



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I423063 B

(45)公告日：中華民國 103 (2014) 年 01 月 11 日

(21)申請案號：098126864

(22)申請日：中華民國 98 (2009) 年 08 月 10 日

(51)Int. Cl. : G06F19/00 (2011.01)

(30)優先權：2008/08/08 美國 61/087,586

(71)申請人：納維吉尼克有限公司(美國) NAVIGENICS, INC. (US)
美國(72)發明人：摩爾 史帝芬 M MOORE, STEPHEN M. (US)；尼倫伯格 麥可 A NIERENBERG,
MICHAEL A. (US)；喬治 尚 E GEORGE, SEANE. (US)；勾莫 羅莉 A GOMER,
LAURIE A. (US)

(74)代理人：陳長文

(56)參考文獻：

TW 200626895A

TW 200722147A

TW 200731148A

WO 2006/065658A2

WO 2008/067551A2

審查人員：劉季涵

申請專利範圍項數：20 項 圖式數：8 共 0 頁

(54)名稱

用於個人化行為計劃之方法及系統

METHODS AND SYSTEMS FOR PERSONALIZED ACTION PLANS

(57)摘要

本發明提供基於個體之基因組分布，用於個人行為計劃之方法及系統。方法包括評估個體基因型與至少一種疾病或病狀之間的相關性及提供用於個人行為計劃之評級系統。本文亦揭示激勵及鼓勵個人改善其健康及康樂之獎勵。

The present disclosure provides methods and systems for personal action plans based on an individual's genomic profile. Methods include assessing the association between an individual's genotype and at least one disease or condition and providing rating systems for an individual's action plan. Incentives to motivate and encourage people to improve their health and well-being are also disclosed herein.

(無元件符號說明)

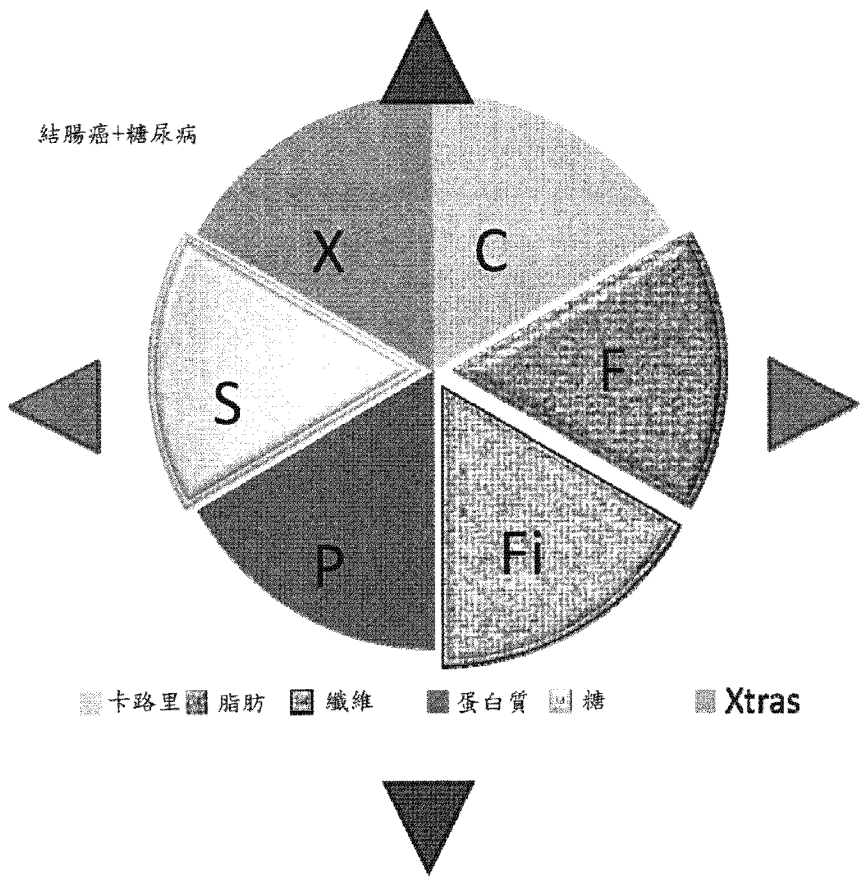


圖4A

六、發明說明：

本申請案主張2008年8月8日申請之美國臨時申請案第61/087,586號之優先權，該申請案係以引用的方式全部併入本文中。

【先前技術】

諸如單核苷酸多態現象(SNP)、突變、缺失、插入、重複、微衛星及其他之基因組遺傳變異與諸如疾病或病狀之各種表型相關。可鑑別個體遺傳變異且使其相關聯以判定個體對不同表型之傾向性，藉此產生個人化表型分布。

個體之表型分布提供對個體具有某一表型之風險或可能性之個人化評估，且個體可能會關注醫學以及生活方式選擇以便降低或增加特定病狀之風險。個體可受益於整合個體基因組分布之個人化行為計劃，該個人化行為計劃亦可進一步涵蓋非遺傳因素，諸如過去及現在之環境及生活方式因素。

因此，個人化行為計劃提供個體之定製方法或其健康護理管理，以在促進其健康及康樂方面作出知情及適當選擇。因此，對向個體及其健康護理管理者提供將個人基因組分布整合在便於遵守之行為計劃中以便進行適當醫學及生活方式選擇且可視情況具有激勵個體遵守其個人化行為計劃之獎勵的系統存在需要。本文所揭示之實施例滿足此等需要且亦提供相關優勢。

【發明內容】

本發明提供基於個體之基因組分布產生個人化行為計劃

之方法及系統。亦提供激勵個體過更健康生活之方法及系統，包括促進個體執行其行為計劃之方法。

本文描述個人化行為計劃中多種建議之評級系統，其中各建議給予一個等級。等級可由電腦產生或確定。各等級對應於給予個體之等級，其中給予個體之等級係基於個體之基因組分布。給予個體之等級可基於遺傳複合指數 (Genetic Composite Index, GCI) 或 GCI 加個體之得分。在一些實施例中，等級係由電腦基於 GCI 或 GCI 加由電腦確定之得分產生。電腦可隨後向個體或個體之健康護理管理者輸出等級。可藉由使用高密度 DNA 微陣列、諸如即時 PCR 之基於 PCR 之方法或其組合自個體擴增遺傳樣本來獲得基因組分布。

等級可為數字、顏色、字母或其組合，且等級可用於多種建議，諸如(但不限於)一或多個非醫藥建議。非醫藥建議可為鍛煉方案、鍛煉活動、飲食計劃或其組合。非醫藥建議亦可為諸如食物、維生素及其類似物類型之營養素。此外，等級可為以二元系統代表之評級系統之部分，例如等級可為兩個符號中之一者。

本文亦揭示向個體提供個人化行為計劃中之建議等級之方法，其包含獲得個體之基因組分布及確定個體之至少一個等級，其中該等級係基於基因組分布。在一些實施例中，向個體提供個人化行為計劃中之建議等級之方法包含產生 GCI 或 GCI 加個體得分及確定個體之至少一個等級，其中該等級係基於 GCI 或 GCI 附加得分。

本文亦提供激勵個體改善其健康之方法，其包含獲得該個體之基因組分布，產生個體之個人化行為計劃，使用於個體之至少一種獎勵與達成個人化行為計劃中之建議相關聯，及當實現該達成時授予個體該獎勵。在一些實施例中，激勵個體改善其健康之方法包含獲得該個體之基因組分布，產生個體之至少一個GCI或GCI附加得分，使用於個體之至少一種獎勵與至少一個GCI或GCI附加得分相關聯，及當達成改善時授予個體該獎勵。在一些實施例中，個人化行為計劃係由電腦產生或確定。舉例而言，電腦可產生個體之GCI或GCI附加得分，且隨後使用GCI或GCI附加得分來產生個人化行為計劃。可隨後由電腦向個體或個體之健康護理管理者輸出個人化行為計劃。

在一些實施例中，獎勵係由雇主、朋友或家庭成員提供。因此，在一些實施例中，個體為雇員且獎勵可為該個體之雇主對健康儲蓄帳戶之繳款、額外假期或該個體之醫療計劃的雇主補貼增加。

獎勵亦可為現金、醫藥產品、健康產品、健身俱樂部會員資格、醫學隨訪、醫療裝置、GCI或GCI附加得分更新、個人化行為計劃更新或線上社區之會員資格。在一些實施例中，獎勵為醫藥產品、健康產品、健身俱樂部會員資格、醫學隨訪、醫療裝置、GCI或GCI附加得分更新、個人化行為計劃更新或線上社區之會員資格的折扣、補貼或補償。在其他實施例中，獎勵為經由線上社區獲得之支持。

以引用的方式併入

本說明書中所述之所有公開案、專利及專利申請案係以引用的方式併入本文中，其引用之程度如該等公開案、專利或專利申請案各自特定且個別地指示以引用的方式併入本文中一般。

【實施方式】

本文所揭示實施例之新穎特徵在隨附申請專利範圍中詳細闡明。對本發明之特徵及優勢之更好理解將藉由參考以下實施方式及附圖而獲得，該實施方式闡明使用本文之揭示內容之原理的說明性實施例。

本文揭示基於個體之基因組分布產生個人化行為計劃之方法及系統。亦提供激勵個體過更健康生活之方法及系統，包括促進個體執行其行為計劃之方法。

基因組分布

個體基因組分布含有基於遺傳變異或標記之有關個體基因之資訊。遺傳變異可形成構成基因組分布之基因型。該等遺傳變異或標記包括(但不限於)單核苷酸多態現象(SNP)、單個及/或多個核苷酸重複、單個及/或多個核苷酸缺失、微衛星重複(具有典型的5-1,000個重複單元之小數目核苷酸重複)、二核苷酸重複、三核苷酸重複、序列重排(包括移位及複製)、複本數變化(特定基因座之損失或增加)及其類似情形。其他遺傳變異包括染色體重複及移位以及著絲點及端粒重複。

基因型亦可包括單體型(haplotype)及雙體型(diplo type)。

在一些實施例中，基因組分布可具有至少100,000、300,000、500,000或1,000,000個基因型。在一些實施例中，基因組分布可實質上為個體之完整基因組序列。在其他實施例中，基因組分布為個體之完整基因組序列之至少60%、80%或95%。基因組分布可為個體之完整基因組序列之約100%。含有目標之遺傳樣本包括(但不限於)未擴增基因組DNA或RNA樣本或經擴增DNA(或cDNA)。目標可為含有尤其相關遺傳標記之基因組DNA之特定區域。

為獲得基因組分布，可自個體之生物樣本分離個體之遺傳樣本。生物樣本包括可分離諸如RNA及/或DNA之遺傳物質的樣本。該等生物樣本可包括(但不限於)血液、毛髮、皮膚、唾液、精液、尿、糞便材料、汗水、口腔及各種身體組織。組織樣本可由個體直接收集，例如口腔樣本可藉由個體沿其面頰內部用拭子擦拭來獲得。諸如唾液、精液、尿、糞便材料或汗水之其他樣本亦可由個體本人提供。其他生物樣本可由諸如抽血者、護士或醫師之健康護理專家取得。舉例而言，血液樣本可由護士自個體抽取。可由健康護理專家進行組織生檢，且健康護理專家亦易於獲得商業套組以便有效地獲得樣本。可移除呈小圓柱狀之皮膚或可使用針來移除小型組織或流體樣本。

亦可向個體提供樣本收集套組。該套組可含有個體生物樣本之樣本收集容器。套組亦可提供個體直接收集其自身樣本之說明，諸如提供多少毛髮、尿、汗水或唾液。套組亦可含有對個體要求組織樣本由護理專家獲取之說明。套

組可包括可由第三方獲取樣本之地點，例如套組可提供給接著自個體收集樣本之健康護理機構。套組亦可提供待發送給樣本處理機構之樣本的返回包裝，遺傳物質係在該樣本處理機構自生物樣本分離。

可根據數種熟知生物化學及分子生物學方法中之任一種自生物樣本分離DNA或RNA之遺傳樣本，參見例如 *Sambrook* 等人，*Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York) (1989)。亦存在數種用於自生物樣本分離DNA或RNA之市售套組及試劑，諸如(但不限於)可自DNA Genotek、Gentra Systems、Qiagen、Ambion及其他供應商獲得者。口腔樣本套組易在市面上購得，諸如Epicentre Biotechnologies之MasterAmp™口腔擦拭DNA提取套組，自血液樣本提取DNA之套組同樣易於購得，諸如Sigma Aldrich之Extract-N-Amp™。可藉由以蛋白酶消化組織且加熱，離心樣本且使用苯酚-氯仿萃取不合意物質，使DNA留在水相中來自其他組織獲得DNA。該DNA可隨後藉由乙醇沈澱進一步分離。

舉例而言，可使用來自DNA Genotek之DNA自收集套組自唾液分離基因組DNA。個體可收集唾液試樣以便使用該套組進行臨床處理，且樣本可便利地在室溫下儲存及運送。將樣本遞送至用於處理之適當實驗室後，藉由加熱變性，且通常使用由收集套組供應商提供之試劑在50°C下蛋白酶消化樣本至少1小時，來分離DNA。接著離心樣本且將上清液用乙醇沈澱。將DNA離心塊懸浮於適於後續分析

之緩衝液中。

可使用RNA作為遺傳樣本，例如可自mRNA鑑別所表現之遺傳變異。mRNA包括(但不限於)前mRNA轉錄產物、轉錄產物加工中間物、準備供轉譯之成熟mRNA及基因轉錄產物，或來源於mRNA轉錄產物之核酸。轉錄產物加工可包括拼接、編輯及降解。如本文中所用，來源於mRNA轉錄產物之核酸係指mRNA轉錄產物或其子序列最終充當合成模板之核酸。因此，自mRNA逆轉錄之cDNA、自cDNA擴增之DNA、自擴增DNA轉錄之RNA等全部來源於mRNA轉錄產物。RNA可使用在此項技術中已知之方法自數種身體組織中的任一種分離，諸如使用可自PreAnalytiX獲得之PAXgene™血液RNA系統自未分級分離全血分離RNA。通常，使用mRNA來逆轉錄cDNA，該cDNA接著用於或經擴增用於基因變異分析。

基因組分布分析之前，可自由RNA逆轉錄之DNA或cDNA擴增遺傳樣本。可藉由許多方法(其中許多採用PCR)擴增DNA。參見例如 *PCR Technology: Principles and Applications for DNA Amplification* (H. A. Erlich編，Freeman Press, NY, N.Y., 1992)；*PCR Protocols: A Guide to Methods and Applications* (Innis等人編，Academic Press, San Diego, Calif., 1990)；Mattila等人，*Nucleic Acids Res.* 19, 4967 (1991)；Eckert等人，*PCR Methods and Applications* 1, 17 (1991)；*PCR* (McPherson等人編，IRL Press, Oxford)及美國專利第4,683,202號、第4,683,195號、第4,800,159號、第4,965,188號及第

5,333,675號，該等文獻各自出於所有目的以引用的方式全部併入本文中。

其他適當擴增方法包括連接酶鏈反應(LCR)(例如 Wu 及 Wallace, *Genomics* 4, 560 (1989) ; Landegren 等人, *Science* 241, 1077 (1988) 及 Barringer 等人, *Gene* 89:117 (1990))、轉錄擴增(Kwoh 等人, *Proc. Natl. Acad. Sci. USA* 86:1173-1177 (1989) 及 WO88/10315)、自主序列複製(Guatelli 等人, *Proc. Nat. Acad. Sci. USA*, 87:1874-1878 (1990) 及 WO90/06995)、選擇性擴增目標聚核苷酸序列(美國專利第 6,410,276 號)、共同序列引發之聚合酶鏈反應(CP-PCR)(美國專利第 4,437,975 號)、任意引發之聚合酶鏈反應(AP-PCR)(美國專利第 5,413,909 號、第 5,861,245 號)、基於核酸序列之擴增(NASBA)、滾環擴增(RCA)、多重置換擴增(MDA)(美國專利第 6,124,120 號及第 6,323,009 號)及環至環擴增(C2CA)(Dahl 等人 *Proc. Natl. Acad. Sci* 101:4548-4553 (2004))。(參見美國專利第 5,409,818 號、第 5,554,517 號及第 6,063,603 號，其各自以引用的方式併入本文中)。可使用之其他擴增法描述於美國專利第 5,242,794 號、第 5,494,810 號、第 5,409,818 號、第 4,988,617 號、第 6,063,603 號及第 5,554,517 號及美國第 09/854,317 號中，其各自係以引用的方式併入本文中。

可使用數種方法中之任一種進行基因組分布之產生。在此項技術中已知數種鑑別遺傳變異之方法，且其包括(但不限於)藉由數種方法中之任一種進行 DNA 定序、基於 PCR

之方法、片段長度多態現象檢定(限制片段長度多態現象(RFLP)、裂解片段長度多態現象(CFLP))、使用等位基因特異性寡核苷酸作為模板之雜交法(例如本文另外描述之TaqMan檢定及微陣列)、使用引子延伸反應之方法、質譜分析(諸如MALDI-TOF/MS法)及其類似方法，諸如Kwok, *Pharmacogenomics* 1:95-100 (2000)中所述。其他方法包括侵染法，諸如單路及雙路侵染檢定(例如可自Third Wave Technologies, Madison, WI獲得且描述於Olivier等人, *Nucl. Acids Res.* 30:e53 (2002)中)。

舉例而言，可使用高密度DNA陣列產生基因組分布。該等陣列可購自Affymetrix及Illumina(參見Affymetrix GeneChip® 500K檢定手冊，Affymetrix, Santa Clara, CA(以引用的方式併入)；Sentrix® humanHap650Y基因分型晶片，Illumina, San Diego, CA)。可使用高密度陣列產生包含為SNP之遺傳變異的基因組分布。舉例而言，可藉由使用Affymetrix基因組廣泛性人類SNP陣列6.0對多於900,000種SNP進行基因分型產生SNP分布。或者，可藉由使用Affymetrix基因晶片人類繪圖500K陣列組來經由全基因組取樣分析測定多於500,000種SNP。在此等檢定中，使用限制酶消化、銜接子連接之人類基因組DNA經由單引子擴增反應擴增人類基因組之子集。通常，接著斷裂所擴增之DNA且測定樣本之品質，隨後變性及標記樣本以便與經塗布石英表面上特定位置之具有DNA探針的微陣列雜交。監測作為所擴增DNA序列之函數的與各探針雜交之標記之

量，藉此產生序列資訊及所得SNP基因分型。

在此項技術中熟知高密度陣列之使用，且若以商業方式獲得，則根據製造商之說明進行。舉例而言，使用Affymetrix基因晶片可包括以NspI或StyI限制性核酸內切酶消化經分離基因組DNA。接著，使經消化DNA與NspI或StyI銜接子寡核苷酸連接，該等銜接子分別與NspI或StyI限制性酶切之DNA黏接。接著藉由PCR擴增連接後含有銜接子之DNA以產生如由凝膠電泳所證實約200與1100鹼基對之間的經擴增DNA片段。純化符合擴增標準之PCR產物且定量以供斷裂。以脫氧核糖核酸酶I斷裂PCR產物以實現最佳DNA晶片雜交。斷裂後，如由凝膠電泳所證實，DNA片段應小於250個鹼基對，且平均為約180個鹼基對。接著使用末端脫氧核苷酸轉移酶以生物素化合物標記符合斷裂標準之樣本。接著使所標記片段變性且隨後於基因晶片250K陣列中雜交。雜交後，將陣列染色，隨後在由抗生蛋白鏈菌素藻紅素(SAPE)染色、接著以生物素化抗抗生蛋白鏈菌素抗體(山羊)進行抗體擴增之步驟及最終以抗生蛋白鏈菌素藻紅素(SAPE)染色組成之三步法中掃描。標記後，將陣列用陣列保持緩衝液覆蓋且隨後以例如掃描儀(諸如Affymetrix基因晶片掃描儀3000)掃描。

掃描高密度陣列後，可根據製造商之指南進行資料分析。舉例而言，在使用Affymetrix基因晶片情況下，藉由使用基因晶片操作軟體(GCOS)或藉由使用Affymetrix基因晶片指令控制台(Affymetrix GeneChip Command

Console™)進行原始資料之獲取。獲取原始資料後，接著使用基因晶片基因分型分析軟體(GTYPE)進行分析。可排除GTYPE檢出率小於某一百分比之樣本。舉例而言，可排除小於約70%、75%、80%、85%、90%或95%之檢出率。接著，以BRLMM及/或SNIper算法分析檢驗樣本。排除BRLMM檢出率小於95%或SNIper檢出率小於98%之樣本。最終，進行相關性分析，且排除SNIper品質指數小於0.45及/或Hardy-Weinberg p值小於0.00001之樣本。

作為DNA微陣列分析之替代或除DNA微陣列分析之外，亦可藉由其他基於雜交之方法，諸如使用TaqMan方法及其變化形式偵測諸如SNP及突變之遺傳變異。本文揭示之方法可使用TaqMan PCR、迭代TaqMan及即時PCR(RT-PCR)之其他變化形式，諸如Livak等人，*Nature Genet.*, **9**, 341-32 (1995)及Ranade等人*Genome Res.*, **11**, 1262-1268 (2001)中所描述者。在一些實施例中，將諸如SNP之特定遺傳變異之探針標記以形成TaqMan探針。探針長度通常約為至少12、15、18或20個鹼基對。其長度可在約10與70、15與60、20與60或18與22個鹼基對之間。以諸如螢光團之報告體標記標記探針之5'端且以該標記之淬滅劑標記3'端。報告體標記可為當密切接近(諸如探針之長度)淬滅劑時螢光受抑制或淬滅之任何螢光分子。舉例而言，報告體標記可為螢光團，諸如6-羧基螢光素(6-carboxyfluorescein)(FAM)、四氯螢光素(tetrachlorfluorescein)(TET)或其衍生物，及淬滅劑四甲基若丹明(tetramethylrhodamine)

(TAMRA)、二氫環吡咯并吡啶三肽(MGB)或其衍生物。

由於報告體螢光團與淬滅劑密切接近，由探針長度分開，因此使螢光淬滅。當探針與樣本中諸如包含SNP之序列之目標序列時黏接，具有5'至3'外切核酸酶活性之DNA聚合酶(諸如Taq聚合酶)可延伸引子且外切核酸酶活性裂解探針，使報告體自淬滅劑分離，且因此報告體可發螢光。可在諸如RT-PCR中重複該過程。TaqMan探針通常與位於經設計以擴增序列之兩個引子之間的目標序列互補。因此，由於各探針可與新近產生之PCR產物雜交，因此PCR產物之累積可與所釋放螢光團之累積相關聯。可量測所釋放螢光團且可測定所存在目標序列之量。用於高產量基因分型之RT-PCR方法，諸如。

亦可藉由DNA定序鑑別遺傳變異。可使用DNA定序來定序個體基因組序列之實質部分或全部。通常，常用DNA定序係基於聚丙烯醯胺凝膠分級分離來分辨鏈終止片段之群體(Sanger等人，*Proc. Natl. Acad. Sci. USA* 74:5463-5467 (1977))。已開發且持續開發替代性方法來增加DNA定序之速度及簡易性。舉例而言，高產量及單個分子定序平台可自454 Life Sciences(Branford, CT)(Margulies等人，*Nature* 437:376-380 (2005))；Solexa(Hayward, CA)；Helicos BioSciences Corporation(Cambridge, MA)(2005年6月23日申請之美國申請案第11/167046號)及Li-Cor Biosciences(Lincoln, NE)(2005年4月29日申請之美國申請案第11/118031號)購得或正由該等公司開發。

產生個體基因組分布後，數位儲存該分布。該分布可以安全方式數位儲存。諸如在電腦可讀媒體上以電腦可讀格式編碼基因組分布以作為資料組之部分儲存且可以資料庫形式儲存，其中基因組分布可「存入」且可在以後再次存取。資料組包含複數個資料點，其中各資料點與一個個體有關。各資料點可具有複數個資料要素。一個資料要素為用以鑑別個體基因組分布之唯一標識符。該唯一標識符可為條碼。另一資料要素為基因型資訊，諸如個體基因組之SNP或核苷酸序列。對應於基因型資訊之資料要素亦可包括於資料點中。舉例而言，若基因型資訊包括由微陣列分析鑑別之SNP，則其他資料要素可包括微陣列SNP鑑別編號。或者，若基因型資訊係藉由其他方式鑑別，諸如藉由RT-PCR方法(諸如TaqMan檢定)，則資料要素可包括螢光含量、引子資訊及探針序列。其他資料要素可包括(但不限於)SNP rs編號、多態核苷酸、基因型資訊之染色體位置、資料之品質規格、原始資料檔案、資料影像及所提取強度得分。

諸如身體資料、醫學資料、種族、家系(ancestry)、地理(geography)、性別、年齡、家族史、已知表型、人口統計資料、暴露資料、生活方式資料、行為資料及其他已知表型之個體特定因素亦可作為資料要素併入。舉例而言，因素可包括(但不限於)個體出生地、父母及/或祖父母、親戚家系、住址、籍貫(ancestors' location of residence)、環境條件、已知健康狀況、已知藥物相互作用、家庭健康狀

況、生活條件、飲食、鍛煉習慣、婚姻狀況及身體量測，諸如體重、身高、膽固醇含量、心率、血壓、葡萄糖含量及其他在此項技術中已知之量測值。個體親戚或長輩(諸如父母及祖父母)之上述因素亦可作為資料要素併入且用以判定個體對某一表型或病狀的風險。

可自個體之調查表或健康護理管理者獲得特定因素。隨後可在需要時自「所存入」分布獲取資訊且使用。舉例而言，在個體基因型相關性之初始評估中，將分析個體整體資訊(通常為整個基因組或自整個基因組獲取之SNP或其他基因組序列)之基因型相關性。在後續分析中，可在需要或適當時自所儲存或存入基因組分布獲取整體資訊或其部分。

相關性及表型分布

使用基因組分布來產生表型分布。基因組分布通常數位儲存且易於在任一時間點獲取以產生表型分布。藉由應用使基因型與表型相關聯或相關之規則產生表型分布。規則係基於證明基因型與表型之間的相關性之科學研究制定。相關性可由具有一位或一位以上專家之委員會管理或驗證。藉由對個體之基因組分布應用該等規則，可判定個體基因型與表型之間的相關性。個體之表型分布將具有此判定。該判定可為個體基因型與指定表型之間的正相關，使得個體具有指定表型或將發展該表型。或者，可判定個體不具有或將不發展指定表型。在其他實施例中，判定可為個體具有或將發展表型之風險因子、估算值或機率。

該等判定可基於許多規則進行，例如可將複數個規則應用於基因組分布以判定個體基因型與特定表型之相關性。該等判定亦可併有個體特定因素，諸如種族、性別、生活方式(例如飲食及鍛煉習慣)、年齡、環境(例如住址)、家庭病史、個人病史及其他已知表型。可藉由修改現有規則以包涵此等因素來進行特定因素之併入。或者，可藉由該等因素產生獨立規則，且在已應用現有規則後將其應用於個體表型判定。

表型可包括任何可量測之性狀或特徵，諸如對某一疾病之易感性或對藥物治療之反應。可包括之其他表型為身體及精神性狀，諸如身高、體重、發色、眼睛顏色、曬傷易感性、體型、記憶力、智力、樂觀程度及總體性情。表型亦可包括與其他個體或生物體之遺傳比較。舉例而言，個體可能會關注其基因組分布與名人之基因組分布之間的相似性。其亦可具有與諸如細菌、植物或其他動物之其他生物體相比較的基因組分布。總之，關於個體判定之相關表型之集合構成個體之表型分布。

可自科學文獻獲得遺傳變異與表型之間的相關性。遺傳變異之相關性可藉由分析已測試個體之群體中是否存在一或多個所關注表型性狀及其基因型分布來判定。回顧分布中各遺傳變異或多態現象之等位基因以判定特定等位基因之存在或不存在是否與所關注性狀相關。可藉由標準統計學方法進行相關性測定且記錄遺傳變異與表型特徵之間的統計學上顯著的相關性。舉例而言，可判定多態現象A處

等位基因 A1 之存在與心臟病相關。作為另一實例，可發現多態現象 A 處等位基因 A1 與多態現象 B 處等位基因 B1 之組合存在與增加之癌症風險相關。分析之結果可公開於同級回顧文獻中，由其他研究組驗證及/或由諸如遺傳學家、統計學家、流行病學家及醫師之專家委員會分析，且亦可管理。舉例而言，US 公開案第 20080131887 號及 PCT 公開案第 WO/2008/067551 號揭示之相關性可用於本文所述之實施例中。

或者，相關性可由所儲存之基因組分布產生。舉例而言，具有儲存基因組分布之個體亦可具有亦儲存之已知表型資訊。分析儲存之基因組分布及已知表型可產生基因型相關性。舉例而言，250 位具有儲存基因組分布之個體亦具有其先前已診斷患有糖尿病之儲存資訊。進行基因組分布分析且與無糖尿病個體之對照組相比較。接著判定先前診斷患有糖尿病之個體具有與對照組相比較高的具有特定遺傳變異之比率，且可作出特定遺傳變異與糖尿病之間的基因型相關性。

基於該經驗證的遺傳變異與特定表型之相關性制定規則。可基於如美國公開案第 20080131887 號及 PCT 公開案第 WO/2008/067551 號所揭示相關之基因型與表型產生規則，且一些規則可併有諸如性別或種族之其他因素以產生效應估算值。由規則產生之其他量度可為估算出之相對風險增加。效應估算值及估算出之相對風險增加可來自公開文獻或自公開文獻計算。或者，規則可基於由儲存之基因

組分布及先前已知表型產生之相關性。

遺傳變異可包括 SNP。當 SNP 存在於單一位點時，通常可預測在一個位點攜帶特定 SNP 等位基因之個體在其他位點攜帶特定 SNP 等位基因。SNP 與使個體易患疾病或病狀之等位基因之相關性經由連鎖不平衡發生，其中兩個或兩個以上基因座處等位基因之非隨機相關在群體中以比自經由重組而隨機形成所預期較高或較低的頻率發生。

諸如核苷酸重複或插入之其他遺傳標記或變異亦可與已展示與特定表型相關之遺傳標記連鎖不平衡。舉例而言，核苷酸插入與表型相關連且 SNP 與核苷酸插入連鎖不平衡。基於 SNP 與表型之間的相關性制定規則。亦可基於核苷酸插入與表型之間的相關性制定規則。由於一個 SNP 之存在可提供某一風險因子，另一者可提供另一風險因子且當組合時可能會增加風險，因此任一規則或兩種規則均可應用於基因組分布。

疾病易感性等位基因經由連鎖不平衡與 SNP 之特定等位基因或 SNP 之特定等位基因的組合共分離。SNP 等位基因沿染色體之特定組合稱為單體型，且其組合存在之 DNA 區域可稱為單體型區塊。儘管單體型區塊可由一個 SNP 組成，但單體型區塊通常代表在個體中顯示低單體型多樣性且通常具有低重組頻率之一系列連續 2 個或 2 個以上 SNP。可藉由鑑別處於單體型區塊中之一個或一個以上 SNP 進行單體型之鑑別。因此，通常可使用 SNP 分布來鑑別單體型區塊，而不必要求鑑別指定單體型區塊中之所有 SNP。

日益知曉SNP單體型模式與疾病、病狀或身體狀態之間的基因型相關性。對於指定疾病，將已知患有疾病之一組個人的單體型模式與無該疾病之一組個人相比較。藉由分析許多個體，可測定群體中多態現象之頻率，且接著可將此等頻率或基因型與諸如疾病或病狀之特定表型相關聯。已知SNP-疾病相關性之實例包括年齡相關之黃斑退化中補體因子H之多態現象(Klein等人，*Science*: 308:385-389, (2005))及與肥胖症相關之INSIG2基因附近的變異(Herbert等人，*Science*: 312:279-283 (2006))。其他已知SNP相關性包括包含諸如CDKN2A及B之9p21區域之多態現象，諸如與心肌梗塞相關聯的rs10757274、rs2383206、rs13333040、rs2383207及rs10116277(Helgadottir等人，*Science* 316:1491-1493 (2007)；McPherson等人，*Science* 316:1488-1491 (2007))。

SNP可為功能或非功能性的。舉例而言，功能性SNP對細胞功能具有影響，藉此產生表型，而非功能性SNP在功能上無表現，但可與功能性SNP連鎖不均衡。SNP亦可為同義或不同義的。同義SNP為不同形式產生相同多肽序列之SNP且為非功能性SNP。若SNP產生不同多肽，則SNP不同義且可或可不具有功能性。亦可使用用以鑑別為2個或2個以上單體型之雙體型中之單體型的SNP或其他遺傳標記使與雙體型相關之表型相關聯。關於個體之單體型、雙體型及SNP分布之資訊可在個體之基因組分布中。

通常，對於待基於與另一與表型相關之遺傳標記連鎖不均衡的遺傳標記產生之規則而言，遺傳標記之 r^2 或 D' 得分

(此項技術中常用於判定連鎖不平衡之得分)大於0.5。該得分可大於約0.5、0.6、0.7、0.8、0.90、0.95或0.99。因此，用以使表型與個體基因組分布相關聯之遺傳標記可與與表型相關聯之功能性或公開SNP相同或不同。在一些實施例中，測試SNP可能尚未鑑別，但使用公開SNP資訊，可基於諸如TaqMan之另一檢定鑑別等位基因差異或SNP。舉例而言，公開之SNP為rs1061170，但測試SNP未經鑑別。可使用公開之SNP藉由LD分析鑑別測試SNP。或者，可不使用測試SNP，且改為使用TaqMan或其他相當檢定來評估具有測試SNP之個體基因組。

測試SNP可為「直接」或「標籤」SNP。直接SNP為與公開或功能性SNP相同之測試SNP。舉例而言，在歐洲及亞洲使用SNP rs1073640，直接SNP可用於FGFR2與乳癌之相關性，其中較小等位基因為A且另一等位基因為G(*Easton*等人，*Nature* 447:1087-1093 (2007))。亦在歐洲及亞洲，可為用於FGFR2與乳癌之相關性之直接SNP之另一公開或功能性SNP為rs1219648(*Hunter*等人，*Nat. Genet.* 39:870-874 (2007))。標籤SNP為測試SNP不同於功能或公開SNP。標記SNP亦可用於其他遺傳變異，諸如用於以下之SNP：CAMTA1(rs4908449)、9p21(rs10757274、rs2383206、rs13333040、rs2383207、rs10116277)、COL1A1(rs1800012)、FVL(rs6025)、HLA-DQA1(rs4988889、rs2588331)、eNOS(rs1799983)、MTHFR(rs1801133)及APC(rs28933380)。

SNP之資料庫可公開獲自例如國際人類基因組單體型圖

計劃(International HapMap Project)(參見 www.hapmap.org，國際人類基因組單體型圖協定(The International HapMap Consortium), Nature 426:789-796 (2003)及國際人類基因組單體型圖協定(The International HapMap Consortium), Nature 437:1299-1320 (2005))、人類基因突變資料庫(Human Gene Mutation Database, HGMD)公共資料庫(參見 www.hgmd.org)及單核苷酸多態現象資料庫(dbSNP)(參見 www.ncbi.nlm.nih.gov/SNP/)。此等資料庫提供SNP單體型或能夠測定SNP單體型模式。因此，此等SNP資料庫能夠檢驗諸如癌症、發炎性疾病、心血管疾病、神經退化性疾病及傳染性疾病之多種疾病及病狀之根本遺傳風險因子。該等疾病或病狀可治療，其中目前存在治療及療法。治療可包括預防性治療以及改善症狀及病狀之治療，包括生活方式改變。

亦可檢驗諸如身體性狀、生理學性狀、精神性狀、情緒性狀、種族、家系及年齡之許多其他表型。身體性狀可包括身高、發色、眼睛顏色、身體或諸如毅力、耐久力及敏捷性之性狀。精神性狀可包括智力、記憶力或學習力。種族及家系可包括鑑別祖先或種族，或個體祖先之起源地。年齡可為判定個體之真實年齡，或個體遺傳學使其相對於普通人群定位之年齡。舉例而言，個體真實年齡38歲，然而其遺傳學可判定其記憶能力或身體健康可為平均28歲。另一年齡性狀可為個體之預計壽命(projected longevity)。

其他表型亦可包括非醫學情形，諸如「娛樂」表型。此

等表型可包括與諸如外國顯要人物、政治家、名人、發明家、運動員、音樂家、藝術家、商人及諸如罪犯之聲名狼藉個體之著名個體相比較。其他「娛樂」表型可包括與諸如細菌、昆蟲、植物或非人類動物之其他生物體相比較。舉例而言，個體可能想知道其基因組分布與其寵物狗或前總統相比如何。

將該等規則應用於儲存基因組分布以產生表型分布。舉例而言，來自公開來源或儲存基因組分布之相關性資料可形成規則或測試之基礎，以應用於個體基因組分布。該等規則可包涵關於測試SNP及等位基因及效應估算值(諸如OR或優勢率(95%置信區間)或平均值)之資訊。效應估算值可為基因型風險，諸如同型合子風險(homoz或RR)，異型合子風險(heteroz或RN)及非風險同型合子(homoz或NN)。效應估算值亦可為攜帶者風險(carrier risk)，其為RR或RN相對於NN。效應估算值可基於等位基因，諸如等位基因風險，一實例為R相對於N。亦可存在2、3、4或4個以上基因座基因型效應估算值(例如對於兩基因座效應估算值，9種可能的基因型組合而言，為RRRR、RRNN等)。

病狀之估算風險可基於如美國公開案第20080131887號及PCT公開案第WO/2008/067551號所列之SNP。在一些實施例中，病狀風險可基於至少一個SNP。舉例而言，個體阿爾茨海默氏症(Alzheimers, AD)、結腸直腸癌(CRC)、骨關節炎(OA)或剝脫性青光眼(XFG)之風險評估可基於1種SNP(例如，對於AD而言為rs4420638、對於CRC而言為

rs6983267、對於 OA 而言為 rs4911178 且對於 XFG 而言為 rs2165241)。對於諸如肥胖症(BMIOB)、格雷氏症(Graves' disease, GD)或血色素沉著(HEM)之其他病狀而言，個體之估算風險可基於至少 1 種或 2 種 SNP(例如，對於 BMIOB 而言為 rs9939609 及 / 或 rs9291171；對於 GD 而言為 DRB1*0301 DQA1*0501 及 / 或 rs3087243；對於 HEM 而言為 rs1800562 及 / 或 rs129128)。對於諸如(但不限於)心肌梗塞(MI)、多發性硬化症(MS)或牛皮癬(PS)之病狀而言，可使用 1、2 或 3 種 SNP 來評估個體之病狀風險(例如，對於 MI 而言為 rs1866389、rs1333049 及 / 或 rs6922269；對於 MS 而言為 rs6897932、rs12722489 及 / 或 DRB1*1501；對於 PS 而言為 rs6859018、rs11209026 及 / 或 HLAC*0602)。對於估算個體之多動腿症候群(RLS)或乳糜瀉(CeID)風險而言，可使用 1、2、3 或 4 種 SNP(例如，對於 RLS 而言為 rs6904723、rs2300478、rs1026732 及 / 或 rs9296249；對於 CeID 而言為 rs6840978、rs11571315、rs2187668 及 / 或 DQA1*0301 DQB1*0302)。對於前列腺癌(PC)或狼瘡(SLE)而言，可使用 1、2、3、4 或 5 種 SNP 來估算個體之 PC 或 SLE 風險(例如，對於 PC 而言為 rs4242384、rs6983267、rs16901979、rs17765344 及 / 或 rs4430796；對於 SLE 而言為 rs12531711、rs10954213、rs2004640、DRB1*0301 及 / 或 DRB1*1501)。對於估算個體之黃斑退化(AMD)或類風濕性關節炎(RA)之終生風險而言，可使用 1、2、3、4、5 或 6 種 SNP(例如，對於 AMD 而言為 rs10737680、rs10490924、rs541862、

rs2230199、rs1061170及/或rs9332739；對於RA而言為rs6679677、rs11203367、rs6457617、DRB*0101、DRB1*0401及/或DRB1*0404)。對於估算個體之乳癌(BC)之終生風險而言，可使用1、2、3、4、5、6或7種SNP(例如，rs3803662、rs2981582、rs4700485、rs3817198、rs17468277、rs6721996及/或rs3803662)。對於估算個體之克隆氏病(Crohn's disease, CD)或第2型糖尿病(T2D)之終生風險而言，可使用1、2、3、4、5、6、7、8、9、10或11種SNP(例如，對於CD而言為rs2066845、rs5743293、rs10883365、rs17234657、rs10210302、rs9858542、rs11805303、rs1000113、rs17221417、rs2542151及/或rs10761659；對於T2D而言為rs13266634、rs4506565、rs10012946、rs7756992、rs10811661、rs12288738、rs8050136、rs1111875、rs4402960、rs5215及/或rs1801282)。在一些實施例中，用作判定風險之基礎的SNP可與如上所述之SNP或諸如美國公開案第20080131887號及PCT公開案第WO/2008/067551號之其他SNP連鎖不平衡。

個體之表型分布可包含許多表型。詳言之，藉由本文所揭示之方法評估患者之疾病或其他病狀之風險(諸如包括新陳代謝、功效及/或安全性之可能藥物反應)允許對多種無關疾病及病狀之易感性進行預測或診斷分析，而不管是帶症狀、症狀前或為無症狀個體，包括具有一或多個疾病/病狀易感性等位基因之攜帶者。因此，此等方法提供個體對疾病或病狀之易感性之一般性評估，而絲毫未預先構想

特意測試特定疾病或病狀。舉例而言，本文揭示之方法允許基於個體基因組分布評估個體對美國公開案第20080131887號及PCT公開案第WO/2008/067551號中所列之數種病狀中之任一種之易感性。此外，該等方法允許評估個體對一或一種以上表型或病狀之估算終生風險或相對風險。

評估提供2或2種以上此等病狀之資訊，且可包括至少3、4、5、10、15、18、20、25、30、35、40、45、50、100種或甚至更多此等病狀。表型之單一規則可應用於單基因表型。一種以上規則亦可應用於單一表型，諸如單一基因內之多種遺傳變異影響具有該表型之機率的多基因表型或單基因表型。

初始篩選個體患者之基因組分布後，當諸如SNP之其他遺傳變異已知時，經由與該等其他遺傳變異相比較可產生(或可獲得)個體基因型相關性之更新。舉例而言，可進行定期更新，例如藉由瀏覽新基因型相關性之科學文獻之一或多個普通熟習遺傳學領域者每日、每週或每月進行更新。可隨後藉由此領域之一或多個專家之委員會進一步驗證新基因型相關性。

新規則可包涵無現有規則之基因型或表型。舉例而言，發現與任何表型均無關之基因型與新穎或現有表型有關。新規則亦可用於先前無相關基因型之表型之間的相關性。亦可判定具有現有規則之基因型及表型的新規則。舉例而言，存在基於基因型A與表型A之間的相關性之規則。新

研究揭示基因型B與表型A相關，且制定基於此相關性之新規則。另一實例為發現表型B與基因型A相關聯，且因此可制定新規則。

亦可針對基於已知相關性，但公開科學文獻中最初未鑑別之發現制定規則。舉例而言，據報告基因型C與表型C相關連。另一公開案報告基因型D與表型D相關連。表型C與D為相關症狀，例如表型C可為呼吸急促且表型D為肺活量小。可經由統計方式使用具有基因型C及D及表型C及D之個體之現有儲存基因組分布或藉由進一步研究發現及驗證基因型C與表型D或基因型D與表型C之間的相關性。可隨後基於新發現且驗證之相關性產生新規則。在另一實施例中，可研究具有特定或相關表型之許多個體之儲存基因組分布以確定個體所共有之基因型，且可判定相關性。可基於此相關性產生新規則。

亦可制定規則以修訂現有規則。舉例而言，可藉由諸如種族、家系、地理、性別、年齡、家族史或個體之任何其他已知表型之已知個別特徵部分判定基因型與表型之間的相關性。可制定基於此等已知個別特徵之規則且將其併入現有規則中以提供修訂規則。待應用修訂規則之選擇應取決於個體之特定個別因素。舉例而言，規則可基於當個體具有基因型E時，該個體具有表型E之機率為35%。然而，若個體為特定種族，則機率為5%。可基於此結果產生新規則且應用於具有彼特定種族之個體。或者，可應用判定值為35%之現有規則，且隨後應用基於彼表型之種族的另

一規則。可由科學文獻確定或基於對儲存基因組分布之研究確定基於已知個別特徵之規則。當發展新規則時，可添加新規則且應用於基因組分布，或其可定期應用，諸如至少一年一次。

當技術進步允許更精細分辨率SNP基因組分布時，亦可擴展個體疾病風險之資訊。如以上所指示，使用用於掃描500,000種SNP之微陣列技術可易於產生初始SNP基因組分布。考慮到單體型區塊之性質，此數字將個體基因組中所有SNP之代表性分布計算在內。但是，存在約1000萬種經估算通常存在於人類基因組中之SNP(國際人類基因組單體型圖計劃；www.hapmap.org)。當技術進步允許以更精細詳細程度對SNP進行可行的成本有效之分辨時，諸如1,000,000、1,500,000、2,000,000、3,000,000或更多種SNP之微陣列，或全基因組定序，可產生更詳細SNP基因組分布。同樣，計算分析方法之進步將使得能夠對更精細SNP基因組分布進行成本有效之分析且更新SNP-疾病相關性之主要資料庫。

在一些實施例中，可自個體收集「現場調用(field deployed)」機制，且併入個體之表型分布中。舉例而言，個體可具有基於遺傳資訊產生之初始表型分布。所產生之初始表型分布包括個人行為計劃中所報告之不同表型之風險因子以及建議治療或預防措施。分布可包括關於某一病狀之可用藥物治療之資訊，及/或關於飲食變化或鍛煉方案之建議。個體可選擇造訪醫師或遺傳學顧問或經由網路

入口或電話聯繫醫師或遺傳學顧問以討論其表型分布。個體可決定採取某一行為過程，例如採取特定藥物治療、改變其飲食及個人行為計劃上建議之其他可能行為。個體可隨後呈遞生物樣本以評估其身體狀況變化及可能的風險因子變化。

個體可具有由直接呈遞生物樣本至產生基因組分布及表型分布之機構(或相關機構，諸如產生遺傳分布及表型分布之實體的合約機構)判定之變化。或者，個體可使用「現場調用」機制，其中個體可呈遞其唾液、血液或其他生物樣本至其家庭偵測裝置中、藉由第三方分析，且傳輸資料以併入另一表型分布中。舉例而言，個體可接收基於報告個體心肌梗塞(MI)之終生風險增加之遺傳學資料之初始表型報告。該報告亦可具有降低MI風險之預防措施建議，諸如降膽固醇藥物及飲食改變。個體可選擇聯繫遺傳學顧問或醫師來討論報告及預防措施且決定改變其飲食。採用新飲食一段時間後，個體可造訪其私人醫師以量測其膽固醇含量。可傳輸(例如經由網際網路)新資訊(膽固醇含量)至具有基因組資訊之實體，且使用新資訊產生具有心肌梗塞及/或其他病狀新風險因子之個體之新表型分布。

個體亦可使用「現場調用」機制或直接機制測定對特定藥物治療之個體反應。舉例而言，個體可對所量測藥物具有反應，且可使用該資訊確定更有效治療。可量測資訊包括(但不限於)代謝物含量、葡萄糖含量、離子含量(例如鈣、鈉、鉀、鐵)、維生素、血細胞計數、體重指數

(BMI)、蛋白質含量、轉錄產物含量、心率等，可藉由易於獲得之方法測定且可包括在演算法中以組合初始基因組分布來確定修改之綜合風險估算得分。風險估算得分可為GCI得分。

遺傳複合指數(GCI)

在一些實施例中，組合且分析關於多種遺傳標記或變異體與一或多種疾病或病狀之相關性之資訊以產生遺傳複合指數(GCI)得分。此得分併有已知風險因子以及諸如等位基因頻率及疾病發病率之其他資訊及假定。可使用GCI來定性估算疾病或病狀與一組遺傳標記之組合效應的相關性。可使用GCI得分向未受遺傳學培訓之個人提供基於當前科學研究其疾病個體風險與有關群體相比如何之可靠(亦即穩固)、可以理解及/或直觀的意義。

可使用GCI得分產生GCI附加得分。本文所揭示之方法包涵使用GCI得分，且一般技術者將易於認識到替代如本文所述之GCI得分使用GCI附加得分或其變化形式。GCI附加得分可含有所有GCI假定，包括病狀之風險(諸如終生風險)、年齡決定之發病率及/或年齡決定之發生率。個體之終生風險可隨後計算為GCI附加得分，其與個體之GCI得分除以平均GCI得分成比例。可由具有類似祖先背景之一群個體，例如一群白種人、亞洲人、東印度人或具有共同祖先背景之其他群體測定平均GCI得分。群體可包含至少5、10、15、20、25、30、35、40、45、50、55或60個個體。在一些實施例中，可自至少75、80、95或100個個體

測定平均值。可藉由測定個體之GCI得分，以GCI得分除以平均相對風險且乘以病狀或表型之終生風險來測定GCI附加得分。舉例而言，使用美國公開案第20080131887號及PCT公開案第WO/2008/067551號之資料，可測定個體之GCI或GCI附加得分。可使用該等得分產生關於個體之表型分布中一或多種病狀之遺傳風險(諸如估算終生風險)之資訊。該等方法允許計算一或多種表型或病狀之估算終生風險或相對風險。單一病狀之風險可基於一或多種SNP。舉例而言，表型或病狀之估算風險可基於至少2、3、4、5、6、7、8、9、10、11或12種SNP，其中用於估算風險之SNP可為公開SNP、測試SNP或兩者。

可產生所關注之各疾病或病狀之GCI得分。可收集此等GCI得分以形成個體之風險分布。GCI得分可數位儲存以便其易於在任一時間點獲得以便產生風險分布。風險分布可分為寬的疾病種類，諸如癌症、心臟病、代謝障礙、精神病症、骨骼疾病或年齡發作病症。寬的疾病種類可進一步分為子類。舉例而言，對於諸如癌症之寬種類，癌症子類可諸如藉由類型(肉瘤、癌瘤或白血病等)或組織特異性(神經、乳房、卵巢、睪丸、前列腺、骨骼、淋巴結、胰腺、食道、胃部、肝臟、腦、肺、腎臟等)列出。另外，風險分布可顯示關於當調整個體年齡或各種風險因子時，如何預測GCI得分改變之資訊。舉例而言，特定疾病之GCI得分可考慮所採用之飲食或預防措施(戒菸、藥物攝取、二次根治性乳房切除術、子宮切除術及其類似措施)

之變化效應。

可產生個體之GCI得分，其提供關於個體之獲得至少一種疾病或病狀或對至少一種疾病或病狀之易感性之風險的易於理解之資訊。可產生單一疾病或病狀或許多疾病或病狀的一或多個GCI得分。可藉由線上入口獲得該或該等GCI得分。或者，可以紙張形式提供該或該等GCI得分，而後續更新亦以紙張形式提供。紙張形式可郵寄給個體或其健康護理管理者或親自提供。

用於產生不同基因座之組合效應之穩固GCI得分之方法可基於所研究之各基因座之報告個體風險。舉例而言，鑑別所關注疾病或病狀且隨後查詢包括(但不限於)資料庫、專利公開案及科學文獻之資訊源中的關於疾病或病狀與一或多個遺傳基因座之相關性的資訊。使用品質標準管理且評估此等資訊源。在一些實施例中，評估過程包括多個步驟。在其他實施例中，評估資訊源之多個品質標準。使用來源於資訊源之資訊鑑別所關注之各疾病或病狀之一或多個遺傳基因座的優勢率或相對風險。

在一替代實施例中，至少一個遺傳基因座之優勢率(OR)或相對風險(RR)自資訊源不可得或不可獲得。接著使用以下計算RR：(1)相同基因座之多個等位基因之報告OR；(2)來自諸如HapMap資料組之資料組之等位基因頻率；及/或(3)來自可用來源(例如CDC、國家健康統計中心(National Center for Health Statistics)等)之疾病/病狀發病率，以推算出所關注之所有等位基因之RR。在一實施例中，單獨

或獨立地估算相同基因座之多個等位基因的OR。在一較佳實施例中，組合相同基因座之多個等位基因的OR以說明不同等位基因之OR之間的相關性。在一些實施例中，使用既定疾病模型(包括(但不限於)諸如乘法、加法、哈佛大學(Harvard)改進模型、顯性效應之模型)產生根據所選模型代表個體之風險的中間得分。

可使用之方法分析多種所關注疾病或病狀之模型且將由此等不同模型獲得之結果相關聯；此使可能因選擇特定疾病模型引入之可能誤差降至最小。此方法使獲自資訊源之發病率、等位基因頻率及OR之估算值之合理誤差對計算相對風險之影響最小。不受理論限制，由於發病率估算值對RR之影響的「線性」或單調性質，因此不正確估算發病率對最終等級得分具有極少或無影響，但限制條件為對所有產生報告之個體總是應用相同模型。

本文所述之方法亦可考慮環境/行為/人口資料作為其他「位點(locus)」。在相關方法中，該等資料可獲自諸如醫學或科學文獻或資料庫之資訊源(例如吸菸w/肺癌之相關性，或自保險業健康風險評估)。本文亦揭示關於一或多種複合疾病所產生之GCI得分。複合疾病可受多種基因、環境因素及其相互作用影響。當研究複合疾病時，可能需要分析大量可能相互作用。可使用諸如Bonferroni校正的用以校正多重比較之程序產生GCI得分。或者，當測試獨立或顯示特殊類型之相關性時，可使用西姆斯檢驗(Simes's test)控制總體顯著性水準(亦稱為「族系誤差

率」)(Sarkar S., *Ann Stat* 26:494-504 (1998))。若對於 $1, \dots, K$ 中之任一 k , $p_{(k)} \leq \alpha k/K$, 則西姆斯檢驗拒絕所有 K 測試特異性虛無假設為真之全局虛無假設(Simes, R. J., *Biometrika* 73:751-754 (1986))。

在多基因及多環境因素分析之情形中可使用之其他實施例控制錯誤發現率-亦即錯誤拒絕之拒絕虛無假設的預期比例。當虛無假設之一部分可假定為錯誤(如在微陣列研究中)時, 此方法可尤其適用。Devlin等人(*Genet. Epidemiol.* 25:36-47 (2003))提出當在多基因座相關性研究中測試大量可能基因 $\times$ 基因相互作用時控制錯誤發現率之Benjamini及Hochberg(*J. R. Stat. Soc. Ser. B* 57:289-300 (1995))遞升程序之變化形式。Benjamini及Hochberg程序係關於西姆斯檢驗; 設定 $k^* = \max k$, 使得 $p_{(k)} \leq \alpha k/K$, 其拒絕對應於 $p_{(1)}, \dots, p_{(k^*)}$ 之所有 k^* 虛無假設。實際上, 當所有虛無假設為真時, Benjamini及Hochberg程序簡化為西姆斯檢驗(Benjamini及Yekutieli, *Ann. Stat.* 29:1165-1188 (2001))。

本文亦提供個體分級, 其中基於中間得分與個體之群體相比較以產生最終等級得分來將個體分級, 最終等級得分可表示為群體中之等級, 諸如第99百分位數或第99、第98、第97、第96、第95、第94、第93、第92、第91、第90、第89、第88、第87、第86、第85、第84、第83、第82、第81、第80、第79、第78、第77、第76、第75、第74、第73、第72、第71、第70、第69、第65、第60、第55、第50、第45、第40、第40、第35、第30、第25、第

20、第15、第10、第5或第0百分位數。等級得分可顯示為諸如第100至第95百分位數、第95至第85百分位數、第85至第60百分位數或第100與第0百分位數之間的任一子範圍之範圍。個體亦可以四分位數分等級，諸如上第75四分位數或最低第25四分位數。個體亦可與群體之平均值或中值得分相比分等級。

在一實施例中，與個體相比較之群體包括來自各種地理及民族背景之許多個人，諸如全球性群體。或者，與個體相比較之群體限於特定地理、家系、種族、性別、年齡(例如，胎兒、新生兒、幼兒、青少年、少年、成人、老人)或疾病病況(例如，帶症狀、無症狀、攜帶者、早期發作、晚期發作)。在一些實施例中，與個體相比較之群體來源於公開及/或私人資訊源中所報告之資訊。

使用多步驟方法產生GCI得分。舉例而言，最初對於各研究病狀，自各遺傳標記之優勢率計算相對風險。對於所有發病率值 $p=0.01$ 、 0.02 、... 0.5 ，基於發病率及HapMap等位基因頻率計算HapMap CEU群體之GCI得分。若GCI得分在不同發病率下相同，則所考慮之唯一假定為存在乘法模型(multiplicative model)。否則，會判定模型對發病率敏感。獲得無檢出值(no-call value)之任何組合的HapMap群體中得分之相對風險及分布。對於各新個體，將個體得分與HapMap分布相比較且所得得分為個體在此群體中之等級。歸因於在該方法中所作出之假定，報告得分之分辨率可能較低。將群體分割為分位數(3-6個格(bin))且所報告

格應為個體等級所屬之格。基於諸如各疾病之得分分辨率的考慮因素，不同疾病之格之編號可不同。在不同HapMap個體之得分持平之情況下，應使用平均等級。

較高GCI得分可解釋為獲得病狀或疾病或診斷患有病狀或疾病之風險增加的指示。通常使用數學模型推算出GCI得分。GCI得分可基於說明關於群體及/或疾病或病狀之隱含資訊之不完全性質之數學模型。數學模型可包括至少一個假定作為計算GCI得分之基礎之部分，其中該假定包括(但不限於)：提供優勢率值之假定；病狀之發病率已知的假定；群體中之基因型頻率已知的假定；及/或顧客來自與研究所用之群體及HapMap相同之家系背景的假定；合併風險為個體遺傳標記之不同風險因子之乘積的假定。GCI亦可包括基因型之多基因型頻率為各SNP或個別遺傳標記之等位基因頻率之乘積(例如，不同SNP或遺傳標記在群體中獨立)的假定。

乘法模型

可在以下假定下計算GCI得分：歸於遺傳標記組之風險為歸於個別遺傳標記之風險的乘積。因此，該等不同遺傳標記獨立於其他遺傳標記造成疾病之風險。形式上存在具有風險等位基因 $r_1, \dots, r_k$ 及非風險等位基因 $n_1, \dots, n_k$ 之 k 個遺傳標記。在SNP i 中，三個可能的基因型值表示為 $r_i r_i$ 、 $n_i r_i$ 及 $n_i n_i$ 。可由向量 $(g_1, \dots, g_k)$ 描述個體之基因型資訊，其中根據位置 i 中之風險等位基因之編號， g_i 可為 0、1 或 2。若以 r_i 表示，則將位置 i 之異型接合基因型之相對風險與相同位

置之同型接合非風險等位基因相比較。換言之，
 $\lambda \frac{i}{1} = \frac{P(D|n_i r_i)}{P(D|n_i n_i)}$ 。類似地， $r_i r_i$ 基因型之相對風險表示為
 $\lambda \frac{i}{2} = \frac{P(D|n_i r_i)}{P(D|n_i n_i)}$ 。在乘法模型下，假定具有基因型 $(g_1, \dots, g_k)$ 之個體的風險為 $GCI(g_1, \dots, g_k) = \prod_{i=1}^k \lambda_{g_i}^i$ 。

估算相對風險

在另一實施例中，已知不同遺傳標記之相對風險且可使用乘法模型用於風險評估。然而，在一些涉及相關性研究之實施例中，研究設計防止報告相對風險。在一些病例對照研究中，不能在不作進一步假定之情況下直接自資料計算相對風險。替代報告相對風險，通常報告基因型之優勢率(OR)，其為指定風險基因型($r_i r_i$ 或 $n_i r_i$)之情況下攜帶疾病之勝率相對於指定風險基因型之情況下未攜帶疾病之勝率。形式上，

$$OR_i^1 = \frac{P(D|n_i r_i)}{P(D|n_i n_i)} \cdot \frac{1 - P(D|n_i n_i)}{1 - P(D|n_i r_i)}$$

$$OR_i^2 = \frac{P(D|r_i r_i)}{P(D|n_i n_i)} \cdot \frac{1 - P(D|n_i n_i)}{1 - P(D|r_i r_i)}。$$

自優勢率得到相對風險可能會要求其他假定。諸如假定已知或估算出整個群體中之等位基因頻率 $a = f_{n_i n_i}$ 、 $b = f_{n_i r_i}$ 及 $c = f_{r_i r_i}$ (此等頻率可自諸如包括120個染色體之HapMap資料組之當前資料組估算)及/或已知疾病之發病率 $p = p(D)$ 。自前三個方程式可得到：

$$p = a \cdot P(D|n_i n_i) + b \cdot P(D|n_i r_i) + c \cdot P(D|r_i r_i)$$

$$OR_i^1 = \frac{P(D|n_i r_i)}{P(D|n_i)} \cdot \frac{1 - P(D|n_i r_i)}{1 - P(D|n_i)}$$

$$OR_i^2 = \frac{P(D|r_i n_i)}{P(D|n_i n_i)} \cdot \frac{1 - P(D|r_i n_i)}{1 - P(D|n_i r_i)}。$$

藉由定義相對風險，除以項 $pP(D|n_i n_i)$ 之後，第一個方程式可重寫為：

$$\frac{1}{P(D|n_i n_i)} = \frac{a + b\lambda_1^i + c\lambda_2^i}{p}$$

且因此，後兩個方程式可重寫為：

$$OR_i^1 = \lambda_1^i \cdot \frac{(a-p) + b\lambda_1^i + c\lambda_2^i}{a + (b-p)\lambda_1^i + c\lambda_2^i}$$

(1)

$$OR_i^2 = \lambda_2^i \cdot \frac{(a-p) + b\lambda_1^i + c\lambda_2^i}{a + b\lambda_1^i + (c-p)\lambda_2^i}。$$

應注意，當 $a=1$ (非風險等位基因頻率為1) 時，方程式系統1相當於 Zhang 及 Yu (*JAMA*, 280:1690-1691 (1998)) 中之 Zhang 及 Yu 式，該文獻係以引用的方式全部併入本文中。與 Zhang 及 Yu 式相對照，一些實施例考慮群體中可能會影響相對風險的等位基因頻率。另外，與獨立地計算各相對風險相反，一些實施例考慮相對風險之相互關聯性。

方程式系統1可重寫為具有至多四個可能解之兩個二次方程式。可使用梯度下降演算法解答此等方程式，其中起始點經設定為優勢率，例如 $\lambda_1^i = OR_i^1$ 及 $\lambda_2^i = OR_i^2$ 。

舉例而言：

$$f_1(\lambda_1, \lambda_2) = OR_1^1(a + (b - p)\lambda_1^i + c\lambda_2^i) - \lambda_1^i \cdot ((a - p) + b\lambda_1^i + c\lambda_2^i)$$

$$f_2(\lambda_1, \lambda_2) = OR_2^2(a + b\lambda_1^i + (c - p)\lambda_2^i) - \lambda_2^i \cdot ((a - p) + b\lambda_1^i + c\lambda_2^i)$$

得到此等方程式之解相當於得到函數 $g(\lambda_1, \lambda_2) = f_1(\lambda_1, \lambda_2)^2 + f_2(\lambda_1, \lambda_2)^2$ 之最小值。

因此，

$$\frac{dg}{d\lambda_1} = 2f_1(\lambda_1, \lambda_2) \cdot b \cdot (\lambda_2 - OR_2) + 2f_2(\lambda_1, \lambda_2)(2b\lambda_1 + c\lambda_2 + a - OR_1b - p + OR_1p)$$

$$\frac{dg}{d\lambda_2} = 2f_2(\lambda_1, \lambda_2) \cdot c \cdot (\lambda_1 - OR_1) + 2f_1(\lambda_1, \lambda_2)(2c\lambda_2 + b\lambda_1 + a - OR_2c - p + OR_2p)。$$

在此實例中，藉由設定 $x_0 = OR_1$ 、 $y_0 = OR_2$ ，設定值 $[\varepsilon] = 10^{-10}$ 為演算法之公差常數。在迭代 i 中，定義

$$\gamma = \min \left\{ 0.001, \frac{x_{i-1}}{[\varepsilon] + 10 \left| \frac{dg}{d\lambda_1}(x_{i-1}, y_{i-1}) \right|}, \frac{y_{i-1}}{[\varepsilon] + 10 \left| \frac{dg}{d\lambda_2}(x_{i-1}, y_{i-1}) \right|} \right\}, \text{ 接著設定}$$

$$x_i = x_{i-1} - \gamma \frac{dg}{d\lambda_1}(x_{i-1}, y_{i-1})$$

$$y_i = y_{i-1} - \gamma \frac{dg}{d\lambda_2}(x_{i-1}, y_{i-1})$$

重複迭代直至 $g(x_i, y_i) < \text{公差}$ ，其中公差在所提供代碼中經設定為 10^{-7} 。

在此實例中，此等方程式提供 a 、 b 、 c 、 p 、 OR_1 及 OR_2 之不同值的正確解。

相對風險估算之穩固性

在一些實施例中，量測不同參數(發病率、等位基因頻率及優勢率誤差)對相對風險之估算值的影響。為量測等位基因頻率及發病率估算值對相對風險值之影響，計算(在HWE下)一組不同優勢率及不同等位基因頻率之值的相對風險且關於0至1範圍內之發病率值將此等計算結果繪圖。另外，對於固定發病率值而言，可將所得相對風險作為風險-等位基因頻率之函數繪圖。在 $p=0$ 之情況下， $\lambda_1=OR_1$ 且 $\lambda_2=OR_2$ 且在 $p=1$ 之情況下， $\lambda_1=\lambda_2=0$ 。此可直接由方程式計算。另外，在一些實施例中，當風險等位基因頻率較高時， λ_1 更接近線性函數且 λ_2 更接近具有有限二階導數之凹函數。在極限中，當 $c=1$ 時， $\lambda_2=OR_2+p(1-OR_2)$ 且 $\lambda_1=OR_1-\frac{(OR_1-1)pOR_2}{OR_2(1-p)+pOR_1}$ 。若 $OR_1 \approx OR_2$ ，則後者亦接近線性函數。當風險-等位基因頻率較低時， λ_1 及 λ_2 逼近函數 $1/p$ 之行為。

在極限中，當 $c=0$ 時， $\lambda_1=\frac{OR_1}{1-p+pOR_1}$ 、 $\lambda_2=\frac{OR_2}{1-p+pOR_2}$ 。此指示對於高風險等位基因頻率，不正確發病率估算值將不會顯著影響所得相對風險。另外，對於低風險-等位基因頻率，若用 $p'=ap$ 之發病率值取代正確發病率 p ，則所得相對風險將至多偏離 $\frac{1}{\alpha}$ 倍。

計算GCI得分

在一實施例中，藉由使用代表相關群體之參考集計算GCI。此參考集可為HapMap或另一基因型資料組中之群體之一。

在此實施例中，GCI係如下計算：對於 k 個風險位點中之每一者，使用方程式系統1由優勢率計算相對風險。接

著，計算參考集中各個體之乘法得分。乘法得分為 s 之個體的GCI為參考資料組中得分 $s' \leq s$ 之所有個體的分數。舉例而言，若參考集中50%之個體具有小於 s 之乘法得分，則個體之最終GCI得分應為0.5。若不同基因型或單體型組合之優勢率或相對風險已知(在一些情況下，此等可見於文獻)，則可歸納GCI以說明SNP-SNP相互作用。

如本文所述，在GCI得分中可使用乘法模型，然而，其他模型可用於確定GCI得分之目的。其他適當模型包括(但不限於)：

加法模型。在加法模型下，具有基因型 $(g_1, \dots, g_k)$ 之個體之風險假定為 $GCI(g_1, \dots, g_k) = \sum_{i=1}^k \lambda_{g_i}^i$ 。

歸納加法模型。在歸納加法模型下，假定存在函數 f ，使得具有基因型 $(g_1, \dots, g_k)$ 之個體的風險為 $GCI(g_1, \dots, g_k) = \sum_{i=1}^k f(\lambda_{g_i}^i)$ 。

哈佛大學改進得分(Het)。此得分來源於Colditz等人(*Cancer Causes and Controls*, 11:477-488 (2000))，其係全部併入本文中。Het得分基本上為歸納加法得分，儘管對優勢率值而非相對風險運算函數 f 。此可用於難以估算相對風險之情形。為定義函數 f ，中間函數 g 定義為：

$$g(x) = \begin{cases} 0 & 1 < x \leq 1.09 \\ 5 & 1.09 < x \leq 1.49 \\ 10 & 1.49 < x \leq 2.99 \\ 25 & 2.99 < x \leq 6.99 \\ 50 & 6.99 < x \end{cases}$$

接著，計算量 $het = \sum_{i=1}^k p_{het}^i g(OR_i^i)$ ，其中 p_{het}^i 為參考群體中之

SNP i 中的異型接合個體之頻率。接著，函數 f 定義為 $f(x)=g(x)/het$ ，且哈佛大學改進得分 (Het) 簡單定義為 $\sum_{i=1}^k f(OR_{g_i}^i)$ 。

哈佛大學改進得分 (Hom)。此得分類似於 Het 得分，例外為值 het 由值 $hom = \sum_{i=1}^k p_{hom}^i g(OR_i^i)$ 置換，其中 p_{hom}^i 為具有同型接合風險等位基因之個體的頻率。

最大優勢率。在此模型中，假定遺傳標記之一 (具有最大優勢率之遺傳標記) 提供整個組之組合風險的下限。形式上，具有基因型 $(g_1, \dots, g_k)$ 之個體的得分為 $GCI(g_1, \dots, g_k) = \max_{i=1}^k OR_{g_i}^i$ 。

得分之間的比較描述於實例 1 中且 GCI 得分評估描述於實例 2 中。

擴展模型至任意數目之變異體

可擴展模型至任意數目之可能變異體存在的情況。先前考慮因素涉及存在三種可能變異體 (nn, nr, rr) 之情況。通常，當多 SNP 相關性已知時，任意數目之變異體可見於群體中。舉例而言，當兩個遺傳標記之間的相互作用與病狀相關聯時，存在九種可能的變異體。此產生 8 個不同優勢率值。

為歸納初始式，可假定存在 $k+1$ 種可能的變異體 $a_0, \dots, a_k$ ，其中頻率為 $f_0, f_1, \dots, f_k$ 、實測優勢率為 $1, OR_1, \dots, OR_k$ 及未知相對風險值為 $1, \lambda_1, \dots, \lambda_k$ 。另外，假定關於 a_0 量測所有相對風險及優勢率，且因此 $\lambda_i = \frac{P(D|a_i)}{P(D|a_0)}$ 及 $OR_i = \frac{P(D|a_i)}{P(D|a_0)} \cdot \frac{1-P(D|a_i)}{1-P(D|a_0)}$ 。基於

$$p = \sum_{i=0}^k f_i P(D|a_i),$$

確定：

$$OR_i = \lambda_i \frac{\sum_{i=0}^k f_i \lambda_i - p}{\sum_{i=0}^k f_i \lambda_i - p}.$$

此外，若設定 $C = \sum_i f_i \lambda_i$ ，則此舉得到方程式：

$$\lambda_i = \frac{C \cdot OR_i}{C - p + OR_i p},$$

且因此得到：

$$C = \sum_{i=0}^k f_i \lambda_i = \sum_{i=0}^k \frac{C \cdot OR_i f_i}{C - p + OR_i p}$$

或

$$1 = \sum_{i=0}^k \frac{OR_i f_i}{C - p + OR_i p}.$$

後者為具有單變數(C)之方程式。此方程式可產生許多不同解(基本上，多達 $k+1$ 種不同解)。可使用諸如梯度下降之標準優化工具得到 $C_0 = \sum f_i t_i$ 之最接近解。

本文亦提供用於定量風險因子之穩固得分框架。儘管不同遺傳模型可產生不同得分，但結果通常相關。因此，風險因子之定量通常不取決於所用模型。

估算相對風險病例對照研究

本文亦揭示在病例對照研究中由多個等位基因之優勢率

估算相對風險之方法。與先前方法相對照，該方法考慮等位基因頻率、疾病發病率及不同等位基因之相對風險之間的相關性。量測該方法對模擬病例對照研究之效能，發現其極為精確。

方法

在測試特定 SNP 與疾病 D 之相關性的情況下，R 及 N 表示此特定 SNP 之風險及非風險等位基因。P(RR|D)、P(RN|D) 及 P(NN|D) 表示在個人分別對於風險等位基因而言為同型接合型、異型接合型或對於非風險等位基因而言為同型接合型之情況下染上疾病之機率。使用 f_{RR} 、 f_{RN} 及 f_{NN} 表示群體中三種基因型之頻率。使用此等定義，相對風險定義為：

$$\lambda_{RR} = \frac{P(D|RR)}{P(D|NN)}$$

$$\lambda_{RN} = \frac{P(D|RN)}{P(D|NN)}$$

在病例對照研究中，可估算值 P(RR|D)、P(RR|~D)，亦即病例及對照中 RR 之頻率，以及 P(RN|D)、P(RN|~D)、P(NN|D) 及 P(NN|~D)，亦即病例及對照中 RN 及 NN 之頻率。為估算相對風險，可使用貝葉斯定律 (Bayes law) 得到：

$$\lambda_{RR} = \frac{P(RR|D)f_{NN}}{P(NN|D)f_{RR}}$$

$$\lambda_{RN} = \frac{P(D|RN)f_{NN}}{P(D|NN)f_{RR}}$$

因此，若基因型之頻率已知，則可使用彼等頻率計算相

對風險。群體中基因型之頻率無法由病例對照研究本身計算，因為其取決於群體中疾病之發病率。詳言之，若疾病發病率為 $p(D)$ ，則：

$$f_{RR} = P(RR|D)p(D) + P(RR|\sim D)(1-p(D))$$

$$f_{RN} = P(RN|D)p(D) + P(RN|\sim D)(1-p(D))$$

$$f_{NN} = P(NN|D)p(D) + P(NN|\sim D)(1-p(D))$$

當 $p(D)$ 足夠小時，基因型之頻率可近似對照群體中基因型之頻率，但發病率高時此值將不為精確估算值。然而，若給出參考資料組(例如，HapMap [引證])，則可基於參考資料組估算基因型頻率。

大多數當前研究並不使用參考資料組估算相對風險，且僅報告優勢率。優勢率可書寫為：

$$OR_{RR} = \frac{P(RR|D)P(NN|\sim D)}{P(NN|D)P(RR|\sim D)}$$

$$OR_{RN} = \frac{P(RN|D)P(NN|\sim D)}{P(NN|D)P(RN|\sim D)}$$

由於通常無需具有群體中等位基因頻率之估算值，因此優勢率通常有利；為計算優勢率，通常需要病例及對照中之基因型頻率。

在一些情況下，儘管基因型資料本身不可得，但可用諸如優勢率之匯總資料。此為基於先前病例-對照研究之結果進行統合分析(meta-analysis)之情況。在此情況下，顯示自優勢率得出相對風險之方法。使用以下方程式適用的事實：

$$p(D) = f_{RR}P(D|RR) + f_{RN}P(D|RN) + f_{NN}P(D|NN)$$

若此方程式除以 $P(D|NN)$ ，得到：

$$\frac{p(D)}{p(D|NN)} = f_{RR}\lambda_{RR} + f_{RN}\lambda_{RN} + f_{NN}$$

此允許以以下方式書寫優勢率：

$$OR_{RR} = \frac{P(D|RR)(1 - P(D|NN))}{P(D|NN)(1 - P(D|RR))} = \lambda_{RR} \frac{\frac{p(D)}{p(D|NN)} - p(D)}{\frac{p(D)}{p(D|NN)} - p(D)\lambda_{RR}} =$$

$$\lambda_{RR} \frac{f_{RR}\lambda_{RR} + f_{RN}\lambda_{RN} + f_{NN} - p(D)}{f_{RR}\lambda_{RR} + f_{RN}\lambda_{RN} + f_{NN} - p(D)\lambda_{RR}}$$

藉由類似計算，得到以下方程式系統：

$$OR_{RR} = \lambda_{RR} \frac{f_{RR}\lambda_{RR} + f_{RN}\lambda_{RN} + f_{NN} - p(D)}{f_{RR}\lambda_{RR} + f_{RN}\lambda_{RN} + f_{NN} - p(D)\lambda_{RR}}$$

$$OR_{RN} = \lambda_{RN} \frac{f_{RR}\lambda_{RR} + f_{RN}\lambda_{RN} + f_{NN} - p(D)}{f_{RR}\lambda_{RR} + f_{RN}\lambda_{RN} + f_{NN} - p(D)\lambda_{RN}}$$

方程式 1

若優勢率、群體中基因型之頻率及疾病發病率已知，則可藉由解答此方程式組得出相對風險。

應注意此等方程式為兩個二次方程式，且因此其最多有四個解。然而，如以下所示，此方程式通常存在一個可能解。

應注意當 $f_{NN}=1$ 時，方程式系統 1 相當於 Zhang 及 Yu 式；然而，此處考慮群體之等位基因頻率。此外，該方法考慮兩種相對風險相互依賴之事實，而先前方法提出獨立地計算各相對風險。

多等位基因基因座之相對風險。若考慮多種標記或其他多種等位基因變異體，則計算稍微複雜。 a_0 、 a_1 、...、 a_k 表示為可能的 $k+1$ 個等位基因，其中 a_0 為非風險等位基因。假定群體中 $k+1$ 個可能的等位基因之等位基因頻率為 f_0 、 f_1 、 f_2 、...、 f_k 。對於等位基因 i ，相對風險及優勢率定義為：

$$\lambda_i = \frac{P(D|a_i)}{P(D|a_0)}$$

$$OR_i = \frac{P(D|a_i)(1-P(D|a_0))}{P(D|a_0)(1-P(D|a_i))} = \lambda_i \frac{1-P(D|a_0)}{1-P(D|a_i)}$$

對於疾病發病率適用以下方程式：

$$p(D) = \sum_{i=0}^k f_i P(D|a_i)$$

因此，藉由使方程式之兩側除以 $p(D|a_0)$ ，得到：

$$\frac{p(D)}{P(D|a_0)} = \sum_{i=0}^k f_i \lambda_i$$

得到：

$$OR_i = \lambda_i \frac{\sum_{i=0}^k f_i \lambda_i - p(D)}{\sum_{i=0}^k f_i \lambda_i - \lambda_i p(D)},$$

藉由設定 $C = \sum_{i=0}^k f_i \lambda_i$ ，結果為 $\lambda_i = C \cdot \frac{OR_i}{p(D)OR_i + C - p(D)}$ 。因此藉

由定義 C ，其為：

$$1 = \sum_{i=0}^k f_i \frac{\lambda_i}{C} = \sum_{i=0}^k \frac{f_i OR_i}{p(D)OR_i + C - p(D)}$$

此為具有單變量 C 之多項式方程式。 C 確定後，確定相對風險。多項式具有次數 $k+1$ ，且因此預期其具有至多 $k+1$ 個解。然而，由於方程式之右手側作為 C 之函數嚴格遞減，因此此方程式可通常僅有一個解。接著使用二分搜尋法找到解，因為該解限於 $C=1$ 與 $C = \sum_{i=0}^k OR_i$ 之間。

相對風險估算之穩固性。量測各不同參數(發病率、等位基因頻率及優勢率誤差)對相對風險之估算值的影響。為量測等位基因頻率及發病率估算值對相對風險值之影響，自一組不同優勢率、不同等位基因頻率值(在HWE下)計算相對風險且關於0至1範圍內之發病率值將此等計算結果繪圖。

另外，對於固定發病率值而言，可將所得相對風險作為風險-等位基因頻率之函數繪圖。明顯地，在所有情況下，當 $p(D)=0$ 時， $\lambda_{RR}=OR_{RR}$ 且 $\lambda_{RN}=OR_{RN}$ ，且當 $p(D)=1$ 時， $\lambda_{RR}=\lambda_{RN}=0$ 。此可直接由方程式1計算。另外，當風險等位基因頻率較高時， λ_{RR} 逼近線性行為，且 λ_{RN} 逼近具有有限二階導數之凹函數。當風險-等位基因頻率較低時， λ_{RR} 及 λ_{RN} 逼近函數 $1/p(D)$ 之行為。此意謂對於高風險-等位基因頻率而言，錯誤發病率估算值通常將不會嚴重影響所得相對風險。

優勢率相對於相對風險。在流行病學文獻中，通常認為相對風險為風險之直觀且有教益之量度。然而，在全面性全基因組相關性研究中，在病例-對照研究之情形中不能直接計算出相對風險。可通常經由前瞻性研究估算相對風

險，其中長期研究一組健康個體。相反，通常在病例-對照研究中報告優勢率。優勢率為病例中相對於對照中攜帶風險等位基因之勝率之間的比率。對於罕見疾病而言，優勢率充分近似於相對風險；然而，對於常見疾病而言，優勢率可產生風險之誤導估算值，其中甚至當風險增加值較小時，優勢率亦可能相當高。

相對終生風險相對於相對風險。相對風險暗中假定對照目前均未患病。當估算患病機率時，此具有相關性。然而，若關注於生命歷程中之風險估算，或個體發展病狀之終生風險，則考慮一些對照最終將發展疾病之事實。相對終生風險經定義為在攜帶風險等位基因 r 之個體的生命期中發展病狀之風險與在攜帶非風險等位基因之個體的生命期中發展病狀之風險之間的比率。此不同於相對風險在基於發病率資訊之病例-對照研究中的標準使用。

可能的 $k+1$ 個等位基因表示為 $a_0, a_1, \dots, a_k$ ，其中 a_0 為非風險等位基因。假定群體中 $k+1$ 個可能的等位基因之等位基因頻率為 $f_0, f_1, f_2, \dots, f_k$ 。進一步假定所研究之個體可分成三個組：CA、Y及Z。CA表示病例，而Y及Z為對照。與來自Z之個體相反，假定來自Y之個體最終將發展病狀。亦由CO表示Y及Z之聯合，且由D表示Y及CA之聯合。假定 $|Y| = \alpha |CO| = \alpha(|Y| + |Z|)$ ，其中 α 為壽命期間將發展病狀之對照的分數。應注意 α 之上限為平均終生風險。 α 可能小於平均壽命，此取決於疾病發作年齡及對照年齡。

現在相對風險及優勢率可表示為：

$$\lambda_i = \frac{P(CA \vee Y | a_i)}{P(CA \vee Y | a_0)}$$

$$OR_i = \frac{P(a_i | CA)P(a_0 | CO)}{P(a_0 | CA)P(a_i | CO)}$$

優勢率可書寫為：

$$\begin{aligned} OR_i &= \frac{P(a_i | CA)P(a_0 | CO)}{P(a_0 | CA)P(a_i | CO)} = \frac{P(a_i | CA)}{P(a_0 | CA)} \cdot \frac{\alpha P(a_0 | Y) + (1-\alpha)P(a_0 | Z)}{\alpha P(a_0 | Y) + (1-\alpha)P(a_0 | Z)} = \\ &= \frac{P(CA | a_i)}{P(CA | a_0)} \cdot \frac{\alpha P(Y | a_0) + (1-\alpha)P(Z | a_0)}{\alpha P(Y | a_i) + (1-\alpha)P(Z | a_i)} = \\ &= \frac{P(CA | a_i)}{P(CA | a_0)} \cdot \frac{\alpha P(CA | a_0) + (1-\alpha)P(Z | a_0)}{\alpha P(CA | a_i) + (1-\alpha)P(Z | a_i)} \end{aligned}$$

基於貝葉斯定律自第一列推導至第二列，而第三列基於 CA 及 Y 基本上為相同群體，且因此 $P(CA|a_i)=P(Y|a_i)$ 之事實。現在使用事實 $P(Z|a_i)=1-P(CA|a_i)$ ，得到：

$$OR_i = \frac{P(CA | a_i)}{P(CA | a_0)} \cdot \frac{(2\alpha-1)P(CA | a_0) + 1-\alpha}{(2\alpha-1)P(CA | a_i) + 1-\alpha} = \lambda_i \cdot \frac{(2\alpha-1)P(CA | a_0) + 1-\alpha}{(2\alpha-1)P(CA | a_i) + 1-\alpha}$$

如前， $p(D) = \sum_{i=0}^k f_i P(D|a_i)$ ，其中 $p(D)$ 為平均終生風險。因此，使用方程式 $C = \frac{p(D)}{P(CA|a_0)} = \sum_{i=0}^k f_i \lambda_i$ ，且優勢率可重寫為：

$$OR_i = \lambda_i \cdot \frac{(2\alpha-1)P(D) + (1-\alpha)C}{(2\alpha-1)P(D)\lambda_i + (1-\alpha)C}$$

因此，若 C 給定，則可藉由指定下式得到相對終生風險：

$$\lambda_i = \frac{(1-\alpha)C \cdot OR_i}{(2\alpha-1)P(D)(1-OR_i) + (1-\alpha)C}$$

可藉由解答以下方程式得到 C：

$$1 = \sum_{i=0}^k f_i \frac{\lambda_i}{C} = \sum_{i=0}^k \frac{f_i(1-\alpha)OR_i}{(2\alpha-1)p(D)(1-OR_i) + (1-\alpha)C}$$

可驗證藉由定義 C 及優勢率， $C > (2\alpha-1)p(D)(OR_i-1)$ 。因此，右手側為 C 之遞減函數，且其可藉由應用二分搜尋法找出。

基於 GCI 之終生風險估算值。 GCI 基本上提供與具有所有相關 SNP 中之非風險等位基因之個體相比，個體之相對風險。為計算個體之終生風險，可取具有平均終生風險之個體之終生風險之乘積，且將此乘積除以群體之平均終生風險。此計算與平均終生風險及相對風險之定義一致。為計算平均終生風險，列舉所有可能基因型，且加和經計算為各單一 SNP 中變異體之相對風險之積的相對風險。

個人化行為計劃

本文揭示之個人化行為計劃提供基於個體之基因組分布的有意義、可實施資訊以改善個體之健康或安康。行為計劃提供考慮到特定基因型相關性而有益於個體之行為過程，且可包括投與治療性處理、監測治療之潛在需要或治療之作用或使生活方式在飲食、鍛煉及其他個人習慣/活動方面改變，其可基於個體之基因組分布而個人化為個人化行為計劃。或者，可對個體給予基於其基因組分布之特定等級，且另外視情況包括其他資訊，諸如家族史、現有生活習慣及地理，諸如(但不限於)工作條件、工作環境、人際關係、居家環境及其他。可併入之其他因素包括種

族、性別及年齡。各種飲食及鍛煉預防策略及其與降低疾病或病狀風險之相關性的優勢率亦可併入評級系統中。

此外，個人化行為計劃可針對個體修改或更新。例如若個體或其健康護理管理者最初要求諸如訂購計劃之自動更新，則可將修改或更新之個人化行為計劃自動發送給個體或其健康護理管理者。或者，可僅當個體或其健康護理管理者要求時才發送更新之個人化行為計劃。個人化行為計劃可基於許多因素修改或更新。舉例而言，可分析個體之更多遺傳相關性，且使用結果在初始個人化行為計劃上修改現有建議、添加其他建議或去除建議。在一些實施例中，個體可改變某些生活習慣/環境，或具有更多關於家族史、現有生活習慣及地理之資訊，諸如(但不限於)工作條件、工作環境、人際關係、居家環境及其他，或希望包括其更新年齡以獲得併有此等變化之個人化行為計劃。舉例而言，個體可遵守其初始個人化行為計劃，諸如降低其飲食或醫藥治療中之膽固醇，且因此可修改其個人化行為計劃建議或其心臟病風險或易感性可降低。

基於個體遵守個人化行為計劃上之建議，個人化行為計劃亦可具有預測之未來建議，或個體可進行或想起其他變化。舉例而言，個體之年齡增加將導致骨質疏鬆症風險增加，但取決於鈣量或諸如個人化行為中之彼等習慣的其他生活習慣，風險可降低。

個人化行為計劃可以具有個體表型分布及/或基因組分布之單次報告形式，報告給個體或其健康護理管理者。或者，可分別報告個人化行為計劃。個體可隨後執行其個人化行為計劃上之推薦行為。個體可選擇在執行其計劃上之任一行為之前諮詢其健康護理管理者。

所提供之個人化行為計劃亦可將許多病狀特異性資訊合併為合併行為步驟組。個人化行為計劃可合併包括(但不限於)各病狀發病率、與各病狀相關之疼痛的相對量及各病狀之治療類型的因素。舉例而言，若個體具有高心肌梗塞風險(例如，表示為較高GCI或GCI附加得分)，則個體可具有包括增加水果、蔬菜及穀物消耗之個人化行為計劃。然而，個體亦可具有乳糜瀉易感性，因此具有麵筋過敏症。因此，可禁止增加小麥消耗，且在個人化行為計劃中指示。

個人化行為計劃可提供醫藥(其包括指定處方藥物、保健食品及其類似物)建議、非醫藥建議或兩者。舉例而言，個人化行為計劃可包括建議醫藥作為預防法，諸如降膽固醇藥物用於易患心肌梗塞之個體，且諮詢醫師。個人化行為計劃亦可提供非醫藥建議，諸如遵守個人化生活方式計劃，其包括基於個體之基因組分布之包括鍛煉方案與飲食計劃。

個人化行為計劃建議可具有特定評級、標記或分類系統。各建議可藉由數字、顏色及/或字母方案或值分級或分類。建議可被分類且進一步分級。可使用許多變化，

諸如不同評級方案(在一或多個評級方案中使用字母、數字或顏色；字母、數字及/或顏色之組合；不同類型建議)。

舉例而言，確定個體基因組分布，且基於其基因組分布，將個人化行為計劃上對個體之建議分類成3個組：「A」表示不利或負作用，「N」表示中性或無顯著作用，且「B」表示有利或正作用。使用此系統作為一實例，對於個體而言分類為A之療法將包括對個體具有不利作用之藥物，分類為N之療法對個體將不具有任何顯著正或負作用，且分類為B之療法應有益於個體健康。使用相同分類系統，飲食計劃亦可分組為A、B、N。舉例而言，使個體過敏或應尤其避免之食物(例如，由於個體易患糖尿病或齲齒而應避免糖)應分類為A。對個體健康無顯著作用之食物可分類為N。尤其有益於個體之食物可分類為B，例如若個體具有高膽固醇，則低膽固醇食物應分類為B。個體鍛煉方案亦可基於相同系統。舉例而言，個體可能易患心臟問題且應避免劇烈鍛煉，且因此跑步可為A活動，而以某種步速散步或漫步可分類為B。對於一個個體而言，站立一段時間可為N，但對於易患靜脈曲張之另一個體而言為A。

此外，在A、N或B之各類別中，自最低至最高影響可存在其他分類級別，諸如1至5。舉例而言，療法可分類為A1，其指示輕微副作用，諸如稍有噁心，而A2將指示療法將引起嘔吐，而A5療法將引起嚴重不利作用，諸如過敏

性休克。相反地，B1將對個體具有些微正作用，而B5將對個體具有顯著正影響。舉例而言，若個體易患肺癌，或在成長時暴露於二手菸，則不吸菸個體可為B5，而不易患肺癌之個體可以該因素作為B4。

不同類別亦可由不同顏色表示(例如A可為紅色調)且表示對個體健康之低至高作用，色澤範圍可介於淡至深紅色色調，淡表示低副作用，至深紅色表示對個體健康具有嚴重不利作用。該系統亦可為顏色、數字或字母之連續範圍。例如類別可為A至G，其中A表示嚴重負面影響個體健康之食物、療法、生活習慣、環境及其他因素，而D表示具有最小正或負作用之因素，且G將表示高度有益於個體健康，而非具有A、N及B及/或其子類。或者，數字或顏色亦可表示影響個體健康之食物、療法、生活習慣、環境及其他因素之連續範圍，而非具有A至G。

在一些實施例中，可將個人化行為計劃中之特定療法、醫藥或其他生活要素分類、標記或評級。舉例而言，個體可具有包括鍛煉方案及飲食計劃之個人化行為計劃。鍛煉方案可包括一或多個等級或類別。舉例而言，鍛煉方案之等級可在諸如表1之A至E範圍內，其中各字母對應於一或多種類型之鍛煉，包括關於屬於各級別之活動類型、時間長度、指定時間段內之次數，且因此關於個體之推薦鍛煉方案之資訊。

表 1：鍛煉方案：心血管活動

等級	選擇1	選擇2	選擇3	選擇4
A	輕快散布2.5 mph， 每週3次， 每次20分鐘	游泳4圈， 每週3次	騎自行車5 mph， 每週3次， 每次20分鐘	輕快散布2.5 mph， 每週2次， 每次20分鐘 騎自行車5 mph， 每週1次， 每次20分鐘
B	漫步3.5 mph， 每週3次， 每次20分鐘	游泳6圈， 每週3次	騎自行車8 mph， 每週3次， 每次20分鐘	漫步3.5 mph， 每週2次， 每次20分鐘 騎自行車8 mph， 每週1次， 每次20分鐘
C	跑步4 mph， 每週3次， 每次20分鐘	游泳8圈， 每週3次	騎自行車10 mph， 每週3次， 每次20分鐘	跑步4 mph， 每週2次， 每次20分鐘 騎自行車10 mph， 每週1次， 每次20分鐘
D	跑步5 mph， 每週3次， 每次25分鐘	游泳10圈， 每週3次	騎自行車15 mph， 每週3次， 每次30分鐘	跑步5 mph， 每週2次， 每次25分鐘 騎自行車15 mph， 每週1次， 每次20分鐘
E	跑步6 mph， 每週3次， 每次30分鐘	游泳12圈， 每週3次	騎自行車15 mph， 每週3次， 每次40分鐘	跑步5 mph， 每週2次， 每次30分鐘 騎自行車15 mph， 每週一次， 每次40分鐘

在一實施例中，基於個體之基因組分布，個人化行為計劃對於個體而言可具有等級A，且因此個體之推薦鍛煉方案對於其心血管鍛煉而言將選自表1之列A。類似地，用於重量訓練之類似系統可為個體鍛煉方案之部分，且用於等級A之重量訓練選擇應推薦給個體。在一些實施例中，諸如(但不限於)個體現有飲食、鍛煉及其他個人習慣/活動之

因素，視情況諸如家族史、現有生活習慣及地理之其他資訊，諸如(但不限於)工作條件、工作環境、人際關係、居家環境、種族、性別、年齡及其他因素可併入個體基因組分布中以確定個體之鍛煉方案等級。此外，由於已知且併有個體生活習慣變化或更多因素，個體等級可改變，例如若個體遵守個人化行為計劃上之起始於等級A的推薦活動，則個體可要求評估且確定個體現在處於等級B之更新之個人化行為計劃。或者，個體個人化行為計劃可提供個體應考慮由等級A移至等級B以便使其最健康之時間安排。

個人化行為計劃亦可具有飲食計劃評級系統。舉例而言，飲食計劃之等級可為範圍介於1至5之系統，其中各數字對應於建議個體飲食中具有的脂肪、纖維、蛋白質、糖及其他營養素之特定組別、特定部分大小、卡路里數及/或個體應作為其飲食的其他食物之組別。基於個體之基因組分布，個人化行為計劃對個體給予2級，且因此個體之推薦飲食計劃應選擇2級下的飲食選擇。

在另一實施例中，可將個別食物分類。舉例而言，給予2級之個體應選擇亦分類為2之特定食物。舉例而言，特定蔬菜、肉、水果、奶品及其他食物可分類為2，而其他並非如此。舉例而言，蘆筍可為等級2之蔬菜，而甜菜為3且因此個體飲食中應包括較多蘆筍而非甜菜。

在另一實施例中，基於基因組分布，將應遵守之飲食類型的建議等級給予個體，該建議等級為個體飲食中應具有

之食物類型的營養素類型之細分類別。該等級可呈包括形狀、顏色、數字及/或字母之視覺表示形式。該等級可呈包括形狀、顏色、數字及/或字母之視覺表示形式。舉例而言，發現個體易患結腸癌及糖尿病，且給予如圖4A所示之表示個體飲食中應具有之推薦食物類型中之不同營養素的比例之符號(亦參見實例3)。諸如(但不限於)特定水果、蔬菜、碳水化合物、肉、奶製品及其類似物之不同類型的食物由諸如圖4B-4D所示之相同方案表示。諸如圖4A所述，具有最接近類似於所給予個體之等級符號之食物應為個體之推薦食物。

在一些實施例中，諸如(但不限於)個體現有飲食、鍛煉及其他個人習慣/活動之因素，視情況諸如家族史、現有生活習慣及地理之其他資訊，諸如(但不限於)工作條件、工作環境、人際關係、居家環境、種族、性別、年齡及其他因素可併入個體基因組分布中以產生個人化行為計劃，且因此影響個體飲食計劃之既定等級。此外，由於已知且併有個體生活習慣變化或更多因素，個體等級可變化。舉例而言，若個體遵守個人化行為計劃上之起始於為極低膽固醇飲食之飲食計劃等級1的推薦活動，則個體可要求併有個體已有之生活習慣變化的更新之個人化行為計劃，以使得個體具有提高之膽固醇含量，該更新之個人化行為計劃可展示個體現在可更適於遵守等級2之飲食計劃，或可自等級1及2之飲食計劃選擇。或者，個體之初始個人化行為計劃可提供個體何時應考慮自等級1移至等級2，或基於

不同等級之不同飲食計劃之間的時程改變其飲食計劃之時間安排，以便使其最健康。

個人化行為計劃之等級可用於不同評級系統之組合。舉例而言，可使用具有等級A至E之鍛煉方案系統及具有等級1至5之飲食計劃系統向個體提供其個人化行為計劃中之A1等級。因此，推薦個體遵守等級A之鍛煉方案及等級1之飲食計劃。或者，對於鍛煉及飲食方案可使用單一評級系統。舉例而言，可對個體給予諸如個人化行為計劃中之等級C之特定等級，以使得個體之推薦鍛煉及飲食方案均在C類別下。在其他實施例中，亦包括諸如其他生活活動及習慣之其他類型的建議。舉例而言，除鍛煉及飲食方案以外，諸如療法、工作環境類型、社會活動類型之其他建議亦可涵蓋在單一評級系統之下。或者，對於其他建議可使用不同評級系統。舉例而言，字母可用於推薦鍛煉方案，數字用於飲食方案且顏色用於醫藥建議。

在一些實施例中，使用二元評級系統，以使得將建議類型成對分組。該系統可類似於邁爾斯-布里格斯類型指標(Myers Briggs Type Indicator, MBTI)系統。在MBTI系統中，存在4對首選項或二分法，且個體被置於各對中之一者中。個體首選項為1)外向或內向；2)感覺或直覺；3)思想或感受；及4)判斷或感知。可使用系統變化確定個體改善其健康及康樂之基於個體基因組分布的建議。

舉例而言，對於飲食而言個體可為A或B，其中A表示某一類型之營養素混合物且B為不同混合物。或者，可將特

定類型之食物分組成A或B。個體可具有用於鍛煉方案之另一二元分類法，諸如H或L，其中H表示個體應參與高衝擊力鍛煉且L表示低衝擊力活動。因而，個體可分類為AH。另一二元分類法可為社交。舉例而言，個體可能在遺傳上傾向於善於社交(S)或不善交際(U)，且因而建議可包括個體應避免或設法降低壓力且增加其健康及康樂之活動類型或人群。

亦可更新個人化行為計劃以包括基於已知之資訊的因素，該資訊包括科學資訊或來自個體之資訊，諸如「現場調用」或直接機制，例如代謝物含量、葡萄糖含量、離子含量(例如鈣、鈉、鉀、鐵)、維生素、血細胞計數、體重指數(BMI)、蛋白質含量、轉錄產物含量、心率等，其可藉由易於獲得之方法測定且當其已知、諸如藉由即時監測得知時，可包括在個人化行為計劃中。可例如基於個體遵守該計劃修改個人化行為計劃，個體遵守該計劃亦可影響個體所具有的一或多種病狀易感性。舉例而言，可更新個體之GCI得分。

社區及激勵

本發明提供基於個體基因組分布之表型分布及個人化行為計劃，以使得個體充分瞭解其健康及康樂，及個體須改善其健康之定製選擇。本文亦提供社區，諸如線上社區，其可向執行個人化行為計劃之個體提供支持及激勵。個體例如藉由遵守個人化行為計劃而改善健康之激勵亦可包括經濟獎勵。

個體可參與社區，諸如線上社區，其中個體或其健康護理管理者獲得個體之基因組分布、表型分布及/或個人化行為計劃。個體可選擇使基因組分布、表型分布及/或個人化行為計劃可用於由社區之全部、社區之子集經由個人線上入口查看或社區均不能查看。朋友、家人或同事可為線上社區之部分。舉例而言，在此項技術中已知諸如 <https://change4fire.com> 之線上社區，其用於激勵個體實現其目標。在本發明中，個體參與或為支持且刺激個體使用表型分布作為基線(諸如 GCI 得分)或藉由實現個人化行為計劃上之目標改善其健康及康樂之線上社區之成員。線上社區可限於個體之朋友、家人或同事或朋友、家人及同事之組合。個體亦可包括先前並不認識之線上社區的其他成員。線上社區亦可為雇主發起之社區。個體可與其他具有類似表型分布、行為計劃之個體形成組且彼此激勵以實現目標。個體可與線上社區之其他個體設定競爭，以提高其 GCI 得分及/或實現個人化行為計劃上之目標。

舉例而言，諸如 GCI 得分及個人化行為計劃之個體報告可由線上社區中個體之家人及朋友查看。個體可具有選擇可查看及/或獲得其報告之人員的選擇或選項。線上版本可包含含有個人化行為計劃上之項目的清單或重要指標，其中個體可標明其個人化行為計劃之成績或進展。GCI 得分可隨遺傳資訊變化而更新且反映在線上報告上。個體亦可輸入可能已改變之因素，諸如生活方式變化、鍛煉方案變化、飲食變化、醫藥治療及其他，其亦可改變個體之報

告。家人及朋友可查看個體之進展以及個體壽命變化及其可能反映或改變個體之風險或易感性之程度。線上入口可允許個體查看初始及後續報告。個體亦可接收來自其朋友及家人之反饋及評論。家人及朋友可留下支持及激勵評論。

線上社區亦可向藉由個人化行為計劃取得進展及/或降低疾病風險或易感性進而改善健康之個體提供獎勵。獎勵亦可提供給不在線上社區中之個體。舉例而言，當個體諸如藉由其個人化行為計劃取得進展，藉此降低疾病風險及/或易感性進而達到某些目標時，雇主發起之線上社區可提供雇主補貼更多、提供額外假期或向個體之健康儲蓄帳戶繳款之健康計劃。或者，社區無須線上，且個體向替雇主處理健康計劃之指定人員呈遞其個人化行為計劃進展及/或疾病易感性降低之證據。

亦可使用其他獎勵激勵個體藉由降低其疾病易感性及/或遵守個人化行為計劃改善其健康。當個體達到某些目標時，其可接收獎勵償還點，諸如疾病風險降低某一百分比或數值，或自一種類別移至另一類別(亦即，較高風險至較低風險)或實現個人化行為計劃中之某些目標。舉例而言，個體可實現某一數值之風險降低以在某一時間段內實現最大疾病風險降低、實現個人化行為計劃上之目標或實現個人化行為計劃上之大多數目標。

或許藉由購買且作為獎賞提供給降低疾病風險或易感性及/或實現個人化行為計劃上之目標的個體，朋友、家人

及/或雇主可提供點及/或獎賞。個體亦可由於在他人(諸如具有相同目標之另一同事，或朋友、家人或線上社區成員之組)之前達到目標而接收點/獎賞。舉例而言，實現某一數值之風險降低、在某一時間段內實現疾病風險之最大降低、實現個人化行為計劃上之目標或實現個人化行為計劃上之大多數目標的第一個人。個體可接收現金或現金償還點作為獎賞。其他獎賞可包括醫藥產品、保健品、健身俱樂部會員資格、水療法、醫學程序、健康監測裝置、遺傳學測試、旅行及其他，諸如本文所述之服務訂購或上述項目之折扣、補貼或補償。

獎勵可由朋友、家人及雇主贊助。醫藥公司、健身俱樂部、醫療裝置公司、水療區及其他亦可贊助獎勵。贊助權可交換做廣告或徵集，例如醫藥公司可能對獲得個體基因組分布資料或臨床試驗感興趣。此外，可使用獎勵來鼓勵個體參與激勵個體改善健康之社區，諸如本文所述之線上社區。

獲取分布及個人化行為計劃

可將含有基因組分布、表型分布及其他與表型及基因組分布相關的諸如個人化行為計劃之資訊之報告提供給個體。諸如看護者、醫師及遺傳學顧問之健康護理管理者及提供者亦可獲取報告。該報告可經列印、保存在電腦上或線上查看。或者，可以紙張形式提供分布及行為計劃。其可呈紙張形式或呈電腦可讀格式(諸如在某一時間線上可讀)，而後續更新由紙張、電腦可讀格式或線上提供。分

布及行為計劃可編碼在電腦可讀媒體上。

可藉由線上入口、經由使用電腦及網際網路網站個體可易於存取之資訊源、電話或其他允許類似獲取資訊之方式獲取基因組分布、表型分布以及個人化行為計劃。線上入口可視情況為安全線上入口或網站。其可提供與其他安全及非安全網站之連接，例如至具有個體表型分布之安全網站，或諸如用於共有特定表型之個體的留言板之非安全網站之連接。

報告可具有個體之GCI得分或GCI附加得分(如本文所述，報告GCI得分亦將涵蓋報告GCI附加得分或兩者之方法)。舉例而言，可使用顯示器觀看一或多種病狀之得分。可使用螢幕(諸如電腦監視器或電視螢幕)觀看顯示器，諸如相關資訊之個人入口。在另一實施例中，顯示器為諸如列印頁之靜態顯示器。顯示器可包括(但不限於)一或多種以下各物：格(諸如1-5、6-10、11-15、16-20、21-25、26-30、31-35、36-40、41-45、46-50、51-55、56-60、61-65、66-70、71-75、76-80、81-85、86-90、91-95、96-100)、彩色或灰度級梯度、溫度計(thermometer)、量表、圓餅圖、直方圖或條形圖。在另一實施例中，使用溫度計顯示GCI得分及疾病/病狀發病率。溫度計可顯示隨所報告之GCI得分變化之程度，例如溫度計可顯示GCI得分增加時之比色變化(諸如，由較低GCI得分之藍色逐漸變為較高GCI得分之紅色)。在相關實施例中，溫度計顯示隨所報告之GCI得分變化之程度與風險等級增加時之比色變

化。

亦可藉由使用聽覺反饋將個體 GCI 得分輸送至個體。舉例而言，聽覺反饋可為風險等級為高或低之語言指示。聽覺反饋亦可為諸如數字、百分位數、範圍、四分位數或與群體之平均或中值 GCI 得分比較之特定 GCI 得分的敘述。在一實施例中，活人親自或經諸如電話(固定電話、行動電話或衛星電話)之電信裝置或經由個人入口輸送聽覺反饋。亦可藉由諸如電腦之自動系統輸送聽覺反饋。可作為互動式語音響應(IVR)系統之部分輸送聽覺反饋，該互動式語音響應系統為允許電腦偵測使用正常電話之語音及按鍵音的技術。個體可經由 IVR 系統與中央伺服器互動。IVR 系統可以預先錄製或動態產生之音訊作出響應以與個體互動且向其提供風險等級之聽覺反饋。個體可播放 IVR 系統應答之數字。視情況輸入鑑別碼、安全碼或進行語音識別方案後，IVR 系統可要求個體選擇諸如按鍵音或語音菜單之菜單中之選項。一個此等選項可向個體提供其風險等級。

個體之 GCI 得分可使用顯示器觀看，且使用聽覺反饋諸如經個人入口輸送。此組合可包括 GCI 得分之視覺顯示及聽覺反饋，其討論 GCI 得分與個體總體健康狀況之相關性及可能的預防措施，諸如其個人化行為計劃。

個體可獲取不同的報告選項。舉例而言，諸如線上入口之線上存取點可允許個體基於其基因組分布顯示單一表型或一種以上表型。訂購者亦可具有不同查看選項，例如

「快速查看」選項，以提供單一或多種病狀之簡明概括。亦可選擇「全面查看」選項，其中提供各類別之較多細節。舉例而言，可能存在關於個體發展表型之可能性的更詳細統計，關於典型症狀或表型之更多資訊(諸如醫學病況之代表症狀)，或諸如身高的身體非醫學病況之範圍，或關於基因及遺傳變異之更多資訊，諸如在例如全世界或不同國家或不同年齡範圍或性別內之群體發生率。舉例而言，許多病狀之估算終生風險之匯總可在「快速查看」選項中，而諸如前列腺癌或克羅恩氏病之特定病狀的更多資訊可為其他查看選項。不同查看選項可存在不同組合及變化。

個體所選擇之表型可為醫學病況，且報告中之不同療法及症狀可連接至含有關於療法之其他資訊的其他網頁。舉例而言，藉由點擊藥物，其將通向含有關於劑量、成本、副作用及有效性之資訊的網站。亦可比較藥物與其他療法。網站亦可含有通向藥物製造商網站之連接。另一連接可向訂購者提供產生醫藥基因組分布之選項，其將包括諸如基於基因組分布，其對藥物之可能反應的資訊。亦可提供至藥物之替代的連接，諸如預防措施(諸如健身及體重減輕)且亦可提供至飲食補充劑、飲食計劃及附近健身俱樂部、醫療診所、健康及安康提供者、日常水療及其類似物之連接。亦可提供教育及資訊視訊、可用療法匯總、可能的補救及一般建議。

線上報告亦可提供連接以安排在場醫師或遺傳諮詢約見

之時間或訪問線上遺傳學顧問或醫師，以便向訂購者提供要求更多關於其表型分布之資訊之機會。線上報告上亦可提供至線上遺傳諮詢及醫師詢問之連接。

在另一實施例中，報告可具有「娛樂」表型，諸如個體基因組分布與諸如艾伯特愛因斯坦(Albert Einstein)之名人的相似性。報告可顯示個體基因組分布與愛因斯坦之基因組分布之間的相似性百分比，且可進一步顯示愛因斯坦及個體之預測IQ。其他資訊可包括一般人群之基因組分布及IQ與該個體及愛因斯坦相比如何。

在另一實施例中，報告可顯示與個體基因組分布相關聯之所有表型。在其他實施例中，報告可僅顯示與個體基因組分布正相關之表型。在其他形式中，個體可選擇顯示表型之某些子組，諸如僅醫學表型，或僅可治療醫學表型。舉例而言，可治療表型及其相關基因型可包括克羅恩氏病(與IL23R及CARD 15相關)、第1型糖尿病(與HLA-DR/DQ相關)、狼瘡(與HLA-DRB1相關)、牛皮癬(HLA-C)、多發性硬化症(HLA-DQA1)、格雷氏症(HLA-DRB1)、類風濕性關節炎(HLA-DRB1)、第2型糖尿病(TCF7L2)、乳癌(BRCA2)、結腸癌(APC)、間歇性記憶(KIBRA)及骨質疏鬆症(COL1A1)。個體亦可選擇在其報告中顯示表型之子類，諸如對於醫學病況而言僅顯示發炎疾病，或對於非醫學病況而言僅顯示身體性狀。在一些實施例中，個體可選擇顯示所有病狀，藉由突出彼等病狀、僅突出具有高風險之病狀或僅突出具有低風險之病狀計算個體之估算風險。

個體呈遞且傳達給個體之資訊可為安全且機密的，且獲取該等資訊可由個體控制。來源於複合基因組分布之資訊可作為管理機構批准、可以理解、醫學上相關及/或高影響力資料供應給個體。資訊亦可廣泛關注，而非醫學上相關。資訊可藉由包括(但不限於)入口介面及/或郵遞之數種方式安全地傳達給個體。更佳地，藉由個體安全且機密存取之入口介面將資訊可靠地(倘若如此，則由個體選擇)提供給個體。較佳地藉由線上、網際網路網站存取，或在替代形式中藉由電話或允許私人、安全且易於獲得存取之其他方式提供該種介面。藉由經網路傳輸資料將基因組分布、表型分布及報告提供給個體或其健康護理管理者。

因此，可產生報告之一代表性實例邏輯裝置可包含諸如圖 5(500)所示之電腦系統(或數位裝置)。電腦系統可接收且儲存基因組分布、分析基因型相關性、基於基因型相關性之分析產生規則、運用該等規則至基因組分布且產生表型分布、個人化行為計劃及報告。舉例而言，可自電腦系統獲得及輸出個人化行為計劃。電腦系統 500 可理解為可自媒體 511 及/或網路埠 505 讀取指示之邏輯設備，網路埠 505 可視情況連接至具有固定媒體 512 之伺服器 509。諸如圖 5 所示之系統可包括 CPU 501、碟片驅動器 503、諸如鍵盤 515 之及/或滑鼠 516 之可選輸入裝置及可選監視器 507。可經由指定通信媒體至局部或遠端位置之伺服器實現資料通信。通信媒體可包括傳輸及/或接收資料之任何方式。舉例而言，通信媒體可為網路連接、無線連接或網際網路

連接。該連接可經全球資訊網提供通信。預見可經該等網路或連接傳輸與本發明有關之資料以便由接收方522接收及/或檢查。接收方522可為(但不限於)個體、健康護理提供者或健康護理管理者。在一實施例中，電腦可讀媒體包括適於傳輸生物樣本或基因型相關性分析結果之媒體。媒體可包括關於個體之表型分布及/或個體之行為計劃的結果，其中該結果係使用本文所述之方法獲得。

個人入口可充當接收及評估基因組資料之個體的主要介面。入口可使得個體能夠跟蹤其樣本自收集至測試及結果之進展。個體經由入口存取，基於其基因組分布引入常見遺傳疾病之相對風險。個體可選擇經由入口將何種規則應用於其基因組分布。

在一實施例中，一或多個網頁將具有表型之列表及鄰接於各表型之訂購者可選擇以便包括在其表型分布中的框。表型可與關於該表型之資訊相連接，以幫助訂購者進行關於其希望包括在表型分布中之表型之知情選擇。網頁亦可具有根據例如可治療或不可治療疾病之疾病組編制之表型。舉例而言，個體可僅選擇可治療表型，諸如HLA-DQA1及乳糜瀉。訂購者亦可選擇顯示表型之症狀前或症狀後療法。舉例而言，個體可選擇症狀前療法可治療之表型(在增加之篩選範圍之外部)，對於乳糜瀉而言，為無麩質飲食之症狀前療法。另一實例可為對於阿爾茨海默氏症，選擇斯達汀(statin)、鍛煉、維生素及精神活動之症狀前療法。血栓症為另一實例，其症狀前療法為避免口服避

孕藥且避免長時間靜坐。具有批准之症狀後療法之表型之一實例為濕性AMD，其與CFH相關，其中個體可獲得用於其病狀之雷射療法。

亦可根據疾病或病狀之類型或種類編制表型，例如神經病學、心血管、內分泌、免疫學等。表型亦可分組為醫學及非醫學表型。網頁上表型之其他分組可根據身體性狀、生理學性狀、精神性狀或情緒性狀進行。網頁可進一步提供藉由選擇一個框選擇一組表型之部分。舉例而言，選擇所有表型、僅醫學上相關表型、僅非醫學上相關表型、僅可治療表型、僅不可治療表型、不同疾病組或「娛樂」表型。「娛樂」表型可包括與名人或其他著名個體或其他動物或甚至其他生物體相比較。可用於比較之基因組分布的列表亦可提供在網頁上以便由個體選擇進而與個體基因組分布相比較。

線上入口亦可提供搜尋引擎以幫助個體到達入口、搜尋特定表型或搜尋其表型分布或報告所揭示之特定術語或資訊。入口亦可提供獲取夥伴服務及供應產品之連接。亦可提供至用於具有共同或類似表型之個體的支持組、留言板及閒聊室之其他連接。線上入口亦可提供至具有更多關於個體表型分布中之表型之資訊之其他網站的連接。線上入口亦可提供允許個體與朋友、家人、同事或健康護理管理者共享表型分布及報告之服務，且可選擇在表型分布中展示其希望與朋友、家人、同事或健康護理管理者共享的表型種類。

表型分布及報告向個體提供個人化基因型相關性。使用基因型相關性產生個人化行為計劃，該計劃向個體提供決定其個人健康護理及生活方式選擇的增加之知識及機會。若在遺傳變異與具有可用療法之疾病之間發現強相關性，則偵測遺傳變異可幫助決定開始疾病療法及/或監測個體。在存在統計學上顯著之相關性，但並不視為強相關性之情況下，個體可與個人醫師一起回顧資訊且決定有利的適當行為過程。考慮到特定基因型相關性而有益於個體之潛在行為過程包括投與治療性處理、監測潛在治療需要或治療之作用或使生活方式在飲食、鍛煉及其他個人習慣/活動方面改變，其可基於個體之基因組分布個人化為個人化行為計劃。諸如現有習慣及活動之其他個人資訊亦可併入個人化行為計劃中。舉例而言，諸如乳糜瀉之可治療表型可具有無麩質飲食且可提供於個人化行為計劃中之症狀前療法。同樣，基因型相關性資訊可應用於醫藥基因組學以預測個體對用特定藥物或藥物方案治療將具有之可能反應，諸如特定藥物治療之可能功效或安全性。

基因型相關性資訊亦可與遺傳諮詢協作使用以建議夫妻考慮生育，及母親、父親及/或子女之潛在遺傳問題。遺傳學顧問可向個體提供關於顯示高特定病狀或疾病風險之表型分布之資訊及支持。其可解釋關於病症之資訊、分析遺傳模式及復發風險且評論訂購者之可用選項。遺傳學顧問亦可提供支持性諮詢以指定訂購者去社區或政府支持服務。遺傳諮詢可包括有特定訂購計劃。遺傳諮詢選項亦可

包括彼等在請求24小時內預訂且在諸如傍晚、星期六、星期日及/或假日之非慣例時間期間可用的選項。

個體入口亦可促進輸送除初始篩選外的附加信息。可告知個體關於與其個人遺傳分布有關之新科學發現，諸如關於其當前或潛在病狀之新療法或預防策略之資訊。新發現亦可輸送給其健康護理管理者。該新發現可併入更新或修正個人行為計劃中。個體或其健康護理提供者可藉由電子郵件被告知關於個體表型分布中之表型之新基因型相關性及新研究。舉例而言，「娛樂」表型之電子郵件可發送給個體，例如電子郵件可告知其基因組分布與亞伯拉罕林肯 (Abraham Lincoln) 77%一致且其他資訊可經由線上入口獲得。

本文亦提供用於通知訂購者新相關性或修正相關性、新規則或修正規則及新報告或修正報告之電腦代碼，例如關於通知新預防及安康資訊、關於正開發之新療法的資訊或新可用療法的代碼。本發明亦提供用於產生新規則、修改規則、組合規則、定期更新規則、設定新規則、安全地維持基因組分布之資料庫、應用規則至基因組分布以確定表型分布、產生個人化行為計劃及報告之電腦代碼系統，包括授予具有不同訂購之個體不同程度之存取及選項之電腦代碼。

訂購

可諸如藉由電腦產生人類或非人類個體之基因組分布、表型分布及報告，包括個人化行為計劃。舉例而言，個體

可包括其他哺乳動物，諸如牛類、馬類、羊類、犬類或貓類。個體可為個人之寵物，且寵物之主人可能需要個人行為計劃以增加其寵物之健康及壽命。個體或其健康護理管理者可為訂購者。如本文所述，訂購者為藉由購買或支付一或多種服務來訂購服務之人類個體。服務可包括(但不限於)一或多種以下：確定其或諸如訂購者之子女或寵物之另一個體的基因組分布，獲得表型分布，更新表型分布及基於其基因組及表型分布獲得包括個人化行為計劃之報告。

訂購者可選擇提供基因組及表型分布或報告給其健康護理管理者，諸如醫師或遺傳學顧問。藉由訂購者列印出一個待給予健康護理管理者之複本，或經由線上入口(諸如經由線上報告上之連接)將其直接發送給健康護理管理者，健康護理管理者可直接獲取基因組及表型分布。

可產生訂購者及非訂購者之基因組分布且數位儲存在諸如電腦可讀媒體上，但獲取表型分布及報告(諸如經由電腦輸出)可限於訂購者。舉例而言，將藉由電腦產生及輸出之至少一個GCI得分的獲取權提供給訂購者，而非非訂購者。在另一變化形式中，訂購者與非訂購者均可經電腦獲取其基因型及表型分布，但非訂購者之獲取有限或所產生報告有限，而訂購者可完全獲取且可產生完整報告。在另一實施例中，訂購者與非訂購者最初可完全獲取，或具有完整初始報告，但僅訂購者可獲取基於其儲存基因組分布之更新報告。舉例而言，將獲取權提供給非訂購者，其

中其對至少一個其GCI得分獲取受限，或其可具有關於所產生之至少一個其GCI得分之初始報告，但僅購買訂購情況下產生更新報告。諸如看護者、醫師及遺傳學顧問之健康護理管理者及提供者亦可獲取至少一個個體之GCI得分。

其他訂購模型可包括提供表型分布之模型，其中訂購者可選擇應用所有現有規則至其基因組分布或應用現有規則之子集至其基因組分布。舉例而言，其可選擇僅應用可治療疾病表型的規則。訂購可具有種類，以使得單一訂購種類內存在不同級別。舉例而言，不同級別可取決於訂購者希望與其基因組分布相關聯之表型的數目，或可獲取表型分布之人數。

另一訂購級別可將個體特異性因素(諸如已知表型，諸如年齡、性別或病史)併入其表型分布。另一基本訂購級別可允許個體產生至少一個疾病或病狀GCI得分。若歸因於用以產生至少一個GCI得分之分析之變化，使至少一個GCI得分存在任何變化，則此級別變化可進一步允許個體指定自動更新至少一個待產生之疾病或病狀GCI得分。在一些實施例中，可藉由電子郵件、語音訊息、文字訊息、郵遞或傳真通知個體自動更新。

訂購者亦可產生具有表型分布以及關於表型之資訊(諸如關於表型之遺傳學及醫學資訊)的報告。個體可獲取之不同資訊量可取決於其具有之訂購級別。舉例而言，個體可具有之不同查看選項可取決於其訂購級別，諸如對於非

訂購者或較基本訂購而言為快速查看，但彼等具有完全訂購者可獲取全面查看。

舉例而言，不同訂購級別可具有資訊可獲取性之不同變化或組合，包括(但不限於)群體中表型之發病率、關於相關性所用之遺傳變異、引起表型之分子機制、表型之療法、表型之治療選項及預防措施，其可包括於報告中。在其他實施例中，報告亦可包括諸如個體基因型與諸如名人或其他著名人士之其他個體之間的相似性之資訊。關於相似性之資訊可為(但不限於)同源性百分比、相同變異體之數目及可能類似之表型。此等報告可進一步含有至少一個GCI得分。

若線上獲取報告，則基於訂購級別之其他選項可包括至具有關於表型之其他資訊之其他網站的連接，至具有相同表型或一或多種類似表型之個人之線上支持組及留言板的連接，至線上遺傳學顧問或醫師的連接，或安排與遺傳學顧問或醫師之電話或在場約見時間的連接。若報告呈紙張形式，則資訊可為上述連接之網址或電話號碼及遺傳學顧問或醫師之地址。訂購者亦可選擇哪些包括在其表型分布中及哪些資訊包括在其報告中。諸如看護者、醫師、精神病學家、心理學家、治療學家或遺傳學顧問之個體健康護理管理者或提供者亦可獲取表型分布及報告。訂購者可能夠選擇該個體之健康護理管理者或提供者是否可獲取表型分布及報告或其部分。

另一訂購級別可為產生初始表型分布及報告後數位維持

個體之基因組分布，且向訂購者提供自最近研究產生具有更新相關性之表型分布及報告之機會。訂購者可具有自最近研究產生具有更新相關性之風險分布及報告之機會。由於研究揭示基因型與表型、疾病或病狀之間的新相關性，因此將基於此等新相關性發展新規則且其可應用於已儲存且維持之基因組分布。新規則可使先前未與任何表型相關之基因型相關聯，使基因型與新表型相關聯，改進現有相關性，或基於基因型與疾病或病狀之間的新發現之相關性提供調整GCI得分之基礎。可經由電子郵件或其他電子方式告知訂購者新相關性，且若關注表型，則其可選擇以該新相關性更新其表型分布。訂購者可選擇其在指定時間內(例如三個月、六個月或一年)償付各次更新、許多次更新或無限次更新之訂購。另一訂購級別可為無論何時基於新相關性產生新規則，訂購者均可自動更新其表型分布或風險分布，而非個體選擇何時更新其表型分布或風險分布。

訂購者亦可向非訂購者介紹產生關於表型與基因型之間的相關性之規則、確定個體之基因組分布、應用規則至基因組分布且產生個體之表型分布之服務。訂購者推薦可給予訂購者低價服務訂購或更新其現有訂購。經推薦個體可自由獲取有限次數或具有折扣訂購價格。

以下實例說明且解釋本文所述之實施例。本發明之範疇不受該等實例限制。

實例

實例1：GCI得分之間的比較

基於多種模型在HapMap CEU群體中計算與T2D相關之10種SNP的GCI得分。相關SNP為rs7754840、rs4506565、rs7756992、rs10811661、rs12804210、rs8050136、rs1111875、rs4402960、rs5215、rs1801282。對於此等SNP中之每一者而言，在文獻中報告三種可能基因型之優勢率。CEU群體由三十個母親-父親-子女三人組組成。使用來自此群體之60位父母以避免相關性。排除10種SNP之一中均無檢出的個體之一，產生59個個體組。接著使用數種不同模型計算各個體之GCI等級。

不同模型產生此資料組之高度相關結果。計算各對模型之間的斯皮爾曼相關性(Spearman correlation)(表2)，其展示乘法及加法模型具有0.97之相關性係數，且因此使用加法或乘法模型時GCI得分穩固。類似地，哈佛大學改進得分與乘法模型之間的相關性為0.83，且哈佛大學得分與加法模型之間的相關性係數為0.7。然而，使用最大優勢率作為遺傳學得分產生由一種SNP定義之二分得分。此等結果總體指示提供使模型相關性最低之穩固框架的得分分級。

表2：模型對之間的CEU資料之得分分布之斯皮爾曼相關性

	乘法	加法	Harv-Het	Harv-Hom	MAX OR
乘法	1	0.97	0.83	0.83	0.42
加法	0.97	1.	0.7	0.7	0.6
Harv-Het	0.83	0.7	1	1	0
Harv-Hom	0.83	0.7	1	1	0
MAX OR	0.42	0.6	0	0	1

量測T2D發病率變化對所得分布之影響。發病率值在

0.001至0.512之間變化。對於T2D之情況，觀察到不同發病率值產生相同個體順序(斯皮爾曼相關性 >0.99)，因此可假定0.01之人工固定發病率值。

實例2：估算GCI

使用 WTCCC 資料 (*Wellcome Trust Case Control Consortium, Nature. 447:661-678 (2007)*) 測試 GCI 框架。此資料組含有分成8個群體之約14,000個個體之基因型。8個群體由7個攜帶7種不同疾病之病例的群體及1個對照群體組成。使用 Affymetrix 500k 基因晶片將所有個體基因分型。對於7種不同疾病中之3種第2型糖尿病、克羅恩氏病及類風濕性關節炎而言，在 Affymetrix 500k 基因晶片上搜尋通過校對標準設定之 SNP 中的與初始公開 SNP 具有 $r^2=1$ 之 SNP。對於第2型糖尿病而言發現8種 SNP、對於克羅恩氏病而言發現9種 SNP，且對於類風濕性關節炎而言發現5種 SNP。

使用受試者工作曲線(ROC)(*The Statistical Evaluation of Medical Tests for Classification and Prediction, MS Pepe. Oxford Statistical Science Series, Oxford University Press (2003)*) 估算 GCI 作為病狀之分類器測試的能力。較佳地，將存在臨限值 t ，使得若個體之 GCI 得分大於 t ，則個體必然為病例，且若個體之 GCI 得分小於 t ，則個體必然為對照。在如上所述之三種病例對照組中計算所有個體之 GCI 得分。接著基於由 GCI 得分臨限值定義的二元測試，將真陽性率作為假陽性率之函數繪圖。最終，計算所得圖之曲

線下面積(AUC)。對於隨機診斷測試而言，AUC為0.5，且對於理想測試而言，AUC為1。

為具有用以比較之基線，使用計算平衡SNP之間的相互作用以擬合數據之最佳模型之邏輯回歸。若SNP為 s_1 、 s_2 、...、 s_n ，則模型假定邏輯值為 $X=a_1s_1+a_2s_2+...+a_ns_n+a_{12}s_{12}+...+a_{n-1,n}s_{n-1,n}$ ，其中 s_{ij} 為 s_i 與 s_j 之間的相互作用。經擬合機率用作風險之估算值且產生此等風險估算值之ROC曲線。應注意，此模型考慮SNP之間的成對相互作用，且其因此應至少與GCI得分一般精確。

所有三種疾病之GCI及邏輯回歸之AUC十分相似(表2)，得出以下結論：SNP-SNP相互作用並未增加風險評估之實質資訊，至少對於此等疾病及此等SNP而言未增加。因此，只要不存在來自先前研究之關於該種相互作用的證據，即可證明SNP-SNP相互作用可忽視之假定。

將GCI ROC曲線與理論疾病模型相比較。此疾病模型假定疾病受環境與遺傳因素兩者影響，且兩種因素獨立。 $P=G+E$ ，其中 G 為遺傳風險且 E 為環境風險。第一個模型假定 $G \sim N(0, \sigma_G)$ 且 $E \sim N(0, \sigma_E)$ ，且若對於固定值 α 而言 $P > \alpha$ ，則個體壽命期間將發展病狀。使用遺傳率為 $\sigma_G/(\sigma_G+\sigma_E)$ 且平均終生風險為 $\Pr(P > \alpha)$ 之約束固定 σ_G 、 σ_E 及 α 。由於各測試病狀之遺傳率及平均終生風險已知，因此可根據疾病設定模型之參數。基於此模型自分布 P 產生100000個隨機樣本。接著假定各個體之 G 已知(但 E 未知，且因此疾病狀態未知)且基於 G 產生ROC曲線。此代表完全瞭解遺傳風險且

可量測所有個體之遺傳風險的最佳情況。

亦產生 $G=\lambda X+Y$ 之此模型的變化，其中 $Y\sim N(0, \sigma_Y)$ 且 $X\sim B(2,p)$ 。在此情況下， X 對應於一種具有較大影響之 SNP 且 Y 對應於許多其他較小遺傳影響。藉由適當地設定參數 λ 、 σ_Y 及 p ，可控制較大影響 SNP 之相對風險。此等相對風險對於風險-風險基因型而言為 4，且對於異型合子而言設定為 2。

如自表 3 及圖 1-3 可見，邏輯回歸及 GCI 之 AUC 十分接近，且其邊界均遠離隨機測試。然而，理論最佳情況顯然可比當前估算值提供更多資訊。基於此等圖，當前科學知識能夠以可有教益方式估算個體之疾病風險，GCI 之 AUC 比未使用資訊之隨機測試高 20-40% 可作為證據。

表 3：不同 ROC 曲線之曲線下面積 (AUC)

疾病	遺傳率	平均終生風險	最佳情況 AUC	GCI AUC	邏輯回歸 AUC
第2型糖尿病	64%	25%	0.902	0.596750	0.603873
克羅恩氏病	80%	0.56%	0.982	0.654024	0.646273
類風濕性關節炎	53%	1.54%	0.944	0.674906	0.688608

實例 3：個人化行為計劃

自唾液樣本獲得基因組分布且以 GCI 得分產生表型分布。報告亦包括具有如表 4 所示之建議之個人化行為計劃。

表 4：個人化行為計劃

	改進鍛煉 (Mod. Exercise)	減輕 體重	阿司匹林 (Aspirin)	8小時 睡眠	考慮 斯達汀	增加 纖維	避免 麩質	抗氧化 劑及 $\omega 3$ (Anti ox. & Omega3)
心肌梗塞	Y	Y	Y	Y	Y	Y	N	Y
乳糜瀉	N	N	N	N	N	N	Y	N
結腸癌	Y	Y	Y	N	N	Y	N	Y
糖尿病	Y	Y	N	N	Y	Y	N	N
肥胖症	Y	Y	N	Y	N	Y	N	N
阿爾茨海默氏症	Y	N	Y	Y	Y	N	N	Y

儘管本文中已展示且描述本發明之較佳實施例，但對於熟習此項技術者而言該等實施例顯然僅以實例方式提供。在不悖離本發明之情況下，熟習此項技術者現將可作出多種變更、變化及取代。應瞭解，可使用本文中所述之本發明之實施例之各種替代形式實施該等實施例。希望以下申請專利範圍界定本發明之範疇且藉此涵蓋此等實施例之範疇內的方法及結構及其等效物。

【圖式簡單說明】

圖 1 為克羅恩氏病 (Crohn's Disease) 之受試者工作特徵 (Receiver Operating Characteristic, ROC) 曲線之圖。底線對應於隨機預期且頂線對應於遺傳變異已知時之理論預期。第一條中線對應於 GCI，而第二條中線對應於藉由邏輯回歸獲得之曲線；

圖 2 為第 2 型糖尿病之 ROC 曲線之圖。底線對應於隨機預期且頂線對應於遺傳變異已知時之理論預期。第一條中線對應於 GCI，而第二條中線對應於藉由邏輯回歸獲得之曲

線；

圖3為類風濕性關節炎之ROC曲線之圖。底線對應於隨機預期且頂線對應於遺傳變異已知時之理論預期。第一條中線對應於GCI，而第二條中線對應於藉由邏輯回歸獲得之曲線；

圖4表示個體的基於基因組分布之評級系統。A)表示易患結腸癌及糖尿病之個體的個人化行為計劃之食物評級；B)表示使用此評級系統無小麵包之普通漢堡；C)表示使用此評級系統之椰菜；且D)表示使用此評級系統之蘋果；及

圖5表示用於分析且經網路傳輸基因組及表型分布及個人化行為計劃之系統的示意圖。

【主要元件符號說明】

500	電腦系統
501	CPU
503	碟片驅動器
505	網路埠
507	監視器
509	伺服器
511	媒體
512	固定媒體
515	鍵盤
516	滑鼠
522	接收方

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※申請案號：98126864

※申請日：98.8.10

※IPC 分類：~~C12Q, G06F~~

一、發明名稱：(中文/英文)

G06F 19/00 (2011.01)

用於個人化行為計劃之方法及系統

METHODS AND SYSTEMS FOR PERSONALIZED ACTION PLANS

二、中文發明摘要：

本發明提供基於個體之基因組分布，用於個人行為計劃之方法及系統。方法包括評估個體基因型與至少一種疾病或病狀之間的相關性及提供用於個人行為計劃之評級系統。本文亦揭示激勵及鼓勵個人改善其健康及康樂之獎勵。

三、英文發明摘要：

The present disclosure provides methods and systems for personal action plans based on an individual's genomic profile. Methods include assessing the association between an individual's genotype and at least one disease or condition and providing rating systems for an individual's action plan. Incentives to motivate and encourage people to improve their health and well-being are also disclosed herein.

八、圖式：

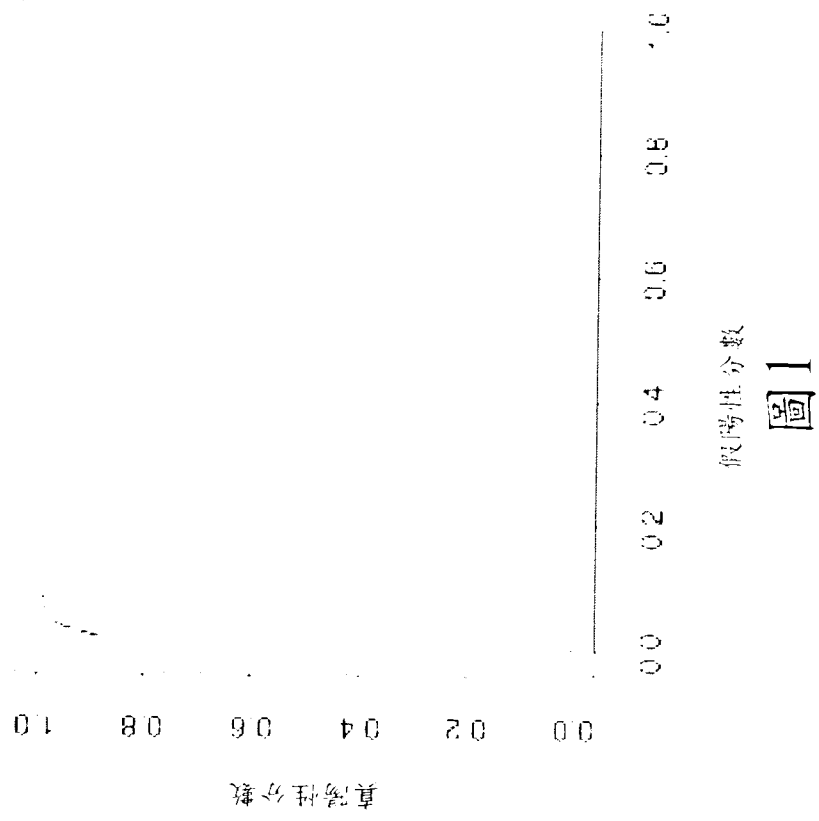


圖 1

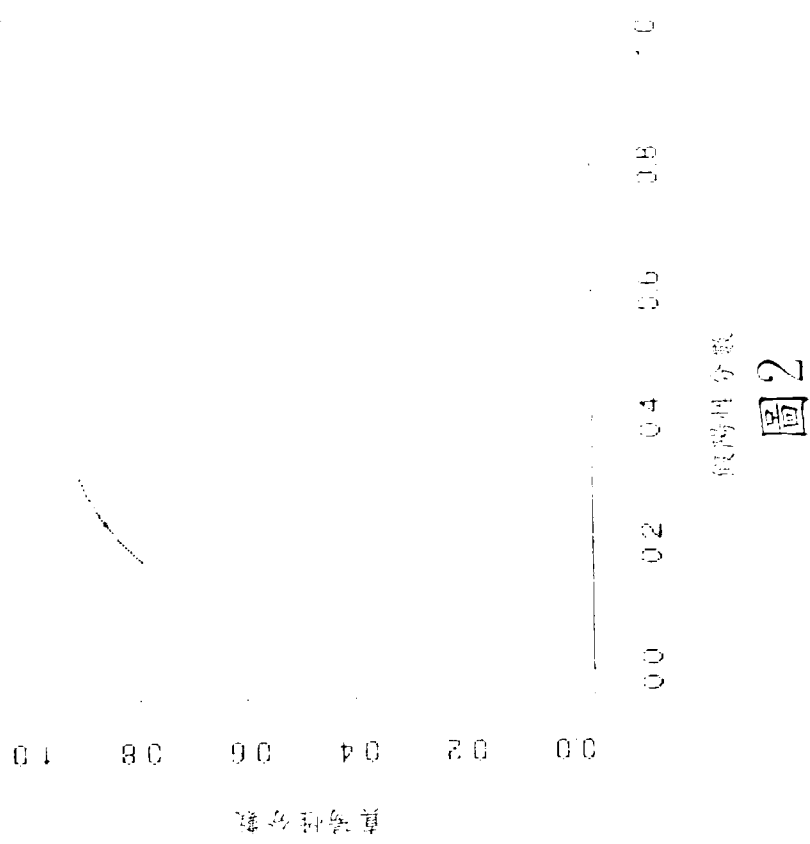


圖2

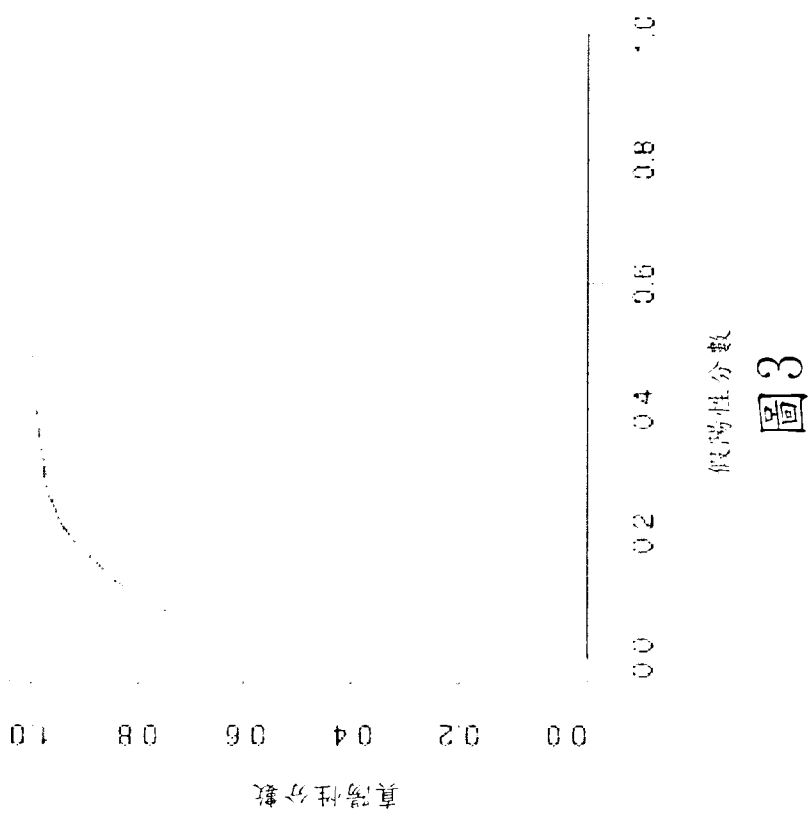


圖3

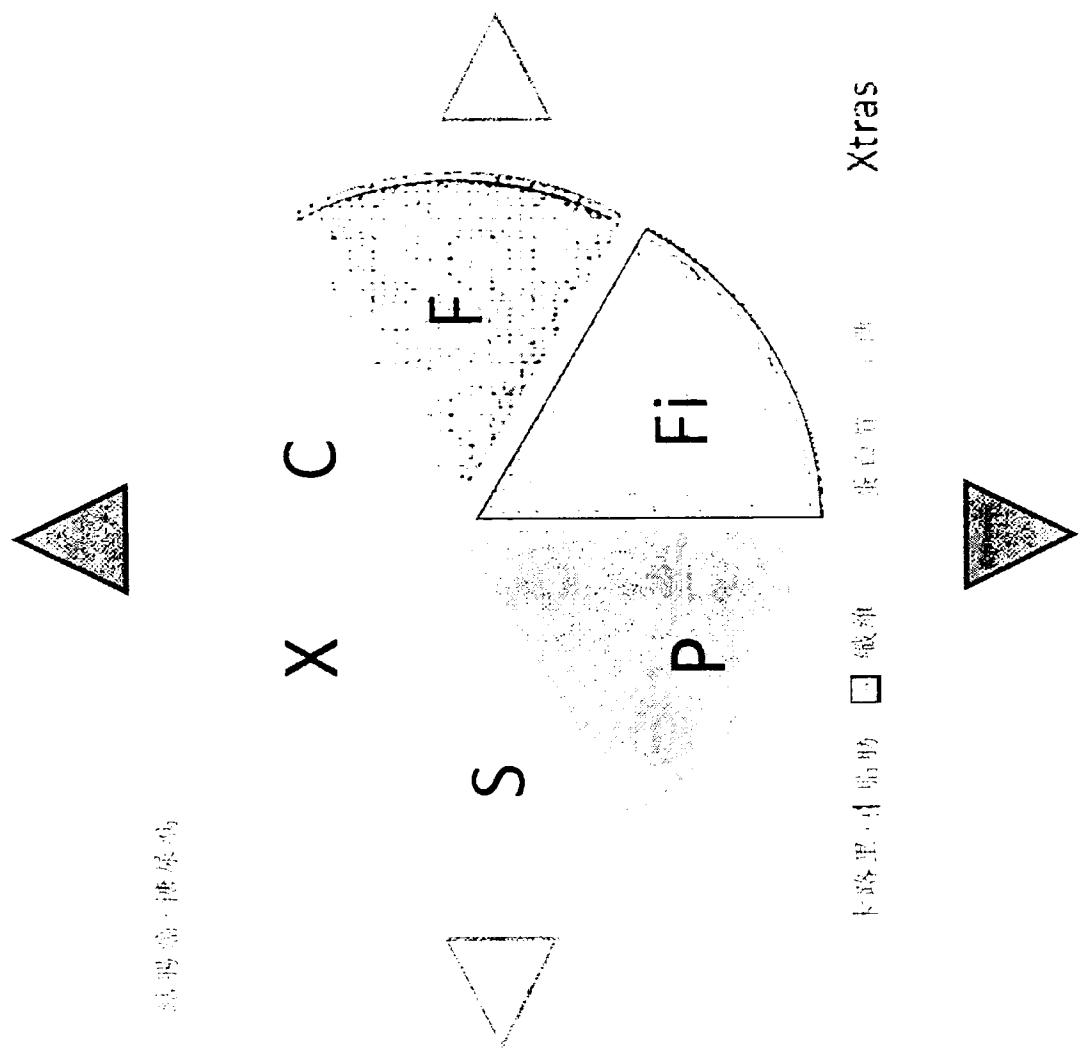
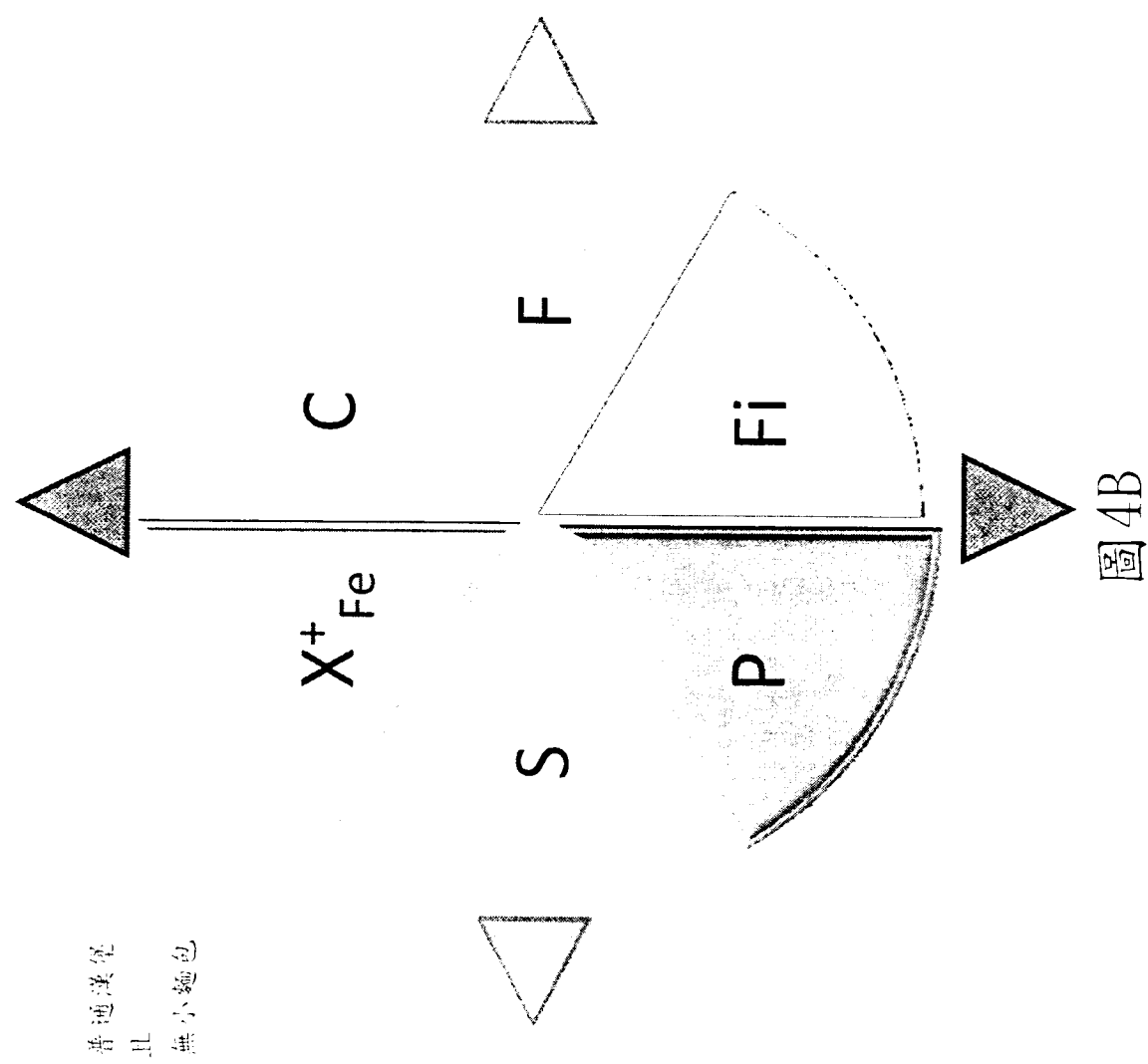
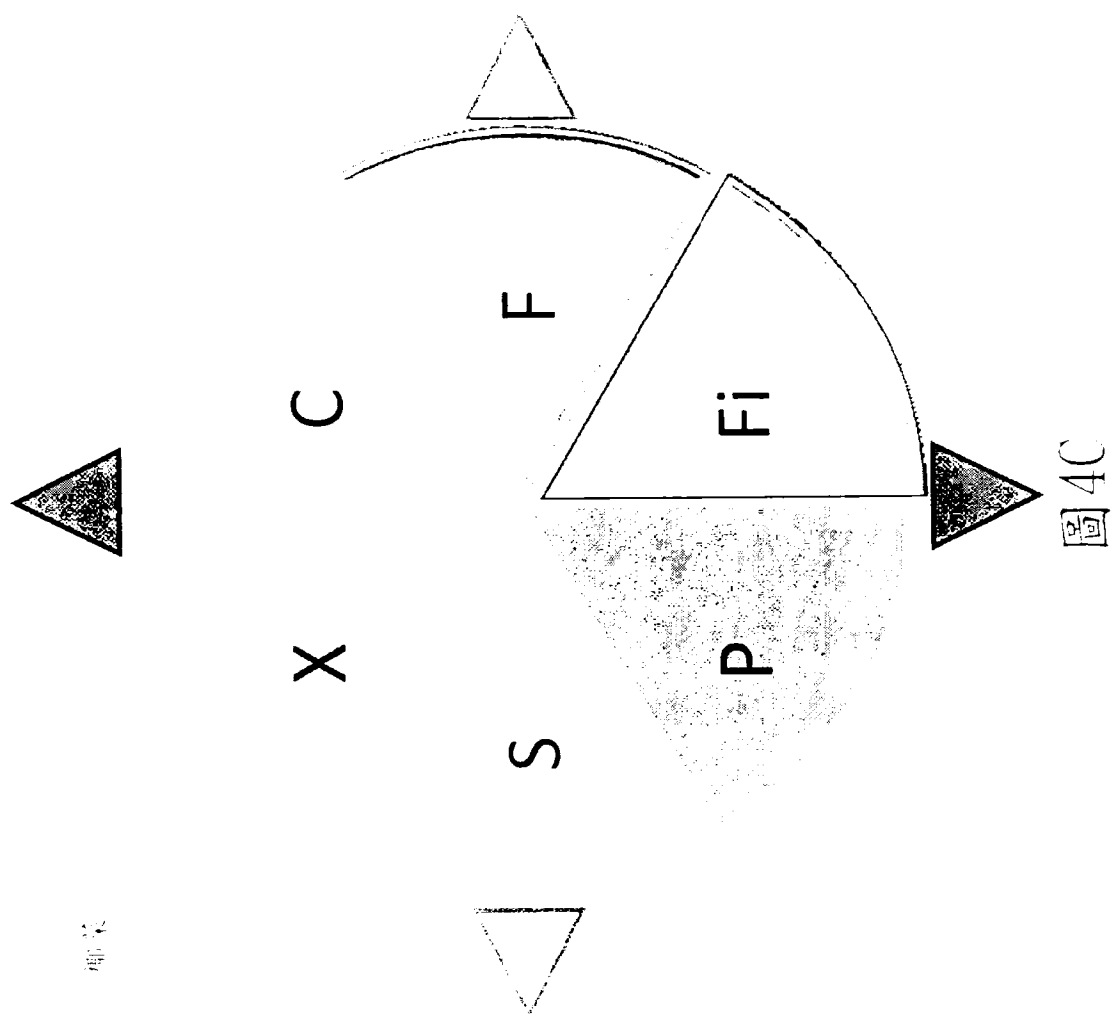
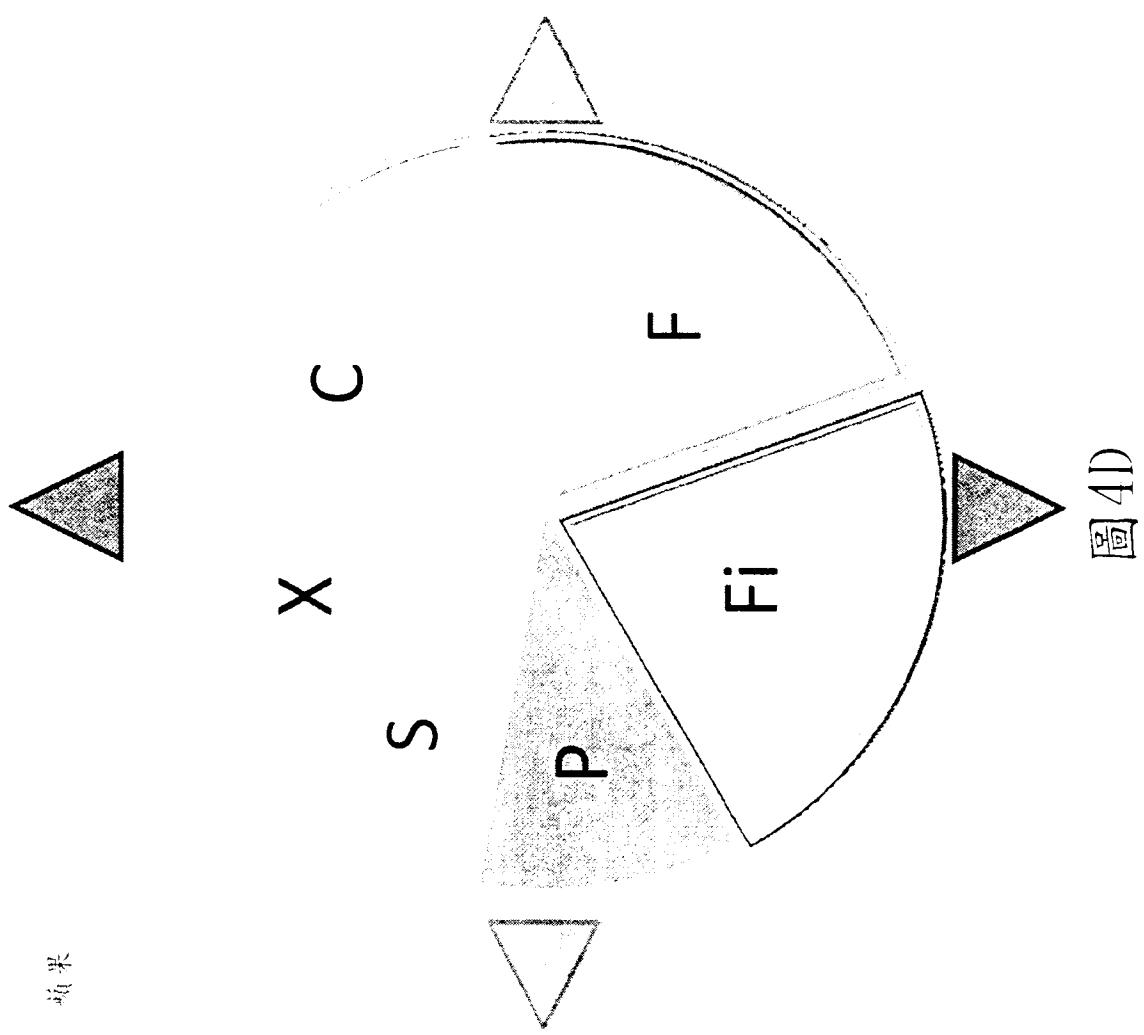


圖 4A







碩果

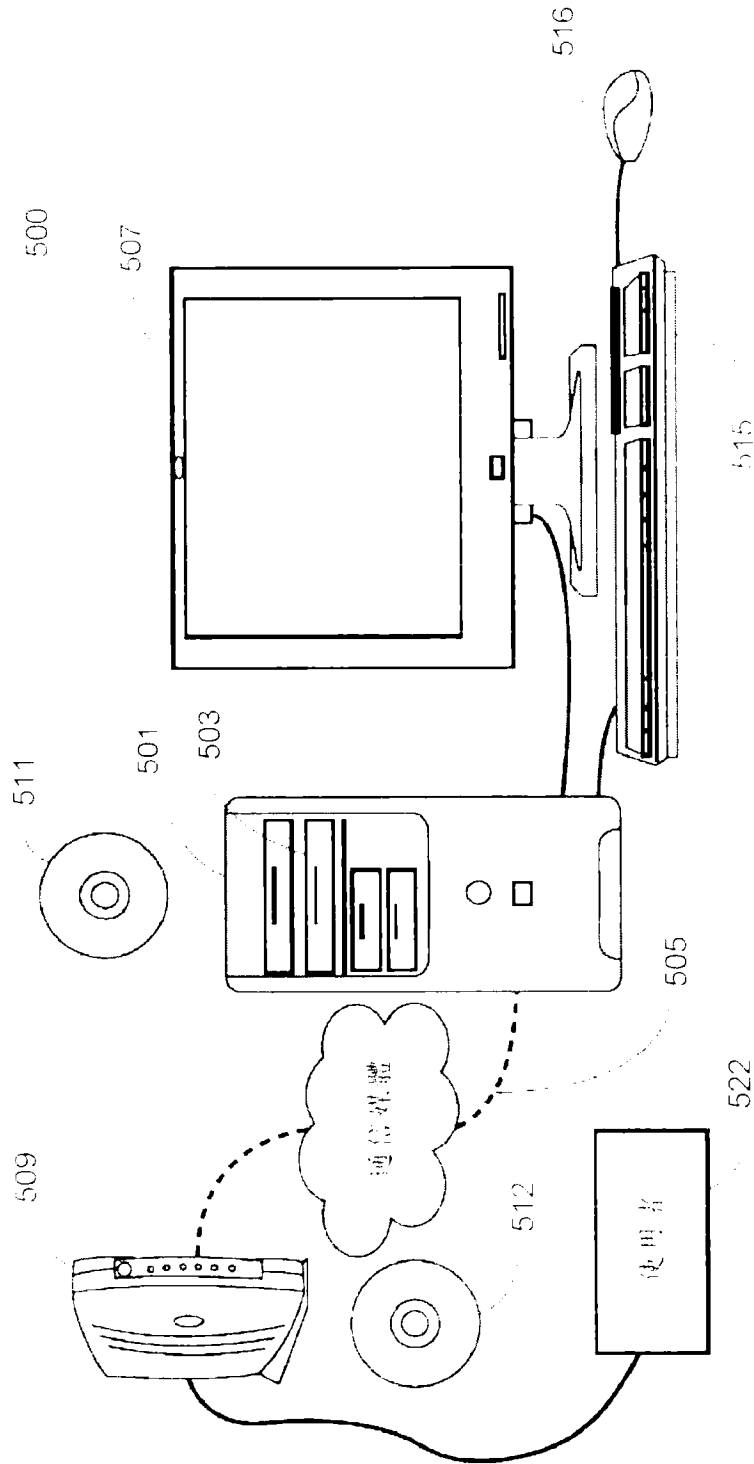


圖5

四、指定代表圖：

(一)本案指定代表圖為：第(4A-4D)圖。

(二)本代表圖之元件符號簡單說明：

(無 元件符號說明)

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

(無)

七、申請專利範圍：

1. 一種用於個人化行為計劃中之多種建議的評級系統，其包含至少一種評級及至少一種獎勵：

其中對該等建議中之每一者給予等級；

其中該等等級中之每一者對應於給予個體之等級；

其中給予該個體之該等級係由電腦基於該個體之基因組分布確定，

其中獎勵與各個建議相關聯。

2. 如請求項1之評級系統，

其中給予該個體之該等級係由電腦基於該個體之遺傳複合指數(GCI)或GCI附加得分確定。

3. 如請求項1或2之評級系統，其中該等級為數字、顏色、字母或其組合。

4. 如請求項1或2之評級系統，其中建議包括醫藥建議。

5. 如請求項1或2之評級系統，其中該非醫藥建議為鍛煉方案。

6. 如請求項1或2之評級系統，其中該非醫藥建議為飲食計劃。

7. 如請求項1或2之評級系統，其中該評級系統係由二元系統代表。

8. 如請求項1或2之評級系統，其中該基因組分布係藉由使用高密度DNA微陣列或基於PCR之方法獲得。

9. 如請求項1或2之評級系統，其中該基因組分布係藉由自該個體擴增遺傳樣本獲得。

10. 一種向個體提供個人化行為計劃中之建議之等級的方法，其包含：
 - (a) 獲得該個體之基因組分布；
 - (b) 確定該個體之至少一個等級，其中該等級係由電腦基於該基因組分布確定；
 - (c) 將自該電腦輸出之該等級報告給該個體或該個體之健康護理管理者；及
 - (d) 將至少一種建議與獎勵相關聯。
11. 一種向個體提供個人化行為計劃中之建議之等級的方法，其包含：
 - (a) 使用電腦產生該個體之GCI或GCI附加得分；
 - (b) 確定該個體之至少一個等級，其中該等級係由該電腦基於該GCI或GCI附加得分確定；
 - (c) 將自該電腦輸出之該等級報告給該個體或該個體之健康護理管理者；及
 - (d) 將至少一種建議與獎勵相關聯。
12. 如請求項10或11之方法，其中該等級係由顏色、字母或數字表示。
13. 如請求項10或11之方法，其中該等級對應於個人化行為計劃上之建議。
14. 如請求項10或11之方法，其中建議包括醫藥建議。
15. 如請求項10或11之方法，其中該非醫藥建議為鍛煉活動。
16. 如請求項10或11之方法，其中該非醫藥建議為飲食計

劃。

17. 如請求項10或11之方法，其中該非醫藥建議為營養素。
18. 如請求項10或11之方法，其中該等級係基於二元系統。
19. 如請求項10之方法，其中該基因組分布係藉由使用高密度DNA微陣列或基於PCR之方法獲得。
20. 如請求項10之方法，其中該基因組分布係藉由自該個體擴增遺傳樣本獲得。