

明 細 書

発明の名称：

生体情報処理方法および装置、記録媒体並びにプログラム

技術分野

[0001] 本技術は、生体情報処理方法および装置、記録媒体並びにプログラムに関する、特に将来のヒト生体状態の変化を予想し制御する生体情報処理方法および装置、記録媒体並びにプログラムに関する。

背景技術

[0002] 新しい治療法、診断法、予防法の開発の生産性の低下は、健康に関する社会の問題を解決する上で大きな問題となっている。この生産性の低下は、基礎生物学で得られた膨大な知見が、臨床上の問題を解決するのに有効に利用できないことが原因である(非特許文献1)。最新の国際調査によると、世界の現在の糖尿病人口は2億8500万人に増加した。この糖尿病人口は、2030年までに北米の人口よりも多い4億3500万人を超えると予測されている。これは世界の成人の糖尿病有病率が7%に近づいていることを意味している。

[0003] ヒトは60兆の細胞からなり、1つ1つの細胞には60億塩基対からなるDNA(Deoxyribonucleic acid)情報が格納される複雑さを持ち、さらに生涯で 10^{16} 回の細胞分裂を行うという動的なシステムである。独立した2人の間には、数百万の塩基置換に加えて、DNA配列の欠損、重複、逆位というDNA構造上の多様性が観察される。

[0004] 生物学の特性の1つに階層性がある。人体はDNAやタンパク質などの分子から構成される「細胞」を基盤とし、さらに様々な細胞から構成されるサブリシステム、組織、臓器によっても特徴づけられる。この階層性は環境からの刺激に対する応答反応の時間スケールを多様化する。細胞内の情報伝達はミリ秒や秒単位で進み、細胞間の情報伝達は分や時間のスケールで進み、細胞増殖や分化のプロセスは日時の単位で進行する。このような時間スケールの多様性は、生体に時間の遅れを生じさせる。

- [0005] このような複雑で多面的な生命現象の中で、基礎生物学は生物を機械的なシステムとして表象することを目的に、生体を構成する分子間の因果関係を明らかにしてきた。これは、多数のパラメータの中から特定の1つの分子パラメータのみを選択して変化させるという手法により実施されてきた。
- [0006] 一方の臨床科学では、疾患に関連する症状の緩和に焦点をあてて治療してきた。
- [0007] また、臨床上の効果を見るために、投薬などの介入と臨床的転帰の間の規則や関係式といった経験則を、特定の時間における生物学的指標から推定していくことが実施されている。
- [0008] 現代医療における重要な課題は、各人に適した健康管理や治療の提供にある。統計学的に安全性、有効性が実証されている治療法や医薬も、一旦上市されて大規模の集団に利用されると、一部の患者に対して副作用や無効が観察されるようになる。これは治療を受けた患者の遺伝的多様性を1つの原因としている。遺伝的多様性による疾患発症率の差異に関しては、最近「ゲノムワイド関連解析(Genome wide association study; GWAS)」により分析が進められている(非特許文献2)。

先行技術文献

非特許文献

- [0009] 非特許文献1 : Nature 453; 840-9, 2008
- [0010] 非特許文献2 : J. Hum. Genet. Doi : 10.1038/jhg.2010.19

発明の概要

発明が解決しようとする課題

- [0011] しかし、現在の臨床では、限られた生物学的指標を限定された時間でのみ測定し、それを統計学的に平均化するために、集団を構成する多様な人々が時間とともに獲得する多様な特性が結果的に捨象されている。加えて基礎生物学がミクロスケールの問題に焦点を当てるのとは対照的に、臨床研究では、身体的な障害、生理的機能の消失、特定の疾患や身体状態など個体レベル

のマクロスケールの問題を臨床的転帰として取り扱っている。このような、基礎生物学と臨床研究において、階層間のギャップが存在する。

[0012] また、現在の治療戦略は疾患を時間とともに獲得される多様化を捨象して組み立てざるを得ないため、いずれの治療戦略も対症療法となる。しかし疾患は症状を緩和することで元の健康状態に戻るわけではない。

[0013] これらの問題は現在の基礎生物学や臨床の概念では、将来のヒト生体状態の変化を予想し制御することができず、有効な予防法、治療法を開発することが困難であることを意味している。

[0014] 本技術はこのような状況に鑑みてなされたものであり、将来のヒト生体状態の変化を予想し制御することができるようにするものである。

課題を解決するための手段

[0015] 本技術の一側面は、生体の分子の発現量を特定の時間間隔で測定する測定ステップと、測定された時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割する分割ステップと、前記ベースライン成分の変動、または前記周期的成分の振幅若しくは周期の変動から、前記時系列データの定常領域を同定し、さらに同定した前記定常領域間の因果関係を同定する同定ステップと、同定された前記定常領域間の因果関係から、前記生体の状態変化を推定する推定ステップとを含む生体情報処理方法である。

[0016] 前記同定ステップは、さらに生体の状態変化と外部環境との関係を同定し、前記推定ステップは、さらに前記同定された生体の状態変化と外部環境との関係から、前記生体の疾患発症を推定することができる。

[0017] 前記測定ステップでは、全身状態、局所状態、および染色体状態を表象する前記分子を測定することができる。

[0018] 前記分子は、前記生体の血中の分子とすることができる。

[0019] 前記分子は、メタボリックシンドロームに関連する分子とすることができる。

[0020] 前記分子は、培地中の分子とすることができる。

[0021] 本技術の一側面においては、生体の分子の発現量が特定の時間間隔で測定

され、測定された時系列データが、周期的成分、環境刺激応答成分、およびベースライン成分に分割され、ベースライン成分の変動、または周期的成分の振幅若しくは周期の変動から、時系列データの定常領域が同定されるとともに、さらに同定した定常領域間の因果関係が同定される。同定された定常領域間の因果関係から、生体の状態変化が推定される。

[0022] 本技術の一側面の生体情報処理装置、記録媒体およびプログラムは、上述した本技術の一側面の生体情報処理方法に対応する生体情報処理装置、記録媒体およびプログラムである。

[0023] 本技術の他の側面は、対象者の生体の生体分子の細胞記憶情報を取得する細胞記憶情報取得ステップと、前記対象者の生体の環境条件に関する環境情報を取得する環境情報取得ステップと、前記対象者の生体を構成する遺伝子配列に関する遺伝子情報を取得する遺伝子情報取得ステップと、取得された前記細胞記憶情報、前記環境情報、および前記遺伝子情報に関し、類似する前記対象者以外の生体の情報を検索する検索ステップとを含む生体情報処理方法である。

[0024] 前記遺伝子情報は、一塩基多型および構造多型とすることができる。

[0025] 前記検索は、前記細胞記憶情報、前記環境情報、および前記遺伝子情報の相関関係における類似性に基づいて行うことができる。

[0026] 前記細胞記憶情報取得ステップは、前記対象者の生体の分子の発現量を特定の時間間隔で測定する測定ステップと、測定された時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割する分割ステップと、前記ベースライン成分の変動、または前記周期的成分の振幅若しくは周期の変動から、前記時系列データの定常領域を同定し、さらに同定した前記定常領域間の因果関係を同定する同定ステップとを含むことができる。

[0027] 本技術の他の側面においては、対象者の生体の生体分子の細胞記憶情報が取得され、対象者の生体の環境条件に関する環境情報が取得され、対象者の生体を構成する遺伝子配列に関する遺伝子情報が取得され、取得された細胞記憶情報、環境情報、および遺伝子情報に関し、類似する対象者以外の生体

の情報が検索される。

[0028] 本技術の他の側面の生体情報処理装置、記録媒体およびプログラムは、上述した本技術の他の側面の生体情報処理方法に対応する生体情報処理装置、記録媒体およびプログラムである。

発明の効果

[0029] 以上のように、本技術の一側面および他の側面によれば、将来のヒト生体状態の変化を予想し制御することができる。

図面の簡単な説明

[0030] [図1]生体情報処理装置の一実施形態の構成を示すブロック図である。

[図2]健康状態予測制御処理を説明するフローチャートである。

[図3]定常領域をノードとした因果関係に関するモデルを示す図である。

[図4]生体情報処理装置の他の実施形態の構成を示すブロック図である。

[図5]培養細胞の細胞状態の予測制御処理説明するフローチャートである。

[図6]細胞のモデルの構成を示す図である。

[図7]細胞の状態変化を示す図である。

[図8]情報の因果関係を説明する図である。

[図9]生体情報処理装置の他の実施形態の構成を示すブロック図である。

[図10]他の健康状態予測制御処理を説明するフローチャートである。

[図11]生活習慣の因果関係に関するモデルを示す図である。

[図12]細胞記憶の履歴を表す行列を示す図である。

発明を実施するための最良の形態

[0031] 以下、本技術を実施するための形態（以下、実施形態と称する）について説明する。なお、説明は以下の順序で行う。

1. 経過
2. 定義
3. 第1の実施形態
4. 第2の実施形態
5. 第3の実施形態

- 6. 第4の実施形態
- 7. 第5の実施形態
- 8. 第6の実施形態
- 9. 第7の実施形態
- 10. 第8の実施形態
- 11. 第9の実施形態

[0032] <経過>

[0033] 最初に本技術の経過について説明すると次のようになる。

[0034] 本技術の開発者は、ヒト生体の時間発展を表象し、その未来の変化の方向を予想し制御するための方法を鋭意検討した。その結果、従来から生命を定式化するのに利用されてきた遺伝子型、表現型、外部環境の3つのパラメータに加えて、「時間発展する細胞記憶」という新しいパラメータを導入することにより、時間発展するヒト生体の問題を解決できることが見出された。また本技術の開発者らは「時間発展する細胞記憶」を定式化するために、各細胞で発現している双安定スイッチの特性を持った転写因子の細胞記憶が、エピジェネティクス修飾とジェネティックス修飾により時間とともに変更される点に注目した。エピジェネティクス修飾は、DNAのメチル化やヒストン修飾からなり、ジェネティックス修飾はDNAの変異や構造変化からなる。

[0035] 本技術の開発者らは、ジェネティックス修飾ならびにエピジェネティクス修飾が、双安定スイッチにより制御される標的遺伝子産物の発現量を改変することで細胞状態を変化させるという「時間発展する細胞記憶モデル」を構築した。さらに本技術の開発者らは、「時間発展する細胞記憶モデル」に従い、染色体上で働く転写因子、ジェネティックス修飾、エピジェネティックス修飾からなる「染色体状態」という概念を構築した。

[0036] 次に本技術の開発者らは、この「染色体状態」を全身状態あるいは局所状態の観測値から推定するために、ヒト生体システムを構成する3つの異なる階層である全身状態、局所状態、および「染色体状態」を連結させた。このため、本技術の開発者らは、全身状態、局所状態、および「染色体状態」の

各階層に対応するマクロモデル、メソモデル、およびミクロモデルを構築した。さらに本技術の開発者は、この「生体階層連結モデル」により、マクロモデル、メソモデル、およびミクロモデルを細胞単位で連結し定式化する方法を提案する。

[0037] 次に本技術の開発者は、ヒト生体の時間発展を担う「染色体状態」を観測するために、「生体階層連結モデル」を時系列モデルに統合する方法の開発を検討し、「生体状態空間モデル」を用いた新しい統合方法を見出した。そして、本技術の開発者は、その「生体状態空間モデル」を用いて全身状態を反映した分子マーカーの時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割することで、「染色体状態」を抽出する方法を提案する。

[0038] 次に本技術の開発者は、本技術によりはじめて表象することが可能となったヒト生体の時間発展的变化のデータを利用した制御モデルの開発を鋭意検討した。そして本技術の開発者は、最初の外部環境からの入力、その後の外部環境からの入力に対する反応性を変えることを意味する「構築された変化」という概念を見出し、「時間発展する細胞記憶」を「構築された変化」により制御するという「動的構築モデル」を提案する。

[0039] また本技術の開発者は、「動的構築モデル」を臨床上の問題に応用するために鋭意検討を行い、生体で産生される分子の発現量に関する時系列データを、「生体状態空間モデル」を用いて定常領域を同定するための「生体局所定常領域モデル」の方法を提案した。さらに本技術の開発者は、生体で産生される分子の発現量の時系列変化の中に見出される生体局所定常領域をノードとして利用した生体状態追跡と因果構造を分析する方法を提案する。

[0040] <定義>

[0041] 次に、本技術の実施形態の説明に用いる用語の定義について説明する。

[0042] ヒト生体の時間発展とは、ヒト生体の状態や機能が時間とともに不可逆に変化するプロセスを指す。ヒトは受精の瞬間から発生、誕生、成長、老化という不可逆的な変化を経て死亡する。疾患の発症も潜在的な変化を受けた発

症の前段階、発症、特定の生理機能障害、特定の生理機能消失、身体障害、死へと展開する。これはヒト生体におけるある瞬間と別の瞬間の時間が同質ではないことを意味する。

[0043] 「時間発展する細胞記憶」とは、転写因子により構築された細胞記憶が、細胞への環境入力や偶然の結果として導入されるジェネティックス修飾ならびにエピジェネティックス修飾により変更されることを示す。従来、細胞記憶は外界からの刺激により獲得された細胞の状態変化が、外界からの刺激が消失した後も維持される現象として定義されてきた(Cell 140: 13-18, 2010)。細胞状態とは細胞に発現している分子の種類、量、修飾特性により表象され、細胞状態の維持ならびに変化は転写因子により司られている。

[0044] すなわち従来の細胞記憶は、細胞の状態が転写因子の持つ双安定スイッチの特性により特定の均衡状態を与えるという概念は与えたものの、スイッチのオン・オフという均一な時間のみしか与えることができなかった。しかし実際には時間とともに導入されるジェネティックス修飾やエピジェネティックス修飾が、転写因子により与えられる細胞状態を時間とともに変化させることで、細胞の状態を転写因子のみで想定した細胞状態よりもはるかに多様化させる。

[0045] すなわち「時間発展する細胞記憶」においては、ジェネティックス修飾やエピジェネティックス修飾という継承可能な履歴により細胞状態の多様性を定式化することが可能となり、「時間発展する細胞記憶」は、時間の不均一性を取り扱えるようになった点で従来の細胞記憶とは本質的に異なっている。

[0046] 従来の細胞記憶の分子的な基盤をなすのが転写因子である。転写因子の回路が非線形で双安定の特性を示すとき、分子の状態はオンあるいはオフという均衡状態をとり、その記憶は細胞分裂を経ても継承される。転写因子とプロモータの結合部位の間に形成される親和性、複数の転写因子間の協調性あるいは多量体化は、転写因子に非線形性の特性を与える。

[0047] この非線形性は、転写因子の反応に閾値様の特性を与えることで、一過的

な外乱に抵抗して発現量を継続させる。これに加えて正のフィードバック、あるいは正または負の2重フィードバックによって十分に大きなヒル係数を獲得した転写因子は、双安定スイッチの特性を獲得し、もともとの入力がなくなったあとも状態変化がロックされる。すなわち、従来の細胞記憶の骨格をなすのが、細胞で発現している転写因子の種類、量、翻訳後修飾である。

[0048] 「時間発展する細胞記憶」の分子的な基盤を担うのは、ヒトゲノムで観察される一塩基多型、DNA配列の一部の欠損、重複、コピー数変化などのジェネティックス修飾とDNAのメチル化、ヒストン修飾、タンパク質の変性などのエピジェネティックス修飾である(Nature Review Genetics 7: 85-97, 2006, Cell 128: 655-658, 2007)。

[0049] ジェネティックス修飾、エピジェネティックス修飾は、それぞれ直接的あるいは間接的に、転写因子の非線形性や双安定スイッチの特性に関連するパラメータを変化させることで、従来の細胞記憶の機能を修飾する。

[0050] ジェネティックス修飾に関しては、一塩基多型と疾患発症率の関係が全ゲノム連鎖解析法によって解析されている。その解析法を用いて2型糖尿病の関連遺伝子として19の遺伝子が同定されている。しかし、この変異を持つ糖尿病患者は全体の1%に過ぎない(Nature 462: 307-314, 2009)。同様に乳がん患者全体で観察されるBRCA1/2遺伝子の変異は3%に過ぎない。一卵性双生児と二卵性双生児の寿命を比較した試験では、遺伝子配列が寿命に貢献する割合は、疫学調査から15%乃至25%程度であることが示されている(Hum. Genet 97: 319-323, 1996)。

[0051] エピジェネティックス修飾とは、DNAの配列変化を伴わない継承可能な遺伝子機能の改変と定義される。DNAのメチル化とクロマチンタンパク質の化学修飾がエピジェネティックスの分子的な実体である。それに加えてプリオン蛋白やアミロイド蛋白などタンパク質の変性による細胞機能の改変も広義のエピジェネティックスと分類される。エピジェネティックスという概念は、発生のプロセスで遺伝情報がどのように表現型へと変化されるかを説明するために、ワディントンにより提唱された。

- [0052] しかしエピジェネティックスの機能は、組織や細胞特異的な遺伝子発現の確立や維持にだけ働いているわけではなく、栄養状態、社会的ストレス、化学物質による刺激などがエピジェネティックス修飾を導入する。ヒトの疫学的研究から、胎児や新生児の間に受けた環境からの刺激が、エピジェネティックスを介して成人後の慢性疾患の発症率に影響を及ぼすことが報告されている(Stem Cell Res. 4; 157-164, 2010)。
- [0053] このようなエピジェネティックス修飾は、細胞分化に関連する発生型エピジェネティックスと区別するために、環境型エピジェネティックスと呼ばれる。本技術の開発者は、環境型エピジェネティックスの機能は、「外部環境情報変化に基づきDNA情報を更新する働き」と捉えられることを見出した。
- [0054] 転写因子の機能の変更に働くジェネティックス修飾とエピジェネティックス修飾は、それぞれ継承される時間のスケールが異なっている。ジェネティックス修飾は両親から子供へと継承される細胞記憶の主因である。一方、エピジェネティックス修飾は、ある世代の中で獲得される細胞記憶として中心的な役割を担っている。
- [0055] しかしトランスジェネレーション・エピジェネティックスの発見は、生殖系列に導入された環境型エピジェネティックスが発生の初期化に抵抗して、両親から子供へと継承されることが示されている(Stem Cell Res. 4; 157-164, 2010)。また体細胞に導入されたDNAの変異や転位などの新たなジェネティックス修飾は、ガン形成の重要な要因と考えられている。
- [0056] 「時間発展する細胞記憶のモデル」とは、ジェネティックス修飾、エピジェネティックス修飾により経時的に変化する「時間発展する細胞記憶」を定式化するために新たに導入した概念である。時間とともに変化する転写因子の発現特性、ジェネティックス修飾、エピジェネティックス修飾は「染色体状態」として統合される。
- [0057] これは従来の細胞記憶が1つの転写因子に対してオン／オフという2つの状態のみを定義しているのとは異なり、本技術の染色体状態では、転写因子による標的遺伝子産物に対する誘導ならびに抑制作用が離散的に変化する。

染色体状態の概念を導入することで、細胞状態は染色体状態と細胞の外部環境状態の2つの時系列変化の連結として定式化される。さらに「時間発展する細胞記憶」は「染色体状態」の時系列変化の関数として定式化される。

[0058] 「生体階層連結モデル」とは、「時間発展する細胞記憶」を個人レベルの健康や疾患の予想に利用するために求められる「染色体状態」と「個体状態」の連結のために新たに提案されるモデルである。ヒト個体は階層的に秩序立てられた連続体であり、分子から細胞が構成され、細胞の上に組織や器官などのさらに複雑なものが構築されている。このような階層構造に基づき人体の状態は全身状態、局所状態、および染色体状態に分類可能であり、各状態に対応してマクロ、メソ、ミクロの3つの異なるスケールのモデルを構築することができる。

[0059] マクロモデルとは、全身を1つの状態として捉えるモデルである。従って全身に拡散する分子の発現特性によりその状態を規定することができる。全身に拡散する分子の代表例は、内分泌系や免疫系のホルモン、増殖因子、サイトカインである。これに加えて自律神経系の制御に働くアドレナリン、ノルアドレナリンも全身の状態を規定する。

[0060] メソモデルは、生体内の同質な局所空間の状態を捉えるモデルである。各臓器や組織は1つのまとまりとして同質の状態を形成している。炎症症状などは一般に局所で特異的に起こる。これは炎症部位を形成する多数の細胞から構成される。一方組織の細胞新生は少数の組織幹細胞の増殖と分化により制御されている。従って、メソモデルとは1つの細胞、炎症組織、1つの組織、臓器のように、異なる数や種類の細胞から構成される。この局所状態は局所環境内で発現しているオートクリン、パラクリン分子群に加えて全身性の分子により規定される。

[0061] ミクロモデルは、1つ1つの細胞内で構成される染色体の状態を捉えるモデルである。ミクロモデルを細胞状態としない理由は、細胞状態は転写因子により駆動される顕在的な細胞記憶しか表象できず、ジェネティックス修飾やエピジェネティックス修飾からなる潜在的な変化が捨象されるからである。

。染色体状態はジェネティックス修飾、エピジェネティックス修飾特性、ならびに転写因子の発現特性により規定される。

[0062] 染色体の状態の中でエピジェネティックス修飾と転写因子の発現特性は、染色体を持つ細胞の局所環境と全身性の環境により制御される。全身性の分子を産生する細胞も、全身状態に加えて細胞の局所状態ならびに染色体状態により制御され、局所環境の分子であるオートクリン、パラクリン因子を産生する細胞は、全身状態、その細胞の局所環境と染色体状態により制御される。すなわち、分析により「染色体状態」、局所状態、全身状態が、互いの変化をと共に決定し、ともに方向付け、共に時間発展することが定式化されることが明らかになった。この定式化を「生体階層連結モデル」と称する。

[0063] 人は社会的な環境、本人の行動特性などに依存して、様々な異なる環境からの刺激を受ける。これら外界からの刺激は、体細胞の様々なシステムにより認識され、人体の内部環境変化へと翻訳される。そのために、刺激を受容する体細胞システムの個人差により、同じ環境からの刺激は必ずしも同じ内部環境の変化を誘導しない。このことは環境要因の定量測定だけでは人体の状態変化は正確に予想できないことを意味している。環境からの刺激はその特性から、物理的ストレス（気温、酸素、紫外線など）、化学ストレス（内分泌かく乱物質、発がん物質など）、社会心理的ストレス、運動ストレス、栄養ストレス（過食、飢餓）、感染ストレス、傷害ストレスなどに分類することができる。

[0064] 環境からの刺激は、受容システムにより感知された後に、異なる時間遅れをもって全身状態、局所状態、「染色体状態」の変化へと翻訳される。従って「時間発展する細胞記憶」を全身レベルの記憶に展開するためには、時間を連続分布として捉えるのではなく、離散分布として捉えることが必要である。時間遅れの不安定性は、生体内ではフィードバックシステムにより制御される。従って、あるフィードバックシステムから新しいフィードバックシステムへの変化が捉えられる程度の時間単位で状態変化を捉えられれば、時間遅れの問題は捨象することができる。

- [0065] 「生体状態空間モデル」とは、「生体階層連結モデル」を時系列モデルに応用するために、新たに提案されるモデルである。時系列モデルで用いられる様々なモデルは、一般に状態空間モデルにより統一的に取り扱うことができる（時系列解析入門 北川源四郎 岩波書店 第9章、11章、12章）。また時系列解析の多くの問題が状態空間モデルの状態推定の問題として定式化できる。
- [0066] 状態空間モデルは、システムモデル（ x ）と、観測モデル（ y ）の2つのサブモデルから構成され、2通りの解釈が可能となる。観測モデルを、観測された時系列データ y_n が観測される仕組みを表現する回帰モデルと考えると、システムモデル（ x ）の状態 x_n はその回帰定数となる。この場合システムモデルは、回帰係数の時間的变化の様子を表現するモデルとなる。一方、状態 x_n を推定すべき信号と考えると、システムモデルは信号の発生メカニズムを表すモデル、観測モデルはその信号を実際に観測するときに信号が変換されノイズが加わる様子を表象している。
- [0067] 「動的構築モデル」において、 x_n が実際の人体の状態を表現し、 y_n が測定可能な細胞外因子群の発現量に関する多変量の時系列ベクトルと仮定し、状態 x_n を推定すべき信号と考える。「生体階層連結モデル」を臨床予想に応用する場合、すべての「染色体状態」と局所状態を測定することは現実的ではない。しかし、ホルモン、増殖因子、サイトカインなど全身を循環する細胞外因子は測定可能である。
- [0068] 「生体階層連結モデル」は、「染色体状態」、局所状態、全身状態が、互いの変化をともに決定し、ともに方向付け、実質的に共時間発展していることを明らかにした。これは、全身を循環する細胞外因子の時系列変化のデータの中に「染色体状態」、局所状態の変化が含有されていることを意味する。すなわち、人体状態 x_n の全身性の血中因子からなる多変数の時系列データである y_n には、局所状態ならびに「染色体状態」の成分が直接的あるいは間接的に反映されている。この定式化法を「生体状態空間モデル」と称する。

[0069] 全身を循環する細胞外因子の時系列データ y_n は、季節調節モデルを用いて周期的成分に、多重線形モデルを用いて環境刺激応答成分に、多項式平滑化スプラインモデルを用いてベースラインの成分に、それぞれ分解することが可能である。細胞記憶は生体で産生される分子の発現に関する周期的成分の振幅と周波数、刺激応答成分の最大発現量、長期的なベースラインの変化に反映される。従って、「生体階層連結モデル」と「生体状態空間モデル」を組み合わせることで、「時間発展する細胞記憶」を具体的にかつ時系列で表象することが可能となる

[0070] 「動的構築モデル」は、時間発展するヒト生体に対する制御理論である。生体の時間発展の機能の1つは、最初の環境からの刺激が、その後の同様な種類の刺激に対する反応性を変えることにある。その反応性変化の分子実体を、本明細書では「構築された変化」と称する。この定式化により時間発展するヒト生体を制御することが可能となる。

[0071] 「構築された変化」の中で、「生体にとって致死的でないストレスが加わった場合に、次の強いストレスへの応答性が改善する」という現象が、一般にホルミーシスと呼ばれる (Toxicology and Applied Pharmacology 222: 122-128, 2007)。現在までのところホルミーシスの分子メカニズムは明らかにされていない。本技術の「時間発展する細胞記憶モデル」は、ホルミーシスをはじめとする「構築された変化」が、ジェネティックス修飾とエピジェネティックス修飾による動的な細胞記憶が分子基盤を担っていることを明らかにした。これにより、慢性疾患の発症は、本技術の「時間発展する細胞記憶」がその後のストレスとミスマッチすることを原因とする、と捉えることが可能となる。

[0072] 「構築された変化」は、動的平衡モデルや動的非平衡モデルという従来の制御モデルが遺伝子決定論による固定された遺伝的アルゴリズムを前提としているのとは異なる。すなわち「構築された変化」は、遺伝子の機能が環境からの入力や偶然の要因から変化し、その変化の一部が未来に継承されることで遺伝的に規定されたアルゴリズムが変化することを前提としている。従

って、「構築された変化」に対しては新しい制御理論が必要となる。

[0073] 遺伝子決定論では、表現型の時間変化 ($P(t)$) は、一定の遺伝子型 (G_x) と、時間とともに変化する環境要因 ($E(t)$) の積として、次のように表現される。

$$P(t) = G_x \times E(t) \quad (1)$$

[0074] 一方の「構築された変化」に対しては、表現型の時間変化 ($P(t)$) は、本技術の「時間発展する細胞記憶」を組み込んだ時間とともに変化する染色体状態 ($C(t)$) と、時間とともに変化する環境要因 ($E(t)$) の積として、次のように表現される。

$$P(t) = C(t) \times E(t) \quad (2)$$

式 (2) で表される生物モデルを「動的構築モデル」と称する。

[0075] 「生体局所定常モデル」とは、本技術の「動的構築モデル」を用いた、健康ならびに疾患の制御のための新たなデータ処理方法である。生体で産生される分子の発現量に関する時系列データは一般に非定常である。ただし、非定常であっても、時間区間を適当な小区間に分割することで、各小区間では定常と仮定することができる。このようなシステムの機能とは独立に計算により時間区間を適当な小区間に分割し、各小区間を定常と仮定することができる時系列のモデルを、一般に局所定常モデルという（時系列解析入門 北川源四郎 岩波書店 第8章）。

[0076] 「生体局所定常モデル」では細胞記憶の状態変化をそれぞれ別の定常状態として割り振るために、生体内の細胞や培養細胞で発現する分子の発現量変化の時系列データを、まず「生体状態空間モデル」を用いて周期的成分、環境刺激応答成分、およびベースラインの傾向成分に分解した後に、局所定常領域が同定される。細胞記憶は、ベースライン成分の変動、周期的成分の振幅ならびに周期、環境刺激応答成分の最大発現量ならびに最小発現量に影響を与える。この中で環境からの入力強度と独立して同定できるのは、ベースライン成分の変動と周期的成分の振幅ならびに周期の3つのカテゴリである。

[0077] 「生体局所定常モデル」では、この3つのカテゴリより局所定常領域を同定する。局所定常領域は一般に、元の定常状態、新しい定常状態、元の定常状態から新しい定常状態への変化の3つに分類できる。本技術の「生体局所定常モデル」は、時系列変化を単に機械的に定常領域に分解するのではなく、生物の特性に基づき定常領域を特定する点で、従来の局所定常モデルとは根本的に異なる。

[0078] 「局所定常領域をノードとして利用した生体状態追跡と因果構造を分析する方法」とは、時間発展するヒト生体に状態変化を誘導する内的な要因を同定することで、健康維持や疾患発症を予想し制御する手法である。疾患は、潜在的な変化を獲得した無症状のステージから始まり、急性症状、慢性化、一部の生理機能の障害、一部の生理機能の消失、身体障害、寝たきり、死へと非定常に時間発展する。すなわち疾患の時間発展も、定常状態と見なせる時間区分の連続的な展開として近似することができる。

[0079] 健康状態や疾患の状態変化を予想し制御するためには、まず全身性で複雑な様相を示す疾患の局所定常状態領域を、細胞から産生される分子を代理指標としてコード化することが必要である。これは疾患と関連する臓器、組織、サブシステムの細胞から発現される分子の時系列データを、局所定常領域の観点から分割し、疾患の表現型と関連させることを意味する。この関連付けに有効な分子の種類は、生物学的知識により構築した因果構造モデルを事前確率として、多数の個人から生体内分子の発現量の時系列変化のデータをベイズ法により分析し最適なものを選抜することができる。従って、細胞から産生される様々な分子の中でも血中の分子は、血液サンプルを用いて比較的容易に時系列データを取得することが可能であり、本技術の「動的構築モデル」の応用に有効である。

[0080] 血中分子を測定するための機器として、グルコースメーターは、プロトタイプとなるデバイスと考えられる（フリースタイルライト、アボット社）。グルコースメーターは、痛針を用いて採取した少量の血液から血中グルコース量を測ることを可能としている。このシステムはグルコース以外の血中分

子の測定にも応用できる。一方、経皮的に血中の分子を測定する方法も最近開発されている。

[0081] 血中分子の発現量の時系列変化から抽出された定常領域の時間区分は、多様な疾患に関連した状態をコード化し、その状態変化を追跡するのに有効なだけでなく、疾患発症の原因である因果関係を解析するのにも利用できる。健康管理の問題は、健康状態から無症状の疾患初期ステージへと移行する原因となる血中分子の変化を同定し、疾患の発症を未然に防ぐのに利用することである。また、疾患の管理の問題とは、初期の疾患を合併症へと誘導する原因となる血中分子の変化を同定し、合併症の発症を予防に利用することである。

[0082] 本技術の「動的構築モデル」による制御を行うには、生体内分子群の発現量の変化の間に存在する因果関係をまず生物学的知識を用いてモデル化する。そして、複数の患者の生体内の分子の時系列変化のデータから定常領域を抽出して同定したノード間の確率構造／因果構造に関するグラフ構造の作製に利用する。その後、多数の健常人や患者の生体分子の発現変化の時系列データを利用して最適なグラフ構造に改良する。グラフ構造の変化は回帰分析における共変量の選択問題として解くことができる。構築された最適なグラフ構造は、生体分子の発現変化を経時的な測定を継続することで、各個人の健康管理や疾患管理を目的とした未来の生体状態の変化を予想し、制御することに利用することができる。

[0083] <第1の実施形態>

[動的構築モデルの健康状態、疾患状態の予想と制御への応用]

[0084] 糖尿病、ガン、免疫疾患、認知症、循環器病に対する有効な予防法を、個人の多様性や多様な過去の履歴に基づき開発することは、現代医療に求められる重要な挑戦である。本技術の「動的構築モデル」を用いることで、これらの疾患群にたいして個別予防を実施することが可能となる。

[0085] 図1は、健康状態を予測制御する生体情報処理装置1の構成を示すブロック図である。この生体情報処理装置1は、選択部11、決定部12、測定部

13、作成部14、分割部15、同定部16、および推定制御部17により構成されている。

[0086] 選択部11は、測定分子を選択する。決定部12は、分子測定の間隔を決定する。測定部13は、測定を実行する。作成部14は、グラフ作成する。分割部15は、成分を分割する。同定部16は、定常領域や因果構造を同定したり、分子マーカの因果関係と外部環境の関係を同定する。推定制御部17は、状態変化を推定制御する。

[0087] 図2は、健康状態を予測制御する処理を説明するフローチャートである。以下、図2を参照して、図1の生体情報処理装置1が実行する予測制御処理について説明する。

[0088] ステップS1において選択部11は、メタボリックシンドロームに関連する予想と制御のための測定分子を選択する。すなわち、発現量の測定を行う血中分子の種類が選択される。

[0089] 本技術の「動的構築モデル」は、メタボリックシンドロームから循環器系疾患へと展開する一連の疾患の発症の予想と制御に応用することができる。糖尿病はメタボリックシンドロームから循環器系疾患へと展開する一連の疾患の中に位置づけられる(Nature 444; 839-888, 2006)。代謝性疾患は食生活の偏りや運動不足をはじめとする生活習慣の問題から、肥満、内臓肥満、アジポネクティンの分泌不全、インスリン抵抗性などが惹起される。病態はその後、高血圧、食後高血糖、脂質異常症へと進展する。ここまでのステージがメタボリックシンドロームと呼ばれる。

[0090] メタボリックシンドロームは内臓脂肪での炎症を誘導し、様々な臓器の再生系を活性化し、酸化ストレスを上昇させる。その結果、脂肪肝や非アルコール性脂肪肝炎を惹起する。その後、膵機能障害、インスリン分泌不全が生じることで糖尿病が発症する。

[0091] 糖尿病は腎症、網膜症、神経症、閉塞性動脈硬化症、脳血管障害、虚血性心疾患を誘導し、腎透析、失明、下肢切断、脳卒中、認知症、心不全へと発展し死に至る。特にメタボリックシンドロームにより誘導される炎症は、ガ

ンや神経変性疾患の原因となる。肥満や2型糖尿病に本技術の「動的構築モデル」を応用する場合には、以下に示した全身性の血中因子を分子マーカーとして、その発現量の時系列データを用いることができる(Nature 444; 839-888, 2006)。

[0092] まず脂肪組織と視床下部下垂体の相互作用に関連する分子群は、摂食、代謝、脂質保存、エネルギーバランスに重要な役割を担っていることから、食生活の乱れや運動不足の状態を表象し、分析するのに利用できる。例えば慢性の社会的なストレスは、脂肪組織と視床下部下垂体副腎系の間の相互作用にエピジェネティックな影響を与え、副腎からのグルコルチコイド、アルドステロン産生に変化を及ぼす。その結果、食欲や睡眠の機能を変化させる。

[0093] この変化を捉えるには、血中のグルコルチコイド、アルドステロンに加えて、下垂体から産生されるACTH、脂肪組織から産生されるレプチン、アジポネクティン、ヴィスファティン、オメンチンの発現量の時系列変化を測定することが有用である。肥満、内臓肥満、アジポネクティンの分泌不全、インスリン抵抗性の表象と制御には、脂肪組織から産生されるレスチン、RBP4、膵臓から産生されるインスリン、グルカゴン、アミリン、GLPなどの発現量の時系列変化を測定することが有用である。メタボリックシンドロームから内臓脂肪での炎症への展開は、肝臓から産生されるCRP、PAI-1, NEFA, VLDL, LDL-ox、免疫細胞や脂肪組織から産生されるTNF- α 、IL-6、MCP1の発現量の時系列変化を測定することが有用である。

[0094] ステップS2において選択部11は、炎症性疾患の予想と制御のための測定分子を選択する。

[0095] リューマチ、全身性エリテマトーデス(SLE)に代表される膠原病、潰瘍性大腸炎、多発性硬化症をはじめとする免疫性疾患も、過去の感染履歴、腸内細菌の特性などが原因で発症することが知られている。自然免疫に働く顆粒球、単球、マクロファージ、樹状細胞、ナチュラルキラー細胞、獲得免疫に働くT細胞、ヘルパーT細胞 (TH0、TH1、TH2)、CD8+ 陽性T細胞、制御性T細胞

胞、B細胞から産生される特徴的なサイトカインの変化を分子マーカーとして追跡することは、免疫系疾患の予想と制御に有用である。サイトカインとしてはインタロイキン、インターフェロン、ケモカイン、増殖因子、細胞傷害因子などがあげられる。

[0096] ステップS 3において決定部1 2は、分子測定の間隔を決定する。

[0097] 分子マーカーの発現量を測定する時間間隔は短い方が望ましいが、測定環境によっては日、週、月程度の長い間隔の時系列データも利用できる。多くの分子が24時間周期のサーカディアンリズムで発現量を変動することから、1日以上の間隔で測定する場合には、一日の中の特定の時間を定めてデータを取得することが必要である。取得された時系列データは、いずれの時間間隔で取得されたものであっても、およそ100程度の離散的なデータを1つのセットとして分析を行うことが望ましい。従って、精度を求めるには1日に少なくとも2〜3回、望ましくは1日50〜100回程度の測定が求められる。

[0098] ステップS 4において測定部1 3は、測定を実行する。ここでは生体の分子としての血中の分子の発現量が測定される。具体的には、全身状態、局所状態、および染色体状態を表象する分子が測定される。この測定は、ステップS 3で決定された間隔で行われる。

[0099] 血中の分子マーカーを測定するための機器としては、痛針を用いることができる。この痛針を用いて少量の血液を採取することができる。また分子によっては、経皮的に血中の分子を測定する方法も可能である。分子の種類や量は測定する分子に対する選択的な抗体を蛍光標識することで実施できる。

[0100] ステップS 5において作成部1 4は、時系列変化のグラフ作成をする。具体的には、以下に示される式(5)乃至式(11)に基づくグラフが作成される。また、ステップS 6において分割部1 5は、成分を分割する。すなわち、生体状態空間モデルを用いて時系列データから3つの成分が抽出される。3つの成分は、例えば周期的成分、環境刺激応答成分、およびベースライン成分とすることができる。

[0101] 分子マーカーの時系列の測定データから、予想と制御のために必要な情報

を取り出すことが必要である。その方法として、複数の血中分子から構成される多変量の時系列データを「状態空間モデル」として統一的に取り扱うことが望ましい。「状態空間モデル」はシステムモデルと観測モデルの2つのサブモデルから構成される。一般にこの2つのサブモデルは次のように条件確率を用いて表現する。

$$[0102] \quad x_n \sim q(x_n | x_{n-1}) \quad (3) \quad (\text{システムモデル})$$

$$y_n \sim r(y_n | x_n) \quad (4) \quad (\text{観測モデル})$$

[0103] ここで y_n は多変量の観測された時系列、 x_n は直接には測定できない k 次元のベクトル、 q と r は、それぞれ x_{n-1} により与えられる x_n 、ならびに x_n により与えられる y_n の条件分布を示す。本技術の「生体状態空間モデル」では、 x_n が実際の人体の状態を表現し、 y_n が測定可能な血中分子の発現量を示す。初期状態ベクトル x_0 は分布が $P(x_0 | y_0)$ により配置されたとすると、疾患の状態の予想とは $P(x_n | y_n)$ の評価、すなわち観測値 y_n により与えられる x_n の分布、ならびに初期分布を求めることである。

[0104] メタボリックシンドロームから糖尿病をへて、循環器系疾患へと進展するプロセスを表象するのに利用できる分子群、免疫系疾患に関連する分子群、内分泌系の分子群などの測定可能な血中分子のそれぞれの時系列データ $y(t)$ は、次式で表される。すなわち、3つの異なる成分からなるモデルで表現することができる。

$$[0105] \quad y(t) = s(t) + x(t) + b(t) + v(t) \quad (5)$$

$s(t)$: 周期的成分

$x(t)$: 環境刺激応答成分

$b(t)$: ベースライン成分

$v(t)$: 観測誤差

[0106] 環境刺激応答成分 $x(t)$ は、多重線形モデルとして次のように定式化できる。

$$x(t) = F(t) x(t-1) + v_x(t) \quad (6)$$

$$i = 1, 2, 3, \dots, n$$

[0107] 上記式において、 $F(t)$ は環境刺激に対する出力の変換関数を示し、 $v_x(t)$ は環境刺激の変化を示す。変換関数のパラメータは、自己回帰モデルにユールウォーカー法、最小二乗法、あるいはPARCO法を適応することで最適な値を選択する。

[0108] ベースライン成分 $b(t)$ は多項式平滑化スプラインモデルにより、次数 m の回帰モデルとして次のように定義できる。

[数1]

$$b(t) = \sum_{i=1}^m a_i b(t-i) + \varepsilon(t) \quad \cdots (7)$$

a_i は回帰係数、 $\varepsilon(t)$ はノイズ成分を示す。

[0109] さらにベースライン成分 $b(t)$ は次のように表記できる。

$$b(t) = H(t, t-1) b(t-1) + V(t, t-1) \quad (8)$$

$H(t, t-1)$ は $m \times m$ の行列であり、 $V(t, t-1)$ は m 次元のノイズに関する行列である。このモデルを用いてベースラインが平滑化するような最適な関数を選択する。

[0110] 本技術の「動的構築モデル」では、周期的成分 $s(t)$ の主要な構成要因は24時間周期のサーカディアンリズムである。このような周期成分は1周期の間に p 個の観測値が得られる場合は、次式が近似的に満たされることになる。

$$s(t) = s(t-p) \quad (9)$$

[0111] これを時間遅れオペレーター G を用いて表すと、次式が近似的に満たされる。

$$(1 - G^p) s(t) = 0 \quad (10)$$

従って次数 l の周期的な成分は白色雑音 $v(t)$ により、次のような季節調整モデルとして定式化できる。

[数2]

$$\left(\sum_{i=0}^{p-1} G^i\right) s(t) = r(t) \quad \cdots (11)$$

[0112] この処理により、血中分子の時系列データは、周期的成分、環境刺激応答成分、およびベースライン成分に分割することができる。

[0113] ステップS7において同定部16は、定常領域を同定する。すなわち、「生体局所定常モデル」を用いて定常領域が同定される。

[0114] 局所定常領域は時系列データを周期的成分、環境刺激応答成分、ベースライン成分のモデルで解析することで見出すことができる。「動的構築モデル」においては「時間発展する細胞記憶」を原因とする定常状態の変化と新しい定常状態の獲得に焦点を合わせることが有用である。細胞記憶は、ベースライン成分の変動、周期的成分の振幅ならびに周期、環境刺激応答成分の最大発現量ならびに最小発現量に影響を与える。この中で環境からの入力 of 強度と独立して同定できるのは、ベースライン成分の変動と周期的成分の振幅ならびに周期の変動である。この点に注目して、ベースライン成分の変動、周期的成分の振幅の変動、周期的成分の周期の変動の少なくとも1つから局所定常領域を同定する。

[0115] 血中分子群の時系列変動の中には、24時間周期のサーカディアンリズムに加えて、月や年の単位で繰り返し現れる季節変動のパターンが存在する。これをベースラインの変化として解釈すると誤った予想や制御方法を導く危険性がある。このような季節性の成分はベースライン成分から抽出し、長期間における周期成分として分割する。

[0116] ステップS8において同定部16は、因果関係を同定する。すなわち、各分子の定常状態間の因果関係が作成される。因果関係の同定には、動的構築モデルを用いることができる。

[0117] この方法により抽出された「時間発展する細胞記憶」に関係する血中の分子群の定常領域は、時系列に従って直接的あるいは間接的な因果関係を形成

している。この因果関係を疾患の進展の制御に応用するために、定常領域をノードとして、確率構造／因果構造に関するグラフ構造を作製することで因果構造を表現する。

- [0118] 図3は、5つの血中分子A, B, C, D, Eの4つの定常領域 y_1 乃至 y_4 をノードとした因果関係に関するモデルを示す図である。図3において、 Y_i (A) 乃至 Y_i (E) は、それぞれ血中分子A乃至Eの定常領域 y_i ($i = 1, 2, 3, 4$) のノードであり、 x_0 乃至 x_4 は、生体の状態を表している。状態 x_0 , x_1 は、健康状態を表しており、状態 x_2 , x_3 は、症状未発現状態を表し、状態 x_4 は症状発現状態を表している。番号は、小さい値から大きい値の方向に向かう時間の経過を表している。
- [0119] このモデルでは、ノード Y_4 (A) とノード Y_4 (D) が疾患の状態を代替するマーカーであるとともに原因となっている。ノード Y_4 (A) の誘導にはノード Y_3 (A) とノード Y_3 (C) が働いており、ノード Y_4 (D) の誘導にはノード Y_3 (E) が働いている。同様に、ノード Y_3 (A), Y_3 (C), Y_3 (E) の誘導には、それぞれノード Y_2 (C), Y_2 (A), Y_2 (D) が働いている。さらにノード Y_2 (A), Y_2 (C), Y_2 (D) の誘導には、それぞれノード Y_1 (A), Y_1 (B), Y_1 (D) が働いている。
- [0120] この生体分子の定常領域間の因果構造に関するグラフ構造の推定は次のように定式化することができる。まず生体分子Molecule 1の定常状態の値を確率変数とみなす。すなわち、ある状況下でMolecule 1の量は、確率変数 X_1 の実現値であると考え。本実施形態の生体局所定常モデルにより処理された観測データは、ある状況下で p 種類の血中分子の確率変数の実現値を計測したデータと見なすことができる。様々な時間、状況において定常状態が観測されるので、局所定常モデルにより処理された観測データは、データ行列として表現することができる。すなわち、 n 個の異なる定常状態からなる観測データは、 $p \times n$ の行列サイズとなる。この行列データ間の依存関係を推定することが本解析の目的となる。
- [0121] データ行列から確率変数間の依存関係を表すグラフを推定する方法として

は、ブーリアンネットワーク、ベイジアンネットワーク、グラフィカルガウシアンモデル、常微分方程式など様々な数理モデルが提案されており、本実施形態に利用することができる。なかでも、確率変数間に非閉路有向グラフで表される依存関係があると仮定し、さらにこの非閉路有向グラフの構造とノード間の依存関係にマルコフ連鎖律を仮定することで、同時確率は各確率変数に対してその直接の親確率変数集合を与えた下での条件確率の積によって表現されるベイジアンネットワークによる推定が、本実施形態に有用である。

[0122] ベイジアンネットワークのより詳細な適応は、DNAチップのデータ解析に用いられた手法を応用することができる（玉田嘉紀、井元清哉、宮野悟 統計数理 第54巻 第2号 333-356）。具体的には本実施形態の定常状態の発現量をDNAチップのmRNAの発現量に対応させることで実施する。本実施形態の定常状態グラフ構造推定のために現在の生理学モデルを事前知識として利用することができる。

[0123] グラフ構造はまず個人レベルで構築される。現在の生理学モデルは十分に個別の多様性や多様化に対応できるものではないことから、グラフ構造推定に誤りが生じる可能性がある。グラフ構造を最適化するには、多数の個別グラフ構造を比較し、似た構造を有する小集団に分類し、グラフ構造を再選択する。これにより最適なグラフ構造が構築される。

[0124] 多数の患者の共通の症状に到る因果の経路は個人により異なる可能性が存在する。従って、最適なグラフ構造は1つに収斂させるのではなく、同じ因果構造を持つ集団ごとに最適なものを決定する。

[0125] ステップS9において同定部16は、分子マーカーの因果関係と外部環境の関係を同定する。すなわち、各分子間の因果関係と環境の関係が作成される。外部環境としては生活習慣の時系列変化の測定値を用いることができる。

[0126] 環境からの入力が定量的に測定できる場合には、異なる時期に受けた異なる強度の刺激がどのように血中分子の反応性に影響を与えるかを分析するこ

とで、「構築された変化」を表象し、制御に応用することができる。例えば負のストレスレベルはDRM (Daily Reconstructive Method) (1日構成法)により測定することができる。これは1日をイベントごとに平均24程度に分割し、それぞれのポジティブ、ネガティブの値を0～6までの数値で段階的に表記していくものである。

[0127] 食事の影響は、各食事の内容を電子的に記録することで、カロリー、炭水化物(糖質)、脂肪、タンパク質、ビタミン、フィトケミカルなどの摂取情報を測定することができる。またサプリメントなどの摂取量も測定することが可能である。これらの環境刺激と血中の分子の発現量の変化に相関が観察されれば、その環境刺激を目的とする血中分子の発現量の変化、すなわち生体の状態変化の制御に利用することができる。

[0128] ステップS10において推定制御部17は、状態変化を推定制御する。すなわち、ヒトの将来の生体の状態変化が推定され、予防方法を含む介入方法が提案される。

[0129] 状態変化の推定は、集団をステップS8で求めた定常領域間の因果構造の相似性により分類する。これにより疾患が進行した患者から取得した因果構造より、同じグループに分類される疾患予備群の未来の変化を推定することが可能となる。また、疾患発症前の潜在的变化から疾患を発症せずに健康状態に回復した因果構造を、疾患発症を予防する制御方法に利用できる。制御のためにはステップS9で同定した分子マーカー(すなわち、内部環境)の因果関係と外部環境の関係が利用できる。

[0130] <第2の実施形態>

[動的構築モデルの培養細胞の標準化への応用]

[0131] 図4は、培養細胞の細胞状態を予測制御する生体情報処理装置51の構成を示すブロック図である。この生体情報処理装置51は、選択部61、測定部62、決定部63、分割部64、同定部65、および推定制御部66により構成されている。

[0132] 選択部61は、測定分子を選択する。測定部62は、測定を実行する。決

定部 6 3 は、測定間隔を決定する。分割部 6 4 は、成分を分割する。同定部 6 5 は、定常領域を同定したり、分子マーカー間の因果構造や、分子マーカーと環境条件の因果構造を同定する。推定制御部 6 6 は、培養細胞の細胞状態を推定制御する。

[0133] 図 5 は、培養細胞の細胞状態を予測制御する処理を説明するフローチャートである。以下、図 5 を参照して、図 4 の生体情報処理装置 5 1 が実行する予測制御処理について説明する。

[0134] ステップ S 3 1 において選択部 6 1 は、測定分子を選択する。

[0135] 培養細胞が培養に伴いその性質を変化させることは、細胞を用いた評価系の構築や細胞を有効成分とする治療剤の開発において大きな障害となっている。培養細胞の状態変化の予想と制御には、メソスケールの状態推定の問題と定式化することができる。メソスケールの状態とは、培地中に添加された成分ならびに培養細胞から産生される分子により形成される局所細胞外環境により表現される。従って培養細胞の状態予想と制御には、培養細胞から産生される分子群と培地中の分子の時系列変化を測定する。

[0136] 培養細胞から産生される分子は、分泌性のものだけではなく細胞表面に発現する分子群を含むが、細胞内情報伝達分子や転写因子は含まない。培地中の成分としてはインスリン、あるいはIGF-Iの量の測定が有用である。細胞から生産される分子としては、インスリン、IGF-Iに加えてFGF-2, EGF, PDGFをはじめとした増殖因子、インテグリンファミリー分子などが有用である。

[0137] ステップ S 3 2 において測定部 6 2 は、測定を実行する。

[0138] 培養細胞から産生される分子ならびに培地中の分子は、培地中より少量のサンプルを取得することで測定する。また細胞から産生される分子の中で、細胞外に発現する分子により捕捉され培地中に浮遊しないものについては、分子を蛍光標識した後、培養細胞の画像解析により定量する。また測定の困難な分子に関しては、細胞内のレポーターシステムを利用して、分泌蛋白や膜蛋白の遺伝子発現により代替測定を実施する。

[0139] ステップ S 3 3 において決定部 6 3 は、測定間隔を決定する。

[0140] 測定の間隔は10分から1時間程度が望ましい。レポート系を用いた場合には連続的にデータを取得する。

[0141] ステップS34において分割部64は、成分を分割する。またステップS35において同定部65は、定常領域を同定する。

[0142] 取得された分子マーカーの時系列データより情報を引き出すために、培養細胞から産生される分子と培地中の分子からなる多変量の時系列は、状態空間モデルとして統一的に取り扱う。状態空間モデルはシステムモデルと観測モデルの2つのサブモデルから構成される。一般にこの2つのサブモデルは次のように、条件付き確率を用いて表現する。

[0143] $x_n \sim q(x_n | x_{n-1})$ (12) (システムモデル)

$y_n \sim r(y_n | x_n)$ (13) (観測モデル)

[0144] ここで y_n は多変量の観測された時系列、 x_n は直接には測定できない k 次元のベクトル、 q と r はそれぞれ x_{n-1} により与えられる x_n ならびに x_n により与えられる y_n の条件分布を示す。「生体状態空間モデル」では x_n が実際の培養細胞の状態を表現し、 y_n が測定可能な培地中の分子ならびに培養細胞から産生される分子の発現量を示す。初期状態ベクトル x_0 は分布が $p(x_0 | y_0)$ により配置されたとすると、培養細胞の状態の予想とは $p(x_n | y_n)$ の評価、すなわち観測値 y_n により与えられる x_n 分布、ならびに初期分布を求めることと定式化できる。

[0145] 培養細胞から産生される分子群の時系列データ y_n は、季節調節モデルを用いて周期的成分に、多重線形モデルを用いて環境刺激応答成分に、多項式平滑化スプラインモデルを用いてベースラインの成分に、それぞれ分解し、抽出することが可能である。培養細胞の細胞記憶は、周期的成分の振幅と周波数、環境刺激応答成分の最大発現量、ベースラインの変化に反映される。この中から、周期的成分の振幅と周波数ならびにベースラインの変化に関する局所定常領域を、「時間発展する細胞記憶」として注目する。

[0146] ステップS36において同定部65は、分子マーカー間の因果関係を同定する。またステップS37において同定部65は、分子マーカーと環境条件

の因果関係を同定する。

[0147] 周期的成分の振幅と周波数の変化、ベースラインの変化、培地中に添加した人工成分の時間変化ならびに環境刺激応答成分は、時系列に従って直接的あるいは間接的な因果関係を形成している。この因果関係を培養細胞の状態変化の制御に応用するために、様々な培地中の分子群ならびに培養細胞から産生される分子群の局所定常状態をノードとして、確率構造／因果構造に関するグラフ構造を作製することで因果構造を表現する。

[0148] グラフ構造は従来の生物学的な知識に基づき作製した後に、時系列データを利用して最適なグラフ構造に改良する。グラフ構造の変化は回帰分析における共変量の選択問題として解析する。培養細胞はその由来により同じ状態変化に到る因果の経路がそれぞれ異なる可能性がある。従って、最適なグラフ構造は1つに収斂させるのではなく、同じ因果構造を持つ培養細胞の集団ごとに最適なものを決定する。この集団の分類においては、「生体状態空間モデル」の状態変化特性が利用できる。また培地成分、細胞培養密度、酸素分圧、温度などの環境条件の変更による介入に対する因果構造も別途作成する。

[0149] ステップS 3 8において推定制御部6 6は、培養細胞の細胞状態を推定制御する。

[0150] 培養による細胞の状態変化の推定と制御は、同じ組織由来の細胞を近似した条件で培養した過去の因果構造から実施することができる。また細胞の状態を求める方向への制御は、培地成分、細胞培養密度、酸素分圧、温度などの環境条件の変更と細胞の状態変化の因果構造から実施することができる。

[0151] <第3の実施形態>

[0152] 次に、血中に分泌される分子Aを産生する細胞のモデルを検討する。この細胞は、図6に示されるように、次の特徴を有する。

- (1) 定常状態ではサーカディアンリズムにより制御される。
- (2) 環境からの刺激に応答して誘導発現する。
- (3) 分子Aの受容体ならびにリプレッサーRによりポジティブならび

にネガティブなフィードバックを受ける。

(4) リプレッサー R がエピジェネティックス修飾されることで、その発現が抑制される。

[0153] この細胞から血中に分泌される分子 A の時系列変化は、図 7 の A の y のように変化する。

この時系列の測定データ $y(t)$ は、次のように仮定される。

$$[0154] \quad y(t) = s(t) + x(t) + b(t) + v(t) \quad (14)$$

$s(t)$: 周期的成分

$x(t)$: 環境刺激応答成分

$b(t)$: ベースライン成分

$v(t)$: 観測誤差

[0155] 環境刺激応答成分 $x(t)$ は、多重線形モデルとして次のように定式化した。

$$x(t) = F(t) x(t-1) + v(t) \quad (15)$$

[0156] 上記式の $F(t)$ は細胞の環境刺激に対する出力の変換関数を示し、 $v(t)$ は環境刺激を示す。変換関数のパラメータは自己回帰モデルを用いて最適なものを探索した。

[0157] ベースライン成分 $b(t)$ は次数 2 の回帰モデルとして次のように定義し、利用した。

$$b(t) = a_{t-1} b(t-1) + a_{t-2} b(t-2) + v_B(t) \quad (16)$$

ここで a_j は回帰係数、 $v_B(t)$ はノイズ成分を示す。そうすると $b(t)$ は次のように表記できる。

$$b(t) = H(t, t-1) b(t-1) + V(t, t-1) \quad (17)$$

[0158] ここで $H(t, t-1)$ は 2×2 の行列であり、 $V(t, t-1)$ は 2 次元のノイズである。関数はベースラインが平滑化するような最適化のものを選択した。

[0159] 周期的成分 $s(t)$ は、24時間周期のサーカディアンリズムを基盤とするので、1周期の間に p 個の観測値が得られる場合は、周期成分は

$$s(t) = s(t - p) \quad (18)$$

を近似的に満たすことになる。これを時間遅れオペレーター G を用いて表すと、次式が近似的に満たされる。

$$(1 - G^p) s(t) = 0 \quad (19)$$

[0160] 従って次数 l の周期的な成分は白色雑音 $v(t)$ により、次のような季節調整モデルとして定式化できる。

[数3]

$$\left(\sum_{j=0}^{p-1} G^j \right) s(t) = v(t) \quad \cdots (20)$$

[0161] このモデルを用いて図7のAの y の時系列データを、周期的成分 s 、環境刺激応答成分 x 、およびベースライン成分 b に分解した。なお、図7のAにおける i は誘導成分を表している。

[0162] 次にベースラインの変動特性から、 $Y_1(A)$ 、 $Y_2(A)$ 、 $Y_3(A)$ の3つの定常領域に分解した。細胞には時間 k から l の間に環境からの入力があり、 $Y_2(A)$ の環境刺激応答成分が、この入力に対する直接の出力である。また、本自己制御モデルでは分子Aの時間変化区分 $Y_2(A)$ は因子Aの発現状態が $Y_1(A)$ から $Y_3(A)$ へと変化した原因となっている。この関係はA分子の発現量変化に関する定常領域をノードしたグラフ構造により、図7のBのように、 $Y_2(A)$ から $Y_3(A)$ への変化として示される。

[0163] 以上のようにして、各実施形態においては、ヒト生体の時間発展的な変化を定式化し、健康管理、疾患予防、疾患根治などに利用できる生体状態の変化を予想し制御することができる。

[0164] また、従来の手法では取り扱うことのできなかった長期間の生活習慣を原因とする慢性疾患の発症ならびに慢性疾患の進展を制御することに応用することができる。さらに個人間の共通性と異質性を同時に扱うことができるこ

とから、個別医療に有効なツールとなる。また細胞培養により誘導される細胞の時間発展による多様化を制御し、細胞治療や細胞評価系の標準化にも応用することが可能である。

[0165] <第4の実施形態>

[0166] 胎内栄養障害を原因として起こる胎児発育不全により低体重で生まれた子供の多くは、2歳までに遅れを取り戻す成長を示すことで、標準体重に生まれた子供と同じ体格となる。誕生後の食事として低カロリー食が与えられると、遅れを取り戻す成長が抑制され、2歳以降は高カロリー食によっても成長が取り戻せない。このような遅れを取り戻す急激な成長を示す場合に、将来インスリン抵抗性が観察されるリスクが高まる(Rotteveel, J. et al. European Journal of Endocrinology 158; 899-904, 2008)。その原因として体格を取り戻しても、胎内の栄養環境に適応するためにエピジェネティックスにより変化した膵臓の β 細胞の分化効率の低下、インスリンシグナルの応答性、食欲レベル、ストレス応答性などが取り戻されないことが推察される。

[0167] しかし遅れを取り戻す成長を示した胎児発育不全による低体重児が、すべてインスリン抵抗性を発症するわけではない。一人一人のインスリン分泌能、インスリンシグナル応答性、食欲レベル、ストレス応答性は、遅れを取り戻す成長のような急激な変化は示さないものの、生活習慣により変化することから、血中インスリン量、血中レプチン量、血中グルココルチコイド量、血糖値、ならびにカロリー摂取量などを本実施形態の方法を用いてモニターすることで、肥満傾向、インスリン抵抗性発症傾向などを予想し制御することが可能となる。

[0168] 具体的には食事による摂取カロリー、食後のインスリン分泌量、血糖値の時間変化の関係を毎日測定し比較することで、インスリン分泌の摂取カロリーあたりの生産量ならびに、インスリン応答性の強度の変化が、ある定常状態から別の定常状態に変化したことを測定することが可能となる。定常状態の変化はインスリン抵抗性を改善する方向の変化と悪化させる方向の変化に分類できる。インスリン分泌量ならびにインスリン応答性の強度の増加は改

善方向の変化となり、この逆は悪化させる方向への変化となる。

[0169] グルココルチコイドやレプチンの分泌量も、本実施形態の方法を利用することで、ある定常状態から別の定常状態への変化としてはじめて定式化することができる。一定のカロリー摂取量に対するレプチンの生産量が低いことは、摂食抑制を弱め肥満傾向を高める。グルココルチコイド量が高いことは、複数の作用部位を介して肥満傾向や循環器系疾患発症傾向を高めることから、グルココルチコイドの分泌量の低下ならびにレプチンの分泌量を増加させる方向の定常状態の変化は、インスリン抵抗性ならびに関連疾患の発症傾向を改善する方向の変化となる。

[0170] 本実施形態の特徴の一つは、様々な生体内の分子の時間変化を幅広く追跡することにある。従って本実施形態は、インスリン分泌、インスリン応答性、グルココルチコイド分泌、レプチン分泌を、ある定常状態から別の定常状態への変化させることに関係する未知の生体内分子を同定することにも利用できる。また生活習慣の記録と該新規関連生体分子の関係を、本実施形態の方法を利用して明らかにすることで、インスリン抵抗性や肥満を改善するための新しい改善方法を提案することが可能となる。

[0171] <第5の実施形態>

[0172] デヒドロエピアンドロステロンに硫酸が付加されたDHEA-Sの血中量は、男女ともに6〜7歳ごろから増加しはじめ、12〜13歳ごろにピークに達し、13〜25歳ごろまでその高値を維持し、以後加齢とともに直線的に減少する。長期追跡研究では、男性の血中DHEA-S値は、死亡率や心血管疾患の発症と逆相関することが、世界各地の複数の研究で明らかになっている（柳瀬敏彦 アンチ・エイジング医学 vol5, No1, 42-46, 2009）。しかし、現在のところ、何がDHEA-Sの生産量を低下させるかは明らかになっていない。本実施形態により、DHEA-Sの分泌量をある定常状態から別の定常状態に変化させることに関与する生体分子ならびに該生体分子の変動と関係する生活習慣を明らかにし、DHEA-Sの減少を抑制する方法の新しい提案を行うことが可能となる。

[0173] <第6の実施形態>

[0174] 老化遅延や寿命の延長に最も有効な方法は、カロリー制限である。カロリー制限とは、単に食事の量を減らすのではなく、至適な栄養を摂取しつつ行うカロリー制限である。しかし、至適な栄養とはなにか、あるいはどのレベルまでカロリーを制限したらいいのかについての科学的なエビデンスは十分に存在しない。本実施形態は、個人レベルで、至適な栄養摂取レベルやカロリー摂取レベルを明らかにすることに利用できる。

[0175] 例えば糖尿病予備軍の患者にカロリー制限を施す場合、脂肪細胞の質的変化の指標となるアジポネクティン産生量、活性酸素産生量、炎症の指標となるc-反応性蛋白(CRP)の時系列変化を測定することができる。これにより、脂肪細胞の状態、活性酸素ストレス状態、炎症状態が、ある定常状態から別の定常状態へと変化するものとして定式化することができる。

[0176] また該定常状態から別定常状態への変化に関連する新しい生体分子の同定にも本実施形態は利用できる。新しく見出された生体分子マーカーと食事に関する時系列記録は、栄養摂取量ならびにカロリー摂取量を個別に最適化することに利用できる。

[0177] <第7の実施形態>

[0178] ストレスに対する感受性の違いは、慢性ストレスに伴う鬱の発症や自殺を誘導する。ストレスに対する感受性は、海馬から扁桃核に投射するグルココルチコイド応答性のニューロンによるネガティブフィードバックの強度に依存する。また該ニューロンのネガティブフィードバックの強度は、グルココルチコイド受容体遺伝子のエピジェネティックス修飾により左右される。自殺者と対照群の当該ニューロンのエピジェネティックス修飾を比較すると、幼児虐待や育児放棄を受けた者は、そのグルココルチコイド受容体遺伝子がエピジェネティックス修飾により抑制されていることが示されている(Hyman, S. Nature Neuroscience Vol. 12, No. 3, 241-243, 2009)。

[0179] ストレス感受性のレベルを把握することは、鬱を発症しやすい人を定量的に把握し、適切な環境を提供することで、鬱の発症を防ぐことが可能となる。本実施形態を利用して、血中グルココルチコイドの時系列変化を10分程

度の短い間隔で測定すると、その時系列変化は日内変動成分と刺激依存的な成分に分割することができる。日内変動成分から、測定日の血中コレステロールの平均的な分泌量を計算することができる。また刺激依存的な成分からは、最大分泌量と持続時間を測定することにより、ネガティブフィードバックの相対強度を計算することができる。従って、本実施形態を適応することで、ストレスの感受性を測定することが可能であり、さらにストレス感受性が生活習慣とともに変化していくプロセスを追跡することも可能である。

[0180] 鬱の発症メカニズムに関する最も有名な仮説は、神経栄養因子であるBDNF (Brain Derived Neurotrophic Factor)の低下を原因とするものである(Shi, Y. et al. Psychiatry and Clinical Neuroscience 64, 249-254, 2010)。BDNFはt-PA（組織型プラスミノゲン活性化因子）によりpro-BDNF（BDNF前駆物質）から産生される。すなわち、本実施形態を利用して、血中のBDNF, pro-BDNF, t-PAの時系列変化を観測し、ある定常状態から別の定常状態への変化を同定することで、鬱の発症や治療効果を推定することが可能となる。また本実施形態は、鬱の発症と予防に関係する新しい生体分子マーカーを見出すことにも利用できる。

[0181] <第8の実施形態>

[0182] 伴侶の介護などの慢性的な過労と精神的なストレスは感染症の頻度を高め、創傷治癒の力を弱め、高血圧や心臓疾患の頻度を向上させる。このような慢性的な過労と精神的なストレスに関係するバイオマーカーとして、炎症性サイトカインIL-6（インターロイキン6）がある。IL-6は加齢依存的に発症する循環器系疾患、骨粗鬆症、II型糖尿病、癌、歯周病、虚弱体質、臓器不全などに関連することも知られている。IL-6は鬱、負の感情、ストレスを伴う経験より誘導発現する。また血中のIL-6の加齢に伴う増加は介護に従事しているものでは介護に従事していないものと比較して顕著に増加する(Kiecolt-Glaser JK et al. Proc. Natl. Acad. Sci. USA 100: 9090-9095)。従ってIL-6の発現量の変化を追跡することで、加齢に伴う疾患やストレスに伴う疾患の発症を予想し予防することが可能である。

[0183] <第9の実施形態>

[0184] 人体は開放系のシステムであり、身体状態の未来は過去の単なる繰り返し

ではない。本技術の細胞記憶は、生体が履歴に従い時間発展する過程を、生

体分子の測定のみから捕捉し、定量化することを可能とする。この定量化さ

れた時間に伴う変化を、個別化ヘルスケアに応用するには、個人の多様性と

多様化に基づき、適切な未来の生活習慣の選択肢を提供することが必要であ

る。

[0185] また、本技術に基づき個々人の生体状態測定を開始する前の過去の多様性

と多様化の実体を推定するためには、個々人が両親から継承するDNA配列情報

が有用である。各個体の現在の多様性は、遺伝的な多様性と過去の生活習慣

により形成される。従って、本技術の時間発展する生体分子の細胞記憶情報

、DNA多型情報、生活習慣に基づく環境情報の3つを統合することで、健康の

未来予想の精度を高めることが可能である。

[0186] 塩基配列技術の進歩により個々人の全ゲノム配列を分析することが可能と

なってきた。このような配列解析からは、個々人の持つ一塩基の多型と構造

の多型を検出することが可能である。遺伝的多様性と健康や病気の関係は、

染色体異常症、単一遺伝子疾患、多因子遺伝子疾患、突発型疾患の4種類に

分類することができる (トンプソン&トンプソン 「遺伝医学 第7版」

メディカル・サイエンス・インターナショナル 2009年)。

[0187] ヒトの体細胞は23対46本の染色体からなり、そのうち22対は男女差のない

常染色体である。残りの1対は性染色体と呼ばれる。女性には2つのX染色体

を、男性は1つのX染色体と1つのY染色体を持つ。染色体異常の発症頻度

は高く、新生児1000人あたり約6人が罹患する。染色体異常のなかでも多い

のが染色体の数が増える異常である。染色体の数が1つ増えたものはトリソ

ミーと呼ばれる。例えば第21染色体のトリソミーはダウン症候群という病気

を発症する。約800人に1人がダウン症候群で生まれる。

[0188] ダウン症候群で生まれる頻度は母親の年齢とともに指数関数的に増加し、4

5歳以上では約1%の子供がダウン症候群で生まれ来る。ダウン症候群で主に

問題になるのは精神発達の遅れである。しかし、それ以外にも白血病のリスクが増加し、多くのダウン症候群の患者が40歳を過ぎるとアルツハイマー型認知症を発症する。第21染色体のトリソミーが起こる原因の90%は母方の卵子を形成する減数分裂のプロセスにある。従って、第21染色体のトリソミーという異常は受精の瞬間から存在する。しかし症状が現れる時期は様々であり、遺伝的原因の成立と病気の発症の間には「時間的な遅れ」が存在する。

[0189] 単一遺伝子疾患はゲノム上の遺伝子の変異を原因として発症する疾患であり、メンデル型遺伝病と呼ばれる。メンデル型遺伝病には現在まで3917種類が報告されている。遺伝形式には優性と劣性がある。優性とは母親か父親のどちらかから遺伝子変異を継承しただけで発症する遺伝形式である。一方の劣性とは両親の両方から遺伝子変異を継承したときにだけ発症する遺伝形式を意味する。集団の約2%が生涯のいずれかの時期に単一遺伝子疾患であることに気づくと言われている。

[0190] 単一遺伝子疾患は主に小児期に発症する疾患であるが、単一遺伝子疾患の約10%弱は思春期以降に症状が出る。さらに1%以下ではあるが、生殖期間の終わった後の高齢期に発症する場合がある。発症の時期が人生のいずれの時期であったとしても、そこに変異の獲得と症状の発現の間には「時間の遅れ」が観察される。

[0191] 先天異常のような出生時に観察される疾患、心筋梗塞、ガン、糖尿病、リウマチ、精神疾患、認知症などの中高年で発症する慢性疾患は多因子疾患と呼ばれる。このようなタイプの疾患が出生時に出現する頻度は1000人あたり約50人である。しかし慢性疾患に罹患する人が多いので、集団でみると、1000人あたり約600人になる。発症には複数の遺伝的要因が組み合わさり、さらにそこにはある特定の環境要因への日常的な暴露や、偶然の暴露が関与していると考えられている。このような多因子疾患の遺伝的な影響についてはゲノムワイド関連解析（GWAS）という手法により解析されている。

[0192] メンデル型遺伝変異と多因子疾患の中間に位置するのが突発型疾患である。パーキンソン病やアルツハイマー病は家族性のもの以外に突発型があり、

このような突発型変異の同定には多因子疾患とは異なり個別ゲノム配列解析を用いる必要がある。

[0193] 染色体異常、メンデル型単一遺伝子疾患、多因子疾患、突発型疾患のいずれを観ても、変異の獲得と発症の間にはなんらかの時間の遅れが存在する。また複数の環境要因が即時的な応答反応として疾患を発症させているという説明も適切ではない。遺伝要因と環境要因が揃ってから疾患が発症するまでの間にも「時間遅れ」がある。すなわち、遺伝的要因と環境要因が細胞記憶として生体に潜在的変化を起こし、それらがある組み合わせを形成したときに症状として顕在化する。

[0194] 細胞記憶は、遺伝的多型の上に環境条件の変化が加わり獲得されることから、環境条件を定量的に記録することが健康の未来予想に有効である。環境条件のうちストレス、食事習慣、運動習慣、感染履歴、治療履歴などの外部環境は定量的に測定することが可能である。

[0195] ストレスとは個々人が環境に適応していないと認識することであり、同じ環境条件でもそれをストレスと感じるか感じないかは人によって異なる。従ってストレスは環境条件だけから把握することはできない。ストレスは、例えばDaily Reconstruction Method (DRM法)により記録することが有効である (Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. 「A survey method for characterizing daily life experience: the day reconstruction method.」 Science. 306:1776-1780, 2004年)。

[0196] この方法は1日のイベントを24程度に分割し、それぞれの心的なポジティブ・ネガティブ値を記録する。それによって各瞬間のストレス対処の状況を把握することができる。このような記録を継続することで、ストレスを、数時間から1日程度の短期間のストレス、数日から1カ月程度の中期的ストレス、数か月以上におよぶ長期ストレスに分類することができる。

[0197] 食事習慣に関しては、食事内容を食材に分けて記録することで、カロリー、炭水化物比率、タンパク質比率、脂肪の種類と量、ビタミン、フィトケミカル、微量元素（鉄、亜鉛、銅、コバルト、ヨウ素、セレン、マンガン、モ

リブデン、クロム、ボロン、バナジウム)、飲酒(アルコール)などの摂取量を記録することができる。

[0198] 運動量に関しては、加速度センサーをもつ携帯可能な機器により歩行、走行の質や量を自動的に測定することが可能である。

[0199] 感染履歴や治療・投薬履歴は、かかりつけの医師の電子カルテデータにより把握することが可能である。また個人で摂取している漢方薬や一般医薬は個人により記録することが可能である

[0200] 以上のように個々人の遺伝的多様性は、染色体異常症、単一遺伝子疾患、多因子遺伝子疾患、突発型疾患の原因となるDNAの一塩基多型ならびに構造多型を用いることで分類することが可能である。環境条件の時間に伴う変化に関しては、ストレス状態、食事による摂取成分、運動量、感染履歴、治療・投薬履歴により定量的に記述することができる。

[0201] 本技術の時間発展する生体分子の細胞記憶情報、DNA多型情報、生活習慣に基づく環境情報の3つを統合するために、次の3つの小集団の間の関係を、条件付き確率を用いて明らかにする。3つの小集団とは、遺伝的多様性に基づき分割された小集団、環境条件の変化に基づき分割された小集団、および生体分子の細胞記憶情報に基づき分割された小集団である。遺伝的多様性に基づき分割された小集団は、染色体異常症、単一遺伝子疾患、多因子遺伝子疾患、突発型疾患の原因となるDNAの一塩基多型並びに構造多型などの遺伝的多様性に基づく集団である。環境条件の変化に基づき分割された小集団は、ストレス状態、食事習慣、運動習慣、感染履歴、治療・投薬履歴などの環境条件に基づく集団である。生体分子の細胞記憶情報に基づき分割された小集団は、図2のフローチャートに示される処理により作製された、図3に代表される生体状態の変化に基づく集団である。

[0202] このような集団の条件付きの関係は、図8に示されるように、大きく次の3つに分類することができる。

i) 遺伝的多様性と環境条件の関係

ii) 環境条件と生体状態の関係

iii) 遺伝的多様性と生体状態の関係

- [0203] 本技術を個別化ヘルスケアに応用するということは、生体の分子の発現量の定常領域間の因果関係として示されるヘルスケアの対象者の過去の細胞記憶のデータに基づき、健康を維持するための未来のオプションを提供することである。このような未来のオプションを見つける方法としては、遺伝子情報の履歴、環境条件の履歴、生体状態の履歴など、総合的に対象者と似た履歴を持つ他者（すなわち対象者以外）の生活習慣の記録を参考にすることがあげられる。この方法は、図8に示される個々人の生体分子の細胞記憶情報、DNA多型情報などの遺伝子情報、生活習慣に基づく環境情報の相関関係における類似性を明らかにすることで実施可能である。
- [0204] このようなデータの類似性は、既存のクラスタリングの手法を応用することで判定することができる。クラスタリングの手法は大きく、分割型クラスタリングと階層型クラスタリングに分類できる。具体的には、A. K. Jain, M. N. Murthy and P. J. Flynn, Data Clustering: A Review, ACM Computing Reviews, (1999). や Ying Zhao and George Karypis, “Hierarchical Clustering Algorithms for Document Datasets”, Data Min. Knowl. Discov. 10(2): pp. 141-168 (2005) に記載の方法を応用することができる。
- [0205] この実施形態の処理を行う場合、生体情報処理装置101は、図9に示されるように構成される。この生体情報処理装置101は、選択部111、決定部112、測定部113、作成部114、分割部115、同定部116、検出部117、検索部118、データベース119、および推定制御部120を有している。選択部111、決定部112、測定部113、作成部114、分割部115、同定部116、および推定制御部120の機能は、図1に示した選択部11、決定部12、測定部13、作成部14、分割部15、同定部16、および推定制御部17と基本的に同様である。従って、その説明は繰り返しになるので省略する。
- [0206] 検出部117は、遺伝子情報を検出する。検索部118は、データベース119から類似する履歴を検索する。

[0207] データベース 119 は、多くの個人の情報を、図 8 を参照して上述した 3 つの小集団のいずれかに分類して記憶する。すなわち、上述した図 2 の処理を行って得られた多くの人の検査結果の情報が記憶される。またこのとき、上述した 3 つの小集団のいずれに分類するのかを決定するのに必要な情報を得るための測定・検査が追加される。具体的には、ストレス状態、食事習慣、運動習慣、感染履歴、治療・投薬履歴などの環境条件の測定・検査、および生体分子の細胞記憶情報の測定・検査以外に、DNA の一塩基多型並びに構造多型などの遺伝的多様性を特定する測定・検査が追加される。

[0208] この実施形態の処理は、図 10 のフローチャートに示される。以下、図 10 を参照してこの実施形態の処理を説明する。

[0209] ステップ S 101 においてデータベース生成処理が実行される。すなわち、多数の個人に対して、図 2 のフローチャートに示される処理が、各個人の DNA の一塩基多型並びに構造多型などの遺伝的多様性を特定する測定・検査とともに行われ、その結果が、データベース 119 に記憶される。ただし、個々人のプライバシーを秘匿する形式で情報が記憶される。

[0210] 上述したように、このデータベース 119 に記憶された情報は、遺伝的多様性に基づく集団、環境条件の変化に基づく集団、および生体分子の細胞記憶情報に基づく集団に分類される。そして各集団の次の 3 つの関係が、条件付き確率を用いて分析される。

i) 遺伝的多様性と環境条件の関係

ii) 環境条件と生体状態の関係

iii) 遺伝的多様性と生体状態の関係

[0211] なお、説明の便宜上、データベース生成処理を図 10 のフローチャートの健康状態予測処理の 1 ステップとして記載したが、何らかの方法によりデータベースが既に生成され、存在する場合には、それを利用することができる。この場合、このデータベース生成処理は省略することができる。

[0212] 図 10 のステップ S 102 乃至 S 109 の処理は、基本的に図 2 のステップ S 1 乃至 S 8 と同様の処理である。繰り返しになるので、これらについて

は簡単に説明する。すなわちステップS 1 0 2において選択部1 1 1は、ヘルスケアの対象者の測定分子を選択し、ステップS 1 0 3において決定部1 1 2は、分子測定の間隔を決定する。ステップS 1 0 4において測定部1 1 3は、ステップS 1 0 2の処理で選択された測定分子を、ステップS 1 0 3の処理で決定された間隔で測定する。

[0213] ステップS 1 0 5において作成部1 1 4は、時系列変化のグラフ作成をする。すなわち、上述した式(5)乃至式(11)に基づくグラフが作成される。また、ステップS 1 0 6において分割部1 1 5は、成分を分割する。すなわち、生体状態空間モデルを用いて時系列データから、周期的成分、環境刺激応答成分、およびベースライン成分が分割される。

[0214] ステップS 1 0 7において同定部1 1 6は、定常領域を同定する。すなわち、「生体局所定常モデル」を用いて定常領域が同定される。ステップS 1 0 8において同定部1 1 6は、因果関係を同定する。すなわち、各分子の定常状態間の因果関係が作成される。因果関係の同定には、動的構築モデルを用いることができる。

[0215] ステップS 1 0 9において同定部1 1 6は、分子マーカークの因果関係と外部環境の関係を同定する。すなわち、各分子間の因果関係と環境の関係が作成される。外部環境としては生活習慣の時系列変化の測定値を用いることができる。これらの測定値は対象者が自ら入力したり、測定機器などから入力される。

[0216] 以上の処理により、対象者の生体の細胞記憶情報と環境情報が取得される。

[0217] ステップS 1 1 0において検出部1 1 7は、DNA多型を検出する。すなわち検出部1 1 7は、ヘルスケアの対象者の持つ遺伝子の遺伝子情報として、一塩基の多型と構造の多型を検出する。

[0218] 以上のようにして対象者の履歴が得られると、次にステップS 1 1 1において検索部1 1 8は、データベース1 1 9に記憶されている情報の中から、対象者の履歴と類似する他者の履歴を検索する。すなわち上述したように、

個々人の生体分子の細胞記憶情報、DNA多型情報などの遺伝子情報、生活習慣に基づく環境情報の相関関係における類似性が判断され、最も類似する履歴が検索される。

[0219] 類似する履歴の検索についてさらに図11と図12を参照して説明する。例えば図11に示されるように、ヘルスケアの対象者の5つの血中分子A乃至Eがその目的に応じて選択され、測定されているものとする。図3における場合と同様に、図11は、5つの血中分子A、B、C、D、Eの4つの定常領域 y_{11} 乃至 y_{14} をノードとした因果関係に関するモデルを示す図である。 $Y_i(A)$ 乃至 $Y_i(E)$ は、それぞれ血中分子A乃至Eの定常領域 y_i ($i = 11, 12, 13, 14$) のノードであり、 x_{10} 乃至 x_{14} は、生体の状態を表している。状態 x_{10} 、 x_{11} は、健康状態を表しており、状態 x_{12} 、 x_{13} は、症状未発現状態を表し、状態 x_{14} は症状発現状態を表している。番号は、小さい値から大きい値の方向に向かう時間の経過を表しており、その順番に対象者の生活習慣が記録されたものである。

[0220] ノード $Y_{14}(A)$ とノード $Y_{14}(D)$ が疾患の状態を代替するマーカーであるとともに原因となっている。ノード $Y_{14}(A)$ の誘導にはノード $Y_{13}(A)$ とノード $Y_{13}(C)$ が働いており、ノード $Y_{14}(D)$ の誘導にはノード $Y_{13}(E)$ が働いている。同様に、ノード $Y_{13}(A)$ 、 $Y_{13}(C)$ 、 $Y_{13}(E)$ の誘導には、それぞれノード $Y_{12}(C)$ 、 $Y_{12}(A)$ 、 $Y_{12}(D)$ が働いている。さらにノード $Y_{12}(A)$ 、 $Y_{12}(C)$ 、 $Y_{12}(D)$ の誘導には、それぞれノード $Y_{11}(A)$ 、 $Y_{11}(B)$ 、 $Y_{11}(D)$ が働いている。

[0221] 各定常領域において細胞記憶が導入された場合を+、導入されなかった場合を-で表記すると、図11における対象者の細胞記憶の履歴は、図12の行列として示すことができる。図12における+と-は、それぞれ図11のノード $Y_{11}(A)$ 乃至 $Y_{14}(E)$ の状態に対応する。

[0222] さらに血中分子A乃至Eに関連する遺伝子の遺伝子情報として、5種類の関連分子G(A)乃至G(E)が検出されたとする。ゲノム解析の結果から、それぞれの遺伝子にn種類の多様性が比較する集団に存在する場合、集団

はゲノム情報から $n! \times 5$ （!は階乗を意味する）通りの小集団（図11において左の枠に示す集団）に分類することができる。

[0223] 類似する履歴の検索を実行する場合、先ずデータをゲノム情報に従って小集団に分類した後に、図12に示される行列の相同性から近似したデータを選抜する。最も相同性が高い場合は、行列のすべての項目が一致する場合である。部分的に近似している場合には、その大きさに従って近似性に順位が付けられる。行列のデータには生活習慣の記録が貼り付けられているので、対象者は近似した他者の生活習慣情報を参考に未来の生活習慣を選択することができる。

[0224] ステップS112において推定制御部120は、状態変化を推定制御する。すなわち、対象者の将来の生体の状態変化が推定され、予防方法を含む介入方法が提案される。

[0225] 最も単純な提案は、ステップS111の検索結果をそのまま対象者に提示することである。すなわち、ステップS111の検索処理では、対象者の履歴と類似する他者の履歴が検索される。この検索された他者が何らかの疾患あるいは疾病を有している場合には、対象者も同様の疾患・疾病を有するに至る確率が高い。そこで対象者は、例えば生活習慣を変えるなどして、環境条件を他者と異なるものとし、同様の疾患・疾病に至ることを予防することができる。

[0226] 勿論さらに進んで、対象者の健康を維持するための未来のオプションを積極的に構築し、提供するようにすることもできる。例えば、ストレス、食事習慣、運動習慣、感染履歴、治療履歴などの環境条件のうち、より因果関係の強い要素を特定し、それを変更することを提案することもできる。さらに遺伝子に関する要因も変更できるようになった場合には、その変更を提案することも可能となる。

[0227] 以上のように、ヘルスケアの対象者の細胞記憶情報、環境情報、および遺伝子情報に類似する、対象者以外の生体のそれらの情報を検索することで、対象者が健康を維持するのに適切な情報を確実に得ることが可能になる。

[0228] なお、本明細書において、フローチャートの処理を記述するステップは、その順序に沿って時系列的に行われる処理はもちろん、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理も含むものである。

[0229] また、本技術の実施形態は、上述した実施形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。例えば2以上の任意の実施形態を組み合わせることも可能である。

符号の説明

[0230] 1 生体情報処理装置, 11 選択部, 12 決定部, 13 測定部, 14 作成部, 15 分割部, 16 同定部, 17 推定制御部, 51 生体情報処理装置, 61 選択部, 62 測定部, 63 決定部, 64 分割部, 65 同定部, 66 推定制御部

請求の範囲

- [請求項1] 生体の分子の発現量を特定の時間間隔で測定する測定ステップと、
測定された時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割する分割ステップと、
前記ベースライン成分の変動、または前記周期的成分の振幅若しくは周期の変動から、前記時系列データの定常領域を同定し、さらに同定した前記定常領域間の因果関係を同定する同定ステップと、
同定された前記定常領域間の因果関係から、前記生体の状態変化を推定する推定ステップと
を含む生体情報処理方法。
- [請求項2] 前記同定ステップは、さらに生体の状態変化と外部環境との関係を同定し、
前記推定ステップは、さらに前記同定された生体の状態変化と外部環境との関係から、前記生体の疾患発症を推定する
請求項1に記載の生体情報処理方法。
- [請求項3] 前記測定ステップでは、全身状態、局所状態、および染色体状態を表象する前記分子を測定する
請求項1に記載の生体情報処理方法。
- [請求項4] 前記分子は、前記生体の血中の分子である
請求項1に記載の生体情報処理方法。
- [請求項5] 前記分子は、メタボリックシンドロームに関連する分子である
請求項1に記載の生体情報処理方法。
- [請求項6] 前記分子は、培地中の分子である
請求項1に記載の生体情報処理方法。
- [請求項7] 生体の分子の発現量を特定の時間間隔で測定する測定部と、
測定された時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割する分割部と、
前記ベースライン成分の変動、または前記周期的成分の振幅若しくは

は周期の変動から、前記時系列データの定常領域を同定し、さらに同定した前記定常領域間の因果関係を同定する同定部と、

同定された前記定常領域間の因果関係から、前記生体の状態変化を推定する推定部と

を備える生体情報処理装置。

[請求項8]

生体の分子の発現量を特定の時間間隔で測定する測定ステップと、測定された時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割する分割ステップと、

前記ベースライン成分の変動、または前記周期的成分の振幅若しくは周期の変動から、前記時系列データの定常領域を同定し、さらに同定した前記定常領域間の因果関係を同定する同定ステップと、

同定された前記定常領域間の因果関係から、前記生体の状態変化を推定する推定ステップと

をコンピュータに実行させるプログラムが記録されている記録媒体。

[請求項9]

生体の分子の発現量を特定の時間間隔で測定する測定ステップと、測定された時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割する分割ステップと、

前記ベースライン成分の変動、または前記周期的成分の振幅若しくは周期の変動から、前記時系列データの定常領域を同定し、さらに同定した前記定常領域間の因果関係を同定する同定ステップと、

同定された前記定常領域間の因果関係から、前記生体の状態変化を推定する推定ステップと

をコンピュータに実行させるプログラム。

[請求項10]

対象者の生体の生体分子の細胞記憶情報を取得する細胞記憶情報取得ステップと、

前記対象者の生体の環境条件に関する環境情報を取得する環境情報取得ステップと、

前記対象者の生体を構成する遺伝子配列に関する遺伝子情報を取得する遺伝子情報取得ステップと、

取得された前記細胞記憶情報、前記環境情報、および前記遺伝子情報に関し、類似する前記対象者以外の生体の情報を検索する検索ステップと

を含む生体情報処理方法。

[請求項11] 前記遺伝子情報は、一塩基多型および構造多型である
請求項10に記載の生体情報処理方法。

[請求項12] 前記検索は、前記細胞記憶情報、前記環境情報、および前記遺伝子情報の相関関係における類似性に基づいて行われる
請求項11に記載の生体情報処理方法。

[請求項13] 前記細胞記憶情報取得ステップは、
前記対象者の生体の分子の発現量を特定の時間間隔で測定する測定ステップと、
測定された時系列データを、周期的成分、環境刺激応答成分、およびベースライン成分に分割する分割ステップと、
前記ベースライン成分の変動、または前記周期的成分の振幅若しくは周期の変動から、前記時系列データの定常領域を同定し、さらに同定した前記定常領域間の因果関係を同定する同定ステップと
を含む請求項12に記載の生体情報処理方法。

[請求項14] 対象者の生体の生体分子の細胞記憶情報を取得する細胞記憶情報取得部と、
前記対象者の生体の環境条件に関する環境情報を取得する環境情報取得部と、
前記対象者の生体を構成する遺伝子配列に関する遺伝子情報を取得する遺伝子情報取得部と、
取得された前記細胞記憶情報、前記環境情報、および前記遺伝子情報に関し、類似する他の生体の情報を検索する検索部と

を備える生体情報処理装置。

[請求項15] 対象者の生体の生体分子の細胞記憶情報を取得する細胞記憶情報取得ステップと、

前記対象者の生体の環境条件に関する環境情報を取得する環境情報取得ステップと、

前記対象者の生体を構成する遺伝子配列に関する遺伝子情報を取得する遺伝子情報取得ステップと、

取得された前記細胞記憶情報、前記環境情報、および前記遺伝子情報に関し、類似する他の生体の情報を検索する検索ステップと

をコンピュータに実行させるプログラムが記録されている記録媒体

。

[請求項16] 対象者の生体の生体分子の細胞記憶情報を取得する細胞記憶情報取得ステップと、

前記対象者の生体の環境条件に関する環境情報を取得する環境情報取得ステップと、

前記対象者の生体を構成する遺伝子配列に関する遺伝子情報を取得する遺伝子情報取得ステップと、

取得された前記細胞記憶情報、前記環境情報、および前記遺伝子情報に関し、類似する他の生体の情報を検索する検索ステップと

をコンピュータに実行させるプログラム。