函数可作为选择亚群特征性基因的标准。在图6中，“方形”细胞28和“菱形”细胞29的该函数值分别为 $20 \times 5\% = 1$ 和 $4 \times 25\% = 1$ 。一个使用的标准是，其他细胞类型的该函数值不应超过目标细胞类型的该函数值。另一细胞类型(例如，“菱形”细胞亚群29)中转录本A表达的上限值为目标细胞类型(例如，“方形”细胞亚群28)的比例细胞计数与表达同样转录本的其他细胞类型的比例细胞计数的比值的函数。该条件尤其适用于确定CD4T细胞和CD8T细胞的亚群特征性转录本相关。

[0125] 该临界值比值作为选择亚群特征性基因的标准被表示为X50 (“方形”细胞：“菱形”细胞)。该临界值比值X50 (“方形”细胞：“菱形”细胞)也适用于通过其他方法如进行了强度数据整体归一化的微阵列或测序数据集获得的基因表达数据。

[0126] D. 仅两个亚群表达时的通用计算方式(formulation)

[0127] 在一些实施方案中，如果基因仅由细胞混合物样品中的两个亚群表达，可取的是使用较高的临界值来选择细胞类型或亚群特征性基因。例如，X75表示特定基因四分之三的转录本是由特定目标细胞类型提供的，或信噪比为3。下述公式说明了在确定和选择亚群特征性基因的过程中，如何确定所需的表达同样目标基因的两个亚群之间的表达差别倍数。

[0128] 使Ep1为特定细胞亚群1，例如CD4T细胞中特定基因的转录丰度(TA)。P1为细胞混合物样品中亚群1的比例细胞计数。Ep2为特定细胞亚群2，例如CD8T细胞中特定基因的转录丰度。P2为细胞混合物样品中亚群2的比例细胞计数。在该实例中，X为亚群1相比较亚群2的分离的部分中基因的TA倍数差别，因而等于 $Ep1/Ep2$ 。C%是相对于细胞混合物样品中转录本的总量，来源自亚群1的理想百分比。

[0129] 亚群1产生的基因转录本量等于 $Ep1 \times P1$ 。亚群2产生的基因转录本量等于 $Ep2 \times P2$ 。为了确定对于比例细胞计数P1和P2的具体值X值，以及期望的C%，可使用下述方程式：

$$[0130] \frac{Ep1 * P1}{Ep2 * P2} = \frac{C\%}{(1-C\%)} \quad (8)$$

[0131] 因此,给定C%时的X值可通过下述方程式获得:

$$[0132] \frac{Ep1}{Ep2} = \frac{C\%}{(1-C\%)} * \frac{P2}{P1} \quad (9)$$

[0133] 例如,外周血样品中CD4和CD8T细胞典型的比例细胞计数分别为20%和10%。为特定细胞类型亚群分析(例如,直接LS-TA检测)指定所需的C%。例如,C%可指定为75%,以便由亚群1提供的75%的转录本作为亚群特征性基因的标准。以下标准可用来确定C%为75%(即X75临界值)的CD4和CD8T细胞亚群的亚群特征性基因。

[0134] 除CD4和CD8外,该基因不在其他细胞类型表达。那么,对于CD4特征性基因,X75(CD4对CD8)=0.75/(1-0.75)*0.1/0.2=3*1/2=1.5。因此,CD4细胞成分中具有高出1.5倍的差别化表达的基因可用作外周血直接LS-TA检测中的CD4特征性基因。对于CD8特征性基因,X75(CD8对CD4)=3*2=6。因此,CD8细胞成分中具有高出6倍的差别化表达的基因可用作外周血直接LS-TA检测中的CD8特征性基因。该实例说明了,在细胞混合物样品中,一些细胞类型中的亚群特异性基因表达较容易确定。例如,可得到更多的亚群特征性基因用于CD4亚群直接LS-TA检测。

[0135] E. 可得到所有亚群的表达水平的情况

[0136] 图7是根据本发明的实施方案的关于细胞混合物样品的另一个图示,当仅可进行亚群但不能进行细胞混合物样品的基因表达测量时,其用以建立用于选择细胞类型亚群特征性基因的特性。该假定的情况类似于图6,但此时靶转录本由所有3种细胞类型表达。

[0137] 细胞混合物样品30(例如,外周全血)具有3种细胞类型,其以不同的形状表示,包括“方形”38、“菱形”39和“圆形”37。同样,“方形”细胞38为目标细胞类型亚群,正在为其建立细胞类型特异性TA检测。在细胞混合物样品30中,大于50%的转录本A由“方形”细胞类型38提供。因此,转录本A满足“方形”细胞类型38的细胞类型特征性转录本的临界值标准X50。

[0138] 细胞混合物样品30可分离为成分32、34和36。每个成分的转录丰度和比例细胞计数可用于确定是否整个细胞混合物样品中至少50%的转录本来自亚群38。在本实例中,确定了相对于HK基因的比例31a-c,即亚群38(比例细胞计数P(1)为5%)为35:1(Ep(1)),亚群39(比例细胞计数P(2)为25%)为4:1(Ep(2)),和亚群37(比例细胞计数P(3)为70%)为1:1(Ep(3))。为了确定是否满足X50标准,Ep(1)*P(1)≥Ep(2)*P(2)+Ep(3)*P(3),或更一般地:

$$[0139] Ep(1) * P(1) \geq \sum_{i=2}^N Ep(i) * P(i) \quad (10)$$

[0140] N为对应于表达各个基因的亚群数目的整数。对于高于50%的标准,可相应地修改上述公式。例如,对于75%,可在右边乘以系数3。通常对于临界值C%,可用系数C%/(1-C%)乘以方程式(10)的右边。

[0141] 图3、6和7中的实例同时适用多个样品。然后可以对样品组取平均值(或其他统计值)。平均值可用于确定平均来说特定基因是否为特征性的。例如,可确定样品组中,某亚群

是否至少表达了预定百分比的基因转录本。

[0142] F. 确定参照基因

[0143] 亚群的特征性参照基因可选定为在一组亚群特征性基因中具有最低生物学变异的特征性基因。所述亚群特征性基因组不必为特定亚群的所有特征性基因的完整组。在一个实施方案中，所述生物学变异要求低于临界值。可通过评估该组内每个特征性基因的生物学变异，然后选出具有最低变异(或可能是具有比靶基因更低的变异的任何参照基因)的特征性基因来确定参照基因。可使用多于一个特征性基因作为参照基因，其中参照基因组被有效地被看作是单个基因。然后将靶基因的TA与参照基因组的TA比较以确定方法200的步骤260中的参数。每当检测到任何一个参照基因的转录本，参照基因组的TA将增加。因而，像ddCT的可选表达，参照基因中一些可归类为一个归一化因子，例如通过取多个参照基因TA的几何平均数。

[0144] G. 确定靶基因

[0145] 通常，靶基因是已知或期望与特定状况例如疾病状态相关联的基因。在分析开始期间，该关联性可能是未知的。但一旦知道关联性，确定的靶基因可用于诊断中。可按与可能使用多于一个参照基因类似的方式使用多于一个靶基因(即，靶基因组)。

[0146] IV. 参数计算

[0147] 可使用亚群的靶基因表达水平相对于参照基因表达水平的某种数学表达式确定步骤260中计算的参数。基于该参数与特定细胞类型亚群中分离的成分的表达水平的关联性，该参数可用于评估特定细胞类型亚群(例如B淋巴细胞)中该靶基因的表达。该参数可为累加的指数，其中将多个参照基因或靶基因的TA合并为一个单一量，然后作为所述第一量或第二量用于数学表达式中来确定参数。

[0148] 已经证明该新参数适用于直接LS-TA，如两个基因或两组基因的转录丰度比率，当二者均满足X(50)标准或更好时，该新参数与纯化后的特定白细胞亚群中靶基因的转录丰度显著相关。下文讨论的图9显示了该关联性的实例。因此，本文的实施方案可测定白细胞亚群特异性基因表达，而无需富集或纯化所述目标白细胞亚群。

[0149] 本发明实施方案也可涉及定量至少两个基因(包括细胞类型特征性靶基因和另一个细胞类型特征性参照基因)或两组这类基因的转录丰度，用于定量目标靶基因的相对表达。两个基因或两组基因的定量可在细胞混合物样品如外周全血样品中进行。所述两个基因或两组基因可按这样的方式选择，其满足X50标准或更好。

[0150] 可将所述参数与一个或多个临界值比较以确定细胞混合物的分类。所述临界值可从已知为健康或者说正常的个体样品确定。所述分类可以包括所述细胞混合物是处于病变状态还是非病变状态。例如，低于临界值(cutoff)的范围值可能与疾病状态相关，并且高于临界值的范围值可能与非病变状态相关，或者反之亦然。

实施例

[0151] V-1. 实施例1(B淋巴细胞)

[0152] 所提供的下列实施例仅用于说明的目的，而非限制本发明。本领域技术人员将很容易理解，可对多个非关键性参数进行更改或修饰以产生基本上相同或类似的结果。

[0153] A. 制备

[0154] 血液样品采集自22名健康志愿者。对所有样品进行处理以通过磁珠细胞分离收集B淋巴细胞。在以下材料中进行候选参照基因的定量：(1)用Tizol处理的外周全血，(2)收集于PAX管中的外周血，(3)通过聚蔗糖(Ficoll)制剂收集的PBMC，和(4)通过两步法逆转录实时PCR纯化的B淋巴细胞。

[0155] 尽管红细胞和血小板都不是有核细胞，其均含有基因转录本。然而，相比有核的造血亚群即白细胞亚群，二者仅有少量的基因的转录本。在一个实施方案中，可将外周血中的基因和由无核细胞群表达的转录本滤除。该步骤仍然留有不同亚群的细胞混合物。尤其是当细胞类型特征性靶基因和细胞类型特征性参照基因均以高水平在红细胞或血小板中表达时，可进行该滤除。可挑选亚群靶基因和亚群参照基因，以便其在红细胞或血小板中不具有可观的表达水平。

[0156] B. 确定具有选定的差别表达值(X)的基因

[0157] 在每个样品制备物中都进行常规管家基因的定量。在每个样品制备物中都进行候选白细胞亚群特征性参照基因的定量。通过 $\Delta\Delta CT$ ($\Delta - \Delta$ 临界值循环)法确定所有样品的相对丰度。

[0158] 使用成对的跨外显子引物进行基因表达分析。用Roche Transcriptor First Strand cDNA试剂盒(Roche)进行逆转录。用Roche480仪器进行实时定量PCR。用 $\Delta - \Delta$ 临界值循环($\Delta\Delta CT$)法确定测试样品相对于标准样品的靶基因表达。标准样品是提取自两名志愿者的合并的全血样品。因此，该 $\Delta\Delta CT$ 被用于计算B细胞成分的靶基因表达高于全血中其表达的倍数差别。以用于分析纯化B细胞的TA的常规方式将GAPDH用作通用管家基因。

[0159] 通过QPCR分析表I中的下述基因，以确定B细胞亚群的特征性基因。

基因	RefSeq ID	引物	产物大小
APOBEC3F	NM_145298	AATTATGCATTCTGCACCG CCATAGGCTTTCGGTAGSGTT	110
DCAL1	NM_172004	AAAACTGTTCGGACTTCccc ATGCACCTTCAGTCTTGG	111
FBXO35	NM_032736	TACCTTCCCTGGTGTGCT CAGCTGCTCTCCCTAGTGC	101
GAPDH	NM_002046	CAATGACCCCTCATGGACC GACAAGCTTCCCGTTCTCAG	106
KCNG1	NM_002237	ACCTCTCCGTCAAGCACCT AGGAGGAACCTCCAGGGAGAA	127
KLHL14	NM_020805	CTCCCCAGCAATTGGTTC AGTTTCCACCTCCACAAACG	109
LAF4	NM_002285	CCAAGCTCTCAAGTTAGTC ACTTTGCCAGGTGCTTGAAT	136
Prickle1	NM_001144881	TGCAGAACTGCTCAAACCCAC GTTTCACACTCAAGGCACCA	120
SNX22	NM_024798	GCTTGGAGGCTTACATCCAG AGTTGCTAGCCTGGGGTCT	111
TNFRSF13C	NM_052945	GTGGGTCTGGGAGA6CTGG GATTCCCGGAGACAGAAATGA	126

[0161] 表I:带有ID、引物和产物大小的基因列表。

[0162] 比较分离的B细胞样品中与全血样品中候选的B细胞亚群特征性基因的平均差别表达,如图8A的表800中所示。因为在该实例中用于计算ddCT的标准样品为来自两名个体的合并的全血样品。在纯化的B细胞样品观察到的常规ddCT(靶基因相对于GAPDH)代表了倍数形式的差别表达的幅度。在一个实施方案如本实例中,可很方便地确定差别表达的幅度。因此,均值和标准差(St.Dev.)的单位为高于全血样品的倍数单位。CV%是指样品间的变异系数(coefficient of variation)。显示了针对两个标准X50和X80的测试结果。注意,是或否取决于混合物中B细胞群的细胞计数。对于KLHL14和TNFRSF13C,其超过全血的平均倍数差别的95%置信区间上限达到了X50标准,因此,也选用了它们。

[0163] 占平均外周全血样品转录本的50%的基因转录本(即达到X50标准,其要求B细胞成分样品中的表达高于10倍)包括KCNG1、KLHL14、刺痛感蛋白(Prickle)和TNFRSF13C。有两个基因达到了X80标准:DLCAL1和FREB。X80表示平均外周全血液样品中80%的转录本由B细胞亚群表达的转录本提供;B细胞成分中的表达需要至少16倍高的倍数差别。这些X50和X80标准基于外周全血样品中B淋巴细胞的5%的平均百分比。这些基因转录本在其他白细胞亚群中以较低的水平表达,包括CD4T细胞、CD8T细胞和粒细胞。因此,它们是B细胞亚群特异性的特征性基因。

[0164] C.选择参照基因

[0165] 本发明实施方案可用新的方式来表示细胞混合物样品中的亚群特异性表达水平。没有使用常规定义的通用管家基因作为内部参照来定量测定靶基因的转录丰度,而是使用新类型的内部参照基因。本文称其为亚群特征性参照基因,如B淋巴细胞特征性参照基因。

[0166] 例如,可从特征性基因的候选名单中挑选白细胞亚群特征性参照基因,这基于其转录本相对丰度在亚群中有低水平的样品间变异量(between-sample variance),例如B淋巴细胞样品。然后该白细胞亚群特征性参照基因可用作细在胞混合物(全血)样品中与另一靶基因的参照,以测定所述靶基因的白细胞亚群特异性表达。

[0167] 这些亚群(例如,白细胞)参照基因能够满足两个标准:(1)亚群中的表达标准,例如,X50或更高的标准,如上所述;(2)各个个体(例如,不同人)间的低生物学变异或变异量(variation或variance)。在开发B淋巴细胞亚群特异性基因表达方法的实例中,图8B所示为基因(如从图8A的表800中所确定)的表850,所述基因为满足X50标准或更好的基因:KCNG1、KLHL14、刺痛感蛋白、TNFRSF13C、DLCAL1和FREB。

[0168] 表850显示了从左至右按升序排列的亚群特征性基因的CV%(生物学变异)。显示了使用GAPDH作为通用管家基因的常规基因表达如 $\Delta\Delta CT$ 。TNFRSF13C具有最小的生物学变异。因此,其在接下来的实施例中被选定为亚群参照基因。KLHL14和FREB也可用于其他应用的潜在内部参照基因。在其他情况下,多于一个细胞类型特征性或白细胞亚群参照基因的累加指数也可用作内部参照。在相对基因表达分析的应用中,有各种方式和公式来表示多个参照基因(Vandesompele et al. (2009) Chapter4. Reference gene validation software for improved normalization. In Real-Time PCR: Current Technology and Applications. Editors Julie Logan, Kirstin Edwards, Nick Saunders. Publisher: Caister Academic Press; 1 edition)。

[0169] D.靶基因的定量

[0170] 亚群特征性靶基因和亚群特征性参照基因可用于直接确定外周全血样品中的白细胞亚群特异性表达。通过 $\Delta\Delta CT$ 法，根据全血(细胞混合物)样品中靶基因相对于白细胞亚群特征性参照基因的相对丰度进行的靶基因定量，反映了或与在纯化的B细胞样品中通过 $\Delta\Delta CT$ 法确定的该靶基因相对于常规管家基因的相对丰度相关联。因而，通过使用合适的靶基因和合适的细胞类型特征性参照基因，可以确定全血(细胞混合物样品)的单个白细胞亚群中靶基因的相对丰度，而无需预先进行细胞分离。

[0171] 在本实施例中，KCNG1被选定为亚群靶基因。因此，外周血的B淋巴细胞中KCNG1的表达水平即是所需的值。将常规使用纯化的方法与直接测量细胞混合物中亚群靶基因和参照基因的相对丰度的方法进行了比较。通过直接在外周全血液样品中进行基因表达分析，获得了新的相对基因表达参数。使用与上文所述相同的引物。用Roche480仪进行实时定量PCR，将 $\Delta-\Delta$ 临界值循环($\Delta\Delta CT$)法用作亚群特征性靶基因KCNG1相对于亚群特征性参照基因TNFRSF13C的基因表达水平或TA的新参数。

[0172] 图9为图表900，其显示了直接LS-TA、 $\Delta\Delta CT$ (具体地，外周全血样品中的KCNG1相对于外周全血样品中的TNFRSF13C的相对丰度)的该新参数和常规LS-TA $\Delta\Delta CT$ (纯化的B细胞亚群中的KCNG1相对于纯化的B细胞亚群中的GAPDH)之间的相关性。沿着水平轴，相对于纯化的B细胞成分中的管家基因表示靶基因(KCNG1)的相对表达。靶基因的相对表达(获自22名志愿者的纯化的B细胞样品中，KCNG1使用GAPDH作为内部参考)被看作用于确定新的直接LS-TA参数的准确性的金标法。

[0173] 决定系数(R^2)超过0.5($p<0.01$)。在一些实施方案中，数值0.5被用作临界值以确定特定直接LS-TA检测是否提供了足够的准确性，以用于检测亚群中亚群靶基因的表达。在其他实施方案中，直接LS-TA检测的性能可通过相关系数(r)的统计显著性来评估。图表900显示了 $\Delta\Delta CT$ (外周全血样品中KCNG1/TNFRSF13C)和 $\Delta\Delta CT$ (纯化的B细胞亚群中KCNG1/GAPDH)之间的线性关系。作为该相关性的结果，本发明实施方案可通过直接分析外周全血的细胞混合物样品中的基因表达，确定白细胞亚群样品中选定靶基因的基因表达。

[0174] 本实施例使用了外周血中B淋巴细胞的基因表达，其占外周全血(WB)样品中细胞组分的5%。因此，该方法显示对代表外周全血样品中约5%的所有白细胞的白细胞亚群是可行的。因为结果表明了直接LS转录丰度分析(Y-轴)和通过常规细胞分选分离的细胞中的基因表达(X-轴)之间的显著相关性，这证明无需预先分离B淋巴细胞，即可实现对B淋巴细胞的KCNG1基因表达的评估。本文中描述的实施方案也适用其他细胞混合物样品类型，包括但不限于实体器官的组织活检样品。

[0175] 以下包含有更多应用全血直接LS-TA检测来确定血液亚群中的基因表达水平的实施例。表II、III、IV和V提供了根据不同的目标亚群，可用于全血直接LS-TA检测的候选亚群特征性基因。

[0176] V-2. 实施例2(总T淋巴细胞)

[0177] 关于方法和对象，本实施例中包含了30名征得同意的健康个体的血液样品。使用了两种样品，其为(1)采集后立即用Trizol试剂(Invitrogen，美国)处理的外周全血，和(2)从每名个体分离的CD3总T细胞(或富集的)样品。对于CD3+ve总T细胞亚群的分离，从新鲜采集的全血样品制备PBMC。然后加入CD3微珠(Miltenyi Biotec)。按照生产商方案的指导孵育后，洗涤混合物并使其通过LS柱(Miltenyi Biotec)内的磁场。进行另外多轮柱内洗涤后，

在移除磁场后通过洗脱收集CD3+ve T细胞亚群的富集/纯化的样品。通过双色免疫荧光流式细胞术分析等分的富集样品以确认CD3+ve T细胞的富集。剩余部分用Trizol试剂处理并于-80°C保存备用。使用该磁性分选法，亚群富集的样品中CD3+ve T细胞的百分比超过80%。

[0178] 下述候选基因的TA经逆转录后通过实时QPCR检测。遵循实施例1(B淋巴细胞)部分和部分VI. 实验步骤中描述的一般实验步骤。没有如实施例1(B淋巴细胞)那样使用以Trizol试剂处理的合并的全血液样品作为标准样品用于计算ddCT，而是使用了另一种血液样品。富集的T细胞的差别化表达不能通过富集的T细胞样品的ddCT值直接揭示。因此，高于全血样品的差别表达是单独地针对每个参与者来确定，其为富集的细胞样品的ddCT值对全血液样品的ddCT值的比率。然后从这些比率计算概括性统计值(均值和标准差)。

[0179] 通过QPCR分析了4个常规的管家基因(ACTB、GAPDH、RPS18、RPL31)，它们的结果通过取4个 Δ -CT值的几何平均值(GeoMean)进行概括。使用测试样品和对照样品间的CT值差别通过PCR效率校正法计算基因的 Δ -CT(dCT)值。

[0180] 为了确定CD3+ve T淋巴细胞亚群的亚群特征性基因，计算了CD3+ve T细胞富集的样品中候选基因的TA(表示为候选基因相对于4个管家基因的几何平均值的ddCT)对来自相同个体对应全血样品中的那些基因的TA的比率。候选的CD3+ve T细胞亚群特征性基因的这些相对表达比值(X,倍数)如图10A的表1000所示。除PRKCQ外的所有候选基因均满足X50标准，其要求T细胞亚群中的表达比全血中的表达高2倍。CD+ve T细胞亚群的富集样品中的生物学变异如图10B的表1050所示。且GIMAP7具有低的生物学变异并被用作亚群参照基因。

[0181] 对这7个基因(除不满足X50标准的PRKCQ和用作亚群参照基因的GIMAP7外)进行全血直接LS-TA检测。确定了来自全血直接LS-TA检测和分离的CD3+ve T细胞中对应的基因表达水平(TA)的结果之间的相关性。5个基因的细胞类型特异性表达与通过全血直接LS-TA检测得到的结果具有强正相关性，其决定系数(R^2)大于0.5，代表了大于0.7的相关系数(r)。直接LS-TA检测和亚群分离后(金标法)确定的基因表达之间的关联性如图10C的表1060所示。

[0182] 全血直接LS-TA检测和分离的CD3+ve T细胞中对应的基因表达水平(TA)之间的关联性如图11A和11B所示。图11A显示了使用GIMAP7作为亚群参照基因的MAF基因全血直接LS-TA检测和分离的CD3+ve T细胞中MAF表达水平之间的关联性的结果。两个分析之间显示了强的关联性， $R^2=0.61$ 。

[0183] 两个或更多个亚群参照基因可同时用于所述全血直接LS-TA检测。在本实施例中，计算了GIMAP7和LRRN3的dCT值的几何平均值(GeoM1)。然后，基于该值确定MAF的相对丰度。图11B显示了该结果与金标法分析结果的相关性。获得了改善的决定系数(R^2)，现为0.6649。

[0184] V-3. 实施例3(粒细胞)

[0185] 关于方法和对象，本实施例中包含了8名征得同意的健康个体的血液样品。在本实施例中使用了两种样品，其为(1)采集后立即用Trizol试剂(Invitrogen, 美国)处理的外周全血，和(2)从每名个体分离的粒细胞(或富集的)样品。按照标准方案通过聚蔗糖溶液从其他有核细胞中分离粒细胞亚群。室温静置并离心后，从聚蔗糖-RBC交界面收集粒细胞。洗涤几轮并离心后，收集小球(pellet)形式的富集/纯化的粒细胞亚群样品，并用Trizol试剂处理并于-80°C保存。

[0186] 下述候选基因的TA经逆转录后通过实时QPCR检测：AQP1、BCL2A1、CEACAM6、

CEACAM8、Clorf24、CSF2、DEFA4、FCGR3B、IL-4、IL-8、IL8RB、LTF、MMP25、MS4A3、RNASE3、SOD2。遵循实施例1(B淋巴细胞)部分和实验步骤中描述的一般实验步骤。通过QPCR分析4个常规的管家基因基因(ACTB、GAPDH、RPS18、RPL31)，它们的结果通过取全血和分离的粒细胞样品的常规TA分析的4个 Δ -CT值的几何平均值(GeoMean)来进行概括。

[0187] 为了确定粒细胞亚群的亚群特征性基因，计算了粒细胞富集的样品中候选基因的TA(表示为候选基因相对于4个管家基因的几何平均值的ddCT)对来自相同个体对应全血液样品的那些基因的TA的比值。候选的粒细胞亚群特征性基因的这些相对表达(X,倍数)如图12A的表1200所示。所有候选基因均满足X50标准，其要求分离的粒细胞样品中亚群基因的TA等于或高于全血中的TA。粒细胞亚群的富集样品中的生物学变异如图12B的表1250所示。且AQP1和SOD具有低的生物学变异并被用作亚群参照基因。

[0188] 对这些基因进行全血直接LS-TA检测(除AQP1和SOD2外，因为它们被用作亚群参照基因)。确定了全血直接LS-TA检测和分离的粒细胞中对应的基因表达水平(TA)的结果之间的关联性。5个基因的细胞类型特异性表达与全血直接LS-TA检测的结果具有强正相关性，其决定系数(R^2)大于0.5，表示相关系数(r)大于0.7。关联程度如图12C的表1260所示。这4个基因可通过直接LS-TA法进行分析，其为CEACAM6、CEACAM8、CSF2、IL8和RNASE3。图13的图表显示了RNASE3的直接LS-TA检测(y轴)和使用纯化的粒细胞样品通过金标法定量的相同基因(x轴)之间的关联性。

[0189] V-4. 实施例4(微阵列数据)

[0190] 在上述总T淋巴细胞的全血直接LS-TA的实施例中，使用GIMAP7和LRRN3作为亚群参照基因来测量CD3+ve总T细胞的基因表达。其通过实时QPCR分析验证。例如，使用实时QPCR，细胞混合物样品中特定亚群靶基因的TA可通过该亚群靶基因的dCT对亚群参照基因的dCT的比率来确定。本实施例为一个实施方案，其中可将微阵列数据用于直接LS-TA检测。

[0191] 通过微阵列平台测量的基因表达数据通常表示为信号强度，并且其通常进行对数转换后再呈现结果。在微阵列数据背景中的直接LS-TA检测是使用对数转换的同样原理的扩展。对数转换后， \log (分离的亚群中靶基因的信号强度)可与[\log (细胞混合物样品中亚群靶基因的信号强度)− \log (细胞混合物样品中亚群参照基因的信号强度)]相关联。

[0192] 在总T淋巴细胞的全血直接LS-TA的实施例中，GIMAP7和LRRN3已经表明为准确的亚群参照基因。此处，将微阵列数据集用于证明使用微阵列数据进行直接LS-TA的应用和性能。Laudanski等人通过微阵列研究了3种造血样品中的基因表达，包括细胞混合物样品(总白细胞)和2个亚群样品(富集的T细胞和富集的单核细胞)。可通过登录GSE5580在基因表达综合数据库(Gene Expression Omnibus)获得该数据库(Laudanski, et al. 2006. "Cell-specific Expression and Pathway Analyses Reveal Alterations in Trauma-related Human T Cell and Monocyte Pathways." Proceedings of the National Academy of Sciences of the United States of America 103 (42) (October 17) : 15564–15569.)。这些样品采集自7名对照对象和7名患者，且将来自每名个体的血液样品处理至3种细胞类型或细胞混合物以用于微阵列分析。微阵列数据归一化后，将信号强度数据进行对数转换。检索和分析T细胞亚群的实施例中研究的这些基因的数据，包括BCL11B、ITK、PRKCQ、GIMAP7、CD3G、GZMK、MAF、NELL2和LRRN3。然而，GIMAP7不符合该微阵列平台。因此，将LRRN3用作亚群参照基因。

[0193] 使用这些微阵列数据进行的直接LS-TA如图14A-14C所示。在图14A中，GZMK基因的

直接LS-TA计算为GZMK (206666_at) 和LRRN3 (取2个探针: 209840_s_at和209841_s_at的均值) 的信号强度的对数之间的差别。结果与如图14A中x-轴上所示分离的T细胞样品中的表达具有强相关性 ($R^2=0.65$)。MAF (206363_at) 和NELL2 (206666_at) 基因的直接LS-TA检测的进行也都是相当的, 如图14B和14C所示。这些结果证明直接LS-TA很容易扩展至微阵列数据。且其也同样适用于通过测序方法如采用大量平行测序技术的RNA-seq产生的转录组数据。

[0194] VI. 实验步骤

[0195] 以下为用于上述实施例的一些实验步骤。这些步骤的执行可用于不同的实施方案, 且对可能使用的其他步骤没有限制性。

[0196] 对于PBMC样品的制备, 将血液样品与等量的PBS缓冲液混合, 然后向其中缓慢加到3ml的Ficoll-Paque plus上 (GE Health, Cat. No. 17-1440-02)。将管在无制动下以500x g离心25min。将 (PBMC) 中间层收集至新的15ml管中, 并用PBS洗涤两次。然后以100x g离心10min。将管底部的PBMC团转移至具有750μl Trizol LS试剂 (Invitrogen, Cat. No. 10296-028) 的2ml微量离心管。

[0197] 对于CD4/CD8/CD14/CD19亚群成分样品的分离, 将50μl CD4/CD8/CD14/CD19微珠 (Miltenyi Biotec, Cat. No. 130-045-101/130-045-201/130-050-201/130-050-301) 加入至如上所述制备的PBMC样品中。将管充分混合, 并于2-8°C震荡下孵育15min。孵育后, 通过加入10ml洗涤缓冲液 (PBS+1%FBS) 洗涤细胞, 并于室温以445x g离心10min。移除上清液后, 加入洗涤缓冲液至总体积为3ml以悬浮细胞。将LS柱 (Miltenyi Biotec, Cat. No. 130-042-401) 置于合适的MACS分离器的磁场中, 并首先用3ml分离缓冲液 (Miltenyi Biotec, Cat. No. 130-091-221) 将其平衡。然后使细胞悬浮液通过平衡的LS柱。以3ml的分离缓冲液洗涤柱3次。通过从磁性分离器移除LS柱, 并使用柱塞以5ml分离缓冲液洗脱, 收获特定亚群。以500x g离心5min获得细胞团, 重悬浮细胞团并用Trizol LS试剂处理, 并于-80°C储存直至进行分析。

[0198] 对于将外周血收集至PAXgene管, 从志愿者采集3ml全血, 并存储于PAXgene血液RNA管 (QIAGEN, Cat. No. 762165) 中。将其于-80°C储存直至进行分析。

[0199] 对于Trizol处理的全血样品, 将等分的300μl全血样品直接与900μl Trizol LS试剂混合。将其于-80°C储存直至进行分析。

[0200] 对于Trizol处理的RNA样品提取物, 使用QIAGEN RNeasy微型试剂盒 (Cat. No. 74106) 进行RNA提取。将样品均质化并于室温孵育样品后, 每500μl体积的Trizol样品加入60μl BCP。震荡并于室温静置2-3min后, 以12,000x g于4°C离心5min。将上层水相转移至新管, 然后加入300μl 100%的异丙醇, 并立即通过移液器吹打充分混合。将多达700μl的该混合物加入至置于2ml收集管内的RNeasy微型柱中。流过后重复加入样品直至将所有样品加入到柱内。将QIAGEN无RNA酶的DNA酶试剂组 (Cat. No. 79254) 用于DNA酶处理。将两轮40μl和30μl的预热的无RNA酶水用于洗脱。用Nanodrop系统测量RNA的浓度。通过凝胶检查RNA完整性。将RNA样品进行逆转录, 或将标记好的样品于-80°C的冰箱中储存备用。

[0201] 对于来自PAXgene血液样品的RNA提取物, 使用QIAGEN PAXgene血液RNA试剂盒 (Cat. No. 762174) 从PAXgene血液RNA管 (QIAGEN, Cat. No. 762165) 提取RNA。遵循试剂盒提供的标准方案。用Nanodrop系统测量RNA的浓度。通过凝胶检查RNA完整性。将RNA样品进行逆转录, 或将标记好的样品于-80°C的冰箱中储存备用。

[0202] 对于逆转录,用Roche Applied Science Applied Science Transcriptor第一链cDNA合成试剂盒(Cat No.04897030001)进行逆转录。在制备步骤中,将多达1 ug提取的总RNA与2.5uM寡聚(dT)、60uM随机引物和10ug/ml MS2RNA(Roche Applied Science Applied Science, Cat No.10165948001)混合。于65℃保持10min后,加入4μl Transcriptor RT反应缓冲液(5X)、1mM每种dNTP、0.5μl每种RNA酶保护抑制剂(终浓度20U)和Transcriptor逆转录酶(终浓度10U)。然后,于25℃保持10min,再于55℃保持30min用于cDNA合成。最后,通过加热至85℃保持5min使Transcriptor逆转录酶失活。一些实施例将Takara PrimeScript® RT试剂的试剂盒(Perfect Real Time)(Cat No.HRR037C,Takara,日本)用于反应混合物,其包括2μl 5x PrimeScript® 缓冲液、0.5μl 寡聚dT引物(终浓度25pmol)、0.5μl 随机引物(终浓度50pmol)和0.5μl PrimeScript® RT酶。然后加入6.5μl 总RNA,并于37℃孵育10min以合成cDNA,然后进行失活步骤。逆转录后,将cDNA于-80℃储存。

[0203] 在一个实施方案中,使用两步法逆转录定量PCR(rt-QPCR)。也可使用一步法rt-QPCR。可使用多个造血亚群特征性参照基因,如同在其他使用常规管家参照基因的rt-QPCR设置中。(Logan et al. 2009. Real-Time PCR: Current Technology and Applications (Book), Publisher: Caister Academic Press. Publication date: January 2009. ISBN: 978-1-904455-39-4)。也可使用其他分析方法,如方差分析(Tichopad et al. 2009. Design and optimization of reverse-transcription quantitative PCR experiments. Clin Chem. 55 (10): 1816-23)。

[0204] 在实时定量PCR的应用中,基因表达(转录丰度)的测量通常表示为靶基因(目标基因)的表达水平对管家基因的比率。管家基因是那些通常由各种组织或细胞以稳定水平表达的基因。

[0205] 在实时定量PCR的应用中,基因表达(转录丰度)的测量可归类为:(1)绝对定量,或(2)相对定量。使用给定浓度的基因转录本的校准标准(例如,每单位体积的拷贝数),通过比较测试样品和校准物之间的结果,有可能确定样品中存在的转录本拷贝数。另一方面,更常用的是相对定量。目标基因(也称靶基因)的表达水平(或TA)表示为相对于(作为比率)一个或多个管家基因(HK基因)的表达水平(TA)。为了校正靶基因和HK基因的扩增方案之间PCR效率的差别,使用了称为 $\Delta - \Delta CT$ (ddCT)的方法。在 $\Delta - \Delta CT$ 法中,将通过实时PCR获得的测试样品的临界循环值针对对照样品(其数据对所有测试样品通用)进行比较。因此, $\Delta - \Delta CT$ 结果反映了靶基因针对对照样品中该基因的相对表达水平。 ΔCT (dCT)是指一个基因在测试样品和对照样品之间的临界循环值的差别。

[0206] 对于实时QPCR,使用Roche Applied Science Applied Science实时PCR体系通过实时PCR定量基因表达。使用了Lightcycler480 SYBR Green I Master Mix试剂盒(Roche Applied Science Applied Science,Cat.No.,04887352001)。每个反应的总体积为10μl,包括5μl 2x Master Mix、0.5μl 每种引物(10uM)和4μl cDNA模板(逆转录步骤后进行10倍稀释)。在一些实施例中,使用了Takara SYBR Premix Ex Taq™(T1i RNase H)试剂盒(Cat No.HRR420D,Takara,日本)。所述PCR Master混合物包含:5μl 2x Premix Ex Taq™(T1i RNase H)、0.2μl 每种引物(终浓度0.2uM)、2μl cDNA和2.6μl ddH₂O,以配制10μl工作体积。所有QPCR反应均在自动平台Biomek® 2000和Biomek® NX自动系统(Beckman Coulter,

Fullerton, 美国加州) 准备。在Roche Light Cycler480仪上用系统相关软件(Light Cycler®480软件, v1.5) 进行实时PCR。按如下进行扩增的标准方案: 起初于95℃加热反应混合物5min, 然后进行如下循环: 95℃变性10sec, 60℃退火10sec, 以及72℃延伸10sec, 重复45次。热循环步骤后, 进行熔解程序以检测任何非特异性产物。将一系列连续稀释的cDNA样品用作标准对照, 梯度为4倍稀释至256倍稀释。此外, 10倍稀释的阳性对照和非模板RT阴性对照也包括在实时PCR中。所有反应均按一式两份进行。

[0207] 尽管实施例使用实时定量PCR执行。可使用定量测量RNA或DNA分子的任何其他方法。除了使用以效率校正来说明临界循环值(CT) 的比较性临界值法($\Delta \Delta CT$ 或 $ddCT$) , 也可使用其他方法, 包括但不限于, 绝对标准曲线法和相对标准曲线法(Pfaffl, M W. 2001. "A new mathematical model for relative quantitation in real-time RT-PCR." Nucleic Acids Research 29 (9) (May1) :e45) 和Pfaffl et al., 2002. "Real-time RT-PCR quantification of insulin-like growth factor (IGF)-1, IGF-1receptor, IGF-2, IGF-2receptor, insulin receptor, growth hormone receptor, IGF-binding proteins1, 2and3in the bovine species." Domestic Animal Endocrinology 22 (2) (April) :91-102)。

[0208] VII. 特征性基因的列举

[0209] A. 挖掘可公开获得的数据以确定亚群特征性基因

[0210] 另外的确定亚群特征性基因的方法为挖掘可公开获得的微阵列数据集。在样品实验中很少有数据集定量全血样品和造血亚群成分。一个实例是Julia et al 2009. "Identification of candidate genes for rituximab response in rheumatoid arthritis patients by microarray expression profiling in blood cells (通过血细胞中的微阵列表达谱确定风湿性关节炎患者体内的利妥昔单抗响应的候选基因)" Pharmacogenomics (药物基因组学) 10 (10) (October) :1697-1708. doi:10.2217/pgs.09.99。其可获自NCBI GEO, 网址www.be-md.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE15316。

[0211] 使用与如上所述相同的标准来选择B细胞亚群特征性基因, 其为X50 (B细胞对全血)。由于比例细胞计数为5%, 其要求B细胞成分中候选基因的表达水平是其在全血液样品中的10倍。下述表II提供了确定的B细胞亚群特征性基因名单。

[0212]

微阵列探针标识符	基因符号	NCBI 数据库中的核苷酸序列
GI_10835070-S	CD74	NM_004355.1
GI_11038671-A	CD79A	NM_021601.1
GI_11038675-A	CD79B	NM_021602.1
GI_11415027-S	TCL1A	NM_021966.1
GI_13904868-S	RPS29	NM_001032.2
GI_14916502-A	RPS24	NM_001026.2
GI_15431308-S	RPS7	NM_001011.2
GI_15812219-S	RPL31	NM_000993.2
GI_17017970-S	RPL26	NM_000987.2
GI_18641371-S	HLA-DRB3	NM_022555.3
GI_18641372-S	HLA-DRB4	NM_021983.3
GI_18641377-S	HLA-DOB	NM_002120.2
GI_21071007-S	TCN1	NM_001062.2
GI_23110953-S	CTSG	NM_001911.2
GI_23110986-A	MS4A1	NM_021950.2
GI_23110992-S	MS4A3	NM_006138.3
GI_23238191-S	TNFRSF17	NM_001192.2
GI_24797068-S	HLA-DQB1	NM_002123.2
GI_26665892-S	HLA-DRB5	NM_002125.3
GI_28178862-S	OSM	NM_020530.3
GI_28416954-S	AZU1	NM_001700.3
GI_31317226-S	EGR1	NM_001964.2
GI_31317229-S	FCER1A	NM_002001.2
GI_31542945-S	CXXC5	NM_016463.5
GI_31982886-S	GATA2	NM_032638.3
GI_32189367-S	IGJ	NM_144646.2
GI_32481214-S	CD19	NM_001770.3
GI_32484980-S	BANK1	NM_017935.2
GI_33469981-S	BLK	NM_001715.2
GI_34147598-S	FCER2	NM_002002.3
GI_38455401-S	LCN2	NM_005564.2
GI_39753969-S	CAMP	NM_004345.3
GI_39930348-S	KIAA0746	NM_015187.1

微阵列探针标识符	基因符号	NCBI 数据库中的核苷酸序列
GI_41197088-S	HLA-DQA1	NM_371812.1
GI_42544162-S	FREB	NM_032738.3
GI_42794770-A	TXND5	NM_030810.2
GI_4502446-S	BPI	NM_001725.1
GI_4502650-S	CD22	NM_001771.1
GI_4502662-S	CD37	NM_001774.1
GI_4503302-S	DEFA4	NM_001925.1
GI_4503548-S	ELA2	NM_001972.1
GI_4507176-S	SPIB	NM_003121.1

[0213] [0214] 表II:候选的B淋巴细胞亚群特征性基因

[0215] [0216] B.用E-TABM-633数据集得到的X75名单

也对可公开获得的基因表达数据集进行了数据挖掘,其中各种造血细胞亚群通过微阵列进行了分析(Watkins et al. 2009 "A HaemAtlas: characterizing gene expression in differentiated human blood cells." Blood 113 (19) (May7) :e1-9)。该数据集可获自:www.ebi.ac.uk/arrayexpress/experiments/E-TABM-633。

[0217] 对于在外周全血样品的直接LS-TA检测的设置中选择CD4T细胞亚群特征性基因,使用了X75(CD4细胞对CD8细胞),其表明CD4亚群特征性基因在CD4T细胞中所需的表达为其在CD8T细胞中的1.5倍高。滤除掉不由粒细胞表达的基因后,被定义为具有T细胞中的1/8或1/16,表III显示了候选的CD4T细胞亚群特征性基因的名单。与其他表中的特征性基因一样,给定亚群的靶基因和参照基因可选自该名单。

[0218]

基因符号	NCBI 数据库中的核苷酸序列
ABLIM1	NM_006720.3
BCL2	NM_000633.2
CCR7	NM_001838.2
CD5	NM_014207.2
HS.13262	Hs.13262
HS.567464	Hs.567464

基因符号	NCBI 数据库中的核苷酸序列
HS.579530	Hs.579530
LDHB	NM_002300.3
LEF1	NM_016269.2
LOC388344	XM_371023.4
MAF	NM_005360.3
MYC	NM_002467.3
PIK3IP1	NM_052880.3
PRKCA	NM_002737.2
TNFRSF7	NM_001242.3
TRAT1	NM_016388.2

[0219]

[0220] 表III: 如通过挖掘E-TABM-633数据集获得的以X75标准确定的候选CD4亚群特征性基因

[0221] 对于确定CD8亚群特征性基因, 使用X66 (CD8对CD4) 作为标准。这些基因在CD8细胞中的表达高出其在CD4细胞中的4倍。滤除不由粒细胞表达的基因后, 被定义为具有T细胞中的1/8, 且排除那些在NK细胞中的表达高出其在CD8T细胞中的2倍的基因, 表IV显示了CD8T细胞亚群特征性基因的名单。

基因符号	NCBI 数据库中的核苷酸序列
NKG7	NM_005601.3
DUSP2	NM_004418.2
KLRK1	NM_007360.1
GZMH	NM_033423.2
NCR3	NM_147130.1
CD8A	NM_001768.4
EOMES	NM_005442.2
CCL5	NM_002985.2
CTSW	NM_001335.2

[0222]

[0223] 表IV: 如通过挖掘E-TABM-633数据集获得的以X66标准确定的候选CD8亚群特征性基因

[0224] C. 用E-TABM-633数据集得到的X50的名单

[0225] 对于粒细胞(或甚至对于更具体的中性粒细胞),由于其为外周血中最普遍的细胞类型,当粒细胞的比例细胞计数为50%并且另一亚群为5%时,X50(粒细胞对另一亚群)能够满足亚群中相对于细胞混合物基因差别化表达水平为0.1倍。有超过1000个基因显示在粒细胞中的表达高出其在其他造血亚群中的2倍。仅前150个候选的粒细胞亚群特征性基因显示于下述表格中。该表中许多编码各种核糖体蛋白的基因为有前景的候选粒细胞亚群特征性参照基因。

基因符号	NCBI数据库中的核苷酸序列
CLC	NM_001828.4
FCGR3B	NM_000570.2
ALPL	NM_000478.2
MMP9	NM_004994.2
PGLYRP1	NM_005091.1
MME	NM_007288.1
CCR3	NM_178329.1
CMTM2	NM_144673.2
IL8RA	NM_000634.2
MMP25	NM_022718.2
CYP4F3	NM_000896.1
PBEF1	NM_005746.1
S100P	NM_005980.2
CAMP	NM_004345.3
IL8RB	NM_001557.2
PRSS33	NM_152891.1
TNFAIP6	NM_007115.2
TFF3	NM_003226.2

[0226]

基因符号	NCBI 数据库中的核苷酸序列
SLPI	NM_003064.2
DEFA3	NM_005217.2
OLIG2	NM_005806.2
DEFA1	NM_004084.2
NOV	NM_002514.2
FFAR2	NM_005306.1
IL5RA	NM_175728.1
ANXA3	NM_005139.1
SLC25A37	NM_016612.1
PROK2	NM_021935.2
IL8	NM_000584.2
CA4	NM_000717.2
LRG1	NM_052972.2
TCN1	NM_001062.2
LOC653600	XM_928349.1
KRT23	NM_015515.3
PANX2	NM_052839.2
PBEF1	NM_182790.1
GPR97	NM_170776.3
JL1R2	NM_173343.1
MXD1	NM_002357.2
KCNJ15	NM_002243.3
CEACAM8	NM_001816.2
TGM3	NM_003245.2
KCNJ2	NM_000891.2
GPR109B	NM_006018.1
ALOX15	XM_937556.1
IFIT2	NM_001547.3
GPR109A	NM_177551.3
LOC653117	XM_931656.1
VNN3	NM_001024460.1
LCN2	NM_005564.2
ORM1	NM_000607.1

[0227]

[0228]

基因符号	NCBI 数据库中的 核苷酸序列
CCL23	NM_005064.3
CEACAM3	NM_001815.1
INDO	NM_002164.3
CEBPE	NM_001805.2
FBXL13	NM_145032.2
ADM	NM_001124.1
DEFA4	NM_001925.1
C3ORF34	NM_032898.2
CACNG6	NM_031897.2
ROPN1L	NM_031916.2
PSG11	NM_002785.2
HIST1H2BE	NM_003523.2
HIST1H2BC	NM_003526.2
C1ORF24	NM_022083.1
CEACAM1	NM_001024912.1
ZDHHC18	NM_032283.1
REPS2	XM_942570.1
GPR44	NM_004778.1
IL1R2	NM_004633.3
FPR2	NM_001462.3
RNASE3	NM_002935.2
CEACAM6	NM_002483.3
SLC45A4	XM_944850.1
AQP9	NM_020980.2
LOC654103	XM_939368.1
ARG1	NM_000045.2
HS.193406	Hs.193406
EPHB1	NM_004441.3
PHOSPHO1	NM_178500.2
LOC653778	XM_929667.1
HS.524705	Hs.524705
IFIT3	NM_001549.2
MANSC1	NM_018050.2

[0229]

基因符号	NCBI 数据库中的核苷酸序列
C1ORF24	NM_052966.1
RNF24	NM_007219.2
REPS2	NM_004726.1
MGAM	NM_004568.1
GPR97	XM_936582.1
GNG10	NM_001017998.1
SOD2	NM_001024465.1
LTF	NM_002343.2
STX3	NM_004177.3
EMR3	NM_032571.2
NRBF2	NM_030759.1
HCG27	NM_181717.1
OR10G3	NM_001005465.1
IL1B	NM_000576.2
ALOX15	NM_001140.3
HS.482814	Hs.482814
MBOAT7	NM_024298.2
PSG9	NM_002784.2
LOC283547	XM_378454.3
CEACAM1	NM_001712.3
HS.131087	Hs.131087
HS.282800	Hs.282800
TNFRSF10C	NM_003841.2
PTGS2	NM_000963.1
IL1RAP	NM_002182.2
LOC642112	XM_936252.1
LOC651738	XM_944898.1
ACSL1	NM_001995.2
LOC552891	NM_004125.2
TRIB1	NM_025195.2
ADORA3	NM_000677.2
DGAT2	NM_032564.2
HSPA6	NM_002155.3

基因符号	NCBI 数据库中的核苷酸序列
LOC643313	XM_933030.1
FRAT2	NM_012083.2
FAM174A	NM_198507.1
HS.569349	Hs.569349
ASPRV1	NM_152792.1
OXER1	NM_148962.3
HS.576627	Hs.576627
HS.540131	Hs.540131
XPO6	NM_015171.1
RNF149	NM_173647.2
VNN2	NM_004665.2
MAK	NM_005906.3
IFRD1	NM_001550.2
LOC642684	XM_926137.1
SLC2A11	NM_001024938.1
LOC653503	XM_929619.1
VNN3	NM_078625.2
LOC651612	XM_940794.1
HECW2	NM_020760.1
SMPD3	NM_018667.2
LOC649154	XM_938223.1
STEAP4	NM_024636.1
HS.161796	Hs.161796
ABTB1	NM_172027.1
F2RL1	NM_005242.3
KIAA1324	NM_020775.2
DHRS9	NM_005771.3
TMEM154	NM_152680.1
ORF1-FL49	NM_032412.2
DPEP3	NM_022357.1
CXCL1	NM_001511.1
C19ORF22	NM_138774.2
TMCC3	NM_020698.1

[0230]

[0231] 表V:挖掘E-TABM-633数据集确定的前150个候选的粒细胞亚群特征性基因

[0232] D.用GSE5580数据集得到的X50的名单

[0233] 对于确定CD3+ve总T细胞的亚群特征性基因,X50 (T细胞对全血) 为2。其要求候选

的亚群基因在总T细胞成分中的表达高出其在全血中的2倍。Laudanski等人的微阵列表达数据集具有关于来自患者和对照的总白细胞样品(作为全血液样品的替代物)和富集的总T细胞样品的数据。其可获自:www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5580。有超过250基因满足该标准,作为候选的总T细胞亚群基因。仅前75个基因被列举于此。

探针 ID	基因符号	NCBI 数据库中的核苷酸序列
209795_at	CD69	NM_001781
205590_at	RASGRP1	NM_005739
208798_x_at	GOLGA8A	NM_015003
204891_s_at	LCK	NM_005356
205798_at	IL7R	NM_002185
203828_s_at	IL32	NM_004221
204777_s_at	MAL	NM_002371
213539_at	CD3D	NM_000732
211796_s_at	TRBC1 /// TRBC2	NG_001333
210972_x_at	TRAC /// TRAJ17 /// TRAV20	
206337_at	CCR7	NM_001838
210915_x_at	TRBC1	NG_001333
201656_at	ITGA6	NM_000210
211675_s_at	MDFIC	
202207_at	ARL4C	NM_005737
202524_s_at	SPOCK2	NM_014767
212400_at	FAM102A	
221558_s_at	LEF1	NM_016269
213958_at	CD6	NM_006725
217838_s_at	EVL	NM_016337
39248_at	AQP3	NM_004925
221756_at	PIK3IP1	NM_052880
201697_s_at	DNMT1	NM_001379
206150_at	CD27	NM_001242

[0234]

[0235]

211623_s_at	FBL	NM_001436
204960_at	PTPRCAP	NM_005608
200953_s_at	CCND2	NM_001759
213649_at	SRSF7	NM_006276
218764_at	PRKCH	NM_006255
212414_s_at	GLYR1 /// SEPT6	NM_015129
201030_x_at	LDHB	NM_002300
217871_s_at	MIF	NM_002415
217950_at	NOSIP	NM_015953
205133_s_at	HSPE1	NM_002157
201892_s_at	IMPDH2	NM_000884
212995_x_at	MZT2B	
200826_at	SNRPD2	NM_004597
209440_at	PRPS1	NM_002764
214439_x_at	BIN1	NM_004305
215091_s_at	GTF3A	NM_002097
203385_at	DGKA	NM_001345
202144_s_at	ADSL	NM_000026
221691_x_at	NPM1	NM_002520
201922_at	NSA2	NM_014886
200705_s_at	EEF1B2	NM_001959
200642_at	SOD1	NM_000454
203408_s_at	SATB1	NM_002971
212071_s_at	SPTBN1	NM_003128
217848_s_at	PPA1	NM_021129
221488_s_at	CUTA	NM_015921
217802_s_at	NUCKS1	NM_022731
205255_x_at	TCF7	NM_003202
217807_s_at	GLTSCR2	NM_015710
214280_x_at	HNRNPA1	NM_002136
207721_x_at	HINT1	NM_005340
56197_at	PLSCR3	NM_020360
200943_at	HMGN1	NM_004965
217969_at	C11orf2	NM_013265
201177_s_at	UBA2	NM_005499
207023_x_at	KRT10	NM_000421
209104_s_at	NHP2	NM_017838
201092_at	RBBP7	NM_002893
200064_at	HSP90AB1	NM_007355
210027_s_at	APEX1	NM_001641
201947_s_at	CCT2	NM_006431
212426_s_at	YWHAQ	NM_006826
201812_s_at	C4orf46 /// TOMM7	NM_019059
208697_s_at	EIF3E	NM_001568

216570_x_at	LOC100510735 // RPL29	
218495_at	UXT	NM_004182
200610_s_at	NCL	NM_005381
200982_s_at	ANXA6	NM_001155
211937_at	EIF4B	NM_001417
209503_s_at	PSMC5	NM_002805
201064_s_at	PABPC4	NM_003819

[0237] 表VI:通过挖掘GSE5580数据集以X50标准确定的候选CD3+ve总T细胞亚群特征性基因。

[0238] VIII.试剂盒

[0239] 本发明实施方案可提供组合物和试剂盒,用于实施本文所描述的方法,以评估含有多个亚群的细胞混合物的第一细胞亚群中亚群靶基因的基因组表达。所述基因组表达可通过用作生物标志的参数来表示。例如,所述参数可与一个或多个临界值比较,以获得用于诊断或预后的分类,例如,疾病或未病的分类。

[0240] 用于执行用于确定含有多个亚群的细胞混合物的第一细胞亚群中靶基因的基因组表达的分析的试剂盒通常包括至少一个用于与亚群靶基因(及其转录本)特异性杂交的寡核苷酸和至少一个与亚群参照基因(及其转录本)特异性杂交的寡核苷酸。任选地,该寡核苷酸标记有可检测的部分(moiety)。

[0241] 在一些情况下,所述试剂盒可包含至少两条寡核苷酸引物,其可用于扩增所述亚群靶基因和亚群参照基因的至少一部分。代替或除引物以外,试剂盒可包含用于检测亚群靶基因转录本的第一标记探针和用于检测亚群参照基因转录本的至少一部分的第二标记探针。

[0242] 通常,所述试剂盒还提供操作说明书,以指导用户分析测试样品和评估测试对象体内的肝脏生理学或病理学状态。

[0243] IX.计算机系统

[0244] 图15所示为可用于本发明实施方案的系统和方法的示例性计算机系统1500的方框图。

[0245] 本文提到的任何计算机系统均可利用任何合适数目的子系统。这类子系统的实例显示于图15的计算机设备1500中。在一些实施方案中,计算机系统包括单个的计算机设备,其中所述子系统可为所述计算机设备的组件。在其他实施方案中,计算机系统可包括多个计算机设备,每个设备为一个带有内部组件的子系统。

[0246] 图15中所示的子系统通过系统总线1575互连。显示了另外的子系统如打印机1574、键盘1578、固定硬盘1579、连接至显示适配器1582的监视器1576和其他组件。偶联至I/O控制器1571的外围设备和输入/输出(I/O)设备可通过本领域已知的任何数目的方式如串行接口1577来连接至所述计算机系统。例如,可使用串行接口1577或外部接口1581将计算机系统1500连接至广域网路如互联网、鼠标输入设备或扫描仪。通过系统总线1575的连接允许中央处理器1573与每个子系统交流,并控制来自系统内存1572或固定硬盘1579的指令的执行,以及子系统之间的信息交换。系统内存1572和/或固定硬盘1579可体现为计算机可读介质。本文提到的任何数值均可从一个组件输出至另一个组件,且可输出至用户。

[0247] 计算机系统可包括多个相同的组件或子系统,例如,通过外部接口1581或通过内

部接口连接到一起。在一些实施方案中,计算机系统、子系统或设备可在网络上交流。在这类情况下,一台计算机可看作一个客户端,且另一台计算机可看作服务器,其中各自可为相同计算机系统的一部分。客户端和服务器可各自包含多个系统、子系统或组件。

[0248] 应理解,任何本发明的实施方案均可使用硬件(例如专用集成电路或现场可编程门阵列)和/或使用具有通用可编程处理程序的计算机软件,按模块化或集成方式以控制逻辑的形式来执行。基于本文提供的公开和教导,本领域技术人员应知道和理解使用硬件和硬件与软件的组合执行本发明实施方案的其他方式和/或方法。

[0249] 本申请描述的任何软件组件或功能均可作为由处理程序使用任何合适的计算机语言如Java、C++或Perl语言执行的软件代码来实施,例如常规或面向对象的技术。所述软件代码可在计算机可读介质上存储为系列指令或命令用于存储和/或传递,合适的介质包括随机存取存储器(RAM)、只读存储器(ROM)、磁性介质如硬盘驱动器或软盘,或者光学介质如光盘(CD)或DVD(数字多用盘)、闪存等。计算机可读介质可为这类存储或传递装置的任何组合。

[0250] 这类程序还可被编码和使用适于通过有线、光学和/或无线网络传递的载波信号传递,所述无线网络遵守各种协议,包括互联网。如此,根据本发明实施方案的计算机可读介质可用以这类程序编码的数据信号创建。用程序代码编码的计算机可读介质可包装于兼容装置内或单独从其他装置提供(例如,通过互联网下载)。任何这类计算机可读介质均可存在于或安装于单个计算机程序产品(例如硬盘驱动器、CD或整个计算机系统)内,且可位于或安装于系统或网络内部的不同计算机程序产品内。计算机系统可包括监视器、打印机或用于将本文提到的任何结果提供给用户的其他合适的显示器。

[0251] 本文描述的任何方法均可全部或部分地通过包括可配置为执行所述步骤的处理程序的计算机系统实施。因而,本发明实施方案可涉及配置为执行本文描述的任何方法的步骤的计算机系统,可能还有执行各自步骤或各自步骤组的不同组件。尽管以编号的步骤呈现,本文方法中的步骤可同时执行或以不同的顺序执行。另外,这些步骤的部分可与来自其他方法的部分其他步骤一起使用。同样,所有或部分步骤可为任选的。另外,任何方法的任何步骤均可使用用于执行这些步骤的模块、电路或其他方式执行。

[0252] 在不偏离本发明实施方案的实质和范围下,特定实施方案的具体细节可按任何合适的方式组合。然而,本发明其他实施方案可涉及与每个单独的方面或这些单独方面的具体组合相关的特定实施方案。

[0253] 提供了上述对本发明示例性实施方案的描述以用于说明和描述的目的。其并非旨在作穷举性描述或将本发明限制于所描述的具体形式,且鉴于以上教导还可有许多修饰和改变。选择和描述实施方案是为了最好地解释本发明的原理及其应用,从而使本领域技术人员能够最好地在各种实施方案中并使用适于特定应用的各种修改利用本发明。

[0254] 除非有明确的相反说明,所陈述的“一(个/种)(a/an)”或“所述(the)”旨在表示“一(个/种)或多(个/种)”。

[0255] 参考文献包括:

[0256] Lu P, Nakorchevskiy A, Marcotte EM: Expression deconvolution: a reinterpretation of DNA microarray data reveals dynamic changes in cell populations. Proc Natl Acad Sci USA 2003, 100:10370-10375;

- [0257] Shen-Orr SS,Tibshirani R,Khatri P,Bodian DL,Staedtler F,Perry NM,et al.Cell type-specific gene expression differences in complex tissues.Nat.Methods.2010Apr;7(4):287-9;
- [0258] Zhao Y,Simon R.Gene expression deconvolution in clinical samples.Genome Med.2010;2(12):93;
- [0259] U.S.Patent Publication US2006/0292572by Robert O Stuart;
- [0260] Foti,Maria,Paola Ricciardi-Castagnoli, and Francesca Granucci.2007."Gene expression profiling of dendritic cells by microarray."Methods in Molecular Biology(Clifton,N.J.)380:215-224;
- [0261] Kobayashi,Scott D,Dan E Sturdevant, and Frank R DeLeo.2007."Genome-scale transcript analyses in human neutrophils."Methods in Molecular Biology(Clifton,N.J.)412:441-453;
- [0262] Pike-Overzet,Karin,Dick de Ridder,Tom Schonewille, and Frank J T Staal.2009."DNA microarray studies of hematopoietic subpopulations."Methods in Molecular Biology(Clifton,N.J.)506:403-421.doi:10.1007/978-1-59745-409-4_27;
- [0263] Fan,Hongtao, and Priti S Hegde.2005."The transcriptome in blood: challenges and solutions for robust expression profiling."Current Molecular Medicine5(1) (February):3-10;
- [0264] Mohr,Steve, and Choong-Chin Liew.2007."The peripheral-blood transcriptome:new insights into disease and risk assessment."Trends in Molecular Medicine13(10) (October):422-432.doi:10.1016/j.molmed.2007.08.003;
- [0265] Vartanian,Kristina,Rachel Slottke,Timothy Johnstone,Amanda Casale,Stephen R Planck,Dongseok Choi,Justine R Smith,James T Rosenbaum, and Christina A Harrington.2009."Gene expression profiling of whole blood: comparison of target preparation methods for accurate and reproducible microarray analysis."BMC Genomics10:2.doi:10.1186/1471-2164-10-2;
- [0266] Weber,Daniel Gilbert,Swaantje Casjens,Peter Rozynek,Martin Lehnert,Sandra Zilch-Schönweis,Oleksandr Bryk,Dirk Taeger,et al.2010."Assessment of mRNA and microRNA Stabilization in Peripheral Human Blood for Multicenter Studies and Biobanks."Biomarker Insights5:95-102;
- [0267] Chaussabel,Damien,Virginia Pascual, and Jacques Banchereau.2010."Assessing the human immune system through blood transcriptomics."BMC Biology8:84.doi:10.1186/1741-7007-8-84;
- [0268] Mesko,Bertalan,Szilard Poliska, and Laszlo Nagy.2011."Gene expression profiles in peripheral blood for the diagnosis of autoimmune diseases,"Trends in Molecular Medicine17(4) (April):223-233.doi:10.1016/j.molmed.2010.12.004;
- [0269] Laudanski et al."Cell-specific expression and pathway analyses reveal alterations in trauma-related human T cell and monocyte pathways."

Proc.Natl.Acad.Sci.U.S.A.2006Oct17;103(42):15564-9;

[0270] Whitney, et al 2003."Individuality and Variation in Gene Expression Patterns in Human Blood."Proceedings of the National Academy of Sciences of the United States of America 100(4) (February 18):1896-1901;

[0271] Palmer, et al. 2006."Cell-type Specific Gene Expression Profiles of Leukocytes in Human Peripheral Blood."BMC Genomics 7:115;

[0272] Watkins, et al 2009."A HaemAtlas:Characterizing Gene Expression in Differentiated Human Blood Cells."Blood 113(19) (May 7):e1-9;

[0273] Bryant, et al. 2009."Detection of Gene Expression in an Individual Cell Type Within a Cell Mixture Using Microarray Analysis."PloS One 4(2):e4427;

[0274] Butt and Swaminathan, Biomarkers, Feb/2011;

[0275] 美国专利公开2007/0037144、US2007/0020618、US20110070581;以及

[0276] 美国专利第7,598,031号、第6,365,352号和第7,888,030号,以上所有均被并入用于所有目的。

[0277] 上述提到的所有专利、专利申请、出版物和描述均通过引用全部完整并入本文用于所有目的。均不承认其为现有技术。

100

外周全血样品中的造血亚群	比例
嗜中性粒细胞	~50%-65%
单核细胞	~5%
淋巴性粒细胞	~3-5%
淋巴细胞亚群	
-B细胞	~5%
-CD4 T细胞	~15-20%
-CD8 T细胞	~10%
NK细胞	~5%

图1A

150

外周血单核细胞样品中的造血亚群	比例
单核细胞	~30%
淋巴细胞亚群	
-B细胞	~10%
-CD4 T细胞	~30%
-CD8 T细胞	~20%
NK细胞	~10%

图1B

200

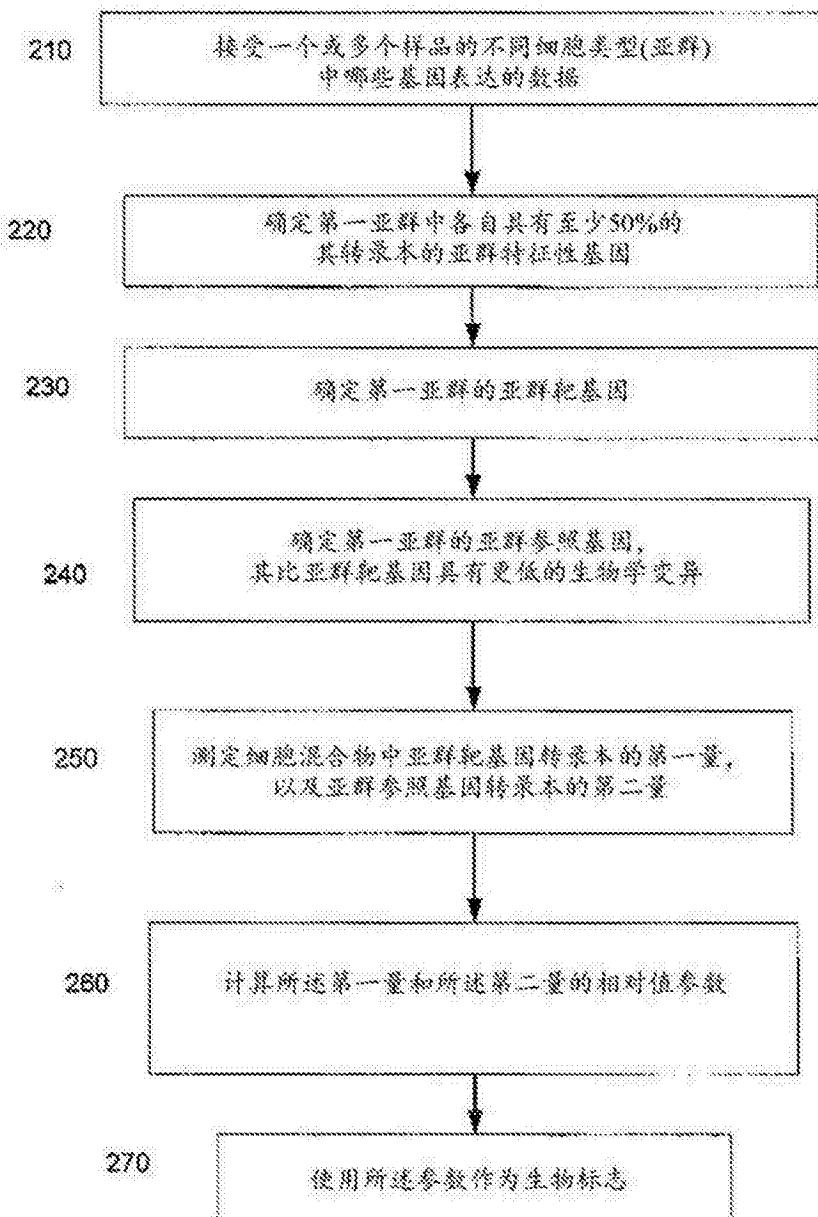


图2

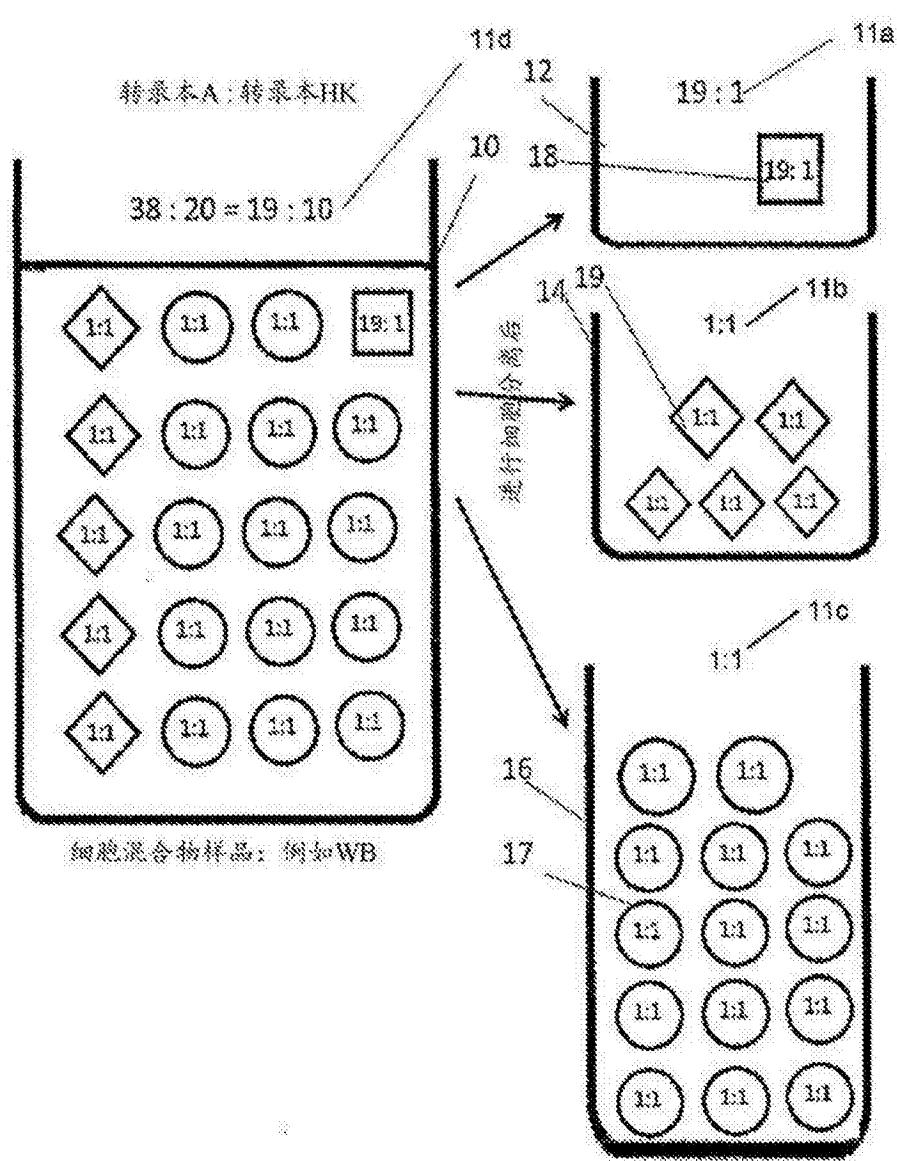


图3

410

420

430

外周全血	各种亚群的 比例细胞计数(P)	X50(亚群：细胞混合物) 当所述亚群在细胞混合物样品中 提供50%的转录本时，倍数
嗜中性粒细胞	~30%-65%	1
单核细胞	~5%	10
嗜酸性粒细胞	~3-5%	10 - 16.7
淋巴细胞亚群		
-B细胞	~5%	10
-总T细胞	~23%	3
-CD4 T细胞	~15-20%	2.5 - 3.3
-CD8 T细胞	~10%	5
-NK细胞	~5%	10

400

图4A

460

470

480

外周血单核细胞	各种亚群的 比例细胞计数(P)	X50(亚群：细胞混合物) 当所述亚群在细胞混合物样品中 提供50%的转录本时，倍数
单核细胞	~30%	1.7
淋巴细胞亚群		
-B细胞	~10%	5
-总T细胞	~50%	1
-CD4 T细胞	~30%	1.7
-CD8 T细胞	~20%	2.5
-NK细胞	~10%	5

450

图4B

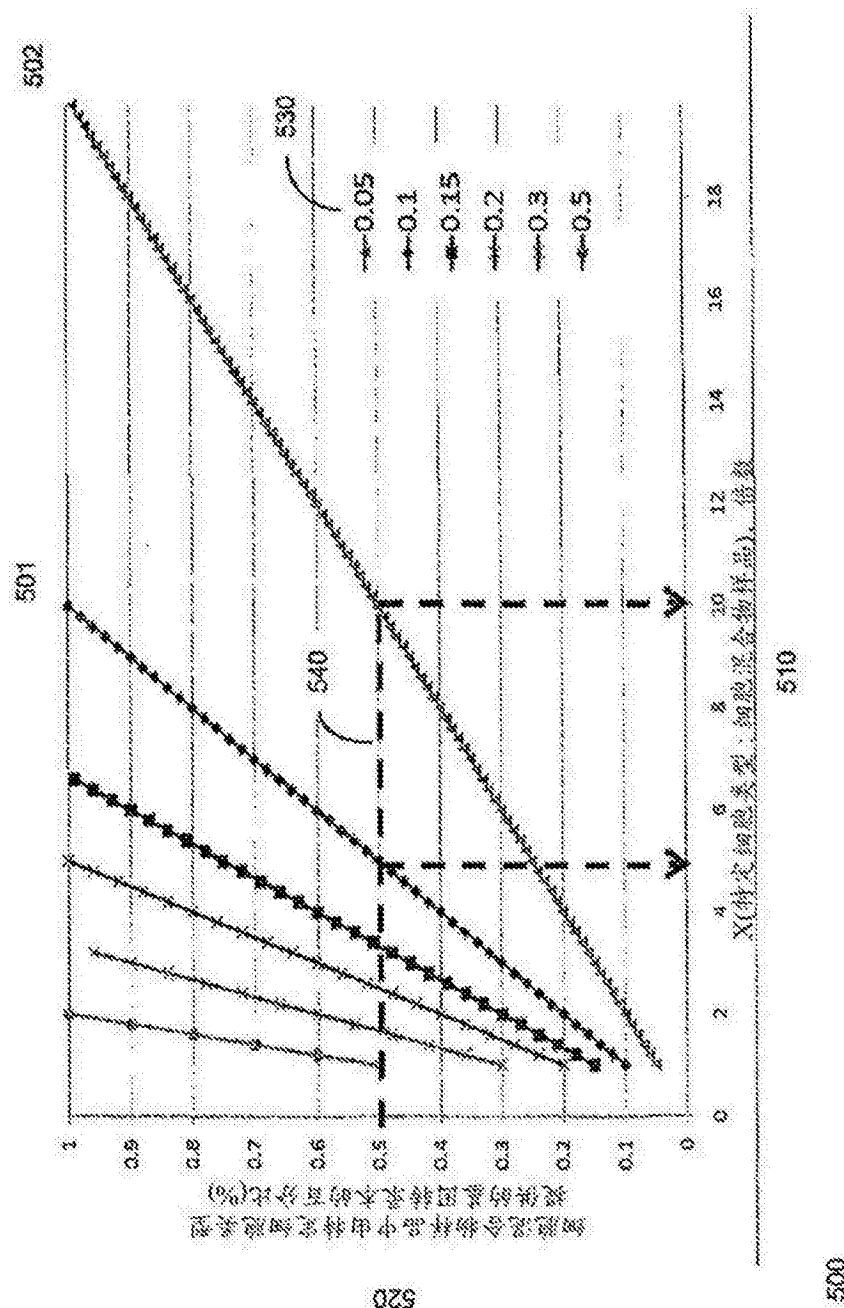


图5

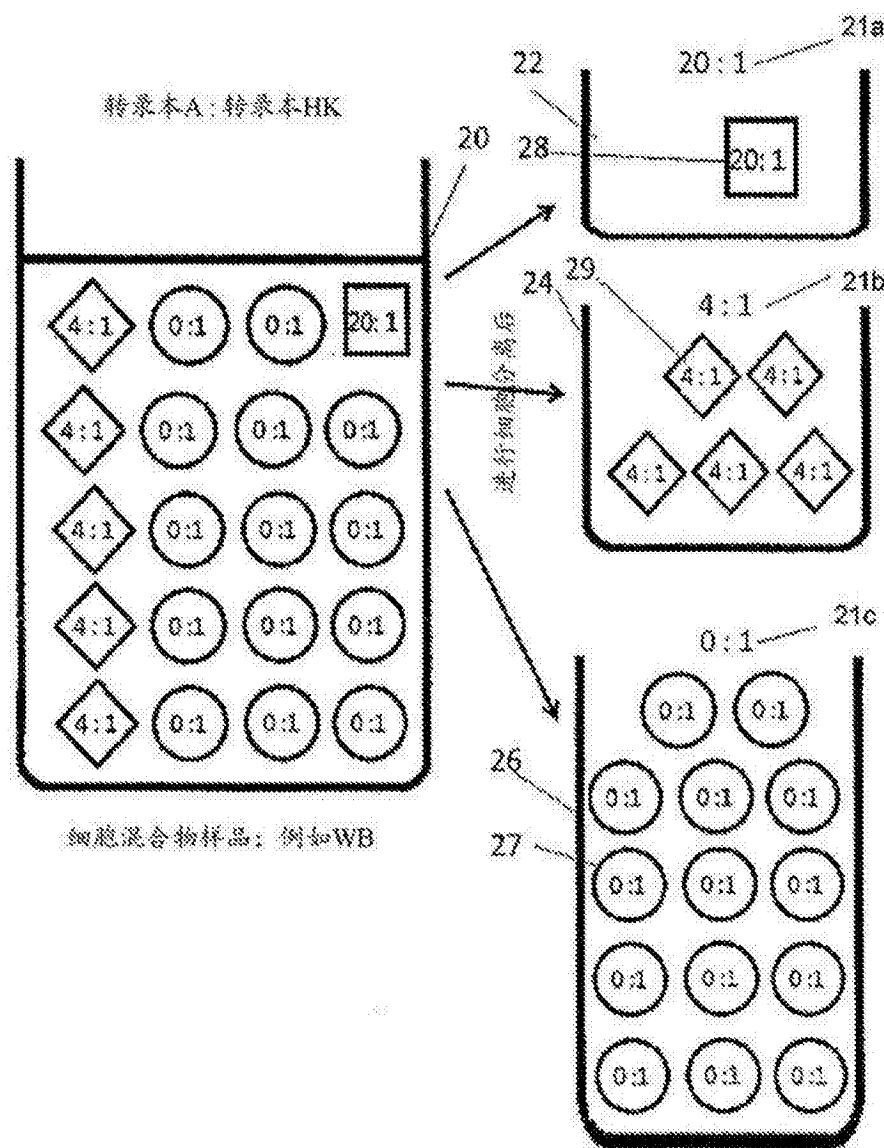


图6

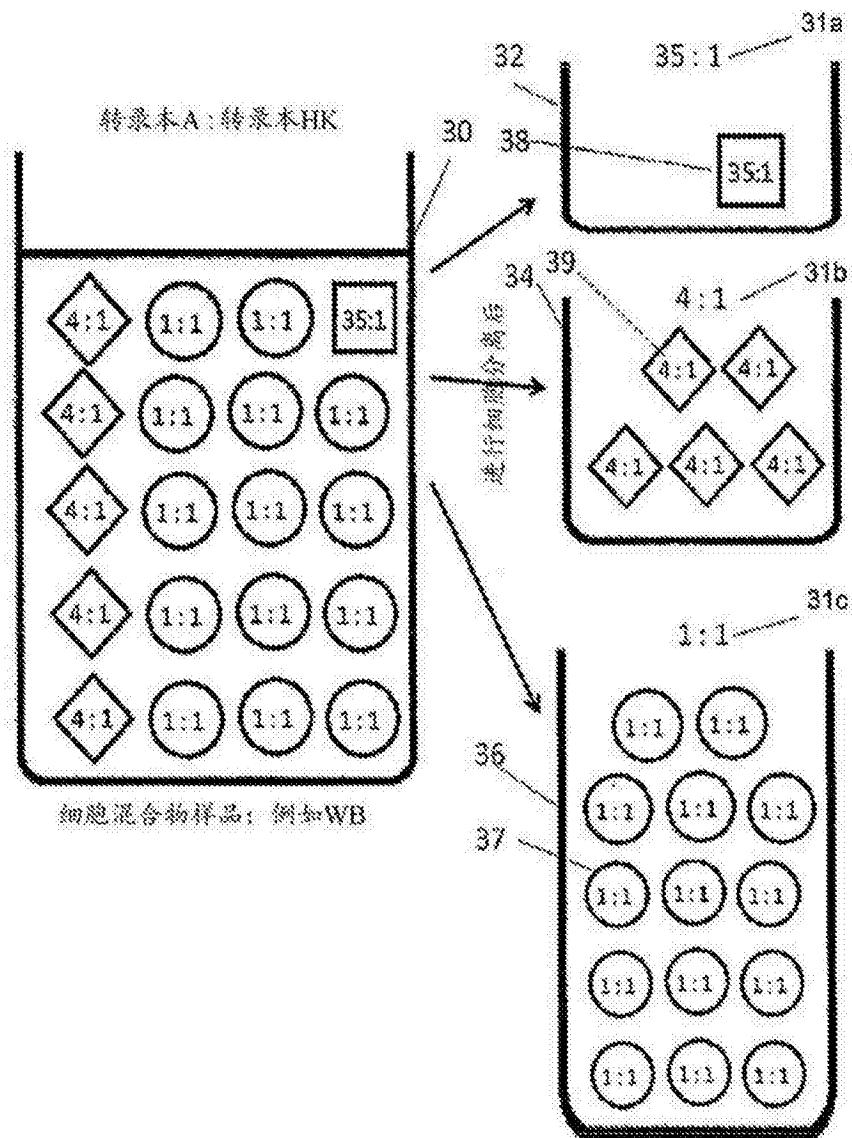


图7

X ₁ %	IAT4	APOBEC3F	BCL2L1	FRE3	KCNG1	KHH14	LRRK2	MMP22	TSPY13C
0.00	0.77	2.58	32.98	38.51	18.54	3.94	11.58	6.78	6.32
St. Dev.	3.84	1.15	21.90	19.72	12.82	3.18	7.18	3.31	3.10
CV%	56.70%	44.37%	62.45%	54.03%	83.11%	61.91%	62.18%	48.81%	51.44%
X50 级%	%	%	%	%	%	%*	%	%	%
X80 级%	%	%	%	%	%	%	%	%	%

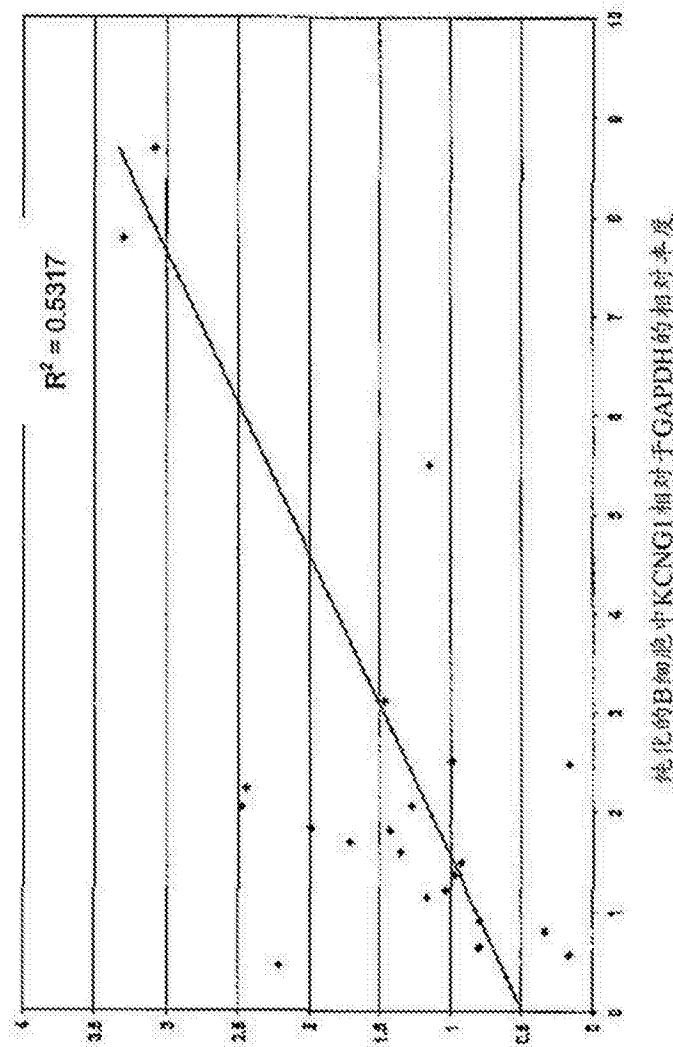
800

图8A

AMCT	KH14	FREE	羟基乙酸 (HGA)	DCA1	GCN1
3.921543	3.943549	26.61327	11.64839	22.899363	15.54784
3.104125	5.181469	16.72221	7.109861	22.296298	12.8219
C%	51.91%	84.63%	62.18%	68.43%	63.11%
CV%	51.91%	84.63%	62.18%	68.43%	63.11%

850

图8B



外周血中KNCG1与TNFSF13C的相对表达

图9

X_候选基因的 (倍数)	SC1138	ITK	PRKCQ	CD3G	GMAP7	GZMK	LRRK3	MAF	NELL2
均值	1.88	2.16	1.81	2.23	2.34	1.84	3.20	1.72	2.72
St. dev.	1.67	1.17	0.68	0.86	1.39	0.62	1.11	0.82	1.24
均值的 95% C.I. (上界)	2.48	2.58	1.85	2.54	2.84	2.07	3.60	2.01	3.15
X50 标准	是	是	否	是	是	是	是	是	是

1000

图10A

采集自30个个体的富集的CD3+ve细胞样品中ddCT的差异									
ddCT的TA	AC138	ITK	PRKCB	CD3G	GIMAP7	G2M8	LRRN3	MAF	NELL2
均值	1.91	0.72	0.52	1.29	1.70	1.00	3.50	0.75	0.61
St. dev.	1.14	0.22	0.21	0.37	0.33	0.37	0.49	0.31	0.23
%	60%	30%	34%	28%	39%	37%	33%	42%	37%

1050

图10B

富集的CD3+T细胞样品中直接LS-TA检测和基因表达之间的相关性(R ²)							
	BCL11B	ITK	CD3G	GZMK	LRRN3	MAF	NELL2
使用GIMAP7进行直接LS-TA检测	0.10	0.23	0.57	0.37	0.56	0.61	0.58
使用GeoM1进行直接LS-TA检测	0.14	0.27	0.57	0.53	0.53	0.66	0.66

1060

图10C

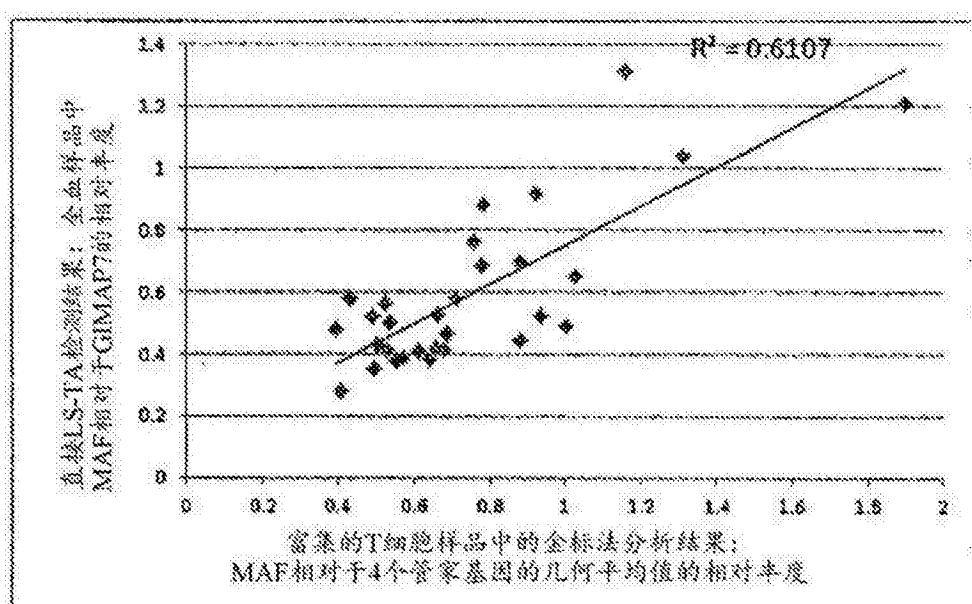


图11A

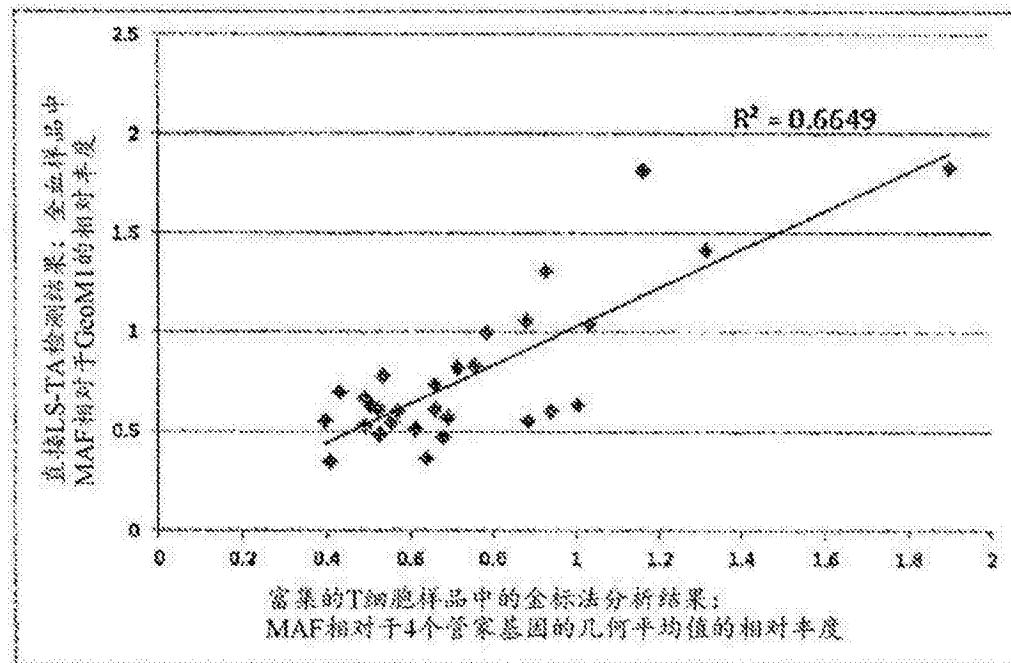


图11B

全血和分离的粒细胞中基因表达水平的比较								
X _i 候选基因 的倍数	AQP1	BCL2A1	CEACAM6	CEACAM8	Clorf24	C3f2	DEFA4	FCGR3B
均值	9.12	6.94	7.71	10.41	6.40	3.72	5.60	8.26
St. dev.	5.83	5.32	4.19	31.84	4.44	5.33	5.47	5.76
均值的95% C.I. (上限)	30.45	8.18	8.58	17.81	7.41	5.28	6.85	9.58
	IL-4	IL-8	IL8RB	LTF	MMP25	MS4A3	SNASE3	SOD2
均值	61.06	26.21	7.77	3.42	39.25	1.21	5.66	6.23
St. dev.	332.13	25.41	5.73	3.43	7.24	0.97	5.01	4.05
均值的95% C.I. (上限)	140.59	32.64	9.06	4.23	41.90	1.44	6.80	7.22

1200

图12A

ddCT 的TA	分离的细胞样品中ddCT的差异							
	AQP1	BCL2A1	CEACAM6	CEACAM8	Clerf24	C9F2	DFFA4	FCGR3B
均值	13.03926	14.37686	7.042252	1724.801	35.06796	4.786464	8.149293	36.48976
St. dev.	7.527057	13.60841	6.678042	10676.64	11.51584	24.63601	7.487139	24.60146
CV%	58%	95%	99%	619%	76%	515%	92%	93%
	IL-4	IL-8	IL8R8	UTP	MMP25	MS4A3	RNASE3	SOD2
均值	281.65568	3.259216	9.431543	3.015567	65.83617	3.3065583	22.86173	4.501045
St. dev.	1663.348	5.040303	8.099712	3.991855	44.98885	2.648231	67.79684	2.932598
CV%	591%	223%	86%	132%	68%	80%	295%	64%

1250

图12B

相关性(R2)	BCL2A1	CEACAM6	CEACAM8	Clerf24	C9F2	DFFA4	FCGR3B
使用AQP1进行直接LS-TA检测	0.02	0.43	0.99	0.04	1.00	0.01	0.04
使用SOD2进行直接LS-TA检测	0.07	0.56	0.95	0.05	0.98	0.12	0.06
	IL-4	IL-8	IL8R8	UTP	MMP25	MS4A3	RNASE3
使用AQP1进行直接LS-TA检测	0.13	0.85	0.11	0.40	0.03	0.38	0.57
使用SOD2进行直接LS-TA检测	0.09	0.92	0.20	0.38	0.64	0.17	0.63

1260

图12C

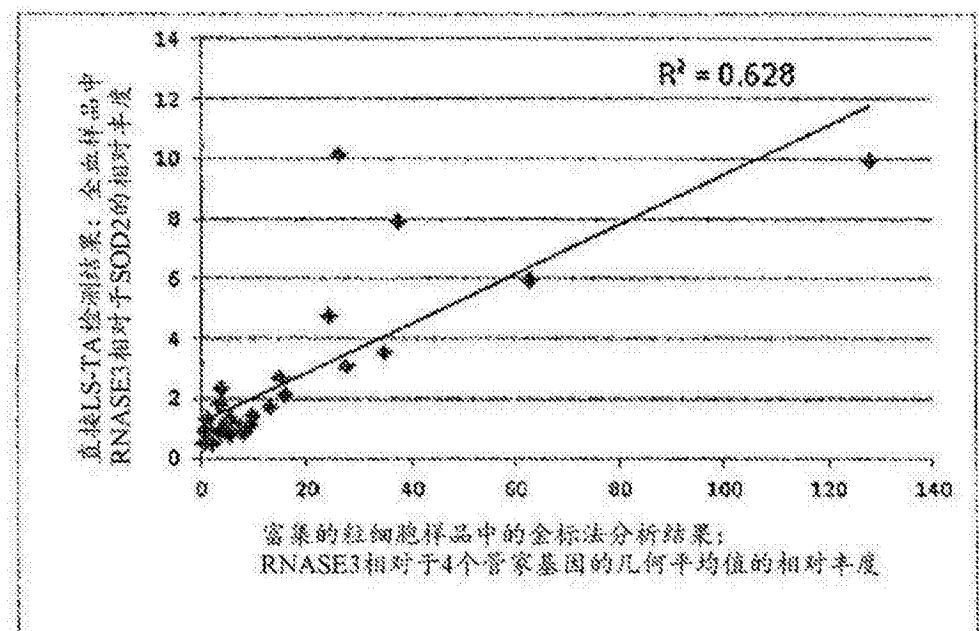


图13

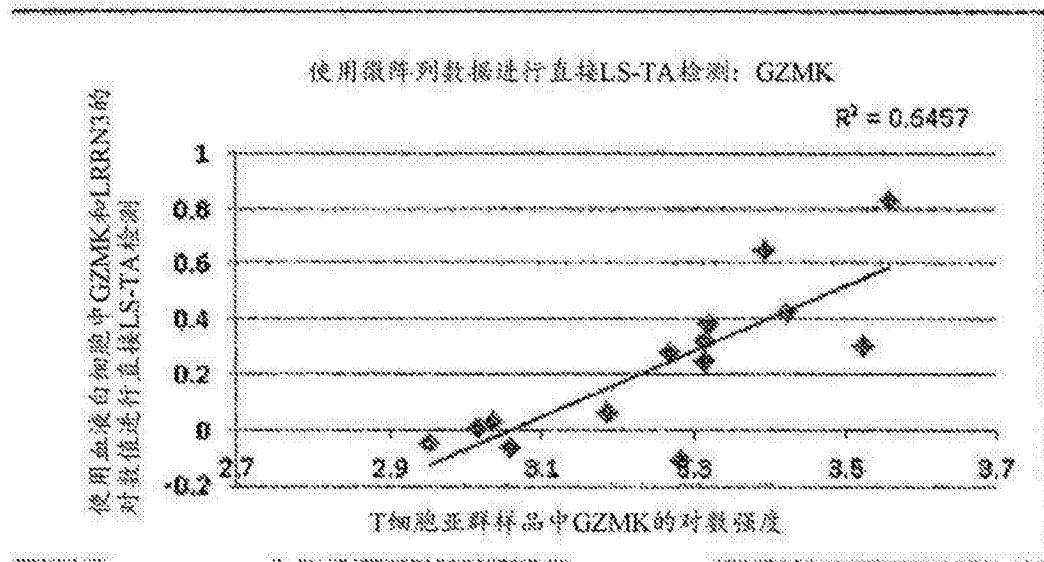


图14A

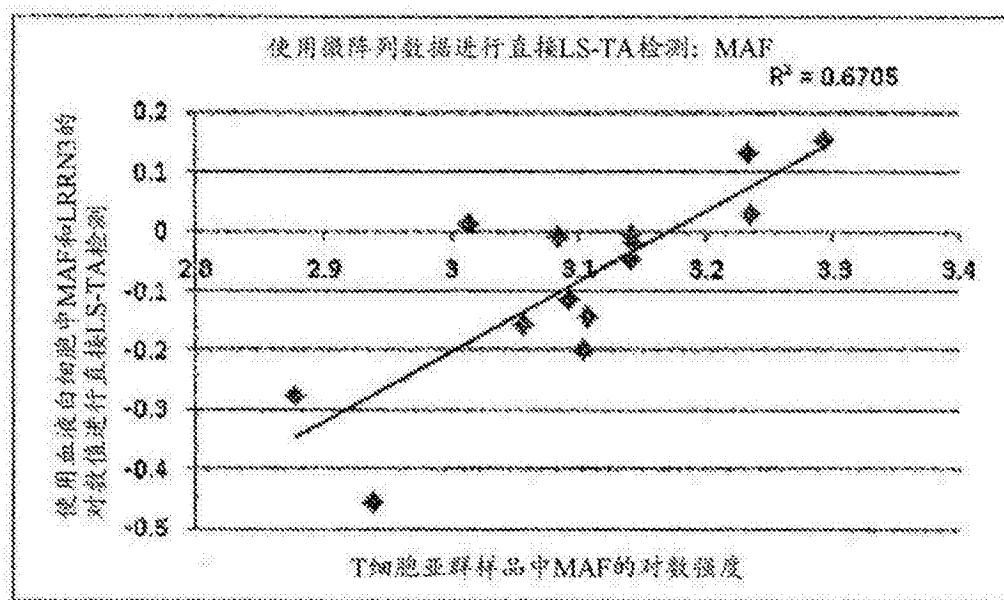


图14B

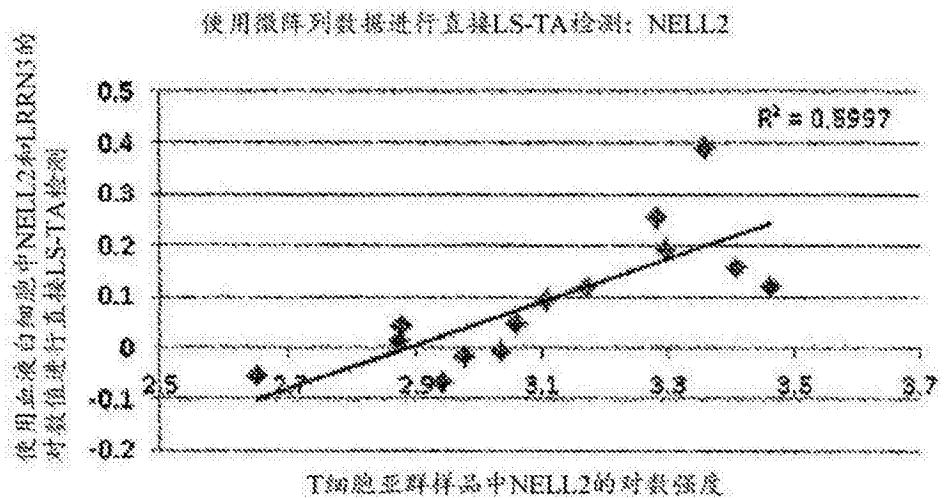


图14C

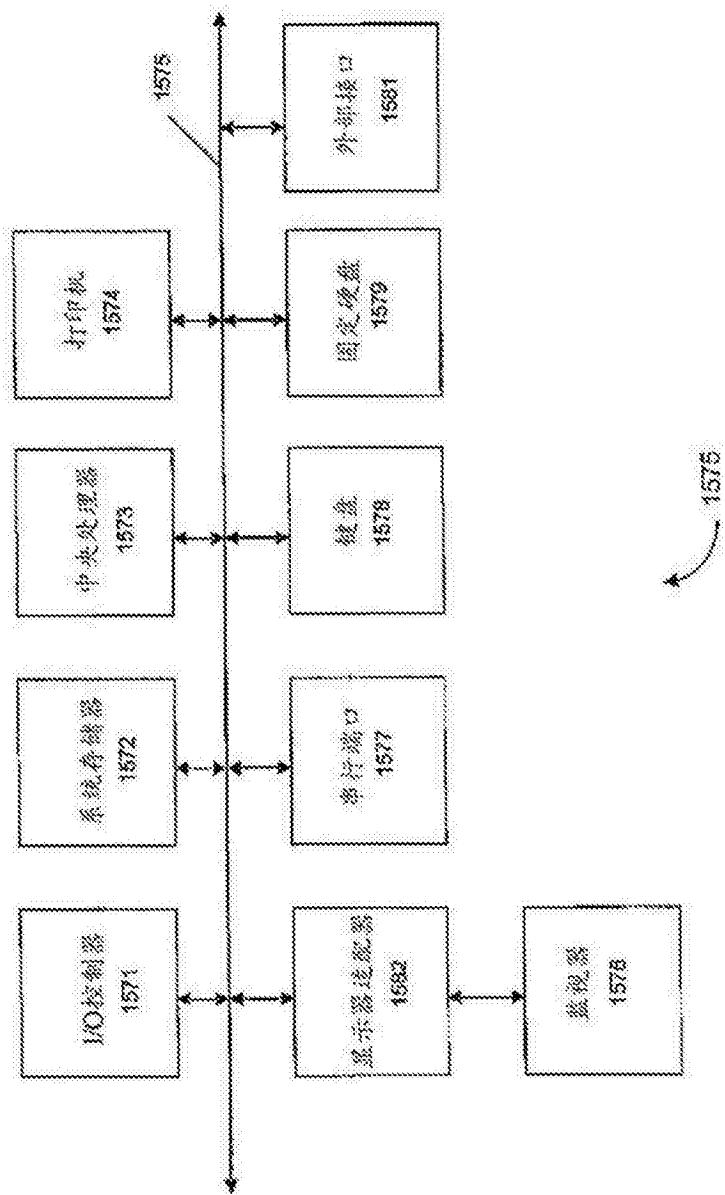


图15