

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2018-200322

(P2018-200322A)

(43) 公開日 平成30年12月20日(2018.12.20)

(51) Int.Cl.		F I			テーマコード(参考)
<b>GO 1 N 33/574 (2006.01)</b>		GO 1 N 33/574		A	2 G O 4 5
<b>GO 1 N 33/68 (2006.01)</b>		GO 1 N 33/68			4 B O 2 9
<b>GO 1 N 33/53 (2006.01)</b>		GO 1 N 33/53		D	
<b>C 1 2 M 1/34 (2006.01)</b>		C 1 2 M 1/34			

審査請求 有 請求項の数 9 O L 外国語出願 (全 86 頁)

(21) 出願番号 特願2018-153283 (P2018-153283)  
 (22) 出願日 平成30年8月16日(2018.8.16)  
 (62) 分割の表示 特願2014-508183 (P2014-508183)の分割  
 原出願日 平成24年4月30日(2012.4.30)  
 (31) 優先権主張番号 61/480,802  
 (32) 優先日 平成23年4月29日(2011.4.29)  
 (33) 優先権主張国 米国(US)  
 (31) 優先権主張番号 61/619,279  
 (32) 優先日 平成24年4月2日(2012.4.2)  
 (33) 優先権主張国 米国(US)

(特許庁注:以下のものは登録商標)

1. L u m i n e x

(71) 出願人 510067957  
 キャンサー・ブリヴェンション・アンド・  
 キュア, リミテッド  
 アメリカ合衆国インディアナ州46360  
 , ミシガン・シティ, ワシントン・ストリ  
 ート 105  
 (74) 代理人 100078282  
 弁理士 山本 秀策  
 (74) 代理人 100113413  
 弁理士 森下 夏樹  
 (74) 代理人 100181674  
 弁理士 飯田 貴敏  
 (74) 代理人 100181641  
 弁理士 石川 大輔

最終頁に続く

(54) 【発明の名称】 分類システムおよびそのキットを使用した肺疾患の同定および診断方法

(57) 【要約】

【課題】分類システムおよびそのキットを使用した肺疾患の同定および診断方法を提供すること。

【解決手段】本発明は、肺疾患、例えば非小細胞肺癌または反応性気道疾患の診断に有用なバイオマーカーおよびバイオマーカーの組み合わせを提供する。これらバイオマーカーの測定値は、個人が肺疾患を患っていることに関する見込みの決定を補助するための分類システム、例えばサポートベクトルマシンまたはアダプストに入力されている。これらのバイオマーカーおよびバイオマーカーの組み合わせを検出するための薬剤を含むキット、ならびに肺疾患の診断を補助するシステムもまた提供する。

【選択図】図1A

図1A

正常集団での発光強度レベル					
バイオマーカー	平均	標準偏差	バイオマーカー	平均	標準偏差
アミノネカタン	2259.55	1504.98	IL-17	37.60	13.52
ヒスタミン	124.37	90.22	IL-1e	92.36	385.32
PAI-1 <sup>+</sup>	335.10	371.63	IFN-γ	23.30	9.28
S E-ケラチタン	66.34	47.58	G-CSF	23.84	4.12
sVCAM-1	1753.31	513.04	GM-CSF	28.60	12.57
sICAM-1	2955.75	1574.35	TNF-α	43.52	15.71
MPO	592.75	1503.40	MCP-1	46.39	29.47
CRP	8384.00	6292.94	IL-12 (p40)	24.93	7.26
SAA	2202.71	4017.36	MIP-1α	41.04	27.61
SAP	2300.71	570.75	MIP-1β	21.28	9.86
レクチン <sup>+</sup>	2204.67	2771.57	VEGF	40.80	11.45
GLP-1	56.21	102.46	IL12 (p70)	7.56	36.69
アミリン (α)	142.47	339.44	IL-13	11.41	57.04
C-ペプチド	4748.58	1971.09	MMP-2	155.73	439.49
インスリン	278.12	424.50	MMP-1	103.52	179.41
Sfag	58.81	148.38	MMP-3	668.81	2442.77
sPSI	15.13	16.89	エオシノキリン	195.78	208.16
MIF	52.08	112.39	レプティジン	3767.80	3885.08
IL-1β	23.71	9.52	IP-10	543.43	820.15
IL-2	12.26	44.02	MMP-9	1090.51	1599.16
IL-1m	23.11	19.55	MMP-13	0.35	35.67
IL-4	114.59	72.53	PAI-1 <sup>+</sup>	33.94	98.33
IL-5	18.37	18.35	I-TAC	20.16	154.83
IL-6	40.99	56.56	MMP-12	12.41	27.69
IL-7	15.82	6.61	HGF	206.22	619.73
TGF-β	31.46	8.00	MMP-7	1004.07	577.82
アミンオキシダーゼ	13.35	3.55	HGF	30.91	49.02
IL-8	239.47	629.22	± CD40 リガンド	153.50	434.04
IL-10	13.06	5.63	MMP-8	70.26	318.85
IL-15	24.53	4.27			

## 【特許請求の範囲】

## 【請求項 1】

対象における非小細胞肺癌の有無の見込みの判定を補助する方法であって、以下の工程

：  
(a) 該対象由来の生理学的試料中の複数のバイオマーカーの発現レベルを測定する工程、であって、

ここで、該バイオマーカーはペプチド、タンパク質、翻訳後修飾を受けたペプチド、翻訳後修飾を受けたタンパク質、またはそれらの組み合わせであり；ここで該複数のバイオマーカーは、肝細胞増殖因子（「HGF」）、レプチン（「LEP」）、IL-4、IL-5、IL-7、IL-8、IL-10、IL-12（p70）、単球走化性タンパク質1（「MCP-1」）、レジスチン（「RETN」；「xcp1」）、マクロファージ遊走性阻止因子（「MIF」）、ミエロペルオキシダーゼ（「MPO」）、血清アミロイドA（「SAA」；「SAA1」）、白血病抑制因子（「LIF」）、インターフェロン誘導性モノカイン（「MIG」）、IL-9、IL-16、血小板由来成長因子サブユニットBホモ二量体（「PDGFB B」）、T細胞由来の好酸球走化性物質（「RANTES」）、可溶性Fasリガンド（「sFasL」）、TNFR-I I、TNF受容体1（「TNFR-I」）、可溶性リガンド（「CD40リガンド」）、可溶性細胞間接着分子（「sICAM」）、アポリポタンパク質（「Apo」）A1、ApoA2、ApoB、ApoC2、ApoE、CD40、Dダイマー、第VII因子、第VII因子、第X因子、プロテインC、組織プラスミノゲン活性化因子（「TPA」）、脳由来神経栄養因子（「BDNF」）、Bリンパ球化学誘引物質（「BLC」）、ケモカイン（C-X-Cモチーフ）リガンド1（「GRO-1」）、皮膚T細胞誘引ケモカイン（「CTACK」）、エオタキシン-2、エオタキシン-3、グランザイム-B、I-TAC（「CXCL11」；「ケモカイン（C-X-Cモチーフ）リガンド11」、インターフェロン誘導T細胞化学誘引物質、単球特異的ケモカイン3（「MMP-3」）、マクロファージコロニー刺激因子（「MCSF」）、マクロファージ炎症性タンパク質-3（「MIP-3」）、マトリックスメタロプロテアーゼ（「MMP」）1、MMP2、MMP3、MMP7、MMP8、MMP9、MMP12、MMP13、CD40、神経成長因子（「NGF」）、上皮増殖因子（「EGF」）、エオタキシン（「CCL11」）、フラクタルキン、塩基性線維芽細胞増殖因子（「FGF-塩基性」）、顆粒球コロニー刺激因子（「G-CSF」）、顆粒球マクロファージコロニー刺激因子（「GM-CSF」）、インターフェロン（「IFN」）、IFN-、IFN-2、IFN-、インターロイキン（「IL」）1a、IL-1、IL-1ra、IL-2、IL-2ra、IL-3、IL-6、IL-12（p40）、IL-13、IL-15、IL-17、IL-17a、IL-17F、IL-20、IL-21、IL-22、IL-23（p19）、IL-27、IL-31、IP-10、マクロファージ炎症性タンパク質（「MIP」）1a、MIP-1、好中球活性化ペプチド78（「ENA-78」）、オステオプロテゲリン（「OPG」）、胎盤増殖因子（「PIGF」）、幹細胞成長因子（「SCGF」）、ストロマ細胞由来因子1（「SDF-1」）、NF-B活性化可溶性受容体リガンド（「sRANKL」）、サバイビン、トランスフォーミング増殖因子a（「TGFA」）、TGFA、腫瘍壊死因子a（「TNFa」）、TNF-、TNF受容体1（「TNFR-I」）、TNFR-I I、TNF関連性アポトーシス誘導性リガンド（「TRAIL」）、トロンボポエチン（「TPO」）、血管内皮成長因子（「VEGF」）、インスリン（「Ins」）、C-ペプチド、グルカゴン様タンパク質-1/アミリン（「GLP-1/アミリン」）、アミリン（全）、グルカゴン、アディポネクチン、プラスミノゲン活性化因子阻害因子1（「PAI-1」；「セルピン」）（活性型/全）、sFas、sE-セレクトリン、可溶性血管細胞接着分子（「sVCAM」）、C反応性タンパク質（「CRP」）、および、血清アミロイドP（「SAP」）からなる群から選択アルゴリズムによって選択され；

該生理学的試料は、全血、血漿、血液血清、またはそれらの組み合わせである、工程；

10

20

30

40

50

ならびに

(b) 分類システムを利用し、該バイオマーカー測定値に基づいて該試料を分類する工程であって、ここで該分類システムは、ランダムフォレストまたはアダブースト、あるいはランダムフォレストおよびアダブーストの集合から選択される機械学習システムであり、ここで該試料の該分類は、該対象における非小細胞肺癌の存在または発生の指標である、工程を包含する、方法。

【請求項2】

対象における喘息の有無の見込みの判定を補助する方法であって、以下の工程：

(a) 該対象由来の生理学的試料中の複数のバイオマーカーの発現レベルを測定する工程、であって、

ここで、該バイオマーカーはペプチド、タンパク質、翻訳後修飾を受けたペプチド、翻訳後修飾を受けたタンパク質、またはそれらの組み合わせであり；ここで該複数のバイオマーカーは、肝細胞増殖因子(「HGF」)、レプチン(「LEP」)、IL-4、IL-5、IL-7、IL-8、IL-10、IL-12(p70)、単球走化性タンパク質1(「MCP-1」)、レジスチン(「RETN」；「xcp1」)、マクロファージ遊走性阻止因子(「MIF」)、ミエロペルオキシダーゼ(「MPO」)、血清アミロイドA(「SAA」；「SAA1」)、白血病抑制因子(「LIF」)、インターフェロン誘導性モノカイン(「MIG」)、IL-9、IL-16、血小板由来成長因子サブユニットBホモ二量体(「PDGFB B」)、T細胞由来の好酸球走化性物質(「RANTES」)、可溶性Fasリガンド(「sFasL」)、TNFR-I I、TNF受容体1(「TNFR-I」)、可溶性リガンド(「CD40リガンド」)、可溶性細胞間接着分子(「sICAM」)、アポリタンパク質(「Apo」)A1、ApoA2、ApoB、ApoC2、ApoE、CD40、Dダイマー、第VII因子、第VII因子、第X因子、プロテインC、組織プラスミノゲン活性化因子(「TPA」)、脳由来神経栄養因子(「BDNF」)、Bリンパ球化学誘引物質(「BLC」)、ケモカイン(C-X-Cモチーフ)リガンド1(「GRO-1」)、皮膚T細胞誘引ケモカイン(「CTACK」)、エオタキシン-2、エオタキシン-3、グランザイム-B、I-TAC(「CXCL11」；「ケモカイン(C-X-Cモチーフ)リガンド11」、インターフェロン誘導T細胞化学誘引物質、単球特異的ケモカイン3(「MMP-3」)、マクロファージコロニー刺激因子(「MCSF」)、マクロファージ炎症性タンパク質-3(「MIP-3」)、マトリックスメタロプロテアーゼ(「MMP」)1、MMP2、MMP3、MMP7、MMP8、MMP9、MMP12、MMP13、CD40、神経成長因子(「NGF」)、上皮増殖因子(「EGF」)、エオタキシン(「CXCL11」)、フラクタルキン、塩基性線維芽細胞増殖因子(「FGF-塩基性」)、顆粒球コロニー刺激因子(「G-CSF」)、顆粒球マクロファージコロニー刺激因子(「GM-CSF」)、インターフェロン(「IFN」)、IFN-、IFN-2、IFN-、インターロイキン(「IL」)1a、IL-1、IL-1ra、IL-2、IL-2ra、IL-3、IL-6、IL-12(p40)、IL-13、IL-15、IL-17、IL-17a、IL-17F、IL-20、IL-21、IL-22、IL-23(p19)、IL-27、IL-31、IP-10、マクロファージ炎症性タンパク質(「MIP」)1a、MIP-1、好中球活性化ペプチド78(「ENA-78」)、オステオプロテゲリン(「OPG」)、胎盤増殖因子(「PIGF」)、幹細胞成長因子(「SCGF」)、ストロマ細胞由来因子1(「SDF-1」)、NF-B活性化可溶性受容体リガンド(「sRANKL」)、サバイピン、トランスフォーミング増殖因子a(「TGFA」)、TGFA、腫瘍壊死因子a(「TNFa」)、TNF-、TNF受容体1(「TNFR-I」)、TNFR-I I、TNF関連性アポトーシス誘導性リガンド(「TRAIL」)、トロンボポエチン(「TPO」)、血管内皮成長因子(「VEGF」)、インスリン(「Ins」)、C-ペプチド、グルカゴン様タンパク質-1/アミリン(「GLP-1/アミリン」)、アミリン(全)、グルカゴン、アディポネクチン、プラスミノゲ

10

20

30

40

50

ン活性化因子阻害因子1(「PAI-1」;「セルピン」)(活性型/全)、sFas、sE-セレクチン、可溶性血管細胞接着分子(「sVCAM」)、C反応性タンパク質(「CRP」)、および、血清アミロイドP(「SAP」)からなる群から選択アルゴリズムによって選択され;

該生理学的試料は、全血、血漿、血液血清、またはそれらの組み合わせである、工程;ならびに

(b)分類システムを利用し、該バイオマーカー測定値に基づいて該試料を分類する工程であって、ここで該分類システムは、ランダムフォレストまたはアダブースト、あるいはランダムフォレストおよびアダブーストの集合から選択される機械学習システムであり、ここで該試料の該分類は、該対象における喘息の指標である、工程を包含する、方法。

10

【請求項3】

対象における肺疾患の有無の見込みの判定を補助する方法であって、以下の工程:

(a)該対象由来の生理学的試料中において、反応性気道疾患の兆候と非小細胞肺癌の区別を補助する複数のバイオマーカー、反応性気道疾患の指標である複数のバイオマーカー、および、非小細胞肺癌の指標である複数のバイオマーカーの発現レベルを測定する工程、であって、

ここで、該バイオマーカーはペプチド、タンパク質、翻訳後修飾を受けたペプチド、翻訳後修飾を受けたタンパク質、またはそれらの組み合わせであり;ここで該複数のバイオマーカーは、肝細胞増殖因子(「HGF」)、レプチン(「LEP」)、IL-4、IL-5、IL-7、IL-8、IL-10、IL-12(p70)、単球走化性タンパク質1(「MCP-1」)、レジスチン(「RETN」;「xcp1」)、マクロファージ遊走性阻止因子(「MIF」)、ミエロペルオキシダーゼ(「MPO」)、血清アミロイドA(「SAA」;「SAA1」)、白血病抑制因子(「LIF」)、インターフェロン誘導性モノカイン(「MIG」)、IL-9、IL-16、血小板由来成長因子サブユニットBホモ二量体(「PDGFB B」)、T細胞由来の好酸球走化性物質(「RANTES」)、可溶性Fasリガンド(「sFasL」)、TNFR-I I、TNF受容体1(「TNFR-I」)、可溶性リガンド(「CD40リガンド」)、可溶性細胞間接着分子(「sICAM」)、アポリポタンパク質(「Apo」)A1、ApoA2、ApoB、ApoC2、ApoE、CD40、Dダイマー、第V I I因子、第V I I I因子、第X因子、プロテインC、組織プラスミノゲン活性化因子(「TPA」)、脳由来神経栄養因子(「BDNF」)、Bリンパ球化学誘引物質(「BLC」)、ケモカイン(C-X-Cモチーフ)リガンド1(「GRO-1」)、皮膚T細胞誘引ケモカイン(「CTACK」)、エオタキシン-2、エオタキシン-3、グランザイム-B、I-TAC(「CXCL11」;「ケモカイン(C-X-Cモチーフ)リガンド11」、インターフェロン誘導T細胞化学誘引物質、単球特異的ケモカイン3(「MMP-3」)、マクロファージコロニー刺激因子(「MCSF」)、マクロファージ炎症性タンパク質-3(「MIP-3」)、マトリックスメタロプロテアーゼ(「MMP」)1、MMP2、MMP3、MMP7、MMP8、MMP9、MMP12、MMP13、CD40、神経成長因子(「NGF」)、上皮増殖因子(「EGF」)、エオタキシン(「CCL11」)、フラクタルキン、塩基性線維芽細胞増殖因子(「FGF-塩基性」)、顆粒球コロニー刺激因子(「G-CSF」)、顆粒球マクロファージコロニー刺激因子(「GM-CSF」)、インターフェロン(「IFN」)、IFN-、IFN-2、IFN-、インターロイキン(「IL」)1a、IL-1、IL-1ra、IL-2、IL-2ra、IL-3、IL-6、IL-12(p40)、IL-13、IL-15、IL-17、IL-17a、IL-17F、IL-20、IL-21、IL-22、IL-23(p19)、IL-27、IL-31、IP-10、マクロファージ炎症性タンパク質(「MIP」)1a、MIP-1、好中球活性化ペプチド78(「ENA-78」)、オステオプロテゲリン(「OPG」)、胎盤増殖因子(「PIGF」)、幹細胞成長因子(「SCGF」)、ストロマ細胞由来因子1(「SDF-1」)、NF-B活性化可溶性受容体リガ

20

30

40

50

ド(「sRANKL」)、サバイビン、トランスフォーミング増殖因子α(「TGFα」)、TGF-β、腫瘍壊死因子α(「TNFα」)、TNF-β、TNF受容体1(「TNFR-I」)、TNFR-II、TNF関連性アポトーシス誘導性リガンド(「TRAIL」)、トロンボポエチン(「TPO」)、血管内皮成長因子(「VEGF」)、インスリン(「Ins」)、C-ペプチド、グルカゴン様タンパク質-1/アミリン(「GLP-1/アミリン」)、アミリン(全)、グルカゴン、アディポネクチン、プラスミノゲン活性化因子阻害因子1(「PAI-1」;「セルピン」)(活性型/全)、sFas、sE-セレクチン、可溶性血管細胞接着分子(「sVCAM」)、C反応性タンパク質(「CRP」)、および、血清アミロイドP(「SAP」)からなる群から選択アルゴリズムによって選択され;

10

該生理学的試料は、全血、血漿、血液血清、またはそれらの組み合わせであり;

該複数のバイオマーカーは同一ではない、工程;ならびに

(b)分類システムを利用し、該バイオマーカー測定値に基づいて該試料を分類する工程であって、ここで該分類システムは、ランダムフォレストまたはアダブースト、あるいはランダムフォレストおよびアダブーストの集合から選択される機械学習システムであり、ここで該試料の該分類は、

- (i) 該対象における喘息と非小細胞肺癌;
  - (ii) 該対象における喘息の有無;および
  - (iii) 該対象における非小細胞肺癌の有無;
- の兆候を区別することを補助する、工程;

20

(c) 該分類において見いだされた状態に基づいて、該対象が、喘息を有するか、非小細胞肺癌を有するか、または、疾患を有さないかを判定する、工程を包含する、方法。

【請求項4】

前記機械学習システムは、ランダムフォレストおよびアダブーストの集合である、請求項1~3のいずれか一項に記載の方法。

【請求項5】

前記機械学習システムは、ランダムフォレストである、請求項1~3のいずれか一項に記載の方法。

【請求項6】

前記機械学習システムは、アダブーストである、請求項1~3のいずれか一項に記載の方法。

30

【請求項7】

前記複数のバイオマーカーは、肝細胞増殖因子(「HGF」)、レプチン(「LEP」)、IL-4、IL-5、IL-7、IL-8、IL-10、IL-12(p70)、単球走化性タンパク質1(「MCP-1」)、レジスチン(「RETN」;「xcp1」)、マクロファージ遊走性阻止因子(「MIF」)、ミエロペルオキシダーゼ(「MPO」)、血清アミロイドA(「SAA」;「SAA1」)、白血病抑制因子(「LIF」)、インターフェロン誘導性モノカイン(「MIG」)、IL-9、IL-16、血小板由来成長因子サブユニットBホモ二量体(「PDGFBβ」)、T細胞由来の好酸球走化性物質(「RANTES」)、可溶性Fasリガンド(「sFasL」)、TNFR-II、TNF受容体1(「TNFR-I」)、可溶性リガンド(「CD40リガンド」)、および可溶性細胞間接着分子(「sICAM」)からなる群から選択される、請求項1~3のいずれか一項に記載の方法。

40

【請求項8】

前記複数の、少なくとも4種類、少なくとも6種類、少なくとも10種類または少なくとも18種類のバイオマーカーを含む、請求項1~3のいずれか一項に記載の方法。

【請求項9】

バイオマーカーが免疫アッセイにより測定される、請求項1~3のいずれか1項に記載の方法。

50

## 【発明の詳細な説明】

## 【技術分野】

## 【0001】

## 発明の背景

## (a) 技術分野

〔001〕 本発明は、バイオマーカーおよびそれらのキット、ならびにそれらのバイオマーカーに基づいて、疾患の有無の見込みの決定を補助するシステムを使用した、肺疾患の検出、同定、および診断に関する。より具体的には、本発明は、特定のバイオマーカーの発現レベルを測定し、これらの測定値を、サポートベクトルマシンなどの分類システムに入力することによる、非小細胞肺癌および反応性気道疾患の診断に関する。

10

## 【背景技術】

## 【0002】

## (b) 関連する文献の記載

## ヒト肺組織の病理

〔002〕 何百万人もの米国人が、喘息や肺癌などの呼吸器系の症状に罹患している。実際のところ、米国肺協会（登録商標）は、2000万人もの米国人が喘息にかかっていると報告している。米国がん協会は、2007年の1年間だけでも、呼吸器系の癌の新規発症例が229,400件で、呼吸器系の癌による死亡例が164,840件だったと推定した。まだ癌が限局性である時に癌が検出されていれば、全癌例の5年生存率は46%であるが、肺癌患者の5年生存率は僅か13%である。それに対応して、肺癌は16%しか疾患が広がる前には発見されない。肺癌は癌細胞の病理に基づいて、通常、主に2種類に分類されている。それぞれの種類には、癌性に転換した細胞型に応じた名前が付けられている。小細胞肺癌はヒト胚組織の小細胞に由来し、非小細胞肺癌は通常、小細胞型でない肺癌全てを含む。一般的に、全ての非小細胞型に対する治療が同じであることから、非小細胞肺癌とまとめられている。また、非小細胞肺癌、すなわちNSCLCは肺癌全体の約75%を占めている。

20

## 【0003】

〔003〕 肺癌患者の生存率が低いことは主に、肺癌を早期に診断することが難しいという事実による。肺癌を診断する、つまりあるヒトに肺癌が存在するかを同定するための現行の方法は、X線撮影法、コンピューター断層（CT）撮影法、および腫瘍の有無を物理的に確認するための同様の肺試験に限られている。そのため、肺癌の診断は、症状への反応だけによって行われることが多く、これは、疾患がヒトに存在してから十分な時間存在していたこと、および物理的に検出可能な量を生成するのに十分な時間が経過したことの証拠である。

30

## 【0004】

〔004〕 同様に、喘息を検出するための現行の方法も、通常、再発性の喘鳴、咳、および胸苦しさなどの症状が顕在化してしばらく経ってから行われる。現在行われている喘息の検出方法は、一般的に、肺活量測定または負荷試験などの肺機能試験に限られている。さらにこれらの試験は、他の病理つまり反応性気道疾患、例えば慢性閉塞性肺疾患（COPD）、気管支炎、肺炎、およびうっ血性心不全を除くために、他の多くの試験と一緒に行われるよう、医師によって指示されることが多い。

40

## 【0005】

## 分類システム

〔005〕 パターンを認識し、関係のないデータにすぎない他の情報をも含む大量のデータベースから重要な情報を抽出することを可能にするために、機械学習法のような、データ解析およびデータマイニングのための様々な分類システムが広く探索されている。学習機械は、汎化するために既知の分類を使用して学習させることが可能なアルゴリズムを含むものである。その後、訓練した学習機械アルゴリズムを結果が分かっていない事例に応用して、結果を予測することが、すなわち、学習したパターンに応じてデータを分類することができる。ニューラルネットワーク、隠れマルコフモデル、信念ネットワークおよ

50

びカーネル法に基づく分類器、例えばサポートベクトルマシンを含む機械学習法は、データが大量であること、パターンにノイズが多いことおよび一般理論がないことを特徴とする問題に有用である。

【0006】

〔006〕 問題のパターン分類、回帰およびクラスタリングの成功例の多くは、パターンのペアの類似性を決定するためのカーネルに基づいている。これらのカーネルは通常、実数のベクトルとして表すことが可能なパターン用に定義される。例えば、線形カーネル、ラジアル基(radial basis)カーネルおよび多項カーネルは全て、実際のベクトルのペアの類似性を測定するものである。このようなカーネルは、データが実数の列として、この方法において最もよく表される場合に適している。カーネルの選択は、特徴空間におけるデータ表現の選択と対応している。使用法の多くで、パターンは様々な階層を有している。これらの階層を利用して、学習アルゴリズムの性能を改善することができる。機械学習法で一般的に生じる階層データの種類の例としては、文字列、文書、木、グラフ、例えばウェブサイトまたは化学分子、シグナル、例えばマイクロアレイ発現プロファイル、スペクトル、画像、時空的データ、関係型データおよび生化学的濃度などがある。

10

【0007】

〔007〕 分類システムは医薬分野で使用されてきた。例えば、病状の発生を診断および予測する方法が、様々なコンピューターシステムや分類システム、例えばサポートベクトルマシンを利用して提唱されてきた。例えば、米国特許第7,321,881号；同第7,467,119号；同第7,505,948号；同第7,617,163号；同第7,676,442号；同第7,702,598号；同第7,707,134号；および同第7,747,547号を参照のこと。これらはここで参照することによりその全体が本明細書に組み込まれる。しかしながらこれらの方法の精度は、ヒトの胚組織での病理、例えば非小細胞肺癌および/または反応性気道疾患の診断および/または予測では、高いレベルではない。

20

【0008】

〔008〕 このように当該分野では、ヒトの胚組織の病理を、特に発症の初期段階で診断するための単純で信頼のおける方法がない。さらに、特定の肺組織の病理の有無を示すことができる血液検査も今のところない。そのため、疾患の進行の初期に肺癌の存在を確認するための方法の開発が望まれている。同様に、臨床上明らかな症状が現れるよりも前に、喘息および非小細胞肺癌を診断するための方法、ならびに喘息や非小細胞肺癌と他の肺疾患、例えば感染を区別するための方法の開発も望まれている。

30

【発明の概要】

【課題を解決するための手段】

【0009】

発明の好ましい態様の概要

〔009〕 本発明は、様々な分類システム、例えばサポートベクトルマシンを利用して、対象における特定のバイオマーカーを評価するための強固な方法を提供することで、これらの需要に対処する。

【0010】

〔010〕 本発明は、対象を生理学的に特徴付ける方法を提供し、この方法は、最初に対象の生理学的試料を収集する工程；次いでその試料での複数のバイオマーカーのバイオマーカー測定値を決定する工程；および最後に、分類システムを利用し、バイオマーカー測定値に基づいて、試料を分類する工程を含み、ここでこの試料の分類は、この対象の生理的な状況もしくは状態、または疾患状況の変化と関係がある。通常、この分類システムは機械学習システムであり、好ましくはカーネルまたは分類および分類システムに基づく回帰木であり、より好ましくはサポートベクトルマシン(SVM)またはアダブーストである。

40

【0011】

〔011〕 一態様では、この生理学的に特徴付ける方法は、対象における非小細胞肺

50

癌の有無を示す診断、または非小細胞肺癌の発生段階を示す診断を提供する。別の態様では、この生理学的に特徴付ける方法は、対象における反応性気道疾患、例えば喘息または閉塞性肺疾患を示す診断を提供する。さらに別の態様では、この生理学的に特徴付ける方法は、対象における肺疾患を示す診断を提供し、ここで複数のバイオマーカ-には、前記試料中の反応性気道疾患と非小細胞肺癌を示すもの、反応性気道疾患を示すものとなる複数のバイオマーカ-、および非小細胞肺癌を示すものとなる複数のバイオマーカ-の区別を補助するマーカ-が含まれ、この複数バイオマーカ-は同一のものではなく、かつ、試料は、試料を3通りに分類するシステム、すなわち、対象の(i)反応性気道疾患と非小細胞肺癌を示すもの；(ii)反応性気道疾患の有無を示すもの；および(iii)非小細胞肺癌の有無を示すものの区別を補助するシステムを使用して、バイオマーカ-測定値に基づいて分類され、その結果、対象は、どの条件が3つの分類中の2つに見られるかに応じて、(1)反応性気道疾患であるか、(2)非小細胞肺癌であるか、または(3)疾患に罹患していないと決定される。

10

20

30

40

50

**【0012】**

〔012〕 本発明はまた、バイオマーカ-のセットそれぞれの複数のバイオマーカ-測定値を含む試験データを分類する方法を提供し、この方法は、哺乳類の試験対象におけるバイオマーカ-のセットに関する複数のバイオマーカ-測定値を含む試験データを受信する工程；次いで、電子的に記憶させた学習データベクトルのセットで訓練したサポートベクトルマシンの電子表示を使って試験データを評価する工程、ここで各学習データベクトルは個々の哺乳動物を表し、かつ、対応する哺乳動物に関するバイオマーカ-のセットの各バイオマーカ-のバイオマーカ-測定値を含み、さらに各学習データベクトルは対応する哺乳動物の病状に関する分類を含み、最後に、評価する工程に基づいて、哺乳類の試験対象の分類を出力する工程、を含む。好ましくは、哺乳類の試験対象はヒトである。別の形態では、評価する工程は、電子的に記憶されている学習データベクトルのセットにアクセスすることを含む。

**【0013】**

〔013〕 別の形態において本発明は、バイオマーカ-のセットそれぞれの複数のバイオマーカ-測定値を含む試験データを分類するために、サポートベクトルマシンを訓練してモデルを生成する方法を提供し、この方法は、電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関するバイオマーカ-のセットの各バイオマーカ-のバイオマーカ-測定値を含み、各学習データベクトルはさらに対応するヒトの病状に関する分類を含み、および、電子的に記憶されている学習データベクトルのセットを使用して、サポートベクトルマシンの電子表示を訓練する工程、を含む。続いて、本発明は、ヒト試験対象におけるバイオマーカ-のセットに関する複数のバイオマーカ-測定値を含む試験データを受信する工程、および訓練したサポートベクトルマシンの電子表示（すなわち、訓練したサポートベクトルマシンによって生成されたモデル）を利用して試験データを評価する工程；および、最後に、評価する工程に基づいたヒト試験対象の分類を出力する工程、を提供する。

**【0014】**

〔014〕 さらに別の態様では、本発明は、バイオマーカ-のセットそれぞれの複数のバイオマーカ-測定値を含む試験データを分類する方法を提供する。この方法は、ヒト試験対象に関する試験データを受信する工程を含み、ここで試験データは、バイオマーカ-のセットの少なくとも各バイオマーカ-のバイオマーカ-測定値を含んでいる。この方法はまた、電子的に記憶されている、学習データベクトルの第一のセットを使用して訓練したサポートベクトルマシンの電子表示を使用して試験データを評価する工程を含み、ここで各学習データベクトルの第一のセットの学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関するバイオマーカ-のセットの少なくとも各バイオマーカ-のバイオマーカ-測定値を含んでいる。学習データベクトルの第一のセットの各学習データベクトルはまた、対応するヒトの病状に関する分類を含む。この方法はさらに、評価する工程

に基づいたヒト試験対象の分類を出力する工程を含む。この態様では、バイオマーカのセットに含まれているそれぞれのバイオマーカは、(A)順序付けられたバイオマーカの初期セグメントは、第二セットの学習データベクトルの補正済み分類のパーセンテージに関しては、順序付けられたバイオマーカの他の初期セグメントの中の極大であり、およびここで第二セットの学習データベクトルの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトのバイオマーカのセットの少なくとも各バイオマーカのバイオマーカ測定値を含み、第二セットの学習データベクトルの各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでいる、それぞれのバイオマーカに関する2群の濃度測定値のマージン分布の中心傾向の関数によって最大から最小に順序付けられたバイオマーカの該初期セグメントの中にあるか、または(B)(A)で同定したバイオマーカの該初期セグメント中に含まれているバイオマーカに関する一次相互作用器であるかのいずれかである。

10

20

30

40

50

**【0015】**

〔015〕 通常、本発明の方法は、病状の有無、ここで病状は好ましくは肺疾患であり、より好ましくは非小細胞肺癌または反応性気道疾患、例えば喘息のいずれかである、に関して試験対象を分類する。バイオマーカ測定値は、実施例で記載のバイオマーカから選択された、少なくとも1つのタンパク質の血漿濃度測定値を含んでいてもよい。好ましくは、バイオマーカ測定値は、少なくとも4種類の別個のバイオマーカの血漿濃度を含むか、あるいはバイオマーカ測定値は、少なくとも6種類の別個のバイオマーカ、またはさらに少なくとも10種類の別個のバイオマーカ、または少なくとも18種類の別個のバイオマーカの血漿濃度を含み得る。学習ベクトルのセットは、少なくとも30のベクトル、50のベクトル、またはさらに100のベクトルを含み得る。一形態では、分類器は、線形カーネル、ラジアル基底カーネル、多項カーネル、統一(uniform)カーネル、三角カーネル、Epanechnikovカーネル、四次多項(quartic)(二乗重み)カーネル、三次三乗(tricube)(三乗重み)カーネル、およびコサインカーネルから選択される1つまたは複数のカーネル関数を含むサポートベクトルマシンである。別の形態では、分類器は、アダプーストを使用して、例えば、ID3またはC4.5に基づく最初の分類器から発展させる。

**【0016】**

〔016〕 本発明はまた、バイオマーカのセットそれぞれの、複数のバイオマーカ測定値を含む試験データを分類するためのシステムを提供する。ここでこのシステムはコンピューターを含み、コンピューターは電子的に記憶されている学習データベクトルのセットを使って訓練することが可能なサポートベクトルマシンの電子表示を含んでおり、各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関するバイオマーカのセットの、各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり、電子的に記憶されている学習データベクトルのセットはコンピューターに操作可能に接続されている。コンピューターはまた、ヒト試験対象におけるバイオマーカのセットに関する複数のバイオマーカ測定値を含む試験データを受け取るように構成されており、コンピューターはさらに、訓練後のサポートベクトルマシンの電子表示を使って試験データを評価するように、かつ、この評価に基づいたヒト試験対象の分類を出力するように構成されている。

**【0017】**

〔017〕 別の態様において本発明は、バイオマーカのセットそれぞれの、バイオマーカ測定値を含む試験データを分類するシステムを提供する。ここでこのシステムはコンピューターを含み、コンピューターは順に、試験対象の病状に関して試験データを分類ために訓練したサポートベクトルマシンの電子表示を含み、訓練は電子的に記憶されている学習データベクトルのセットに基づくものであり、各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関するバイオマーカのセットの各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでいる。コンピューターは、ヒト試験対象におけるバイオマーカのセット

に関する複数のバイオマーカー測定値を含む試験データを受け取るように構成されており、コンピューターはさらに、訓練したサポートベクトルマシンの電子表示を使って試験データを評価するように、かつ、この評価に基づいたヒト試験対象の分類を出力するように構成されている。

【0018】

〔018〕 いずれの態様においても、本発明のこのシステムは、上述した方法のいずれを実施するのにも適している。特定の形態においては、システムのいずれの態様のコンピューターもさらに、(a) バイオマーカーの上位セットに含まれている各バイオマーカーに関して、それぞれのバイオマーカーに関する2群の濃度測定値のマージン分布間の距離を計算し、それによって複数の距離を生成し；(b) その距離に従って、バイオマーカーの上位セットに含まれているバイオマーカーを順序付け、それによって順序付けられたバイオマーカーのセットを生成し；(c) 順序付けられたバイオマーカーのセットの、複数の初期セグメントのそれぞれに関して、学習データに基づいてモデルフィットの指標を計算し；(d) 順序付けられたバイオマーカーのセットの好ましい初期セグメントが選択されるように、順序付けられたバイオマーカーのセットの初期セグメントを、モデルフィットの最大値型指標に従って選択し；(e) バイオマーカーの空集合から開始し、バイオマーカーの下位セットを生成するために、順序付けられたバイオマーカーのセットの、好ましい初期セグメントからさらなるバイオマーカーを再帰的に追加し、ここで各さらなるバイオマーカーは、(1) 好ましい初期セグメントに含まれている残りのバイオマーカーの中でも、それを追加することで、モデルフィットが最大に改良される場合、および(2) それを追加することで、モデルフィットが少なくとも予め設定した閾値までは改良される場合に、既存のバイオマーカーの下位セットに追加され；および(f) それ以上バイオマーカーを追加しても、予め設定しておいた閾値より、そのモデルフィット指標を超えるモデルフィットの指標が得られない場合には、既存のバイオマーカーの下位セットへのバイオマーカーの追加を停止し、それにより、バイオマーカーの下位セットを選択する、ために構成されている論理を使用して、バイオマーカーの上位セットからバイオマーカーのセットを選択するために構成されていてもよい。

10

20

【0019】

〔019〕 本明細書で提供する方法およびシステムは、一般的に90%を越える精度(例えば、感度および特異度)で、肺の病理(例えば、癌や喘息)を、診断および予測することができる。これらの結果は、非小細胞肺癌などの肺の病理を診断および予測するための、現在使用可能な方法に有意な進歩をもたらすものである。

30

【図面の簡単な説明】

【0020】

【図1A】 実施例1の正常(NO)集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図1B】 実施例1の非小細胞肺癌(LC)集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図1C】 実施例1の喘息(AST)集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

40

【図1D】 実施例1の、NO集団に対するAST集団中の、NO集団に対するLC集団中の、およびLC集団に対するAST集団中の、各バイオマーカーの蛍光強度の平均の変化(%)を示している。

【図2A】 実施例1の正常(NO)女性集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図2B】 実施例1の非小細胞肺癌(LC)女性集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図2C】 実施例1の喘息(AST)女性集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図2D】 実施例1の、NO女性集団に対するAST集団中の、NO女性集団に対するL

50

C 集団中の、および LC 女性集団に対する AST 集団中の、各バイオマーカーの蛍光強度の平均の変化 (%) を示している。

【図 3 A】実施例 1 の正常 (NO) 男性集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図 3 B】実施例 1 の非小細胞肺癌 (LC) 男性集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図 3 C】実施例 1 の喘息 (AST) 男性集団中のバイオマーカーの、蛍光強度レベルの平均、および標準偏差と相対的標準偏差を示している。

【図 3 D】実施例 1 の、NO 男性集団に対する AST 集団中の、NO 男性集団に対する LC 集団中の、および LC 男性集団に対する AST 集団中の、各バイオマーカーの蛍光強度の平均の変化 (%) を示している。

【図 4】実施例 1 の、AST 女性集団と比較した AST 男性集団中の、LC 女性集団と比較した LC 男性集団中の、および NO 女性集団と比較した NO 男性集団中の、各バイオマーカーの蛍光強度の平均の変化 (%) を示している。

【図 5】様々な分子と HGF (肝細胞増殖因子) との関係を示している。この図は ARIADN E PATHWAY STUDIO (登録商標) で作成した。

【図 6】アダプーストの ROC 曲線を示している。

【図 7】SVM の ROC 曲線を示している。

【図 8】男性に限定したアダプーストの ROC 曲線を示している。

【図 9】女性に限定したアダプーストの ROC 曲線を示している。

【図 10】アダプーストモデルに基づく選択した様々なプロットを示している。

【図 11】男性についてのアダプーストモデルに基づいた、選択した様々なプロットを示している。

【図 12】女性についてのアダプーストモデルに基づいた、選択した様々なプロットを示している。

【図 13】アダプーストモデルの精度分布を示している。

【図 14】アダプーストモデルの感度の分布を示している。

【図 15】アダプーストモデルの特異度の分布を示している。

【発明を実施するための形態】

【0021】

〔044〕本発明は、バイオマーカーを使用した、様々な肺疾患の検出、同定、および診断方法に関する。これらの方法は、特定のバイオマーカーのバイオマーカー測定値を決定する工程、および分類システム中のこれらバイオマーカー測定値を使用して、個人が非小細胞肺癌および/または反応性気道疾患 (例えば、喘息、慢性閉塞性肺疾患など) を有する見込みを決定する工程を含む。本発明はまた、肺疾患の見込みの決定を補助するためのシステムの構成要素として、これらのバイオマーカーを検出するための検出剤を含むキット、またはこれらのバイオマーカーのバイオマーカー測定値を検出するための手段を提供する。

【0022】

〔045〕非小細胞肺癌または喘息であると診断された患者集団、ならびに非小細胞肺癌および/または喘息ではないことが、医師による確認によって診断された患者集団に由来する血漿中の、59種類の選択したバイオマーカーの発現レベルを測定することによって、代表的なバイオマーカーを同定した。この方法については実施例 1 で詳細に説明する。

【0023】

定義

〔046〕本明細書で使用する場合、「バイオマーカー」または「マーカー」とは、その生体システムの生理学的な状態の特徴を示すものとして、客観的に測定することが可能な生物学的な分子である。本開示の目的では、生物学的な分子には、イオン、低分子、ペプチド、タンパク質、翻訳後修飾を受けたペプチドおよびタンパク質、ヌクレオシド、R

10

20

30

40

50

N A や D N A を含むヌクレオチドおよびポリヌクレオチド、糖タンパク質、リポタンパク質、ならびにこれらを共有的におよび非共有的に修飾した様々な分子が含まれる。生物学的な分子には、生物学的なシステムの機能にとって天然の、生物学的なシステムの機能の特徴、および/または生物学的なシステムの機能にとって必須の、これらの実体のいずれもが含まれる。バイオマーカーの多くはポリペプチドであるが、バイオマーカーはポリペプチドとして発現される遺伝子産物の翻訳前の形態を表す m R N A であってもまたは修飾された m R N A であってもよく、あるいは、それらはポリペプチドの翻訳後修飾を含んでいてもよい。

【 0 0 2 4 】

〔 0 4 7 〕 本明細書で使用する場合、「バイオマーカー測定値」とは、疾患の有無を特徴付けるのに有用なバイオマーカーに関する情報である。そのような情報は、濃度もしくは濃度の割合である測定された値を含んでいてもよく、あるいは、組織もしくは体液中のバイオマーカーの発現の量であっても、もしくは発現を量的に示すものであってもよい。各バイオマーカーは、ベクトル空間中の次元として表すことができ、ここで各ベクトルは、特定の対象と関連する複数のバイオマーカー測定値から成り立っている。

10

【 0 0 2 5 】

〔 0 4 8 〕 本明細書で使用する場合、「下位セット(サブセット)」とは適切な下位セットであり、「上位セット(スーパーセット)」とは適切な上位セットである。

〔 0 4 9 〕 本明細書で使用する場合、「対象」は、いずれの動物も意味するが、好ましくは哺乳動物であり、例えば、ヒトなどである。多くの態様において対象は、肺疾患を罹患している、または肺疾患を罹患するリスクのあるヒト患者である。

20

【 0 0 2 6 】

〔 0 5 0 〕 本明細書で使用する場合、「生理学的な試料」には、体液および組織に由来する試料が含まれる。体液には、全血、血漿、血液血清、喀痰、尿、汗、リンパ液、および肺胞洗浄が含まれる。組織試料には、固形肺組織または他の固形組織由来の生検、リンパ節生検組織、転移性病巣の生検が含まれる。生理学的な試料の収集方法は良く知られている。

【 0 0 2 7 】

〔 0 5 1 〕 本明細書で使用する場合、「検出剤」には、本明細書に記載のバイオマーカーを特異的に検出する試薬およびシステムが含まれる。検出剤には、試薬、例えば抗体、核酸プローブ、アプタマー、レクチン、または特定のマーカーもしくは目的の試料に含まれている可能性のある特定のマーカーと他のマーカーとを区別するのに有効なマーカーに特に親和性を有する他の試薬、およびシステム、例えば、上述した結合させた試薬あるいは固定した試薬を使用するセンサーなどのセンサーを含む。

30

【 0 0 2 8 】

一次相互作用器

〔 0 5 2 〕 生命を維持するのに必要な、細胞のおよび有機的な、多くの生理学的な機能を促進し、制御するために、生物学的な分子は互いに相互作用しなくてはならない。これらの相互作用は、一種のコミュニケーションと考えることができる。このコミュニケーション中では、多種多様な生物学的な分子をメッセージと見なすことが可能である。これらの分子は、シグナル伝達機能に必須の部分として、他の種類の生物学的な分子を含む、多様な標的と必然的に相互作用する。

40

【 0 0 2 9 】

〔 0 5 3 〕 相互作用している分子の一種は一般に、受容体として知られている。それら受容体は、これも相互作用する分子であるリガンドと結合する。別の型の直接的な分子間相互作用には、コファクターまたはアロステリックエフェクターと酵素との結合がある。これらの分子間相互作用は、細胞および生物の必須の生命機能の実行および制御に一緒になって機能する、シグナル伝達分子のネットワークを形成する。本発明で使用する用語では、これらの相互作用分子はそれぞれバイオマーカーである。本発明の特定のバイオマーカーは、特定のバイオマーカーのレベルと協調して、そのレベルが上昇または低下する

50

他のバイオマーカーと生理学的に関連している。これら他の関連のあるバイオマーカーは、本発明の特定のバイオマーカーについて、「一次相互作用器」と呼ばれる。

【0030】

〔054〕 「一次相互作用器」は、特定の生物学的な分子と直接相互作用する分子実体である。例えば、薬物であるモルヒネはオピエート受容体と直接相互作用し、最終的に、疼痛の感覚を低減させる。従って、オピエート受容体は、「一次相互作用器」の定義の下では、一次相互作用器である。一次相互作用器は、前記バイオマーカーが相互作用するコミュニケーション経路中で、前記バイオマーカーのすぐ上流およびすぐ下流の両方で隣り合っている分子を含む。これらの実体には、直接的な（または間接的な）調節、発現、化学反応、分子合成、結合、プロモーター結合、タンパク質修飾および分子輸送を含むがこれらには限定されない関係によって結合され得るタンパク質、核酸および低分子が包含される。それらのレベルが協調しているバイオマーカーのグループは、当業者、ならびに生理学および細胞生物学に精通した人々には良く知られている。実際のところ、特定のバイオマーカーに関する一次相互作用器は当該分野で公知であり、かつ、様々なデータベースや、利用可能なバイオインフォマティクスのソフトウェア、例えばARIADNE PATHWAY STUDIO（登録商標）、ExPASY Proteomics Server Qlucore OmicsExplorer、Protein Prospector、PQuad、ChEMBLなどを利用して見出すことができる。（例えば、ARIADNE PATHWAY STUDIO（登録商標）、Ariadne、Inc.、<www.ariadne.genomics.com>またはChEMBLデータベース、European Bioinformatics Institute、European Molecular Biology Laboratory、<www.ebi.ac.uk>を参照のこと）。

10

20

【0031】

〔055〕 一次相互作用器バイオマーカーとは、その発現レベルが別のバイオマーカーと協調しているバイオマーカーである。そのため、特定のバイオマーカーのレベルに関する情報（「バイオマーカー測定値」）を、特定のバイオマーカーと協調している一次相互作用器のレベルを測定することで導き出すことが可能である。当然のことながら、特定のバイオマーカーの挙動と一致する、定義され、かつ、再現性のある方法によって、当業者は、特定のバイオマーカーの代わりにまたは特定のバイオマーカーに加えて使用される一次相互作用器のレベルが多様になることを確認する。

【0032】

〔056〕 本発明は、特定のバイオマーカーの一次相互作用器を利用して実施される可能性のある、あるいは実施される、本明細書に記載のいかなる方法をも提供する。例えば、本発明のいくつかの態様は、HGFのバイオマーカー測定値を決定する工程を含む、生理学的に特徴付ける方法を提供する。よって、本発明また、HGFの一次相互作用器のバイオマーカー測定値を決定する工程を含む、生理学的に特徴付ける方法を提供する。HGFの一次相互作用器には、図5で同定した一次相互作用器（例えば、INS、EGF、MIF）が含まれるがこれらには限定されない。従って、本発明の想定内では、特定のバイオマーカー測定値を、特定のバイオマーカーの一次相互作用器で置き換えてもよい。

30

【0033】

バイオマーカー測定値の決定

〔057〕 通常、バイオマーカー測定値は典型的にはタンパク質またはポリペプチドである発現産物の定量的な測定に関する情報である。本発明は、バイオマーカー測定値をRNA（翻訳前）レベルでまたはタンパク質レベル（翻訳後修飾も含まれ得る）で決定することを想定している。具体的には、本発明は、転写、翻訳、翻訳後修飾のレベル、またはタンパク質分解の範囲もしくは程度の上昇または低下が反映されるバイオマーカー濃度の変化を決定することを想定しており、ここでこれらの変化は、特定の病状または疾患の進行と関連している。

40

【0034】

〔058〕 正常な対象で発現している多くのタンパク質は、疾患または状態に応じて、例えば、非小細胞肺癌または喘息を呈している対象ではその発現がより多くまたはより少なくなる。当業者であれば、疾患の多くが、複数の異なるバイオマーカーに変化を及ぼ

50

すことを理解するだろう。よって疾患は、複数のマーカーの発現パターンによって特徴付けることができる。複数のバイオマーカーの発現レベルを決定することによって発現パターンの観測が容易になり、そのようなパターンは、個々のバイオマーカーを検出するよりも感度が高く、かつ、精度の高い診断を提供する。あるパターンでは、いくつかの特定のバイオマーカーが異常に亢進しており、同時に他の特定のバイオマーカーが異常に低下していてもよい。

【0035】

〔059〕 本発明に従って、対象から、その試料におけるバイオマーカー測定値が、その試料を提供した対象におけるそのバイオマーカーの濃度に比例していることを確保する方法によって、生理学的な試料を採取する。測定された値が、試料中のバイオマーカーの濃度の比例となるように、測定を行う。これらの要件を満たす試料の採取方法および測定方法の選択は当該分野において一般的な技術の範囲内である。

10

【0036】

〔060〕 当業者には当然のことであるが、個々のバイオマーカーについては、バイオマーカー測定値を決定するための様々な方法が当該分野で知られている。Instrumental Methods of Analysis, Seventh Edition, 1988を参照のこと。それらの決定は、複合的なまたはマトリックスを使用した形態、例えばマルチプレックス免疫アッセイ (multiplexed immunoassay) によって実施され得る。

【0037】

〔061〕 バイオマーカー測定値を決定するための多数の方法が当該分野で知られている。そのような決定手段には、放射免疫アッセイ、酵素結合免疫吸着アッセイ (ELISA)、可視光または紫外光の吸光度を利用した放射検出もしくはスペクトル検出を伴う高速液体クロマトグラフィー、質量分析による定性的および定量的な解析、ウェスタンブロット、放射性プローブ、蛍光プローブ、もしくは化学発光プローブまたは核を検出する手段による定量的な可視化を伴う一次元もしくは二次元ゲル電気泳動、吸光度の測定もしくは蛍光の測定を利用した抗体を使用した検出法、いくつかの化学発光レポーターシステムのいずれかの発光を用いた定量、酵素アッセイ、免疫沈降または免疫捕捉アッセイ、固相および液相免疫アッセイ、タンパク質アレイもしくはチップ、DNAアレイもしくはチップ、プレートアッセイ、識別を可能にする結合親和性を有する分子、例えばアプタマーや分子インプリントポリマーを使用したアッセイ、ならびに、他の任意の好適な技術および記載した検出法に関するいずれもの遂行意志または器具操作による、バイオマーカーの濃度を定量的分析的に決定する他のいかなる方法も含まれるがこれらには限定されない。

20

30

【0038】

〔062〕 バイオマーカー測定値を決定する工程は、当該分野で公知のいずれの手段によっても、特に本明細書で議論した手段によって実施することができる。好ましい態様では、バイオマーカー測定値を決定する工程は、抗体を利用した免疫アッセイを実施する工程を含む。当業者であれば、本発明で使用するのに適切な抗体を容易に選択することができる。選択される抗体は、好ましくは、目的の抗原に選択的 (すなわち、特定のバイオマーカーに選択的) であり、前記抗原に高い結合特異度を有し、かつ、他の抗原との交差反応性が最小限のものである。抗体の目的の抗原へと結合する能力は、例えば、既知の方法によって、例えば酵素結合免疫吸着アッセイ (ELISA)、フローサイトメトリー、および免疫組織化学によって決定することができる。さらに、抗体は、目的の抗原に対して、相対的に高い結合特異度を有するべきである。抗体の結合特異度は、既知の方法によって、例えば免疫沈降によって、またはインビトロ結合アッセイ、例えば放射免疫アッセイ (RIA) もしくは ELISA によって決定することができる。目的の抗原と高い結合特異度で結合することができ、かつ、交差反応性が最も小さい抗体を選択する方法は、例えば、参照により本明細書に組み込まれる、米国特許第7,288,249号で開示されている。肺疾患を示すバイオマーカーのバイオマーカー測定値は、サポートベクトルマシンなどの分類システムへの入力データとして使用することができる。

40

【0039】

50

〔063〕 それぞれのバイオマーカーはベクトル空間における次元として表すことができ、ここで各ベクトルは、特定の対象と関連のある複数のバイオマーカー測定値から成り立っている。従って、ベクトル空間の次元性は、バイオマーカーのセットの大きさと対応する。複数のバイオマーカーのバイオマーカー測定値のパターンを、様々な診断法および予後方法に使用することができる。本発明は、そのような方法を提供する。代表的な方法には、分類システム、例えばサポートベクトルマシンが含まれる。

【0040】

分類システム

〔064〕 本発明はとりわけ、継続的に分布している複数のバイオマーカーに基づいて、肺の病理を癌または喘息であると予測することに関する。いくつかの分類システム（例えば、サポートベクトルマシン）については、予測は3工程プロセスとなり得る。第一の工程では、予め設定しておいたデータセットの特徴を記述することによって分類器を構築する。これは「学習工程」であり、「学習」データについて行われる。

10

【0041】

〔065〕 学習データベースは、各個人の病状についての分類と関係のある、複数のヒトに関する複数のバイオマーカー測定値を反映している、コンピューターに実装されたデータの記録である。記憶されるデータの形式は、単層ファイル、データベース、表、または当該分野で知られている、他のいずれの検索可能なデータ記憶形式であってよい。例示的な態様では、試験データは複数のベクトルとして記憶され、ここで各ベクトルは個々のヒトに対応し、各ベクトルは複数のバイオマーカーに関する複数のバイオマーカー測定値と、ヒトの病状についての分類を合わせて含んでいる。通常、各ベクトルは、各バイオマーカー測定値の見出しを複数のバイオマーカー測定値中に含んでいる。学習データベースは、認可された実体（例えば、ヒト使用者またはコンピュータープログラム）によって遠隔操作で検索できるように、ネットワーク、例えばインターネットと接続されていてもよい。あるいは、学習データベースは、ネットワークから独立したコンピューター中に配置されていてもよい。

20

【0042】

〔066〕 第二の工程、これは任意であるが、では、分類器を「確認」データベースに適用し、感度や特異度などの、精度に関する様々な測定値を観測する。例示的な態様では、学習データベースの一部分のみを学習工程に使用し、学習データベースの残りの部分を確認データベースとして使用する。第三の工程では、ある対象に由来するバイオマーカー測定値を、対象に関する計算した分類（例えば、病状）を出力する分類システムにかける。

30

【0043】

〔067〕 分類器を構築するための複数の方法が当該分野で知られており、これらには、決定木、ベイジアン分類器、ベイジアン信念ネットワーク、k - 最近傍法、事例ベース推論、およびサポートベクトルマシン（Han J & Kamber M, 2006, 第六章, Data Mining, Concepts and Techniques, 第二版, Elsevier アムステルダム）が含まれる。好ましい態様では、本発明は、サポートベクトルマシンの使用に関する。しかしながら本明細書に記載したように、当該分野で知られているいずれの分類システムを使用してもよい。

40

【0044】

サポートベクトルマシン

〔068〕 サポートベクトルマシン（SVM）は当該分野において公知である。例えば、サポートベクトルマシンを使用して、病態の発症を診断する方法および予測する方法が提唱されている。例えば、参照によりその全体が本明細書に組み込まれる、米国特許第7,505,948号；同第7,617,163号；および同第7,676,442号を参照のこと。

【0045】

〔069〕 一般的にSVMは、各対象につき、バイオマーカーの1種類のk - 次元ベクトル（k - タプルと呼ばれる）に基づいて、各n個の対象を2つ以上の疾患クラスに分

50

類するのに使われるモデルを提供する。SVMでは最初に、カーネル関数を使用して、同じかより高次元の空間へとk-タプルを変換する。カーネル関数は、クラスが元々のデータ空間で分離されるよりも、超平面によってより大きく分離される空間にデータを写像する。クラスを分ける超平面を決定するためには、疾患クラス境界の最も近くにあるサポートベクトルのセットを選択してもよい。次いで超平面を、不適当な予測に罰金を科す費用関数の境界内でサポートベクトルと超平面との間の距離が最大になるように、既知のSVM技術で選択する。この超平面が予測の観点からデータを最適に分離する平面の1つである (Vapnik, 1998 *Statistical Learning Theory*, New York: Wiley)。その後、新たな観測はいずれも、観測されたデータの分布と超平面の関係に基づいて、目的のクラスのいずれか1つに属していると分類される。3つ以上のクラスについては、全てのクラスについてこのプロセスをペアワイズで行い、それらの結果を合わせて、全てのクラスを分離する規則を生成する。

10

【0046】

〔070〕 例示的な態様では、ガウスのラジアル基底関数 (RBF) (Vapnik、1998) として知られているカーネル関数を使用する (RBF関数は

【0047】

【数1】

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\sigma \|\mathbf{x} - \mathbf{x}'\|^2)$$

【0048】

であり、式中、

【0049】

【数2】

 $\mathbf{x}$ 

【0050】

および

【0051】

【数3】

 $\mathbf{x}'$ 

【0052】

は2つのk-タプルである)。RBFは、いくつかの定義されている他のカーネル関数、例えば多項カーネルまたはシグモイドカーネルから選択するための前提知識がない場合に使用されることが多い (Han J. & Kamber M., 343頁)。RBFでは、元々の空間を無限次元の新しい空間に写像する。この問題に関する議論および統計処理言語Rへのその実装は、Karatzoglou et al.に見られる (*Support Vector Machines in R*, *Journal of Statistical Software*, 2006)。本明細書に記載の全てのSVM統計計算は、統計処理用のソフトウェアプログラム用言語・環境であるR 2.10.0 ([www.r-project.org](http://www.r-project.org))を使用して実施した。kernelabパッケージに含まれているksvm()関数を使用してSVMに適用した。

30

【0053】

〔071〕 以下に、サポートベクトルマシンに関するいくつかの表記法 (Cristianini N, Shawe-Taylor J. *An Introduction to Support Vector Machines and other kernel-based learning methods*, 2000, p. 106)、ならびに彼らが複数のクラスに由来する観測を区別した方法の概要を説明する。

40

【0054】

〔072〕 以下のような学習試料または学習データベースを与えられたとする。

【0055】

【数4】

$$S = ((x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)) \quad (1)$$

50

【 0 0 5 6 】

式中、

【 0 0 5 7 】

【 数 5 】

$i=1, \dots, l$

【 0 0 5 8 】

ならば、

【 0 0 5 9 】

【 数 6 】

$x_i$

【 0 0 6 0 】

はバイオマーカー測定値のベクトルであり、および

【 0 0 6 1 】

【 数 7 】

$y_i$

【 0 0 6 2 】

は

【 0 0 6 3 】

【 数 8 】

$x_i$

【 0 0 6 4 】

がどのクラス（例えば、正常、非小細胞肺癌、喘息）に属しているかを示すものであり、特徴空間はカーネル

【 0 0 6 5 】

【 数 9 】

$K(\mathbf{x}, \mathbf{z})$

【 0 0 6 6 】

によって暗黙的に定義されている。指標

【 0 0 6 7 】

【 数 1 0 】

$\alpha^*$

【 0 0 6 8 】

が、

【 0 0 6 9 】

【 数 1 1 】

$$\text{最大化 } W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j \left( K(\mathbf{x}_i, \mathbf{x}_j) + \frac{1}{C} \delta_{ij} \right) \quad (2)$$

$$\text{制約条件 } \sum_{i=1}^l \alpha_i y_i = 0, \alpha_i \geq 0.$$

【 0 0 7 0 】

で表される二次の最適化問題を解決すると仮定すると、

【 0 0 7 1 】

【 数 1 2 】

$$f(\mathbf{x}) = \sum_{i=1}^l y_i \alpha_i^* K(\mathbf{x}_i, \mathbf{x}) + b^* \quad (3)$$

10

20

30

40

50

【 0 0 7 2 】

であるため、

【 0 0 7 3 】

【 数 1 3 】

 $b^*$ 

【 0 0 7 4 】

は、

【 0 0 7 5 】

【 数 1 4 】

10

 $\alpha_i^* \neq 0$ 

【 0 0 7 6 】

の任意の

【 0 0 7 7 】

【 数 1 5 】

 $i$ 

【 0 0 7 8 】

に関し、

【 0 0 7 9 】

【 数 1 6 】

20

$$y_i f(x_i) = 1 - \frac{\alpha_i^*}{C} \quad (4)$$

【 0 0 8 0 】

となるように選択される。よって、

【 0 0 8 1 】

【 数 1 7 】

30

$$\text{sgn}(f(\mathbf{x})) \quad (5)$$

【 0 0 8 2 】

で得られる決定規則は、

【 0 0 8 3 】

【 数 1 8 】

$$\begin{aligned} & \text{最小化} && (6) \\ & \xi, w, b && \langle w \cdot w \rangle + C \sum_{i=1}^I \xi_i^2 \\ \text{制約条件} &&& \end{aligned}$$

40

$$y_i (\langle w \cdot x_i \rangle + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

【 0 0 8 4 】

で与えられる最適化問題を解決するカーネル

【 0 0 8 5 】

【 数 1 9 】

 $K(\mathbf{x}, \mathbf{z})$ 

【 0 0 8 6 】

によって暗黙的に定義されている特徴空間中の超平面と等価になり、ここでスラック変数

50

【 0 0 8 7 】

【 数 2 0 】

 $\xi$   
 $S_i$ 

【 0 0 8 8 】

は、幾何マージン

【 0 0 8 9 】

【 数 2 1 】

$$\gamma = \left( \sum_{i \in S_v} \alpha_i^* - \frac{1}{C} \langle \alpha^* \cdot \alpha^* \rangle \right)^{\frac{1}{2}} \quad (7)$$

10

【 0 0 9 0 】

に対して定義され、ここで

【 0 0 9 1 】

【 数 2 2 】

 $S_v$ 

【 0 0 9 2 】

は、

【 0 0 9 3 】

【 数 2 3 】

 $\alpha > 0$ 

【 0 0 9 4 】

の指数の集合である（対応する

【 0 0 9 5 】

【 数 2 4 】

 $\mathbf{x}_i$ 

【 0 0 9 6 】

がサポートベクトルと呼ばれる）。

30

〔 0 7 3 〕 データを分類するクラスが 2 つ以上ある場合には、モデルにクラス間のペアワイズ（一連の下位モデル）を適用する。この場合、各下位モデルは特定のクラスに投票するものである。観測は、最も得票の多かったクラスに属していると判断される。

【 0 0 9 7 】

〔 0 7 4 〕 本発明の新規側面を形成する 1 つのカーネル関数は、

【 0 0 9 8 】

【 数 2 5 】

$$K(\mathbf{x}_i, \mathbf{x}_j) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{p\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left( 1 + \frac{\langle (\mathbf{x}_i - \mathbf{x}_j) \cdot (\mathbf{x}_i - \mathbf{x}_j) \rangle}{\nu} \right)^{-\frac{\nu+1}{2}} \quad (8)$$

40

【 0 0 9 9 】

と定義され、ここで

【 0 1 0 0 】

【 数 2 6 】

 $p$ 

【 0 1 0 1 】

はベクトル

【 0 1 0 2 】

50

【数 2 7】

$x_j$

【0 1 0 3】

の長さであり、かつ、

【0 1 0 4】

【数 2 8】

ひ

【0 1 0 5】

は予め設定しておいた定数（自由度）である。

【0 7 5】

10

式（8）が実際に有効なカーネルであることを確認するために、

【0 1 0 6】

【数 2 9】

$$\mathbf{K} = \left( K(x_i, x_j) \right)_{i,j=1}^n \quad (l=2)$$

【0 1 0 7】

で定義されている行列

【0 1 0 8】

【数 3 0】

20

$\mathbf{K}$

【0 1 0 9】

を考える。

【0 1 1 0】

【数 3 1】

$$-a_1 a_2 < \frac{(a_1^2 + a_2^2)}{4K(x_1, x_2)} \quad (9)$$

【0 1 1 1】

と注記することで、この行列が正定値（positive definite）であると示すことができる

30

。この場合、

【0 1 1 2】

【数 3 2】

$a_1$

【0 1 1 3】

および

【0 1 1 4】

【数 3 3】

$a_2$

【0 1 1 5】

40

は、

【0 1 1 6】

【数 3 4】

$\mathbf{x}_1 \neq \mathbf{x}_2$

【0 1 1 7】

に関する任意の二次元ベクトル

【0 1 1 8】

【数 3 5】

$\mathbf{a}$

50

【0119】

の要素である。帰納法および

【0120】

【数36】

$l=2$

【0121】

の場合と同様の論理を用いることで、行列

【0122】

【数37】

**K**

【0123】

が正定値であること、およびマーサーの定理によって、関数

【0124】

【数38】

$K(\mathbf{x}, \mathbf{z})$

【0125】

が有効なカーネル関数であることが分かる。

【076】 他の適したカーネル関数には、線形カーネル、ラジアル基底カーネル、多項カーネル、統一カーネル、三角カーネル、Epanechnikovカーネル、四次多項 (quartic) (二乗重み) カーネル、三次三乗 (tricube) (三乗重み) カーネル、およびコサインカーネルが含まれるがこれらには限定されない。

【0126】

他の分類システム

【077】 サポートベクトルマシンは、データについて使用することが可能な、多くの潜在性のある分類器のうちの1つである。非限定的な例として、そして以降で議論するように、例えばナイーブベイズ分類器、分類木、 $k$ -最近傍法などの他の方法を同じデータについて使用して学習させ、サポートベクトルマシンを検証してもよい。

【0127】

ナイーブベイズ分類器

【078】 一連のベイズ分類器とは、ベイズの定理、

【0128】

【数39】

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (10)$$

【0129】

に基づく一連の分類器である。

【079】 この種の分類器は全て、ある観測に関するデータに基づいて、その観測があるクラスに属する確率を見出そうとするものである。最も高い確率を有するクラスに、それぞれの新しい観測が割り当てられる。

【0130】

【080】 理論的には、ベイズ分類器の誤り率は一連の分類器の中で最も低い。実際には、ベイズ分類器を適用した場合にデータに関して作成される仮定の違反により、誤り率が常に最低であるということではない。

【0131】

【081】 ナイーブベイズ分類器は、ベイズ分類器の一例である。ナイーブベイズでは、それぞれのクラスが、データを与えるその他のクラスから独立していると仮定することで、分類に使用される確率計算を単純化する。

10

20

30

40

50

## 【0132】

〔082〕 ナイーブベイズ分類器は、実装が容易なこと、および分類速度が速いことから、多くの著名な反スパムフィルターに使用されているが、実際にはその仮定が満たされることが少ないという欠点をもつ。

## 【0133】

〔083〕 本明細書で議論するように、ナイーブベイズ分類器を実装するためのツールは、統計処理用ソフトウェア計算言語・環境であるRに使用可能である。例えば、R言語のパッケージである「e1071」の第1.5-25版は、ナイーブベイズ分類器を作成・処理・利用するためのツールを含んでいる。

## 【0134】

## ニューラルネット

〔084〕 ニューラルネットは、辺とそれらの重みが、各頂点のそれらが結合しているその他の頂点に及ぼす影響を表している重み付き有向グラフとして考えることができる。ニューラルネットは、入力層（データで形成されている）と出力層（値。この例では予測されるクラス）の2つの部分から構成される。入力層と出力層の間は、隠れ頂点のネットワークとなっている。ニューラルネットを設計した方法にもよるが、入力層と出力層の間には複数の頂点がある場合がある。

## 【0135】

〔085〕 ニューラルネットは人工知能やデータマイニングで広く利用されているが、ニューラルネットで作成されるモデルがデータを過剰適応させる（つまり、モデルが現在データを非常によく適応させるが、未来データはあまりよく適応しない）危険性がある。本明細書で議論するように、ニューラルネットを実装するためのツールは、統計処理用ソフトウェア計算言語・環境であるRに使用可能である。例えば、R言語のパッケージである「e1071」の第1.5-25版は、ニューラルネットを作成・処理・利用するためのツールを含んでいる。

## 【0136】

## k - 最近傍法

〔086〕 最近傍法は、メモリーベースの分類器の一部である。これらは、新しい観測を分類するために、学習セットの中に何が入っていたかを「覚えて」おく必要のある分類器である。最近傍法は当てはめるためのモデルを必要としない。

## 【0137】

〔087〕 k - 近傍 (knn) 分類器を構築するためには、以下の工程がとられる。

1. 分類する観測から、学習セット中の各観測への距離を計算する。距離の計算には任意の妥当な尺度を使用してもよいが、ユークリッド距離およびマハラノビス距離が使用されることが多い（マハラノビス距離とは、観測の中に含まれる変数間の共分散を考慮する尺度である）。
2. 各クラスに属すk最近傍観測の中にある観測の数を数える。
3. 新しい観測を、最も数の多かったクラスに割り当てる。

## 【0138】

〔088〕 最近傍アルゴリズムには、2点間の距離を計算する必要があるため、カテゴリカルデータを取り扱うという問題があるが、これは、任意の2クラス間の距離を人為的に定義することで克服することができる。この種のアルゴリズムはまた、スケールや尺度の変化に対して感度がよい。これらの問題を念頭においても、最近傍アルゴリズムは、特に大量のデータセットに対して非常に強力となり得る。

## 【0139】

〔089〕 本明細書で議論するように、k - 最近傍法を実装するためのツールは、統計処理用ソフトウェア計算言語・環境であるRに使用可能である。例えば、R言語のパッケージである「e1071」の第1.5-25版は、k - 最近傍法を作成・処理・利用するためのツールを含んでいる。

## 【0140】

10

20

30

40

50

## 分類木

〔090〕 分類木とは、特徴選択が組み込まれた翻訳処理が容易な分類器である。分類木は、各下位空間に含まれている1つのクラスからの観測の割合が最大となるような方法で、データ空間を再帰的に分割する。

【0141】

〔091〕 データ空間を再帰的に分割するプロセスによって、それぞれの端点に試験される条件をもつ二分木が作成される。葉に到達するまで木の枝を追っていくことによって、新しい観測を分類する。それぞれの葉では、ある確率が、所定のクラスに属している観測に割り当てられる。新しい観測は、最も高い確率を有するクラスに分類される。

【0142】

〔092〕 分類木は本質的に、その属性が統計処理言語で構成されている決定木である。これらは非常に柔軟であるが、ノイズが非常に多い（他の方法に比べて誤りの分散が大きい）。

【0143】

〔093〕 本明細書で議論するように、分類木を実装するためのツールは、統計処理用ソフトウェア計算言語・環境であるRに使用可能である。例えば、R言語のパッケージである「tree」の第1.0-28版は、分類木を作成・処理・利用するためのツールを含んでいる。

【0144】

## ランダムフォレスト

〔094〕 分類木は一般的にノイズが多い。ランダムフォレストは、多くの木の平均をとることで、このノイズを低下させようとするものである。その結果、ある分類木と比較して、誤りの分散が小さくなった分類木が得られる。

【0145】

〔095〕 以下のアルゴリズムを使用して、森を大きくする：

1. Bが森を構成する木の数である

【0146】

【数40】

$\mathcal{B} = \{1 \text{ to } B\}$

【0147】

について、

a. ブーストラップサンプルを生成する（ブーストラップサンプルとは、観測データと同じ数の観測を有する観測データからの置き換えによって導き出されたサンプルである）

b. このブーストラップサンプルについて分類木、

【0148】

【数41】

$T_b$

【0149】

を作成する。

2. セット

【0150】

【数42】

$\{T_b\}_{b \in \mathcal{B}}$

【0151】

を出力する。このセットがランダムフォレストである。

〔096〕 ランダムフォレストを使って新しい観測を分類するには、ランダムフォレストに含まれているそれぞれの分類木を使って新しい観測を分類する。分類木のうち、新しい観測が最も頻繁に分類されるクラスが、その新しい観測をランダムフォレストが分類

10

20

30

40

50

するクラスである。

【0152】

〔097〕 翻訳処理の評価においては、ランダムフォレストは分類木で見られる多くの問題を低減する。

〔098〕 本明細書で議論するように、ランダムフォレストを実装するためのツールは、統計処理用ソフトウェア計算言語・環境であるRに使用可能である。例えば、R言語のパッケージである「randomForest」の第4.6-2版は、ランダムフォレストを作成・処理・利用するためのツールを含んでいる。

【0153】

アダブースト（適応的ブースティング）

〔099〕 アダブーストは、各対象につき、測定値の1種類の $k$ -次元ベクトル（ $k$ -タプルと呼ばれる）に基づいて、各 $n$ 個の対象を2つ以上の疾患クラスに分類するのに使われる方法を提供する（アダブーストは、技術的には、ある観測が属するクラスが2つの場合にのみ機能する。 $g > 2$ の場合には、観測を「なし」のクラスに分類する（ $g/2$ ）のモデルを生成する必要がある。その後、これらのモデルから得られた結果を合わせて、特定の観測が属するクラスを予測する）。アダブーストは、予測性能が良くないが、無作為よりはましな、一連の「弱い」分類器を利用し、それらを組み合わせて優れた分類器を生成する（この文脈における予測性能とは、誤って分類される観測として定義される）。アダブーストで使用される弱い分類器は、分類木および回帰木（CART）である。CARTはデータ空間を、その領域中に分布している新しい観測の全てが特定の分類名に割り当てられている領域に、再帰的に分割する。アダブーストは、重み付けられているデータセットに基づいて、一連のCARTを構築する。ここで重みは、直前の反復で使用された分類器の性能に依存する（Han J & Kamber M, 2006, 第六章, Data Mining, Concepts and Techniques, 第二版, Elsevier アムステルダム）。

10

20

【0154】

データを分類する方法

〔0100〕 本発明は、個人から得たデータ（試験データ、すなわち、バイオマーカー測定値）を分類する方法を提供する。これらの方法は、学習データを準備または収集する工程、および個人から得た試験データを上述した分類システムのうちの1つを利用して評価する（学習データと比較して）工程を含む。好ましい分類システムは学習機械であり、例えば、サポートベクトルマシンまたはアダブースト分類器である。分類システムは、試験データに基づき、個人の分類を出力する。

30

【0155】

〔0101〕 試験データは、いかなるバイオマーカー測定値であってもよく、例えば複数のバイオマーカーの血漿濃度の測定値であってよい。一態様において本発明は、それぞれのバイオマーカーのセットの、複数の血漿濃度の測定値であるバイオマーカー測定値を含んでいる試験データの分類方法を提供し、この方法は、（a）電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルまたは $k$ -タプルは個々のヒトを表しており、かつ、バイオマーカー測定値、例えば、対応するヒトに関する各バイオマーカーのセットの、血漿濃度の測定値を含んでおり、学習データはさらに、対応するそれぞれのヒトの病状についての分類を含んでおり；（b）電子的に記憶されている学習データベクトルのセットを使用して、サポートベクトルマシンの電子表示を訓練する工程；（c）ヒト試験対象に関する複数の血漿濃度の測定値を含む試験データを受信する工程；（d）サポートベクトルマシンの電子表示を使用して、試験データを評価する工程；および（e）評価する工程に基づいたヒト試験対象の分類を出力する工程、を含む。別の態様では、本発明は、各バイオマーカーのセットの複数の血漿濃度の測定値であるバイオマーカー測定値を含む試験データを分類する方法を提供し、この方法は、（a）電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルまたは $k$ -タプルは個々のヒトを表しており、かつ、バイオマーカー測定値、例えば対応するヒトに関する各バイオマーカーのセットの、血漿濃度の測定

40

50

値を含み、学習データはさらに、対応するそれぞれのヒトの病状についての分類を含んでおり；(b) アダプストを介した分類器を構築するために、電子的に記憶されている学習データベクトルのセットを使用する工程；(c) ヒト試験対象に関する複数の血漿濃度の測定値を含む試験データを受け取る工程；(d) アダプスト分類器を使用して試験データを評価する工程；および(e) 評価する工程に基づいたヒト試験対象の分類を出力する工程、を含む。本発明による出力は、電子表示にヒトが読み取り可能な形式で表示する工程を含む。

【0156】

〔0102〕 病状についての分類は、病状の有無であってよい。本発明による病状は、肺疾患、例えば非小細胞肺癌または反応性気道疾患（例えば、喘息）となり得る。

10

〔0103〕 学習ベクトルのセットは、少なくとも20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、105、110、115、120、125、130、135、140、145、150、またはそれ以上のベクトルを含み得る。

【0157】

〔0104〕 当然のことながら、データを分類する方法は、本明細書に記載のいずれの方法においても使用することができる。具体的には、本明細書に記載のデータを分類する方法は、生理学的に特徴付ける方法、および肺疾患、例えば非小細胞肺癌および反応性気道疾患（例えば、喘息）を診断するための方法において使用することができる。

【0158】

少ない数のバイオマーカーを使用した、データの分類

〔0105〕 本発明はまた、減数したセットのバイオマーカーを含む、データ（例えば個人から得た試験データ）を分類する方法を提供する。つまり、選択したバイオマーカーの下位セットに関するバイオマーカー測定値の下位セットだけを含めることで学習データを少なくすることができる。同様に、試験データも、選択した同じセットのバイオマーカーに由来するバイオマーカー測定値の下位セットだけに限定することができる。

20

【0159】

〔0106〕 一態様では、本発明は、各バイオマーカーのセットの複数の血漿濃度の測定値であるバイオマーカー測定値を含む試験データを分類する方法を提供し、この方法は、(a) 電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルは個々のヒトを表しており、かつ、対応するヒトに関するバイオマーカーのセットの、各バイオマーカーのバイオマーカー測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；(b) バイオマーカーのセットから、バイオマーカーの下位セットを選択する工程；(c) 学習機械、例えばサポートベクトルマシンの電子表示を、電子的に記憶されている学習データベクトルのセットのバイオマーカーの下位セットからのデータを使用して訓練する工程、；(d) ヒト試験対象に関する複数の血漿濃度の測定値を含む試験データを受け取る工程；(e) 学習機械の電子表示を使用して試験データを評価する工程；および(f) 評価する工程に基づいたヒト試験対象の分類を出力する工程、を含む。

30

【0160】

〔0107〕 好ましい態様では、バイオマーカーの下位セットを選択する工程は、(i) バイオマーカーのセットに含まれているそれぞれのバイオマーカーについて、バイオマーカーの2群の濃度測定値のマージン分布間の距離を計算すること、それによって、複数の距離が生成され；(ii) その距離に従って、バイオマーカーのセットに含まれているバイオマーカーを順序付けること、それによって、順序付けられたバイオマーカーのセットが生成され；(iii) 順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについて、学習データに基づいて、学習機械用のモデルフィットの指標を計算すること；(iv) モデルフィットの最大値型指標に従って、順序付けられたバイオマーカーのセットの初期セグメントを選択すること、それによって、順序付けられたバイオマーカーのセットの好ましい初期セグメントが選択され；(v) バイオマーカーの空集合から開始し、バイオマーカーの下位セットを生成するために、順序付けられたバイオマーカーのセットの、好ましい初期セグメントからさらなるバイオマーカーをモデルに再帰

40

50

的に追加すること、ここで各さらなるバイオマーカーは、(a) 好ましい初期セグメントに含まれている残りのバイオマーカーの中でも、それを追加することで、モデルフィットが最大に改良される場合、および(b) それを追加することで、モデルフィットが少なくとも予め設定した閾値までは改良される場合に、既存のバイオマーカーの下位セットに追加され；および(vi) それ以上バイオマーカーを追加しても、予め設定しておいた閾値より、そのモデルフィット指標を超えるモデルフィットの指標が得られない場合には、既存のバイオマーカーの下位セットへのバイオマーカーの追加を停止すること、それによって、バイオマーカーの下位セットが選択される、を含む。

**【0161】**

〔0108〕 本明細書に記載の方法、キット、およびシステムは、選択した複数のバイオマーカーのバイオマーカー測定値を決定する工程を含む場合がある。好ましい形態では、この方法は、実施例に記載したバイオマーカーの、任意の3種類のバイオマーカーの下位セットのバイオマーカー測定値を決定する工程を含む。あるいは、この方法は、実施例に記載したバイオマーカーの、少なくとも4、5、6、または7種類の特定のバイオマーカーの下位セットのバイオマーカー測定値を決定する工程を含む。あるいは、この方法は、実施例に記載したバイオマーカーの、少なくとも8、9、10、11、12、または13種類の特定のバイオマーカーの下位セットのバイオマーカー測定値を決定する工程を含む。あるいは、この方法は、実施例に記載したバイオマーカーの、少なくとも14、15、16、17、18、19、20、またはそれ以上(例えば、59)の種類の特定のバイオマーカーの下位セットのバイオマーカー測定値を決定する工程を含む。無論、さらなるバイオマーカーのバイオマーカー測定値を、目的の疾患に関連があるうとなかろうと、同時に決定することも本発明の想定内であることを当業者は理解する。これらさらなるバイオマーカー測定値の決定は、本発明による対象の分類を妨害するものではない。

**【0162】**

〔0109〕 バイオマーカーの下位セットは、本明細書に記載の減数方法によって決定することができる。例えば、本発明は、モデルフィットの最高値型指標に寄与し、それによって高い予測精度を維持するバイオマーカーの下位セットを見出すための、様々なモデル選択アルゴリズム(例えば、F<sub>SSFS</sub>)を提供する。実施例7~10に、減数した、特定のバイオマーカーの下位セットのモデルを示す。

**【0163】**

〔0110〕 好ましい形態では、バイオマーカーは、モデルフィットの最大値型指標に寄与しているバイオマーカーを含む、コンピューターで計算した下位セットから選択される。これらのバイオマーカーが含まれている限り、本発明は、必ずしも貢献しない、少数のさらなるバイオマーカーを含むことを排除しない。モデルを本明細書に記載したように導き出す限り、そのようなさらなるバイオマーカー測定値を分類モデルに含めても、試験データの分類は妨害されない。他の態様では、対象に関し、4、5、6、7、8、9、10、12、15、20、25、30、35、40または50以下のバイオマーカーのバイオマーカー測定値を決定し、また、同じ数のバイオマーカーが学習段階でも使用される。

**【0164】**

〔0111〕 別の形態では、選択バイオマーカーは、モデルフィットの指標への寄与が最も低かったバイオマーカーを除いた、コンピューターで計算したバイオマーカーの下位セットから選択される。これらの選択されたバイオマーカーが含まれている限り、本発明は、必ずしも貢献しない、少数のさらなるバイオマーカーを含むことを排除しない。モデルを本明細書に記載したように導き出す限り、そのようなさらなるバイオマーカー測定値を分類モデルに含めても、試験データの分類は妨害されない。他の態様では、対象に関し、4、5、6、7、8、9、10、12、15、20、25、30、35、40または50以下のバイオマーカーのバイオマーカー測定値を決定し、また、同じ数のバイオマーカーが学習段階でも使用される。

**【0165】**

〔 0 1 1 2 〕 また、本明細書に記載のバイオマーカーの様々な組み合わせを、本明細書に記載のキットの設計方法ならびにキットおよびシステムに適用可能であることも理解される。別の態様では、観測または試験データを訓練したモデルを使用して分類するために、学習機械、例えばサポートベクトルマシンで使用するバイオマーカーの数を、任意の数のクラスに拡張された L e e の F \_ S S F S 方法 ( L e e 、 2 0 0 9 ) で減数する。F \_ S S F S 方法は、( i ) モデル中に保持されるのに適した候補である変数のセットを決定し；および ( i i ) クラス間の変数の値の分離を定量化する F - スコアに基づいて、候補を選択する。(ここで、

【 0 1 6 6 】

【 数 4 3 】

$\bar{x}_i$

【 0 1 6 7 】

の変数の F - スコアは

【 0 1 6 8 】

【 数 4 4 】

$$\frac{\sum_{j=1}^g (\bar{x}_i - \bar{x})^2}{\sum_{j=1}^g \left( \frac{1}{n_j - 1} \sum_{k=1}^{n_j} (x_{ki} - \bar{x}_i)^2 \right)}$$

【 0 1 6 9 】

として定義され、式中、

【 0 1 7 0 】

【 数 4 5 】

$g$

【 0 1 7 1 】

はクラスの数であり、

【 0 1 7 2 】

【 数 4 6 】

$n_j$

【 0 1 7 3 】

はクラス

【 0 1 7 4 】

【 数 4 7 】

$j$

【 0 1 7 5 】

に由来する観測の数である。) この変数のセットには、前向きモデル選択が、学習機械の精度の向上に基づいてモデルに加えられた変数とともに適用される。本明細書で例示するように、変数はバイオマーカーであり、クラスは肺病理の分類である。代表的な学習機械としては、SVMやアダブースト分類器が挙げられる。

【 0 1 7 6 】

〔 0 1 1 3 〕 現在のところ、バイオマーカーの下位セットを選択するための別の技術も開示されている。この技術の代表的なアルゴリズムには以下の工程が含まれる。

1 . 以下の工程は、学習データベクトルのセットに関するものであり、それぞれのセットには分類が含まれる。バイオマーカーのセット

【 0 1 7 7 】

【 数 4 8 】

$G$

【 0 1 7 8 】

10

20

30

40

50

に含まれている各バイオマーカー

【 0 1 7 9 】

【 数 4 9 】

$g_r$  ( $r=1, \dots, p$ )

【 0 1 8 0 】

について、各学習データベクトル、と関連がある経験的な分類によって定義した 2 クラスのバイオマーカー

【 0 1 8 1 】

【 数 5 0 】

$g_r$

【 0 1 8 2 】

のマージン分布間の距離メトリックを、以下のように計算する。

【 0 1 8 3 】

【 数 5 1 】

$$F_r = \frac{\sum_{s=1}^m (\tilde{x}_{r,s} - \tilde{x}_r)}{\sum_{s=1}^m (x_{r,s(0.75)} - x_{r,s(0.25)})} \quad (11)$$

【 0 1 8 4 】

式 ( 1 1 ) の記号

【 0 1 8 5 】

【 数 5 2 】

$m$

【 0 1 8 6 】

は、検討中のクラスの数である。学習機械、例えばサポートベクトルマシンを使用する多くの例において、 $m = 2$  である。記号

$\tilde{x}_r$

【 0 1 8 7 】

は、学習データベクトルのセットに含まれている  $g_r$  のバイオマーカー測定値の中央値を表す。記号

【 0 1 8 8 】

【 数 5 3 】

$\tilde{x}_{r,s}$

【 0 1 8 9 】

は、バイオマーカー  $g_r$  の  $s$  - 番目のクラスの中央値を表し、ここで各クラスは、学習データベクトルの分類によって定義される。記号

【 0 1 9 0 】

【 数 5 4 】

$x_{r,s(0.75)}$

【 0 1 9 1 】

および

【 0 1 9 2 】

10

20

30

40

【数 5 5】

 $x_{r,s(0.25)}$ 

【0 1 9 3】

はそれぞれ、 $s$  - 番目のクラスの分布の第一および第三四分位数を示す（バイオマーカー  $g_r$  について）。経験的な学習データベクトル分類を使用して  $s$  で表示される 2 つのクラスを定義することの代替法は、各学習ベクトルを別個のクラスに分類するための全てのバイオマーカーを含むサポートベクトルマシンの初回の実行を利用することであることに留意する。

2. 式 (11) に従って、最大から最小の順番でバイオマーカーを順序付ける。

3. 式 (12) によって、カットオフ指数のセットを定義する。

【0 1 9 4】

【数 5 6】

$$K = \{K_t : K_t = \lfloor tp/4 \rfloor, \text{ where } t \in \{1, \dots, c\} \text{ for some } c \geq \log_2 p\} \quad (12)$$

【0 1 9 5】

ここで  $p$  はベクトルの次元である。

4. それぞれの

【0 1 9 6】

【数 5 7】

 $K_t \in K$ 

【0 1 9 7】

について、訓練した学習機械に関連するモデル中の第一の

【0 1 9 8】

【数 5 8】

 $K_t$ 

【0 1 9 9】

バイオマーカーを保持し、スコアに従って順序付け（降順に）、およびモデルフィットの指標（例えば感度または精度）を計算する。言い換えれば、バイオマーカーの初期セグメントを保持し、モデル中で (11) に従って順序付け、およびモデルフィットの使用、例えば正しく分類された試験ベクトルのパーセンテージを計算する（モデルフィットの他の指標としては、精度、感度、特異度、陽性予測値、および陰性予測値が挙げられる。例えば表 2 を参照のこと）。これを、 $K$  に含まれている各

【0 2 0 0】

【数 5 9】

 $K_t$ 

【0 2 0 1】

のバイオマーカーの各初期セグメントについて（すなわち、式 (11) に従って第一のバイオマーカーから

【0 2 0 2】

【数 6 0】

 $K_t$ 

【0 2 0 3】

番目のバイオマーカーまで) 実行する。

【0 2 0 4】

10

20

30

40

【数 6 1】

 $K'$ 

【0 2 0 5】

を

【0 2 0 6】

【数 6 2】

 $K_r$ 

【0 2 0 7】

にする。ここでは、その

【0 2 0 8】

【数 6 3】

 $K_r$ 

【0 2 0 9】

に關係するモデルが、モデルフィットの最大値型指標を有するようにする。

5 . セットを定義する。

【0 2 1 0】

【数 6 4】

 $G' = \{g_r \mid F_r \geq K'\}$ 

(13)

【0 2 1 1】

6 . バイオマーカーを含まないモデルから開始し、現在モデルには入っていない各

【0 2 1 2】

【数 6 5】

 $g_r \in G'$ 

【0 2 1 3】

に

【0 2 1 4】

【数 6 6】

 $g_r$ 

【0 2 1 5】

を加え、モデルフィットの選択指標を計算する。その後、このバイオマーカーを除き、セット中の次のバイオマーカーを加える。

7 . 工程 6 でモデルフィットを最もよく改善したバイオマーカーを、その改善が予め設定しておいた閾値を上回る場合に追加する。例えば、モデルフィットの指標を正しく分類された試験ベクトルのパーセンテージとすれば、閾値は、0 . 0 0 0 5、0 . 0 0 0 1、0 . 0 0 5、0 . 0 0 1、0 . 0 5、0 . 0 1、0 . 5、または 0 . 1 となり得る。直前の文では、数を絶対値として、つまり、パーセンテージ (%) なしで表した。そのため、例えば、0 . 0 0 0 5 は 0 . 0 5 % である。

8 . 工程 7 でバイオマーカーを追加した場合には工程 5 に戻る。そうでなければ、アルゴリズムの直前の反復からのセット、または工程 7 で定義した追加への基準をどのバイオマーカーも満たさない場合には空集合が、このモデルで使用される、減数されたバイオマーカーのセットである。従って、工程 6 および 7 の第一の反復では、単一のバイオマーカーがモデルに追加され (バイオマーカーがいずれも閾値の基準を満たさない場合を除いて)、次いで、閾値の基準に従ってプロセスが停止されるまで、さらなるバイオマーカーがその後の各反復で追加される。よって、工程 6 ~ 8 は、減数されたバイオマーカーのセット

10

20

30

40

50

を選択するための再帰的アルゴリズムを提供する。

【0216】

〔0114〕 従って上記工程1および2は、マージン分布に従って、バイオマーカーに順序を付けることに向けられる。具体的には、学習ベクトルのセットに含まれている2クラスのバイオマーカー測定値のマージン分布の中心傾向（例えば、中央値）の間の距離に従って、バイオマーカーに順序を付けることができる（中央値の代わりに、別の中心傾向、例えば形態または手段を使用してもよい）。各クラスは分類に対応し、およびこれらの分類は、学習データ自体に含まれている経験的な分類から得てもよく、またはこれらの分類は、全てのバイオマーカーを使用した学習機械の初回の実行から得てもよい。従って、バイオマーカーは、経験的なものであっても、または学習機械の初回の実行によって生成されたものであったとしても、分類に対応する2つのクラス間のバイオマーカー測定値の識別性能の関数として順位付けられる。

10

【0217】

〔0115〕 上記工程3、4および5は、選択された初期セグメントが、その他の初期セグメントの中で、学習ベクトルのセットに関する最もよいモデルフィットを有するように、マージン分布によって降順に順位付けられたバイオマーカーの初期セグメントを選択することに向けられている。この初期セグメントは、工程6、7および8に従って、最終的な減数されたバイオマーカーのセットをそこから選択するための、バイオマーカーの普遍集合として機能する。

【0218】

〔0116〕 工程6、7および8は、バイオマーカーのない基本事例から開始して、再帰的にモデルにバイオマーカーを追加することに向けられている。順番に追加されたバイオマーカーは、マージン分布による順序を考慮せず、モデルフィットへの寄与に従って選択される。工程の基本は、モデルに加えられるバイオマーカーの空集合を考慮することである。再帰工程では、さらなるバイオマーカーを追加するかどうかを決定するために、モデルに含まれている現在のバイオマーカーのセットと一緒に、残りの各バイオマーカーについて学習機械が生成される。残りのバイオマーカーのうち、既存のバイオマーカーに追加した時に、もっとも正確な学習機械に対応するものが、続いて行う追加への候補である。候補バイオマーカーのモデルフィットへの寄与が閾値を上回る限り、それらは順次追加される。この連続的なバイオマーカーの追加プロセスは、残りのバイオマーカーのうちの最もよいものが、予め設定しておいた閾値を超えてモデルフィットを改善しなくなるまで続けられる。

20

【0219】

〔0117〕 要約すると、このプロセスは工程1～5でバイオマーカーの最初の普遍集合を選択することに始まり、次いで、工程6、7および8に従って、この普遍集合から、最終的な減数されたバイオマーカーのセットを選択することに続く。

30

【0220】

〔0118〕 あるいは、工程5で定義する上位セットとなる工程6で定義する最初のモデルを変更し、上位セットから各バイオマーカーを追加する代わりに各バイオマーカーを1つずつ除き、次いでモデルフィットの指標を計算することで、減数されたバイオマーカーのセットを導き出すこともできる。従って、工程7を変更して、モデルフィットの指標が予め設定しておいた閾値を下回るまでは縮小されないように、モデルフィットの指標の縮小が最も小さいバイオマーカーを除く。その後、工程7で追加するバイオマーカーがなくなる代わりに、工程7で除くバイオマーカーがなくなることを停止条件とする工程8に従う。

40

【0221】

〔0119〕 上述した、バイオマーカーの下位セット選択アルゴリズムは、データから予測される好ましいバイオマーカーのセットを含む学習機械を提供することに加えて、考慮するバイオマーカーの結合および相関を明らかにすることができる。これを達成するためには、上述したアルゴリズムの工程7における閾値を排除し、次いで、アルゴリズム

50

の各反復の時点で、直前の反復によって示唆されているモデルと正確に関連しているマージン分布の改善によって付けられた順位またはそれぞれの反復と直前の反復との間の精度の向上に従って追加されたバイオマーカーを保持する。

【0222】

〔0120〕 当然のことながら、減数したバイオマーカーのセットまたはバイオマーカーの下位セットを使用してデータを分類する方法は、本明細書に記載した方法のいずれにも使用することができる。具体的には、本明細書に記載の少ない数のバイオマーカーを使用してデータを分類する方法を、生理学的に特徴付ける方法および肺疾患、例えば非小細胞肺癌および反応性気道疾患（例えば、喘息）を診断する方法に使用してもよい。減らした数のバイオマーカー以外のバイオマーカーを追加してもよい。これらのさらなるバイオマーカーは、診断に寄与してもしなくても、または診断を強化してもしなくてもよい。

10

【0223】

〔0121〕 診断的アッセイまたは予後アッセイで使用するためのバイオマーカーの選択は、特定のバイオマーカーとそれらの一次相互作用器との間の既知の相互作用によって容易になる可能性がある。本発明者らによって同定された多くのバイオマーカーは、全てではなくとも、細胞または器官の様々なコミュニケーション経路に関わっている。コミュニケーション経路中の1つの構成要素の正常からの逸脱は、そのコミュニケーション経路中の他のメンバーにおける関連する逸脱によってもたらされると予想される。当業者であれば容易に、様々なデータベースや使用可能なバイオインフォマティクスソフトウェアを使用して、コミュニケーション経路のメンバーに関連付けることができる（例えば、AR IADNE PATHWAY STUDIO（登録商標）、Ariadne, Inc.、<www.ariadne.genomics.com>またはChEMBLデータベース、European Bioinformatics Institute、European Molecular Biology Laboratory、<www.ebi.ac.uk>を参照のこと）。複数のバイオマーカーのレベルを決定することに基づいた診断的方法、ここで複数のバイオマーカーには、他のものとは別のコミュニケーション経路に関わっているいくつかのバイオマーカーが含まれる、はバイオマーカーレベルを測定することによってもたらされる情報を最大化するだろう。代替的な態様では、選択した下位セット中の任意のバイオマーカーを、同じコミュニケーション経路中の別のバイオマーカー（すなわち、そのバイオマーカーの一次相互作用器）で置き換えてもよい。サポートベクトルマシンの態様では、バイオマーカーを一次相互作用器で置き換えることは、置き換えたバイオ測定値を使用してサポートベクトルマシンを再学習させることを含み得る。

20

30

【0224】

生理学的に特徴付ける方法

〔0122〕 本発明は、以下に記載する様々な集団中に含まれている個々のヒトを、生理学的に特徴付ける方法に向けられている。本明細書で使用する場合、本発明による生理学的に特徴付ける方法には、特定の肺疾患を診断する方法、個々のヒトが治療的介入に反応する見込みを予測する方法、個々のヒトがそれぞれの肺疾患を発症するリスクを有しているか否かを決定する方法、患者の疾患重篤度を分類する方法、および共通したいくつかの症状を呈している患者を区別する方法が含まれる。これらの方法は通常、本明細書に記載の特定のバイオマーカーのバイオマーカー測定値を決定すること、および分類システム、例えばサポートベクトルマシン中のこれらの値を利用し、これらの生理的な特徴付けのうちの1つに従って個々のヒトを分類することに依存している。

40

【0225】

A. 母集団の特徴付け

〔0123〕 本発明は、対象を生理学的に特徴付ける方法を提供し、この方法は、対象の生理学的な試料中の複数のバイオマーカーのバイオマーカー測定値を決定する工程を含み、ここで、複数のマーカーの発現パターンと生理的な状況もしくは状態、または疾患状況の変化（例えば、非小細胞肺癌の期）もしくは状態には相関がある。好ましい態様では、複数のバイオマーカーの発現パターンは、肺疾患例えば非小細胞肺癌または反応性気道疾患を示すものであるか、または反応性気道疾患であるかまたは非小細胞肺癌であるかの

50

区別を補助するものである。好ましくは、複数のバイオマーカーは、機械学習アルゴリズム、例えばサポートベクトルマシンを介した学習データの解析に基づいて選択される。学習データには、多数の対象についての複数のバイオマーカー、ならびに個々の対象に関する疾患の分類情報（例えば、式（1）の  $y_i$ ）、および場合により、対象の他の特徴、例えば性別、人種、年齢、喫煙歴、職歴などが含まれる。別の好ましい態様では、バイオマーカーの発現パターンと、対象が特定の疾患または状態を有しているまたは有する可能性があることについての見込みの高さには相関がある。より好ましい態様では、対象の複数のバイオマーカーのバイオマーカー測定値を決定する方法は、対象が、肺疾患、例えば非小細胞肺癌または反応性気道疾患（例えば、喘息）を発症している、罹患している、または罹患する可能性があることについての見込みの上昇を検出する。発現パターンは、パターン認識について当該分野で公知のいかなる技術によって特徴付けてもよい。複数のバイオマーカーには、実施例1に記載したバイオマーカーのいかなる組み合わせも含まれ得る。

10

20

30

40

50

#### 【0226】

〔0124〕 一態様では、対象は、非小細胞癌または反応性気道疾患（例えば、喘息、慢性閉塞性肺疾患など）の肺疾患のリスクを有している。「リスクを有している」対象には、無症状ではあるが、個人歴もしくは家族歴、行動、疾患原因物質への曝露（例えば、発癌物質）、または他の複数の原因によって、その集団の大部分よりも疾患を発症する可能性の高い個人が含まれる。「リスクを有している」個人は、これまで、個人について決定したリスク因子を統合することで同定されていた。本発明は、関連するバイオマーカーのバイオマーカー測定値を決定することで、「リスクを有している」個人の強力な特徴付けを提供する。

#### 【0227】

〔0125〕 上述した態様は、実施例に記載するバイオマーカーの一覧によって例示される。当然のことながら、これらのバイオマーカーの下位セット、例えば実施例1～9に記載の下位セットを、記載したいずれの態様で使用してもよい。当業者の自由裁量で、他のバイオマーカーのバイオマーカー測定値を含めることもできる。

#### 【0228】

##### B．男性集団の特徴付け

〔0126〕 好ましい態様では、本発明は、男性対象を生理学的に特徴付ける方法を提供し、この方法は、男性対象の生理学的な試料中の複数のバイオマーカーのバイオマーカー測定値を決定する工程を含み、ここで、複数のマーカーの発現パターンと生理的な状況もしくは状態、または疾患状況の変化（例えば、非小細胞肺癌の期）もしくは状態には相関がある。別の好ましい態様では、複数のバイオマーカーの発現パターンは、肺疾患、例えば非小細胞肺癌または反応性気道疾患を示すものとなるか、または反応性気道疾患であるかもしくは非小細胞肺癌であるかの区別を補助するものである。好ましくは、問題となっている病状を有していると同定された多数の男性対象およびその疾患を有していないことが分かっている同程度の数の男性対象のバイオマーカー測定値を含む学習データの収集に基づいて、複数のバイオマーカーを選択する。次いで、この学習データを、機械学習アルゴリズム、例えばサポートベクトルマシンで解析する。発現パターンは、パターン認識について当該分野で公知のいかなる技術によって特徴付けてもよい。複数のバイオマーカーには、実施例、例えば実施例1～5または7～8に記載したバイオマーカーのいかなる組み合わせも含まれ得る。

#### 【0229】

〔0127〕 一態様では、男性対象は、非小細胞癌または反応性気道疾患（例えば、喘息、慢性閉塞性肺疾患など）の肺疾患のリスクを有している。「リスクを有している」対象および個人については上で説明した。

#### 【0230】

##### C．女性集団の特徴付け

〔0128〕 本発明はまた、女性対象を生理学的に特徴付ける方法を提供する。好まし

い態様では、本発明は、女性対象の生理学的に特徴付ける方法を提供し、この方法は、女性対象の生理学的な試料中の複数のバイオマーカのバイオマーカ測定値を決定する工程を含み、ここで、複数のマーカの発現パターンと生理的な状況もしくは状態、または疾患状況の変化（例えば、非小細胞肺癌の期）もしくは状態には相関がある。別の好ましい態様では、複数のバイオマーカの発現パターンは、肺疾患、例えば非小細胞肺癌または反応性気道疾患を示すものであるか、または反応性気道疾患であるかもしくは非小細胞肺癌であるかの区別を補助するものである。これらの態様での方法は、上述した方法と同様であるが、学習データセットに含まれる対象は女性となる。複数のバイオマーカには、実施例、例えば実施例 1 ~ 4、6 ~ 7、および 9 に記載したバイオマーカのいかなる組み合わせも含まれ得る。

10

【0231】

〔0129〕

一態様では、女性対象は、非小細胞癌または反応性気道疾患（例えば、喘息、慢性閉塞性肺疾患など）の肺疾患のリスクを有している。「リスクを有している」対象および個人については上で説明した。

【0232】

肺疾患

〔0130〕 本発明は、肺疾患についての様々な診断方法および予後方法を提供する。具体的には、本発明は、反応性気道疾患、具体的には過剰応答性  $TH_2$  細胞および  $TH_1$  細胞と関連がある疾患を診断する方法を提供する。反応性気道疾患には、喘息、慢性閉塞性肺疾患、アレルギー性鼻炎、嚢胞性線維症、気管支炎、または、様々な生理的な刺激および/または環境刺激に対する過剰な反応を表している他の疾患が含まれる。具体的には、本発明は、喘息および慢性閉塞性肺疾患の診断方法、より具体的には喘息の診断方法を提供する。

20

【0233】

〔0131〕 また、本発明は非小細胞肺癌の診断方法も提供する。この方法は、本明細書に記載の複数のバイオマーカのバイオマーカ測定値を決定する工程を含み、ここでバイオマーカは、非小細胞肺癌の有無または発達を示すものである。例えば、本明細書に記載のバイオマーカのバイオマーカ測定値を使用して、非小細胞肺癌の進行の程度、前癌病変の有無、または非小細胞肺癌の期を決定することができる。

30

【0234】

〔0132〕 具体的な態様では、対象は、非小細胞肺癌または反応性気道疾患の 1 つまたは複数の症状を呈している個人から選択される。症状には、咳、息切れ、喘鳴、胸痛、および咯血；腕の外側を走る肩痛または嘔声を引き起こす声帯の麻痺；嚥下を難しくする可能性のある食道の進入、が含まれ得る。大気道が閉塞すると、肺の一部が虚脱して感染を生じ、その結果、膿瘍または肺炎が引き起こされる可能性がある。骨への転移は強い痛みを生じ得る。脳への転移は、霧視、頭痛、痙攣、または、一般的に脳卒中に関連する症状、例えば身体の各部における知覚の弱まりまたは消失などの神経症状を引き起こす場合がある。肺癌では、腫瘍細胞からのホルモン様物質の生産による症状が引き起こされることが多い。NSCLC で多く見られる腫瘍随伴性症候群には、血中のカルシウムを上昇させる、副甲状腺ホルモン様物質の生産がある。喘息は通常、例えば咳、特に夜間の、喘鳴、息切れおよび胸苦しさ、疼痛または圧迫の症状を引き起こす。従って、喘息の症状の多くが NSCLC の症状と共通していることが明らかである。

40

【0235】

反応性気道疾患を診断する方法

〔0133〕 本発明は、以下に記載する様々な集団中に含まれている、個々のヒトの反応性気道疾患を診断する方法に向けられている。これらの方法は通常、本明細書に記載したように、特定のバイオマーカのバイオマーカ測定値を決定すること、および分類システム、例えばサポートベクトルマシンを使用して、バイオマーカ測定値を分類することに依存している。

50

## 【0236】

## A．母集団の決定

〔0134〕 本発明は、対象の反応性気道疾患を診断する方法を提供し、この方法は、(a)対象の生理学的試料を収集する工程；(b)前記試料中の複数のバイオマーカのバイオマーカ測定値を本明細書に記載したように決定する工程；および(c)分類システムを利用し、バイオマーカ測定値に基づいて試料を分類する工程を含み、ここで試料の分類が、対象における反応性気道疾患を示すものとなる。

## 【0237】

〔0135〕 好ましい態様では、本発明は、対象の反応性気道疾患を診断する方法を提供し、この方法は、対象の生理学的な試料中の複数のバイオマーカのバイオマーカ測定値を決定する工程を含み、ここで複数のマーカの発現パターンが、反応性気道疾患または反応性気道疾患の段階に関連する変化を示すものとなる。好ましくは、複数のバイオマーカは、機械学習アルゴリズム、例えばサポートベクトルマシンを介した学習データの解析に基づいて選択される。学習データには、多数の対象についての複数のバイオマーカ、ならびに個々の対象に関する疾患の分類、および場合により、対象の他の特徴、例えば性別、人種、年齢、喫煙歴、職歴などが含まれる。別の好ましい態様では、バイオマーカの発現パターンと、対象が反応性気道疾患を有しているまたは有する可能性があることについての見込みの高さには相関がある。発現パターンは、パターン認識について当該分野で公知のいかなる技術によって特徴付けてもよい。複数のバイオマーカには、実施例1に記載したバイオマーカのいかなる組み合わせも含まれ得る。

10

20

## 【0238】

〔0136〕 少なくとも1つの態様では、対象は、反応性気道疾患のリスクを有している。一態様では、反応性気道疾患と関連がある特定のバイオマーカのバイオマーカ測定値が個人について決定され、正常集団について予測されるレベルとは異なるレベルによって、その個人が「リスクを有している」ことが示唆される。別の態様では、対象は、反応性気道疾患の症状のうちの1つまたは複数呈している個人から選択される。

## 【0239】

## B．男性集団の決定

〔0137〕 本発明は、男性対象の反応性気道疾患を診断する方法を提供する。これらの態様の方法は上述したものと同様であるが、学習データおよび試料の両方で、対象は男性である。

30

## 【0240】

## C．女性集団の決定

〔0138〕 本発明は、女性対象の反応性気道疾患を診断する方法を提供する。これらの態様の方法は上述したものと同様であるが、学習データおよび試料の両方で、対象は女性である。

## 【0241】

## 非小細胞肺癌を診断する方法

〔0139〕 本発明は、以下に記載する様々な集団中に含まれている、個々のヒトの非小細胞肺癌を診断する方法に向けられている。これらの方法は通常、本明細書に記載したように、特定のバイオマーカのバイオマーカ測定値を決定すること、および分類システム、例えばサポートベクトルマシンを使用して、バイオマーカ測定値を分類することに依存している。

40

## 【0242】

## A．母集団の決定

〔0140〕 本発明は、対象の非小細胞肺癌を診断する方法を提供し、この方法は、(a)対象の生理学的試料を収集する工程；(b)前記試料中の複数のバイオマーカのバイオマーカ測定値を、本明細書に記載したように決定する工程；および(c)分類システムを利用し、バイオマーカ測定値に基づいて試料を分類する工程を含み、ここで試料の分類が、対象における非小細胞肺癌有無または発達を示すものとなる。

50

## 【0243】

【0141】 好ましい態様では、本発明は、対象の非小細胞肺癌を診断する方法を提供し、この方法は、対象の生理学的な試料中の複数のバイオマーカのバイオマーカ測定値を決定する工程を含み、ここで複数のマーカの発現パターンが、非小細胞肺癌または非小細胞肺癌病状に関連する変化（すなわち、臨床的ステージまたは診断的ステージ）を示すものとなる。好ましくは、複数のバイオマーカは、機械学習アルゴリズム、例えばサポートベクトルマシンを介した学習データの解析に基づいて選択される。学習データには、多数の対象についての複数のバイオマーカ、測定値ならびに個々の対象に関する疾患の分類、および場合により、対象の他の特徴、例えば性別、人種、年齢、喫煙歴、職歴などが含まれる。別の好ましい態様では、バイオマーカの発現パターンと、対象が非小細胞肺癌を有しているまたは有する可能性があることについての見込みの高さには相関がある。発現パターンは、パターン認識について当該分野で公知のいかなる技術によって特徴付けてもよい。複数のバイオマーカには、実施例1に記載したバイオマーカのいかなる組み合わせも含まれ得る。

10

## 【0244】

【0142】 一態様では、対象は、非小細胞肺癌のリスクを有している。別の態様では、対象は、非小細胞肺癌の症状のうちの1つまたは複数呈している個人から選択される。

## 【0245】

## B．男性集団の決定

【0143】 本発明は、男性対象の非小細胞肺癌を診断する方法を提供する。これらの態様の方法は上述したものと同様であるが、学習データおよび試料の両方で、対象は男性である。

20

## 【0246】

## C．女性集団の決定

【0144】 本発明は、女性対象の非小細胞肺癌を診断する方法を提供する。これらの態様の方法は上述したものと同様であるが、学習データおよび試料の両方で、対象は女性である。

## 【0247】

## 非小細胞肺癌と反応性気道疾患を区別する方法

【0145】 本発明は、以下に記載する様々な集団中に含まれている、個々のヒトの肺疾患を診断する方法に向けられている。これらの方法は通常、反応性気道疾患と非小細胞肺癌を示すものを区別する特定のバイオマーカのバイオマーカ測定値を決定すること、および分類システム、例えばサポートベクトルマシンを使用して、バイオマーカ測定値を分類することに依存している。

30

## 【0248】

## A．母集団の決定

【0146】 本発明は、対象の肺疾患を診断する方法を提供し、この方法は、前記対象中の複数のバイオマーカのバイオマーカ測定値を決定する工程を含み、ここで前記複数のバイオマーカのバイオマーカ測定値は、反応性気道疾患と非小細胞肺癌を示すものの区別を補助するものである。一態様において対象は、反応性気道疾患および/または非小細胞肺癌を罹患していると診断されている対象である。例えば、その診断は、対象の生理学的な試料中の少なくとも1つのバイオマーカのバイオマーカ測定値によって行われており、ここで少なくとも1つのバイオマーカのバイオマーカ測定値は、反応性気道疾患および/または非小細胞肺癌を示すものである。

40

## 【0249】

【0147】 また、本発明は、(a)対象の生理学的試料を収集する工程；(b)前記試料中の反応性気道疾患と非小細胞肺癌を示すもの、反応性気道疾患を示すものとなる複数のバイオマーカ、および非小細胞肺癌を示すものとなる複数のバイオマーカの区別を補助する複数のバイオマーカのバイオマーカ測定値を、本明細書に記載したよう

50

に決定する工程、ここで前記複数のバイオマーカーは同一のものではなく；(c)分類システムを利用し、バイオマーカー測定値に基づいて試料を分類する工程、ここで試料の分類は、(i)反応性気道疾患と非小細胞肺癌を示すもの；(ii)反応性気道疾患の有無を示すもの；および(iii)対象における非小細胞肺癌の有無を示すものの区別を補助するものであり；ならびに(d)どの条件が3つの分類のうち2つに見られるかに応じて、対象が、(1)反応性気道疾患に罹患している；(2)非小細胞肺癌に罹患している；または(3)疾患に罹患していないことを決定する工程を含む、対象の肺疾患を診断する方法も提供する。

【0250】

〔0148〕 好ましくは、複数のバイオマーカーは、機械学習アルゴリズム、例えばサポートベクトルマシンを介した学習データの解析に基づいて選択される。学習データには、多数の対象についての複数のバイオマーカー測定値、ならびに個々の対象に関する疾患の分類、および場合により、対象の他の特徴、例えば性別、人種、年齢、喫煙歴、職歴などが含まれる。好ましい態様では、発現パターンと、対象が非小細胞肺癌または反応性気道疾患を有していることについての見込みの高さには相関がある。発現パターンは、パターン認識について当該分野で公知のいかなる技術によって特徴付けてもよい。複数のバイオマーカーには、実施例1に記載したバイオマーカーのいかなる組み合わせも含まれ得る。

10

【0251】

〔0149〕 一態様では、対象は、非小細胞肺癌および/または反応性気道疾患のリスクを有している。別の態様では、対象は、非小細胞肺癌および/または反応性気道疾患の症状のうち1つまたは複数呈している個人から選択される。

20

【0252】

〔0150〕 本発明はまた、対象が、非小細胞肺癌または反応性気道疾患を発症するまたはそれらに罹患するリスクに関する見込みの区別を補助する診断法も提供し、この方法は、(a)非小細胞肺癌または反応性気道疾患のリスクを有している対象の生理学的試料を収集する工程、(b)前記対象が非小細胞肺癌または反応性気道疾患を有するリスクの見込みの区別を補助する、前記試料中の、前記対象における複数のバイオマーカーのバイオマーカー測定値を、本明細書に記載したように決定する工程；(c)分類システムを利用し、バイオマーカー測定値に基づいて試料を分類する工程、ここで試料の分類は、(i)反応性気道疾患と非小細胞肺癌を示すもの；(ii)反応性気道疾患の有無；および(iii)対象における非小細胞肺癌の有無を示すものの区別を補助するものであり；ならびに(d)どの条件が3つの分類のうち2つに見られるかに応じて、対象が、(1)反応性気道疾患の発症または罹患のリスクを有している；(2)非小細胞肺癌の発症または罹患のリスクを有している；または(3)疾患に罹患していないことを決定する工程を含む。

30

【0253】

〔0151〕 好ましくは、複数のバイオマーカーは、機械学習アルゴリズム、例えばサポートベクトルマシンを介した学習データの解析に基づいて選択される。学習データには、多数の対象についての複数のバイオマーカー測定値、ならびに個々の対象に関する疾患の分類、および場合により、対象の他の特徴、例えば性別、人種、年齢、喫煙歴、職歴などが含まれる。好ましい態様では、発現パターンと、対象が非小細胞肺癌または反応性気道疾患を有していることについての見込みの高さには相関がある。発現パターンは、パターン認識について当該分野で公知のいかなる技術によって特徴付けてもよい。複数のバイオマーカーには、実施例1に記載したバイオマーカーのいかなる組み合わせも含まれ得る。

40

【0254】

〔0152〕 一態様では、対象は、非小細胞肺癌または反応性気道疾患の症状のうち1つまたは複数呈している個人から選択される。「リスクを有している」対象に関連する方法については上述されており、かつ、本明細書ではそれらに関する方法が企図され

50

る。

【0255】

B．男性集団の決定

〔0153〕 本発明は、男性対象の肺疾患を診断する方法を提供する。これらの態様の方法は上述したものと同様であるが、学習データおよび試料の両方で、対象は男性である。

【0256】

C．女性集団の決定

〔0154〕 本発明は、女性対象の肺疾患を診断する方法を提供する。これらの態様の方法は上述したものと同様であるが、学習データおよび試料の両方で、対象は女性である。

【0257】

特徴付けに関するシステムの設計方法

A．母集団

〔0155〕

本発明はさらに、対象の肺疾患を診断するためのシステムの設計方法を提供し、この方法は、(a)複数のバイオマーカを選択する工程；(b)前記複数のバイオマーカーのバイオマーカー測定値を決定する手段を選択する工程；および(c)バイオマーカー測定値を決定するための前記手段と対象が肺疾患に罹患しているという見込みを決定するためのバイオマーカー測定値を解析するための手段とを含むシステムを設計する工程、を含む。

【0258】

〔0156〕 本発明はまた、非小細胞肺癌を診断するためのシステムの設計方法を提供し、この方法は、(a)複数のバイオマーカを選択する工程；(b)前記複数のバイオマーカーのバイオマーカー測定値を決定するための手段を選択する工程；および(c)バイオマーカー測定値を決定するための前記手段および対象が非小細胞肺癌に罹患しているという見込みを決定するための、バイオマーカー測定値を解析するための手段を含むシステムを設計する工程、を含む。

【0259】

〔0157〕 本発明はまた、対象の反応性気道疾患を診断するためのシステムの設計方法を提供し、この方法は、(a)複数のバイオマーカを選択する工程；(b)前記複数のバイオマーカーのバイオマーカー測定値を決定するための手段を選択する工程；および(c)バイオマーカー測定値を決定するための前記手段および対象が反応性気道疾患に罹患しているという見込みを決定するための、バイオマーカー測定値を解析するための手段を含むシステムを設計する工程、を含む。

【0260】

〔0158〕 本発明はまた、対象の非小細胞肺癌または反応性気道疾患を診断するためのシステムの設計方法を提供し、この方法は、(a)複数のバイオマーカを選択する工程；(b)前記複数のバイオマーカーのバイオマーカー測定値を決定する手段を選択する工程；および(c)バイオマーカー測定値を決定するための前記手段および対象が反応性気道疾患に罹患しているという見込みを決定するための、バイオマーカー測定値を解析するための手段を含むシステムを設計する工程、を含む。好ましい方法では、複数のバイオマーカーは、非小細胞肺癌を示すものとなるバイオマーカー、反応性気道疾患を示すものとなるバイオマーカー、および非小細胞肺癌と反応性気道疾患との区別を補助するバイオマーカーを含む。

【0261】

〔0159〕 上述した方法において工程(b)および(c)は、(b)前記複数のバイオマーカーを検出するための検出剤を選択する工程、および(c)複数のバイオマーカーを検出するための前記検出剤を含むシステムを設計する工程、によって行ってもよく、あるいはこれらの工程によって行われる。

10

20

30

40

50

## 【0262】

## B．男性集団

〔0160〕 本発明はまた、男性対象の肺疾患の診断を補助するためのシステムの設計方法を提供する。これらの態様の方法は上述したものと同様である。

## 【0263】

## C．女性集団

〔0161〕 本発明はまた、女性対象の肺疾患の診断を補助するためのシステムの設計方法を提供する。これらの態様の方法は上述したものと同様である。

## 【0264】

## キット

〔0162〕 本発明は、本明細書に記載の複数のバイオマーカのバイオマーカ測定値を決定するための手段を含むキットを提供する。本発明はさらに、本明細書に記載の複数のバイオマーカを検出するための検出剤を含むキットも提供する。

## 【0265】

〔0163〕 複数のバイオマーカには、非小細胞肺癌を示すものとなるバイオマーカ、反応性気道疾患を示すものとなるバイオマーカ、および/または非小細胞肺癌と反応性気道疾患との区別を補助するバイオマーカが含まれ得る。好ましくは、これらのバイオマーカは、本明細書に記載の方法によって決定された、少ない数のバイオマーカのセットである。

## 【0266】

〔0164〕 本発明はまた、(a)非小細胞肺癌を示すものとなる複数のバイオマーカのバイオマーカ測定値を決定するための第一の手段；および(b)反応性気道疾患を示すものとなる複数のバイオマーカのバイオマーカ測定値を決定するための第二の手段を含むキットを提供し、ここで(a)および(b)における前記バイオマーカは同一のものではない。

## 【0267】

〔0165〕 本発明はまた、(a)非小細胞肺癌を示すものとなる複数のバイオマーカを検出するための検出剤；および(b)反応性気道疾患を示すものとなる複数のバイオマーカを検出するための検出剤を含むキットを提供し、ここで(a)および(b)における前記バイオマーカは同一のものではない。

## 【0268】

〔0166〕 本発明はまた、(a)非小細胞肺癌を示すものとなる複数のバイオマーカのバイオマーカ測定値を決定するための第一の手段；(b)反応性気道疾患を示すものとなる複数のバイオマーカのバイオマーカ測定値を決定するための第二の手段；および(c)非小細胞肺癌と反応性気道疾患との区別を補助する複数のバイオマーカのバイオマーカ測定値を決定するための第三の手段を含むキットを提供し、ここで(a)、(b)、および(c)における前記バイオマーカは同一のものではない。

## 【0269】

〔0167〕 本発明はまた、(a)非小細胞肺癌を示すものとなる複数のバイオマーカを検出するための検出剤；(b)反応性気道疾患を示すものとなる複数のバイオマーカを検出するための検出剤；および(c)複数の非小細胞肺癌と反応性気道疾患との区別を補助するバイオマーカを検出するための検出剤を含むキットを提供し、ここで(a)、(b)、および(c)における前記バイオマーカは同一のものではない。

## 【0270】

〔0168〕 当然のことながら、本発明が、特定の複数のバイオマーカを検出することを必要とするいかなる方法のために、上述したバイオマーカのいかなる特定の組み合わせを検出するための手段を含むキットを想定していることが理解される。

## 【0271】

## システム

〔0169〕 本発明は、本発明の方法の実施を補助するシステムを提供する。例示的な

10

20

30

40

50

システムには、学習データセットおよび/または試験データセットを記憶するための記憶装置、および学習機械、例えばアダプスト分類器またはSVMを実行するためのコンピューターが含まれる。コンピューターは、データベースから学習データセットを収集するため、学習データセットを前処理するため、前処理した試験データセットを使用して学習機械を訓練するため、および、訓練した学習機械の試験出力の受信に应答して、試験出力が最適解かどうかを決定するために、試験出力を後処理するためにも操作可能であってよい。そのような前処理には、例えば、明かに誤っている記載を検出・削除するために目視で点検すること、適切な標準量で割ることによってデータを正規化すること、およびそれぞれのアルゴリズムで使用するために、適切な形式のデータを確保することが含まれ得る。例示的なシステムにはまた、手元のない提供源から試験データセットと学習データセットを受信するためのコミュニケーション装置を含んでいてもよい。そのような場合には、コンピューターは、学習データセットを前処理する前に、学習データセットを記憶装置に記憶するために、また、試験データセットを前処理する前に、試験データセットを記憶装置に記憶するために、操作可能であってよい。例示的なシステムはまた、後処理した試験データを表示するための表示装置を含んでいてもよい。例示的なシステムのコンピューターはさらに、上述したさらなるそれぞれの機能を実行するために操作可能であってよい。

10

## 【0272】

〔0170〕 本明細書で使用する場合、用語「コンピューター」は、少なくとも1つのメモリーを使用する、少なくとも1つのハードウェアプロセッサを含むと理解される。少なくとも1つのメモリーは、一連の指示を記憶し得る。指示は、メモリーまたはコンピューターのメモリーに、恒久的にまたは一時的に記憶させることができる。プロセッサは、データを処理するために、メモリーに記憶されている指示を実行する。一連の指示には、特定のタスク、例えば本明細書に記載のタスクを実施する、様々な指示が含まれ得る。そのような、特定のタスクを実施するための一連の指示は、プログラム、ソフトウェアプログラム、または単にソフトウェアとして特徴付けられ得る。

20

## 【0273】

〔0171〕 上述したように、コンピューターは、メモリーに記憶されている指示を実行して、データを処理する。このデータの処理は、例えば、使用者またはコンピューターの使用者によるコマンドに应答したもの、その前の処理に应答したもの、別のコンピューターおよび/または他の任意の入力による要求に应答したものとなる可能性がある。

30

## 【0274】

〔0172〕 態様を少なくとも部分的に実装するために使用されるコンピューターは、汎用コンピューターであってよい。しかしながら、コンピューターは、専用コンピューター、マイクロコンピューターを含むコンピューターシステム、ミニコンピューターまたはメインフレーム、例えば、プログラムしたマイクロプロセッサ、マイクロコントローラー、周辺用集積回路要素、CSIC(顧客仕様集積回路)もしくはASIC(特定用途向け集積回路)または他の集積回路、論理回路、デジタルシグナルプロセッサ、プログラム可能な論理装置、例えばFPGA、PLD、PLAもしくはPAL、あるいは、本発明の過程の少なくともいくつかの工程を実装することが可能な他の任意の装置もしくは構成などの、多種多様な他の技術のいずれを使用してもよい。

40

## 【0275】

〔0173〕 当然のことながら、本発明の方法を実践するために、プロセッサおよび/またはコンピューターのメモリーが、物理的に同じ地理学上の場所にある必要はない。つまり、コンピューターによって使用されるプロセッサおよびメモリーはそれぞれ、地理学上の別の位置にあってもよく、かつ、任意の適切な方法によってコミュニケーションするように接続されていてもよい。加えて、プロセッサおよび/またはメモリーはそれぞれ、設備の物理的に異なる部品から構成されていてもよい。従ってプロセッサは、1箇所にある、単一の部品の設備であって、かつ、メモリーが別の場所にある、別の単一の部品の設備である必要もない。つまり、例えば、プロセッサが2箇所の物理的に別の場所

50

にある、2つ以上の部品の設備であってもよいことが想定される。設備の2つ以上の別個の部品は、適切な様式で、例えばネットワークで接続され得る。加えて、メモリーは、2箇所以上の物理的位置にある2つ以上の部分を含んでいてもよい。

【0276】

〔0174〕 様々なコンピューター、プロセッサおよび/またはメモリー間のコミュニケーションを提供するために、ならびに本発明のプロセッサおよび/またはメモリーを任意の他の実体とコミュニケーションさせるために、例えば、さらなる指示を得るか、または、例えば遠隔にあるメモリー記憶にアクセスし、それらを使用するために、様々な技術を使用してもよい。そのようなコミュニケーションを提供するのに使われるそのような技術には、ネットワーク、インターネット、イントラネット、エクストラネット、LAN、イーサネット、または例えば、コミュニケーションを提供する任意のクライアントサーバーシステムが含まれるだろう。そのようなコミュニケーション技術は、任意の好適なプロトコル例えばTCP/IP、UDP、またはOSIなどを使用し得る。

10

【0277】

〔0175〕 さらに、本発明の実装および操作に使用されるコンピューターの指示または一連の指示は、コンピューターがその指示を読み込むのに適した形式である。

〔0176〕 いくつかの態様では、使用者が、態様を少なくとも部分的に実装するのに使用されるコンピューターまたは機械と連動することを可能にするための、様々なユーザーインターフェースを使用することができる。ユーザーインターフェースは、ダイアログスクリーンの形式であってもよい。ユーザーインターフェースはまた、マウス、タッチスクリーン、キーボード、音声読み取り装置、音声認識装置、ダイアログスクリーン、メニューボックス、リスト、チェックボックス、トグルスイッチ、押しボタン、または、コンピューターが一連の指示を処理するのに連れて、使用者がコンピューターの操作に関する情報を受信することを可能にする、および/または使用者がコンピューターに情報を提供することを可能にする他のいかなる装置を含んでいてもよい。従って、ユーザーインターフェースは、使用者とコンピューターとの間のコミュニケーションを提供する、任意の装置である。ユーザーインターフェースを解して使用者がコンピューターに提供する情報は、例えば、コマンド、データ選択、または他のいくつかの入力の形式であってもよい。

20

【0278】

〔0177〕 例えば、情報を伝搬・受信するために、本発明のユーザーインターフェースが、ヒト使用者と交流する以上に、別のコンピューターと交流するかもしれないことも想定される。従って、その他のコンピューターはユーザーとして特徴付けられる可能性もある。さらに、本発明のシステムおよび方法に含まれるユーザーインターフェースが、部分的には別のコンピューターと交流し、同時に、部分的にはヒト使用者と交流することも想定される。

30

【実施例】

【0279】

〔0178〕 以下に記載する実施例は、本明細書で開示する本発明の様々な方法を例示するために提供するものであり、本発明を多少なりとも限定する目的で提供されるものではない。

40

【0280】

実施例1 - データの収集とサポートベクトルマシンを使用した解析

試料の採取

〔0179〕 ヒトの血液試料を志願者から採取した。非小細胞肺癌または喘息を罹患しているか分かっていない個人から、288の試料を採取した。これらの試料は正常な集団を含み、本明細書では「正常集団」と呼ぶ。喘息に罹患していることが分かっており、そのように医師によって診断された個人から180の血液試料を採取した。これらの試料は喘息の集団を含み、本明細書では「喘息集団」と呼ぶ。非小細胞肺癌に罹患していることが分かっており、そのように医師によって診断された個人から360の血液試料を採取した。これらは肺癌の集団を含み、本明細書では「肺癌集団」と呼ぶ。試料の人口統計学

50

的および条件を以下の表に示す。

【 0 2 8 1 】

【 表 1 - 1 】

人口統計学的データ		正常 (N=288)	喘息 (N=180)	癌 (N=360)	計 (N=828)
年齢(才)	範囲	18~60 (N=87)	38~80 (N=178)	38~92 (N=354)	18~92 (N=619)
性別	男性	165	67	245	477
	女性	122	111	114	347
	不明	1	2	1	4
人種	白人	5	168	222	395
	アフリカ系アメリカ人	82	0	0	82
	ヒスパニック	0	8	4	12
	アジア系	0	2	3	5
	不明	201	2	131	334

10

20

【 0 2 8 2 】

【 0 1 8 0 】 発現レベルの変化が肺癌または喘息と関連するだろうと考えられるバイオマーカーを選択するための研究を行った。本明細書で使用する場合、「肺癌」は、非小細胞肺癌として知られる肺癌を包含することを意味する。研究、方法論および得られたデータはこれ以降に記載され、かつ、参照によりその全体が本明細書に組み込まれる国際公開第 2 0 1 0 / 1 0 5 2 3 5 号パンフレットに示されている。

30

【 0 2 8 3 】

【 0 1 8 1 】 以下に示す 5 9 種類のバイオマーカーを選択し、これらについて試験を行った：CD 4 0、肝細胞増殖因子（「HGF」）、I - T A C（「CXCL11」；「ケモカイン（C - X - Cモチーフ）リガンド11」、インターフェロン誘導T細胞化学誘引物質）、レプチン（「LEP」）、マトリックスメタロプロテアーゼ（「MMP」）1、MMP 2、MMP 3、MMP 7、MMP 8、MMP 9、MMP 1 2、MMP 1 3、CD 4 0可溶性リガンド（「CD 4 0リガンド」）、上皮増殖因子（「EGF」）、エオタキシン（「CCl11」）、フラクタルキン、顆粒球コロニー刺激因子（「G - CSF」）、顆粒球マクロファージコロニー刺激因子（「GM - CSF」）、インターフェロン（「IFN」）、インターロイキン（「IL」）1、IL - 1、IL - 1 r a、IL - 2、IL - 4、IL - 5、IL - 6、IL - 7、IL - 8、IL - 1 0、IL - 1 2（p 4 0）、IL - 1 2（p 7 0）、IL - 1 3、IL - 1 5、IL - 1 7、IP - 1 0、単球走化性タンパク質 1（「MCP - 1」）、マクロファージ炎症性タンパク質（「MIP」）1、MIP - 1、トランスフォーミング増殖因子（「TGF」）、腫瘍壊死因子（「TNF」）、血管内皮成長因子（「VEGF」）、インスリン（「Ins」）、C - ペプチド、グルカゴン様タンパク質 - 1 / アミリン（「GLP - 1 / アミリン」）、アミリン（全）、グルカゴン、アディポネクチン、プラスミノゲン活性化

40

50

因子阻害因子1(「PAI-1」;「セルピン」)(活性型/全)、レジスチン(「RETN」;「xcp1」)、sFas、可溶性Fasリガンド(「sFasL」)、マクローファージ遊走性阻止因子(「MIF」)、sE-セレクチン、可溶性血管細胞接着分子(「sVCAM」)、可溶性細胞間接着分子(「sICAM」)、ミエロペルオキシダーゼ(「MPO」)、C反応性タンパク質(「CRP」)、血清アミロイドA(「SAA」;「SAA1」)、および血清アミロイドP(「SAP」)。

【0284】

データ収集

〔0182〕 血漿標本を、自動化され、ビーズを使用した定量的なマルチプレックス免疫アッセイであるLuminexのxMAP技術にかけることによって、正常、喘息および肺癌それぞれの集団の血漿標本について、59種類のバイオマーカーのスクリーニングを行った。

10

【0285】

〔0183〕 バイオマーカーのスクリーニングには、LuminexのxMAP技術と共に、複数のその他のアッセイキット、すなわちMilliporeのヒトサイトカイン/ケモカイン(カタログ番号MPXHCYT0-60K、ヒトエンドクリン(カタログ番号HENDO-65K)、ヒト血清アディポカイン(カタログ番号HADKI-61K)、ヒトセプシス/アポトーシス(カタログ番号HSEP-63K)、ヒト心血管パネル1(カタログ番号HCVD1-67AK)およびヒト心血管パネル2(HCVD2-67BK)、R&D Systems, IncのヒトフルオロカインMAPプロファイリングベースキットB(カタログ番号LUB00)およびヒトフルオロカインMAP MMPプロファイリングベースキット(カタログ番号LMP000)を使用した。PanomicsのProcartaサイトカインキット(カタログ番号PC1017)も使用した。PAI-1およびレプチンに関する抗体は、2種類のキットのものを使用した。PAI-1<sup>A</sup>およびLeptin<sup>1</sup>についての抗体はMilliporeによって製造されているものを使用した。PAI-1<sup>B</sup>についての抗体はPanomics製ものを使用した。

20

【0286】

〔0184〕 マルチプレックス免疫アッセイで生じた蛍光強度レベルを、それぞれの集団の各血漿標本についての59種類のバイオマーカーに関するバイオマーカー測定値として記録した。記録された蛍光強度は、試料中の対応するバイオマーカーの濃度に比例し、かつ、試料が採取された時点での個人におけるその発現の割合にも比例している。それぞれの集団の各バイオマーカーと関連する蛍光強度レベルの平均、標準偏差、および相対的標準偏差を計算した。図1Aから1Cに、正常(NO)、非小細胞肺癌(LC)、および喘息(AST)集団における各バイオマーカー測定値の平均、標準偏差、および相対的標準偏差を示し、図1Dにはこれらの集団のうちの2つの集団における、特定のバイオマーカー測定値のレベルの変化の平均を示す。

30

【0287】

〔0185〕 得られたデータを性別によっても分けた。

〔0186〕 図2A~2Cでは、正常(NO)、非小細胞肺癌(LC)、および喘息(AST)の女性集団におけるバイオマーカーの蛍光強度レベルの平均を示している。図2Dは、喘息対正常の女性集団、肺癌対正常の女性集団、および喘息対肺癌の女性集団における各バイオマーカー測定値の平均の変化(%)を示す。

40

【0288】

〔0187〕 男性集団に関する同様の情報を図3A~3Dに示す。

〔0188〕 次に、女性集団と男性集団とを比較した。図4では、喘息の女性集団と比較した喘息の男性、肺癌の女性集団と比較した肺癌の男性集団、および正常な女性集団と比較した正常な男性集団における、各バイオマーカー測定値の平均の変化(%)を示している。

【0289】

〔0189〕 Luminexアッセイから得られたデータを、医師の診断に基づく、

50

患者の経験的な分類によって同定された、特定の患者の試料における各バイオマーカーの蛍光強度のデータと共に、データ記憶装置に電子的に記録した。

【0290】

データ解析

〔0190〕 サポートベクトルマシンアルゴリズムを介したデータの解析は、以下の工程によって遂行された。

1. 記憶装置から処理装置へデータセットを読み込んだ。
2. データを前処理し、モデル選択アルゴリズムおよびサポートベクトルマシンで使用するのに適したものにした。
3. データを無作為に2群、つまり学習セットと確認セットに分けた。
4. 学習データセットについてサポートベクトルマシンアルゴリズムを実行し、モデルを生成した。本明細書に記載の全てのSVM統計的計算は、統計処理用のソフトウェアプログラム用言語・環境であるR 2.10.0 (www.r-project.org) を使用して行った。kermlabパッケージに含まれているksvm()関数を使ってSVMを当てはめた。
5. 前の工程で生成したモデルを使って確認データセットを後処理し、予測分類を算出した。予測分類を、試験セット試料の経験的な分類と比較し、モデルフィットの指標、例えば精度、感度、特異度、陽性予測値および陰性予測値を算出した。ここで感度とは、対象が病気の場合にその対象が病気であると予測される確率であり、特異度とは、対象が病気でない場合にその対象が病気でないとして予測される確率であり、陽性予測値とは、対象が病気であると予測される場合にその対象が病気である確率であり、陰性予測値とは、対象が病気でないとして予測される場合にその対象が病気でない確率であり、および精度とは、正しい予測の確率である。

10

20

30

40

【0291】

〔0191〕 787の対象が全データを有した。これらのデータは解析にのみ使用した。学習セットは398対象、試験セットは389対象であった。

〔0192〕 完全なデータセットについて解析を行った場合には、389対象のうち344対象が正しく分類され、その場合の精度は0.88 (SE: 0.017) であった(表1を参照のこと)。癌対象では、その他の対象との比較でみると、サポートベクトルマシンの感度は0.98 (SE: 0.007) であり、陰性予測値は0.99 (SE: 0.008) であった(表2を参照のこと)。

【0292】

【表1-2】

表1. モデル1 (全クラス) : 分割表 (N=389)

予測される肺 病理の分類	実際の肺病理の分類			計	精度(SE <sup>1</sup> )
	喘息	癌	正常		
喘息	79	4	1	81	0.88(0.017)
癌	13	148	12	173	
正常	4	14	117	135	
計	93	166	130	389	

<sup>1</sup>標準誤差

【0293】

## 【表 2】

表 2. モデル1 (全クラス) : 性能統計 (N=389)

病理 <sup>1</sup>	統計項目	評価(SE <sup>2</sup> )	95%信頼区間	統合 95%信頼区間
喘息	真の陽性の割合 (TPF または感度)	0.86(0.018)	(0.78, 0.92)	(0.77, 0.93)×(0, 0.02) <sup>3</sup>
	偽陽性の割合(FPF)	0.01(0.004)	(0, 0.02)	
	特異度(1-FPF)	0.99(0.004)	(0.98, 1)	(0.77, 0.93)×(0.98, 1) <sup>4</sup>
	陽性予測値(PPV)	0.98(0.008)	(0.92, 0.99)	(0.92, 0.99)×(0.93, 0.98) <sup>5</sup>
	陰性予測値(NPV)	0.96(0.01)	(0.93, 0.98)	
癌	真の陽性の割合 (TPF または感度)	0.98(0.007)	(0.95, 0.99)	(0.95, 1)×(0.04, 0.11) <sup>3</sup>
	偽陽性の割合(FPF)	0.08(0.014)	(0.05, 0.12)	
	特異度(1-FPF)	0.92(0.014)	(0.88, 0.95)	(0.95, 1)×(0.88, 0.95) <sup>4</sup>
	陽性予測値(PPV)	0.9(0.015)	(0.85, 0.94)	(0.85, 0.94)×(0.97, 1) <sup>5</sup>
	陰性予測値(NPV)	0.99(0.006)	(0.96, 1)	
正常	真の陽性の割合 (TPF または感度)	0.95(0.011)	(0.9, 0.98)	(0.89, 0.98)×(0, 0.03) <sup>3</sup>
	偽陽性の割合(FPF)	0.01(0.005)	(0, 0.03)	
	特異度(1-FPF)	0.99(0.005)	(0.97, 1)	(0.89, 0.98)×(0.97, 1) <sup>4</sup>
	陽性予測値(PPV)	0.98(0.008)	(0.94, 0.99)	(0.94, 0.99)×(0.95, 0.99) <sup>5</sup>
	陰性予測値(NPV)	0.97(0.008)	(0.95, 0.99)	

<sup>1</sup>感度、FPFおよび特異度に関し、これが実際の病理である。PPVおよびNPVについては、これは予測した病理である。

<sup>2</sup>標準誤差

<sup>3</sup> (TPF、FPF)

<sup>4</sup> (感度、特異度)

<sup>5</sup> (PPV、NPV)

## 【0294】

## 実施例 2 - 肺癌データの解析

【0193】 喘息対象のデータを除いた実施例 1 からのデータセットにも、サポートベクトルマシンを当てはめた。癌対象および病気でない対象のデータのみを含むデータセットについて、実施例 1 で説明したように工程 1 ~ 5 を行った。得られたサポートベクトルマシンの感度は 0.92 (SE: 0.016) であり、特異度は 0.92 (SE: 0.015) であった (表 3 および 4 を参照のこと)。

## 【0295】

10

20

30

40

50

## 【表3】

表3. モデル1 (癌対正常) : 分割表 (N=296)

病的分類	実際の肺病理の分類			感度(SE)	特異度(SE)
	癌	正常	計		
癌	153	10	163	0.92(0.016)	0.92(0.015)
正常	13	120	133		
計	166	130	296		

10

<sup>1</sup>標準誤差

【0296】

【表4-1】

表4. モデル1 (癌対正常) : 性能統計 (N=296)

統計項目	予測(SE <sup>2</sup> )	95%信頼区間	統合 95%信頼区間
精度	0.92(0.016)	(0.89, 0.95)	
真の陽性の割合(TPF または感度)	0.92(0.016)	(0.87, 0.96)	(0.87, 0.95)×(0.03, 0.12) <sup>3</sup>
偽陽性の割合(FPF)	0.08(0.015)	(0.04, 0.13)	
特異度(1-FPF)	0.92(0.015)	(0.87, 0.96)	(0.87, 0.96)×(0.87, 0.96) <sup>4</sup>
陽性予測値(PPV)	0.94(0.014)	(0.89, 0.97)	(0.89, 0.97)×(0.86, 0.96) <sup>5</sup>
陰性予測値(NPV)	0.9(0.017)	(0.84, 0.94)	

20

30

<sup>2</sup>標準誤差<sup>3</sup> (TPF、FPF)<sup>4</sup> (感度、特異度)<sup>5</sup> (PPV、NPV)

【0297】

## 実施例3 - 別の試験セットを用いた解析

【0194】 実施例1でLuminoxアッセイから収集したデータを、実施例1で説明した工程1~5を使用して、再度解析した。個々の試料に由来するデータを、新しい学習セットと試験セットに無作為に割り当てた。学習セットには398対象、試験セット

40

【0298】

【0195】 試料が喘息、癌、または正常であるかを予測するのに、実施例1に記載した59種類のバイオマーカーと、併せて性別について考慮した。結果を以下に示す。

【0299】

【表 4 - 2】

予測した肺病	実際の肺病理の分類				
理の分類	喘息	癌	正常	計	精度(SE)
喘息	80	1	0	81	0.94 (0.012)
癌	12	164	9	185	
正常	1	1	121	123	
計	93	165	130	389	

10

全59種類のバイオマーカーと性別

癌、正常、および喘息 (N=389)

## 【0300】

## 実施例 4 - 肺癌データの解析

〔0196〕 喘息対象のデータを除いた実施例 3 からの学習データセットにも、サポートベクトルマシンを当てはめた。サポートベクトルマシン。喘息患者のデータを除いた確認データセットを使用して、実施例 1 で説明したデータ解析プロトコルの工程 1 ~ 5 を行うことで、以下の結果が得られた。

20

## 【0301】

## 【表 4 - 3】

予測した肺病	実際の肺病理の分類				
理の分類	癌	正常	計	感度(SE)	特異度(SE)
癌	165	7	172	0.99 (0.004)	0.95 (0.013)
正常	1	123	124		
計	166	130	296		

30

全59種類のバイオマーカーと性別

癌および正常 (N=296)

## 【0302】

## 実施例 5 - 肺癌データの解析 (男性)

〔0197〕 実施例 3 のデータセットを使用して、男性試料が癌かまたは正常かを予測するために、59種類のバイオマーカーについての検討を行った。データを実施例 1 の5工程プロトコルによって解析した。結果を以下に示す。

40

## 【0303】

【表 4 - 4】

予測した肺病	実際の肺病理の分類			感度(SE)	特異度(SE)
理の分類	癌	正常	計		
癌	116	4	120	1 (0)	0.94 (0.018)
正常	0	62	62		
計	116	66	182		

10

全59種類のバイオマーカー；男性のみ

癌および正常；男性のみ (N=182)

## 【0304】

〔0198〕 次いで、男性試料および女性試料が癌かまたは正常であるかの予測に関し、生成したモデルを検討した。結果を以下に示す。

## 【0305】

【表 4 - 5】

予測した肺病	実際の肺病理の分類			感度(SE)	特異度(SE)
理の分類	癌	正常	計		
癌	164	29	193	0.99 (0.006)	0.78 (0.024)
正常	2	101	103		
計	166	130	296		

20

全59種類のバイオマーカー；男性のみ

癌および正常；男性 (N=182) および女性 (N=114)；計 (N=296)

30

## 【0306】

実施例 6 - 肺癌データの解析 (女性)

〔0199〕 実施例 3 からのデータを使用して、女性試料が癌であるか、または正常であるかの予測について、59種類のバイオマーカーについての検討を行った。実施例 1 の解析プロトコルの工程 1 ~ 5 を女性患者のみからのデータに適用した。結果を以下に示す。

## 【0307】

【表 4 - 6】

予測した肺病	実際の肺病理の分類			感度(SE)	特異度(SE)
理の分類	癌	正常	計		
癌	50	2	52	1 (0)	0.97 (0.016)
正常	0	62	62		
計	50	64	114		

10

全59種類のバイオマーカー；女性のみ

癌および正常；女性のみ (N=114)

## 【0308】

〔0200〕 次に、男性試料および女性試料が癌かまたは正常であるかの予測に関し、同じモデルを検討した。結果を以下に示す。

## 【0309】

## 【表 4 - 7】

予測した肺病	実際の肺病理の分類			感度(SE)	特異度(SE)
理の分類	癌	正常	計		
癌	163	16	179	0.98 (0.008)	0.88 (0.019)
正常	3	114	117		
計	166	130	296		

20

全59種類のバイオマーカー；女性のみ

癌および正常；男性 (N=182) および女性 (N=114)；計 (N=296)

30

## 【0310】

実施例 7 - 選択アルゴリズム (バイオマーカー；癌および正常)

〔0201〕 実施例 1～6の結果は、59種類のバイオマーカーを含むモデルに関するものである。本明細書で議論しているように、選択アルゴリズムを使用して、予測精度を有意に低下させることなく、バイオマーカーの数を減らすことができる。バイオマーカー選択アルゴリズムを実行して、サポートベクトルマシンで使用するバイオマーカーを見つけ出した。

## 【0311】

〔0202〕 上述した8工程のバイオマーカー選択アルゴリズムを使用し、2つの肺病理分類(癌、正常)を特徴付けるための4種類のバイオマーカーモデル(EGF、SCD40リガンド、IL-8、およびMMP-8)を選択した。実施例1からのデータを5工程プロトコールに従って処理した。ただし、工程2の前処理工程では、選択アルゴリズムによって選抜された4種類のバイオマーカー以外のすべてのバイオマーカーを除いた。以下に記載したように、モデルフィットの指標は、95%の精度、93%の感度、および87%の特異度を示した。

40

## 【0312】

【表 4 - 8】

予測した肺病		実際の肺病理の分類			
理の分類	癌	正常	計	感度(SE)	特異度(SE)
癌	155	17	172	0.93 (0.014)	0.87 (0.02)
正常	11	113	124		
計	166	130	296		

10

59種類のバイオマーカーから選択した、4種類の変数 (EGF、sCD40リガンド、IL-8、MMP-8) の下位セット。

癌および正常 (N=296)

## 【0313】

実施例 8 - 選択アルゴリズム (バイオマーカーおよび男性 ; 癌および正常)

【0203】 実施例 7 に記載した、バイオマーカーの数を制限する過程を、実施例 1 からの、男性患者に関する値のみを含む下位セットに適用した。8工程のバイオマーカー選択アルゴリズムを使用して、以下に示すように、100%の精度、100%の感度、および100%の特異度で男性における2つの肺病理分類 (癌および正常) を特徴づけるために、5つのバイオマーカーモデル (EGF、IL-8、Sfas、MMP-9、およびPAI-1) を選択した (もとのデータセットに含まれていた2つの変数をPAI-1と命名した。これはその内の2番目のもので、Panomics社のキットに由来するPAI-1バイオマーカーである)。

20

## 【0314】

## 【表 4 - 9】

予測した肺病		実際の肺病理の分類			
理の分類	癌	正常	計	感度(SE)	特異度(SE)
癌	116	0	116	1(0)	1(0)
正常	0	66	66		
計	116	66	182		

30

59種類のバイオマーカーから選択した、5種類 (EGF、IL-8、Sfas、MMP-9、PAI-1) から構成される下位セット ; 男性のみ

40

癌および正常 ; 男性のみ (N=182)

## 【0315】

【0204】 次に、同様の考察 (すなわち、5種類のバイオマーカーから構成される下位セット、男性) を、男性および女性の試料が癌であるかまたは正常であるかの予測に関して検討した。結果を以下に示す。

## 【0316】

【表 4 - 1 0】

予測した肺病	実際の肺病理の分類			感度(SE)	特異度(SE)
理の分類	癌	正常	計		
癌	157	27	184	0.95 (0.013)	0.79 (0.024)
正常	9	103	112		
計	166	130	296		

10

59種類のバイオマーカーから選択した、5種類（EGF、IL-8、Sfas、MMP-9、PAI-1）から構成される下位セット；男性のみ

癌および正常；男性（N=182）および女性（N=114）；計（N=296）

## 【0317】

実施例 9 - 選択アルゴリズム（バイオマーカーおよび女性；癌および正常）

【0205】 実施例 7 に記載した、バイオマーカーの数を制限する過程を、実施例 1 からの、女性患者に関する値のみを含む下位セットに適用した。8工程のバイオマーカー選択アルゴリズムを使用して、以下に示すように、100%の精度、100%の感度、および100%の特異度で女性における2つの肺病理分類（癌および正常）を特徴づけるために、3つのバイオマーカーモデル（EGF、sCD40リガンド、IL-8）を選択した。

20

## 【0318】

【表 4 - 1 1】

予測した肺病	実際の肺病理の分類			感度(SE)	特異度(SE)
理の分類	癌	正常	計		
癌	50	0	50	1(0)	1(0)
正常	0	64	64		
計	50	64	114		

30

59種類のバイオマーカーから選択した、3種類（EGF、sCD40リガンド、IL-8）から構成される下位セット；女性のみ

癌および正常；女性のみ（N=114）

40

## 【0319】

【0206】 次に、同様の考察（すなわち、3種類のバイオマーカーから構成される下位セット、女性）を、男性および女性の試料が癌であるかまたは正常であるかの予測に関して検討した。結果を以下に示す。

## 【0320】

【表 4 - 1 2】

予測した肺病		実際の肺病理の分類			
理の分類	癌	正常	計	感度(SE)	特異度(SE)
癌	131	7	138	0.79 (0.024)	0.95 (0.013)
正常	35	123	158		
計	166	130	296		

10

59種類のバイオマーカーから選択した、3種類（EGF、sCD40リガンド、IL-8）から構成される下位セット；女性のみ

癌および正常；男性（N=182）および女性（N=114）；計（N=296）

## 【0321】

## 実施例10 - データの収集とアダプストを使用した解析

## データに関する問題

〔0207〕 受信したデータは、実施例1で記載した、Luminexからのバイオマーカー濃度に関する生の出力であった。Luminexから出力されたデータには、蛍光レベル、イベントの数、集計した蛍光レベル、調整した蛍光レベル、正規化したバイオマーカー濃度（この場合正規化とは、観測された蛍光を濃度の標準曲線に当てはめて、観測された蛍光を濃度に変換することを意味する）、集計し、正規化したバイオマーカー濃度、および調整したバイオマーカー濃度が含まれていた。本明細書に記載した解析には、正規化したバイオマーカー濃度を使用した。タンパク質の定量の試験から、試料はタンパク質の総量の点でほぼ矛盾がなく、そのため、バイオマーカーレベルをさらに正規化する必要はないことがわかった（この場合の正規化とは、試料間でのタンパク質レベルの差を調整するために、係数をかけることを意味している）。

20

## 【0322】

〔0208〕 以下に記載する86種類のバイオマーカーのそれぞれについて、バイオマーカーの定量データを収集した：脳由来神経栄養因子（「BDNF」）、Bリンパ球化学誘引物質（「BLC」）、皮膚T細胞誘引ケモカイン（「CTACK」）、エオタキシン-2、エオタキシン-3、グランザイム-B、肝細胞増殖因子（「HGF」）、I-TAC（「CXCL11」；「ケモカイン（C-X-Cモチーフ）リガンド11」、「インターフェロン誘導T細胞化学誘引物質」）、レプチン（「LEP」）、白血病抑制因子（「LIF」）、マクrofアージコロニー刺激因子（「MCSF」）、インターフェロン誘導性モノカイン（「MIG」）、マクrofアージ炎症性タンパク質-3（「MIP-3」）、神経成長因子（「NGF」）、可溶性リガンド（「CD40リガンド」）、上皮増殖因子（「EGF」）、エオタキシン（「CCL11」）、フラクタルキン、塩基性線維芽細胞増殖因子（「FGF-塩基性」）、顆粒球コロニー刺激因子（「G-CSF」）、顆粒球マクrofアージコロニー刺激因子（「GM-CSF」）、インターフェロン（「IFN」）、IFN-、IFN-2、IFN-、インターロイキン（「IL」）1a、IL-1、IL-1ra、IL-2、IL-2ra、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-12（p40）、IL-12（p70）、IL-13、IL-15、IL-16、IL-17、IL-17a、IL-17F、IL-20、IL-21、IL-22、IL-23（p19）、IL-27、IL-31、IP-10、単球走化性タンパク質1（「MCP-1」）、マクrofアージ炎症性タンパク質（「MIP」）1、MIP-1、好中球活性化ペプチド78（「ENA-78」）、オステオプロテゲリン（「OPG」）、胎盤増殖

30

40

50

因子(「PIGF」)、血小板由来成長因子サブユニットBホモ二量体(「PDGFBB」)、T細胞由来の好酸球走化性物質(「RANTES」)、幹細胞成長因子(「SCGF」)、ストロマ細胞由来因子1(「SDF-1」)、可溶性Fasリガンド(「sFas-リガンド」)、NF- $\kappa$ B活性化可溶性受容体リガンド(「sRANKL」)、サバイン、トランスフォーミング増殖因子(「TGF- $\beta$ 」)、TGF- $\beta$ 、腫瘍壊死因子 $\alpha$ (「TNF- $\alpha$ 」)、TNF- $\alpha$ 、TNF受容体1(「TNFR-I」)、TNF関連性アポトーシス誘導性リガンド(「TRAIL」)、血管内皮成長因子(「VEGF」)、アディポネクチン、プラスミノゲン活性化因子阻害因子1(「PAI-1」;「セルピン」)(活性型/全)、レジスチン(「RETN」;「xcp1」)、sFas、可溶性Fasリガンド(「sFasL」)、マクロファージ遊走性阻止因子(「MIF」)、sE-セレクチン、可溶性血管細胞接着分子(「sVCAM」)、可溶性細胞間接着分子(「sICAM」)、ミエロペルオキシダーゼ(「MPO」)、血清アミロイドA(「SAA」;「SAA1」)。

10

## 【0323】

〔0209〕 それぞれの対象につき、計3試料について、86種類のバイオマーカーをそれぞれ測定した。分類する目的では、各試料に分類を割り当てるモデルを作成した。ある対象のいずれかの試料が癌であると見なされた場合には、その患者は癌を患っていると仮定した。病理を決定する方法は、他の可能な方法、例えば投票よりも控えめである。

20

## 【0324】

〔0210〕 検出の上限を超えたバイオマーカー濃度は、検出の上限と等しいと設定した。検出の下限を下回るバイオマーカー濃度は、検出の下限と等しいと設定し、2の平方根で割った。この解はこの問題に限られるものであり、真のバイオマーカー分布の評価に偏りを生じるものではないと考えられる。これは適宜、検出の上限または下限でのバイオマーカーの値の分布に質点を生成する効果を有する。SVMがノンパラメトリックであり、アダプストが一連の木に基づくものであることから、上述したこのその場しのぎの解の欠点は適用されない。性別、年齢、および喫煙は全ての分類モデルに含められた。

## 【0325】

## 結果

〔0211〕 データ：肺病理分類、 $y$ (NSCLC、正常)、および86-タブルの連続的に分布しているバイオマーカー、 $x$ 、が、3重複(計1634試料、癌：546、癌以外：1088)で行った各544人の対象(癌：180、癌以外：364)について使用可能であった。それぞれの試料についてのデータ( $y$ 、 $x$ )を観測と見なす。544対象(1634試料)を学習セット( $N=209$ ;626試料)と確認セット( $N=335$ ;1008試料)とに無作為に分けた(表5)。

30

## 【0326】

## 【表5】

表5：減数したデータベース中の学習データセットおよび確認データセットのサンプルサイズ

## a. サンプル

肺病理：	学習セット	確認セット	全データ
癌、N (%)	188(30.0)	358(35.5)	546(33.4)
正常、N (%)	438(70.0)	650(64.5)	1088(66.6)
計	626	1008	1634

10

## b. 対象

肺病理：	学習セット	確認セット	全データ
癌、N (%)	63(30.1)	117(34.9)	180(33.1)
正常、N (%)	146(69.9)	218(65.1)	364(66.9)
計	209	335	544

20

## 【0327】

〔0212〕 モデル：本試験では、第3a相にSVMおよびアダブーストを使用した。本明細書で示す結果は、全バイオマーカーと人口統計学的情報を使用するモデルに関する結果である（544対象、対象一人当たり3試料の1634試料、および86種類のバイオマーカー）。バイオマーカーのみ、またはバイオマーカー全体のパネルを含む下位セットとモデルについても検討した。SVMと比較するとアダブーストの性能が非常に優れていたため、精力的にアダブーストを探索した。

30

## 【0328】

〔0213〕 統計学的手法：肺病理分類に伴う各バイオマーカーの分布の変化に関する統計的な有意差をクラスカル・ワリス検定で評価した。全ての統計検定は両側検定で、有意差レベルは5%であった。ジェフリーズ法を使用して、両方の割合に関する95%信頼区間の上限と下限を計算した。試料一つ当たりを基準として解析した場合には自己相関を無視し、全ての解析において、複数の比較に関する補正は行わなかった。

## 【0329】

〔0214〕 モデル性能：モデル性能は、確認セット中に含まれている試料に関するモデルの予測を試験すること、または対象レベルでの試料の予測を集計することのいずれかによって決定することができる。試料レベルでの予測を集計するために、ある対象に由来する1つの試料が癌であると予測された場合には、その対象を癌であると予測した。データを集計するには他の方法もあるが、本実施例では、感度（真陽性率としても知られている）および特異度（偽陽性率としても知られている）を最大にする方法を選択した。

40

## 【0330】

〔0215〕 1種類のバイオマーカー（IP-10）を除いて、全てのバイオマーカーで有意な変化が見られた。試料一つ当たりを基準とした性別に関するバイオマーカーの対比からは、22種類のバイオマーカー（アディポネクチン、IL-27、IL-2ra、IL-31、LIF、MPO、PIGF、SCF、sEセレクチン、sFasリガンド、TNFR-II、ENA-78、エオタキシン、フラクタルキン、GCSF、GM-CSF、IL-15、I-TAC、レプチン、MIP-1b、レジスチン、IL-21）

50

が有意な変化を表したことが示された。データベースに含まれている人種（白人、黒人）に関し、試料一つ当たりを基準としてバイオマーカーを生データで対比すると、53種類のバイオマーカーが有意な変化を示したことが明らかとなった。試料一つ当たりを基準として、データベース中の癌対象とそれ以外の対象を対比すると、1種類のバイオマーカー（IP.10）を除き、全てのバイオマーカーで有意な変化が認められた。

【0331】

〔0216〕 試料一つ当たりを基準とした予測に関しては（表6～9）、総体的に、SVMはアダブーストよりも低い性能しか示さなかった〔SVM：感度＝0.78（0.02）、95%信頼区間（0.74、0.82）、特異度＝0.92（0.01）、95%CI（0.90、0.94）、アダブースト：感度＝0.86（0.02）、95%CI（0.82、0.89）、特異度＝0.93（0.01）、95%CI（0.90、0.94）〕。アダブーストの性能は、男性のみ（表10および11）および女性のみ（表12および13）に限定した場合には同程度であった〔男性のみ：感度＝0.87（0.02）、95%CI（0.82、0.91）、特異度＝0.95（0.01）、95%CI（0.92、0.97）、女性のみ：感度＝0.82（0.03）、95%CI（0.76、0.87）、特異度＝0.94（0.01）、95%CI（0.90、0.96）〕。

10

【0332】

【表6】

表6：モデル性能表—SVM—試料毎

20

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	281	49	330	0.78(0.02)	0.92(0.01)
癌以外	77	601	678		
計	358	650	1008		

30

<sup>1</sup>クラスカル・ワリス検定

【0333】

## 【表7】

表7：モデル性能統計－SVM－試料毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.88(0.01)	(0.85,0.89)	
感度	0.78(0.02)	(0.74,0.82)	
擬陽性の割合	0.08(0.01)	(0.06,0.1)	(0.73,0.83)×(0.06,0.1)
特異度	0.92(0.01)	(0.9,0.94)	(0.73,0.83)×(0.9,0.94)
陽性予測値	0.85(0.02)	(0.81,0.89)	
陰性予測値	0.89(0.01)	(0.86,0.91)	

10

<sup>1</sup>クラスカル・ワリス検定

## 【0334】

## 【表8】

20

表8：モデル性能表－アダブースト－全対象－試料毎

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	308	48	356	0.86(0.02)	0.93(0.01)
癌以外	50	602	652		
計	358	650	1008		

30

<sup>1</sup>クラスカル・ワリス検定

## 【0335】

## 【表 9】

表 9：モデル性能統計—アダブースト—全対象—試料毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.9(0.01)	(0.88, 0.92)	
感度	0.86(0.02)	(0.82, 0.89)	
偽陽性の割合	0.07(0.01)	(0.06, 0.1)	(0.81, 0.9)×(0.05, 0.1)
特異度	0.93(0.01)	(0.9, 0.94)	(0.81, 0.9)×(0.9, 0.95)
陽性予測値	0.87(0.02)	(0.83, 0.9)	
陰性予測値	0.92(0.01)	(0.9, 0.94)	

<sup>1</sup>クラスカル・ワリス検定

【 0 3 3 6 】

【表 1 0】

表 1 0：モデル性能表—アダブースト—男性のみ—試料毎

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	164	15	179	0.87(0.02)	0.95(0.01)
癌以外	24	305	329		
計	188	320	508		

<sup>1</sup>クラスカル・ワリス検定

【 0 3 3 7 】

【表 1 1】

表 1 1：モデル性能統計—アダブースト—男性のみ—試料毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.92(0.01)	(0.9, 0.94)	
感度	0.87(0.02)	(0.82, 0.91)	
偽陽性の割合	0.05(0.01)	(0.03, 0.08)	(0.81, 0.92)×(0.03, 0.08)
特異度	0.95(0.01)	(0.92, 0.97)	(0.81, 0.92)×(0.92, 0.97)
陽性予測値	0.92(0.02)	(0.87, 0.95)	
陰性予測値	0.93(0.01)	(0.89, 0.95)	

<sup>1</sup>クラスカル・ワリス検定

【0338】

【表12】

表12：モデル性能表—アダプースト—女性のみ—試料毎

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	140	21	161	0.82(0.03)	0.94(0.01)
癌以外	30	309	339		
計	170	330	500		

10

<sup>1</sup>クラスカル・ワリス検定

【0339】

【表13】

表13：モデル性能表—アダプースト—女性のみ—試料毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.9(0.01)	(0.87,0.92)	
感度	0.82(0.03)	(0.76,0.87)	
偽陽性の割合	0.06(0.01)	(0.04,0.1)	(0.75,0.88)×(0.04,0.1)
特異度	0.94(0.01)	(0.9,0.96)	(0.75,0.88)×(0.9,0.96)
陽性予測値	0.87(0.03)	(0.81,0.91)	
陰性予測値	0.91(0.02)	(0.88,0.94)	

20

30

<sup>1</sup>クラスカル・ワリス検定

【0340】

【0217】 対象一人当たりについての予測（表14～17）に関しては、総体的に、SVMはアダプーストよりも低い性能しか示さなかった[SVM：感度=0.79(0.04)、95%信頼区間(0.70、0.85)、特異度=0.92(0.02)、95%CI(0.88、0.95)、アダプースト：感度=0.87(0.03)、95%CI(0.80、0.92)、特異度=0.93(0.02)、95%CI(0.88、0.96)]。アダプーストの性能は、男性のみ（表18および19）および女性のみ（表20および21）に限定した場合でも同程度であった[男性のみ：感度=0.95(0.02)、95%CI(0.89、0.98)、特異度=0.87(0.04)、95%CI(0.76、0.93)、女性のみ：感度=0.95(0.02)、95%CI(0.90、0.98)、特異度=0.74(0.06)、95%CI(0.61、0.83)]。

40

【0341】

## 【表 1 4】

表 1 4 : モデル性能表—SVM—対象毎

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	92	17	109	0.79(0.04)	0.92(0.02)
癌以外	25	201	226		
計	117	218	335		

10

<sup>1</sup>クラスカル・ワリス検定

## 【0 3 4 2】

## 【表 1 5】

表 1 5 : モデル性能統計—SVM—対象毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.87(0.02)	(0.83, 0.91)	
感度	0.79(0.04)	(0.7, 0.85)	
偽陽性の割合	0.08(0.02)	(0.05, 0.12)	(0.69, 0.86)×(0.05, 0.13)
特異度	0.92(0.02)	(0.88, 0.95)	(0.69, 0.86)×(0.87, 0.95)
陽性予測値	0.84(0.03)	(0.76, 0.9)	
陰性予測値	0.89(0.02)	(0.84, 0.92)	

20

30

<sup>1</sup>クラスカル・ワリス検定

## 【0 3 4 3】

## 【表 1 6】

表 1 6 : モデル性能表—アダブースト—全対象—対象毎

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	102	16	118	0.87(0.03)	0.93(0.02)
癌以外	15	202	217		
計	117	218	335		

40

<sup>1</sup>クラスカル・ワリス検定

## 【0 3 4 4】

## 【表 17】

表 17：モデル性能統計—アダブースト—全対象—対象毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.91(0.02)	(0.87,0.93)	
感度	0.87(0.03)	(0.8,0.92)	
偽陽性の割合	0.07(0.02)	(0.04,0.12)	(0.79,0.93)×(0.04,0.12)
特異度	0.93(0.02)	(0.88,0.96)	(0.79,0.93)×(0.88,0.96)
陽性予測値	0.86(0.03)	(0.79,0.92)	
陰性予測値	0.93(0.02)	(0.89,0.96)	

10

<sup>1</sup>クラスカル・ワリス検定

## 【0345】

## 【表 18】

表 18：モデル性能表—アダブースト—男性のみ—対象毎

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	102	8	110	0.95(0.02)	0.87(0.04)
癌以外	5	52	57		
計	107	60	167		

20

30

<sup>1</sup>クラスカル・ワリス検定

## 【0346】

## 【表 19】

表 19：モデル性能統計—アダブースト—男性のみ—対象毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.92(0.02)	(0.87,0.95)	
感度	0.95(0.02)	(0.89,0.98)	
偽陽性の割合	0.13(0.04)	(0.07,0.24)	(0.88,0.99)×(0.06,0.26)
特異度	0.87(0.04)	(0.76,0.93)	(0.88,0.99)×(0.74,0.94)
陽性予測値	0.93(0.02)	(0.86,0.96)	
陰性予測値	0.91(0.04)	(0.81,0.97)	

40

<sup>1</sup>クラスカル・ワリス検定

50

【 0 3 4 7 】

【 表 2 0 】

表 2 0 : モデル性能表—アダブースト—女性のみ—対象毎

予測した病理	実際の病理			感度(SE)	特異度(SE)
	癌	癌以外	計		
癌	106	15	121	0.95(0.02)	0.74(0.06)
癌以外	5	42	47		
計	111	57	168		

10

<sup>1</sup>クラスカル・ワリス検定

【 0 3 4 8 】

【 表 2 1 】

表 2 1 : モデル性能表—アダブースト—女性のみ—対象毎

統計項目	評価	95%信頼区間	95%統合信頼区間
精度	0.88(0.02)	(0.82, 0.92)	
感度	0.95(0.02)	(0.9, 0.98)	
偽陽性の割合	0.26(0.06)	(0.17, 0.39)	(0.88, 0.99)×(0.15, 0.41)
特異度	0.74(0.06)	(0.61, 0.83)	(0.88, 0.99)×(0.59, 0.85)
陽性予測値	0.88(0.03)	(0.8, 0.92)	
陰性予測値	0.89(0.04)	(0.77, 0.96)	

20

30

<sup>1</sup>クラスカル・ワリス検定

【 0 3 4 9 】

【 0 2 1 8 】 アダブーストおよび SVM の受信者特性 (ROC) 曲線および曲線下面積 (AUC) を図 1 と 2 に示す。アダブーストの AUC は 0.98 であり、SVM の AUC は 0.96 である。アダブーストの男性のみおよび女性の場合の ROC 曲線を図 3 と 4 に示す。男性の場合の AUC は 0.98 であり、女性の場合の AUC は 0.95 である。アダブーストの変数重要度プロットを図 5 に示す。アダブーストモデルで最も重要な 3 つの変数は、CTACK、MSCF、およびエオタキシン・3 である。男性に限定した場合のアダブーストの変数重要度プロットを図 6 に示す。最も重要な 3 つの変数は、MSCF、CTACK、およびエオタキシン・3 であった。女性に限定した場合のアダブーストの変数重要度プロットを図 7 に示す。最も重要な 3 つの変数は、MSCF、FGF、塩基性、および CTACK であった。

40

【 0 3 5 0 】

【 0 2 1 9 】 アダブーストの性能が、データベースを無作為に学習セットと確認セットに分割したことによる人為的な結果でないことを確認するために、無作為に 200 個の画分を生成し、それぞれの画分にモデルを当てはめた。これら 200 個のモデルの性能分布のを図 8 ~ 10 に要約した。アダブーストの性能は一定して良いように見え、このことから、アダブーストについて記載した性能統計が確認された。

50

## 【0351】

## 考察

〔0220〕 これらのデータは、性別による制限を設けても設けなくても、アダプスト分類器の優れた予測に関する一貫したパターンを示している。他の分析は、a) バイオマーカーのみを使用して癌をモデル化する工程、b) バイオマーカーの最適な下位セットを生成する工程、ここでバイオマーカーは、好ましくは図5-7の結果に基づいて、より好ましくはこれらの図におけるスコアが0.004を上回る、より好ましくはスコアが0.006を上回る、さらに好ましくはスコアが0.008を上回る、一層さらに好ましくはスコアが0.01を上回るバイオマーカーを含み、数が少ないながらも優れた予測の質を有し、c) この実施例の結果を実施例1~9の結果と比較する工程を含むはずである。

10

## 【0352】

## 実施例11 - 非小細胞肺癌の診断試験

〔0221〕 診断情報が望まれている患者から体液試料を得る。試料は好ましくは、血清または血漿である。実施例1~10のうちのいずれか1例からのバイオマーカーそれぞれの、試料中の濃度を決定する。試料中の各バイオマーカーについて測定した濃度を、サポートベクトルマシン中の学習データを使って決定した式に入力する。この式によって導き出された値が正であれば、これは非小細胞肺癌を示すものであり、値が負であれば、これは非小細胞肺癌でないことを示す。

20

## 【0353】

## 実施例12 - 男性対象における非小細胞肺癌の診断試験

〔0222〕 診断情報が望まれている男性患者から体液試料を得る。試料は好ましくは、血清または血漿である。実施例1~5、7~8または10のうちのいずれか1例からのバイオマーカーそれぞれの、試料中の濃度を決定する。試料中の各バイオマーカーについて測定した濃度を、サポートベクトルマシン中の学習データを使って決定した式に入力する。この式によって導き出された値が正であれば、これは非小細胞肺癌を示すものであり、値が負であれば、これは非小細胞肺癌でないことを示す。

## 【0354】

## 実施例13 - 男性対象における非小細胞肺癌の別の試験

〔0223〕 本明細書に記載する多くのバイオマーカーは、全てではなくても、上述した種類のコミュニケーション経路に関与している。バイオマーカーのいくつかは互いに一次相互作用器として関わり合っている。診断アッセイまたは予後アッセイで使用するためのマーカーの選択は、特定のバイオマーカー間およびそれらの一次相互作用器との間における既知の関係を使用することで容易になり得る。ARIADNE PATHWAY STUDIO (登録商標) で生成した、HGF (肝細胞増殖因子) と他のバイオマーカーとの間の既知のコミュニケーションを図5で見ることができる。図5は、sFasL (可溶性Fasリガンド)、PAI-1 (セルピンプラスミノゲン活性化因子阻害因子1) (活性型/全)、Ins (インスリン; C-ペプチドも含む)、EGF (上皮増殖因子)、MPO (ミエロペルオキシダーゼ)、およびMIF (マクロファージ遊走性阻止因子) を含む、HGF (肝細胞増殖因子) の一次相互作用器を示している。他の相互作用器 (一次相互作用器ではない) としては、RETN (レジスチン、xcp1)、SAA1 (血清アミロイドA、SAA)、CCL11 (エオタキシン)、LEP (レプチン) およびCXCL11 (ケモカイン (C-X-Cモチーフ) リガンド11、インターフェロン誘導T細胞 化学誘引物質 (ITAC) または インターフェロン誘導性タンパク質9 (IP-9)) が挙げられる。加えて、図5は、MMP1およびMMP-8 (マトリックスメタロプロテアーゼ1および8) の2種類のバイオマーカーが、HGFを伴うコミュニケーション経路上にはないことを示している。

30

40

## 【0355】

〔0224〕 診断情報が望まれている患者から体液試料を得る。試料は好ましくは、血清または血漿である。試料中の、選択したバイオマーカーだけの濃度を決定する。HG

50

Fがサポートベクトルマシン中で使用するための選択したバイオマーカーのうちの1つであると仮定すると、HGFのいずれかの一次相互作用器（例えば、MIF、EGFなど）の濃度を、HGFと置き換えることができる。そのため、HGFと置き換えた一次相互作用器を含む学習データについて、サポートベクトルマシンが再度実行される。次にこのモデルを患者の試料に適用する。この式によって導き出された値が正であれば、これは非小細胞肺癌を示すものであり、値が負であれば、これは非小細胞肺癌でないことを示す。

【0356】

実施例14 - 非小細胞肺癌と反応性気道疾患との区別

〔0225〕 非小細胞肺癌と反応性気道疾患とを区別するためには、(1)非小細胞肺癌の有無の評価、(2)反応性気道疾患の有無の評価、および(3)非小細胞肺癌または反応性気道疾患の評価、の3つを決定する。これらの評価は以下に記載するように実施する。

10

【0357】

〔0226〕 診断情報が望まれている患者から体液試料を得る。試料は好ましくは、血清または血漿である。実施例1～10のうちのいずれか1例からのバイオマーカーの、試料中の濃度を決定する。試料中の各バイオマーカーについて測定した濃度を、サポートベクトルマシン中の学習データを使って決定した式に入力する。この式によって導き出された値が正であれば、これは非小細胞肺癌を示すものであり、値が負であれば、これは非小細胞肺癌でないことを示す。

【0358】

20

〔0227〕 次に、実施例1～10のうちのいずれか1例からのバイオマーカーの、試料中の濃度を決定する。試料中の各バイオマーカーについて測定した濃度を、サポートベクトルマシン中の学習データを使って決定した式に入力する。この式によって導き出された値が正であれば、これは反応性気道疾患を示すものであり、値が負であれば、これは反応性気道疾患でないことを示す。

【0359】

〔0228〕 次に、実施例1～10のうちのいずれか1例からのバイオマーカーの、試料中の濃度を決定する。試料中の各バイオマーカーについて測定した濃度を、サポートベクトルマシン中の学習データを使って決定した式に入力する。この式によって導き出された値が正であれば、これは非小細胞肺癌を示すものであり、値が負であれば、これは反応性気道疾患を示す。

30

【0360】

〔0229〕 正のスコアおよび負のスコアを分析することによって、これらの結果をさらに評価する。具体的には、その患者が非小細胞肺癌であるか、反応性気道疾患であるか、または疾患に罹患していないかの決定は、どの条件が、3つのスコアのうちの2つで見られるかに依存する。例えば、第一の試験と第三の試験が正であった場合、その患者は非小細胞肺癌に罹患していると診断され得る。第一の試験と第二の試験が負であった場合には、その患者は非小細胞肺癌にも反応性気道疾患にも罹患していないと診断され得る。

【0361】

40

〔0230〕 本開示に含まれる等式、式および関係は、例示的かつ代表的なものであり、本発明を限定することを意図しない。本明細書で開示した任意の所与の式によって説明されるものと同じ現象を表すために、別の式を使ってもよい。具体的には、本発明で開示する式を、誤り訂正タームを加えることで、高位タームを加えることで、または不正確さを考慮すること、別の名称の定数もしくは変数を使用すること、または別の表現を使用することで修正してもよい。式の他の修正、置換、置き換え、または変更を実施してもよい。

【0362】

〔0231〕 本明細書中で言及している全ての出版物、特許、および公開されている特許出願は、その全体が、あたかも個々の出版物、特許または公開特許出願が具体的に

50

つ個別に参照することにより組み込まれると示されているのと同じように、参照することにより本明細書に組み込まれる。

本発明は、例えば、以下を提供する。

(項目1)

(a) 対象の生理学的試料を収集する工程；(b) 前記試料中の複数のバイオマーカーのバイオマーカー測定値を決定する工程；および(c) 分類システムを利用し、該バイオマーカー測定値に基づいて該試料を分類する工程を含む、該対象を生理学的に特徴付ける方法であって、ここで該試料の該分類と、該対象の生理的な状況もしくは状態、または疾患状況の変化には相関がある、前記方法。

(項目2)

(a) 対象の生理学的試料を収集する工程；(b) 前記試料中の複数のバイオマーカーのバイオマーカー測定値を決定する工程；および(c) 分類システムを利用し、該バイオマーカー測定値に基づいて該試料を分類する工程を含む、該対象における非小細胞肺癌の診断方法であって、ここで該試料の該分類は、該対象における非小細胞肺癌の存在または発生を示すものである、前記方法。

(項目3)

(a) 対象の生理学的試料を収集する工程；(b) 前記試料中の複数のバイオマーカーのバイオマーカー測定値を決定する工程；および(c) 分類システムを利用し、該バイオマーカー測定値に基づいて該試料を分類する工程、を含む該対象における反応性気道疾患の診断方法であって、ここで該試料の該分類は、該対象における反応性気道疾患を示すものである、前記方法。(項目4)

(a) 対象の生理学的試料を収集する工程；

(b) 前記試料中の、反応性気道疾患の兆候と非小細胞肺癌の区別を補助する複数のバイオマーカー、反応性気道疾患を示す複数のバイオマーカー、および非小細胞肺癌を示す複数のバイオマーカーのバイオマーカー測定値を決定する工程、ここで前記複数のバイオマーカーは同一のものでなく；

(c) 3つの分類システムを使用して、該バイオマーカー測定値に基づいて該試料を分類する工程、ここで該試料の該分類は、該対象における(i) 反応性気道疾患および非小細胞肺癌；(ii) 反応性気道疾患の有無；および(iii) 非小細胞肺癌の有無をの兆候の区別を補助するものであり；

(d) どの条件が該3つの分類のうち2つに見られるかに応じて、該対象が、(1) 反応性気道疾患に罹患しているか；(2) 非小細胞肺癌に罹患しているか、または(3) 疾患に罹患していないか、を決定する工程、を含む該対象の肺疾患の診断方法。

(項目5)

該分類システムが機械学習システムである、項目1～4のいずれか1項に記載の方法。

(項目6)

該機械学習システムがカーネルを使用する分類システムである、項目5の方法。

(項目7)

該カーネルを使用する分類システムがサポートベクトルマシンである、項目6の方法。

(項目8)

該機械学習システムが分類および回帰システムである、項目5の方法。

(項目9)

該機械学習システムが分類および回帰システムの集合である、項目8の方法。

(項目10)

該機械学習システムがアダプティブである、項目9の方法。

(項目11)

バイオマーカーのセットそれぞれの、複数のバイオマーカー測定値を含む試験データの分類方法であって、  
ヒト試験対象における該バイオマーカーのセットのそれぞれのバイオマーカーに関するバイオマーカー測定値を含む試験データを受信する工程；

10

20

30

40

50

電子的に記憶されている学習データベクトルのセットを使って訓練したサポートベクトルマシンの電子表示を使用して該試験データを評価する工程、ここで各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカのセットの各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；および  
該評価する工程に基づいて、該ヒト試験対象の分類を出力する工程を含む、試験データの分類方法。

(項目12)

バイオマーカのセットそれぞれの、複数のバイオマーカ測定値を含む試験データの分類方法であって、

ヒト試験対象における該バイオマーカのセットの、各バイオマーカに関するバイオマーカ測定値を含む該試験データを受信する工程；

電子的に記憶されている学習データベクトルのセットを使って訓練したアダプスト分類器の電子表示を使用して試験データを評価する工程、ここで、各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカのセットの各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；および、

評価する工程に基づいたヒト試験対象の分類を出力する工程を含む、試験データの分類方法。

(項目13)

バイオマーカのセットそれぞれの、複数のバイオマーカ測定値を含む試験データの分類方法であって、

電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカのセットの各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；

該電子的に記憶されている学習データベクトルのセットを使用して、サポートベクトルマシンの電子表示を訓練する工程；

ヒト試験対象における該バイオマーカのセットに関する複数のバイオマーカ測定値を含む試験データを受信する工程；

該サポートベクトルマシンの該電子表示を使用して、該試験データを評価する工程；および、

該評価する工程に基づいて、該ヒト試験対象の分類を出力する工程を含む、試験データの分類方法。

(項目14)

バイオマーカのセットそれぞれの、複数のバイオマーカ測定値を含む試験データの分類方法であって、

電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカのセットの各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；

該電子的に記憶されている学習データベクトルのセットを使用して、アダプスト分類器の電子表示を訓練する工程；

ヒト試験対象における該バイオマーカのセットに関する複数のバイオマーカ測定値を含む試験データを受信する工程；

該アダプスト分類器の該電子表示を使用して、該試験データを評価する工程；および、該評価する工程に基づいて、該ヒト試験対象の分類を出力する工程を含む、試験データの分類方法。

(項目15)

バイオマーカのセットそれぞれの、複数のバイオマーカ測定値を含む試験データの

10

20

30

40

50

分類方法であって、

電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカのセットの各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；

該バイオマーカのセットから、バイオマーカの下位セットを選択する工程；

該電子的に記憶されている学習データベクトルのセットの、バイオマーカの該下位セットからのデータを使用して、サポートベクトルマシンの電子表示を訓練する工程；

ヒト試験対象に関する複数のバイオマーカ測定値を含む試験データを受信する工程；

該サポートベクトルマシンの該電子表示を使用して、該試験データを評価する工程；および、

該評価する工程に基づいて、該ヒト試験対象の分類を出力する工程を含む、試験データの分類方法であって、ここで、バイオマーカの該下位セットを選択する工程には、

a．該バイオマーカのセットに含まれているそれぞれのバイオマーカについて、プログラムされているコンピューターを使用して、それぞれのバイオマーカに関する2群の濃度測定値のマージン分布間の距離を計算すること、それによって、複数の距離が生成され；

b．該距離に従って、該バイオマーカのセットに含まれている該バイオマーカを順序付けること、それによって順序付けられたバイオマーカのセットが生成され；

c．該順序付けられたバイオマーカのセットの複数の初期セグメントのそれぞれについて、該学習データに基づいてモデルフィットの指標を計算すること；

d．モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカのセットの初期セグメントを選択すること、それによって、該順序付けられたバイオマーカのセットの好ましい初期セグメントが選択され；

e．バイオマーカの空集合から開始し、バイオマーカの該下位セットを生成するために、該順序付けられたバイオマーカのセットの好ましい初期セグメントから、該モデルにさらなるバイオマーカを再帰的に追加すること、ここで各さらなるバイオマーカは、(1)該好ましい初期セグメントに含まれている残りのバイオマーカの中でも、それを追加することで、モデルフィットが最大に改良される場合、および(2)それを追加することで、モデルフィットが少なくとも予め設定した閾値までは改良される場合に、既存のバイオマーカの下位セットに追加され；

f．それ以上バイオマーカを追加しても、該予め設定しておいた閾値より、そのモデルフィット指標を超えるモデルフィットの指標が得られない場合には、既存のバイオマーカの下位セットへのバイオマーカの追加を停止すること、それによって、バイオマーカの下位セットが選択される、が含まれる、前記方法。

(項目16)

バイオマーカのセットそれぞれの、複数のバイオマーカ測定値を含む試験データの分類方法であって、

電子的に記憶されている学習データベクトルのセットにアクセスする工程、ここで各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカのセットの各バイオマーカのバイオマーカ測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；

該バイオマーカのセットから、バイオマーカの下位セットを選択する工程；

該電子的に記憶されている学習データベクトルのセットのバイオマーカの該下位セットからの該データを使用して、アダプスト分類器の電子表示を訓練する工程；

ヒト試験対象に関する複数のバイオマーカ測定値を含む試験データを受信する工程；

該サポートベクトルマシンの該電子表示を使用して、該試験データを評価する工程；および、

該評価する工程に基づく該ヒト試験対象の分類を出力する工程を含む試験データの分類方法であって、ここで、バイオマーカの該下位セットを選択する工程には：

10

20

30

40

50

a . 該バイオマーカーのセットに含まれているそれぞれのバイオマーカーについて、プログラムされているコンピューターを使用して、それぞれのバイオマーカーに関する 2 群の濃度測定値のマージン分布間の距離を計算すること、それによって、複数の距離が生成され；

b . 該距離に従って、該バイオマーカーのセットに含まれている該バイオマーカーを順序付けること、それによって順序付けられたバイオマーカーのセットが生成され；

c . 該順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについて、該学習データに基づいてモデルフィットの指標を計算すること；

d . モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカーのセットの初期セグメントを選択すること、それによって、該順序付けられたバイオマーカーのセットの好ましい初期セグメントが選択され；

e . バイオマーカーの空集合から開始し、バイオマーカーの該下位セットを生成するために、該順序付けられたバイオマーカーのセットの好ましい初期セグメントから、該モデルにさらなるバイオマーカーを再帰的に追加すること、ここで各さらなるバイオマーカーは、( 1 ) 該好ましい初期セグメントに含まれている残りのバイオマーカーの中でも、それを追加することで、モデルフィットが最大に改良される場合、および ( 2 ) それを追加することで、モデルフィットが少なくとも予め設定した閾値までは改良される場合に、既存のバイオマーカーの下位セットに追加され；

f . それ以上バイオマーカーを追加しても、該予め設定しておいた閾値より、そのモデルフィット指標を超えるモデルフィットの指標が得られない場合には、既存のバイオマーカーの下位セットへのバイオマーカーの追加を停止すること、それによって、バイオマーカーの該下位セットが選択される工程、が含まれる、前記方法。

( 項目 1 7 )

バイオマーカーのセットそれぞれの、複数のバイオマーカー測定値を含む試験データの分類方法であって、

ヒト試験対象に関する該バイオマーカーのセットそれぞれのバイオマーカー測定値を含む試験データを受信する工程；

電子的に記憶されている学習データベクトルのセットを使って訓練したサポートベクトルマシンの電子表示を使用して、該試験データを評価する工程、ここで各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカーのセットの各バイオマーカーのバイオマーカー測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり、該バイオマーカーのセットはバイオマーカーの上位セットから選択され；および、

該評価する工程に基づいた該ヒト試験対象の分類を出力する工程を含む、試験データの分類方法。

( 項目 1 8 )

バイオマーカーのセットそれぞれの、複数のバイオマーカー測定値を含む該試験データの分類方法であって、

ヒト試験対象に関する該バイオマーカーのセットそれぞれの各バイオマーカー測定値を含む試験データを受信する工程；

電子的に記憶されている学習データベクトルのセットを使って訓練したアダプスト分類器の電子表示を使用して試験データを評価する工程、ここで、各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関するバイオマーカーのセットの各バイオマーカーのバイオマーカー測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり、該バイオマーカーのセットはバイオマーカーの上位セットから選択され；および

該評価する工程に基づいた該ヒト試験対象の分類を出力する工程を含む、試験データの分類方法。

( 項目 1 9 )

a . バイオマーカーの上位セットに含まれている各バイオマーカーについて、プログラ

ムされているコンピューターを使用して、それぞれのバイオマーカーに関する 2 群の濃度測定値のマージン分布間の距離を計算すること、それによって、複数の距離が生成され；  
b . 該距離に従って、該バイオマーカーのセットに含まれている該バイオマーカーを順序付けること、それによって、順序付けられたバイオマーカーのセットが生成され；  
c . 該順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについて、該学習データに基づいてモデルフィットの指標を計算すること；

d . モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカーのセットの初期セグメントを選択すること、それによって、該順序付けられたバイオマーカーのセットの好ましい初期セグメントが選択され；

e . バイオマーカーの空集合から開始し、バイオマーカーの該下位セットを生成するために、該順序付けられたバイオマーカーのセットの該好ましい初期セグメントから、さらなるバイオマーカーを再帰的に追加すること、ここで各さらなるバイオマーカーは、( 1 ) 該好ましい初期セグメントに含まれている残りのバイオマーカーの中でも、それを追加することで、モデルフィットが最大に改良される場合、および ( 2 ) それを追加することで、モデルフィットが少なくとも予め設定した閾値までは改良される場合に、既存のバイオマーカーの下位セットに追加され；

f . それ以上バイオマーカーを追加しても、該予め設定しておいた閾値より、そのモデルフィット指標を超えるモデルフィットの指標が得られない場合には、既存のバイオマーカーの下位セットへのバイオマーカーの追加を停止すること、それによって、バイオマーカーの下位セットが選択される、を含む方法を使用して、該バイオマーカーのセットの測定値を、バイオマーカーの該上位セットから選択する、項目 17 および 18 に記載の方法。

( 項目 20 )

a . バイオマーカーの上位セットに含まれている各バイオマーカーに関して、それぞれのバイオマーカーに関する 2 群の濃度測定値のマージン分布間の距離を計算し、それによって、複数の距離を生成し；

b . 該距離に従って、バイオマーカーの該上位セットに含まれている該バイオマーカーを順序付け、それによって、順序付けられたバイオマーカーのセットを生成し；

c . 該順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについて、該学習データに基づいてモデルフィットの指標を計算し；

d . モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカーのセットの初期セグメントを選択し、それによって、該順序付けられたバイオマーカーのセットの好ましい初期セグメントを選択し；

e . バイオマーカーの該初期セグメントから開始し、バイオマーカーの下位セットを生成するために、該順序付けられたバイオマーカーのセットの該好ましい初期セグメントから、バイオマーカーを再帰的に除去し、ここで、各バイオマーカーは、( 1 ) 該好ましい初期セグメントに含まれている残りのバイオマーカーのなかでも、それを除去することが、モデルフィットの縮小を最小限に抑える場合、および ( 2 ) それを除去することが、少なくとも、予め設定しておいた閾値まではモデルフィットを縮小させない場合に、バイオマーカーの既存の上位セットから除去され；

f . それ以上バイオマーカーを除去すると、モデルフィットの指標が、該予め設定しておいた閾値によってモデルフィットの指標を下回る低下を生じる場合には、バイオマーカーの既存の上位セットからのバイオマーカーの除去を停止し、それによって、バイオマーカーの該上位セットを選択する、ために構成されている論理を使用して、バイオマーカーの該上位セットから該バイオマーカーのセットを選択するために構成されているコンピューターをさらに含む、項目 17 および 18 に記載のシステム。

( 項目 21 )

該評価する工程が、電子的に記憶されている学習データベクトルのセットにアクセスすることを含む、項目 17 ~ 20 のいずれか 1 項に記載の方法。

( 項目 22 )

該病状についての分類が前記病状の有無である、項目 11 ~ 21 のいずれか 1 項に記載

10

20

30

40

50

の方法。

(項目 23)

該病状が肺疾患である、項目 22 の方法。

(項目 24)

該肺疾患が非小細胞癌である、項目 23 の方法。

(項目 25)

該肺疾患が反応性気道疾患である、項目 23 の方法。

(項目 26)

該反応性気道疾患が喘息である、項目 25 の方法。

(項目 27)

該バイオマーカー測定値が、アポリポタンパク質(「Apo」)A1、ApoA2、ApoB、ApoC2、ApoE、CD40、Dダイマー、第VII因子、第VII因子、第X因子、プロテインC、組織プラスミノゲン活性化因子(「TPA」)、脳由来神経栄養因子(「BDNF」)、Bリンパ球化学誘引物質(「BLC」)、ケモカイン(C-X-Cモチーフ)リガンド1(「GRO-1」)、皮膚T細胞誘引ケモカイン(「CTACK」)、エオタキシン-2、エオタキシン-3、グランザイム-B、肝細胞増殖因子(「HGF」)、I-TAC(「CXCL11」;「ケモカイン(C-X-Cモチーフ)リガンド11」、インターフェロン誘導T細胞化学誘引物質)、レプチン(「LEP」)、白血病抑制因子(「LIF」)、単球特異的ケモカイン3(「MMP-3」)、マクロファージコロニー刺激因子(「MCSF」、インターフェロン誘導性モノカイン(「MIG」)、マクロファージ炎症性タンパク質-3(「MIP-3」)、マトリックスメタロプロテアーゼ(「MMP」)1、MMP2、MMP3、MMP7、MMP8、MMP9、MMP12、MMP13、CD40、神経成長因子(「NGF」)、可溶性リガンド(「CD40リガンド」)、上皮増殖因子(「EGF」)、エオタキシン(「CCL11」)、フラクタルキン、塩基性線維芽細胞増殖因子(「FGF-塩基性」)、顆粒球コロニー刺激因子(「G-CSF」)、顆粒球マクロファージコロニー刺激因子(「GM-CSF」)、インターフェロン(「IFN」)、IFN-、IFN-2、IFN-、インターロイキン(「IL」)1a、IL-1、IL-1ra、IL-2、IL-2ra、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-12(p40)、IL-12(p70)、IL-13、IL-15、IL-16、IL-17、IL-17a、IL-17F、IL-20、IL-21、IL-22、IL-23(p19)、IL-27、IL-31、IP-10、単球走化性タンパク質1(「MCP-1」)、マクロファージ炎症性タンパク質(「MIP」)1a、MIP-1、好中球活性化ペプチド78(「ENA-78」)、オステオプロテゲリン(「OPG」)、胎盤増殖因子(「PIGF」)、血小板由来成長因子サブユニットBホモ二量体(「PDGFBB」)、T細胞由来の好酸球走化性物質(「RANTES」)、幹細胞成長因子(「SCGF」)、ストロマ細胞由来因子1(「SDF-1」)、可溶性Fasリガンド(「Fas-リガンド」)、NF-B活性化可溶性受容体リガンド(「sRANKL」)、サバイピン、トランスフォーミング増殖因子a(「TGFa」)、TGF-、腫瘍壊死因子a(「TNFa」)、TNF-、TNF受容体1(「TNFR-I」)、TNFR-II、TNF関連性アポトーシス誘導性リガンド(「TRAIL」)、トロンボポエチン(「TPO」)、血管内皮成長因子(「VEGF」)、インスリン(「Ins」)、C-ペプチド、グルカゴン様タンパク質-1/アミリン(「GLP-1/アミリン」)、アミリン(全)、グルカゴン、アディポネクチン、プラスミノゲン活性化因子阻害因子1(「PAI-1」;「セルピン」)(活性型/全)、レジスチン(「RETN」;「xcp1」)、sFas、可溶性Fasリガンド(「sFasL」)、マクロファージ遊走性阻止因子(「MIF」)、sE-セレクトリン、可溶性血管細胞接着分子(「sVCAM」)、可溶性細胞間接着分子(「sICAM」)、ミエロペルオキシダーゼ(「MPO」)、C反応性タンパク質(「CRP」)、血清アミロイドA(「SAA」;「SAA1」)、および血清アミロイドP(「SAP」)からなる群よ

10

20

30

40

50

り選択される、少なくとも1つのタンパク質の血漿濃度の測定値を含む、項目11～26のいずれか1項に記載の方法。

(項目28)

該バイオマーカー測定値が、少なくとも4種類の別個のバイオマーカーの血漿濃度を含む、項目27の方法。

(項目29)

該バイオマーカー測定値が、少なくとも6種類の別個のバイオマーカーの血漿濃度を含む、項目27の方法。

(項目30)

該バイオマーカー測定値が、少なくとも10種類の別個のバイオマーカーの血漿濃度を含む、項目27の方法。

(項目31)

該バイオマーカー測定値が、少なくとも18種類の別個のバイオマーカーの血漿濃度を含む、項目27の方法。

(項目32)

該学習ベクトルのセットが、少なくとも30ベクトル、50ベクトル、100ベクトルを含む、項目11～31のいずれか1項に記載の方法。

(項目33)

該サポートベクトルマシンが、線形カーネル、ラジアル基底カーネル、多項カーネル、統一カーネル、三角カーネル、エパネクニコフ(Epanechnikov)カーネル、四次多項(二乗重み)カーネル、三次多項(三乗重み)カーネル、およびコサインカーネルから選択される、1つまたは複数のカーネル関数を含む、項目11、13、15、17、および19～26のいずれか1項に記載の方法。

(項目34)

該アダプスト分類器が、5、10、15、20、25、30、40、50、75または100回の反復を含む、項目12、14、16、および18～26のいずれか1項に記載の方法。

(項目35)

バイオマーカーのセットそれぞれのバイオマーカー測定値を含む試験データを分類するためのシステムであって、

個々のヒトを表し、かつ、対応するヒトに関するバイオマーカーのセットの各バイオマーカーのバイオマーカー測定値を含み、さらに、対応するヒトの病状に関する分類を含んでいる、電子的に記憶されている学習データベクトルのセット；ならびに、

該電子的に記憶されている学習データベクトルのセットを使用して訓練するのに適しているサポートベクトルマシンの電子表示を含んでおり、ヒト試験対象における該バイオマーカーのセットの複数のバイオマーカー測定値を含む試験データを受信するために構成されており、さらに、該サポートベクトルマシンの該電子表示を使用して該試験データを評価し、および、該評価に基づいて該ヒト試験対象の分類を出力するように構成されている、該電子的に記憶されている学習データベクトルのセットに操作可能に接続されているコンピュータを含む、試験データを分類するためのシステム。

(項目36)

バイオマーカーのセットそれぞれのバイオマーカー測定値を含む試験データを分類するためのシステムであって、

個々のヒトを表し、かつ、対応するヒトに関するバイオマーカーのセットの各バイオマーカーのバイオマーカー測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでいる、電子的に記憶されている学習データベクトルのセット；ならびに、

該電子的に記憶されている学習データベクトルのセットを使用して訓練するのに適しているアダプスト分類器の電子表示を含み、ヒト試験対象における該バイオマーカーのセットの複数のバイオマーカー測定値を含む試験データを受信するように構成されており、さ

10

20

30

40

50

らに、該サポートベクトルマシンの該電子表示を使用して該試験データを評価し、および、該評価に基づいて該ヒト試験対象の分類を出力するために構成されている、該電子的に記憶されている学習データベクトルのセットに操作可能に接続されているコンピューターを含む、試験データを分類するためのシステム。

(項目37)

電子的に記憶されている学習データベクトルのセットを使用して訓練したサポートベクトルマシンの電子表示を含むコンピューターを含む、バイオマーカーのセットそれぞれの複数のバイオマーカー測定値を含む試験データを分類するためのシステムであって、各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカーのセットの各バイオマーカーのバイオマーカー測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり、該コンピューターは、ヒト試験対象における該バイオマーカーのセットに関する複数のバイオマーカー測定値を含む試験データを受信するために構成されており、該コンピューターはさらに、該サポートベクトルマシンの該電子表示を使用して該試験データを評価し、および、該評価に基づいて該ヒト試験対象の分類を出力するように構成されている、試験データを分類するためのシステム。

(項目38)

電子的に記憶されている学習データベクトルのセットを使用して訓練したアダプスト分類器の電子表示を含むコンピューターを含む、バイオマーカーのセットそれぞれの複数のバイオマーカー測定値を含む試験データを分類するためのシステムであって、各学習データベクトルは個々のヒトを表し、かつ、対応するヒトに関する該バイオマーカーのセットの各バイオマーカーのバイオマーカー測定値を含み、各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり、該コンピューターは、ヒト試験対象における該バイオマーカーのセットに関する複数のバイオマーカー測定値を含む試験データを受信するために構成されており、該コンピューターはさらに、該サポートベクトルマシンの該電子表示を使用して該試験データを評価し、および、該評価に基づいて該ヒト試験対象の分類を出力するように構成されている、試験データを分類するためのシステム。

(項目39)

a. バイオマーカーの上位セットに含まれている各バイオマーカーに関して、それぞれのバイオマーカーに関する2群の濃度測定値のマージン分布間の距離を計算し、それによって、複数の距離を生成し；

b. 該距離に従って、バイオマーカーの該上位セットに含まれている該バイオマーカーを順序付け、それによって、順序付けられたバイオマーカーのセットを生成し；

c. 該順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについて、該学習データに基づいてモデルフィットの指標を計算し；

d. モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカーのセットの初期セグメントを選択し、それによって、該順序付けられたバイオマーカーのセットの好ましい初期セグメントを選択し；

e. バイオマーカーの空集合から開始し、バイオマーカーの下位セットを生成するために、該順序付けられたバイオマーカーのセットの該好ましい初期セグメントから、さらなるバイオマーカーを再帰的に追加し、ここで各さらなるバイオマーカーは、(1)該好ましい初期セグメントに含まれている残りのバイオマーカーの中でも、それを追加することで、モデルフィットが改良される場合、および(2)それを追加することで、モデルフィットが少なくとも予め設定した閾値までは改良される場合に、既存のバイオマーカーの下位セットに追加され；

f. それ以上バイオマーカーを追加しても、該予め設定しておいた閾値より、そのモデルフィット指標を超えるモデルフィットの指標が得られない場合には、既存のバイオマーカーの下位セットへのバイオマーカーの追加を停止し、それによって、バイオマーカーの該下位セットを選択する、ために構成されている論理を使用して、バイオマーカーの該上位セットから該バイオマーカーのセットを選択するために構成されているコンピューターをさらに含む、項目37のシステム。

10

20

30

40

50

( 項目 4 0 )

- a . バイオマーカーの上位セットに含まれている各バイオマーカーに関して、それぞれのバイオマーカーに関する 2 群の濃度測定値のマージン分布間の距離を計算し、それによって、複数の距離を生成し；
- b . 該距離に従って、バイオマーカーの該上位セットに含まれている該バイオマーカーを順序付け、それによって、順序付けられたバイオマーカーのセットを生成し；
- c . 該順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについて、該学習データに基づいてモデルフィットの指標を計算し；
- d . モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカーのセットの初期セグメントを選択し、それによって、該順序付けられたバイオマーカーのセットの好ましい初期セグメントを選択し；
- e . バイオマーカーの該初期セグメントから開始し、バイオマーカーの下位セットを生成するために、該順序付けられたバイオマーカーのセットの該好ましい初期セグメントから、バイオマーカーを再帰的に除去し、ここで、各バイオマーカーは、( 1 ) 該好ましい初期セグメントに含まれている残りのバイオマーカーのなかでも、それを除去することが、モデルフィットの縮小を最小限に抑える場合、および ( 2 ) それを除去することが、少なくとも、予め設定しておいた閾値まではモデルフィットを縮小させない場合に、バイオマーカーの既存の上位セットから除去され；
- f . それ以上バイオマーカーを除去すると、モデルフィットの指標が、該予め設定しておいた閾値によってモデルフィット指標を下回る低下を生じる場合には、バイオマーカーの既存の上位セットからのバイオマーカーの除去を停止し、それによって、バイオマーカーの該上位セットを選択する、ために構成されている論理を使用して、バイオマーカーの該上位セットから該バイオマーカーのセットを選択するために構成されているコンピューターをさらに含む、項目 3 7 に記載のシステム。

10

20

( 項目 4 1 )

- a . バイオマーカーの上位セットに含まれている各バイオマーカーに関して、それぞれのバイオマーカーに関する 2 群の濃度測定値のマージン分布間の距離を計算し、それによって、複数の距離を生成し；
- b . 該距離に従って、バイオマーカーの該上位セットに含まれている該バイオマーカーを順序付け、それによって、順序付けられたバイオマーカーのセットを生成し；
- c . 該順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについて、該学習データに基づいてモデルフィットの指標を計算し；
- d . モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカーのセットの初期セグメントを選択し、それによって、該順序付けられたバイオマーカーのセットの好ましい初期セグメントを選択し；
- e . バイオマーカーの空集合から開始し、バイオマーカーの下位セットを生成するために、該順序付けられたバイオマーカーのセットの該好ましい初期セグメントから、さらなるバイオマーカーを再帰的に追加し、ここで各さらなるバイオマーカーは、( 1 ) 該好ましい初期セグメントに含まれている残りのバイオマーカーの中でも、それを追加することで、モデルフィットが改良される場合、および ( 2 ) それを追加することで、モデルフィットが少なくとも予め設定した閾値までは改良される場合に、既存のバイオマーカーの下位セットに追加され；
- f . それ以上バイオマーカーを追加しても、該予め設定しておいた閾値より、そのモデルフィット指標を超えるモデルフィットの指標が得られない場合には、既存のバイオマーカーの下位セットへのバイオマーカーの追加を停止し、それにより、バイオマーカーの該下位セットを選択する、ために構成されている論理を使用して、バイオマーカーの該上位セットから該バイオマーカーのセットを選択するために構成されているコンピューターをさらに含む、項目 3 8 のシステム。

30

40

( 項目 4 2 )

- a . バイオマーカーの上位セットに含まれている各バイオマーカーに関して、それぞれ

50

のバイオマーカーに関する 2 群の濃度測定値のマージン分布間の距離を計算し、それによ  
って、複数の距離を生成し；

b . 該距離に従って、バイオマーカーの該上位セットに含まれている該バイオマーカーを  
順序付け、それによって、順序付けられたバイオマーカーのセットを生成し；

c . 該順序付けられたバイオマーカーのセットの複数の初期セグメントのそれぞれについ  
て、該学習データに基づいてモデルフィットの指標を計算し；

d . モデルフィットの最大値型指標に従って、該順序付けられたバイオマーカーのセット  
の初期セグメントを選択し、それによって、該順序付けられたバイオマーカーのセットの  
好ましい初期セグメントを選択し；

e . バイオマーカーの該初期セグメントから開始し、バイオマーカーの下位セットを生成  
するために、該順序付けられたバイオマーカーのセットの該好ましい初期セグメントから  
、バイオマーカーを再帰的に除去し、ここで、各バイオマーカーは、( 1 ) 該好ましい初  
期セグメントに含まれている残りのバイオマーカーのなかでも、それを除去することが、  
モデルフィットの縮小を最小限に抑える場合、および ( 2 ) それを除去することが、少な  
くとも、予め設定しておいた閾値まではモデルフィットを縮小させない場合に、バイオマ  
ーカーの既存の上位セットから除去され；

f . それ以上バイオマーカーを除去すると、モデルフィットの指標が、該予め設定してお  
いた閾値によってモデルフィットの指標を下回る低下を生じる場合には、バイオマーカー  
の既存の上位セットからのバイオマーカーの除去を停止し、それによって、バイオマーカー  
の該上位セットを選択する、ために構成されている論理を使用して、バイオマーカーの  
該上位セットから該バイオマーカーのセットを選択するために構成されているコンピュー  
ターをさらに含む、項目 3 8 に記載のシステム。

( 項目 4 3 )

該病状についての分類が前記病状の有無である、項目 3 5 ~ 4 2 のいずれか 1 項に記載  
のシステム。

( 項目 4 4 )

該病状が肺疾患である、項目 4 3 のシステム。

( 項目 4 5 )

該肺疾患が非小細胞癌である、項目 4 4 のシステム。

( 項目 4 6 )

該肺疾患が反応性気道疾患である、項目 4 4 のシステム。

( 項目 4 7 )

該反応性気道疾患が喘息である、項目 4 6 のシステム。

( 項目 4 8 )

該バイオマーカー測定値が、アポリポタンパク質 ( 「 A p o 」 ) A 1、A p o A 2、A  
p o B、A p o C 2、A p o E、C D 4 0、D ダイマー、第 V I I 因子、第 V I I I 因子  
、第 X 因子、プロテイン C、組織プラスミノゲン活性化因子 ( 「 T P A 」 )、脳由来神  
経栄養因子 ( 「 B D N F 」 )、B リンパ球化学誘引物質 ( 「 B L C 」 )、ケモカイン ( C  
- X - C モチーフ) リガンド 1 ( 「 G R O - 1 」 )、皮膚 T 細胞誘引ケモカイン ( 「 C T  
A C K 」 )、エオタキシン - 2、エオタキシン - 3、グランザイム - B、肝細胞増殖因子  
( 「 H G F 」 )、I - T A C ( 「 C X C L 1 1 」 ; 「ケモカイン ( C - X - C モチーフ)  
リガンド 1 1 」、 「インターフェロン誘導 T 細胞 化学誘引物質」 )、レプチン ( 「 L E  
P 」 )、白血病抑制因子 ( 「 L I F 」 )、単球特異的ケモカイン 3 ( 「 M M P - 3 」 )、  
マクロファージコロニー刺激因子 ( 「 M C S F 」 )、インターフェロン誘導性モノカイン  
( 「 M I G 」 )、マクロファージ炎症性タンパク質 - 3 ( 「 M I P - 3 」 )、マト  
リックスメタロプロテアーゼ ( 「 M M P 」 ) 1、M M P 2、M M P 3、M M P 7、M M P  
8、M M P 9、M M P 1 2、M M P 1 3、C D 4 0、神経成長因子 ( 「 N G F - 」 )  
、可溶性リガンド ( 「 C D 4 0 リガンド」 )、上皮増殖因子 ( 「 E F G 」 )、エオタキシ  
ン ( 「 C C L 1 1 」 )、フラクタルキン、塩基性線維芽細胞増殖因子 ( 「 F G F - 塩基性  
」 )、顆粒球コロニー刺激因子 ( 「 G - C S F 」 )、顆粒球マクロファージコロニー刺激

10

20

30

40

50

因子(「GM-CSF」)、インターフェロン(「IFN」)、IFN-、IFN-2、IFN-、インターロイキン(「IL」)1a、IL-1、IL-1ra、IL-2、IL-2ra、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-12(p40)、IL-12(p70)、IL-13、IL-15、IL-16、IL-17、IL-17a、IL-17F、IL-20、IL-21、IL-22、IL-23(p19)、IL-27、IL-31、IP-10、単球走化性タンパク質1(「MCP-1」)、マクロファージ炎症性タンパク質(「MIP」)1a、MIP-1、好中球活性化ペプチド78(「ENA-78」)、オステオプロテゲリン(「OPG」)、胎盤増殖因子(「PIGF」)、血小板由来成長因子サブユニットBホモ二量体(「PDGFBB」)、T細胞由来の好酸球走化性物質(「RANTES」)、幹細胞成長因子(「SCGF」)、ストロマ細胞由来因子1(「SDF-1」)、可溶性Fasリガンド(「Fas-リガンド」)、NF- $\kappa$ B活性化可溶性受容体リガンド(「sRANKL」)、サバイビン、トランスフォーミング増殖因子a(「TGFa」)、TGF- $\beta$ 、腫瘍壊死因子a(「TNFa」)、TNF- $\alpha$ 、TNF受容体1(「TNFR-I」)、TNFR-II、TNF関連性アポトーシス誘導性リガンド(「TRAIL」)、トロンボポエチン(「TPO」)、血管内皮成長因子(「VEGF」)、インスリン(「Ins」)、C-ペプチド、グルカゴン様タンパク質-1/アミリン(「GLP-1/アミリン」)、アミリン(全)、グルカゴン、アディポネクチン、プラスミノゲン活性化因子阻害因子1(「PAI-1」;「セルピン」)(活性型/全)、レジスチン(「RETN」;「xcp1」)、sFas、可溶性Fasリガンド(「sFasL」)、マクロファージ遊走性阻止因子(「MIF」)、sE-セレクトイン、可溶性血管細胞接着分子(「sVCAM」)、可溶性細胞間接着分子(「sICAM」)、ミエロペルオキシダーゼ(「MPO」)、C反応性タンパク質(「CRP」)、血清アミロイドA(「SAA」;「SAA1」)、および血清アミロイドP(「SAP」)からなる群より選択される少なくとも1つのタンパク質の血漿濃度の測定値を含む、項目35~47のいずれか1項に記載にシステム。

(項目49)

該バイオマーカー測定値が、少なくとも4種類の別個のバイオマーカーの血漿濃度を含む、項目48のシステム。

(項目50)

該バイオマーカー測定値が、少なくとも6種類の別個のバイオマーカーの血漿濃度を含む、項目48のシステム。

(項目51)

該バイオマーカー測定値が、少なくとも10種類の別個のバイオマーカーの血漿濃度を含む、項目48のシステム。

(項目52)

該バイオマーカー測定値が、少なくとも18種類の別個のバイオマーカーの血漿濃度を含む、項目48のシステム。

(項目53)

該学習ベクトルのセットが、少なくとも30ベクトル、50ベクトル、100ベクトルを含む、項目35~52のいずれか1項に記載のシステム。

(項目54)

該サポートベクトルマシンが、線形カーネル、ラジアル基底カーネル、多項カーネル、統一カーネル、三角カーネル、エパネクニコフ(Epanechnikov)カーネル、四次多項(quartic)(二乗重み)カーネル、三次三乗(tricube)(三乗重み)カーネル、およびコサインカーネルから選択される、1つまたは複数のカーネル関数を含む、項目35、37、39、40、または43~53のいずれか1項に記載のシステム。

(項目55)

該アダプティブ分類器が、5、10、15、20、25、30、40、50、75また

10

20

30

40

50

は 100 回の反復を含む、項目 36、38、または 41 ~ 53 のいずれか 1 項に記載のシステム。

(項目 56)

バイオマーカーのセットそれぞれの、複数のバイオマーカー測定値を含む試験データを分類する方法であって、

ヒト試験対象に関する試験データを受信する工程、ここで該試験データは、バイオマーカーのセットの少なくとも各バイオマーカーのバイオマーカー測定値を含んでおり；

電子的に記憶されている学習データベクトルの第一のセットを使用して訓練したサポートベクトルマシンの電子表示を使用して該試験データを評価する工程、ここで、学習データベクトルの該第一のセットの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトの該バイオマーカーのセットの少なくとも各バイオマーカーのバイオマーカー測定値を含み、学習データベクトルの該第一のセットの各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；および、

該評価する工程に基づいた該ヒト試験対象の分類を出力する工程；を含み、

ここで、該バイオマーカーのセットに含まれているそれぞれのバイオマーカーは、(A) 順序付けられたバイオマーカーの初期セグメントが、第二セットの学習データベクトルの補正済み分類のパーセンテージに関しては、該順序付けられたバイオマーカーの他の初期セグメントの中の極大であり、およびここで第二セットの学習データベクトルの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトの該バイオマーカーのセットの少なくとも各バイオマーカーのバイオマーカー測定値を含み、該第二セットの学習データベクトルの各学習データベクトルがさらに対応するヒトの病状に関する分類を含んでいる、それぞれのバイオマーカーに関する 2 群の濃度測定値のマージン分布の中心傾向の関数によって最大から最小に順序付けられたバイオマーカーの該初期セグメントの中にあるか、または (B) (A) で同定したバイオマーカーの該初期セグメント中に含まれているバイオマーカーに関する一次相互作用器であるかのいずれかである、試験データを分類する方法。

(項目 57)

バイオマーカーのセットそれぞれの複数のバイオマーカー測定値を含む試験データを分類する方法であって、

ヒト試験対象に関する試験データを受信する工程、ここで該試験データは、バイオマーカーのセットの少なくとも各バイオマーカーのバイオマーカー測定値を含んでおり；

電子的に記憶されている学習データベクトルの第一のセットを使用して訓練したアダプティブ分類器の電子表示を使用して該試験データを評価する工程、ここで、学習データベクトルの該第一のセットの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトの該バイオマーカーのセットの少なくとも各バイオマーカーのバイオマーカー測定値を含み、学習データベクトルの該第一のセットの各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでおり；および、

該評価する工程に基づいた該ヒト試験対象の分類を出力する工程；を含み、

ここで、該バイオマーカーのセットに含まれているそれぞれのバイオマーカーは、(A) 順序付けられたバイオマーカーの初期セグメントが、第二セットの学習データベクトルの補正済み分類のパーセンテージに関しては、該順序付けられたバイオマーカーの他の初期セグメントの中の極大であり、および、ここで第二セットの学習データベクトルの各学習データベクトルが個々のヒトを表し、かつ、対応するヒトの該バイオマーカーのセットの少なくとも各バイオマーカーのバイオマーカー測定値を含み、該第二セットの学習データベクトルの各学習データベクトルがさらに、対応するヒトの病状に関する分類を含んでいる、それぞれのバイオマーカーに関する 2 群の濃度測定値のマージン分布の中心傾向の関数によって最大から最小に順序付けられたバイオマーカーの該初期セグメントの中にあるか、または (B) (A) で同定したバイオマーカーの該初期セグメント中に含まれているバイオマーカーに関する一次相互作用器であるかのいずれかである、試験データを分類する方法。

10

20

30

40

50

(項目58)

該バイオマーカのセットに含まれているそれぞれのバイオマーカが、空集合から開始して、さらなるバイオマーカを追加することが、該第二の学習データベクトルのセットの正しい分類のパーセンテージを少なくとも20%の閾値まで改良しなくなるまで、該第二の学習データベクトルのセットの正しい分類のパーセンテージを最大限に改良するバイオマーカを前のセットに再帰的に追加することによって生成されるバイオマーカのセットに含まれている、項目56または57に記載の方法。

(項目59)

該閾値が15%である、項目58の方法。

(項目60)

該閾値が10%である、項目58の方法。

(項目61)

該閾値が5%である、項目58の方法。

(項目62)

該閾値が2%である、項目58の方法。

(項目63)

該閾値が1%である、項目58の方法。

(項目64)

該閾値が0.5%である、項目58の方法。

(項目65)

該閾値が0.1%である、項目58の方法。

(項目66)

該閾値が0.01%である、項目58の方法。

(項目67)

学習データベクトルの該第一のセットが、学習データベクトルの該第二のセットと同じである、項目58の方法。

(項目68)

学習データベクトルの該第一のセットの各ベクトルおよび学習データベクトルの該第二のセットが無作為に選択される、項目58~67のいずれか1項に記載の方法。

(項目69)

学習データベクトルの該第一のセットが、少なくとも50のベクトルを含む、項目58~68のいずれか1項に記載の方法。

(項目70)

学習データベクトルの該第二のセットが少なくとも50のベクトルを含む、項目58~69のいずれか1項に記載の方法。

(項目71)

電子記憶装置中の学習データベクトルの該第一のセットにアクセスする工程および電子記憶装置中の学習データベクトルの該第二のセットにアクセスする工程をさらに含む、項目58~70のいずれか1項に記載の方法。

(項目72)

ヒト試験対象に関する試験データを受信するようにプログラムされている電子的なコンピュータを含む、バイオマーカのセットそれぞれの複数のバイオマーカ測定値を含む該試験データを分類するためのシステムであって、該試験データは、該バイオマーカのセットの少なくとも各バイオマーカのバイオマーカ測定値を含んでおり、および電子的に記憶されている学習データベクトルの第一のセットを使って訓練したサポートベクトルマシンの電子表示を使用して該試験データを評価するために、学習データベクトルの第一のセットの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトの該バイオマーカのセットの少なくとも各バイオマーカのバイオマーカ測定値を含み、学習データベクトルの第一のセットの各学習データベクトルは、対応するヒトの病状に関する分類を含み；

10

20

30

40

50

ここで、該コンピューターはさらに、該サポートベクトルマシンの該電子表示に基づいた該ヒト試験対象の分類を出力するためにさらにプログラムされており；

ここで、該バイオマーカのセットに含まれているそれぞれのバイオマーカは：（Ａ）順序付けられたバイオマーカの初期セグメントが、第二セットの学習データベクトルの補正済み分類のパーセンテージに関しては、該順序付けられたバイオマーカの他の初期セグメントの中の極大であり、および、ここで第二セットの学習データベクトルの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトのバイオマーカのセットの少なくとも各バイオマーカのバイオマーカ測定値を含み、第二セットの学習データベクトルの各学習データベクトルはさらに、対応するヒトの病状に関する分類を含んでいる、それぞれのバイオマーカに関する２群の濃度測定値のマージン分布の中心傾向の関数によって最大から最小に順序付けられたバイオマーカの該初期セグメントの中にあるか、または（Ｂ）（Ａ）で同定したバイオマーカの該初期セグメント中に含まれているバイオマーカの一次相互作用器であるか、のいずれかである、前記システム。

（項目 7 3）

ヒト試験対象に関する試験データを受信するようにプログラムされている電子的なコンピューターを含む、バイオマーカのセットそれぞれの複数のバイオマーカ測定値を含む該試験データを分類するためのシステムであって、該試験データは、バイオマーカのセットの少なくとも各バイオマーカのバイオマーカ測定値を含んでおり、および電子的に記憶されている、学習データベクトルの第一のセットを使って訓練したアダプスト分類器の電子表示を使用して該試験データを評価するために、学習データベクトルの第一のセットの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトの該バイオマーカのセットの少なくとも各バイオマーカのバイオマーカ測定値を含み、学習データベクトルの第一のセットの各学習データベクトルは、対応するヒトの病状に関する分類を含み；

ここで、該コンピューターはさらに、該アダプスト分類器の該電子表示に基づいた該ヒト試験対象の分類を出力するためにさらにプログラムされており；

ここで、該バイオマーカのセットに含まれているそれぞれのバイオマーカは：（Ａ）順序付けられたバイオマーカの初期セグメントが、第二セットの学習データベクトルの補正済み分類のパーセンテージに関しては、該順序付けられたバイオマーカの他の初期セグメントの中の極大であり、および、ここで第二セットの学習データベクトルの各学習データベクトルは個々のヒトを表し、かつ、対応するヒトのバイオマーカのセットの少なくとも各バイオマーカのバイオマーカ測定値を含み、第二セットの学習データベクトルの各学習データベクトルがさらに、対応するヒトの病状に関する分類を含んでいる、それぞれのバイオマーカに関する２群の濃度測定値のマージン分布の中心傾向の関数によって最大から最小に順序付けられたバイオマーカの該初期セグメントの中にあるか、または（Ｂ）（Ａ）で同定したバイオマーカの該初期セグメント中に含まれているバイオマーカの一次相互作用器であるか、のいずれかである、前記システム。

（項目 7 4）

該バイオマーカのセットに含まれているそれぞれのバイオマーカが、空集合から開始して、さらなるバイオマーカを追加することが、該第二の学習データベクトルのセットの正しい分類のパーセンテージを少なくとも 20 % の閾値まで改良しなくなるまで、該第二の学習データベクトルのセットの正しい分類のパーセンテージを最大限に改良するバイオマーカを前のセットに再帰的に追加することによって生成されるバイオマーカのセットに含まれている、項目 7 2 または 7 3 に記載のシステム。

（項目 7 5）

該閾値が 15 % である、項目 7 4 のシステム。

（項目 7 6）

該閾値が 10 % である、項目 7 4 のシステム。

（項目 7 7）

該閾値が 5 % である、項目 7 4 のシステム。

10

20

30

40

50

( 項目 7 8 )

該 閾 値 が 2 % である、 項 目 7 4 の シ ス テ ム 。

( 項目 7 9 )

該 閾 値 が 1 % である、 項 目 7 4 の シ ス テ ム 。

( 項目 8 0 )

該 閾 値 が 0 . 5 % である、 項 目 7 4 の シ ス テ ム 。

( 項目 8 1 )

該 閾 値 が 0 . 1 % である、 項 目 7 4 の シ ス テ ム 。

( 項目 8 2 )

該 閾 値 が 0 . 0 1 % である、 項 目 7 4 の シ ス テ ム 。

( 項目 8 3 )

学 習 デ ー タ ベ ク ト ル の 第 一 の セ ッ ト が、 該 学 習 デ ー タ ベ ク ト ル の 第 二 の セ ッ ト と 同 じ で 有 る、 項 目 7 4 の シ ス テ ム 。

( 項目 8 4 )

学 習 デ ー タ ベ ク ト ル の 第 一 の セ ッ ト の 各 ベ ク ト ル お よ び 該 学 習 デ ー タ ベ ク ト ル の 第 二 の セ ッ ト が 無 作 為 に 選 択 さ れ る、 項 目 7 2 ~ 8 3 の い ず れ か 1 項 に 記 載 の シ ス テ ム 。

( 項目 8 5 )

学 習 デ ー タ ベ ク ト ル の 第 一 の セ ッ ト が、 少 なく とも 5 0 の ベ ク ト ル を 含 む、 項 目 7 2 ~ 8 4 の い ず れ か 1 項 に 記 載 の シ ス テ ム 。

( 項目 8 6 )

該 学 習 デ ー タ ベ ク ト ル の 第 二 の セ ッ ト が 少 なく とも 5 0 の ベ ク ト ル を 含 む、 項 目 7 2 ~ 8 5 の い ず れ か 1 項 に 記 載 の シ ス テ ム 。

( 項目 8 7 )

該 電 子 的 コ ン ピ ュ ー タ ー に 操 作 可 能 に 接 続 さ れ て い る 電 子 記 憶 装 置 に 含 ま れ て い る 学 習 デ ー タ ベ ク ト ル の 第 一 の セ ッ ト お よ び 該 電 子 的 コ ン ピ ュ ー タ ー に 操 作 可 能 に 接 続 さ れ て い る 電 子 記 憶 装 置 に 含 ま れ て い る 学 習 デ ー タ ベ ク ト ル の 該 第 二 の セ ッ ト を さ ら に 含 む、 項 目 7 2 ~ 8 6 の い ず れ か 1 項 に 記 載 の シ ス テ ム 。

10

20

【図1A】

図1A

胚盤帯中の遺伝子発現レベル						
バイオマーカー	平均	標準偏差	バイオマーカー	平均	標準偏差	バイオマーカー
アディポネクチン	2259.55	1504.98	IL-17	37.60	13.32	33.95
レジスタチン	124.37	90.22	IL-1e	95.36	185.32	194.13
PAI-1 <sup>A</sup>	335.10	371.63	IFN- $\gamma$	23.30	9.28	39.81
S E-セレクチン	60.34	47.58	G-CSF	23.84	4.12	17.29
sVCAM-1	1753.31	313.04	GM-CSF	28.60	17.57	45.94
sICAM-1	2925.75	1574.55	TNF- $\alpha$	43.52	15.71	36.10
MPO	592.23	1503.40	MCP-1	46.39	29.47	63.51
CRP	8384.00	6292.94	IL-12 (p40)	34.93	7.26	29.11
SA A	2202.71	4017.36	MIP-1a	41.04	27.61	67.27
SAP	2203.71	3703.75	MIP-1 $\beta$	21.28	9.86	46.35
レプティン <sup>1</sup>	2204.62	2771.57	VEGF	40.80	11.45	28.07
GLP-1	56.21	102.46	IL12 (p70)	7.66	36.69	479.00
アミリン (B)	142.47	339.44	IL-13	11.41	57.04	499.86
G-ベータアド	4748.58	1971.09	MMP-2	155.73	439.49	282.21
インスリン	278.12	424.50	MMP-1	103.52	179.41	173.71
Sfns	38.81	148.38	MMP-3	6608.81	2442.77	36.63
sFSI	15.13	16.89	エオシノキリン	196.78	208.46	105.94
MIF	52.08	112.39	レプティン <sup>2</sup>	3767.80	3883.08	103.11
IL-1 $\beta$	23.71	9.52	IP-10	543.43	820.15	150.92
IL-2	12.26	44.02	MMP-9	1090.51	1599.16	146.64
IL-1m	23.11	19.35	MMP-13	9.35	35.67	335.67
IL-4	114.39	72.33	PAI-1 <sup>B</sup>	33.94	98.35	252.58
IL-5	18.37	18.35	I-TAC	30.16	154.83	313.40
IL-6	40.99	56.56	MMP-12	12.41	27.69	233.09
IL-7	15.82	6.61	HGF	206.22	619.73	300.92
TGF- $\alpha$	31.46	8.00	MMP-7	1094.07	577.02	373.23
フクタルキリン	13.36	3.35	EGF	30.91	49.02	138.60
IL-8	239.47	629.22	sCD40リガンド	153.50	434.04	279.12
IL-10	13.06	3.63	MMP-8	70.26	118.85	453.80
IL-15	24.53	4.27				

【図1C】

図1C

胎盤中の遺伝子発現レベル						
バイオマーカー	平均	標準偏差	バイオマーカー	平均	標準偏差	バイオマーカー
アディポネクチン	2963.37	1855.21	IL-17	48.31	24.48	50.68
レジスタチン	290.12	347.80	IL-1e	250.35	734.45	293.38
PAI-1 <sup>A</sup>	791.20	259.58	IFN- $\gamma$	27.57	6.28	23.79
S E-セレクチン	39.47	26.26	G-CSF	27.42	12.61	45.98
sVCAM-1	2580.30	1090.17	GM-CSF	35.60	36.94	75.67
sICAM-1	3202.40	1322.66	TNF- $\alpha$	69.23	272.27	393.15
MPO	3799.80	2433.44	MCP-1	436.56	1969.47	451.13
CRP	10529.32	5894.42	IL-12 (p40)	26.81	8.15	30.41
SA A	3837.04	3965.64	MIP-1a	97.63	647.11	662.84
SAP	2336.72	646.89	MIP-1 $\beta$	26.41	32.51	123.10
レプティン <sup>1</sup>	3009.58	2925.20	VEGF	46.98	18.08	38.48
GLP-1	106.06	513.75	IL12 (p70)	32.48	34.89	168.40
アミリン (B)	274.00	1097.69	IL-13	297.79	362.17	121.62
G-ベータアド	5505.85	2653.46	MMP-2	184.87	248.32	134.32
インスリン	407.57	1178.79	MMP-1	192.84	280.81	145.61
Sfns	76.93	40.19	MMP-3	6150.32	2828.47	45.99
sFSI	31.20	48.27	エオシノキリン	652.21	484.14	74.23
MIF	61.77	80.34	レプティン <sup>2</sup>	4322.85	3756.93	86.91
IL-1 $\beta$	34.22	117.74	IP-10	1575.99	1241.02	78.75
IL-2	12.75	3.38	MMP-9	4097.20	2679.44	65.40
IL-1m	36.25	3.19	MMP-13	9.37	38.54	411.12
IL-4	190.65	318.74	PAI-1 <sup>B</sup>	461.47	358.67	77.72
IL-5	36.72	73.10	I-TAC	783.04	637.12	81.36
IL-6	132.01	548.69	MMP-12	20.79	42.60	204.89
IL-7	23.90	13.78	HGF	1381.06	921.98	65.76
TGF- $\alpha$	38.36	16.97	MMP-7	487.07	493.70	101.36
フクタルキリン	15.87	3.41	EGF	419.64	417.35	99.30
IL-8	829.39	2332.42	sCD40リガンド	175.32	259.04	147.75
IL-10	15.64	4.29	MMP-8	140.35	889.83	634.00
IL-15	27.01	7.96				

【図1B】

図1B

胎盤中の遺伝子発現レベル						
バイオマーカー	平均	標準偏差	バイオマーカー	平均	標準偏差	バイオマーカー
アディポネクチン	3563.76	2288.94	IL-17	46.36	29.90	64.51
レジスタチン	191.29	136.17	IL-1e	122.68	394.79	321.80
PAI-1 <sup>A</sup>	787.03	389.14	IFN- $\gamma$	28.31	9.99	35.29
S E-セレクチン	41.71	45.85	G-CSF	29.55	19.35	63.51
sVCAM-1	1422.20	548.36	GM-CSF	33.26	10.06	30.23
sICAM-1	3121.23	1440.47	TNF- $\alpha$	75.33	25.82	47.69
MPO	2188.38	1952.85	MCP-1	183.88	210.11	114.26
CRP	12951.79	5490.37	IL-12 (p40)	26.01	8.07	31.01
SA A	7383.07	6685.22	MIP-1a	53.41	97.52	182.57
SAP	1663.82	907.73	MIP-1 $\beta$	33.74	81.49	241.55
レプティン <sup>1</sup>	1441.00	1806.76	VEGF	43.93	17.22	30.19
GLP-1	130.29	700.59	IL12 (p70)	46.82	32.33	197.24
アミリン (B)	337.43	1028.44	IL-13	401.00	520.85	129.89
G-ベータアド	3431.59	2603.38	MMP-2	165.31	320.62	193.95
インスリン	373.06	777.90	MMP-1	744.40	766.38	102.95
Sfns	68.60	41.26	MMP-3	8137.21	3111.85	58.24
sFSI	18.15	18.41	エオシノキリン	818.65	576.64	76.44
MIF	76.00	155.45	レプティン <sup>2</sup>	3214.92	3365.20	102.81
IL-1 $\beta$	27.08	4.32	IP-10	1846.73	1381.81	74.82
IL-2	11.66	3.89	MMP-9	3141.11	2612.52	83.30
IL-1m	29.38	9.12	MMP-13	21.06	189.11	898.16
IL-4	154.43	214.60	PAI-1 <sup>B</sup>	268.11	201.13	75.02
IL-5	26.12	43.27	I-TAC	429.88	313.58	73.43
IL-6	74.67	185.61	MMP-12	32.74	168.74	513.37
IL-7	23.67	73.58	HGF	678.43	787.38	116.66
TGF- $\alpha$	37.56	12.62	MMP-7	2070.90	1166.98	56.35
フクタルキリン	15.27	4.57	EGF	455.13	462.40	101.69
IL-8	168.26	286.40	sCD40リガンド	120.25	367.11	305.30
IL-10	20.38	47.38	MMP-8	284.02	494.63	178.15
IL-15	28.90	7.64				

【図1D】

図1D

平均値標準偏差の変化 (%)						
バイオマーカー	胎盤対正常	胎盤対胎盤	胎盤対胎盤	バイオマーカー	胎盤対正常	胎盤対胎盤
アディポネクチン	31.15	57.72	-20.36	IL-17	28.48	4.04
レジスタチン	133.28	53.81	34.97	IL-1e	162.52	28.65
PAI-1 <sup>A</sup>	136.11	134.86	0.53	IFN- $\gamma$	18.31	21.46
S E-セレクチン	-37.99	-30.88	-11.46	G-CSF	15.02	23.95
sVCAM-1	47.17	-18.89	44.88	GM-CSF	24.45	16.28
sICAM-1	-2.83	-5.28	2.52	TNF- $\alpha$	59.14	73.10
MPO	541.07	269.21	42.41	MCP-1	841.00	296.25
CRP	23.59	84.48	-23.01	IL-12 (p40)	7.53	4.30
SA A	65.12	235.18	-103.00	MIP-1a	137.86	30.14
SAP	6.18	-24.40	28.80	MIP-1 $\beta$	24.09	58.53
レプティン <sup>1</sup>	36.51	-34.64	52.12	VEGF	15.13	7.67
GLP-1	88.68	131.77	-22.84	IL12 (p70)	324.00	511.21
アミリン (B)	92.33	126.84	-23.15	IL-13	2509.42	3413.77
G-ベータアド	16.01	-27.73	37.71	MMP-2	18.71	6.15
インスリン	45.55	34.14	8.47	MMP-1	86.29	619.10
Sfns	30.81	16.64	10.83	MMP-3	-7.77	22.02
sFSI	106.17	19.89	41.85	エオシノキリン	231.44	316.02
MIF	18.62	45.93	-23.03	レプティン <sup>2</sup>	14.73	-14.67
IL-1 $\beta$	44.28	14.18	20.86	IP-10	190.61	239.83
IL-2	3.97	-4.95	8.58	MMP-9	275.71	188.04
IL-1m	13.62	27.16	-11.92	MMP-13	-5.75	11.71
IL-4	66.67	35.90	19.01	PAI-1 <sup>B</sup>	1885.19	888.59
IL-5	99.90	42.20	28.87	I-TAC	2496.41	1324.07
IL-6	222.04	82.15	43.44	MMP-12	67.51	163.78
IL-7	47.25	49.61	-1.60	HGF	569.70	238.98
TGF- $\alpha$	21.95	19.41	2.08	MMP-7	-51.49	106.25
フクタルキリン	18.84	14.34	3.78	EGF	1257.75	1372.57
IL-8	19.70	55.21	-29.66	sCD40リガンド	12.74	-22.67
IL-10	10.11	17.84	-7.02	MMP-8	99.75	304.23
IL-15						

【図2A】

図2A

バイオマーカー	平均	標準偏差	標準偏差	バイオマーカー	平均	標準偏差	標準偏差
PAI-1	2116.17	1225.14	57.89	IL-17	41.58	12.84	30.85
PAI-1 <sup>Δ</sup>	158.01	85.09	60.18	IL-1α	83.31	155.36	186.68
PAI-1 <sup>Δ</sup>	499.52	438.77	87.84	IFN-γ	26.73	11.79	44.13
SE-セレクチン	75.37	56.07	74.40	G-CSF	26.24	3.43	13.09
sVCAM-1	1658.49	448.12	27.02	GM-CSF	30.93	4.43	14.32
sICAM-1	3520.37	1512.81	42.97	TNF-α	54.88	13.94	25.40
MPO	1269.64	2084.74	164.20	MCP-1	59.23	39.47	66.53
CRP	7532.14	5561.16	73.83	IL-12(p40)	27.09	6.13	22.64
SAA	1773.59	3197.93	130.31	MIP-1α	20.13	39.77	79.24
SAP	2238.83	555.16	24.80	MIP-1β	26.92	11.02	40.86
レプテン <sup>1</sup>	860.83	1311.22	133.32	VEGF	43.79	9.56	21.83
GLP-1	54.12	99.58	184.01	IL-12(p70)	7.87	11.21	142.41
アミリン(全)	161.50	469.30	200.59	IL-13	27.17	74.25	273.33
α-ペプチド	4648.15	1725.03	42.61	MMP-2	132.15	148.40	112.30
インスリン	287.45	483.63	161.29	MMP-1	182.66	242.24	132.62
Sfas	63.98	42.36	66.21	MMP-3	7497.26	2419.60	32.40
sFSI	13.35	10.94	81.94	エオタキシン	289.67	253.39	87.48
MIF	105.73	157.68	149.14	レプテン <sup>2</sup>	2169.69	2831.50	130.50
IL-1β	26.63	13.82	51.90	IP-10	452.38	410.35	90.67
IL-2	11.04	2.52	21.05	MMP-9	1715.45	2162.04	126.18
IL-1ra	24.48	3.85	15.71	MMP-13	14.07	51.00	362.52
IL-4	105.72	60.09	57.05	PAI-1 <sup>Δ</sup>	66.22	133.55	171.67
IL-5	18.93	19.27	101.80	L-TAC	92.89	207.98	223.91
IL-6	43.44	57.12	131.50	MMP-12	15.55	36.33	233.59
IL-7	17.80	6.28	35.26	HGF	538.75	846.22	157.07
TGF-α	34.76	7.54	21.70	MMP-7	1163.17	547.89	47.09
フラクタルキン	14.96	2.95	19.71	EGF	48.56	64.54	132.92
IL-8	401.98	913.65	221.79	sCD40リガンド	139.86	213.56	152.69
IL-10	14.76	3.07	20.81	MMP-8	222.46	446.02	200.50
IL-15	26.97	3.76	13.93				

【図2B】

図2B

バイオマーカー	平均	標準偏差	標準偏差	バイオマーカー	平均	標準偏差	標準偏差
PAI-1	4041.07	2456.96	60.80	IL-17	45.15	20.95	46.40
PAI-1 <sup>Δ</sup>	165.14	104.34	63.18	IL-1α	101.95	454.84	446.12
PAI-1 <sup>Δ</sup>	680.06	281.74	41.43	IFN-γ	29.47	16.86	36.86
SE-セレクチン	39.28	24.76	63.02	G-CSF	28.03	7.27	25.95
sVCAM-1	1486.20	580.24	59.94	GM-CSF	33.00	6.80	20.61
sICAM-1	3129.94	1388.46	44.36	TNF-α	113.63	639.16	562.50
MPO	2236.00	2021.46	90.41	MCP-1	175.20	224.24	127.99
CRP	12375.56	6171.04	49.87	IL-12(p40)	25.63	9.73	37.96
SAA	6407.43	6756.26	105.44	MIP-1α	67.31	169.27	251.61
SAP	1610.90	912.77	56.66	MIP-1β	41.90	143.28	341.91
レプテン <sup>1</sup>	1836.07	2086.83	113.66	VEGF	41.74	11.81	85.59
GLP-1	112.59	751.23	667.22	IL-12(p70)	38.24	142.89	245.35
アミリン(全)	252.51	919.55	364.69	IL-13	410.97	510.37	124.18
α-ペプチド	3270.49	2492.50	76.21	MMP-2	139.64	205.55	147.20
インスリン	338.74	801.56	236.63	MMP-1	602.07	546.07	90.70
sFSI	64.72	36.09	55.76	MMP-3	7690.73	3309.64	43.03
sFSI	13.51	21.88	161.59	エオタキシン	814.32	373.07	70.37
MIF	55.10	97.46	176.88	レプテン <sup>2</sup>	3933.88	3884.56	98.24
IL-1β	27.70	4.44	16.03	IP-10	1631.25	1066.13	65.36
IL-2	12.03	5.77	47.98	MMP-9	3423.98	2497.43	73.94
IL-1ra	29.50	8.55	28.23	MMP-13	11.66	33.41	286.54
IL-4	130.02	173.21	133.21	PAI-1 <sup>Δ</sup>	244.76	206.20	84.24
IL-5	22.15	44.44	199.26	L-TAC	387.53	514.80	81.23
IL-6	70.54	257.26	365.41	MMP-12	24.93	36.33	145.74
IL-7	31.77	130.07	409.42	HGF	537.31	620.75	152.53
TGF-α	36.97	12.35	33.42	MMP-7	2122.99	1114.85	52.51
フラクタルキン	15.30	4.82	31.48	EGF	422.32	483.79	114.56
IL-8	126.83	168.86	133.13	sCD40リガンド	99.65	261.62	262.55
IL-10	27.03	84.44	312.39	MMP-8	247.64	364.87	147.34
IL-15	28.91	6.61	22.86				

【図2C】

図2C

バイオマーカー	平均	標準偏差	標準偏差	バイオマーカー	平均	標準偏差	標準偏差
PAI-1	2944.64	1614.58	54.83	IL-17	43.39	21.49	49.53
PAI-1 <sup>Δ</sup>	332.29	416.68	125.40	IL-1α	164.86	441.25	267.64
PAI-1 <sup>Δ</sup>	841.92	266.60	31.67	IFN-γ	27.88	6.03	21.64
SE-セレクチン	36.85	25.26	68.56	G-CSF	21.06	4.35	20.67
sVCAM-1	2435.16	971.17	39.88	GM-CSF	28.29	5.43	19.20
sICAM-1	3084.39	1423.90	46.23	TNF-α	43.88	8.91	20.30
MPO	4086.85	2406.72	59.32	MCP-1	337.96	1426.34	333.28
CRP	11594.22	6148.46	53.45	IL-12(p40)	23.53	5.49	23.34
SAA	3764.05	3808.32	101.18	MIP-1α	47.70	29.11	61.04
SAP	2216.83	576.86	26.02	MIP-1β	22.11	6.60	29.87
レプテン <sup>1</sup>	4148.56	3054.11	73.62	VEGF	40.46	9.69	23.95
GLP-1	153.82	647.05	420.66	IL-12(p70)	30.32	33.08	175.08
アミリン(全)	377.66	1265.49	335.09	IL-13	307.87	373.66	121.37
α-ペプチド	5897.41	2865.27	48.59	MMP-2	222.30	297.87	134.00
インスリン	521.74	1464.79	289.73	MMP-1	239.58	326.58	136.31
Sfas	88.02	45.72	51.94	MMP-3	6158.26	2312.49	37.55
sFSI	26.67	43.10	161.59	エオタキシン	676.90	441.51	65.23
MIF	50.56	92.56	183.07	レプテン <sup>2</sup>	5870.26	3702.57	63.07
IL-1β	23.81	4.81	20.19	IP-10	1698.22	1212.85	71.42
IL-2	12.46	3.21	25.76	MMP-9	4522.93	2746.21	60.72
IL-1ra	26.42	5.68	21.51	MMP-13	11.59	48.42	417.84
IL-4	147.54	195.73	132.66	PAI-1 <sup>Δ</sup>	512.52	379.29	74.03
IL-5	27.72	42.32	132.66	L-TAC	890.89	677.56	166.06
IL-6	69.69	142.44	204.39	MMP-12	24.82	51.96	209.35
IL-7	18.13	7.94	43.80	HGF	1549.14	858.93	55.45
TGF-α	32.08	15.94	49.69	MMP-7	517.14	462.20	89.38
フラクタルキン	15.25	2.85	18.70	EGF	405.07	357.97	88.37
IL-8	704.34	1718.95	244.09	sCD40リガンド	213.58	312.95	146.52
IL-10	14.95	4.15	27.78	MMP-8	192.74	1107.98	574.85
IL-15	22.54	4.34	19.27				

【図2D】

図2D

バイオマーカー	前症対正常	前症対正常	前症対正常	バイオマーカー	前症対正常	前症対正常	前症対正常
PAI-1	39.15	90.26	37.43	IL-17	4.37	8.59	-4.04
PAI-1 <sup>Δ</sup>	110.29	4.51	50.50	IL-1α	97.89	22.38	-36.16
PAI-1 <sup>Δ</sup>	68.55	36.14	19.23	IFN-γ	4.32	10.26	-5.70
SE-セレクチン	-51.10	-47.88	-6.59	G-CSF	19.75	6.81	-33.10
sVCAM-1	-6.83	-10.39	-8.97	GM-CSF	-8.54	6.71	-16.67
sICAM-1	-12.38	-11.99	-1.48	TNF-α	-20.04	107.06	-158.93
MPO	213.53	75.11	17.88	MCP-1	369.66	1658.11	48.16
CRP	-52.74	64.28	-7.56	IL-12(p40)	13.17	5.49	-8.95
SAA	112.23	261.27	-70.23	MIP-1α	-4.86	34.37	-41.12
SAP	-0.98	-28.85	-27.33	MIP-1β	-17.86	55.88	-89.54
レプテン <sup>1</sup>	381.92	113.59	55.74	VEGF	-7.61	-6.67	-3.18
GLP-1	184.22	108.05	26.80	IL-12(p70)	285.01	639.61	-92.10
アミリン(全)	153.85	56.35	33.14	IL-13	1033.27	1112.51	-33.49
α-ペプチド	45.83	24.21	11.57	MMP-2	6.82	5.67	31.18
インスリン	-81.51	17.85	35.07	MMP-1	31.16	232.61	-151.36
Sfas	-7.56	1.15	26.47	MMP-3	-17.53	2.99	-24.88
sFSI	99.73	1.15	49.36	エオタキシン	133.68	181.12	-20.30
MIF	-52.18	-47.89	-8.98	レプテン <sup>2</sup>	170.56	81.31	32.99
IL-1β	-10.60	3.99	-16.33	IP-10	378.23	269.44	3.94
IL-2	12.92	8.99	3.48	MMP-9	163.91	92.83	24.29
IL-1ra	7.92	20.61	-11.63	MMP-13	-47.64	171.11	-0.64
IL-4	40.16	23.46	11.87	PAI-1 <sup>Δ</sup>	633.63	269.60	374.21
IL-5	46.45	17.02	28.10	L-TAC	859.97	317.20	-86.50
IL-6	60.44	62.39	-1.22	MMP-12	59.60	60.26	-0.42
IL-7	1.86	78.45	-75.19	HGF	187.55	40.27	65.32
TGF-α	-7.71	6.37	-15.26	MMP-7	-55.54	82.52	-310.52
フラクタルキン	1.99	7.80	8.31	EGF	744.23	767.75	-4.76
IL-8	753.10	424.45	81.69	sCD40リガンド	547.72	-28.73	53.55
IL-10	1.34	83.17	-88.75	MMP-8	13.56	11.32	-28.48
IL-15	-16.41	7.20	-28.25				

【 図 3 A 】

図 3 A

正常男性集団での発光強度									
バイオマーカー	平均	標準偏差	標準偏差	バイオマーカー	平均	標準偏差	標準偏差	標準偏差	標準偏差
アディポネクチン	2375.01	1677.13	70.62	IL-17	34.68	13.33	38.44		
レスレチン	99.39	78.12	78.60	IL-1α	104.49	705.17	196.35		
PAI-1 <sup>A</sup>	213.37	254.21	119.14	IFN-γ	20.78	5.73	27.58		
SE-セレクチン	49.28	36.75	74.57	G-CSF	22.05	3.68	16.68		
sVCAM-1	1822.54	548.47	30.09	GM-CSF	26.91	15.97	59.33		
sICAM-1	3128.49	1607.19	51.37	TNF-α	35.13	11.03	31.38		
MPO	95.40	396.70	415.85	MCP-1	36.83	12.40	33.66		
CRP	8947.58	6700.43	74.89	IL-12(p40)	23.35	7.64	32.74		
SAA	2444.04	4420.98	180.89	MIP-1α	34.36	7.76	22.58		
SAP	2170.66	583.20	26.87	MIP-1β	17.13	6.29	36.73		
レプテン <sup>1</sup>	3159.85	3108.95	98.39	VEGF	38.60	12.27	31.79		
GLP-1	57.97	105.08	181.26	IL-12(p70)	7.53	47.57	631.42		
アミリン (総)	127.91	196.50	153.63	IL-13	-0.07	36.12	-48855.51		
ロニズチド	5243.25	1975.58	37.68	MMP-2	173.98	566.49	325.61		
インスリン	269.95	395.42	146.48	MMP-1	45.54	70.31	154.37		
Sfas	54.83	192.79	351.58	MMP-3	6075.64	2300.75	37.87		
sFSI	16.47	20.16	122.40	エノタキシン	127.35	132.34	103.92		
MIF	12.66	6.60	52.09	レプテン <sup>2</sup>	4899.58	4110.65	83.90		
IL-1β	21.56	2.60	12.05	IP-10	612.33	1020.53	166.66		
IL-2	13.17	58.19	441.69	MMP-9	635.66	724.02	115.90		
IL-1ra	22.06	25.34	114.88	MMP-13	6.92	16.91	244.36		
IL-4	121.09	79.91	65.99	PAI-1 <sup>B</sup>	13.51	73.47	543.75		
IL-5	17.97	17.75	98.78	L-TAC	-15.62	70.30	-450.19		
IL-6	39.17	56.42	144.04	MMP-12	10.06	18.84	187.30		
IL-7	14.36	6.51	45.31	HGF	-38.01	32.84	-86.39		
TGF-α	29.04	7.49	25.78	MMP-7	892.29	570.49	63.94		
フラクタルケン	12.16	3.14	25.82	EGF	17.96	27.01	150.43		
IL-8	120.31	200.43	166.59	αCD40リガンド	167.64	543.74	324.35		
IL-10	11.81	5.52	29.85	MMP-8	-41.37	35.20	-85.08		
IL-15	22.73	3.72	16.36						

【 図 3 C 】

図 3 C

喘息男性集団での発光強度レベル									
バイオマーカー	平均	標準偏差	標準偏差	バイオマーカー	平均	標準偏差	標準偏差	標準偏差	標準偏差
アディポネクチン	2947.72	2190.78	74.32	IL-17	56.69	27.10	47.81		
レスレチン	221.87	167.29	75.40	IL-1α	396.04	1049.31	264.95		
PAI-1 <sup>A</sup>	711.75	724.74	31.58	IFN-γ	27.11	6.72	24.79		
SE-セレクチン	38.26	28.16	73.59	G-CSF	38.02	14.71	38.68		
sVCAM-1	2829.19	1012.66	35.79	GM-CSF	47.83	40.94	85.59		
sICAM-1	3394.91	1123.56	53.10	TNF-α	111.92	444.94	397.55		
MPO	3403.65	2428.77	71.36	MCP-1	606.76	2887.62	475.91		
CRP	8963.97	5125.42	57.18	IL-12(p40)	32.29	8.99	27.84		
SAA	3443.76	4262.10	123.76	MIP-1α	182.01	1059.60	582.15		
SAP	2535.86	714.51	28.18	MIP-1β	33.66	52.04	154.58		
レプテン <sup>1</sup>	1141.66	1299.27	113.81	VEGF	57.87	23.15	40.00		
GLP-1	27.56	38.25	138.81	IL-12(p70)	36.55	57.75	158.03		
アミリン (総)	104.12	110.92	106.53	IL-13	284.25	346.16	121.78		
ロニズチド	4893.34	2138.06	43.69	MMP-2	124.45	107.75	86.58		
インスリン	222.02	317.93	143.20	MMP-1	117.09	156.76	133.88		
Sfas	58.05	16.92	29.15	MMP-3	6119.28	3561.09	38.19		
sFSI	38.99	55.65	162.74	エノタキシン	610.56	552.52	90.46		
MIF	80.03	50.32	62.87	レプテン <sup>2</sup>	1764.32	2112.34	119.72		
IL-1β	51.76	192.52	371.57	IP-10	1500.30	1119.80	86.12		
IL-2	13.27	3.63	27.33	MMP-9	3405.05	2443.42	71.76		
IL-1ra	26.01	4.29	16.50	MMP-13	5.81	7.96	136.94		
IL-4	263.83	449.02	170.54	PAI-1 <sup>B</sup>	381.16	307.44	80.66		
IL-5	52.10	109.08	209.57	L-TAC	610.84	524.00	85.78		
IL-6	237.79	874.22	367.65	MMP-12	14.41	17.40	120.78		
IL-7	32.02	16.91	52.81	HGF	1109.68	967.32	87.17		
TGF-α	48.84	13.25	27.13	MMP-7	-437.39	545.63	124.75		
フラクタルケン	16.93	4.02	23.75	EGF	447.62	505.26	112.88		
IL-8	1050.01	3114.51	296.62	αCD40リガンド	111.41	103.15	92.58		
IL-10	16.83	4.30	25.53	MMP-8	55.11	265.37	481.49		
IL-15	34.84	7.08	20.56						

【 図 3 B 】

図 3 B

喘息男性集団での発光強度レベル									
バイオマーカー	平均	標準偏差	標準偏差	バイオマーカー	平均	標準偏差	標準偏差	標準偏差	標準偏差
アディポネクチン	3308.09	2112.68	63.86	IL-17	46.93	33.34	71.03		
レスレチン	203.24	147.53	72.59	IL-1α	152.63	364.95	275.17		
PAI-1 <sup>A</sup>	837.56	421.88	50.37	IFN-γ	27.78	9.55	34.39		
SE-セレクチン	42.79	32.96	123.79	G-CSF	30.21	22.90	75.83		
sVCAM-1	1392.31	532.93	38.28	GM-CSF	33.37	11.28	33.81		
sICAM-1	3120.40	1469.20	47.08	TNF-α	57.62	17.36	30.13		
MPO	2172.28	1925.67	88.65	MCP-1	187.93	203.98	108.54		
CRP	15263.74	5095.50	38.41	IL-12(p40)	26.17	7.19	27.49		
SAA	7854.70	6623.56	84.33	MIP-1α	47.00	23.96	50.98		
SAP	1689.28	907.93	53.75	MIP-1β	30.00	14.79	49.29		
レプテン <sup>1</sup>	1241.43	1616.24	130.19	VEGF	44.89	19.17	42.71		
GLP-1	138.75	679.26	489.56	IL-12(p70)	41.57	54.79	131.80		
アミリン (総)	377.72	1076.85	285.09	IL-13	397.79	527.21	132.54		
ロニズチド	3318.52	2655.51	75.42	MMP-2	177.06	362.25	204.59		
インスリン	388.05	769.26	198.24	MMP-1	311.18	843.83	104.03		
Sfas	70.28	43.45	61.83	MMP-3	8334.20	2930.06	35.92		
sFSI	20.35	69.16	339.84	エノタキシン	815.97	575.92	70.58		
MIF	85.95	175.61	204.33	レプテン <sup>2</sup>	2860.57	2948.31	103.07		
IL-1β	26.79	4.24	15.84	IP-10	1949.78	1499.56	76.91		
IL-2	11.49	5.97	51.93	MMP-9	3017.93	2667.24	88.38		
IL-1ra	29.34	9.50	32.36	MMP-13	25.44	228.13	896.89		
IL-4	165.98	231.18	139.28	PAI-1 <sup>B</sup>	279.67	198.30	70.91		
IL-5	28.01	42.91	153.18	L-TAC	450.09	314.46	69.87		
IL-6	76.81	141.04	153.63	MMP-12	36.42	203.07	537.58		
IL-7	19.93	9.24	46.38	HGF	746.89	847.21	113.43		
TGF-α	37.83	12.78	33.77	MMP-7	2049.58	1193.44	58.23		
フラクタルケン	15.26	4.16	27.25	EGF	471.66	452.29	96.00		
IL-8	187.89	325.97	173.49	αCD40リガンド	129.51	407.77	314.83		
IL-10	17.14	6.35	37.07	MMP-8	300.58	545.34	181.43		
IL-15	28.87	7.24	25.09						

【 図 3 D 】

図 3 D

喘息男性集団における平均発光強度の変化 (%)											
バイオマーカー	前測定	後測定	変化率 (%)	バイオマーカー	前測定	後測定	変化率 (%)	バイオマーカー	前測定	後測定	変化率 (%)
アディポネクチン	24.11	39.52	12.33	IL-17	4.83	3.88	-20.08	IL-17	4.83	3.88	-20.08
レスレチン	123.21	104.49	8.40	IL-1α	279.84	260.93	-6.51	IL-1α	279.84	260.93	-6.51
PAI-1 <sup>A</sup>	233.88	292.55	17.68	IFN-γ	30.45	33.66	2.44	IFN-γ	30.45	33.66	2.44
SE-セレクチン	-22.36	-13.18	-11.82	G-CSF	72.41	56.96	-20.36	G-CSF	72.41	56.96	-20.36
sVCAM-1	352.23	224.61	-50.79	GM-CSF	77.75	24.62	-68.22	GM-CSF	77.75	24.62	-68.22
sICAM-1	852	41.26	-95.51	TNF-α	218.55	64.00	-70.74	TNF-α	218.55	64.00	-70.74
MPO	3463.91	2172.11	-37.28	MCP-1	1547.33	410.23	-73.51	MCP-1	1547.33	410.23	-73.51
CRP	0.18	-8.26	-47.99	IL-12(p40)	38.29	12.49	-67.64	IL-12(p40)	38.29	12.49	-67.64
SAA	40.90	221.38	128.99	MIP-1α	229.75	36.79	-84.11	MIP-1α	229.75	36.79	-84.11
SAP	16.81	-22.18	-33.38	MIP-1β	96.57	75.16	-21.64	MIP-1β	96.57	75.16	-21.64
レプテン <sup>1</sup>	-63.87	-60.71	3.74	VEGF	49.91	16.29	-67.42	VEGF	49.91	16.29	-67.42
GLP-1	524.7	139.24	-73.43	IL-12(p70)	335.11	451.75	33.34	IL-12(p70)	335.11	451.75	33.34
アミリン (総)	-18.59	155.41	838.18	IL-13	346539.50	-53869.98	-15.26	IL-13	346539.50	-53869.98	-15.26
ロニズチド	-6.67	-32.90									

【 図 4 】

遺伝子	男性集団と女性集団の平均発現量の差 (%) の比較		遺伝子の発現量	遺伝子の発現量	遺伝子の発現量	遺伝子の発現量
	男性の発現量	女性の発現量				
ハイオクサーカー	0.10	-0.90	18.14	18.14	3.98	5.18
アロペクチン	-3.23	-8.92	14.22	14.22	30.08	20.27
PAI-1	15.46	14.12	184.12	184.12	-5.75	2.62
SE-エペリン	3.83	8.93	54.03	54.03	30.57	13.99
AVCAM1	16.18	-6.32	-9.00	-9.00	159.08	14.84
MICAM1	0.07	-0.30	12.53	12.53	155.04	56.20
CD133	5.10	1.15	11.81	11.81	9.47	11.27
CDP	-22.40	7.21	18.53	18.53	27.24	31.07
SAA	8.51	22.59	-27.43	-27.43	201.56	45.91
SNP	14.30	4.87	3.14	3.14	52.27	28.41
レプティン	-72.48	-32.39	-72.76	-72.76	-33.04	7.55
EGF	-82.09	23.23	-6.65	-6.65	11.12	4.52
アミリン (全)	-22.43	49.39	26.27	26.27	-7.67	-3.21
アミリン (自由)	17.03	1.78	-22.29	-22.29	-4.01	26.80
インスリン	1.26	1.26	6.13	6.13	24.04	24.04
SNB	-6.15	30.64	-18.92	-18.92	-0.83	8.24
INS	58.30	75.03	-69.94	-69.94	-27.29	-55.72
IL-1β	117.79	3.26	23.55	23.55	-24.43	-26.09
IL-2	6.47	-4.51	-10.22	-10.22	-24.71	-11.86
IL-1m	1.56	-0.52	11.01	11.01	-40.86	114.12
IL-6	27.95	24.47	13.47	13.47	14.49	14.49
IL-8	87.95	24.47	5.34	5.34	-31.49	-20.45
IL-10	21.21	8.89	10.90	10.90	-11.95	46.11
IL-17	76.56	-37.27	24.56	24.56	-28.37	-117.59
TGF-α	52.29	2.33	19.67	19.67	-15.42	-30.86
アマトルグリン	19.96	-0.76	24.99	24.99	10.50	11.68
IL-8	40.10	48.14	24.11	24.11	-47.84	29.97
IL-11	15.51	-36.60	-5.89	-5.89	-71.41	21.38
IL-13	32.33	-0.19	-1.83	-1.83		-65.74

図 4

【 図 5 】

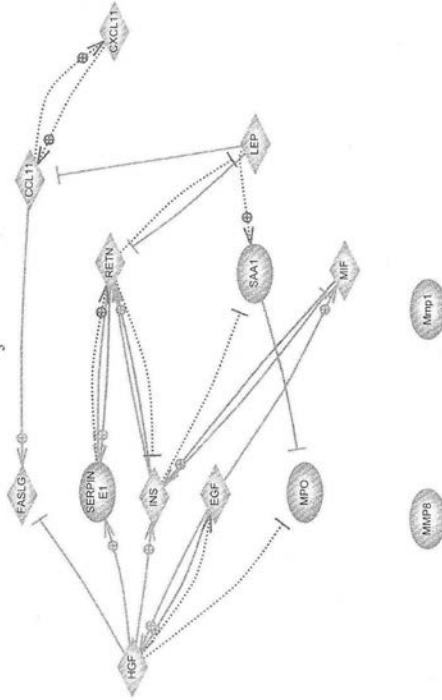
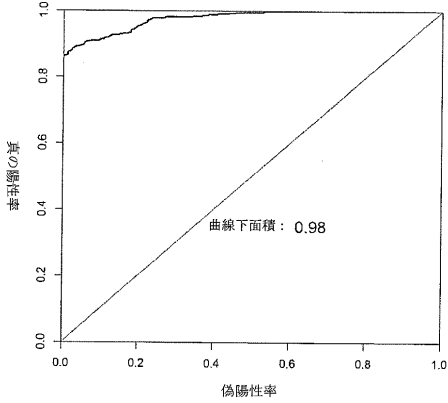


図 5

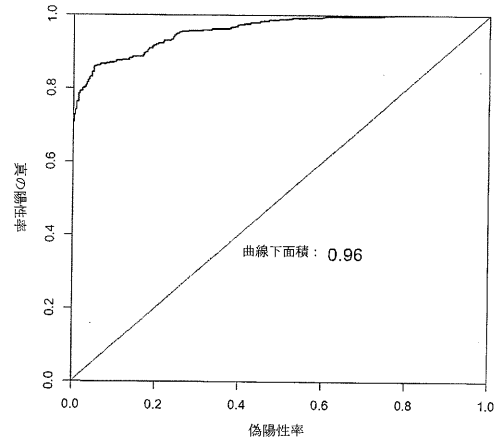
【 図 6 】

図 6 : アダプストの ROC 曲線



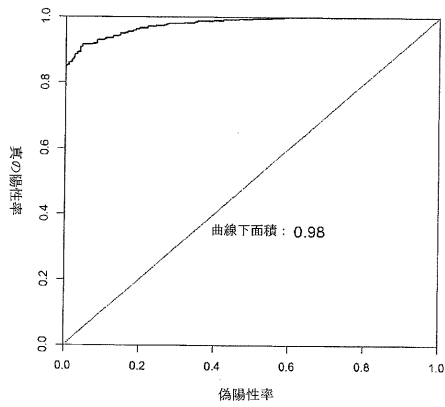
【 図 7 】

図 7 : SVM の ROC 曲線



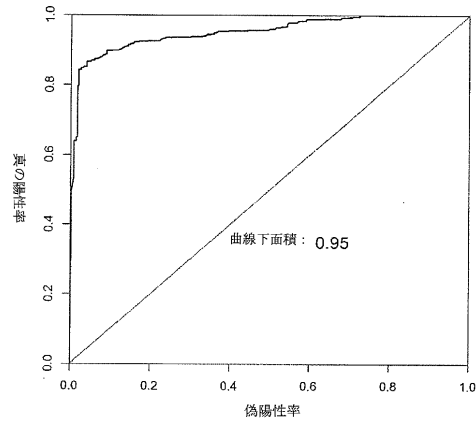
【 図 8 】

図 8 : 男性に限定したアダプストのROC曲線



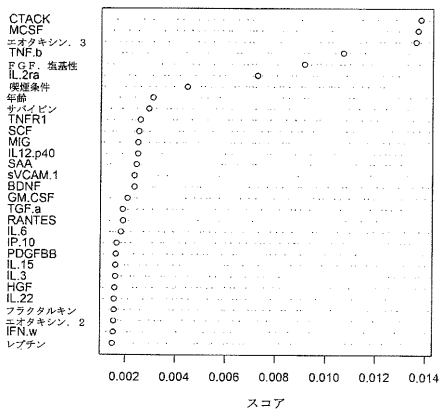
【 図 9 】

図 9 : 女性に限定したアダプストのROC曲線



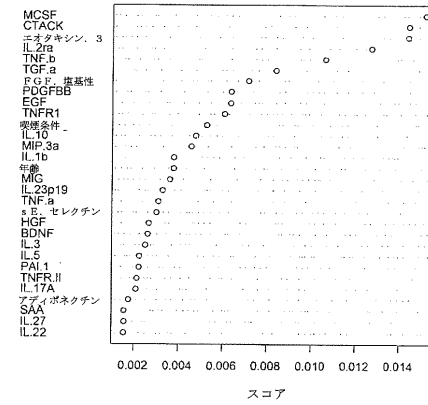
【 図 10 】

図 10 : アダプストモデルに基づく変数重要度プロット



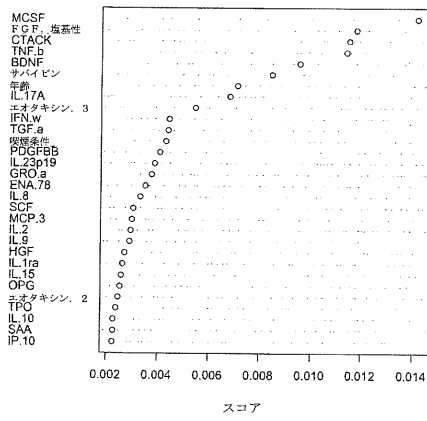
【 図 11 】

図 11 : 男性に限定したアダプストに基づく変数重要度プロット



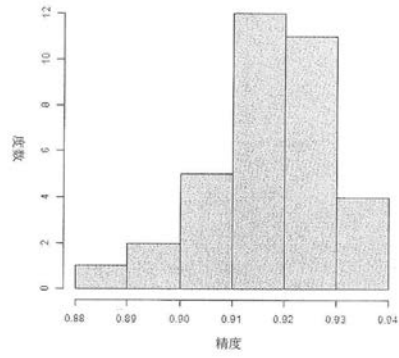
【 図 1 2 】

図 1 2 : 女性に限定したアダプストモデルに基づく変数重要度プロット



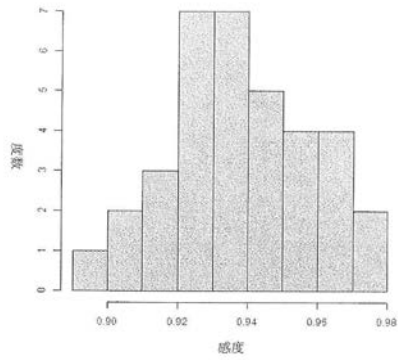
【 図 1 3 】

図 1 3 : アダプストモデルの精度分布



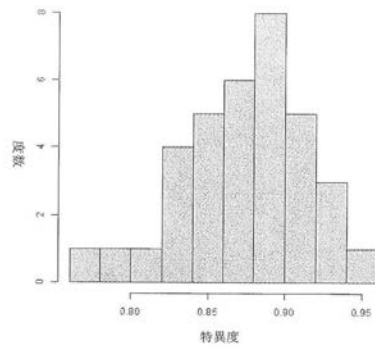
【 図 1 4 】

図 1 4 : アダプストモデルの感度分布



【 図 1 5 】

図 1 5 : アダプストモデルの特異度分布



---

フロントページの続き

(74)代理人 230113332

弁護士 山本 健策

(72)発明者 ロバート・ティー ストリーパー

アメリカ合衆国テキサス州78240, サン・アントニオ, アップル・グリーン・ロード 773  
8

(72)発明者 エルジュビェタ イズビッカ

アメリカ合衆国テキサス州78240, サン・アントニオ, アップル・グリーン・ロード 773  
8

(72)発明者 ジョエル マイカレック

アメリカ合衆国テキサス州78232, サン・アントニオ, ペブル・バレー 2634

(72)発明者 クリス ラウデン

アメリカ合衆国テキサス州78003, バンディアラ, セトラーズ・サークル 109

Fターム(参考) 2G045 AA25 AA26 CA25 CA26 DA36 FB03 JA01 JA03

4B029 AA07 BB11 BB15 BB20 FA15

【外国語明細書】

2018200322000001.pdf

专利名称(译)	使用分类系统鉴定和诊断肺病的方法及其试剂盒		
公开(公告)号	<a href="#">JP2018200322A</a>	公开(公告)日	2018-12-20
申请号	JP2018153283	申请日	2018-08-16
[标]申请(专利权)人(译)	癌症预防和治疗有限公司		
申请(专利权)人(译)	癌症私法化下降和固化有限公司		
[标]发明人	ロバート・ティー・ストリーパー エルジュビエタ・イズビッカ ジョエル・マイカレック クリス・ラウデン		
发明人	ロバート・ティー・ストリーパー エルジュビエタ・イズビッカ ジョエル・マイカレック クリス・ラウデン		
IPC分类号	G01N33/574 G01N33/68 G01N33/53 C12M1/34		
CPC分类号	G01N33/57423 G01N2800/60 G16H50/20 G16H50/70 G16Z99/00 G06F19/00		
FI分类号	G01N33/574.A G01N33/68 G01N33/53.D C12M1/34		
F-TERM分类号	2G045/AA25 2G045/AA26 2G045/CA25 2G045/CA26 2G045/DA36 2G045/FB03 2G045/JA01 2G045/JA03 4B029/AA07 4B029/BB11 4B029/BB15 4B029/BB20 4B029/FA15		
代理人(译)	夏木森下 飯田TakashiSatoshi 石川大介 山本健作		
优先权	61/480802 2011-04-29 US 61/619279 2012-04-02 US		
外部链接	<a href="#">Espacenet</a>		

摘要(译)

要解决的问题：提供使用分类系统识别和诊断肺部疾病的分类系统和方法。本发明提供了生物标志物和生物标志物的组合，其可用于诊断肺病，例如非小细胞肺癌或反应性气道疾病。这些生物标志物的测量已经进入分类系统，例如支持向量机或Adaboost，以帮助确定患有肺病的个体的可能性。还提供了包含用于检测这些生物标志物和生物标志物的组合的试剂的试剂盒，以及用于辅助诊断肺病的系统。背景技术

図 1 A

正常集団での蛍光強度レベル							
バイオマーカー	平均	標準偏差	SD/平均	バイオマーカー	平均	標準偏差	SD/平均
アポロキサニン	2250.55	1504.98	66.61	IL-17	37.60	13.52	35.95
シタビジン	784.37	90.22	72.55	IL-1a	93.36	185.32	194.33
PAI-1	336.10	374.63	110.95	IL-12 (p70)	23.30	9.28	39.81
シロニロキサタン	602.34	475.58	78.55	GM-CSF	23.64	4.15	17.29
sVCAM-1	1753.31	313.04	26.26	GM-CSF	26.60	12.37	45.94
sICAM-1	3293.75	1574.35	47.77	TNF-a	43.32	15.71	36.10
MPO	392.75	1803.40	253.64	MCP-1	46.39	29.47	63.51
CRP	8382.00	6702.94	79.86	IL-12 (p40)	24.95	7.36	29.11
SAA	2202.71	4017.36	182.38	MIP-2a	41.64	27.61	67.27
SAP	2304.71	570.75	25.03	MIP-1a	31.24	4.86	40.35
レゾナン	3204.67	2774.37	125.70	VEGF	40.80	11.45	28.07
GLP-1	36.21	302.46	182.37	IL-12 (p70)	7.66	36.69	479.00
アミロリン	142.47	2390.24	218.28	IL-1b	11.41	37.04	499.26
シーグレット	4748.58	1971.09	41.51	MMP-2	155.73	459.29	282.21
インスリン	278.12	424.50	152.65	MMP-1	103.52	179.41	175.31
Stas	38.51	148.38	252.33	MMP-3	6068.81	2442.77	26.63
sFas	15.13	16.89	117.39	マトリキニン	198.78	208.26	163.94
MIP	52.08	112.39	215.81	レゾナン	3767.80	5835.08	100.11
IL-1b	23.71	9.52	40.19	IP-1a	54.63	820.15	150.92
IL-2	12.26	44.02	358.99	MMP-9	1090.51	1599.16	146.04
IL-1ra	23.11	19.35	83.77	MMP-13	9.95	35.87	358.67
IL-4	114.39	72.33	63.23	PAI-1	38.94	98.35	252.58
IL-5	18.37	18.33	99.20	L-TAC	20.16	184.23	413.40
IL-6	40.99	56.56	137.96	MMP-12	12.41	27.69	223.06
IL-7	15.82	6.61	41.79	MGF	206.22	619.73	300.52
TEF-a	31.46	8.00	25.44	MMP-2	1026.07	577.82	57.33
マトリキニン	15.36	3.55	23.09	sFas	30.91	49.02	158.60
IL-10	13.06	3.63	27.82	CD40リガンド	155.50	454.04	270.12
IL-15	24.53	4.27	17.42	MMP-8	70.20	318.85	433.80