

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5279505号
(P5279505)

(45) 発行日 平成25年9月4日(2013.9.4)

(24) 登録日 平成25年5月31日(2013.5.31)

(51) Int.Cl.

F 1

C 1 2 N	15/09	(2006.01)	C 1 2 N	15/00	F
C 1 2 M	1/00	(2006.01)	C 1 2 M	1/00	A
C 1 2 Q	1/68	(2006.01)	C 1 2 Q	1/68	A
G O 1 N	33/53	(2006.01)	G O 1 N	33/53	M
G O 1 N	37/00	(2006.01)	G O 1 N	37/00	1 0 2

請求項の数 6 (全 67 頁)

(21) 出願番号 特願2008-544529 (P2008-544529)
 (86) (22) 出願日 平成18年12月9日(2006.12.9)
 (65) 公表番号 特表2009-518040 (P2009-518040A)
 (43) 公表日 平成21年5月7日(2009.5.7)
 (86) 国際出願番号 PCT/US2006/046858
 (87) 国際公開番号 W02007/067734
 (87) 国際公開日 平成19年6月14日(2007.6.14)
 審査請求日 平成21年12月7日(2009.12.7)
 (31) 優先権主張番号 60/748,884
 (32) 優先日 平成17年12月9日(2005.12.9)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 11/446,825
 (32) 優先日 平成18年6月5日(2006.6.5)
 (33) 優先権主張国 米国 (US)

(73) 特許権者 509004712
 ベイラー リサーチ インスティテュート
 BAYLOR RESEARCH INS
 TITUTE
 アメリカ国 テキサス75204 ダラス
 スイート501 ライブオークストリー
 ト3310
 (74) 代理人 100107984
 弁理士 廣田 雅紀
 (74) 復代理人 100131093
 弁理士 堀内 真
 (72) 発明者 シヨサベル, ダミアン
 アメリカ合衆国テキサス州75082リチ
 ヤードソン・サウスポイントドライブ45
 32

前置審査

最終頁に続く

(54) 【発明の名称】 末梢血白血球の転写プロファイルのモジュールレベル分析

(57) 【特許請求の範囲】

【請求項1】

以下のステップを含む、疾患分析ツールのためのデータを取得する方法；

【表 1】

転写モジュール	
形質細胞:免疫グロブリン鎖(IGHM, IGHJ, IGL1, IGKC, IGHD)および形質細胞マーカー CD38をコードする遺伝子;	
血小板:血小板タンパク質(ITGA2B, ITGB3, GP6, GP1A/B), およびPPPB(フロー血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子;	
B細胞:B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子(初期B細胞因子(EBF), B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ(BLK)をコードする遺伝子;	
cAMPシグナル伝達経路(JUN, ATF4, CREM, PDE4, NR4A2, VIL2)のレギュレーターおよび標的,ならびにTNF-アルファ媒介型NF- κ B活性化のリプレッサー(CYLD, ASK, TNFAIP3)をコードする遺伝子;	
ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある:CD14, TLR2, MYD88)をコードする遺伝子, この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;	10
シグナル伝達分子, 活性化STAT (PIAS1およびPIAS2)の垂給フィンガー含有インヒビター, または活性化T細胞の核因子NFATC3をコードする遺伝子;	
MHC/リボソームタンパク質/MHCクラスII分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボソームタンパク質(RPLs, RPSs)をコードする遺伝子;	
代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子;	
細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs), 細胞溶解分子(グランザイム, パーホリン, グラニュロシン), ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(GTSW)をコードする遺伝子;	
好中球:好中球顆粒に見いだされる内在性分子(ラクトトランスフェリン, LTF, デフェンシン; DEAF1, 細菌透過性増加タンパク質:RPL, カセリジン抗微生物タンパク質: CAMP...)をコードする遺伝子;	
赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン:ANK1, グリコホリン C:GYPC, ヒドロキシメチルピラン シンターゼ:HMBS, 赤血球関連因子:ERAF)をコードする遺伝子;	20
リボソームタンパク質:リボソームタンパク質(RPLs, RPSs), 真核生物翻訳伸長因子ファミリーメンバー(EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子;	
免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン, サイトキネシスのデデュクター, シンデカン 2, プレキシシンC1, ジストロプレビン)をコードする遺伝子;	
ミエロイド系:M15に関連する, 単球および好中球のようなミエロイド系細胞で発現する遺伝子 (IGT2/CD18 リンボトキシンベータ受容体, ミエロイド関連タンパク質B/14 ホルミルヘフチド受容体 1)を含む;	
ケモカイン様因子スーパーファミリー(CXCLSF8)のメンバーをコードする遺伝子;	
T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンボトキシンベータ IL2-誘導性T細胞キナーゼ, ICF/T細胞分化タンパク質 mat, GATA3, STAT5B)をコードする遺伝子;	
細胞骨格に関連する分子(アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子, また存在するのは, T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1);	
免疫関連細胞表面分子(CD36, CD86, LILRB), サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子;	30
キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKC1, ITPKB, SRPK 2, STK17B, DYRK2, HK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A, KRAS)をコードする遺伝子;	

【表 2】

転写モジュール	
インターフェロン-誘導性:インターフェロン-誘導性遺伝子:抗ウイルス分子(OAS1/2/3/L, GBP1, GIP2, EIF2AK2, PKR, MX1, PML), ケモカイン(CXCL10, IP-10), シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)をコードする遺伝子;	
炎症I:炎症プロセスに関与する分子(IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16), およびアホトシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/G/2A1, GADD45B)をコードする遺伝子;	
炎症II:顆粒球-マクロファージCSFを誘導する,またはそれにより誘導性の(SPI1, IL18, ALOX5 ANPEP),ならびにリソソーム酵素を誘導する,またはそれにより誘導性の(PPT1, CTSS/S, CE S1, NEU1, ASAH1, LAMP2, CAST)の分子をコードする遺伝子;	40
タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ(PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)をコードする遺伝子;	
ヘモグロビン遺伝子(HBA1, HBA2, HBB)をコードする遺伝子;	
T細胞表面マーカー(CD101, CD102, CD103),ならびに血中白血球の回りで偏在的に発現される分子(CXCR1):フラクタルキン受容体, CD47, P-セレクチンリガンド)をコードする遺伝子;	
プロテアソームサブユニット(PSMA2/5, PSMB5/8), ユビキチンタンパク質リガーゼ HIP2, STUB1,ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子;	
幾つかの酵素:アミノメチルトランスフェラーゼ, アルギニルトランスフェラーゼ, アスバラギンシンターゼ, ジアシルグリセロールキナーゼ, イノシトールホスファターゼ, メチルトランスフェラーゼ, ヘリカーゼをコードする遺伝子 および	
プロテインキナーゼ(IPRKPIR, PRKDC, PRKC1)およびホスファターゼ (PTPLB, PPP1R8/2CB), RAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)をコードする遺伝子;	

からなる群から 1 又は複数の遺伝子モジュールを選択するステップ、及び、自己免疫疾患

、ウイルス感染、バクテリア感染、癌および移植拒絶を識別するために、患者から、モジュール内の前記各遺伝子のデータを取得するステップであって、

前記の選択されるモジュールが：

形質細胞：免疫グロブリン鎖（IGHM，IGJ，IGLL1，IGKC，IGHD）および形質細胞マーカーCD38をコードする遺伝子；および

血小板：血小板糖タンパク質（ITGA2B，ITGB3，GP6，GP1A/B）、および血小板由来免疫メディエーターPPPB（プロ-血小板塩基性タンパク質）およびPF4（血小板因子4）をコードする遺伝子；

から選択され、そしてこのモジュールがこれら2つのモジュールに陽性ベクトルを有することにより患者から全身性エリテマトーデスを同定するために使用される方法。

10

【請求項2】

以下のステップを含む、疾患分析ツールのためのデータを取得する方法；

【表3】

転写モジュール	
形質細胞:免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHG)および形質細胞マーカーCD38をコードする遺伝子;	
血小板:血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B);およびPPPB(プロ血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子;	
B細胞:B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子(初期B細胞因子(EBF), B細胞リンカー(BLNK)およびBリンパ系チロシナーゼ(BLK)をコードする遺伝子;	
cAMPシグナル伝達経路(JUND, ATF4, CREM, PDE4, NR4A2, VIL2)のレギュレーターおよび標的、ならびにTNF-アルファ媒介型NF-κB活性化のリプレッサー(CYLD, ASK, TNFAIP3)をコードする遺伝子;	20
ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある。CD14, TLR2, MYD88)をコードする遺伝子、この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;	
シグナル伝達分子、活性化STAT (PIAS1およびPIAS2)の亜鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子;	
MHC/リボソームタンパク質/MHCクラスI分子(HLA-A,B,C,E)+ベータ2-ミクログロブリン(B2M)またはリボソームタンパク質(RPLs, RPSs)をコードする遺伝子;	
代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子;	
細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュロリン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(GTSW)をコードする遺伝子;	
好中球:好中球顆粒に見いだされる内在性分子(ラクトランスフェリン、LTF、デフェンシン、DEAF1、細菌透過性増加タンパク質:RPL,カゼリジン抗微生物タンパク質:CAMP...)をコードする遺伝子;	30
赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン:ANK1, グリコホリン C:GYPC、ヒドロキシメチルピラン シンターゼ:HMBS、赤血球関連因子:ERAF)をコードする遺伝子;	
リボソームタンパク質:リボソームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー(EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子;	
免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(メオシン、サイトキネシスのデドicator、シンデカン 2、プレキシシンC1、ジストロブレピン)をコードする遺伝子;	
ミエロイド系M15に関連する、単球および好中球のようなミエロイド系細胞で発現する遺伝子(ITGB2/CD18 リンボトキシンベータ受容体、ミエロイド関連タンパク質8/14 ホルミルヘフチド受容体 1)を含む;	
ケモカイン様因子スーパーファミリー(CKLF5F8)のメンバーをコードする遺伝子;	
T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンボトキシンベータ IL2-誘導性T細胞キナーゼ,ICF7/T細胞分化タンパク質mal, GATA3, STAT5B)をコードする遺伝子;	40
細胞骨格に関連する分子(アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子、また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1);	
免疫関連細胞表面分子(CD36, CD86, LILRB), サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子;	
キナーゼ(UHMK1, GSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK 2, STK17B, DYRK2, PK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A, KRAS)をコードする遺伝子;	

【表4】

転写モジュール
インターフェロン誘導性: インターフェロン誘導性遺伝子: 抗ウイルス分子 (OAS1/2/3/L, GBP1, GIP2, EIF2AK2/PKR, MX1, PML), ケモカイン (CXCL10/IP-10), シグナル伝達分子 (STAT1, STAT2, IRF7, ISGF3G)をコードする遺伝子;
炎症: 炎症プロセスに関与する分子(IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16), およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/B-2A1, GADD45B)をコードする遺伝子;
炎症II: 顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の (SP11, IL18, ALOX5, ANPEP), ならびにリソソーム酵素を誘導する、またはそれにより誘導性の(PPT1, CTSB/S, CE S1, NEU1, ASAHI, LAMP2, CAST)の分子をコードする遺伝子;
タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ (PI3K)ファミリーメンバー(PIK3CA, PIK32A, PI3K3)をコードする遺伝子; ヘモグロビン遺伝子(HBA1, HBA2, HBE)をコードする遺伝子;
T細胞表面マーカー(CD101, CD102, CD103), ならびに血中白血球の回りで偏在的に発現される分子(CXCR1: フラクタルキン受容体, CD47, P-セクテンリガンド)をコードする遺伝子; プロテアソームサブユニット(PSMA2/5, PSMB5/8), ユビキチンタンパク質リガーゼ HIP2, STUB1, ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子;
幾つかの酵素: アミノメチルトランスフェラーゼ, アルギニルトランスフェラーゼ, アスバラギンシンターゼ, ジアシルグリセロールキナーゼ, イノシトールホスファターゼ, メチルトランスフェラーゼ, ヘリカーゼをコードする遺伝子; および
プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ (PTPLB, PPP1R8/2CB), RAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)をコードする遺伝子;

10

からなる群から1又は複数の遺伝子モジュールを選択するステップ、及び、自己免疫疾患、ウイルス感染、バクテリア感染、癌および移植拒絶を識別するために、患者から、モジュール内の前記各遺伝子のデータを取得するステップであって、

前記の選択されるモジュールが:

20

形質細胞: 免疫グロブリン鎖 (I G H M , I G J , I G L L 1 , I G K C , I G H D) および形質細胞マーカー C D 3 8 をコードする遺伝子; および

血小板: 血小板糖タンパク質 (I T G A 2 B , I T G B 3 , G P 6 , G P 1 A / B)、および血小板由来免疫メディエーター P P P B (プロ - 血小板塩基性タンパク質) および P F 4 (血小板因子 4) をコードする遺伝子;

から選択され、そしてこのモジュールがこれら2つのモジュールに陽性ベクトルも陰性ベクトルも持たないことにより患者からインフルエンザ感染を同定するために使用される方法。

【請求項3】

以下のステップを含む、疾患分析ツールのためのデータを取得する方法;

30

【表 5】

転写モジュール	
形質細胞:免疫グロブリン鎖(IGHM, IGHJ, IGL1, IGKC, IGHG)および形質細胞マーカー CD38をコードする遺伝子;	
血小板:血小板タンパク質(ITGA2B, ITGB3, GP6, GP1A/B), およびPPPB(フロー血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子;	
B細胞:B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子(初期B細胞因子(EBF), B細胞リンカー(BLNK)およびBリンパ系チロシキナーゼ(BLK)をコードする遺伝子;	
cAMPシグナル伝達経路:JUND, ATF4, CREM, PDE4, NR4A2, VIL2)のレギュレーターおよび標的,ならびにTNF-アルファ媒介型NF- κ B活性化のリプレッサー(CYLD, ASK, TNFAIP3)をコードする遺伝子;	
ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある:CD14, TLR2, MYD88)をコードする遺伝子, この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;	10
シグナル伝達分子, 活性化STAT (PIAS1およびPIAS2)の垂給フィンガー含有インヒビター, または活性化T細胞の核因子NFATC3をコードする遺伝子;	
MHC/リボソームタンパク質/MHCクラスII分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボソームタンパク質(RPLs, RPSs)をコードする遺伝子;	
代謝酵素:GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子;	
細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs), 細胞溶解分子(グランザイム, パーホリン, グラニュロシン), ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(GTSW)をコードする遺伝子;	
好中球:好中球顆粒に見いだされる内在性分子(ラクトランスフェリン, LTF, デフェンシン; DEAF1, 細菌透過性増加タンパク質:RPL, カセリジン抗微生物タンパク質: CAMP...)をコードする遺伝子;	
赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン:ANK1, グリコホリン C:GYPC, ヒドロキシメチルピラン シンターゼ:HMBS, 赤血球関連因子:ERAF)をコードする遺伝子;	20
リボソームタンパク質:リボソームタンパク質(RPLs, RPSs), 真核生物翻訳伸長因子ファミリーメンバー(EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子;	
免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン, サイトキネシスのデディケーター, シンデカン 2, プレキシシンC1, ジストロプレブリン)をコードする遺伝子;	
ミエロイド系:M15に関連する, 単球および好中球のようなミエロイド系細胞で発現する遺伝子 (IGT2/CD18 リンボトキシンベータ受容体, ミエロイド関連タンパク質8/14 ホルミルヘフチド受容体 1)を含む;	
ケモカイン様因子スーパーファミリー(CKLFSF8)のメンバーをコードする遺伝子;	
T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンボトキシンベータ IL2-誘導性T細胞キナーゼ, ICF/T細胞分化タンパク質 mat, GATA3, STAT5B)をコードする遺伝子;	
細胞骨格に関連する分子(アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子, また存在するのは, T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1);	
免疫関連細胞表面分子(CD36, CD86, LILRB), サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子;	30
キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKC1, ITPKB, SRPK 2, STK17B, DYRK2, HK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A, KRAS)をコードする遺伝子;	

【表 6】

転写モジュール	
インターフェロン-誘導性: インターフェロン-誘導性遺伝子: 抗ウイルス分子 (OAS1/2, 3/L, GBP1, GIP2, EIF2AK2, PKR, MX1, PML), ケモカイン (CXCL10/IP-10), シグナル伝達分子 (STAT1, STAT2, IRF7, ISGF3G)をコードする遺伝子;	
炎症: 炎症プロセスに関与する分子(IL8, ICAM1, CSRI, CD44, PLAUR, IL1A, CXCL16), およびアポトーシスのレギュレーター(MGL1, FOXO3A, RARA, BCL3/0/2A1, GADD45B)をコードする遺伝子;	
炎症II 顆粒球-マクロファージCSFを誘導する, またはそれにより誘導性の (SPI1, IL18, ALOX5, ANPEP), ならびにリソソーム酵素を誘導する, またはそれにより誘導性の(PPT1, GTSB-S, CE S1, NEU1, ASAH1, LAMP2, CAST)の分子をコードする遺伝子;	40
タンパク質ホスファターゼ(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ (PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)をコードする遺伝子;	
ヘモグロビン遺伝子(HBA1, HBA2, HBE)をコードする遺伝子;	
T細胞表面マーカー(CD101, CD102, CD103), ならびに血中白血球の回りで偏在的に発現される分子(CXCR1): フラクタルキン受容体, CD47, P-セレクチンリガンド)をコードする遺伝子;	
プロテアソームサブユニット(PSMA2/5, PSMB5/B), ユビキチンタンパク質リガーゼ HIP2, STUB1, ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子;	
幾つかの酵素:アミノメチルトランスフェラーゼ, アルギニルトランスフェラーゼ, アスパラギンシンターゼ, ジアシルグリセロールキナーゼ, イノシトールホスファターゼ, メチルトランスフェラーゼ, ヘリカーゼをコードする遺伝子 および	
プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ (PTPLB, PPP1R8/2GB), RAS系遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)をコードする遺伝子;	

からなる群から 1 又は複数の遺伝子モジュールを選択するステップ、及び、自己免疫疾患

、ウイルス感染、バクテリア感染、癌および移植拒絶を識別するために、患者から、モジュール内の前記各遺伝子のデータを取得するステップであって、

前記の選択されるモジュールが：

形質細胞：免疫グロブリン鎖（IGHM，IGJ，IGLL1，IGKC，IGHD）および形質細胞マーカーCD38をコードする遺伝子；および

血小板：血小板糖タンパク質（ITGA2B，ITGB3，GP6，GP1A/B）、およびPPPB（プロ-血小板塩基性タンパク質）およびPF4（血小板因子4）のような血小板由来免疫メディエーターをコードする遺伝子；

から選択され、そしてこのモジュールが形質細胞マーカーに陰性ベクトルを、そして血小板マーカーに陽性ベクトルを有することにより患者から黒色腫を同定するために使用される方法。

10

【請求項4】

以下のステップを含む、疾患分析ツールのためのデータを取得する方法；

【表7】

転写モジュール	
形質細胞:免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHG)および形質細胞マーカーCD38をコードする遺伝子;	
血小板:血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ-血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子;	
B細胞:B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子(初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ(BLK)をコードする遺伝子;	20
cAMPシグナル伝達経路(JUN, ATF4, CREM, PDE4, NR4A2, VIL2)のレギュレーターおよび標的、ならびにTNF-アルファ媒介型NF-κB活性化のリプレッサー(CYLD, ASK, TNFAIP3)をコードする遺伝子;	
ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある。CD14, TLR2, MYD88)をコードする遺伝子。この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;	
シグナル伝達分子、活性化STAT (PIAS1およびPIAS2)の垂糸フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子;	
MHC/リボソームタンパク質/MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボソームタンパク質(RPLs, RPSs)をコードする遺伝子;	
代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子;	
細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グランジュリン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)をコードする遺伝子;	30
好中球:好中球顆粒に見いだされる内在性分子(ラクトトランスフェリン、LTF、デフェンシン、DEAF1)、細菌透過性増加タンパク質:RPL、カゼリジン抗微生物タンパク質: CAMP...をコードする遺伝子;	
赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン、ANK1、グリコホリン C、GYPC、ヒドロキシメチルピラン シンターゼ: HMBS、赤血球関連因子: ERAP)をコードする遺伝子;	
リボソームタンパク質:リボソームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー(EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子;	
免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター、シンデカン 2、プレキシシンC1、ジストロプレビン)をコードする遺伝子;	
ミエロイド系:M15に関連する、単球および好中球のようなミエロイド系細胞で発現する遺伝子(IGT82/CD18 リンホトキシンベータ受容体、ミエロイド関連タンパク質8/14 ホルミルヘフチド受容体 1)を含む;	
ケモカイン様因子スーパーファミリー(CXCLSF8)のメンバーをコードする遺伝子;	
T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ, ICF7,T細胞分化タンパク質 mal, GATA3, STAT5B)をコードする遺伝子;	40
細胞骨格に関連する分子(アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子。また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1);	
免疫関連細胞表面分子(CD36, CD86, LILRB)、サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子;	
キナーゼ(UHMK1, GSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK 2, STK17B, DYRK2, HK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A, KRAS)をコードする遺伝子;	

【表 8】

転写モジュール
インターフェロロン誘導性: インターフェロロン誘導性遺伝子: 抗ウイルス分子 (OAS1/2, 3/L, GBP1, GIP2, EIF2AK2/PKR, MX1, PML), ケモカイン (CXCL10/IP-10), シグナル伝達分子 (STAT1, STAT2, IRF7, ISGF3G)をコードする遺伝子;
炎症: 炎症プロセスに関与する分子(IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16), およびアポトーシスのレギュレーター(MGL1, FOXO3A, RARA, BCL3/0/2A1, GADD45B)をコードする遺伝子;
炎症: 顆粒球-マクロファージCSFを誘導する, またはそれにより誘導性の (SPI1, IL18, ALOX5, ANPEP), ならびにリソソーム酵素を誘導する, またはそれにより誘導性の(PPT1, CTSD/S, CE S1, NEU1, ASAH1, LAMP2, CAST)の分子をコードする遺伝子;
タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ (PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)をコードする遺伝子; ヘモグロビン遺伝子(HBA1, HBA2, HBB)をコードする遺伝子;
T細胞表面マーカー(CD101, CD102, CD103), ならびに血中白血球の回りで偏在的に発現される分子(CXCR1): フラクタルキン受容体, CD47, P-セレクチンリガンド)をコードする遺伝子; プロテアソームサブユニット(PSMA2/5, PSMB5/8), ユビキチンタンパク質リガーゼ HIP2, STUB1, ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子;
幾つかの酵素: アミノメチルトランスフェラーゼ, アルギニルトランスフェラーゼ, アスパラギンシクターゼ, ジアシルグリセロールキナーゼ, イノシトールホスファターゼ, メチルトランスフェラーゼ, ヘリカーゼをコードする遺伝子; および
プロテインキナーゼ(PRKFR, PRKDC, PRKCI)およびホスファターゼ (PTPLB, PPP1R8/2GB), RAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)をコードする遺伝子;

10

からなる群から 1 又は複数の遺伝子モジュールを選択するステップ、及び、自己免疫疾患、ウイルス感染、バクテリア感染、癌および移植拒絶を識別するために、患者から、モジュール内の前記各遺伝子のデータを取得するステップであって、

前記の選択されるモジュールが：

20

形質細胞：免疫グロブリン鎖 (I G H M , I G J , I G L L 1 , I G K C , I G H D) および形質細胞マーカー C D 3 8 をコードする遺伝子；および

血小板：血小板糖タンパク質 (I T G A 2 B , I T G B 3 , G P 6 , G P 1 A / B)、および血小板由来免疫メディエーター P P P B (プロ - 血小板塩基性タンパク質) および P F 4 (血小板因子 4) をコードする遺伝子；

から選択され、そしてこのモジュールがこれら 2 つのモジュールに陰性ベクトルを有することにより患者から移植拒絶を同定するために使用される方法。

【請求項 5】

以下のステップを含む、疾患分析ツールのためのデータを取得する方法；

【表 9】

転写モジュール	
形質細胞:免疫グロブリン鎖(IGHM, IGHJ, IGL1, IGKC, IGHG)および形質細胞マーカー CD38をコードする遺伝子;	
血小板:血小板タンパク質(ITGA2B, ITGB3, GP6, GPIA/B), およびPPPB(フロー血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子;	
B細胞:B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子(初期B細胞因子(EBF), B細胞リンカー(BLNK)およびBリンパ系チロシキナーゼ(BLK)をコードする遺伝子;	
cAMPシグナル伝達経路:JUND, ATF4, CREM, PDE4, NR4A2, VIL2)のレギュレーターおよび標的,ならびにTNF-アルファ媒介型NF-κB活性化のリプレッサー(CYLD, ASK, TNFAIP3)をコードする遺伝子;	
ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある:CD14, TLR2, MYD88)をコードする遺伝子, この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;	10
シグナル伝達分子, 活性化STAT (PIAS1およびPIAS2)の垂給フィンガー含有インヒビター, または活性化T細胞の核因子NFATC3をコードする遺伝子;	
MHC/リボソームタンパク質/MHCクラスII分子(HLA-A,B,C,E,G)+ベータ2-ミクログロブリン(B2M)またはリボソームタンパク質(RPLs, RPSs)をコードする遺伝子;	
代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子;	
細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs), 細胞溶解分子(グランザイム, パーホリン, グラニュロシン), ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)をコードする遺伝子;	
好中球:好中球顆粒に見いだされる内在性分子(ラクトランスフェリン, LTF, デフェンシン; DEAF1, 細菌透過性増加タンパク質:RPL, カセリジン抗微生物タンパク質: CAMP...)をコードする遺伝子;	
赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン:ANK1, グリコホリン C:GYPC, ヒドロキシメチルピラン シンターゼ:HMBS, 赤血球関連因子:ERAF)をコードする遺伝子;	20
リボソームタンパク質:リボソームタンパク質(RPLs, RPSs), 真核生物翻訳伸長因子ファミリーメンバー(EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子;	
免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン, サイトキネシスのデディケーター, シンデカン 2, プレキシシンC1, ジストロプレビン)をコードする遺伝子;	
ミエロイド系:M15に関連する, 単球および好中球のようなミエロイド系細胞で発現する遺伝子 (IGT2/CD18 リンボトキシンベータ受容体, ミエロイド関連タンパク質B/14 ホルミルヘフチド受容体 1)を含む;	
ケモカイン様因子スーパーファミリー(CXCLSF8)のメンバーをコードする遺伝子;	
T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンボトキシンベータ IL2-誘導性T細胞キナーゼ, ICF/T細胞分化タンパク質 mat, GATA3, STAT5B)をコードする遺伝子;	
細胞骨格に関連する分子(アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子, また存在するのは, T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1);	
免疫関連細胞表面分子(CD36, CD88, LILRB), サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll様受容体経路)をコードする遺伝子;	30
キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKC1, ITPKB, SRPK 2, STK17B, DYRK2, HK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A, KRAS)をコードする遺伝子;	

【表 10】

転写モジュール	
インターフェロン誘導性:インターフェロン誘導性遺伝子:抗ウイルス分子(OAS1/2/3/L, GBP1, GIP2, EIF2AK2, PKR, MX1, PML), ケモカイン(CXCL10, IP-10), シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)をコードする遺伝子;	
炎症I:炎症プロセスに関与する分子(IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16), およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/G-2A1, GADD45B)をコードする遺伝子;	
炎症II:顆粒球-マクロファージCSFを誘導する, またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP), ならびにリソソーム酵素を誘導する, またはそれにより誘導性の(PPT1, CTSB/S, CE S1, NEU1, ASAH1, LAMP2, CAST)の分子をコードする遺伝子;	40
タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ(PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)をコードする遺伝子;	
ヘモグロビン遺伝子(HBA1, HBA2, HBB)をコードする遺伝子;	
T細胞表面マーカー(CD101, CD102, CD103), ならびに血中白血球の回りで偏在的に発現される分子(CXCR1):フラクタルキン受容体, CD47, P-セレクチンリガンド)をコードする遺伝子;	
プロテアソームサブユニット(PSMA2/5, PSMB5/8), ユビキチンタンパク質リガーゼ HIP2, STUB1, ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子;	
幾つかの酵素:アミノメチルトランスフェラーゼ, アルギニルトランスフェラーゼ, アスバラギンシンターゼ, ジアシルグリセロールキナーゼ, イノシトールホスファターゼ, メチルトランスフェラーゼ, ヘリカーゼをコードする遺伝子 および	
プロテインキナーゼ(IPRKPIR, PRKDC, PRKC1)およびホスファターゼ(PTPLB, PPP1R8/2CB), RAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)をコードする遺伝子;	

からなる群から 1 又は複数の遺伝子モジュールを選択するステップ、及び、自己免疫疾患

、ウイルス感染、バクテリア感染、癌および移植拒絶を識別するために、患者から、モジュール内の前記各遺伝子のデータを取得するステップであって、

前記の選択されるモジュールが：

形質細胞：免疫グロブリン鎖（ I G H M , I G J , I G L L 1 , I G K C , I G H D ）および形質細胞マーカー C D 3 8 をコードする遺伝子；および

血小板：血小板糖タンパク質（ I T G A 2 B , I T G B 3 , G P 6 , G P 1 A / B ）、および血小板由来免疫メディエーター P P P B （プロ - 血小板塩基性タンパク質）および P F 4 （血小板因子 4 ）をコードする遺伝子；

から選択され、そしてこのモジュールがこれら 2 つのモジュールに陰性ベクトルを有することにより患者からインフルエンザ感染を同定するために使用される方法。

10

【請求項 6】

1 又は複数の転写モジュールの代表である遺伝子の組み合わせを含んでなるカスタマイズされた遺伝子アレイを含んでなる予知遺伝子アレイであって、前記カスタマイズされた遺伝子アレイと接触する患者のトランスクリプトームが、前記転写モジュールに合った 1 又は複数の疾患または状態の予知であり、

前記転写モジュールが、

疾患または状態毎に合う各クラスターからそのモジュールについての遺伝子を選択し；

前記選択した遺伝子を分析から除き；そして

疾患の遺伝子について、遺伝子発現値を選択する工程を繰り返すことにより、1 又は複数の転写モジュールに関する遺伝子発現値を反復して選択し；そして

20

すべての遺伝子クラスターが使用し尽されるまで、各クラスターについてモジュールの作成を繰り返し反復する、ことにより事前に得られる

上記予知遺伝子アレイであって、

【表 1 1】

サブモジュール	プローブの組数	キーワードの選択	評価
M 1.1	69	Ig, 免疫グロブリン、骨、骨髓、PreB、IgM、Mu.	形質細胞: 免疫グロブリン鎖 (IGHM, IGJ, IGLL1, IGKC, IGHD) および形質細胞マーカー CD38 をコードする遺伝子;
M 1.2	96	血小板、接着、凝集、内皮、血管	血小板: 血小板糖タンパク質 (ITGA2B, ITGB3, GP6, GPIA/B), および PPPB (フロー血小板塩基性タンパク質) および PF4 (血小板因子 4) のような血小板由来免疫メディエーターをコードする遺伝子;
M 1.3	47	イムノレセプター、BCR、B細胞、IgG	B細胞: B細胞表面マーカー (CD72, CD79A/B, CD19, CD22) および他の B細胞関連分子: 初期 B細胞因子 (EBF)、B細胞リンカー (BLNK) および Bリンパ系チロシンキナーゼ (BLK) をコードする遺伝子;
M 1.4	87	複製、抑制、修復、CREB、リンパ系、TNF-アルファ	cAMPシグナル伝達経路のレギュレーター および標的 (JUND, ATF4, CREM, PDE4, NR4A2, VIL2), ならびに TNF-アルファ媒介型 NF- κ B 活性化のリプレッサー (CYLD, ASK, TNFAIP3) をコードする遺伝子;
M 1.5	130	単球、樹状突起の、MHC、同時刺激、TLR4、MYD88	ミエロイド系/ミエロイド系細胞により発現される分子 (CD86, CD163, FCGR2A), (その中には病原体の認識に関与するものもある: CD14, TLR2, MYD88) の細胞により発現される分子。この組は TNFファミリーメンバー (TNFR2, BAFF) を含む;
M 1.6	28	垂鉛、フィンガー、Pb3、RAS	シグナル伝達分子、活性化 STAT (PIAS1 および PIAS2) の垂鉛フィンガー含有インヒビター、または活性化 T細胞の核因子 NFATC3 をコードする遺伝子;
M 1.7	127	リボソーム、翻訳、40S、60S、HLA	HMC/リボソームタンパク質/MHCクラス I 分子 (HLA-A,B,C,G,E) + ベータ 2-ミクログロブリン (B2M); またはリボソームタンパク質 (RPLs, RPSs) をコードする遺伝子;
M 1.8	86	代謝、生合成、複製、ヘリカーゼ	代謝酵素 (GLS, NSF1, NAT1) および DNA複製に関与する因子 (PURA, TERF2, EIF2S1) をコードする遺伝子;
M 2.1	72	NK、キラー、細胞溶解、CD8、細胞媒介型、T細胞、CTL、IFN- γ	細胞傷害性細胞: 細胞傷害性 T細胞 および NK細胞表面マーカー (CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子 (グランザイム、パーホリン、グラニュリシン)、ケモカイン (CCL5, XCL1) および CTL/NK-細胞関連分子 (GTSW) をコードする遺伝子;
M 2.2	44	顆粒球、好中球、防御、ミエロイド、骨髓	好中球: 好中球顆粒に見いだされる内在性分子 (ラクトトランスフェリン、LTF、デフェンシン、DEAF1、細菌透過性増加タンパク質 BPI、カセリジン抗菌生体タンパク質 CAMP...) をコードする遺伝子;

10

20

30

【表 1 2】

サブモジュール	プローブの組数	キーワードの選択	評価
M 2 3	94	赤血球、赤、貧血、グロビン、ヘモグロビン	赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン: ANK1、グリコホリン C:GYPC、ヒドロキシメチルピラン シンターゼ: HMBS、赤血球関連因子: ERAF)
M 2 4	118	リボヌクレオタンパク質、60S、核小体、アッセンブリー、延長	リボソームタンパク質: リボソームタンパク質(RP1 s, RPSs)、真核生物翻訳伸長因子ファミリーメンバー (EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子;
M 2 5	242	腺癌、間質の、間葉、樹状突起、運動	免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデティケーター、シンデカン 2、フレキシシンC1、ジストロブレピン)をコードする遺伝子;
M 2 6	110	顆粒球、単球、ミエロイド、ERK、壊死	ミエロイド系細胞 (IGTB2/CD18, リンホトキシンベータ受容体、ミエロイド関連タンパク質 8/14 ホルミルペプチド受容体 1)、単球および好中球で発現する分子をコードする遺伝子;
M 2 7	43	キーワードは抽出されない	ケモカイン様因子スーパーファミリー (CKLFSF8)の1もしくは複数のメンバーをコードする遺伝子;
M 2 8	104	リンパ腫、T細胞、CD4, CD8, TCR、胸腺、リンパ系、IL2	T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ, TCF7, T細胞分化タンパク質 mal, GATA3, STAT5B)をコードする遺伝子;
M 2 9	122	ERK、トランス活性化、細胞骨格、MAPK, JNK	細胞骨格に関連する分子 (アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子。また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1)。
M 2 10	44	ミエロイド、マクロファージ、樹状突起、炎症、インターロイキン	免疫関連細胞表面分子(CD36, CD86, LILRB)、サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子;
M 2 11	77	複製、抑制、RAS、自己リン酸化作用、発癌遺伝子	キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCL, ITPKB, SRPK 2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバ (G3BP, RAB14, RAS A2, RAP2A, KRAS)をコードする遺伝子;
M 3 1	80	ISRF、インフルエンザ、抗ウイルス、IFN-ガンマ、IFN-アルファ、インターフェロン	インターフェロン-誘導性: インターフェロン-誘導性遺伝子:抗ウイルス分子(OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML)、ケモカイン (CXCL10/IP-10)、シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)をコードする遺伝子;

10

20

30

【表 1 3】

サブモジュール	プローブの組数	キーワードの選択	評価
M 3 2	230	TGF-ベータ、TNF、炎症、アポトーシスの、リホ多糖	炎症I: 炎症プロセスに関与する分子(IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3、6/2A1, GADD45B)をコードする遺伝子;
M 3 3	230	顆粒球、炎症、防御、酸化する、リソソームの	炎症II: 顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP)、ならびにリソソーム酵素を誘導する、またはそれにより誘導性(PPT1, GTSB/S, CE S1, NEU1, ASAH1, LAMP2, CAST)の分子をコードする遺伝子;
M 3 4	323	キーワードは抽出されない	タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ(PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)をコードする遺伝子;
M 3 5	19	キーワードは抽出されない	ヘモグロビン遺伝子(HBA1, HBA2, HBB)をコードする遺伝子;
M 3 6	233	補体、宿主、酸化的、細胞骨格、T細胞	T細胞表面マーカー(CD101, CD102, CD103)、ならびに血中白血球の回りで偏在的に発現される分子(CXCR1: フラクタルキン受容体, CD47, P-セレクチンリガンド)をコードする遺伝子;
M 3 7	80	スプライソソーム、メチル化、ユビキチン、ベーターカテニン	プロテアソームサブユニット(PSMA2、5, PSMB5、8); ユビキチンタンパク質リガーゼHIP 2, STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子;
M 3 8	182	CDC, TCR, CREB, グリコシラーゼ	幾つかの酵素: アミノメチルトランスフェラーゼ、アルギニルトランスフェラーゼ、アスハラギンシンテターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼをコードする遺伝子; および
M 3 9	261	クロマチン、チェックポイント、複製、トランス活性化	プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ (PTPLB, PPP1R8/2CB)をコードする遺伝子。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)も含む;

10

20

から選択される 1 又は複数のサブモジュールを含んでなる 3 つの転写モジュールに組織化され、そしてモジュールの 1 又は複数の遺伝子に特異的に結合するプローブを含むアレイ

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明の技術分野

本発明は一般的には細胞の転写プロファイリングに関し、そしてより詳細には白血球の転写発現プロファイルから疾患の診断および予知に関する。

【0002】

長い表

40

本発明には長い表を含み、その全内容は引用により本明細書に編入する。このファイルを含む 2 枚の CD のコピーを横長の方向で添付する。

【発明の背景】

【0003】

遺伝子発現マイクロアレイの広く知られた用途により、生物医学的調査に大きな望みもたれる。この技術は癌患者に予知の確立(非特許文献 1、2、3、4)、そして病原に関与する遺伝子または経路の同定(例えば若年性突発性関節炎の全身的発生の病因におけるインターロイキン - 1 (IL - 1) の役割の発見)を導いた(非特許文献 5)。しかしこれら重大な進歩にもかかわらず、遺伝子発現マイクロアレイ技術は、その初期の騒ぎの期待に添えず、そしてマイクロアレイ平板(platform)から派生した結果は最近

50

、鋭い批判の対象となった(非特許文献6)。中でも主な関心はマイクロアレイのデータが特にノイズの影響を受け易く、そして拡大解釈する場合、懐疑的結果の生成を生む恐れがあるという事実である(非特許文献7)。この疑念はまた、異なる研究室により、およびすべての平板にわたって得られたマイクロアレイデータの評判の悪い再現性からも生じる(非特許文献8、9、10、11、12)。最後にゲノムの全内容における実験結果を解釈するための能力に限界があることも、マイクロアレイ調査における別の障害を構成している(非特許文献13)。

【参考文献】

【0004】

- 【非特許文献1】Golub, T. R. et al. 癌の分子的分類：遺伝子発現のモニタリングによるクラスの見出しおよびクラスの予想 (Molecular classification of cancer: class discovery and class prediction by gene expression monitoring). *Science* 286, 531-7, 1999 10
- 【非特許文献2】Alizadeh, A. A. et al. 遺伝子発現プロファイリングにより同定された拡散大型B細胞リンパ腫 (Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling). *Nature* 403, 503-11, 2000.
- 【非特許文献3】Garber, K. ゲノム創薬。遺伝子発現試験は乳癌の予後を予測する (Genomic medicine. Gene expression tests foretell breast cancer's future). *Science* 303, 1754-5, 2004 20
- 【非特許文献4】van de Vijver, M. J. et al. 乳癌における生存の予測としての遺伝子発現兆候 (A gene-expression signature as a predictor of survival in breast cancer). *N Engl J Med* 347, 1999-2009, 2002
- 【非特許文献5】Pascual, V., Allantaz, F., Arce, E., Punaro, M. & Banchereau, J. 若年性突発性関節炎の全身的発生の病因におけるインターロイキン-1 (IL-1) の役割およびIL-1遮断に対する臨床的応答 (Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade). *J Exp Med* 201, 1479-86, 2005 30
- 【非特許文献6】Michiels, S., Koscielny, S. & Hill, C. マイクロアレイを用いた癌の結果の予測：多数の無作為な検証法 (Prediction of cancer outcome with microarrays: a multiple random validation strategy). *Lancet* 365, 488-92, 2005 40
- 【非特許文献7】Ioannidis, J. P. マイクロアレイおよび分子調査：ノイズ発見？ (Microarrays and molecular research: noise discovery?) *Lancet* 365, 454-5, 2005
- 【非特許文献8】Jarvinen, A. K. et al. 異なる遺伝子マイクロアレイ平板からのデータは比較可能か？ (Are data from different gene expression microarray platforms comparable?) *Genomics* 83, 1164-8, 2004.
- 【非特許文献9】Tan, P. K. et al. 市販されているマイクロアレイ平板からの遺伝子発現測定の評価 (Evaluation of gene expression measurements from commercial microarra 50

y platforms). *Nucleic Acids Res* 31, 5676-84, 2003

【非特許文献10】Bammler, T. et al. 研究室間および平板全体の全体的遺伝子発現分析の標準化 (Standardizing global gene expression analysis between laboratories and across platforms). *Nat Methods* 2, 351-6, 2005

【非特許文献11】Irizarry, R. A. et al. マイクロアレイ平板の多数の研究室による比較 (Multiple-laboratory comparison of microarray platforms). *Nat Methods* 2, 345-50, 2005

10

【非特許文献12】Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R. & Quackenbush, J. マイクロアレイ平板をわたる独立性および再現性 (Independence and reproducibility across microarray platforms). *Nat Methods* 2, 337-44 (2005).

【非特許文献13】Chaussabel, D. 生物医学的文献調査: 'omics' 時代における挑戦および解決 (Biomedical literature mining: challenges and solutions in the 'omics' era). *Am J Pharmacogenomics* 4, 383-93 (2004).

20

【発明の開示】

【0005】

発明の要約

ゲノム調査は、評判の悪いノイズ、解釈が難しく、しかも研究室および平板全体を十分に比較しない転写データの分析で重大な困難に直面している。本発明者は分析の初期段階で生物学的に関連する遺伝子の選択を強調する分析法を開発し、これはマイクロアレイ平板間の矛盾を克服する分析モジュールに統合される。開発された転写モジュールは、膨大な遺伝子発現データベースの分析に使用することができる。この分析から派生する結果は、容易に解釈でき、そして市販のマイクロアレイ平板全体で観察される高い再現性の程度により証明されるように特に強固 (robust) である。

30

【0006】

この分析法に関する応用は、大きな組のPBM C転写プロファイルを調査することを介して具体的に説明される。4742遺伝子に再度グループ分けされる28の転写モジュールが同定された。本発明を使用することにより、疾患はモジュラーレベルで測定される例えば血中白血球における転写の変化の組み合わせにより独自に特徴付けられることを証明することが可能である。実際、血中白血球の転写レベルにおけるモジュール・レベルの変化は、疾患またはサンプルの分子指紋を構成する。

【0007】

本発明は、広い範囲の応用を有する。本発明は任意の生物学系 (例えば末梢血単核細胞 (PBM C)、血液細胞、糞細胞、腹膜細胞、固体の臓器生検材料、切開した腫瘍、一次細胞、細胞株、細胞クローン等) のモジュラー転写成分を特徴付けるために使用することができる。この取り組みを介して集められたモジュラーPBM C転写データは、疾患の重篤度の分子診断的、予知的評価、薬剤処置に対する応答、薬剤毒性等に使用することができる。この取り組みを使用して処理した他のデータは、例えば薬剤化合物の機械作用的実験またはスクリーニングに使用することができる。実際に、データ分析法およびマイニングアルゴリズムは、一般的な遺伝子発現データ分析ソフトウェアで実施することができ、そしてさらに新たな疾患または状態に特異的なモジュールを見だし、開発し、そして試験するために使用することができる。また本発明は薬理遺伝学、分子診断、バイオインフォマティクス等にも使用することができ、ここで詳細な発現データを使用して臨床試験

40

50

中に得ることができる結果を改善することができる（例えばサンプル群を改善し、またはそれからさらに選択することにより）。

【0008】

さらに詳細には本発明は、患者のトランスクリプトームを得：疾患または状態の指標となる1もしくは複数の転写モジュールに基づきトランスクリプトームを分析し：そして患者の疾患または状態を、1もしくは複数の転写モジュールのトランスクリプトーム内の遺伝子発現の存在、不存在またはレベルに基づき決定する、疾患または状態を診断するためのアレイ、装置、システムおよび方法を含む。転写モジュールは：疾患または状態毎に合う各クラスターからそのモジュールについての遺伝子を選択し；選択した遺伝子を分析から除き；そして疾患または状態の亜分画中に集まる遺伝子について遺伝子発現値を選択する工程を繰り返すことにより1もしくは複数の転写モジュールに関する遺伝子発現を反復して選択し；そしてすべての遺伝子クラスターが使用し尽されるまで、各クラスターについてモジュールの作成を繰り返し反復することにより得ることができる。

10

【0009】

本発明と共に使用するために選択されるクラスターの例には、限定するわけではないが発現値クラスター、キーワードクラスター、代謝クラスター、疾患クラスター、感染クラスター、移植クラスター、シグナル伝達クラスター、転写クラスター、複製クラスター、細胞サイクルクラスター、s i R N Aクラスター、m i R N Aクラスター、ミトコンドリアクラスター、T細胞クラスター、B細胞クラスター、サイトカインクラスター、リンホカインクラスター、熱ショッククラスターおよびその組み合わせを含む。本発明を使用して分析するため疾患または状態の例には、例えば自己免疫疾患、ウイルス感染、細菌感染、癌および移植拒絶がある。さらに詳細には、分析のための疾患は1もしくは複数の以下の状態：全身型若年性特発性関節炎、全身性エリテマトーデス、I型糖尿病、肝移植受容者、黒色腫患者および大腸菌、黄色ブドウ球菌のような細菌に感染している患者、インフルエンザAのようなウイルスに感染している患者、およびそれらの組み合わせから選択され得る。特異的アレイはさらに、バイオテクノロジー物質に関連する特異的な疾患または状態を検出するために作成することができる。

20

【0010】

本発明を使用して分析することができる細胞には、例えば末梢血単核細胞（P B M C）、血液細胞、胎児細胞、腹膜細胞、固体の臓器生検材料、切除腫瘍、一次細胞、細胞株、細胞クローンおよびそれらの組み合わせを含む。細胞は単一細胞、細胞の集合、組織、細胞培養物、例えば血液のような体液中の細胞でよい。細胞は組織生検材料、1もしくは複数の分類された細胞群、細胞培養物、細胞クローン、形質転換細胞、生検材料または単一細胞から得ることができる。細胞のタイプは例えば脳、肝臓、心臓、腎臓、肺、脾臓、網膜、骨、ニューロンの、リンパ節、内分泌腺、生殖器官、血液、神経、血管組織および嗅覚上皮細胞でよい。細胞が単離された後、これら細胞からのこれらm R N Aを得、そして個々の遺伝子発現レベルの分析を例えばプローブアレイ、P C R、定量的P C R、ビーズに基づくアッセイおよびそれらの組み合わせを使用して行う。個々の遺伝子発現レベルの分析はさらに、逆転写酵素用の鋳型として細胞から集めたm R N Aから作成したc D N Aを使用して、固体支持体上の核酸のハイブリダイゼーションを使用して行うことができる。

30

40

【0011】

別の態様では、本発明は転写モジュールを同定する方法を含み、この方法は疾患または状態がある1もしくは複数の患者から得た細胞から個々の遺伝子発現レベルを得；クラスターに分割されている表中の各遺伝子について発現値を記録し；各疾患または状態に合う各クラスターからそのモジュールについての遺伝子を選択し；選択した遺伝子を分析から除き；そして疾患または状態の亜分画中に集まる遺伝子について遺伝子発現値を選択する工程を繰り返すことにより1もしくは複数の転写モジュールに関する遺伝子発現を反復して選択し；そしてすべての遺伝子クラスターが使用し尽されるまで、各クラスターについてモジュールの作成を繰り返し反復することによる。本発明で使用するための転写モジュ

50

ールの例は、以下から選択することができる：

【 0 0 1 2 】

【 表 1 】

転写モジュール	
形質細胞:免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含む;	
血小板:血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ-血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む;	
B細胞:B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子:初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ (BLK)をコードする遺伝子を含む;	10
未決定:この組にはcAMPシグナル伝達経路のレギュレーターおよび標的(JUND, ATF4, CREM, PDE4, NR4A2, VIL2)、ならびにTNF-アルファ媒介型NF-KB活性化のリプレッサー(CYLD, ASK, TNFAIP3)をコードする遺伝子を含む;	
ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある:CD14, TLR2, MYD88)をコードする遺伝子。この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;	
未決定:この組にはシグナル伝達分子、活性化STAT(PIAS1およびPIAS2)の垂鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子を含む;	
MHC/リボゾームタンパク質:MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボゾームタンパク質(RPLs, RPSs)をコードする遺伝子によりほとんど排他的に形成される;	20
未決定:代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子を含む;	
細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュリシン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)をコードする遺伝子を含む;	
好中球:この組には好中球顆粒に見いだされる内在性分子(ラクトトランスフェリン: LTF、デフェンシン: DEAF1、細菌透過性増加タンパク質: BPI、カセリシジン抗微生物タンパク質: CAMP…)をコードする遺伝子を含む;	
赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン:ANK1、グリコホリン C:GYPC、ヒドロキシメチルピラン シンターゼ: HMBS、赤血球関連因子: ERAF)をコードする遺伝子を含む;	30
リボゾームタンパク質:リボゾームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー (EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子を含む;	
未決定:このモジュールは免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター(Dedicator)、シンデカン2、プレキシシンC1、ジストロブレピン)をコードする遺伝子を含む;	
ミエロイド系:M 1.5に関連する。単球および好中球のようなミエロイド系細胞で発現する遺伝子(IGTB2/CD18, リンホトキシンベータ受容体, ミエロイド関連タンパク質8/14 ホルミルペプチド受容体 1)を含む;	
未決定:このモジュールはほとんど、未知の機能を持つ転写産物からなる。ケモカイン様因子スーパーファミリー(CKLFSF8)のメンバーを含め、文献に関連するわずか2の遺伝子;	
T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ, TCF7, T細胞分化タンパク質 mal, GATA3, STAT5B)をコードする遺伝子を含む;	40
未決定:細胞骨格に関連する分子 (アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子。また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1)をコードする遺伝子を含む;	

【 0 0 1 3 】

【表 2】

転写モジュール	
未決定:免疫関連細胞表面分子(CD36, CD86, LILRB), サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子を含む;	
未決定:キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK 2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー (G3BP, RAB 14, RAS A2, RAP2A, KRAS)をコードする遺伝子を含む;	
インターフェロン誘導性: この組はインターフェロン誘導性遺伝子:抗ウイルス分子 (OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML)、ケモカイン(CXCL10/IP-10)、シグナル伝達分子 (STAT1, STAT2, IRF7, ISGF3G)をコードする遺伝子を含む;	
炎症I: 炎症プロセスに関与する分子(例えばIL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL 16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B)をコードする遺伝子を含む;	10
炎症II: 顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の (SPI1, IL18, ALOX5, ANPEP)、ならびにリソソーム酵素を誘導する、またはそれにより誘導性の(PPT1, CTSS/S, CES1, NEU1, ASAH1, LAMP2, CAST)の分子をコードする遺伝子を含む;	
未決定:タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ (PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)をコードする遺伝子を含む;	
未決定:少数の転写産物のみからなる。ヘモグロビン遺伝子(HBA1, HBA2, HBB)をコードする遺伝子を含む;	
未決定:この大変大きい組はT細胞表面マーカー(CD101, CD102, CD103)、ならびに血中白血球の回りで偏在的に発現される分子(CXCR1: フラクタルキン受容体, CD47, P-セレクチンリガンド)をコードする遺伝子を含む;	20
未決定:プロテアソームサブユニット(PSMA2/5, PSMB5/8); ユビキチンタンパク質リガーゼ HIP 2, STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子を含む;	
未決定:幾つかの酵素:アミノメチルトランスフェラーゼ、アルギニルトランスフェラーゼ、アスパラギンシンターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼをコードする遺伝子を含む:および	
未決定:プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ (PTPLB, PPP1R8/2CB)をコードする遺伝子を含む。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)も含む;	

【 0 0 1 4 】

ならびにそれらの組み合わせから選択され、ここでサンプル中の遺伝子の発現レベルが、疾患または状態を決定するためにモジュールに対してランク付けされる。

30

【 0 0 1 5 】

本発明は、例えば

【 0 0 1 6 】

【表3】

転写モジュール	
形質細胞:免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含む;	
血小板:血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む;	
B細胞:B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子:初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ (BLK)をコードする遺伝子を含む;	
未決定:この組にはcAMPシグナル伝達経路のレギュレーターおよび標的(JUND, ATF4, CREM, PDE4, NR4A2, VIL2)、ならびにTNF-アルファ媒介型NF- κ B活性化のリプレッサー(CYLD, ASK, TNFAIP3)を含む;	10
ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある:CD14, TLR2, MYD88)を含む。この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;	
未決定:この組にはシグナル伝達分子、活性化STAT(PIAS1およびPIAS2)の亜鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子を含む;	
MHC/リボゾームタンパク質:MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボゾームタンパク質(RPLs, RPSs)をコードする遺伝子によりほとんど排他的に形成される;	
未決定:代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子を含む;	20
細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュリシン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)を含む;	
好中球:この組には好中球顆粒に見いだされる内在性分子(ラクトランスフェリン: LTF、デフェンシン: DEAF1、細菌透過性増加タンパク質: BPI、カセリシジン抗微生物タンパク質: CAMP...)をコードする遺伝子を含む;	
赤血球:ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン: ANK1、グリコホリン C:GYPC、ヒドロキシメチルピラン シンターゼ: HMBS、赤血球関連因子: ERAF)を含む;	
リボゾームタンパク質: リボゾームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー (EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子を含む;	30
未決定:このモジュールは免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター(Dedicator)、シンデカン2、プレキシシンC1、ジストロブレピン)をコードする遺伝子を含む;	
ミエロイド系:M 1.5に関連する。単球および好中球のようなミエロイド系細胞で発現する遺伝子(IGTB2/CD18, リンホトキシンベータ受容体, ミエロイド関連タンパク質8/14 ホルミルペプチド受容体 1)を含む;	
未決定:このモジュールはほとんど、未知の機能を持つ転写産物からなる。ケモカイン様因子スーパーファミリー(CKLFSF8)のメンバーを含め、文献に関連するわずか2の遺伝子;	
T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ, TCF7, T細胞分化タンパク質 mal, GATA3, STAT5B)を含む;	
未決定:細胞骨格に関連する分子 (アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子。また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1)をコードする遺伝子を含む;	40

【0017】

【表4】

転写モジュール	
未決定:免疫関連細胞表面分子(CD36, CD86, LILRB), サイトカイン(IL15)およびシグナル伝達経路に關与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子を含む;	
未決定:キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK 2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー (G3BP, RAB 14, RAS A2, RAP2A, KRAS)を含む;	
インターフェロン誘導性: この組はインターフェロン誘導性遺伝子:抗ウイルス分子(OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML)、ケモカイン(CXCL10/IP-10)、シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)を含む;	
炎症I: 炎症プロセスに關与する分子(例えばIL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B)をコードする遺伝子を含む;	10
炎症II: 顆粒球マクロファージCSFを誘導する、またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP)、ならびにリソソーム酵素を誘導する、またはそれにより誘導性の(PPT1, CTSS/S, CES1, NEU1, ASAH1, LAMP2, CAST)の分子を含む;	
未決定:タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ(PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)を含む;	
未決定:少数の転写産物のみからなる。ヘモグロビン遺伝子(HBA1, HBA2, HBB)をコードする遺伝子を含む;	
未決定:この大変大きい組はT細胞表面マーカー(CD101, CD102, CD103)、ならびに血中白血球の回りで偏在的に発現される分子(CXCR1: フラクタルキン受容体, CD47, P-セレクチンリガンド)を含む;	20
未決定:プロテアソームサブユニット(PSMA2/5, PSMB5/8); ユビキチンタンパク質リガーゼ HIP 2, STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子を含む;	
未決定:幾つかの酵素:アミノメチルトランスフェラーゼ、アルギニルトランスフェラーゼ、アスパラギンシンターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼをコードする遺伝子を含む;および	
未決定:プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ(PTPLB, PPP1R8/2CB)をコードする遺伝子を含む。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4(CD244)も含む;	

【0018】

からなる群から選択される1もしくは複数の遺伝子モジュールを含み、自己免疫疾患、ウイルス感染、細菌感染、癌および移植拒絶の間を識別するために十分である疾患分析ツールを含む。

30

【0019】

1つの態様では、選択されるモジュールが

形質細胞:免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含む;および

血小板:血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ-血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む;

から選択され、そしてこのモジュールがこれら2つのモジュールに陽性ベクトルを有することにより全身性エリテマトーデスを同定するために使用される。

40

【0020】

別の態様では、選択されるモジュールが:

形質細胞:免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含み;および

血小板:血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ-血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む;

から選択されることができ、そしてこのモジュールがこれら2つのモジュールに陽性ベクトルも陰性ベクトルも持たないことによりインフルエンザ感染を同定するために使用され

50

る。

【0021】

別の態様では、選択されるモジュールが：

形質細胞：免疫グロブリン鎖（IGHM，IGJ，IGLL1，IGKC，IGHD）および形質細胞マーカーCD38をコードする遺伝子を含み；および

血小板：血小板糖タンパク質（ITGA2B，ITGB3，GP6，GP1A/B）、およびPPPB（プロ-血小板塩基性タンパク質）およびPF4のような（血小板因子4）

血小板由来免疫メディエーターをコードする遺伝子を含む；

から選択され、そしてこのモジュールが形質細胞マーカーに陰性ベクトルを、そして血小板マーカーに陽性ベクトルを有することにより黒色腫を同定するために使用される。

10

【0022】

別の態様では、選択されるモジュールが：

形質細胞：免疫グロブリン鎖（IGHM，IGJ，IGLL1，IGKC，IGHD）および形質細胞マーカーCD38をコードする遺伝子を含む；および

血小板：血小板糖タンパク質（ITGA2B，ITGB3，GP6，GP1A/B）、およびPPPB（プロ-血小板塩基性タンパク質）およびPF4（血小板因子4）のような

血小板由来免疫メディエーターをコードする遺伝子を含む；

から選択され、そしてこのモジュールがこれら2つのモジュールに陰性ベクトルを有することにより移植拒絶を同定するために使用される。

20

【0023】

別の態様では、選択されるモジュールが：

形質細胞：免疫グロブリン鎖（IGHM，IGJ，IGLL1，IGKC，IGHD）および形質細胞マーカーCD38をコードする遺伝子を含む；および

血小板：血小板糖タンパク質（ITGA2B，ITGB3，GP6，GP1A/B）、およびPPPB（プロ-血小板塩基性タンパク質）およびPF4（血小板因子4）のような

血小板由来免疫メディエーターをコードする遺伝子を含む；

から選択され、そしてこのモジュールがこれら2つのモジュールに陰性ベクトルを有することによりインフルエンザ感染を同定するために使用される。

【0024】

さらに別の態様では、本発明は1もしくは複数の転写モジュールの代表である遺伝子の組み合わせを有するカスタマイズされた遺伝子アレイを含む予知遺伝子アレイであり、ここでカスタマイズされた遺伝子アレイと接触する患者のトランスクリプトームは、転写モジュールに合った1もしくは複数の疾患または状態の予知である。1例では、疾患または状態に対する患者の免疫応答が、転写モジュールと特異的疾患または状態との相関に基づきトランスクリプトームの遺伝子発現の存在、不存在またはレベルに基づき決定される。このアレイは自己免疫疾患、ウイルス感染、細菌感染、癌および移植拒絶の間を識別することができる。このアレイは2以上の転写モジュールに組織化され得る。例えばアレイは：

30

【0025】

【表5】

サブモジュール	プローブの組数	キーワードの選択	評価
M 1.1	69	Ig、免疫グロブリン、骨、骨髄、PreB、IgM、Mu.	形質細胞。免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHD)およびプラスマ細胞マーカーCD38をコードする遺伝子を含む;
M 1.2	96	血小板、接着、凝集、内皮、脈管	血小板。血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む;
M 1.3	47	イムノレセプター BCR、B細胞、IgG	B細胞。B細胞表面マーカー(CD72, CD79 A/B, CD19, CD22)および他のB細胞関連分子: 初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ(BLK)をコードする遺伝子を含む;
M 1.4	87	複製、抑制、修復、CREB、リンパ系、TNF-アルファ	未決定。この組はcAMPシグナル伝達経路(JUND, ATF4, CREM, PDE4, NR4A2, VIL2)のレギュレーターおよび標的、ならびにTNF-アルファ媒介型NF-KB活性化のリプレッサー(CYLD, ASK, TNFAIP3)を含む;
M 1.5	130	単球、樹状突起の、MHC、同時刺激、TLR4、MYD88	ミエロイド系。ミエロイド系の細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある: CD14, TLR2, MYD88)の細胞により発現される分子を含む。この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;
M 1.6	28	亜鉛、フィンガー、P53、RAS	未決定。この組はシグナル伝達分子、活性化STAT(PIAS1およびPIAS2)の亜鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子を含む;
M 1.7	127	リボソーム、翻訳、40S、60S、HLA	MHC/リボソームタンパク質。MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボソームタンパク質(RPLs, RPSs)をコードする遺伝子によりほとんど排他的に形成される;
M 1.8	86	代謝、生合成、複製、ヘリカーゼ	未決定。代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子を含む;
M 2.1	72	NK、キラー、細胞溶解、CD8、細胞媒介型、T細胞、CTL、IFN-g	細胞傷害性細胞。細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュリシン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)を含む;

【0026】

【表6】

サブモジュール	プローブの組数	キーワードの選択	評価	
M 2.2	44	顆粒球、好中球、防御、ミエロイド、骨髄	好中球。この組は好中球顆粒(ラクトランスフェリン: LTF、デフェンシン: DEAF1、細菌透過性増加タンパク質: BPI、カセリシジン抗微生物タンパク質: CAMP…)に見いだされる内在性分子を含む;	
M 2.3	94	赤血球、赤、貧血、グロビン、ヘモグロビン	赤血球。ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン: ANK1、グリコホリン C: GYPC、ヒドロキシメチルピラン シンターゼ: HMBS、赤血球関連因子: ERAF)を含む;	10
M 2.4	118	リボヌクレオタンパク質、60S、核小体、アッセムブリ、延長	リボゾームタンパク質: リボゾームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー (EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子を含む;	
M 2.5	242	腺癌、間質の、間葉、樹状突起、運動	未決定。このモジュールは免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター、シンデカン 2、プレキシシン C1、ジストロプレビン)をコードする遺伝子を含む;	20
M 2.6	110	顆粒球、単球、ミエロイド、ERK、壊死	ミエロイド系。M1.5に関連する。単球および好中球のようなミエロイド系細胞で発現する遺伝子(IGTB2/CD18、リンホトキシンベータ受容体、ミエロイド関連タンパク質8/14 ホルミルペプチド受容体 1)を含む;	
M 2.7	43	キーワードは抽出されない	未決定。このモジュールは、未知の機能の転写産物からほぼ構成される。ケモカイン様因子スーパーファミリー(CKLFSF8)の1メンバーを含め、文献に関連するわずか20個の遺伝子;	30
M 2.8	104	リンパ腫、T細胞、CD4、CD8、TCR、胸腺、リンパ系、IL2	T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ, TCF7, T細胞分化タンパク質 mal, GATA3, STAT5B)を含む;	
M 2.9	122	ERK、トランス活性化、細胞骨格、MAPK, JNK	未決定。細胞骨格に関連する分子(アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子を含む。また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1);	
M 2.10	44	ミエロイド、マクロファージ、樹状突起、炎症、インターロイキン	未決定。免疫関連細胞表面分子(CD36, CD86, LILRB)、サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子を含む;	40

【 0 0 2 7 】

【表7】

サブモジュール	プローブの組数	キーワードの選択	評価
M 2.11	77	複製、抑制、RAS、自己リン酸化作用、発癌遺伝子	未決定。キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK 2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A, KRAS)を含む;
M 3.1	80	ISRE、インフルエンザ、抗ウイルス、IFN-ガンマ、IFN-アルファ、インターフェロン	インターフェロン誘導性。この組はインターフェロン誘導性遺伝子: 抗ウイルス分子 (OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML)、ケモカイン(CXCL10/IP-10)、シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)を含む;
M 3.2	230	TGF-ベータ、TNF、炎症、アポトーシスの、リポ多糖	炎症I。炎症プロセスに関与する分子 (IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B)をコードする遺伝子を含む;
M 3.3	230	顆粒球、炎症、防御、酸化する、リソソームの	炎症II。顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP)、ならびにリソソーム酵素を誘導する、またはそれにより誘導性の(PPT1, CTSB/S, CE S1, NEU1, ASAH1, LAMP2, CAST)の分子を含む;
M 3.4	323	キーワードは抽出されない	未決定。タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ (PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)を含む;
M 3.5	19	キーワードは抽出されない	未決定。わずかな数の転写産物から構成される。ヘモグロビン遺伝子(HBA1, HBA2, HBB)を含む;
M 3.6	233	補体、宿主、酸化的、細胞骨格、T細胞	未決定。この大変大きい組はT細胞表面マーカー(CD101, CD102, CD103)、ならびに血中白血球の回りで偏在的に発現される分子(CXRCR1: フラクタルキン受容体, CD47, P-セレクチンリガンド)を含む;
M 3.7	80	スプライソソーム、メチル化、ユビキチン、ベーターカテニン	未決定。プロテアソームサブユニット(PSMA2/5, PSMB5/8); ユビキチンタンパク質リガーゼ HIP2, STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子を含む;

10

20

30

【0028】

40

【表 8】

サブモジュール	プローブの組数	キーワードの選択	評価
M 3.8	182	CDC、TCR、CREB、グリコシラーゼ	未決定。幾つかの酵素:アミノメチルトランスフェラーゼ、アルギニルトランスフェラーゼ、アスパラギンシンターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼをコードする遺伝子を含む: および
M 3.9	261	クロマチン、チェックポイント、複製、トランス活性化	未決定。プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ(例えばPTPLB, PPP1R8/2CB)をコードする遺伝子を含む。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)も含む;

10

【 0 0 2 9 】

から選択される 1 もしくは複数のサブモジュールを含んでなる 3 つの転写モジュールに組織化されることができ、ここで各々から 1 もしくは複数のプローブがモジュールの 1 もしくは複数の遺伝子に特異的に結合する。

【 0 0 3 0 】

さらに別の本発明には、

【 0 0 3 1 】

20

【表9】

キーワードの選択	転写モジュール	
Ig、免疫グロブリン、骨、骨髄、PreB、IgM、Mu.	形質細胞。免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含む。	
血小板、接着、凝集、内皮、血管	血小板。血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む。	
イムノレセプター、BCR、B細胞、IgG	B細胞。B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子: 初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ(BLK)をコードする遺伝子を含む	10
複製、抑制、修復、CREB、リンパ系、TNF-アルファ	この組はcAMPシグナル伝達経路レギュレーターおよび標的(JUND, ATF4, CREM, PDE4, NR4A2)、ならびにTNF-アルファ媒介型NF-κB活性化のリプレッサー(CYLD, ASK, TNFAIP3)をコードする遺伝子を含む。	
単球、樹状突起の、MHC、同時刺激、TLR4、MYD88	ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある: CD14, TLR2, MYD88)の細胞により発現される分子を含む。この組はTNFファミリーメンバー(TNFR2, BAFF)も含む。	
垂鉛、フィンガー、P53、RAS	未決定。この組はシグナル伝達分子、例えば活性化STAT (PIAS1およびPIAS2)の垂鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子を含む。	
リボソーム、翻訳、40S、60S、HLA	MHC/リボソームタンパク質:MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボソームタンパク質(RPLs, RPSs)をコードする遺伝子によりほぼ排他的に形成される。	20
代謝、生合成、複製、ヘリカーゼ	未決定。代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子を含む。	
NK、キラー、細胞溶解、CD8、細胞媒介型、T細胞、CTL、IFN-γ	細胞傷害性細胞。細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュリシン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)を含む。	
顆粒球、好中球、防御、ミエロイド、骨髄	好中球。この組は好中球顆粒に見いだされる内在性分子(ラクトラン、スフェリン: LTF、デフェンシン: DEAF1、細菌透過性増加タンパク質: BPI、カセリジン抗微生物タンパク質: CAMP...)を含む。	
赤血球、赤、貧血、グロビン、ヘモグロビン	赤血球。ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン: ANK1、グリコホリン C: GYPC、ヒドロキシメチルピランシンターゼ: HMBS、赤血球関連因子: ERAF)を含む。	30
リボヌクレオタンパク質、60S、核小体、アッセムブリー、延長	リボソームタンパク質。リボソームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー(EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子;	
腺癌、間質の、間葉、樹状突起、	未決定。このモジュールは免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター、シンデカン2、プレキシシンC1、ジストロブレピン)をコードする遺伝子を含む。	
顆粒球、単球、ミエロイド、ERK、壊死	ミエロイド系細胞。M1.5に関連する。単球および好中球のようなミエロイド系細胞で発現する分子(IGTB2/CD18, リンホトキシンベータ受容体、ミエロイド関連タンパク質 8/14ホルミルペプチド受容体 1)をコードする遺伝子を含む。	40

【0032】

【表 10】

キーワードの選択	転写モジュール	
キーワードは抽出されない	未決定。このモジュールは未知の機能の転写産物からほとんどなる。ケモカイン様因子スーパーファミリー (CKLFSF8)メンバーを含む文献に関連するわずか20の遺伝子。	
リンパ腫、T細胞、CD4、CD8、TCR、胸腺、リンパ系、IL12	T細胞。T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子、(リンホトキシンベータ IL2-誘導性T細胞キナーゼ, TCF7, T細胞分化タンパク質 mal, GATA3, STAT5B)を含む。	
ERK、トランス活性化、細胞骨格、MAPK, JNK	未決定。細胞骨格に関連する分子 (アクチン関連タンパク質2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子を含む。また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1)。	10
ミエロイド、マクロファージ、樹状突起、炎症、インターロイキン	未決定。免疫関連細胞表面分子(CD36, CD86, LILRB)、サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB, TICAM2-Toll様受容体経路)をコードする遺伝子を含む。	
複製、抑制、RAS、自己リン酸化、発癌遺伝子	未決定。キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2)およびRAS ファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A, KRAS)を含む。	
ISRE、インフルエンザ、抗ウイルス、IFN-ガンマ、IFN-アルファ、インターフェロン	インターフェロン誘導性。この組はインターフェロン誘導性遺伝子: 抗ウイルス分子 (OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML)、ケモカイン (CXCL10/IP-10)、シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)を含む。	20
TGF-ベータ、TNF、炎症、アポトーシスの、リポ多糖	炎症I。炎症プロセスに関与する分子(IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B)をコードする遺伝子を含む。	
顆粒球、炎症、防御、酸化する、リソソームの	炎症II。顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP)、ならびにリソソーム酵素を誘導する、またはそれにより誘導性(PPT1, CTSB/S, CES1, NEU1, ASAH1, LAMP2, CAST)の分子を含む。	
キーワードは抽出されない	未決定。タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ (PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)を含む。	30
キーワードは抽出されない	未決定。少数の転写産物のみよりなる。ヘモグロビン遺伝子(HBA1, HBA2, HBB)含む。	
補体、宿主、酸化的、細胞骨格、T細胞	未決定。この大きい組はT細胞表面マーカー(CD101, CD102, CD103)、ならびに白血球の回りで偏在的に発現される分子(CXCR1: フラクタルキン受容体, CD47, P-セレクチンリガンド)を含む。	
スプライソソーム、メチル化、ユビキチン、ベーターカテニン	未決定。プロテアソームサブユニット(PSMA2/5, PSMB5/8)ユビキチンタンパク質リガーゼ HIP2, STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子を含む。	
CDR、TCR、CREB、グリコシラーゼ	未決定。幾つかの酵素: アミノメチルトランスフェラーゼアルギニル転スフェラーゼ、アスパラギンシンターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼをコードする遺伝子を含む: および	40
クロマチン、チェックポイント、複製、トランス活性化	未決定。プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ (PTPLB, PPP1R8/2CB)をコードする遺伝子を含む。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)も含む;	

【0033】

を含んでなる左カラムから選択される1グループ、および右カラムから選択される1グループの組み合わせ、およびその組み合わせから選択される1もしくは複数の遺伝子モジュールを含む遺伝子分析ツールを含み、ここで疾患または状態を決定するためにモジュールに対してサンプル中の遺伝子発現のレベルがランク付けされる。

【0034】

本発明のアレイ、方法およびシステムは：予想される患者のトランスクリプトームを得；トランスクリプトームを、臨床試験で処置される疾患または状態の指標となる1もしくは複数の転写モジュールに対して比較し；そして患者が臨床試験の良い候補となる見込を、臨床試験での成功と相関する1もしくは複数の転写モジュール内の患者のトランスクリプトームで発現される1もしくは複数の遺伝子の存在、不存在またはレベルに基づき決定することにより、臨床試験の患者を選択するために使用することができる。一般に各モジュールについて、サンプル中の転写産物の比率の和と相関するベクトルを使用することができる。例えば各モジュールがベクトルを含んでなり、そしてここで1もしくは複数の疾患または状態がこの1もしくは複数のベクトルと関連する。したがって各モジュールは、各モジュール内の1もしくは複数の遺伝子の発現レベルと相関するベクトルを含んでなることができる。

10

【0035】

また本発明は、1もしくは複数の疾患の間を識別するために差次的に発現される遺伝子の十分な比率を提供するために、1もしくは複数のモジュールから十分なプローブを含む固体支持体上に固定化された核酸プローブを含むアレイ、例えばカスタムマイクロアレイを含み、このプローブは表3から選択される。例えば固体支持体上に固定化された核酸プローブのアレイ、ここでアレイは；

【0036】

【表 1 1】

モジュールI. D.	転写モジュール	
M 1.1	形質細胞。免疫グロブリン鎖(IGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含む。	
M 1.2	血小板。血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ-血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む。	
M 1.3	B細胞。B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子:初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシキナーゼ(BLK)をコードする遺伝子を含む。	
M 1.4	未決定。この組はcAMPシグナル伝達経路(JUND, ATF4, CREM, PDE4, NR4A2, VIL2)のレギュレーターおよび標的、ならびにTNF-アルファ媒介型NF-KB活性化のリプレッサー(CYLD, ASK, TNFAIP3)を含む。	10
M 1.5	ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある:CD14, TLR2, MYD88)の細胞により発現される分子を含む。この組はTNFファミリーメンバー(TNFR2,BAFF)も含む;	
M 1.6	未決定。この組はシグナル伝達分子、例えば活性化STAT(PLAS1およびPIAS2)の亜鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子を含む。	
M 1.7	MHC/リボゾームタンパク質。MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボゾームタンパク質(RPLs, RPSs)をコードする遺伝子によりほとんど排他的に形成される。	20
M 1.8	未決定。代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA,TERF2,EIF2S1)をコードする遺伝子を含む。	
M 2.1	細胞傷害性細胞:細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュリシン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)を含む。	
M 2.2	好中球。この組は好中球顆粒に見いだされる内在性分子(ラクトランスフェリン: LTF、デフェンシン:DEAF1,細菌透過性増加タンパク質: BPI、カセリジジン抗微生物タンパク質: CAMP...)を含む	
M 2.3	赤血球。ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン:ANK1、グリコホリンC:GYPC、ヒドロキシメチルビルンシンターゼ:HMBS、赤血球関連因子: ERAF)を含む;	30
M 2.4	リボゾームタンパク質。リボゾームタンパク質(RPLs, RPSs)、真核生物翻訳延長因子ファミリーメンバー(EEFs)および核小体タンパク質(NPM1,NOAL2, NAP1L1)をコードする遺伝子を含む。	
M 2.5	未決定。このモジュールは免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター、シンデカン 2、プレキシシンC1, ジストロブレビン)をコードする遺伝子を含む。	
M 2.6	ミエロイド系:単球および好中球のようなミエロイド系細胞で発現する遺伝子(IGTB2/CD18,リンホトキシンベータ受容体,ミエロイド関連タンパク質 8/14ホルミルペプチド受容体 1)を含む。	
M 2.7	未決定。このモジュールは、ほとんど未知の機能を持つ転写産物からなる。ケモカイン様因子スーパーファミリーのメンバー(CKLFSF8)を含め、文献に関連するわずか20の遺伝子。	40

【 0 0 3 7 】

【表 1 2】

モジュールI. D.	転写モジュール
M 2.8	T細胞。T細胞表面マーカー(CD5,CD6,CD7,CD26,CD28,CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ,TCF7,T細胞分化タンパク質mal, GATA3, STAT5B)を含む。
M 2.9	未決定。細胞骨格に関連する分子(アクチン関連タンパク質2/3,MAPK1, MAP3K1, RAB5A)をコードする遺伝子を含む。また存在するのは、T細胞が発現する遺伝子(FAS,ITGA4/CD49D, ZNF1A1)。
M 2.10	未決定。免疫関連細胞表面分子(CD36, CD86, LILRB)、サイトカイン(IL15)およびシグナル伝達経路に関与する分子(FYB,TICAM2-Toll-様受容体経路)をコードする遺伝子を含む。
M 2.11	未決定。キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2,PRKCI, ITPKB, SRPK2, STK17B, DYRK2,PIK3R1, STK4, CLK4, PKN2)およびRASファミリーメンバー(G3BP, RAB14, RAS A2, RAP2A,KRAS)を含む。
M 3.1	インターフェロン-誘導性。この組はインターフェロン誘導性遺伝子:抗ウイルス分子(OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1,PML)、ケモカイン(CXCL10/IP-10)、シグナル伝達分子(STAT1,STAT2, IRF7, ISGF3G)を含む。
M 3.2	炎症I: 炎症プロセスに関与する分子(例えばIL8,ICAM1,C5R1, CD44,PLAUR, IL1A, CXCL16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B)をコードする遺伝子を含む。
M 3.3	炎症II: 顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP)、ならびにリソソーム酵素を誘導する、またはそれにより誘導性(PPT1, CTSB/S, CES1, NEU1,ASAHI, LAMP2, CAST)の分子を含む。
M 3.4	未決定。タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ(PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)を含む。
M 3.5	未決定。少数の転写産物のみからなる。ヘモグロビン遺伝子(HBA1, HBA2, HBB)を含む。
M 3.6	未決定。この大きな組はT細胞表面マーカー(CD101, CD102, CD103)、ならびに血中白血球の回りで偏在的に発現される分子(CXCR1:フラクタルキン受容体,CD47, P-セレクチンリガンド)を含む。
M 3.7	未決定。プロテアソームサブユニット(PSMA2/5, PSMB5/8);ユビキチンタンパク質リガーゼHIP2,STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子を含む。
M 3.8	未決定。幾つかの酵素:アミノメチルトランスフェラーゼ、アルギニルトランスフェラーゼ、アスパラギンシンターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼをコードする遺伝子を含む。
M 3.9	未決定。プロテインキナーゼ(例えばPRKPIR, PRKDC, PRKCI)およびホスファターゼ(PTPLB,PPP1R8/2CB)を含む。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4(CD244)を含む。

10

20

30

【0038】

から選択される少なくとも2組のプローブモジュールを含み、ここで第1組のプローブが、1もしくは複数の疾患にそれぞれ対応する1もしくは複数の質問位置を有する。アレイは100から1000個のプローブを有することができ、そして各プローブは例えば9~21ヌクレオチド長である。組織化されたプローブの組に分離された場合、これらは別個に質問され得る。

40

【0039】

また本発明は、固体支持体に固定化され、少なくとも1対の第1および第2プローブ群を含むアレイを形成する1もしくは複数の核酸プローブを含み、各群が表3により定められる1もしくは複数のプローブを有する。プローブ群は、マイクロアレイ平板全体に一貫した複合転写マーカーベクトルを提供するように選択される。実際に、プローブ群はマイクロアレイ平板全体に一貫した複合転写マーカーベクトルを提供し、そして調節承認に関す

50

る要約を表示するために使用され得る。当業者は、本発明のモジュールを使用して、異なる疾患および/または状態の間を迅速に診断し、または識別するために使用できる1もしくは複数の疾患特異的アレイを迅速に開発することが可能であると考える。

【0040】

発明の詳細な説明

本発明の様々な態様を作成し、そして使用することを以下で詳細に検討するが、本発明は広い様々な特定の内容で具現化され得る多くの応用可能な革新的概念を提供すると考えるべきである。本明細書で検討する特別な態様は、本発明を作成し、そして使用するための特定の方法の単に具体的説明であり、本発明の範囲に限界を定めるものではない。

【0041】

本発明の理解を容易にするために、多くの用語を以下に定義する。本明細書で定義する用語は、本発明に関する当業者が一般に理解している意味を有する。“a”、“an”および“the”のような用語は、単数の物体を指すことのみを意図するのではなく、特定例の具体的説明に使用できる一般的種類を含む。本明細書の実語法は、本発明の具体的態様を説明するために使用されるが、それらの用途は特許請求の範囲で境界を定めることを除き、本発明を限定しない。他に定義しない限り、本明細書で使用するすべての技術的および科学的用語は、本発明が属する当業者に通常に理解されている意味を有する。以下の参照は、本発明で使用される多くの用語の一般的定義を用いて技術の1つを提供する：Singleton et al., 微生物学および分子生物学辞典 (DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY (2nd ed., 1994)); 科学および技術のケンブリッジ辞書 (THE CAMBRIDGE DICTIONARY OF SCIENCE AND TECHNOLOGY) (Walker ed., 1988); 遺伝学の実語解説 (THE GLOSSARY OF GENETICS), 5TH ED., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY (1991).

【0042】

様々な生化学的および分子生物学的方法が当該技術分野では知られている。例えば核酸の分離および精製法は、国際公開第97/10365号、同第97/27317明細書、生物学および分子生物学の研究室技術：核酸プローブでのハイブリダイゼーション (Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes) の第3章、I部、理論および核酸の調製 (Theory and Nucleic Acid Preparation), (P. Tijssen, ed.) エルセビア (Elsevier), N.Y. (1993); 生物学および分子生物学の研究室技術：核酸プローブでのハイブリダイゼーション (Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes) の第3章、I部、理論および核酸の調製 (Theory and Nucleic Acid Preparation), (P. Tijssen, ed.) エルセビア (Elsevier), N.Y. (1993); および Sambrook et al., モレキュラークローニング：ア ラボラトリーマニュアル (Molecular Cloning: A Laboratory Manual), コールドスプリングハーバー出版 (Cold Spring Harbor Press), N.Y., (1989); および分子生物学の現在的手法 (Current Protocols in Molecular Biology), (Ausubel, F.M. et al., eds.) ジョン ウィリー & サンズ社 (John Wiley & Sons, Inc.), ニューヨーク (1987-1999) (追補46 (1999年4月のような追補を含む) に詳細に記載されている。

10

20

30

40

50

【 0 0 4 3 】

バイオインフォマティクスの定義

本明細書で使用する「オブジェクト (o b j e c t) 」は、問題とする任意の項目または情報を称する（一般に原文どおり、名詞、動詞、形容詞、副詞、句、文、記号、数字等を含む）。したがってオブジェクトは関係を形成することができる任意のもであり、また供給源から得られ、同定され、かつ/または調査される任意のものである。「オブジェクト」には限定するわけではないが、遺伝子、タンパク質、疾患、表現型、メカニズム、薬剤等のような問題の実態を含む。幾つかの観点では、オブジェクトは以下にさらに記載するようなデータでよい。

【 0 0 4 4 】

本明細書で使用する「関係」とは、同じ単位（例えば、句、文、明細書中の2以上の行、段落、ウェブページの章、ページ、雑誌、新聞、本等の欄）内でのオブジェクトの共起を指す。これはテキスト、記号、数字およびその組み合わせでよい。

【 0 0 4 5 】

本明細書で使用する「メタデータ コンテンツ (m e t a d a t a c o n t e n t) 」とは、データソース (d a t a s o u r c e) におけるテキストの組織化に関する情報を指す。メタデータは、ダブリン コア (D u b l i n C o r e) メタデータのような標準的メタデータを含んでなることができ、あるいはコレクション - スペシフィック (c o l l e c t i o n - s p e c i f i c) であることができる。メタデータ形式の例には限定するわけではないが、ライブラリータカログに使用されるマシーン リーダブル カタログ (M a c h i n e R e a d a b l e C a t a l o g : M A R C) レコード、リソースディスクリプションフォーマット (R e s o u r c e D e s c r i p t i o n F o r m a t : R D F) およびエクステンシブル マークアップ ランギッジ (E x t e n s i b l e M a r k u p L a n g u a g e : X M L) がある。メタデータ オブジェクトはマニュアルで、または自動化情報抽出アルゴリズムを介して作成することができる。

【 0 0 4 6 】

本明細書で使用する「エンジン」は、他のプログラムの中心または本質的機能を行うプログラムを指す。例えばエンジンは、他のプログラムの全体的操作を調整する操作システムまたはアプリケーションプログラムの中心的プログラムであり得る。また用語「エンジン」は、変更することができるアルゴリズムを含むプログラムも指すことができる。例えば関係を同定することに対するそのアプローチを変更するために、知識発見エンジンを設計し、同定およびランキングの関係の新規法則を反映させることができる。

【 0 0 4 7 】

本明細書で使用する「意味分析 (s e m a n t i c a n a l y s i s) 」とは、例えば接尾辞の除去またはステミング (s t e m m i n g) を介して、またはシソーラスを採用することにより同じ概念を表す言葉の間の関係の確認を指す。「統計分析」とは各用語（言葉、語源、語幹、n - グラム、句等）の発生数の計数に基づく技術を指す。主題に対して無制限に集めると、異なる内容で使用される同じ句は、異なる概念を表すか可能性がある。同時に存在する句の統計的分析は、言葉の意味の曖昧さを解決するために役立つ。 「シンタックス分析 (S y n t a c t i c a n a l y s i s) 」は、品詞分析によりさらに曖昧さを減らすために使用することができる。本明細書で使用するように、1もしくは複数のそのような分析は、さらに一般的に「字句解析 (l e x i c a l a n a l y s i s) 」と呼ばれる。「人工知能 (A I) 」は、コンピューターのような人間以外のデバイスにより、ヒトが価値があると、または「知的」であると見なす作業を行う方法を指す。例には絵を識別すること、会話または書かれた内容を理解すること、および問題を解決することを含む。

【 0 0 4 8 】

様々な情報ファセットをデータフィールド内に見いだすことができる場合でも、本明細書で使用する用語「データベース」は、生の、またはコンパイルされたデータに関する

10

20

30

40

50

貯蔵所を指す。データベースは典型的にはその内容にアクセスでき、管理でき、そして更新できるように組織化される（例えばデータベースは動的である）。また用語「データベース」および「ソース」は、データおよび情報の主な供給源がデータベースであるので、本発明では互換的に使用される。しかし「ソースデータベース」または「ソースデータ」は一般にデータ、例えばオブジェクトを同定し、そして関係を決定するためにシステムに入力される非構造テキストおよび/または構造データを指す。ソースデータベースはリレーショナル(re relational)データベースであってもなくてもよい。しかしシステムデータベースは通常、リレーショナルデータベースまたは同じ均等なタイプのデータベースを含み、これらはオブジェクト間の関係を関連づける値を保存する。

10

【0049】

本明細書で使用する「システムデータベース」および「リレーショナルデータベース」は、互換的に使用され、そして予め定義されたカテゴリーに合うデータを含む1組の表として組織化されたデータの1もしくは複数の集合を称する。例えばデータベースの表は、カラム（例えば属性）に定められる1もしくは複数のカテゴリーを含んでなることができ、同時にデータベースの横行はカラムにより定まるカテゴリーについて、独自のオブジェクトを含むことができる。すなわち遺伝子の同定のようなオブジェクトはその存在、不存在および/または遺伝子発現のレベルについてのカラムを有する。またリレーショナルデータベースの横行は「組」とも呼ばれ、そして一般にそのカラムの値により定められる。リレーショナルデータベースの内容における「ドメイン」は有効な値の範囲

20

【0050】

本明細書で使用する「知識のドメイン」は、システムが作動する実験のエリア、例えばすべての生物医学データを指す。この多様なデータは、1つの分野または調査/研究にのみ精通している（1ドメイン）標準的な人が、一緒にすることはできない事柄を時折、連結することができるので、幾つかのドメインからデータ（例えば生物医学データおよびエンジニアリングデータ）を組み合わせることは有利であると指摘すべきである。「分散データベース」とはネットワークの異なる点に分散または複製され得るデータベースを指す。

【0051】

「データ」および「情報」のような用語は、「情報」および「知識」のようにしばしば互換的に使用される。本明細書で使用する「データ」は、経験的測定値または測定値の組である最も基本的な単位である。データは情報の一助となるように処理されるが、基本的には情報から独立している。対照的に情報は興味から派生し、例えばデータ（単位）は、心血管疾患の危険性と相関し得る変数を見いだす目的で、民族性、性別、身長、体重、および食事について集めることができる。しかし同じデータを使用して式を開発するか、または食事の嗜好についての「情報」、すなわち特定の製品がスーパーマーケットでより良く売れる高い見込みを作成することができる。

30

【0052】

本明細書で使用するように「情報」にはデータの組から生じるまたは引き出される数、文字、数の組、文字の組または結果を含むことができる。次いで「データ」は、情報の測定値または統計および基準単位である。また「情報」は、言葉、記号のような他の種類のデータ、非構造フリーテキストのようなテキスト、コード等を含むことができる。「知識」は、原因および効果をモデル化するために系に十分な理解を与える情報の組として大雑把に定義される。以前の例を拡大するために、人口統計学、性別およびこれまでの買い物に関する情報を使用して、食品販売に関する地域的マーケティング戦略を開発し、一方、購買者による民族性に関する情報を製品の輸入のための指針として使用することができる。データ、情報および知識の間には厳密な境界が無いことに注目することは重要である：3つの用語は時折、等価と考えられる。一般に、データは調査から生じ、情報は相関から生じ、そして知識はモデル化から生じる。

40

50

【 0 0 5 3 】

本明細書で使用する「プログラム」または「コンピュータープログラム」とは、一般に特定のプログラミング言語の規則に従わせる統辞論単位を指し、そしてそれは特定の機能、作業または問題を解決または実行するために必要な「コードセグメント」に分けることができる宣言 (d e c l a r a t i o n) および文 (s t a t e m e n t) またはインストラクション (i n s t r u c t i o n) からなる。プログラミング言語は一般にプログラムを記述する人工言語である。

【 0 0 5 4 】

本明細書で使用する「システム」または「コンピューターシステム」は、一般に1もしくは複数のコンピューター、端末装置およびデータ処理を行うソフトウェアを指す。「ユーザー」または「システムオペレーター」は一般に、データ処理および情報交換の目的で「ユーザーデバイス」(例えばコンピューター、無線デバイス等)を介してコンピューターネットワークを使用する人を含む。「コンピューター」は一般に、人の介入無しに膨大な数の計算操作および論理操作を含む実質的なコンピューター処理を行うことができる基本単位である。

10

【 0 0 5 5 】

本明細書で使用する「アプリケーションソフトウェア」または「アプリケーションプログラム」は、一般にアプリケーションのトラブルの解決に特異的なソフトウェアまたはプログラムを指す。「アプリケーションの問題」は、一般にエンドユーザーにより提起され、そしてその解決のための情報処理を必要とする問題である。

20

【 0 0 5 6 】

本明細書で使用する「自然言語」は、規則が特殊に記載されない現行使用に基づく、例えば英語、スペイン語または中国語を指す。本明細書で使用する「人工言語」は、規則がその使用前に実験的に確立された言語、例えばC、C++、Java、BASIC、FORTRANまたはCOBOLのようなコンピュータープログラム言語を指す。

【 0 0 5 7 】

本明細書で使用する「統計的関連性」は、1もしくは複数のランキングスキームを使用することを指し(O/E比、強度等)、ここで関連性はそれが無作為の機会により期待されるよりも有意に頻繁に起こる場合に、統計的に関連性があると定められる。

【 0 0 5 8 】

本明細書で使用する「調和して (c o o r d i n a t e l y) 調節される遺伝子」または「転写モジュール」という用語は互換的に使用され、特異的遺伝子のグループ分けされた遺伝子発現プロファイル(例えば特異的遺伝子配列と関連したシグナル値)を指す。各転写モジュールは2つの重要なデータ部分 (p i e c e) 、文献調査部分、および遺伝子マイクロアレイから得た実際の経験的な遺伝子発現値データと相関する。転写モジュールに選択された遺伝子の組は、遺伝子発現データの分析に基づく(上記のモジュール抽出アルゴリズム)。さらなる工程はChaussabel, D. & Sher, 文献のプロファイリングによるマイクロアレイ発現データのマイニング (A . M i n i n g m i c r o a r r a y e x p r e s s i o n d a t a b y l i t e r a t u r e p r o f i l i n g) . G e n o m e B i o l 3 , R E S E A R C H 0 0 5 5 (2 0 0 2) , (h t t p : / / g e n o m e b i o l o g y . c o m / 2 0 0 2 / 3 / 1 0 / r e s e a r c h / 0 0 5 5) により教示され、関連の部分は引用により本明細書に編入され、そして発現データは目的の疾患または状態、例えば全身性エリテマトーデス、関節炎、リンパ腫、悪性腫瘍、黒色腫、急性感染、自己免疫障害、自己炎症障害等から得られる。

30

40

【 0 0 5 9 】

以下の表は文献調査部分、または転写モジュールへの貢献を開発するために使用されたキーワードの例を列挙する。当業者は他の用語を他の状態、例えば特異的な癌、特異的な感染性疾患、移植等に選択できると認識する。例えばT細胞活性化に関連する遺伝子およびこれらの遺伝子に関するシグナルを以下にモジュールID " M 2 . 8 " と記載し、ここで特定のキーワード(例えばリンパ腫、T細胞、CD4、CD8、TCR、胸腺、リンパ

50

系、IL2)は、重要なT細胞関連遺伝子、例えばT細胞表面マーカー(CD5、CD6、CD7、CD26、CD28、CD96)；リンパ系細胞により発現される分子(リンホトキシンベータ、IL-2誘導性T細胞キナーゼ、TCF7；およびT細胞分化タンパク質mal、GATA3、STAT5B)を同定するために使用された。次に完全なモジュールは、患者群からのデータをこれらの遺伝子と関連させることにより開発されて(平板、存在/不存在および/またはアップもしくはダウンレギュレーションにかかわらず)、転写モジュールが作成される。場合により、遺伝子プロファイルはこれらの疾患状態およびデータについていかなる特定の遺伝子クラスターとも合わないことがある(この時点で)が、特定の生理学的経路(例えばcAMPシグナル伝達、亜鉛フィンガータンパク質、細胞表面マーカー等)が「未決定」モジュール内に見いだされる。実際に、遺伝子発現データ組は、キーワードの調査と合わせる前に発現が調整された遺伝子を抽出するために使用でき、すなわちいずれかのデータ組を第2のデータ組と相互参照する前に関連させることができる。

【0060】

【表 1 3】

表1. 転写モジュールの例

モジュール I. D. の例	キーワード 選択の例	遺伝子プロファイル評価	
M 1.1	Ig、 免疫グロブリン、 骨、骨髄、PreB、 IgM, Mu.	形質細胞。免疫グロブリン鎖(例えばIGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含む。	
M 1.2	血小板、接着、 凝集、内皮、 脈管	血小板。血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1 A/B)、およびPPPB(プロ血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む。	10
M 1.3	イムレセプター、 BCR、B細胞、 IgG	B細胞。B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子: 初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ (BLK)をコードする遺伝子を含む。	
M 1.4	複製、抑制、 修復、CREB、 リンパ系、 TNF-アルファ	未決定:この組はcAMPシグナル伝達経路のレギュレーターおよび標的(JUND, ATF4, CREM, PDE4, NR4A2, VIL2)、ならびにTNF-アルファ媒介型NF-KB活性化のリプレッサー(CYLD, ASK, TNFAIP3)を含む。	
M 1.5	単球、 樹状突起の、 MHC、同時刺激、 TLR4、MYD88	ミエロイド系:ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある: CD14, TLR2, MYD88)を含む。この組はTNFファミリーメンバー(TNFR2, BAFF)も含む。	20
M 1.6	亜鉛、 フィンガー、 P53、RAS	未決定:この組はシグナル伝達分子、例えば活性化STAT (PIAS1およびPIAS2)の亜鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子を含む。	
M 1.7	リボゾーム、 翻訳、40S、 60S、HLA	MHC/リボゾームタンパク質。MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボゾームタンパク質(RPLs, RPSs)をコードする遺伝子によりほとんど排他的に形成される。	
M 1.8	代謝、生合成、 複製、ヘリカーゼ	未決定:代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子を含む。	30
M 2.1	NK、キラー、 細胞溶解、CD8、 細胞媒介型、 T細胞、CTL、 IFN-g	細胞傷害性細胞。細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュリシン)、ケモカイン(GCL5, XCL1)およびCTL/NK-細胞関連分子(GTSW)を含む。	
M 2.2	顆粒球、 好中球、防御、 ミエロイド、骨髄	好中球。この組は好中球顆粒に見いだされる内在性分子(ラクトランスフェリン: LTF、デフェンシン: DEAF1、細菌透過性増加タンパク質: BPI、カセリシジン抗微生物タンパク質: CAMP…)を含む。	
M 2.3	赤血球、赤、 貧血、グロビン、 ヘモグロビン	赤血球。ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン: ANK1、グリコホリン C: GYPC、ヒドロキシメチルピラン シンターゼ: HMBS、赤血球関連因子: ERAF)を含む。	40

【 0 0 6 1】

【表 1 4】

モジュール I. D. の例	キーワード 選択の例	遺伝子プロファイル評価	
M 2.4	リボヌクレオタンパク質、60S、核小体、アッセンブリー、延長	リボゾームタンパク質。リボゾームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー (EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子を含む。	
M 2.5	腺癌、間質の、間葉、樹状突起、運動	未決定。このモジュールは免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター、シンデカン 2、プレキシシC1, ジストロプレビン)をコードする遺伝子を含む。	10
M 2.6	顆粒球、単球、ミエロイド、ERK、壊死	ミエロイド系。M1.5に関連する。単球および好中球のようなミエロイド系細胞で発現する遺伝子 (IGTB2/CD18、リンホトキシンベータ受容体、ミエロイド関連タンパク質 8/14 ホルミルペプチド受容体 1)を含む；	
M 2.7	キーワードは抽出されない	未決定。このモジュールは、未知の機能の転写産物からほぼ構成される。ケモカイン様因子スーパーファミリー (CKLFSF8)の1メンバーを含め、文献に関連するわずか20個の遺伝子。	
M 2.8	リンパ腫、T細胞、CD4、CD8、TCR、胸腺、リンパ系、IL2	T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ,TCF7, T細胞分化タンパク質 mal, GATA3, STAT5B)を含む。	20
M 2.9	ERK、トランス活性化、細胞骨格、MAPK, JNK	未決定。細胞骨格に関連する分子 (アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子を含む。また存在するのは、T細胞が発現する遺伝子(FAS, ITGA4/CD49D, ZNF1A1)。	
M 2.10	ミエロイド、マクロファージ、樹状突起、炎症、インターロイキン	未決定。免疫関連細胞表面分子(CD36, CD86, LILRB)、サイトカイン(IL15)およびシグナル伝達経路に関与する分子 (FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子を含む。	
M 2.11	複製、抑制、RAS、自己リン酸化作用、発癌遺伝子	未決定。キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1,TAOK1, CALM2, PRKCI, ITPKB, SRPK 2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2) およびRAS ファミリーメンバー (G3BP, RAB14, RAS A2, RAP2A, KRAS)を含む。	30
M 3.1	ISRE、インフルエンザ、抗ウイルス、IFN-ガンマ、IFN-アルファ、インターフェロン	インターフェロン誘導性。この組はインターフェロン誘導性遺伝子:抗ウイルス分子 (OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML)、ケモカイン (CXCL10/IP-10)、シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)を含む。	
M 3.2	TGF-ベータ、TNF、炎症、アポトーシスの、リポ多糖	炎症I。炎症プロセスに関与する分子(IL8, ICAM1, G5R1, CD44, PLAUR, IL1A, CXCL16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B)をコードする遺伝子を含む。	40

【 0 0 6 2 】

【表 15】

モジュール I. D. の例	キーワード 選択の例	遺伝子プロファイル評価
M 3.3	顆粒球、炎症、 防御、酸化する、 リソソームの	炎症II。顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP)、ならびにリソソーム酵素を誘導する、またはそれにより誘導性の(PPT1, CTSS/S, CES1, NEU1, ASAH1, LAMP2, CAST)分子を含む。
M 3.4	キーワードは 抽出されない	未決定。タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ(PI3K)ファミリーメンバー(PIK3CA, PIK32A, PIP5K3)を含む。
M 3.5	キーワードは 抽出されない	未決定。わずかな数の転写産物からなる。ヘモグロビン遺伝子(HBA1, HBA2, HBB)を含む。
M 3.6	補体、宿主、 酸化的、 細胞骨格、 T細胞	未決定。この大変大きい組はT細胞表面マーカー(CD101, CD102, CD103)、ならびに血中白血球の回りで偏在的に発現される分子(CXCR1: フラクタルキン受容体, CD47, P-セレクトリンリガンド)を含む。
M 3.7	スプライソーム、 メチル化、 ユビキチン、 ベーターカテニン	未決定。プロテアソームサブユニット(PSMA2/5, PSMB5/8); ユビキチンタンパク質リガーゼ HIP2, STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子を含む。
M 3.8	CDR、TCR、 CREB、 グリコシラーゼ	未決定。幾つかの酵素:アミノメチルトランスフェラーゼ、アルギニルトランスフェラーゼ、アスパラギンシンターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼをコードする遺伝子を含む
M 3.9	クロマチン、 チェックポイント、 複製、 トランス活性化	未決定。プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ(例えばPTPLB, PPP1R8/2CB)をコードする遺伝子を含む。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)も含む。

10

20

【0063】

生物学的定義

本明細書で使用する「アレイ」という用語は、支持体に1もしくは複数のペプチドまたは核酸プローブが結合した固体支持体または基体を指す。アレイは、異なる既知の位置の基体表面にカップリングされた典型的には1もしくは複数の異なる核酸またはペプチドプローブを有する。これらのアレイは、既知のゲノム、例えばヒトゲノムに基づき10,000; 20,000、30,000または40,000種の異なる同定可能な遺伝子を有することができる「マイクロアレイ」または「遺伝子チップ」とも記載される。これらのパン-アレイ(pan-array)は、サンプル中で発現、または見いだされる全「トランスクリプトーム」または遺伝子の転写プール、例えばDNAレプリコンの相補的組を作成するためにRTおよび/またはRT-PCRにかけることができるRNA、mRNA等として発現される核酸を検出するために使用される。アレイは、非リトグラフおよび/またはフォトリトグラフ法および固相合成法の組み合わせを包含する機械的合成法、光直接合成法等を使用して製造することができる。

30

40

【0064】

これら核酸アレイの合成のための様々な技術がこれまで記載されており、例えば事実上任意の形状の表面上またはさらに多表面上でも作成された。アレイはビーズ、ゲル、ポリマー表面、光ファイバーのようなファイバー、ガラスまたは任意の他の適切な基体上のペプチドまたは核酸であることができる。アレイは診断またはすべてを含むデバイスの他の操作を可能とする様式で包装されることができ、例えば本明細書に編入する米国特許第6,955,788号明細書、関連部分を参照にされたい。

【0065】

本明細書で使用する用語「疾患」とは、細胞に任意の異常な生物学的状態がある生物の

50

生理学的状態を指す。疾患には限定するわけではないが、遺伝的な、遺伝した、感染により生じた、異常な細胞機能、異常な細胞分裂等により生じた細胞、組織、身体機能、系または臓器の妨害、休止または障害を含む。「疾患状態」を導く疾患は、一般に生物系、すなわち疾患の宿主に対して有害である。本発明に関して、疾患または障害と関連する感染（例えばウイルス、バクテリア、真菌、寄生虫等）、炎症、自己炎症、自己免疫、アナフィラキシー、アレルギー、前悪性、悪性、外科的、移植、生理学的等のような任意の生物学的状態は、疾患状態と考えられる。病的状態は一般に疾患状態と等価である。

【0066】

また疾患状態は、異なるレベルの疾患状態に分類することもできる。本明細書で使用するように、疾患または疾患状態のレベルは、疾患または疾患状態の進行、ならびに処置時、最中および後の生理学的応答を反映する任意の尺度である。一般に疾患または疾患状態はレベルまたは段階を通して進行し、ここで疾患の影響は次第に重篤になる。疾患状態のレベルはサンプル中の細胞の生理学的状態により影響を受け得る。

10

【0067】

本明細書で使用する用語「治療」または「治療計画」は、疾患状態を緩和または改変するために取る医学的段階、例えば薬理的、外科的、食事的および/または他の技術を使用して疾患の影響または症状を下げるか、または排除することを意図した過程を指す。治療計画には1もしくは複数の薬剤または外科治療の処方された投与を含むことができる。治療の多くは有益であり、そして疾患状態を下げるが、多くの場合、治療の効果は望ましくない作用、すなわち副作用を有する。また治療の効果は宿主の生理学的状態、例えば年齢、性別、遺伝子、体重、他の疾患状態等により影響され得る。

20

【0068】

本明細書で使用する「薬理学的状態 (state)」または「薬理学的状況 (status)」という用語は、サンプル中の1もしくは複数の核酸の薬理学的状態に影響を及ぼすことができる1もしくは複数の薬剤、外科医術等で処置される、および/またはされたサンプル、例えば薬理的介入により新たに転写、安定化かつ/または脱安定化されたサンプルを指す。サンプルの薬理学的状態は、薬剤処置の前、最中および/または後の生物学的状態の変化に関連し、そして本明細書に教示する診断的または予知的機能に役立つことができる。薬剤処置または外科的治療後の幾つかの変化は、疾患状態と関連するかもしれないし、かつ/または無関係な副作用であるかもしれない。薬理学的状態の変化は、治療期間、処方した薬剤の種類および用量、与える治療過程のコンプライアンスの程度、および/または摂取した処方以外の薬剤の結果であろう。

30

【0069】

本明細書で使用する「生物学的状態」という用語は、発現の変化を分析するために単離および精製した細胞サンプルのトランスクリプトーム（すなわちRNA転写産物の全集合）の状態を指す。生物学的状態は、細胞成分の量および/または活性の測定、形態学的表現型の特性決定、または転写産物の検出法の組み合わせによりサンプル中の細胞の生理学的状態を反映する。

【0070】

本明細書で使用する「発現プロファイル」という用語は、RNA、DNAまたはタンパク質の量 (abundance) または活性レベルの相対的量を指す。発現プロファイルは、例えば任意の数の方法による、および任意の数の遺伝子チップ、遺伝子アレイ、ビーズ、マルチプレックスPCR、定量的PCR、ラン-オンアッセイ、ノーザンブロット分析、ウエスタンブロット分析、タンパク質発現、蛍光活性化細胞ソーティング (FACS)、酵素結合免疫吸着アッセイ (ELISA)、化学発光実験、酵素アッセイ、増殖実験または市販されているものから容易に入手可能な遺伝子発現の測定および/または分析用の他の方法、装置および系を使用した、転写状態または翻訳状態の測定値であることができる。

40

【0071】

本明細書で使用するサンプルの「転写状態」という用語には、サンプル中に存在する R

50

RNA種、特にmRNAの同一性および相対的量を含む。サンプルの全転写状態、すなわちRNAの同一性および存在量ノ組み合わせを本明細書ではトランスクリプトームと呼ぶ。一般にサンプル中のRNA種の全組の全相対的成分の実質的画分が測定される。

【0072】

本明細書で使用する「モジュール転写ベクトル」という用語は、「差次的に発現される遺伝子の比率」を反映する転写発現データを指す。例えば各モジュールについて、少なくとも2つの群（例えば健康な個体对患者）間で差次的に発現される転写産物の比率。このベクトルは2つの群のサンプルの比較により引き出される。第1の分析工程は、各モジュール内で疾患に特異的な転写産物の組の選択に使用される。次いで「発現レベル」がある。所定の疾患に関する群の比較により、各モジュールについて差次的に発現される転写産物のリストが提供される。異なる疾患が異なるサブセットのモジュール転写産物を生じることが見いだされた。この発現レベルを用いて、次に差次的に発現されると同定された遺伝子の疾患特異的サブセットの発現値を平均することにより、1つのサンプルについて各モジュール（1もしくは複数）に関するベクトルを算出することが可能である。この取り組みにより1つのサンプルについて、モジュール発現ベクトルのマップ、例えば本明細書に開示するモジュールマップを作成することが可能となる。これらのベクトルモジュールマップは、各サンプルについて誘導され得る各モジュールの平均化発現レベルを表す（差次的に発現される遺伝子の比に代わり）。

10

【0073】

本発明を使用して、モジュールレベルだけでなく、遺伝子レベルでも疾患を同定し、そして識別することが可能であり；すなわち2つの疾患は同じベクトルを有することができるが（同一比の差次的に発現される転写産物、同一「極性」）、それでもベクトルの遺伝子組成は疾患特異的となり得る。遺伝子レベルの発現は、分析の解像度を大きく上げる明白な利点を提供する。

20

【0074】

さらに本発明は複合転写マーカーを利用する。本明細書で使用する「複合転写マーカー」という用語は、マーカーとして個別の遺伝子を使用することに比べて、複数の遺伝子（モジュールのサブセット）の平均発現値を指す（そしてこれらマーカーの組成は、疾患特異的であることができる）。複合転写マーカー法は、使用者が例えばSLE患者の疾患重篤度を評価するために、または本明細書に開示する発現ベクトルを得るために多変量マイクロアレイスコアを開発できるので独自である。最も重要であるのは、本発明の複合モジュール転写マーカーを使用して、本明細書で見いだされた結果がマイクロアレイ平板全体で再現性があり、これにより調節承認の大きな信頼性を提供することが見いだされた点である。

30

【0075】

本発明と一緒に使用するための遺伝子発現モニタリングシステムには、1もしくは複数の標的疾患に特異的および/またはカスタマイズされた限定数および/または基本数の遺伝子を有するカスタマイズされた遺伝子アレイを含むことができる。通常使用されている一般的なパン-ゲノムアレイとは異なり、本発明は特異的平板を使用することを必要としない過去の遺伝子およびゲノム分析用の一般的パン-アレイの使用を提供するだけでなく、より重要なことは、数千もの他の無関係遺伝子を必要としない分析に最適な遺伝子組を提供するカスタマイズされたアレイの開発を提供する。既存の技術に優る本発明の最適化されたアレイおよびモジュールの1つの明らかな利点は、経費の削減（例えばアッセイ毎の、材料、装置、時間、人的、訓練等の経費）であり、そしてさらに重要なことは、ほとんどのデータが無関係であるパン-アレイを製造するための環境的経費の削減である。本発明のモジュールは、単純な設計、シグナル対ノイズ比を最大にすると同時に最少数のプローブで最適なデータを提供するカスタムアレイを初めて可能にする。分析のための遺伝子総数を削減することにより、例えばほとんど無関係な膨大な量のデータを提供するパン-遺伝子チップの製造中に、フォトリソグラフィ用数千もの高価な白金マスクを製造する必要性を削減することが可能である。本発明を使用して、本発明の限定されたプロ

40

50

ブ組（1もしくは複数）はデジタル光化学アレイ、ボールビーズアレイ、ビーズ（例えば LumineX）、マルチプレックスPCR、定量的PCR、ラン-オンアッセイ、ノーザンブロット分析、あるいはさらにタンパク質の分析は例えばウエスタンブロット分析、2Dおよび3Dゲルタンパク質発現、MALDI、MALDI-TOF、蛍光活性化細胞ソーティング（FACS）（細胞表面または細胞内）、酵素連結免疫吸着アッセイ（ELISA）、化学発光実験、酵素アッセイ、増殖実験または市販されているものから容易に入手可能な遺伝子発現の測定および/または分析用の任意の他の方法、装置および系と一緒に使用されれば、マイクロアレイの必要性を完全に回避することが可能である。

【0076】

本発明の「分子のフィンガープリンティング系」は、異なる細胞または組織、同じ細胞または組織の異なる部分集合、同じ細胞または組織の異なる生理学的状態、同じ細胞または組織の異なる発達段階、あるいは同じ組織の異なる細胞群における発現を、他の疾患および/または正常な細胞対照に対して比較分析することを容易にし、そして行うために使用することができる。幾つかの場合では、正常または野生型発現データは、同時もしくはほぼ同時に分析したサンプルに由来してよく、あるいはそれは既存の遺伝子アレイ発現データベース、例えばNCBI遺伝子発現多目的データベースのような公的なデータベースから得た、または集めた発現データでよい。

【0077】

本明細書で使用するように、用語「差次的に発現する」とは、2以上のサンプル、例えば疾患のサンプルと正常サンプルとの間で変動する細胞の構成成分（例えば核酸、タンパク質、酵素活性等）の測定値を指す。細胞の構成成分はオンもしくはオフ（存在もしくは不存在）、参照に対してアップレギュレートされるか、または参照に対してダウンレギュレートされることができる。遺伝子チップまたは遺伝子アレイを用いた使用には、核酸、例えばmRNAまたは他のRNA（miRNA、siRNA、hnRNA、rRNA、tRNA等）の差次的遺伝子発現を使用して、細胞型または核酸間を識別することができる。最も普通には、細胞の転写状態の測定は、定量的逆転写酵素（RT）および/または定量的逆転写ポリメラーゼ連鎖反応（RT-PCR）、ゲノム発現分析、翻訳後分析、ゲノムDNAに対する修飾、トランスロケーション、in situハイブリダイゼーション等によりなされる。

【0078】

幾つかの疾患状態には、特に疾患状態の初期レベルで細胞または形態学的差異を同定することが可能である。本発明は、細胞そのものの自体の遺伝子のモジュールを調べることにより、さらに重要なことには細胞の通常の生理学的内容、すなわち免疫賦活化、免疫寛容またはさらに免疫エネルギーで作用している免疫エフェクター細胞に由来する遺伝子の細胞性RNAの発現を調べることにより、それら特異的変異または1もしくは複数の遺伝子を同定する必要性を回避する。遺伝的変異は遺伝子群の発現レベルに劇的な変化を生じることができるが、生物系はしばしば他の遺伝子の発現を改変することにより変化を補償する。これらの内部補償応答の結果として、多くの混乱（perturbation）は系において観察可能な表現型の効果が最少かもしれないが、細胞の構成成分の組成には顕著な効果を有する可能性がある。同様に、遺伝子転写産物の実際のコピーは増加も減少もしないかもしれないが、転写産物の寿命もしくは半減期が影響を受けて、タンパク質生産を大きく増加させる可能性がある。本発明は1つの態様では、1つのメッセージおよび/または変異よりもむしろエフェクター細胞（例えば白血球、リンパ球および/またはそれらの亜群）に注目することにより、実際のメッセージを検出する必要性を排除する。

【0079】

当業者は、サンプルが例えば単一細胞、細胞集団、組織、細胞培養物等を含む様々な供給源から得ることができると直ちに考えるだろう。特定の場合、例えば尿、血液、唾液、組織または生検サンプル等に見いだされる細胞から十分なRNAを単離することさえ可能となり得る。特定の状況では、十分な細胞および/またはRNAを：粘膜分泌物、糞、涙、血漿、腹腔液、間質液、硬膜液、脳脊髄液、汗または他の体液から得ることができる。

10

20

30

40

50

核酸供給源、例えば組織もしくは細胞の供給源には、組織生検サンプル、1もしくは複数の分類した細胞群、細胞培養物、細胞クローン、形質転換細胞、生検材料または単一細胞を含む。組織供給源には、例えば脳、肝臓、心臓、腎臓、肺、脾臓、網膜、骨、神経系、リンパ節、内分泌腺、生殖器、血液、神経、脈管組織および嗅覚上皮を含むことができる。

【0080】

本発明は、単独または組み合わせて使用できる以下の基本的構成要素、すなわち1もしくは複数のデータマイニングアルゴリズム：1もしくは複数のモジュール-レベルの分析プロセス；血中白血球の転写モジュールの特性決定；ヒトの疾患の分子的診断/予知のための多変数分析における総 (aggregated) モジュールデータの使用；および/またはモジュール-レベルのデータおよび結果の視覚化を含む。本発明を使用して、さらに1つの多変数スコアに集めることができる複合転写マーカーを開発および分析することも可能である。

10

【0081】

データ獲得速度の爆発的急増も、マイクロアレイデータおよび生物医学的知識を活用するためのマイニングツールおよびアルゴリズムの開発に拍車をかけた。転写系のモジュラー組織および機能を暴露する目的の取り組みは、疾患の強固な分子的兆候の同定に有望な方法を構成する(14-16, 17)。実際にそのような分析は、大規模な転写実験の概念を、個々の遺伝子のレベルまたは遺伝子のリストを越えたマイクロアレイデータの概念化を取るにより転換することができる。

20

【0082】

本発明者は、現在のマイクロアレイに基づく調査が、評判の悪い「ノイズが高い」データ分析、すなわちデータは解釈が難しく、そして研究室および平板全体で十分に比較できないという重大な難題に直面していることを認識した。マイクロアレイデータの分析に広く受け入れられている取り組みは、実験群の間で差次的に発現する遺伝子のサブセットの同定から始まる。次いで使用者は引き続き、パターン発見アルゴリズムおよび既存の科学的知識を使用して生じた遺伝子リストに「意味を見いだす」。

【0083】

平板全体で膨大な変数を扱うというよりは、むしろ本発明者は分析の初期段階で生物学的に関連する遺伝子の選択を強調する方法を開発した。簡単に説明すると、この方法は所定の生物学的系を特徴付ける転写成分の同定を含み、このために改善されたデータマイニングアルゴリズムが開発され、大きなデータの集合から調和して発現される遺伝子または転写モジュールの群を分析、そして抽出した。

30

【0084】

1例では、4742個のプロープ組にさらにグループ分けされる28個の転写モジュールが239個の血中白血球転写プロファイルから得られた。これらのモジュールを形成する遺伝子間の機能的収束(functional convergence)は、文献のプロファイリングを介して証明された。第2段階は、モジュール基準つ転写系の混乱(perurbation)を実験することからなる。この概念を具体的に説明するために、健康な有志および患者から得た白血球転写プロファイルを得、比較し、そして分析した。さらにこの遺伝子フィンガープリンティング法の検証は、公開されているマイクロアレイのデータセットの分析を介して得た。注目すべきは、既存のデータを使用した本発明のモジュラー転写装置、システムおよび方法が、2つの市販されているマイクロアレイ平板全体で高度な再現性を示した点である。

40

【0085】

本発明は転写系のモジュラー分析用に設計された広い応用性の2段階のマイクロアレイデータマイニング法の実行を含む。この新規取り組みは、血中白血球の転写的兆候を特徴付けるために使用され、これは臨床的に関連性のある情報の最もアクセス可能な供給源を構成する。

【0086】

50

本明細書で証明するように、たとえベクトルが2つの疾患について同一(+ / +) (例えばS L EおよびF L Uの両方についてM 1 . 3 = 5 3 %低下)でも、各ベクトルの組成はそれらを区別するために使用できるので、2つのベクトルに基づき2つの疾患の間を測定し、示差し、かつ/またはそれらを識別することが可能である。例えば差次的に発現される転写産物の比率および極性がM 1 . 3の2つの疾患の間で同一であっても、遺伝子組成は疾患特異的となり得る。遺伝子レベルおよびモジュールレベルの分析を組み合わせることにより、解像度がかなり上昇する。さらに2、3、4、5、10、15、20、25、28以上のモジュールを使用して、疾患を区別することが可能である。

【0087】

材料および方法。血液の処理。すべての血液サンプルはクエン酸デキストロース管(B Dバキュテイナー: Vacutainer)に集め、そして処理するために直ちに室温でテキサス州、ダラスのバイロール免疫調査研究所(Baylor Institute for Immunology Research)に運んだ。3~4mlの血液から末梢血単核細胞(PBMC)をFicoll勾配を介して単離し、そして直ちにベータ-メルカプトエタノールを含むRLT試薬中(キアゲン: Qiagen、バレンシア、カリフォルニア州)で溶解し、そしてRNA抽出工程前に-80で保存した。

10

【0088】

マイクロアレイ分析。全RNAはRNeasyキット(キアゲン)を使用して、製造元の使用説明に従い単離し、そしてRNAの完全性をアギレント(Agilent)2100バイオアナライザー(Bioanalyzer)(アギレント、パロアルト、カリフォルニア州)を使用して評価した。

20

【0089】

アフィメトリックス(Affymetrix)GeneChips:これらのマイクロアレイは、石英ウェーハ上のその場で成された短いオリゴヌクレオチドプローブ組を含む。標的の標識化は製造元の標準プロトコルに従って行った(アフィメトリックス社、サンタクララ、カリフォルニア州)。ピオチン化cRNA標的を精製し、そして引き続きアフィメトリックスのHG-U133AおよびU133B GeneChips(>44,000プローブ組)にハイブリダイズさせた。アレイはアフィメトリックスの共焦点レーザーキャナーを使用して走査した。Microarray Suite、バージョン5.0(MAS5.0;アフィメトリックス)ソフトウェアを使用して蛍光ハイブリダイゼーションシグナルを評価し、シグナルを標準化し、そしてシグナル検出コールを評価した。チップあたりのシグナル値の標準化は、GeneChipあたり500の標的強度値に対するMAS5.0のグローバルなスケールリング法を使用して達成された。遺伝子発現分析ソフトウェアプログラムであるGeneSpring、バージョン7.1(アギレント)を使用して、統計分析および階層的クラスタリングを行った。

30

【0090】

イルミナ(Illumina)BeadChip:これらのマイクロアレイは、3μmビーズに結合した50merのオリゴヌクレオチドプローブを含み、これらはマイクロウェルにガラススライドの表面上で付けられた。サンプルを処理し、そしてサービス契約に基づきイルミナ社(サンディエゴ、カリフォルニア州)により得た。標的はイルミナのRNA増幅キット(アンピオン: Ambion、オースチン、テキサス州)を使用して調製した。cRNA標的をSentrrix HumanRef8BeadChip(>25,000プローブ)にハイブリダイズし、これをイルミナのBeadStation500で走査した。イルミナのBeadstudioソフトウェアを使用して蛍光ハイブリダイゼーションシグナルを評価した。

40

【0091】

文献プロファイリング。この実験で使用した文献プロファイリングアルゴリズムは、以前に詳細に記載された(18)。この取り組みは類似のキーワードを共有する遺伝子を連結する。これは階層的クラスタリング、一般的なアンスーパーバイズドパターンディスカバーアルゴリズム(popular unsupervised pattern d

50

iscovery algorithm) を使用して、文献の要約中に存在する用語のパターンを分析する。工程 1：遺伝子：各遺伝子に関して直接関係する公報を同定する文献索引を作成する。工程 2：用語が存在する頻度をテキストプロセッサにより計算する。工程 3：厳密なフィルター基準を使用して関連するキーワードを選択する（すなわち、すべての遺伝子をわたり高いかまたは低い頻度の用語を排除し、そしてわずかな遺伝子についてのみ高い存在のパターンに特徴付けられるいくつかの違いが分かる用語を残す）。工程 4：2 方向の階層クラスタリングが、存在パターンに基づき遺伝子および関連するキーワードをグループ分けし、遺伝子のグループ間に存在する機能的関連性の視覚的表示を提供する。

【0092】

モジュラーデータマイニングアルゴリズム。最初に、単一遺伝子のレベルを越えて生物系の特徴付けを可能にする 1 もしくは複数の転写成分が同定される。協調して調節される遺伝子の組、または転写モジュールが新規マイニングアルゴリズムを使用して抽出され、これは血中白血球マイクロアレイプロファイルの大きな組に適用された（図 1）。全部で 239 種の末梢血単核細胞（PBMC）サンプルからの遺伝子発現プロファイルを、Affymetrix U133A & B GeneChip (> 44,000 プロブ組) を使用して作成した。転写データを 8 つの実験群（全身型若年性特発性関節炎、全身性エリテマトーデス、I 型糖尿病、肝移植受容者、黒色腫患者および急性感染の患者：大腸菌、黄色ブドウ球菌およびインフルエンザ A）について得た。各群に関して、すべての状態をわたりフラッグコールが無い（absent flag call）転写産物にフィルターをかけた。残る遺伝子は階層的クラスタリングにより 30 に分布した（クラスター C1 から C30）。各遺伝子に関するクラスターの割り当ては表に記録され、そして分布パターンはすべての遺伝子間で比較された。モジュールは、すべての実験群中、同じクラスターに属する大きな遺伝子組（すなわち 8 つの実験群中の 8 つで同じクラスターに見いだされた遺伝子）から始めて反復プロセスを使用して選択された。次いでこの選択をこの中心参照パターンから、7/8、6/8 そして 5/8 対の遺伝子を含むように拡大した。生じた遺伝子の組は転写モジュールを形成し、そして選択プールから引き抜かれた。次いでこのプロセスを 2 番目に大きい遺伝子群から始めて繰り返し、進めるにしたがって厳密度のレベルを下げた。この分析により 28 のモジュールに分布した 5348 の転写産物の同定が導かれた（完全なリストは追補資料に提供する）。各モジュールは、選択回数および順序を示す独自の識別名に割り当てられる（すなわち M3.1 は第 3 回目の選択で最初に同定される）。

【0093】

モジュールは明確な「転写挙動」を提示する。同時発現遺伝子が機能的に連結していることは広く想定されている。この「関連によるギルト（guilt by association）」という概念は、遺伝子が多くのサンプルにわたり複雑な発現パターンに従う場合に特に強要される。本発明者は、転写モジュールが整合的な生物単位を形成することを見だし、したがって我々の最初のデータセットで同定される同時発現特性は、独立したサンプルの組で保存されると予想した。データは、21 名の健康な有志の血液から単離された PBMC について得た。これらのサンプルは上記のモジュール選択プロセスでは使用されなかった。

【0094】

図 2 は、4 種のモジュールに関する遺伝子発現プロファイルを示す（図 2：M1.2、M1.7、M2.11 および M2.1）。図 2 のグラフでは、各線が多数のサンプル全体で 1 つの遺伝子の発現レベル（y - 軸）を表す（x - 軸は 21 サンプル）。この例における遺伝子発現の差異は、「健康」な個体間の個体間変動を表す。各モジュール内で遺伝子は整合的な「転写挙動」を表すことが分かった。実際に遺伝子発現の変動は、すべてのサンプルにわたり一貫していると思われた（同じサンプルに関して、すべての遺伝子の発現が上昇し、そしてピークを形成する一方、下降（dip）を形成するすべての遺伝子について他はレベルが低かった）。重要なことは、ピークおよび下降が M1.2、M2.11

10

20

30

40

50

およびM2.1の異なるサンプルについて形成するので、個体間変動はモジュール特異的となるようである。さらに変動の幅も各モジュールで特徴的であり、発現レベルはM2.1、そして特にM1.7よりもM1.2およびM2.11で変動性であった。すなわち我々は転写モジュールが独立的な生物学変数を構成することを見いだす。

【0095】

転写モジュールの機能的特性決定。次いでモジュールを機能的レベルで特性決定した。テキストマイニング法を使用して、各遺伝子について集めた生物医学文献からキーワードを抽出した(18に記載)。分析した4つのモジュールに関連するキーワードの分布は、明らかに異なる(図3)。以下は、特定のモジュールに関連し得るキーワードのリストである。

10

【0096】

M1.2に高度に特異的なキーワードは血小板、凝固または血栓症を含み、そしてITGA2B(インテグリンアルファ2b、血小板糖タンパク質IIb)、PF4(血小板因子4)、SELP(セレクチンP)およびGP6(血小板糖タンパク質6)のような遺伝子に関連した。

【0097】

M1.3に高度に特異的なキーワードはB細胞、免疫グロブリンまたはIgGを含み、そしてCD19、CD22、CD72A、BLNK(B細胞リンカータンパク質)、BLK(Bリンパ系チロシンキナーゼ)、およびPAX5(ペアドボックス遺伝子5、B細胞系特異的アクチベーターのような遺伝子に関連した。

20

【0098】

M1.5に高度に特異的なキーワードは、単球、樹状細胞、CD14またはToll様を含み、そしてMYD88(ミエロイド分化1次応答遺伝子88)、CD86、TLR2(Toll様受容体2)、LILRB2(白血球免疫グロブリン様受容体B2)およびCD163のような遺伝子に関連した。

【0099】

M3.1に高度に特異的なキーワードは、インターフェロン、IFN-アルファ、抗ウイルスまたはISREを含み、そしてSTAT1(シグナル伝達および転写1の活性化)、CXCL10(CXCケモカインリガンド10、IP-10)、OAS2(オリゴアデニレートシンターゼ2)およびMX2(ミクソウイルス耐性2)のような遺伝子に関連した。

30

【0100】

この用語の存在の対比パターンは、各モジュールの驚くべき機能的整合性を表す。同定されたすべてのモジュールに関して文献から抽出した情報は、転写レベルでのPBM系C系の包括的な機能的特性決定を可能とする。28種のサンプルの各PBM系C転写モジュールについて、同定された機能的関連の記載を表2に提供する。

【0101】

【表 1 6】

表2:28の転写モジュールの完全な機能的評価

モジュール I. D.	プローブの 組数	キーワードの 選択	評価
M 1.1	69	Ig、 免疫グロブリン、 骨、骨髄、PreB、 IgM, Mu.	形質細胞。 免疫グロブリン鎖(例えばIGHM, IGJ, IGLL1, IGKC, IGHD)および形質細胞マーカーCD38をコードする遺伝子を含む;
M 1.2	96	血小板、接着、 凝集、内皮、 脈管	血小板。 血小板糖タンパク質(ITGA2B, ITGB3, GP6, GP1A/B)、およびPPPB(プロ血小板塩基性タンパク質)およびPF4(血小板因子4)のような血小板由来免疫メディエーターをコードする遺伝子を含む;
M 1.3	47	イムノレセプター、 BCR、B細胞、IgG	B細胞。 B細胞表面マーカー(CD72, CD79A/B, CD19, CD22)および他のB細胞関連分子:初期B細胞因子(EBF)、B細胞リンカー(BLNK)およびBリンパ系チロシンキナーゼ(BLK)をコードする遺伝子を含む;
M 1.4	87	複製、抑制、 修復、CREB、 リンパ系、 TNF-アルファ	未決定。 この組はcAMPシグナル伝達経路のレギュレーターおよび標的(JUND, ATF4, CREM, PDE4, NR4A2, VIL2)、ならびにTNF-アルファ媒介型NF-KB活性化のリプレッサー(CYLD, ASK, TNFAIP3)を含む;
M 1.5	130	単球、 樹状突起の、 MHC、同時刺激、 TLR4、MYD88	ミエロイド系。 ミエロイド系細胞により発現される分子(CD86, CD163, FCGR2A)(その中には病原体の認識に関与するものもある:CD14, TLR2, MYD88)の細胞により発現される分子を含む。この組はTNFファミリーメンバー(TNFR2, BAFF)も含む;
M 1.6	28	亜鉛、 フィンガー、 P53、RAS	未決定。 この組はシグナル伝達分子、例えば活性化STAT(PIAS1およびPIAS2)の亜鉛フィンガー含有インヒビター、または活性化T細胞の核因子NFATC3をコードする遺伝子を含む;
M 1.7	127	リボゾーム、 翻訳、40S、 60S、HLA	MHC/リボゾームタンパク質。 MHCクラスI分子(HLA-A,B,C,G,E)+ベータ2-ミクログロブリン(B2M)またはリボゾームタンパク質(RPLs, RPSs)をコードする遺伝子によりほとんど排他的に形成される;
M 1.8	86	代謝、生合成、 複製、ヘリカーゼ	未決定。 代謝酵素(GLS, NSF1, NAT1)およびDNA複製に関与する因子(PURA, TERF2, EIF2S1)をコードする遺伝子を含む;
M 2.1	72	NK、キラー、 細胞溶解、CD8、 細胞媒介型、 T細胞、CTL、 IFN-g	細胞傷害性細胞。 細胞傷害性T細胞およびNK細胞表面マーカー(CD8A, CD2, CD160, NKG7, KLRs)、細胞溶解分子(グランザイム、パーホリン、グラニュリシン)、ケモカイン(CCL5, XCL1)およびCTL/NK-細胞関連分子(CTSW)を含む;

10

20

30

40

【 0 1 0 2 】

【表 17】

モジュール I. D.	プローブの 組数	キーワードの 選択	評価	
M 2.2	44	顆粒球、 好中球、防御、 ミエロイド、骨髄	好中球。この組は好中球顆粒(ラクトランスフェリン: LTF、デフェンシン: DEAF1、細菌透過性増加タンパク質: BPI、カセリシジン抗微生物タンパク質: CAMP…)に見いだされる内在性分子を含む;	
M 2.3	94	赤血球、赤、 貧血、グロビン、 ヘモグロビン	赤血球。ヘモグロビン遺伝子(HGBs)および他の赤血球関連遺伝子(赤血球のアルキリン: ANK1、グリコホリン C:GYPC、ヒドロキシメチルピラン シンターゼ: HMBS、赤血球関連因子: ERAF)を含む;	10
M 2.4	118	リボヌクレオ タンパク質、 60S、核小体、 アッセンブリー、 延長	リボゾームタンパク質: リボゾームタンパク質(RPLs, RPSs)、真核生物翻訳伸長因子ファミリーメンバー (EEFs)および核小体タンパク質(NPM1, NOAL2, NAP1L1)をコードする遺伝子を含む;	
M 2.5	242	腺癌、間質の、 間葉、 樹状突起、 運動	未決定。このモジュールは免疫関連(CD40, CD80, CXCL12, IFNA5, IL4R)ならびに細胞骨格関連分子(ミオシン、サイトキネシスのデディケーター、シンデカン 2、プレキシシンC1、ジストロプレビン)をコードする遺伝子を含む;	20
M 2.6	110	顆粒球、単球、 ミエロイド、ERK、 壊死	ミエロイド系。M1.5に関連する。単球および好中球のようなミエロイド系細胞 (IGTB2/CD18、リンホトキシンベータ受容体、ミエロイド関連タンパク質 8/14ホルミルペプチド受容体1)で発現する遺伝子を含む;	
M 2.7	43	キーワードは 抽出されない	未決定。このモジュールは、未知の機能の転写産物からほぼ構成される。ケモカイン様因子スーパーファミリー(CKLFSF8)の1メンバーを含め、文献に関連するわずか20個の遺伝子;	
M 2.8	104	リンパ腫、 T細胞、CD4、 CD8、TCR、 胸腺、リンパ系、 IL2	T細胞:T細胞表面マーカー(CD5, CD6, CD7, CD26, CD28, CD96)およびリンパ系細胞により発現される分子(リンホトキシンベータ IL2-誘導性T細胞キナーゼ.TCF7, T細胞分化タンパク質 mal, GATA3, STAT5B)を含む;	30
M 2.9	122	ERK、 トランス活性化、 細胞骨格、 MAPK, JNK	未決定。細胞骨格に関連する分子 (アクチン関連タンパク質 2/3, MAPK1, MAP3K1, RAB5A)をコードする遺伝子を含む。 また存在するのは、T細胞が発現する遺伝子 (FAS, ITGA4/CD49D, ZNF1A1);	
M 2.10	44	ミエロイド、 マクロファージ、 樹状突起、 炎症、 インターロイキン	未決定。免疫関連細胞表面分子(CD36, CD86, LILRB)、サイトカイン(IL15)およびシグナル伝達経路に参与する分子(FYB, TICAM2-Toll-様受容体経路)をコードする遺伝子を含む;	40
M 2.11	77	複製、抑制、 RAS、自己リン 酸化作用、 発癌遺伝子	未決定。キナーゼ(UHMK1, CSNK1G1, CDK6, WNK1, TAOK1, CALM2, PRKCI, ITPKB, SRPK 2, STK17B, DYRK2, PIK3R1, STK4, CLK4, PKN2)およびRAS ファミリーメンバー (G3BP, RAB14, RAS A2, RAP2A, KRAS)を含む;	

【 0 1 0 3 】

【表 18】

M 3.1	80	ISRE、インフルエンザ、抗ウイルス、IFN-ガンマ、IFN-アルファ、インターフェロン	インターフェロン誘導性。この組はインターフェロン誘導性遺伝子: 抗ウイルス分子 (OAS1/2/3/L, GBP1, G1P2, EIF2AK2/PKR, MX1, PML)、ケモカイン(CXCL10/IP-10)、シグナル伝達分子(STAT1, STAT2, IRF7, ISGF3G)を含む;	
M 3.2	230	TGF-ベータ、TNF、炎症、アポトーシスの、リポ多糖	炎症I。炎症プロセスに関与する分子(IL8, ICAM1, C5R1, CD44, PLAUR, IL1A, CXCL16)、およびアポトーシスのレギュレーター(MCL1, FOXO3A, RARA, BCL3/6/2A1, GADD45B)をコードする遺伝子を含む;	10
M 3.3	230	顆粒球、炎症、防御、酸化する、リゾソームの	炎症II。顆粒球-マクロファージCSFを誘導する、またはそれにより誘導性の(SPI1, IL18, ALOX5, ANPEP)、ならびにリゾソーム酵素を誘導する、またはそれにより誘導性の(PPT1, CTSS/S, CE S1, NEU1, ASAHI, LAMP2, CAST)の分子を含む;	
M 3.4	323	キーワードは抽出されない	未決定。タンパク質ホスフェート(PPP1R12A, PTPRC, PPP1CB, PPM1B)およびホスホイノシチド3-キナーゼ (PI3K)ファミリーメンバー(PIK3CA, PIK3A, PIP5K3)を含む;	20
M 3.5	19	キーワードは抽出されない	未決定。わずかな数の転写産物から構成される。ヘモグロビン遺伝子(HBA1, HBA2, HBB)を含む;	20
M 3.6	233	補体、宿主、酸化的、細胞骨格、T細胞	未決定。この大変大きい組はT細胞表面マーカー(CD101, CD102, CD103)、ならびに血中白血球の回りで偏在的に発現される分子(CXCR1: フラクタルキン受容体, CD47, P-セレクチンリガンド)を含む;	
M 3.7	80	スプライノソーム、メチル化、ユビキチン、ベーターカテニン	未決定。プロテアソームサブユニット(PSMA2/5, PSMB5/8); ユビキチンタンパク質リガーゼ HIP2, STUB1、ならびにユビキチンリガーゼ複合体(SUGT1)の成分をコードする遺伝子を含む;	30
M 3.8	182	CDR、TCR、CREB、グリコシラーゼ	未決定。幾つかの酵素: アミノメチルトランスフェラーゼ、アルギニルトランスフェラーゼ、アスパラギンシンテターゼ、ジアシルグリセロールキナーゼ、イノシトールホスファターゼ、メチルトランスフェラーゼ、ヘリカーゼ... をコードする遺伝子を含む;	
M 3.9	261	クロマチン、チェックポイント、複製、トランス活性化	未決定。プロテインキナーゼ(PRKPIR, PRKDC, PRKCI)およびホスファターゼ (例えばPTPLB, PPP1R8/2CB)をコードする遺伝子を含む。またRAS癌遺伝子ファミリーメンバーおよびNK細胞受容体2B4 (CD244)も含む;	40

【0104】

モジュールに基づくマイクロアレイデータマイニング法。「従来の」マイクロアレイ分析からの結果は、よく知られているようにノイズが高く、そして解釈が難しい。マイクロアレイデータ分析に関して広く受け入れられている取り組みには、3つの基礎段階を含む: 1) 実験群の間で差次的に発現される遺伝子を選択するための統計的試験の使用; 2) 得られた遺伝子リストの間の兆候を同定するために、パターン発見アルゴリズムの適用する; および 3) 文献またはオントロジーデータベースからの知識を使用してデータを解釈する。

【0105】

本発明は分析の初期段階で生物学的に関連する転写産物の選択を強調する新規マイクロ

アレイデータマイニング法を使用する。この第1段階は、例えば上記のモジュラーマイニングアルゴリズムを、詳細な各転写モジュールの徹底的な特性決定に使用する機能的マイニングツールと組み合わせて使用して行うことができる(図4; 上パネル、工程1)。分析は群間の遺伝子発現レベルの差を考慮しない。むしろ代わりに本発明は生物学的変動(例えば患者集団の中の個体間の変動)により生じる複雑な遺伝子発現パターンに注目する。所定の生物学的系に関連する転写成分を定めた後、分析の第2段階には異なる実験群の比較を介して遺伝子発現における変化の分析を含む(図4: 下パネル、工程2)。群比較分析は、各モジュールについて独立して行われる。モジュールレベルでの変化は、有意な基準を満たす遺伝子の比率として表される(図5のパイチャートまたは図6のスポットとして表される)。モジュールレベルでの比較を行うことは、数千の試験を「無作為」な遺伝子の集合について行う場合により生成するノイズの回避を可能とすることに注目されたい。

10

【0106】

ヒト疾患におけるモジュラーPBMC転写プロファイルの混乱。上記のマイクロアレイデータマイニング法の第2段階を具体的に説明するために(図4)、全身性エリテマトーデス(SLE)の18名の児童および急性インフルエンザA感染の16名の児童からなる2つの小児科の患者集団から得たPBMCサンプルに関する遺伝子発現データを得、比較し、そして分析した。各患者の一群はその各々の対照群と合っていた(健康な有志: 11および10名の供与体がそれぞれSLEおよびインフルエンザ群と合った)。図4に示す分析的スキームの後、各個別のモジュールについて患者と健康群との間で、および各モジュールで有意に変化した遺伝子の測定比率の間で統計的な群の比較(図5)を行った。統計的な群の比較法では使用者はノイズを最少量で含む遺伝子の十分定められた群についての分析に集中でき、そして同定可能な生物学的意味をもつことが可能となる。これらの結果のグラフ表示に対する要点を、図4に提供する。

20

【0107】

以下の知見がなされた:(1)M3.1の遺伝子の大きな集団(「インターフェロン関連」)は、FluおよびSLE群の両方で有意なレベルを満たした(それぞれ84%および94%)。この考察ではSLE患者(19)での以前の研究が確認され、そして急性インフルエンザ感染の患者にインターフェロン兆候の存在を同定する。(2)M1.3における均等比の遺伝子(「B-細胞関連」)は、両群(53%)で有意に変化し、50%を越えて2つのリスト間で重複する。今度は遺伝子は健康な群に比べて患者でかなり発現が抑えられた。(3)またモジュールは、2つの疾患を区別することも分かった。モジュール1.1で有意に変化した遺伝子の比率は、SLE患者で39%に達し、そしてFlu患者ではわずか7%であり、これは0.05の有意レベルで、偶然に差次的に発現されるだけであると予想される遺伝子群に大変近い。興味深いことに、このモジュールはほぼ排他的に免疫グロブリン鎖をコードする遺伝子からなり、そして形質細胞と関連した。しかしこのモジュールは明らかにB細胞関連モジュール(M1.3)とは遺伝子発現レベルおよびパターンの両方の意味で異なる(示さず)。(4)モジュールM1.5により具体的に説明されるように、個々のモジュールの遺伝子レベルの分析を使用して、2つの疾患をさらに区別することができる。これもまたM1.3の場合に、ここではモジュールレベルでの差異が存在しないにもかかわらず(図4: 53%の発現を抑えられた転写産物)、FluとSLE群との間の差異は遺伝子レベルで同定できた(M1.3ではわずかに51%の抑えて発現された転写産物が、2つの疾患群で共通した)。これらの例は、モジュラー構造の使用が分析およびマイクロアレイの結果の解釈を簡素化することを具体的に説明している。

30

40

【0108】

モジュールレベルでの遺伝子発現の変化におけるマッピング。データの視覚化は複雑なデータセットの解釈に最高であり、そして我々はモジュールレベルで起こる変化の包括的なグラフによる説明を提供するために調査した。異なる疾患により生じた遺伝子発現レベルでの変化が、28種のPBMC転写モジュールについて表示された(図6)。各疾患群

50

は、年齢および性別が合った健康な供与体からなるその各対照群と比較される（18名のSLE患者、16名の急性インフルエンザ感染、16名の転移性黒色腫、および免疫抑制薬剤処置を受けている16名の肝臓移植受容者が、10から11名の健康な個体からなる対照群と比較された）。モジュールレベルのデータは、格子上に並んだスポットによりグラフで表示され、各位置は異なるモジュールに対応する（各モジュールに関する機能的注釈に関しては表1を参照にされたい）。

【0109】

スポットの強度は、各モジュールについて有意に変化した遺伝子の比を示す。スポットの色は、変化の極性を示す（赤：過剰発現した遺伝子の比、青：抑えて発現された遺伝子の比；過剰および抑制して発現された両遺伝子を含むモジュールは紫から無色となる有意な比で観察された）。この表示はPBMCT転写系の混乱の迅速な評価を可能とする。そのような「モジュールマップ」を各疾患について作成した。4つのマップを比較する時、我々は疾患が独自のモジュールの組み合わせにより特徴付けられることが分かった。実際に、M1.1およびM1.2に関する結果だけで4種のすべての疾患を識別するに十分であった（M1.1/M1.2：SLE = +/+；FLU = 0/0；黒色腫 = -/+；移植 = -/-）。M3.2（「炎症」）の多数の遺伝子がすべての疾患で過剰に発現され（特に移植群で過剰）、一方、M3.1の遺伝子（インターフェロン）は、SLE患者、インフルエンザ感染およびある程度の移植受容者で過剰に発現された。「リボゾームタンパク質」モジュール遺伝子（M1.7およびM2.4）は、SLEおよびFlu群の両方で抑制されて発現した。これら遺伝子の発現レベルは最近、SLE患者での疾患活性に逆相関するようになることが分かった（Bennett et al., 提出）。M2.8は、T細胞を標的とする免疫抑制剤で処置したリンパ球減少SLE患者および移植受容体で抑制して発現されるT細胞転写産物を含む。

【0110】

興味深いことに、各モジュール中の差次的に発現される遺伝子は、主に抑制されて発現されるか、または過剰に発現されるかのいずれかであった（図5および図6）。さらにモジュールは群間の発現レベルにおける変化ではなく、遺伝子発現プロファイルの類似性に基づき純粋に選択された。遺伝子発現の変化は各モジュール内で高度に分極するらしいという事実は、モジュールデータの機能的関連性を表す。このように、本発明は患者の血中白血球転写プロファイルのモジュール分析により疾患のフィンガープリンティングを可能にする。

【0111】

公開されたデータセットにおけるPBMCTモジュールの検証。次いで「第3団」のデータセットにおける上記PBMCT転写モジュールの有効性を試験した。ヒトPBMCT(20)を対象として、運動が遺伝子発現に及ぼす効果を調査したConnolly, et al., からの実験を試験した。

【0112】

簡単に説明すると、サンプルは15名の健康な男性から、30分間の一定の作業速度サイクルのエルゴメトリーを行う前および直後、そして運動が終了してから1時間後に得た。転写プロファイルはアフィメトリックスU133A遺伝子チップを使用して、各々3名の個体の5個のRNAプールについて作成した。生の発現データはNCBI遺伝子発現オムニバスウェブサイト(21)からダウンロードし、そしてモジュール対モジュール基準で遺伝子発現の変化を分析した。図7はモジュールM1.1（「形質細胞」）、M1.7（「リボゾームタンパク質」）およびM2.1（「細胞傷害性細胞」）の転写プロファイルを表す。これら各モジュールの転写挙動は、明らかに異なった。興味深いことに、個体プール（M1.1）、実験条件（M2.1）間に差異が見いだされ、または差異が無かった（M1.7）。これらのデータは、運動直後の細胞傷害性細胞に関連する遺伝子（M2.1）の発現に上昇を、これに続いて回復後のベースラインに匹敵するレベルへの低下を明らかに示す。この知見は、座っている個体を対象として運動後に観察された循環しているナチュラルキラー細胞における上昇と一致する（22、23）。M2.1に含まれる

遺伝子の幾つかは、Connolly, et al., により「炎症応答」の基準に列挙されるが、この著者は細胞組成における変化の可能性と結び付けなかった。「炎症」モジュール(M3.2、M3.3)に属するほんの少数の遺伝子しか、広い範囲の疾患でこれらのモジュールを含んでなる遺伝子の発現レベルが上昇することはないという事実にもかかわらず、運動後に変化することが分かった(Chaussabel et al., 提出)。しかし興味深いことに、第IV段階の黒色腫患者および移植患者で特異的に過剰発現する免疫抑制分子は(Chaussabel et al., 提出)、運動後、一時的に増加することが分かった(示さず、M1.4; 例えばTCF8、CREM、RGS1、TNFAIP3)。

【0113】

この分析からの結果を合わせると、独立した調査群により作成されたデータの内容で、提案されたモジュラーマイニング法の有効性が示される。本発明を使用して、モジュールの転写データはマイクロアレイ平板全体で再現性があることが分かった。

【0114】

最初に、2種類の市販されているマイクロアレイ平板を使用して得たモジュール転写プロファイルと比較した。PBMCは4名の健康な有志および10名の肝移植受容者から供与された14個のサンプルから単離した。同じ起源の全RNAから始めて、標的は独立して作成され、そしてアフィメトリックスU133 Gene Chips (バイロール免疫調査研究所で)、およびイルミナHuman Ref8 Bead Chip (イルミナサービスコアで)を使用して分析された。2つのマイクロアレイ技術の間には基本的な差異が存在する(詳細については方法を参照にされたい)。各製造元により提供されるプローブIDは、遺伝子発現プロファイルに合わせるために使用した独自のID(NCBI Entrez 遺伝子ID)に転換された。モジュールM1.2(「血小板」)、M3.1(「インターフェロン」)およびM3.2(「炎症」)に関して共有する遺伝子組について得たデータは、図8に示される。イルミナのビーズチップで得られたデータから引き出されたプロファイルは、各モジュール内の遺伝子間で大変高レベルの同時発現を示す。この観察は、転写モジュールの選択がアフィメトリックスのGene Chipsを使用して作成された遺伝子発現データにほとんど基づいていたので、特に意味がある。さらに各モジュールについて平均化された遺伝子発現値は、マイクロアレイ平板全体で再現性が高かった(図8)。

【0115】

これらの結果は、モジュールの転写兆候の強さを証明し、そしてモジュールレベルの分析が異なる場所および異なる平板で作成されたマイクロアレイデータの再現性に関して、関心を傾ける潜在能力があることを明らかに示す。

【0116】

マイクロアレイ遺伝子発現データは生物学的系の包括的な、しかし組織化されていない観点を作り出す。マイクロアレイに基づく調査が直面する課題には3つの部分がある:(1)ノイズ、(2)データ解釈および(3)再現性。ノイズに関して、本発明は成功裏に何万もの遺伝子を比較し、これは従来技術の方法では大きな比率のノイズを含む結果を必ず生じる(24)。データ解釈に関しては、本発明は情報の過負荷の問題を克服する。実際に、マイクロアレイのデータを解釈するには調査者が既存の生物医学知識の内容で実験データをゲノム全域の規模で調査する必要があることが多い(13)。さらに不安になることは、ノイズが多いデータの解釈を通して偽の結果が生じる恐れである(7)。最後に再現性に関して、既存の技術の重大な問題は、異なる研究室および平板全体により得られるマイクロアレイの結果の良くない再現性が混乱を起こし続けていることが詳細に記録され、今日まで主要な関心事として残っている(6, 7, 10~12)。

【0117】

主流のマイクロアレイ分析法は、幾つかの理由によりこの3つの論点の取り組みに限定的に成功しただけであった。まず第1に、統計的試験が分析の初期段階の先行条件と考えられるからである。その結果、いったん差次的に発現された遺伝子のリストが生成される

10

20

30

40

50

だけで、生物学的考察が効果を示すようになる。しかし何万もの変数を試験することから生じるデータのサブセットは必ずノイズを含み、したがって解釈することは特に難しい。本発明の系および方法は、モジュールの特徴を決定する場合、細胞の細胞および分子生物学を考慮する。本発明では、第1段階は分析の極めて初期の段階で系の生物学を考慮することであり、これにより数百ものサンプル全体で調和して発現されると判明した機能に連結している遺伝子の組を選択する。次いで統計的試験を生物学的に意味のある遺伝子がかなり濃縮しているモジュールのデータセットに適用する。この取り組みのさらなる利点は、基本単位として転写モジュールを使用することにより、遺伝子レベルの分析を越える点である。転写モジュールは、定めた生物学的系の内容で起こる混乱の分析に関する骨格を構成する。このモジュールデータ形式は、マイクロアレイ実験の解釈の簡素化に役立つ。しかしこれには、膨大な範囲の生物学的変数、例えば異なる実験条件、個体間の変動の下での各実験系の予備的特性決定が必要となり、そして生物材料の経費および入手も限定され得る。

10

【0118】

興味深いことに、モジュールレベルの分析からのデータは、2つの市販されているマイクロアレイ平板をわたって得られる優れた再現性により示されるように、特に確固として示されることが示された。さらにPBM C転写モジュールの多変数分析は、「ゲノムスコア」の確立を導き、これは全身性エリテマトーデスの患者において疾患の重篤度に正確な評価を提供した(Bennett et al., 提出)。信頼性のある血中白血球転写マーカーの同定は、臨床的状況におけるマイクロアレイの応用に向けた重要な段階を構成する。

20

【0119】

多くの細胞型により形成されるサンプルでの作業は、マイクロアレイ遺伝子発現データの分析に複雑性レベルを加える。実際、遺伝子発現レベルの差異は転写活性の変化だけでなく、細胞組成の変化によっても説明され得る。PBM Cサンプルを分析することにより得られるモジュールの兆候はこの事実を反映し、そして細胞成分(血小板 - M1.2、赤血球 - M2.3またはT細胞 - M2.8に関連する遺伝子)を活性化に関連する成分(インターフェロン - M3.1、炎症M3.2またはシグナル伝達 - M2.11に関連する遺伝子)と区別できるようにする。この種の考察は、この内容で行われるマイクロアレイ分析の大半が多細胞サンプルを含むので、患者に基づく調査に関連する。

30

【0120】

アフィメトリックスおよびイルミナの平板により作成されるモジュール発現データは、高度に比較可能である(図9:移植群のピアソン相関係数はM1.2、M3.1およびM3.2についてそれぞれ $R^2 = 0.83$, 0.98 および 0.93 ; $p < 0.0001$)。合わせると、これらの結果はモジュール転写データがマイクロアレイ平板全体で再生できることを証明している。この知見は、「モジュールのマイクロアレイスコア」が使用したマイクロアレイ平板から独立して引き出された患者の疾患重篤度を評価するために使用できることを示すので、重要である。

【0121】

この実験で記載したモジュールレベルマイニング法は、広い範囲の生物学的系に使用することができ、そして腫瘍または固体臓器生検材料のような他の臨床的に関連するサンプルの分析にも良く適している。

40

【0122】

発現レベルベクトルは、表3に提供する1もしくは複数のモジュールおよび/または1もしくは複数の遺伝子から得ることができる。さらに本発明に基づき、当業者には明白であるように、疾患発現プロファイルに依存して、および本発明の方法を使用することにより、モジュールおよびモジュール内の遺伝子を開発し、そしてさらに洗練させることが可能である。例えば必要とされる特異性のレベル、データセットの数、患者の数等に依存して、所定の疾患の内容で差次的に発現する遺伝子を異なる比率で含む1もしくは複数の異なる新規モジュールを使用して、新たなデータに基づく新規モジュールを開発し、転写産物

50

の新たなサブセットに基づくアレイを形成し、そして組織することができ、これは平均発現レベルを表す新規ベクトルを定める。

【0123】

表1、2および3は長い表である。本特許出願は長い表の章を含む。表のコピーはUSPTOウェブサイトから電子形式で利用可能である。また表の電子コピーは要求すればUSPTOから37CFR1.19(b)(3)に説明されている料金で入手することもでき、これにはEFSファイルが付いており、そして表1、2および3は全部、参照により包含されている。

【0124】

本明細書に記載する特定の態様は、具体的説明の目的で示し、本発明を限定するものではないと理解される。本発明の主な特徴は本発明の範囲から逸脱せずに様々な態様で採用することができる。当業者は日常的な実験だけで、本明細書に記載する特異的手順の多くの等価物を認識し、または確認することができる。そのような等価物は本発明の範囲内にあると考えられ、そして特許請求の範囲により網羅される。

10

【0125】

本明細書に挙げたすべての公報および特許出願は、本発明が関連する当業者の技術レベルの指標である。すべての公報および特許出願は、各公報または特許出願が引用により具体的に、そして個別に編入されることが示されれば、同程度まで参照により編入する。

【0126】

特許請求の範囲では、「含んでなる」、「含む」、「持つ」、「有する」、「含有する」、「含む」等のようなすべての移行句は広く解釈され、すなわち含むが限定はしないことを意味する。「からなる」および「から本質的になる」という移行句のみ、それぞれ排他的、または半排他的な移行句となる。

20

【0127】

本明細書に開示し、そして特許請求するすべての組成物および/または方法は、本開示に照らして過度な実験を行わずに作成し、そして実行することができる。本発明の組成物および方法は、好適な態様という意味で記載してきたが、当業者には本発明の概念、精神および範囲から逸脱せずに本明細書に記載する組成物、および/または方法、および方法の段階または段階の順序に変化を適用できることは明らかである。さらに詳細には、化学的および生理学的の両方に関連する特定の作用物質を、本明細書に記載する作用物質と置き換えても、同じまたは類似の結果を達成できることは明らかである。当業者には明白であるすべてのそのような類似の置換および修飾は、添付する特許請求の範囲により定められる本発明の精神、範囲および概念の中にあるものと見なす。

30

【0128】

参考文献

1. Golub, T. R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-7 (1999).
2. Alizadeh, A. A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503-11 (2000).
3. Garber, K. Genomic medicine. Gene expression tests foretell breast cancer's future. *Science* 303, 1754-5 (2004).
4. van de Vijver, M. J. et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347, 1999-2009 (2002).

40

50

5. Pascual, V., Allantaz, F., Arce, E., Punaro, M. & Bancherreau, J. Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade. *J Exp Med* 201, 1479-86 (2005).
6. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365, 488-92 (2005). 10
7. Ioannidis, J. P. Microarrays and molecular research: noise discovery? *Lancet* 365, 454-5 (2005).
8. Jarvinen, A. K. et al. Are data from different gene expression microarray platforms comparable? *Genomics* 83, 1164-8 (2004).
9. Tan, P. K. et al. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 31, 5676-84 (2003). 20
10. Bammler, T. et al. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2, 351-6 (2005).
11. Irizarry, R. A. et al. Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2, 345-50 (2005).
12. Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R. & Quackenbush, J. Independence and reproducibility across microarray platforms. *Nat Methods* 2, 337-44 (2005). 30
13. Chaussabel, D. Biomedical literature mining: challenges and solutions in the 'omics' era. *Am J Pharmacogenomics* 4, 383-93 (2004).
14. Rhodes, D. R. et al. Mining for regulatory programs in the cancer transcriptome. *Nat Genet* 37, 579-83 (2005).
15. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nat Genet* 36, 1090-8 (2004). 40
16. Mootha, V. K. et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34, 267-73 (2003).
17. Segal, E., Friedman, N., Kaminski, N., Regev, A. & Koller, D. From signatures to mode 50

- ls: understanding cancer using microarrays. Nat Genet 37 Suppl, S38 - 45 (2005).
18. Chaussabel, D. & Sher, A. Mining microarray expression data by literature profiling. Genome Biol 3, RESEARCH0055 (2002).
19. Bennett, L. et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med 197, 711 - 23 (2003).
20. Connolly, P. H. et al. Effects of exercise on gene expression in human peripheral blood mononuclear cells. J Appl Physiol 97, 1461 - 9 (2004).
21. Barrett, T. et al. NCBI GEO: mining millions of expression profiles - database and tools. Nucleic Acids Res 33, D562 - 6 (2005).
22. Ogawa, K., Oka, J., Yamakawa, J. & Higuchi, M. A single bout of exercise influences natural killer cells in elderly women, especially those who are habitually active. J Strength Cond Res 19, 45 - 50 (2005).
23. Woods, J. A., Evans, J. K., Wolters, B. W., Ceddia, M. A. & McAuley, E. Effects of maximal exercise on natural killer (NK) cell cytotoxicity and responsiveness to interferon-alpha in the young and old. J Gerontol A Biol Sci Med Sci 53, B430 - 7 (1998).
24. Tuma, R. S. Efforts aimed at reducing noise, data overload in microarrays. J Natl Cancer Inst 97, 1173 - 5 (2005).

【0129】

本発明の特徴および利点をさらに完全に理解するために、添付する図面と一緒に本発明の詳細な説明が参照される。

【図面の簡単な説明】

【0130】

【図1】図1A～1Cは、受けた遺伝子レベルのマイクロアレイデータ分析に含まれる基本的なマイクロアレイデータマイニング法の工程(図1A)、本発明のモジュールマイニング法図1bおよびモジュール抽出アルゴリズムの完全なサイズ表示図1Cを表す。図1Cはモジュール抽出アルゴリズムのさらに詳細な観点を提供し、ここで工程(a)はデータ例が定めた実験系の内容で生成されることを示す(例えばエクスピボのPBMC);工程(b)は、転写プロファイルが幾つかの実験群(例えばS1～8)に関して得られることを表し;工程(c)は各群について、遺伝子が類似の発現プロファイルに基づきxクラスター(例えばx=30)に分布することを表し(K-平均クラスタリングアルゴリズムを使用して);工程(d)は異なる実験群にわたる各遺伝子のクラスター分布が表に記録され、そして分布パターンが合うことを表し;そして工程(e)は、モジュールがすべての実験群をわたり同じクラスターに分布する最大の遺伝子組から始めて、反復プロセスを介して選択されることを示し(8群のうちの8つについて同じクラスター内に見いだされる)。選択はこのコア参照パターンから7/8、6/8および5/8の対合を持つ遺伝子へと拡大される。いったんモジュールが形成されれば、遺伝子は選択プールから引き出さ

れる。次いでこのプロセスを、遺伝子の2番目に大きい群から始めて、厳密度のレベルを漸次下げながら進めて繰り返す。

【図2】独立したサンプル群全体のモジュールの遺伝子発現プロファイル。モジュール間の転写挙動の差異は、21名の健康な有志から得た1組のサンプルで具体的に説明される。サンプルはモジュール選択工程で使用されなかった。グラフは転写プロファイルを表し、各線は多くの条件にわたり1つの転写産物の発現レベル(y-軸)を示す(サンプル、x-軸)。モジュール1.2、1.7、2.1および2.11の転写プロファイルを示す。各遺伝子の発現は、すべてのサンプルをわたり得た測定値の中央値に対して標準化される。

【図3】文献中、キーワードの存在の分布は、4組の調和して発現される遺伝子について得た。要約中、用語の存在レベルは少なくとも10の公報に関連してM3.1、M1.5、M1.3およびM1.2内のすべての遺伝子についてコンピューターで計算した(26,000以上の要約を表す)。キーワードプロファイルは各モジュールについて抽出し、そして選択を使用してこの図を生成した。要約中、キーワードの存在のレベルは色付きの尺度で示し、黄色は高い存在を示す。M3.1はインターフェロンに関連し、M1.5は病原体認識モジュール/ミエロイド系細胞に関連し、M1.3はB細胞に関連し、そしてM1.2は血小板に関連する。

【図4】モジュールのマイクロアレイ分析法。提案されたマイクロアレイデータ分析法には、2つの基本的段階を含む：1. 転写系の特性決定：転写成分は、教示なしの(un-supervised)「クラスタリングメタ分析」を介して抽出する(図1)。各モジュール(独自のID、例えばM1.1により命名される)を形成する遺伝子は、定めた実験系についてすべての条件にわたり一貫した転写挙動を有する。転写モジュールは2つのデジタルID(例えば1.1)により同定される。グラフは多くの条件にわたってモジュールを形成する遺伝子の発現プロファイルを表す(サンプル)。次いで各モジュールは機能的に特徴付けられる(例えば文献プロファイルの分析を介して)。この結果は生物学的に意味がある転写決定基の集合である。2. 系の混乱を研究する：実験群間の比較は、各モジュールについて独立して行う。この分析により、異なる条件について発現レベルの変化を同定することが可能となった(例えば患者および健康な対照からのサンプルを比較する)。各モジュールについて得られた結果をグラフに表す。有意な基準を満たす遺伝子の比率(クラス比較)を丸で示し、赤色は有意に過剰発現した遺伝子の比率であり、そして青色は有意に抑制されて発現した遺伝子の比率である。この理論的例では、 $p < 0.05$ の3/4の遺伝子(75%)がグラフ上に表された。これら遺伝子の2つが過剰に発現し(50% - 赤)、そして1つが抑制されて発現した(25% - 青)。

【図5】患者の血中白血球転写プロファイルの分析である。a) 遺伝子レベルの分析。上パネルはSLEまたは急性インフルエンザ感染の患者およびそれらの各対照との間で差次的に発現される転写産物を同定した統計的比較を表す($p < 0.001$ 、マンホイットニーU検定、ベンジャミンおよびホックバーグの誤発見率(Benjamini and Hochberg False Discovery Rate): SLE = 733転写産物、FLU = 234転写産物)。クラスタリング分析は遺伝子を発現パターンに基づきグループ分けし、そして結果がヒートマップにより表される。下のパネルはモジュールレベルの分析である。各モジュールについて、患者(SLEまたはFLU)および各々の健康な有志のPBMCについて得られた遺伝子発現レベルを比較した($p < 0.05$ 、マンホイットニーの順位検定)。パイチャートは、有意に変化した遺伝子の比率を示す。グラフは有意に変化した遺伝子の転写プロファイルを表し、各線は多くの条件(サンプル、x-軸)をわたり1つの転写産物の発現レベル(y-軸)を示す。各遺伝子の発現は、すべてのサンプル全体で得られた測定値の中央値に対して標準化される。28個のPBMC転写モジュールについて得られた結果を格子中に表す。座標はモジュールのID(例えばM2.8は行M2、カラム8)を示すために使用する。スポットは各モジュールについて有意に変化した遺伝子の比率を表す。赤点：過剰発現した遺伝子の比率、青点：抑制されて発現した遺伝子の比率。機能的解釈は色付きコードにより格子中に示す。

10

20

30

40

50

【図6】疾患により生じる転写の変化のモジュールマップ。各モジュールについて、患者および彼らの健康な対照群に由来する単離したP B M Cで測定された発現レベルを比較した(マン-ホイットニー順位検定、 $p < 0.05$: S L Eの18名の患者と11名の健康な有志との間; 16名の急性インフルエンザ感染の患者と10名の有志との間; 16名の転移性黒色腫患者と10名の有志との間; および16名の肝移植受容体 対 10名の有志との間)。スポットは各モジュールについて有意に変化した遺伝子の比率を表す。赤点: 過剰発現した遺伝子の比率、青点: 抑制されて発現した遺伝子の比率。28個のP B M C転写モジュールについて得た結果は、格子中に示す。座標はモジュールIDを示すために使用する(例えばM2.8は行M2、カラム8)。

【図7】第3団のデータセットの分析。モジュールマイクロアレイデータ分析は、公開されたP B M C遺伝子発現データセットについて行った。実験では運動が遺伝子発現に及ぼす効果を調査した。血液サンプルは15名の個体、運動前(P r e)、運動終了時(E n d)および60分の回復(R e)について得た。転写プロファイルは、3名の個体の5個のプールそれぞれについて作成した。発現プロファイルは3つの転写モジュールについて示す。各遺伝子の発現は、すべてのサンプル全体で得られた測定値の中央値に対して標準化する。文献からのキーワード抽出は緑で示す。

【図8】平板の相互的(C r o s s - p l a t f o r m)検証。健康な供与体および肝臓移植受容体からのP B M Cサンプルを、2つの異なるマイクロアレイ平板で分析した。アフィメトリックスU133A&B GeneChipsおよびイルミナS e n t r i x Human Ref8 BeadChips。同じ全RNAのプールを使用して、ピオチン-標識化cRNA標的を別個に調製した。結果は2つの平板により共有される1組の転写産物について表す(アフィメトリックス: 上パネル; イルミナ; 中段パネル)。各遺伝子の発現は、すべてのサンプル全体から得られた測定値の中央値に対して標準化される。各転写モジュールを形成するすべての遺伝子について、平均化された発現値をアフィメトリックスおよびイルミナの平板の両方について下のパネルに示す。

【図9】マイクロアレイ平板をわたるモジュールレベルの発現データの再現性を含む3つのグラフ。健康な供与体および肝移植受容体からのP B M Cサンプルを、2つの異なるマイクロアレイ平板で分析した: アフィメトリックスU133A&B GeneChipsおよびイルミナS e n t r i x Human Ref8 BeadChips。同じ起源の全RNAを使用して、ピオチン-標識cRNA標的を別個に調製した。標準化された「モジュール発現レベル」は、各モジュールを形成する遺伝子の発現値を平均することにより各サンプルについて得た。アフィメトリックスおよびイルミナの平板により生成されたデータに由来するモジュールの発現レベルは、高度に比較可能であった: M1.2、M3.1およびM3.2についてそれぞれピアソン相関係数 $R^2 = 0.83$ 、 0.98 および 0.93 ; $p < 0.0001$)。

10

20

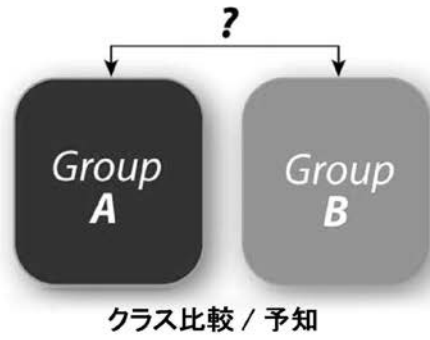
30

【図1A】

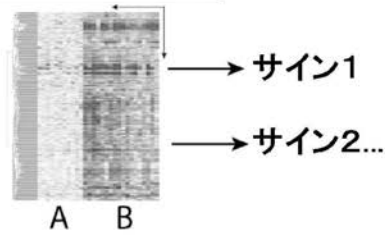
Figure 1A

a. **遺伝子レベルの分析**

I. 統計試験:



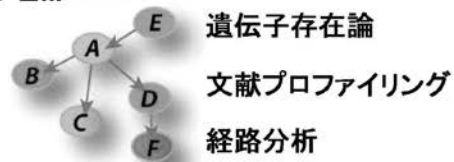
II. パターン発見



III. 機能的注解 / 分析

サイン1

サイン2...

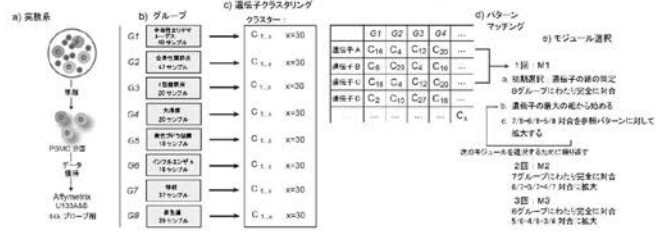


【 図 1 B 】

Figure 1B

b. モジュールレベルの分析

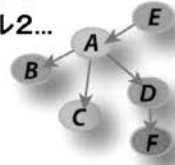
I. モジュール抽出アルゴリズム



II. 機能的注解/分析

モジュール1

モジュール2...

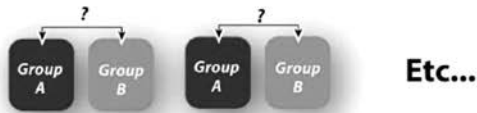
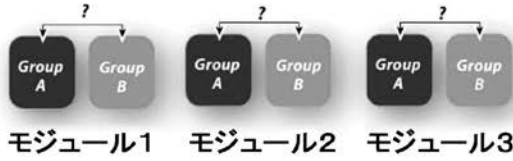


遺伝子存在論

文献ファイリング

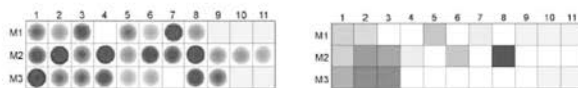
経路分析

III. 統計試験:



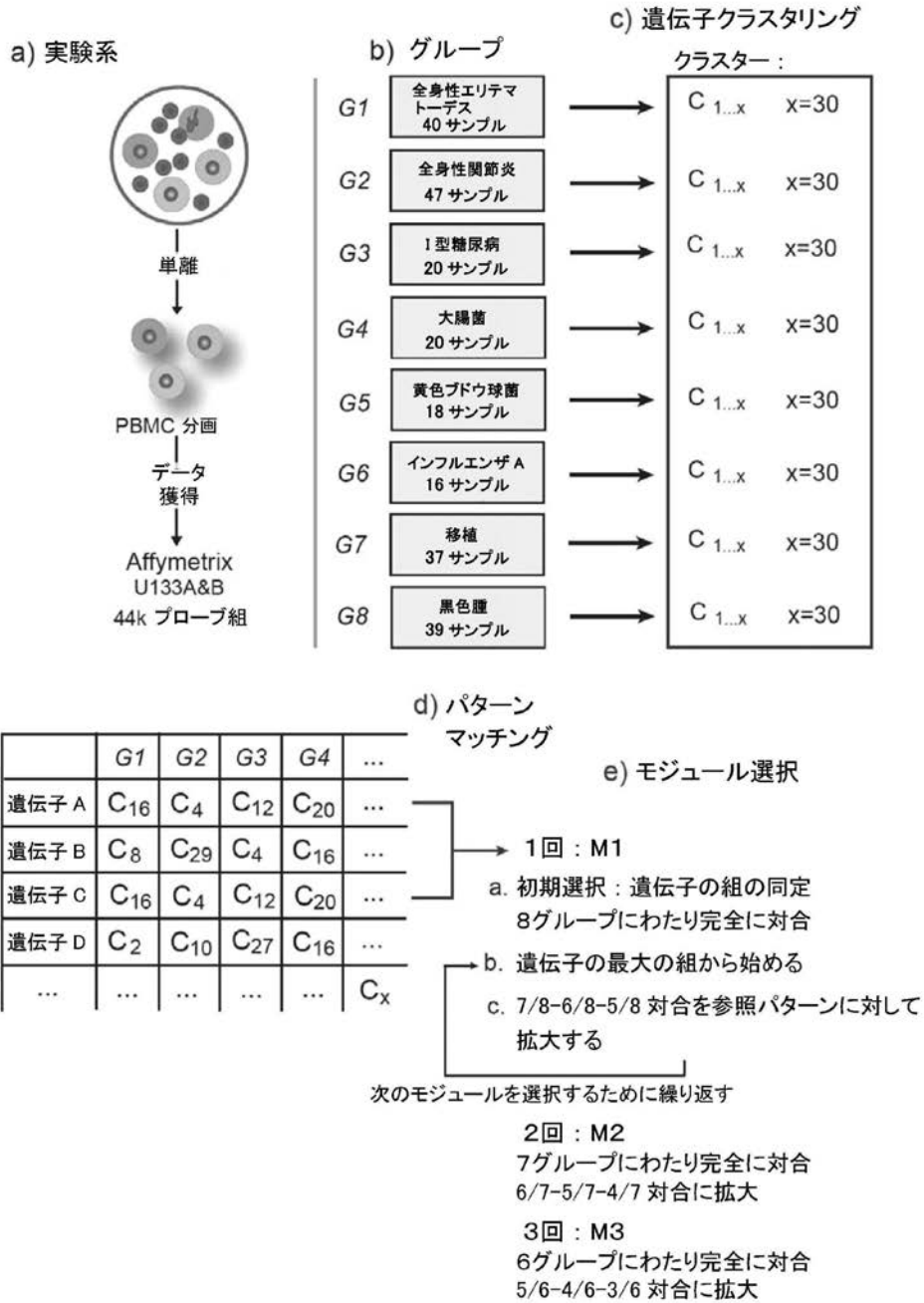
クラス比較 / 予知、モジュール毎

IV. 視覚化 / 解釈:



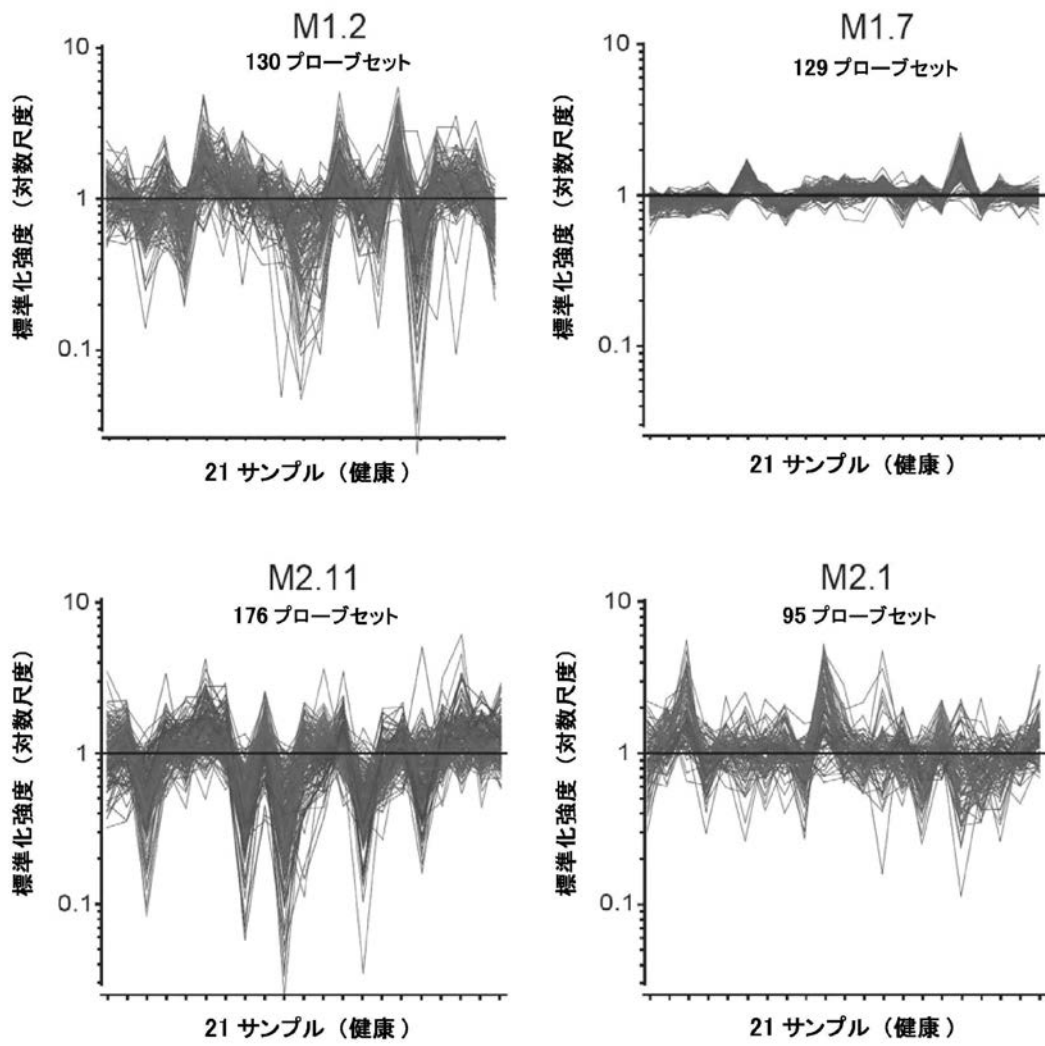
【図1C】

Figure 1C



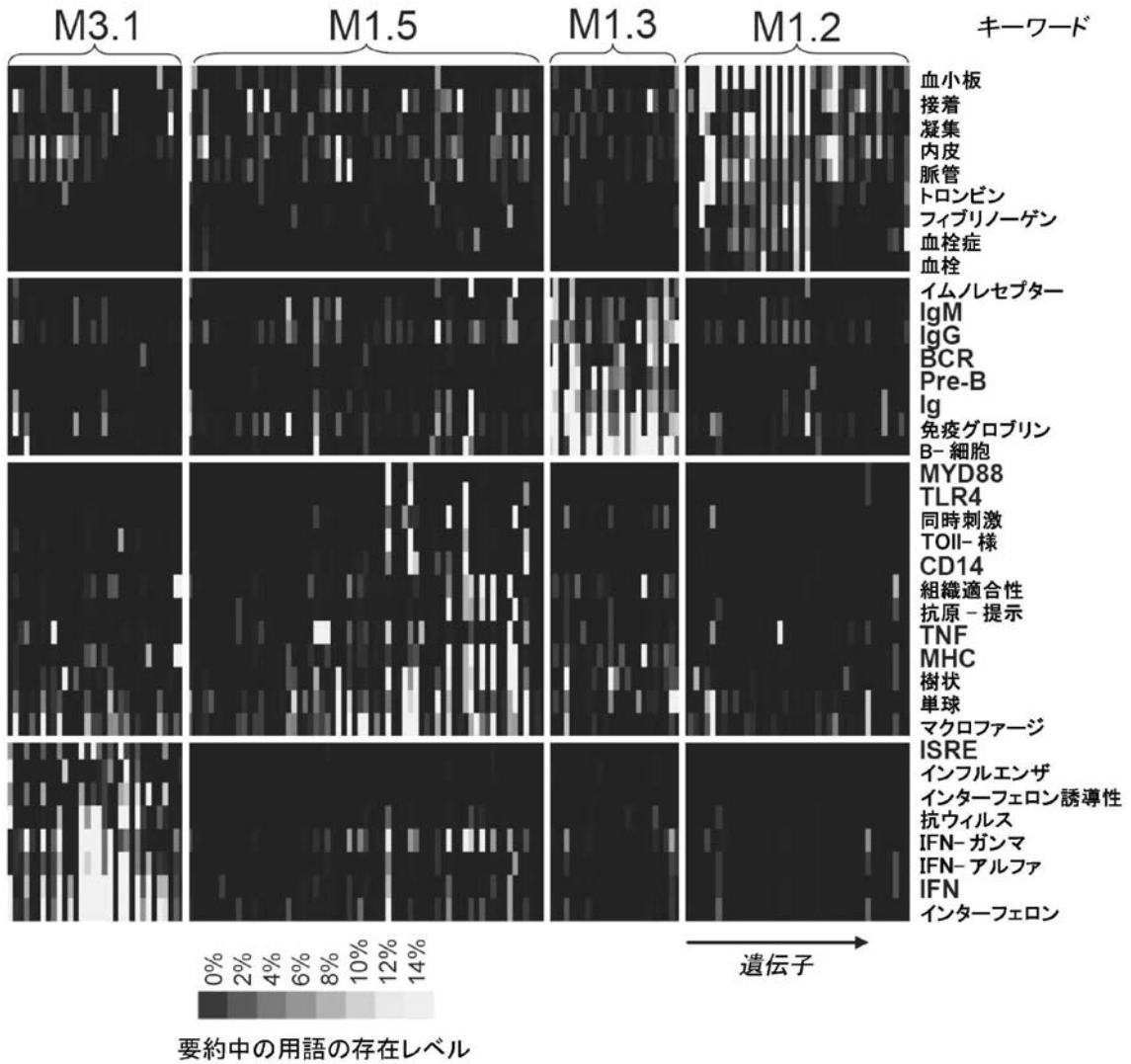
【 図 2 】

Figure 2



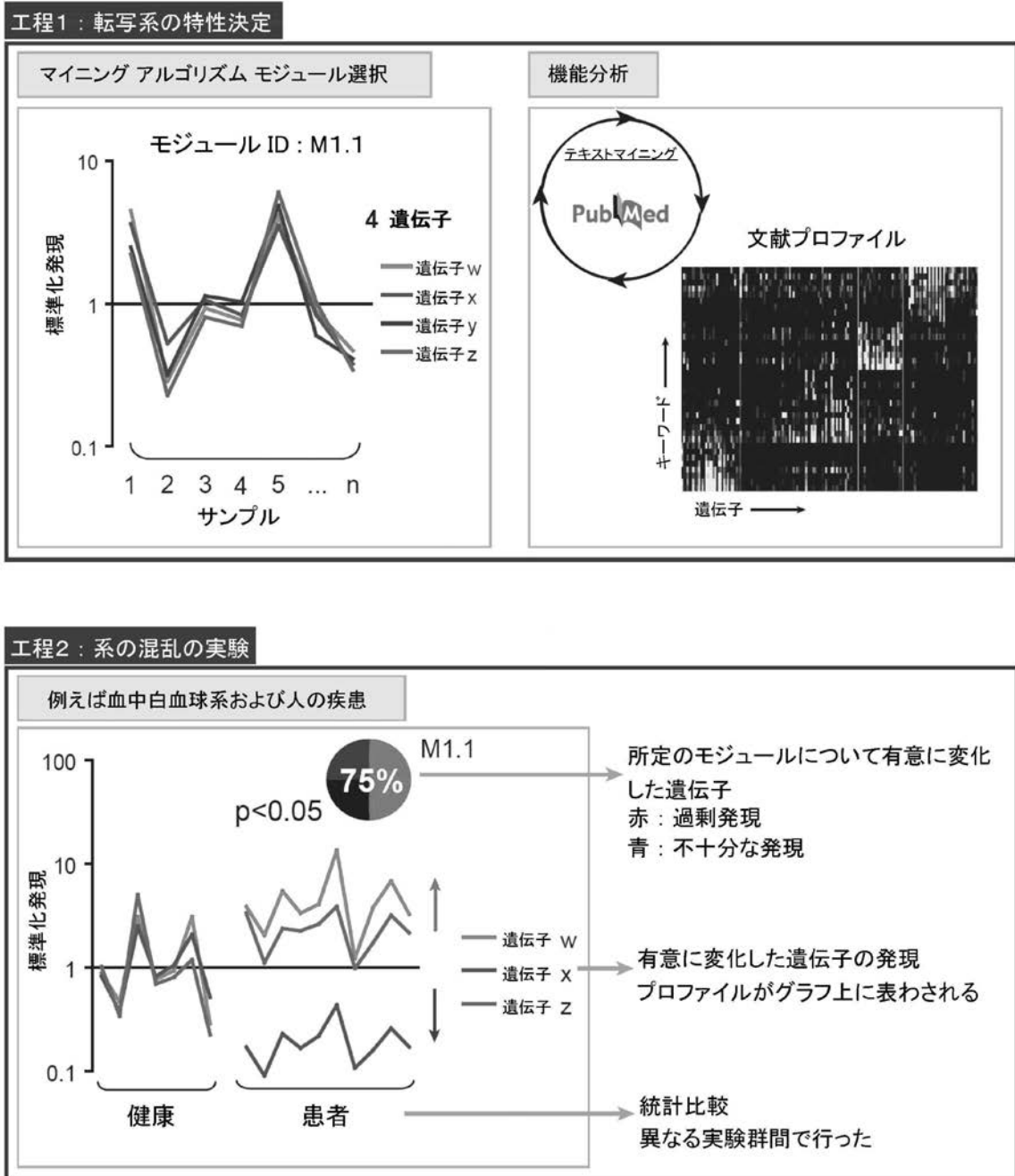
【 図 3 】

Figure 3



【 図 4 】

Figure 4



【 図 5 】

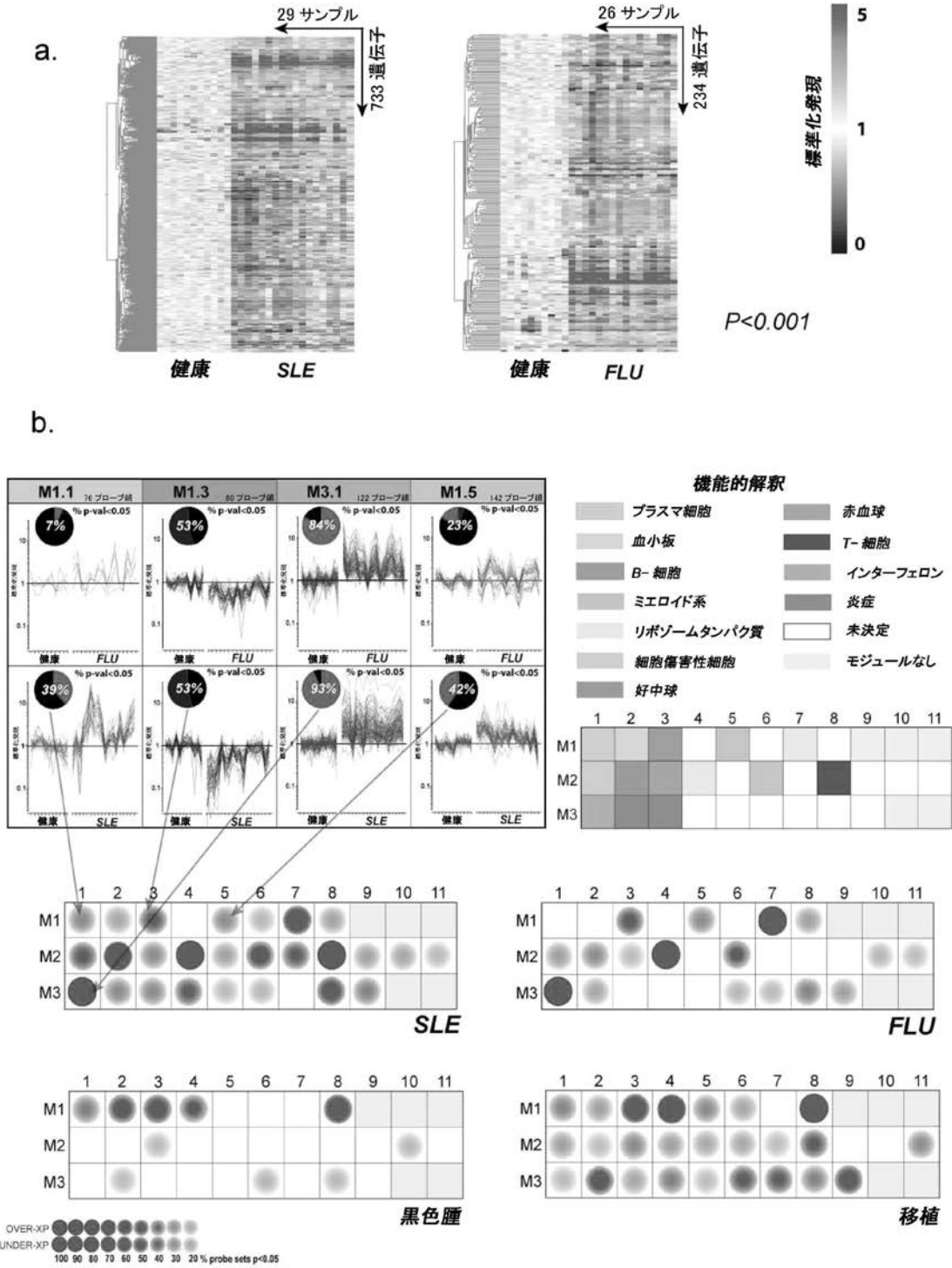
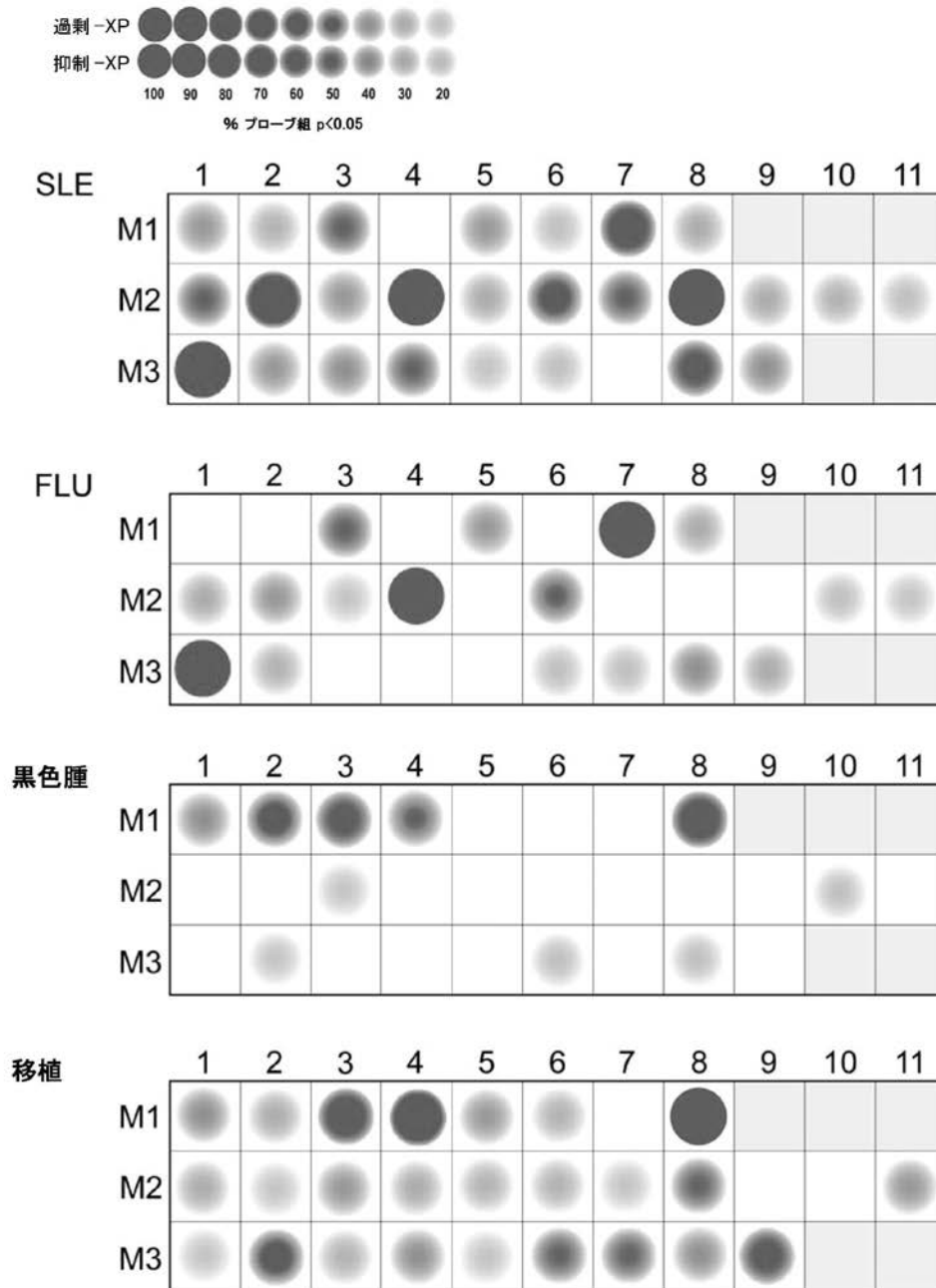


Figure 5

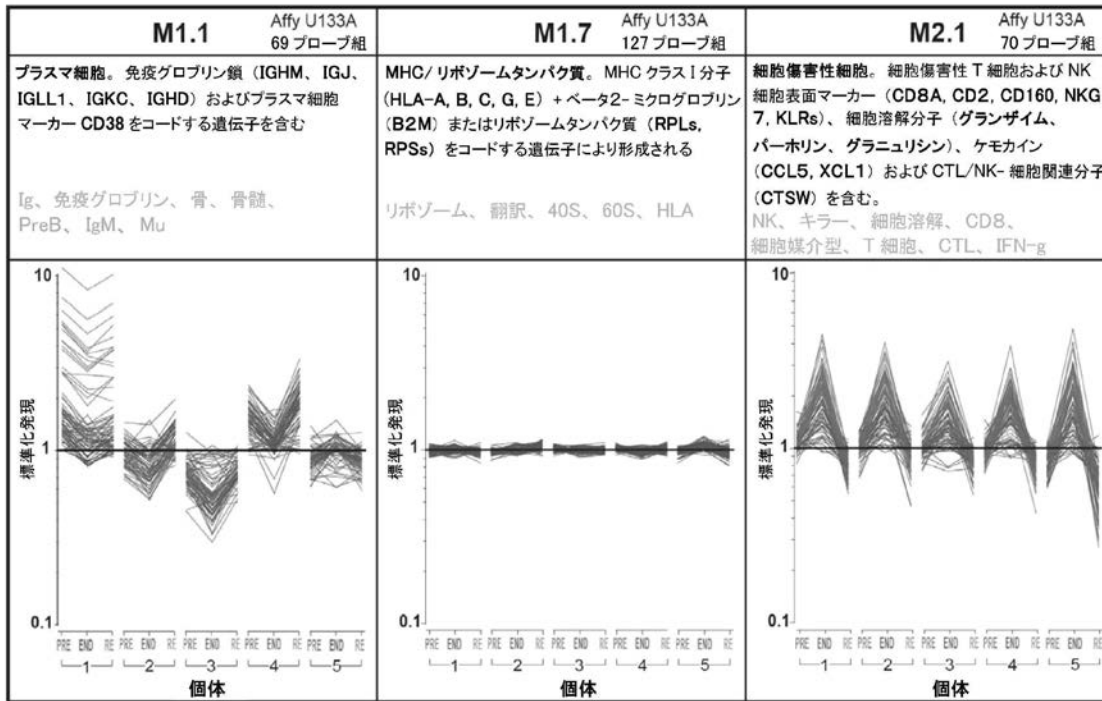
【 図 6 】

Figure 6



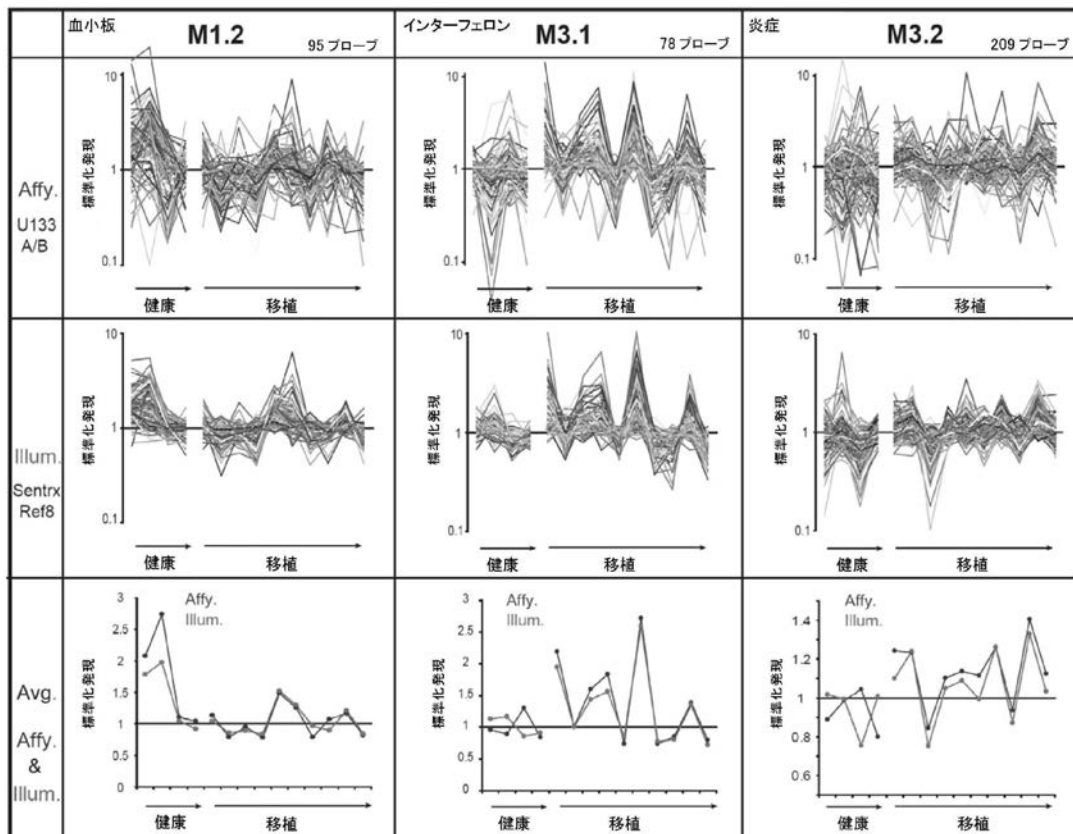
【 図 7 】

Figure 7



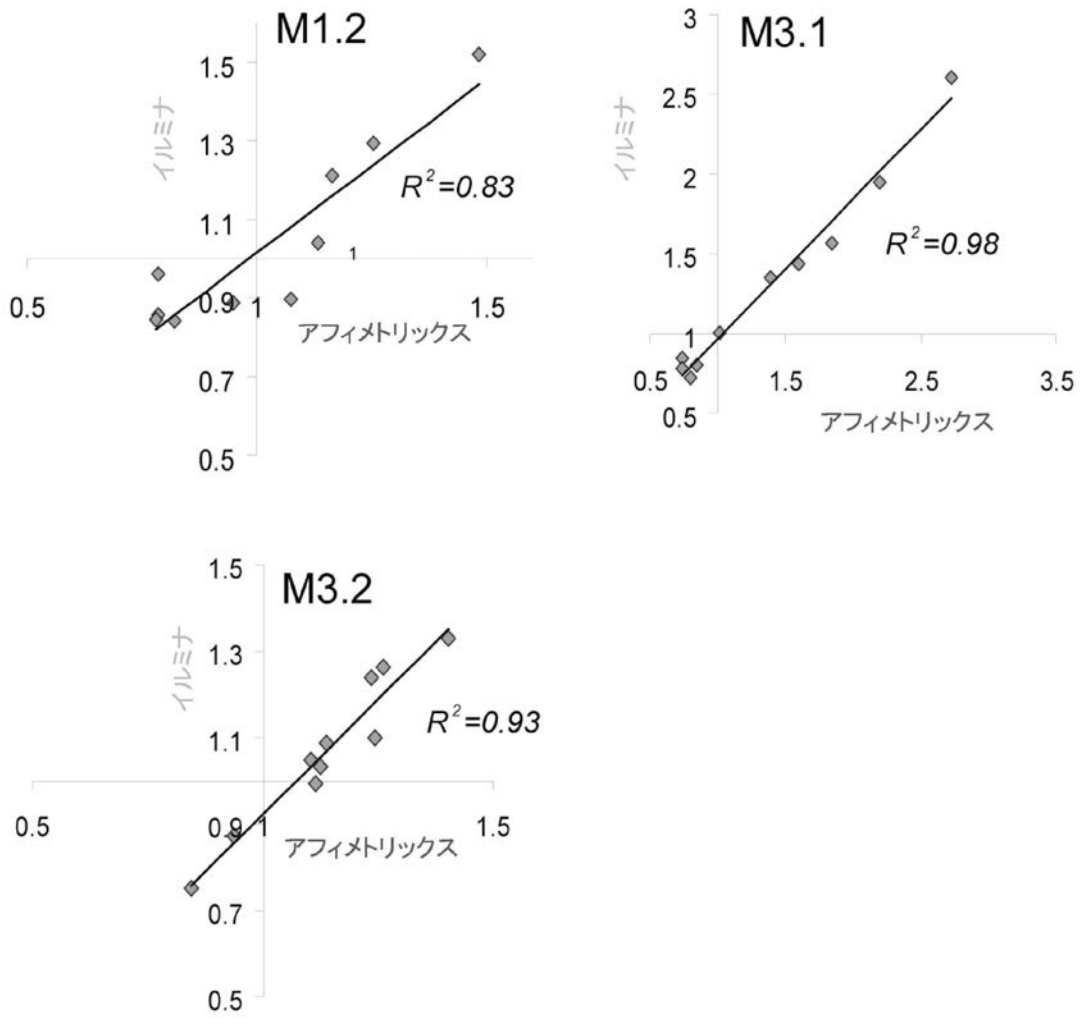
【 図 8 】

Figure 8



【 図 9 】

Figure 9



フロントページの続き

(72)発明者 バンシユロ, ジヤツク・エフ
アメリカ合衆国テキサス州75230ダラス・ノースアブン6730

審査官 柴原 直司

(56)参考文献 米国特許出願公開第2004/0158407(US, A1)
米国特許出願公開第2005/0142573(US, A1)
米国特許出願公開第2003/0114373(US, A1)
米国特許出願公開第2002/0068319(US, A1)
米国特許出願公開第2004/0014044(US, A1)
Eran Segal, et al., A module map showing conditional activity of expression modules in cancer., Nature Genetics, 2004年, Vol.36, No.10, p.1090-1098
Zhao L., et al., TRICLUSTER: an effective algorithm for mining coherent clusters in 3D microarray data., PROCEEDINGS OF THE ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA - SIGMOD 2005, 2005年 6月, P.694-705
WADE C., et al., EBP2 is a member of the yeast RRB regulation, a transcriptionally coregulated set of genes that are required for ribosome and rRNA biosynthesis., Molecular and Cellular Biology, 2001年, Vol.21, No.24, p.8638-8650

(58)調査した分野(Int.Cl., DB名)

C12Q 1/00 - 3/00

CAplus/BIOSIS/MEDLINE(STN)

专利名称(译)	外周血白细胞转录谱的模块水平分析		
公开(公告)号	JP5279505B2	公开(公告)日	2013-09-04
申请号	JP2008544529	申请日	2006-12-09
[标]申请(专利权)人(译)	贝勒研究协会		
申请(专利权)人(译)	贝勒研究仪器希希士		
当前申请(专利权)人(译)	贝勒研究所		
[标]发明人	シヨサベルダミアン バンシユロジャツクエフ		
发明人	シヨサベル,ダミアン バンシユロ,ジャツク・エフ		
IPC分类号	C12N15/09 C12M1/00 C12Q1/68 G01N33/53 G01N37/00		
CPC分类号	C12Q1/6883 C12Q2600/158 G16B20/00 G16B25/00 G16B40/00 Y02A90/24 Y02A90/26		
FI分类号	C12N15/00.F C12M1/00.A C12Q1/68.A G01N33/53.M G01N37/00.102		
代理人(译)	堀内申		
优先权	60/748884 2005-12-09 US 11/446825 2006-06-05 US		
其他公开文献	JP2009518040A JP2009518040A5		
外部链接	Espacenet		

摘要(译)

本发明提供了从一种或多种患有疾病或病症的患者的细胞中获得基因表达水平的方法;记录表中每种基因的表达值,分成簇;通过从每个群体中选择与每种疾病或病症匹配的基因并从分析中去除所选基因来重复选择它;并选择聚集在疾病或病症的亚组分中的基因的基因重复表达值选择的过程;并重复模块的生成,包括与转录模块的开发和使用相关的设备,系统和方法。

サブモジュール	グループの種数	キーワードの選択	評価
M 1.1	68	Ig, 免疫グロブリン, 赤血球, PvdB, IgM, Mu	形質細胞: 免疫グロブリン鎖 (IGHM, IGL, KLL1, IGHG, IGHD) および形質細胞マーカー CD38 をコードする遺伝子。
M 1.2	96	血小板、凝集、凝集、内皮、血管	血小板: 血小板種タンパク質 (ITGA2B, ITGB3, GPIIb/IIIa, GPIIb), および PFA4 (血小板因子4) のような血小板由来免疫マーカーをコードする遺伝子。
M 1.3	47	免疫レセプター, DCIR, B細胞, IgG	巨細胞: 巨細胞表面マーカー (CD72, CD79A, B, CD19, CD22) および他の巨細胞関連分子。初期巨細胞因子 (EPC), 巨細胞シグナル (ELNK) および (B) リンパ系チロシンキナーゼ (BLK) をコードする遺伝子。
M 1.4	87	複製、印刷、修復、CREB1, リンパ系, TNF-α, アルファ	cAMPシグナル伝達経路のレギュレーターおよび標的 JUNB, ATF4, CREM, PDE4, NR3A2, MYL9, ならびに TNF-α 誘発型 NF-κB 活性化の調節因子 (CYLD, ASK, TNFAIP3) をコードする遺伝子。
M 1.5	130	樹皮、樹皮突起の, MHC, 即時刺激, TLR4, MYD88	免疫細胞: 免疫細胞により発現される分子 (CD88, CD133, FCGR2A), (その中には抗原提示の役割に關与するものもある) CD14, IL18, MYD88 の調節因子。発現される分子。この群は TNFファミリーメンバー (TNFR2, HADFD) を含む。
M 1.6	26	亜群、フィンガー、Hsp3, RAS	シグナル伝達分子、活性化 STAT (STAT) および PI3K/Akt の亜群フィンガー含有インヒビター、または活性化 T 細胞の核因子 NFATC3 をコードする遺伝子。
M 1.7	127	リボソーム、翻訳、40S, 60S, HLA	HMIC/リボソームタンパク質/MHCクラスI分子 (HLA-A/B/C/D/E/F) および MHCクラスII (DM, DMAP, DMAP2) またはリボソームタンパク質 (RPL4, RPS3) をコードする遺伝子。
M 1.8	86	代謝、生合成、複製、ヘリカーゼ	代謝酵素 (GLS, NSF1, RAL135) および DNA 複製に關与する因子 (PURA, TERF2, EIF2S1) をコードする遺伝子。
M 2.1	72	NK 細胞、細胞溶解, GDB, 細胞融合, T細胞, CTL, IFN-γ	細胞傷害性 T 細胞: 細胞傷害性 T 細胞および NK 細胞表面マーカー (CD8A, CD8, CD103, NKD2, KLRF1), 細胞溶解分子 (グランザイム、パーフォリン、グザム), ケモカイン (CCL5, XCL1) および CTL/NK-細胞関連分子 (GZMA) をコードする遺伝子。
M 2.2	44	顆粒球、好中球、好中球、好中球、好中球、好中球、好中球、好中球	好中球: 好中球顆粒に見いだされる内在性分子 (ラクトトランスフェリン, LTF, デフェンシン, DEAF1, 免疫反応性糖タンパク質 BPI, カゼリジン抗菌生物タンパク質 (CAMP)) をコードする遺伝子。