(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6758368号
(P6758368)

(45)発行日　令和2年9月23日(2020.9.23)　(24)登録日　令和2年9月3日(2020.9.3)

(51)Int.Cl.		F I		
G06F 16/90	(2019.01)	G06F	16/90	100
G06F 16/28	(2019.01)	G06F	16/28	
G01N 33/48	(2006.01)	G01N	33/48	Z
G06Q 50/22	(2018.01)	G06Q	50/22	

請求項の数 22　(全 32 頁)

(21)出願番号　特願2018-510702(P2018-510702)
(86)(22)出願日　平成28年5月9日(2016.5.9)
(65)公表番号　特表2018-527674(P2018-527674A)
(43)公表日　平成30年9月20日(2018.9.20)
(86)国際出願番号　PCT/US2016/031518
(87)国際公開番号　WO2016/183026
(87)国際公開日　平成28年11月17日(2016.11.17)
審査請求日　平成31年4月5日(2019.4.5)
(31)優先権主張番号　62/158,903
(32)優先日　平成27年5月8日(2015.5.8)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　517389757
フロージョー　エルエルシー
ＦｌｏｗＪｏ，　ＬＬＣ
アメリカ合衆国　オレゴン　アッシュランド　ウィリアムソン　ウェイ　３８５
３８５　Ｗｉｌｌｉａｍｓｏｎ　Ｗａｙ，　Ａｓｈｌａｎｄ，　Ｏｒｅｇｏｎ　９７５２０，　Ｕｎｉｔｅｄ　Ｓｔａｔｅｓ　ｏｆ　Ａｍｅｒｉｃａ
(74)代理人　100094569
弁理士　田中　伸一郎
(74)代理人　100088694
弁理士　弟子丸　健
(74)代理人　100103610
弁理士　▲吉▼田　和彦

最終頁に続く

(54)【発明の名称】データ発見ノード

(57)【特許請求の範囲】

【請求項１】

任意の特定データセットから独立したモデルに従って科学データを処理するための非一時的コンピュータ可読記憶媒体であって、前記非一時的コンピュータ可読記憶媒体は、

データ発見ノードデータ構造と、

複数のプロセッサ実行可能命令と、

を含み、

前記データ発見ノードデータ構造は、（１）反復型科学データ解析を受けるべき科学データの仕様、（２）前記反復型科学データ解析のための出力フォーマットの仕様、及び（３）前記反復型科学データ解析を制御するための複数の動作変数の仕様を含み、指定された前記動作変数は、（ｉ）前記反復型科学データ解析の一部として、指定された前記科学データに対して実行されるべきアルゴリズムの仕様、（ｉｉ）メタデータの仕様であって、指定された前記メタデータが、指定された前記アルゴリズムが指定された前記科学データに適用される条件を定義するように構成される、前記メタデータの仕様、及び（ｉｉｉ）充足変数の仕様であって、指定された前記充足変数が、前記反復型科学データ解析の一部として実行される反復の回数を制御するように構成される、前記充足変数の仕様を含み、

前記命令は、コンピュータのプロセッサによる実行時に、前記コンピュータに、前記データ発見ノードデータ構造を読み出し及び呼び出して、指定された前記動作変数に従って指定された前記科学データに対応する特定データセットに対し前記反復型科学データ解析

を実行させ、指定された前記出力フォーマットで結果を生成させる、
非一時的コンピュータ可読記憶媒体。
【請求項２】
前記命令が、前記プロセッサによる実行時に、前記読み出し及び呼び出し動作の一部として前記コンピュータに、
指定された前記動作変数をメモリにロードさせ、
ロードされた前記メタデータを前記特定データセットに対してテストさせ、
前記メタデータテスト動作に基づいて、前記特定データセットに対して実行されるべき指定されたアルゴリズムを決定させ、
決定された前記アルゴリズムを指定された前記データセットに適用してデータ解析結果を生成させ、
前記結果及び前記反復型科学データ解析についての状態を記憶するメモリ内に一時データオブジェクトを作成させ、
ロードされた前記充足変数に基づいて前記反復型科学データ解析の別の反復が必要かどうかを判断させ、
前記反復型科学データ解析の別の反復が必要であるという判断に応答して、（１）ロードされた前記充足変数が充足したという判断が行われるまで、前記メタデータテスト動作、前記指定されたアルゴリズム決定、前記アルゴリズム適用動作、及び前記別の反復判断動作を繰り返させ、（２）繰り返される前記アルゴリズム適用動作に基づいて前記一時データオブジェクトを更新させ、
前記科学データ解析の別の反復が必要でないという判断に応答して、指定された前記出力フォーマットに従って前記反復型科学データ解析の結果をワークスペースに書き込ませる、
ようにさらに構成される、請求項１に記載の非一時的コンピュータ可読記憶媒体。
【請求項３】
前記非一時的コンピュータ可読記憶媒体上に存在する複数の前記データ発見ノードデータ構造をさらに含み、複数の前記データ発見ノードデータ構造が、互いに対して異なる仕様を含み、前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
ユーザ入力に応答して、前記複数のデータ発見ノードデータ構造の中からデータ発見ノードデータ構造を選択させ、
選択された前記データ発見ノードデータ構造に対して前記読み出し及び呼び出し動作を実行させるようにさらに構成される、
請求項１または２に記載の非一時的コンピュータ可読記憶媒体。
【請求項４】
前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
ユーザ入力に応答して、選択された前記データ発見ノードデータ構造についての指定された前記動作変数のうちの少なくとも１つを定義させるようにさらに構成される、
請求項３に記載の非一時的コンピュータ可読記憶媒体。
【請求項５】
指定された前記アルゴリズムが、外部アルゴリズムを含み、前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
外部アルゴリズムプラグインインタフェースフレームワークを介して前記外部アルゴリズムを呼び出させるようにさらに構成される、
請求項１～４のいずれか１項に記載の非一時的コンピュータ可読記憶媒体。
【請求項６】
指定された前記アルゴリズムが、リモートコンピュータによって実行され、前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
リモートコンピュータプラグインインタフェースフレームワークを介して指定された前記アルゴリズムを呼び出させるようにさらに構成される、
請求項１～５のいずれか１項に記載の非一時的コンピュータ可読記憶媒体。

【請求項７】
前記科学データが、実験からの細胞データを含む、請求項１～６のいずれか１項に記載の非一時的コンピュータ可読記憶媒体。
【請求項８】
前記細胞データが、フローサイトメータによって取得される単一細胞データを含む、請求項７に記載の非一時的コンピュータ可読記憶媒体。
【請求項９】
指定された前記アルゴリズムが、集団同定アルゴリズムを含む、請求項７または８に記載の非一時的コンピュータ可読記憶媒体。
【請求項１０】
指定された前記アルゴリズムが、幾何学的ゲーティングアルゴリズムを含む、請求項７～９のいずれか１項に記載の非一時的コンピュータ可読記憶媒体。
【請求項１１】
科学データを解析するための方法であって、
データ発見ノードデータ構造をデータファイルに適用するステップであって、前記データファイルが、取得機器によって収集される科学データを含み、前記データファイルが、関連付けられるメタデータを有するステップを含み、
前記適用するステップが、
前記データ発見ノードに関連付けられる複数の動作変数及び前記データファイルに関連付けられる前記メタデータをメモリにロードするステップと、
前記メタデータが前記複数の動作変数のうちの１つによって指定されるメタデータルール基準を満たすかどうかを判断するステップと、
前記メタデータが前記メタデータルール基準を満たすという判断に応答して、
前記データファイルに関連付けられる前記科学データをメモリにロードするステップと、
前記データファイルに関連付けられる前記科学データに対して第１の解析アルゴリズムを実行するステップであって、前記複数の動作変数のうちの１つが前記第１の解析アルゴリズムを指定するステップと、
充足変数を定義する一時データオブジェクトを作成するステップと、
前記一時データオブジェクトの充足変数が、前記複数の動作変数のうちの１つによって指定される充足閾値を充足するかどうかを判断するステップと、
前記一時データオブジェクトの充足変数が前記充足閾値を充足しないという判断に応答して、（１）前記データファイルに関連付けられる前記科学データのフルセットまたはサブセットに対して前記第１の解析アルゴリズムまたは第２の解析アルゴリズムのいずれかを実行するステップであって、前記複数の動作変数のうちの１つが、前記第１の解析アルゴリズムまたは前記第２の解析アルゴリズムを生データの前記フルセットまたは前記サブセットに適用するかどうかを定義するステップと、（２）前記第１の解析アルゴリズムまたは前記第２の解析アルゴリズムの前記実行に基づいて前記一時データオブジェクトを更新するステップと、
（１）前記一時データオブジェクトの充足変数が前記充足閾値を充足するかどうかを判断するステップと、（２）前記第１の解析アルゴリズムまたは前記第２の解析アルゴリズムのいずれかを実行するステップと、（３）更新済みの前記一時データオブジェクトの充足変数が前記充足閾値を充足するまで、前記一時データオブジェクトを更新するステップと、を繰り返し実行するステップと、
を含み、
前記方法のステップが、プロセッサによって実行される、方法。
【請求項１２】
前記データ発見ノードが、データベースに保存された予め定義されたデータ構造である、請求項１１に記載の方法。
【請求項１３】

前記データ発見ノードが、ユーザによって作成される、新たに定義されたデータ構造である、請求項１１に記載の方法。

【請求項１４】

非一時的コンピュータ可読記憶媒体であって、

前記非一時的コンピュータ可読記憶媒体は複数のプロセッサ実行可能命令を含み、前記命令が、コンピュータにノードデータ構造をデータファイルに適用させることによって科学データを解析するように、前記コンピュータのプロセッサによる実行のために構成され、

前記データファイルが、取得機器によって収集される科学データを含み、前記データファイルが、関連付けられるメタデータを有し、前記適用動作が、

前記データ発見ノードに関連付けられる複数の動作変数及び前記データファイルに関連付けられる前記メタデータをメモリにロードし、

前記メタデータが前記複数の動作変数のうちの１つによって指定されるメタデータルール基準を満たすかどうかを判断し、

前記メタデータが前記メタデータルール基準を満たすという判断に応答して、

前記データファイルに関連付けられる前記科学データをメモリにロードし、

前記データファイルに関連付けられる前記科学データに対して、前記複数の動作変数のうちの１つが指定する第１の解析アルゴリズムを実行し、

充足変数を定義する一時データオブジェクトを作成し、

前記一時データオブジェクトの充足変数が、前記複数の動作変数のうちの１つによって指定される充足閾値を充足するかどうかを判断し、

前記一時データオブジェクトの充足変数が前記充足閾値を充足しないという判断に応答して、（１）前記データファイルに関連付けられる前記科学データのフルセットまたはサブセットに対して前記第１の解析アルゴリズムまたは第２の解析アルゴリズムのいずれかを実行し、前記複数の動作変数のうちの１つが、前記第１の解析アルゴリズムまたは前記第２の解析アルゴリズムを生データの前記フルセットまたは前記サブセットに適用するかどうかを定義し、（２）前記第１の解析アルゴリズムまたは前記第２の解析アルゴリズムの前記実行に基づいて前記一時データオブジェクトを更新し、

（１）前記一時データオブジェクトの充足変数が前記充足閾値を充足するかどうかとの判断動作と、（２）前記第１の解析アルゴリズムまたは前記第２の解析アルゴリズム実行動作と、（３）更新済みの前記一時データオブジェクトの充足変数が前記充足閾値を充足するまで、前記更新動作を繰り返し実行する、

ように構成される、

非一時的コンピュータ可読記憶媒体。

【請求項１５】

科学データを解析するための方法であって、

複数の動作変数の仕様を受信するステップであって、前記仕様が、（１）充足基準の仕様、（２）第１の解析アルゴリズムの仕様、（３）第２の解析アルゴリズムの仕様、ならびに（４）前記第１の及び第２の解析アルゴリズムが前記科学データに適用されるべき条件の仕様を含むステップと、

前記第１の解析アルゴリズムを指定する前記動作変数、及び前記第１の解析アルゴリズムが前記科学データに適用されるべき条件を指定する前記動作変数に基づいて、前記科学データの少なくとも一部に対して前記第１の解析アルゴリズムを実行するステップと、

前記充足基準が満たされるまで、前記実行ステップの前記結果及び前記動作変数に基づいて前記科学データの少なくとも一部に対し前記第１の解析アルゴリズムまたは第２の解析アルゴリズムを繰り返し実行するステップと、

を含み、

前記方法のステップが、プロセッサによって実行される、方法。

【請求項１６】

メモリ内のデータ発見ノードデータ構造にアクセスするステップであって、前記データ

発見ノードデータ構造が、複数のデータセットのうちのいずれかに対して実行されるべき複数の処理動作を定義するように構成され、前記データ発見ノードデータ構造が、任意の特定データセットから独立した処理のためのモデルとして構成されるステップと、
データセットを選択するステップと、
プロセッサによって、選択された前記データセットについての処理結果を生成するために、アクセスされた前記データ発見ノードデータ構造を選択された前記データセット内のデータに適用するステップと、
を含み、
前記データ発見ノードデータ構造が、（１）充足基準の仕様、（２）第１の解析アルゴリズムの仕様、（３）第２の解析アルゴリズムの仕様、ならびに（４）前記第１の及び第２の解析アルゴリズムがデータセットに適用されるべき条件の仕様を含み、
前記適用するステップは、前記プロセッサが、（１）指定された前記条件に基づいて、指定された前記解析アルゴリズムのうちのどれを実行するかを決定するステップと、（２）選択された前記データセット内のデータに関して、決定された前記解析アルゴリズムを実行するステップと、（３）決定された前記解析アルゴリズムの前記実行が、指定された前記充足基準の充足をもたらしたかどうかを判断するステップと、（４）指定された前記充足基準が充足したという判断が行われるまで、（ｉ）指定された前記条件に基づいて、指定された前記解析アルゴリズムのうちのどれを実行するかを決定するステップと、（ｉｉ）選択された前記データセット内のデータに関して、決定された前記解析アルゴリズムを実行するステップと、（ｉｉｉ）決定された前記解析アルゴリズムの前記実行が、指定された前記充足基準の充足をもたらしたかどうかを判断するステップと、を繰り返すステップと、
を含む、方法。
【請求項１７】
非一時的コンピュータ可読記憶媒体であって、
前記非一時的コンピュータ可読記憶媒体は複数のプロセッサ実行可能な命令を含み、前記命令が、コンピュータに、
複数の動作変数の仕様を受信することであって、前記仕様が、（１）充足基準の仕様、（２）第１の解析アルゴリズムの仕様、（３）第２の解析アルゴリズムの仕様、ならびに（４）前記第１の及び第２の解析アルゴリズムが科学データに適用されるべき条件の仕様を含む、ことと、
前記第１の解析アルゴリズムを指定する前記動作変数、及び前記第１の解析アルゴリズムが前記科学データに適用されるべき条件を指定する前記動作変数に基づいて、前記科学データの少なくとも一部に対して前記第１の解析アルゴリズムを実行することと、
前記充足基準が満たされるまで、前記実行ステップの前記結果及び前記動作変数に基づいて前記科学データの少なくとも一部に対し前記第１の解析アルゴリズムまたは第２の解析アルゴリズムを繰り返し実行することと、
を行わせることによって前記科学データを解析するように、前記コンピュータのプロセッサによる実行のために構成される、
非一時的コンピュータ可読記憶媒体。
【請求項１８】
前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
前記データ発見ノードデータ構造にアクセスすることであって、前記データ発見ノードデータ構造が、複数のデータセットのうちのいずれかに対して実行されるべき複数の処理動作を定義するようにさらに構成され、前記データ発見ノードデータ構造が、任意の特定データセットから独立した処理のためのモデルとしてさらに構成される、ことと、
データセットを選択することと、
選択された前記データセットについての処理結果を生成するために、アクセスされた前記データ発見ノードデータ構造を選択された前記データセット内のデータに適用することと、

を行わせるようにさらに構成される、
請求項１に記載の非一時的コンピュータ可読記憶媒体。
【請求項１９】
前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
データ解析ソフトウェアの実行セッション中にサンプルデータの解析に関するワークスペースを開かせ、
前記外部アルゴリズムへのインタフェースのためのラッパーを作成させ、
前記外部アルゴリズムの呼び出しのための、前記外部アルゴリズムの識別及び前記外部アルゴリズムのためのデータ入力の仕様を含む要求を構成させ、
サンプルデータ、前記サンプルデータについてのメタデータ、及び前記外部アルゴリズムが、前記要求に従って前記外部アルゴリズムを呼び出すために結果を前記外部アルゴリズムに保存すべき出力先を提供させ、
前記出力先において、前記外部アルゴリズムから結果を受信させ、
受信された前記結果を前記ワークスペース内に保存させ、
前記外部アルゴリズムからの前記結果を前記データ解析ソフトウェアワークスペースに統合するために、前記ワークスペースを前記データ解析ソフトウェアに復元させる、
ようにさらに構成される、
請求項１に記載の非一時的コンピュータ可読記憶媒体。
【請求項２０】
前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
外部集団ノードを作成することであって、前記外部集団ノードが、（１）取得機器により収集されるデータを表すデータセット、（２）出力フォルダ先、及び（３）前記データセットのマークアップ言語記述を含み、前記マークアップ言語記述が、（ｉ）前記データセットについてのメタ情報、及び（ｉｉ）前記データセットに到達するために使用される以前の計算の仕様を含む、ことと、
前記外部集団ノードに基づいて外部アルゴリズムを呼び出すことであって、前記外部アルゴリズムが、前記データセットに対する処理を実行するように構成され、前記出力フォルダ先が、前記外部アルゴリズムがその処理からの結果を保存するべき場所を定義する、ことと、
前記データセットに対して前記外部アルゴリズムにより実行される前記処理の結果として、前記外部アルゴリズムから結果を受信することと、
前記外部アルゴリズムからの前記結果を解析ワークスペースファイルに統合することと、
を行わせるようにさらに構成される、請求項１に記載の非一時的コンピュータ可読記憶媒体。
【請求項２１】
前記命令が、前記プロセッサによる実行時に、前記コンピュータに、
（１）取得機器によって収集される科学データに関連付けられるメタデータ、及び（２）リモートコンピュータシステムに接続するためのインタフェースに関連付けられるメタデータを読み出すことと、
前記科学データに関連付けられる前記メタデータ、及び前記インタフェースに関連付けられる前記メタデータに基づいて、生データ解析の方法を指示することと、
前記ワークスペース及び前記インタフェースの状態を保存することと、
前記リモートコンピュータシステムについてのネットワークアドレスを参照すること及び前記リモートコンピュータシステムとの接続を生成することによって、前記インタフェースを通して前記リモートコンピュータシステムと通信することと、
前記ワークスペースに関連付けられるワークスペースデータを前記リモートコンピュータシステムに送信することと、
前記リモートコンピュータシステムからデータを受信することと、
前記リモートコンピュータシステムとの前記接続を終了することと、

を行わせるようにさらに構成される、請求項１に記載の非一時的コンピュータ可読記憶媒体。

【請求項２２】

前記命令が、前記プロセッサによる実行時に、前記コンピュータに、

データ解析ソフトウェアの実行セッション中にサンプルデータの解析に関するワークスペースを開くことと、

前記ワークスペースのデータ記述を読み出すことであって、前記ワークスペースデータ記述は前記サンプルデータについてのメタデータを含む、ことと、

前記データ解析ソフトウェアに関連するプラグインインタフェースフレームワークのデータ記述を読み出すことであって、前記プラグインインタフェースフレームワークは前記コンピュータに第２のコンピュータによる動作関数の実行を呼び出すことを可能にし、前記インタフェースデータ記述は前記第２のコンピュータについてのアドレスを識別するメタデータを含む、ことと、

前記アドレスに基づいて前記プラグインインタフェースフレームワークを介して前記第２のコンピュータへの接続を確立することと、

前記サンプルデータに関して前記第２のコンピュータによる前記動作関数の実行を呼び出すために、前記接続を通して前記第２のコンピュータと通信することと、

前記呼び出しに応答して前記接続を通して前記第２のコンピュータからデータを受信することと、

受信された前記データに基づいて前記ワークスペースを修正することと、

修正された前記ワークスペースに基づいて前記ワークスペースメタデータを修正することと、

前記データ解析ソフトウェアセッション内で、修正された前記ワークスペースをメモリに保存することと、

修正された前記ワークスペースメタデータをメモリに保存することと、

前記セッションを終了することと、

を行わせるようにさらに構成される、請求項１に記載の非一時的コンピュータ可読記憶媒体。

【発明の詳細な説明】

【技術分野】

【０００１】

関連特許出願の相互参照及び優先権主張

本特許出願は、２０１５年５月８日に出願された「Ｄａｔａ　Ｄｉｓｃｏｖｅｒｙ　Ｎｏｄｅｓ」と題する米国仮特許出願第６２／１５８，９０３号の優先権を主張し、その開示全体が参照により本明細書に組み込まれる。

【背景技術】

【０００２】

序論

技術の進歩により、単一細胞実験機器は、以前の世代の機器よりもはるかに多くの情報を生成することが可能である。例えば、フローサイトメータは、各細胞毎に多数のパラメータ（例えば、１０以上のパラメータ）を有する、何千もの個々の細胞を表すデータを生成し得る。その結果、同定され得る可能性のある表現型の数が、指数関数的に増加する。言い換えると、本出願の出願前には、単一細胞アッセイによってもたらされる情報コンテンツは、実質的に増加している。さらに、単一細胞の審査が、細胞及びＤＮＡ修飾毎に何千もの転写産物（ＲＮＡ）分子の照会を含むように拡大されている。例えば、トランスクリプトーム解析全体は、１万個の遺伝子を一度に検査することになる。

【０００３】

より多くのデータを生成するほど、細胞表現型が相互に作用し、または疾患に影響を及ぼす方法、及び他の疾患関連タンパク質を発現する可能性により多くの見通しがもたらされるが、取得機器によって生成されるデータの膨大な量は驚異的であり、最前線の専門家

さえも圧倒し得る。概して、生命科学者は、自身の専門的技術を細胞機能または細胞表現型のセットまたはサブセットに集中させる。例えば、免疫学者は、疾患または免疫細胞機能を理解するために、自身の経験を一握りの細胞表現型に集中させる。一方、システム生物学者は、細胞相互作用及び遺伝子とタンパク質を互いにリンクする経路における豊富な知識を有し得る。細胞の相互作用、同定、及び機能性は、多様かつ複雑な特性の範囲を含むため、一個人が全ての細胞集団の専門家であることを期待するのは、非現実的である。生命科学者の専門的技術は、概して、全てではないくらかの細胞表現型（通常、現在知られる全ての細胞表現型の５０％より少ない）に限定されるため、各細胞表現型が疾患または細胞相互作用とどのように相関するかを専門家が詳細に知らないことから、発見及び診断解析において知識の不一致が生じる。この知識の不一致の結果として、専門家は、取得機器によって取得されたデータの研究を、専門家により強く知られる細胞表現型に集中させることがある。表現型のサブセットへの実験及び研究を限定する際に、解析者は、疾患または細胞機能に非常に大きな影響を有し得る重要な表現型を無視または見落とすことがある。さらに、既知の表現型に集中することによって、取得機器によって収集された大量のデータが、休眠状態及び未使用状態に置かれ得る。

【０００４】

細胞表現型のサブセットに基づいてデータのサブセットを解析することが、実験中の興味深い発見を導き出すことがある。しかしながら、細胞応答は、細胞が複数機能のパターンを発現することを含む場合があり、細胞表現型のサブセットのみを解析することによって、科学者は、他の細胞集団が、細胞の応答または疾患にどのように影響を与えるかを認識し損なう可能性がある。例えば、研究者は、特定の免疫応答において重要である、Ｔ細胞のサブセットを探す実験を行っている場合がある。この例では、Ｔ細胞のサブセットは、４つのパラメータ（マーカーとしても知られる）の組み合わせによって定義され得る。当然ながら、実験の発端では、研究者は、関心のあるＴ細胞のサブセットを同定するのに必要なマーカーの数を意識していない。したがって、より多くの細胞に対してより多くのマーカーを検査することによって、研究者は、罹患率または治療効果と相関する細胞サブセットを発見し得る。また、より多くのデータ解析技術を用いて、研究者は、自分自身の知識の不一致を克服して、疾患または細胞機能において重要である新たな予想外のサブセットを見つけることがある。よって、多くの研究者及び科学者によって表される知識ギャップを補償する技術に対する当該技術分野における必要性が存在する。

【０００５】

従来の技術解決策は、科学者の知識の欠如と実際の細胞応答との間のギャップを十分に埋めていないと、発明者は考えている。例えば、従来技術は、有益な解析ツールを提供することによって研究者の実験を支援し得るが、それらのツールは、データと知識の不一致を埋めるのに依然として十分ではない。従来の発見解決策では、解析者は、依然としてサンプル群に対して手動クラスタリングを実行し解析を適用しなければならない。一方、細胞表現型を検査するための９個のマーカー、メモリ状態を検査する８個のマーカー、及び細胞信号伝達を検査する８個のマーカーを有する実験の例では、可能性のあるクラスタの数は、2^{25}、即ち３３，５５４，４３２個のクラスタであり、それは、手動での解析にはあまりに多すぎるクラスタである。言い換えると、潜在的な表現型の数及び可能性のある２次元表示は、手動の解析ではうまくスケールされない。当然ながら、いくつかの表現型を枝刈りすることが、表現型空間をより管理可能な数に制限するために発生し得る。例えば、生命科学者は、ナイーブ、エフェクタ、セントラルメモリ、及びエフェクタメモリ細胞にさらに分割されるＣＤ３＋ＣＤ４５＋／－／ＨＬＡ－ＤＲ－／ＣＤ１６＋、ＣＤ４＋、及びＣＤ８＋細胞を検査するためにさらなる表現型枝刈りで単一細胞及び生きた無傷細胞に対してプレ処理ゲーティングを実行し得る。しかしながら、この表現型枝刈りされた例でさえ、発見のためにサンプル毎に１６のファイルを手動で操作する必要がある。したがって、狭い焦点を越えて発見に焦点を置いた研究において単一細胞技術を活用しようとする科学者は、困難で、非決定的かつ再現不可能な道に直面する。それゆえに、高度な技術のある専門家の介入なしに高次元データを解析し、生物学的に関連のあるデータを見

つけることが可能な、データ解析ツールを提供する必要性が当該技術分野において存在する。

【０００６】

発明の概要

上記問題を考慮して、本発明が開発された。発明者は、セッション内でリアルタイムに、任意の外部アルゴリズムを呼び出して取り入れ、上記アルゴリズムと対話するためのフレームワーク及びインタフェースを開示する。発明者は、また、再現可能で更新可能なノード及びデータ駆動型解析にこれらのノードを活用することを開示する。データ駆動型解析によって、データ自体が、解析ワークフローにおける反復及び最適化に至るアルゴリズムの選択、変数、及び提示を指示することができる。実施形態の例のこれらの２つの態様を通じて、発見または診断プロセス全体が、特定のデータセットに対して実行されてもよく、それによって、同一の発見または診断プロセス、表現型同定、及び視覚化が、将来の実験においても繰り返され、発行され、検証され、または別の研究者と共有され得るように、発見または診断プロセスを特定データセットから分離する。

【０００７】

本発明のさらなる特徴及び利点、ならびに本発明の多様な実施形態の構造及び動作は、添付図面を参照して以下で詳細に説明される。

【０００８】

本明細書に組み込まれ、かつ本明細書の一部を形成する添付図面は、本発明の実施形態を説明し、説明と共に本発明の原理を説明する役割を果たす。

【図面の簡単な説明】

【０００９】

【図１】実施形態の例についてのシステム図を示す。

【図２Ａ】実施形態の例によるプラグインフレームワーク及びアーキテクチャを示す。

【図２Ｂ】ワークスペースのＸＭＬ記述の例を示す。

【図２Ｃ】プラグインのＸＭＬ記述の例を示す。

【図３】プラグインフレームワーク及びアーキテクチャを用いてリモートコンピュータとインタフェースするための実施態様を示す。

【図４】プラグインフレームワーク及びアーキテクチャを用いて外部アルゴリズムとインタフェースするための実施態様を示す。

【図５Ａ】例示的な実施形態による結果のフィードバックを有するデータ発見ノードプロセスの高レベル表現を示す。

【図５Ｂ】データ発見ノードがどのようにナレッジベースを拡張するために使用され得るかの例を示す。

【図６】データ発見ノードによって実行されるデータ解析フローについてのライフサイクル全体を示す。

【図７】データ発見ノードを作成し、動作変数を設定及び定義するために使用されるユーザインタフェースを示す。

【図８Ａ】データ発見ノードを訓練する専門家及び専門的に訓練されたデータ発見ノードを呼び出す解析者を示す。

【図８Ｂ】データ発見ノードの専門家訓練の例を示す。

【図８Ｃ】データ発見ノードの専門家訓練の例を示す。

【図９】データ発見ノードによって表される決定木を示す。

【発明を実施するための形態】

【００１０】

図１は、システム図を示し、同様の参照番号が同様の要素を示す添付図面を参照する。図１に示されるように、データ取得機器は、取得コンピュータに接続されている。実施形態の例では、取得機器は、フローサイトメータである。ただし、フローサイトメータ以外の機器が、取得機器として使用されてもよいと理解されるべきである。しかしながら、発明者は、本明細書で説明される技術がフローサイトメトリを含む単一細胞技術に関して特

に革新的かつ有用であると考えるため、説明のために、本明細書における実施形態の例としてフローサイトメトリが使用されるものとする。

【００１１】

解析コンピュータは、インターネットを介して、サブネットを介して、イントラネットを介してなど、ネットワーク接続を通してサーバに接続され、またはインターネットを通してクラウドに接続される。いくつかの実施形態では、取得機器は、取得コンピュータに接続されてもよく、取得機器によって取得されたデータは、データを解析コンピュータに転送後、解析コンピュータ上で解析される。

【００１２】

解析コンピュータは、解析ソフトウェアを実行し、解析ソフトウェアは、テストされているサンプルのために取得機器の１つまたは複数のパラメータ（例えば、電圧、流量など）を調整することが可能である。このような解析ソフトウェアは、また、サンプルデータを取得する間初期サンプル情報を表示して、パラメータが正しく設定されているかどうかを評価するためにフィードバックをユーザに提供してもよい。解析ソフトウェアは、取得機器の製造業者に応じて変化してもよい。いくつかの実施形態では、取得コンピュータは、ほとんどのユーザインタフェース項目を含む解析ソフトウェアの簡易バージョンを実行してもよく、サーバもまた、解析ソフトウェアのあるバージョンを含む。この実施形態では、サーバが、取得コンピュータよりも多くの計算リソースを有し得るため、サーバは、重いデータ解析などの処理集約型機能を実行してもよい。

【００１３】

解析ソフトウェアは、取得機器によって解析されているサンプルの結果を示すデータ信号を取得機器から受信してもよく、または、解析ソフトウェアは、取得機器によって収集されるデータを表すデータファイルを受信してもよい。いくつかの実施形態では（例えば、取得機器がフローサイトメータであるとき）、解析ソフトウェアによって生成されるデータは、サンプル内の細胞の数、末梢血単核球細胞（ＰＢＭＣ）の数及び頻度、ＣＤ４＋Ｔ細胞の数、ＣＤ１４細胞の数、ＣＤ７＋細部の数などのうちのいずれかまたは全てを示してもよい。サンプル解析の結果は、１つまたは複数のフローサイトメトリ標準フォーマットファイル（例えば、ＦＣＳまたはＣＳＶファイル）内に含まれてもよい。取得コンピュータは、取得機器によって提供される信号及びデータに基づいて、ＦＣＳファイルを作成する。ただし、特に、取得機器がフローサイトメータではない場合は、他のファイルフォーマットが使用されてもよいと理解すべきである。解析ソフトウェアは、取得機器ＩＤ、患者ＩＤ、取得条件及びパラメータなどのものを示す、サンプルについてのメタデータをさらに生成してもよい。

【００１４】

解析コンピュータは、解析コンピュータが、解析サーバまたはサードパーティサーバなどのリモートコンピュータと通信することを可能にするインタフェースも含む。取得されたデータが転送される他のコンピュータの例として、サーバは、フローサイトメトリ解析専用のリモートサーバであってもよい。リモートサーバの実施形態では、解析または取得コンピュータは、ネットワークを介してサーバにアクセスしてもよい。解析または取得コンピュータは、サードパーティコンピュータシステムまたはサーバとも通信してもよい。解析または取得コンピュータは、集合を同定し、臨床のための追跡識別番号を含むように構成されるアルゴリズム、または取得コンピュータによって生成されるデータを解析し、もしくはデータを処理することが可能な任意の他の外部アルゴリズムなどの、サードパーティアルゴリズムを記憶し、実行してもよい。図１は、解析または取得コンピュータシステムがサードパーティアルゴリズムを記憶及び実行する状況を示しているが、サーバなどのリモートコンピュータは、サードパーティ、即ち「外部」アルゴリズムも実行してもよいと理解されるべきである。取得コンピュータは、必要性及び取得コンピュータによって実行される解析に応じて、複数のリモートコンピュータシステムと通信してもよい。

【００１５】

サーバは、プロセッサ及びメモリ、ならびにデータベースなどのデータストレージを備

える。非一時的コンピュータ可読記憶媒体（メモリなど）上に存在するプロセッサ実行可能命令は、プロセッサにより実行されて本明細書で説明されるタスクを実行してもよい。データベースは、本明細書で説明されるデータ発見ノードデータ構造を記憶してもよい。取得コンピュータは、同様に、プロセッサ及びメモリを備えてもよく、非一時的コンピュータ可読記憶媒体（取得コンピュータのメモリなど）上に存在するプロセッサ実行可能命令が、取得コンピュータのプロセッサにより実行されて、取得コンピュータのために本明細書で説明されるタスクを実行してもよい。

【００１６】

続く説明では、（１）外部ソフトウェアアルゴリズムを呼び出し、取り入れるためのプラグインフレームワーク及びインタフェース、及び（２）データ発見ノードを利用するデータ駆動型発見プロセス、を含むがこれらに限定されない、本明細書で説明される進歩性を有する技術のいくつかの様々な態様について詳しく述べる。

【００１７】

アルゴリズムプラグインフレームワーク及びインタフェース

単一細胞アッセイの研究の中で、科学者及びアルゴリズム研究者は、取得機器によって収集されるデータの解析を合理化する有用な解析アルゴリズムを生成し続けている。例えば、いくつかの外部解析アルゴリズムは、細胞集団を同定するように構成される。

【００１８】

従来、細胞集団同定は、ゲーティングと呼ばれるプロセスを通して手動で行われる。手動ゲーティングは、概して、細胞集団を同定するために、データポイントのセット（クラスタ）の周囲に円または多角形などの形状を、ユーザが手動で描くことを伴う。しかしながら、生命科学のデータ解析における進歩によって、細胞集団を同定することが可能な自動ゲーティングプログラムが生成された。さらに、プロセッサ実行されるアルゴリズムは、人間によって行われる手動の解析よりも高速かつ客観的に、細胞集団を同定し、または他の解析を行うことができるため、細胞集団同定または任意の他のデータ解析ステップのためにコンピュータプロセッサを使用することによって、人間が作り出すいかなるボトルネックまたはバイアスも除去し得る。例として集団同定アルゴリズムが与えられるが、レポートを生成し、または解析結果を視覚化するための外部アルゴリズム、ならびにＳＰＡＤＥ、ＦｌｏｗＭｅａｎｓ、及びＢｉｏｃｏｎｄｕｃｔｏｒプロジェクトの一部として提供されるアルゴリズムなどの高スループットゲノミクス及びフェノミクスデータ解析といった、取得機器によって収集されるデータを科学者が解析し解釈することを助ける、他の種類のデータ解析アルゴリズムが存在する。

【００１９】

集団同定のための外部アルゴリズムに加えて、アルゴリズムプラグインフレームワーク及びインタフェースは、外部サーバまたはリモートコンピュータシステムと通信して、オープンソースデータベースから実験データをダウンロードし、外部データベースから注釈付き実験データをダウンロードし、外部サーバまたはリモートコンピュータシステムが統計値を求めてスキャンし得るようにワークスペースデータをアップロードし、アプリケーションレベルの動作を実行し、または臨床試験のための追跡識別番号を受信してもよい。外部サーバシステムと対話するための能力は、解析ソフトウェアに解析結果の有益なプレ及びポスト処理を提供する。例えば、臨床試験を行う科学者が、試験的な識別番号を必要とする場合、アルゴリズムプラグインフレームワーク及びインタフェースは、外部サーバと通信して、確認のために臨床試験実験結果をアップロードしてもよい。

【００２０】

さらに別の実施形態では、解析ソフトウェア内部のアルゴリズムは、特定のプラットフォームにおいて区画化されてもよく、意図するコンテキスト外ではそれらをアクセス不可能にする。これらの内部であるが、その意図するコンテキスト外ではアクセス不可能なアルゴリズムの例（解析ソフトウェアがＦｌｏｗＪｏであるとき）は、増殖プラットフォームにおける多項式適合、ＦｌｏｗＪｏのコンペンセーションエディタにおける＋／－ピーク検出、またはＦｌｏｗＪｏの細胞周期プラットフォームにおけるガウシアンフィッティ

ングを含んでもよい。本明細書で説明されるアルゴリズムプラグインフレームワーク及びインタフェースは、外部アルゴリズムを解析ソフトウェアに統合するだけでなく、上述した現在の制限されたコンテキスト外の区画化された内部アルゴリズムの使用も可能にする。

【００２１】

プラグインシステムは、外部アルゴリズムが製品内で動作することを可能にしてその機能性を拡張するためのＡＰＩを提供する機構である。外部アルゴリズムは、典型的には、結果として生じるＣＬＲ／ＣＳＶファイル（各行がサンプル内のイベントに対応する）を生成することによって集団を同定するために使用され得るが、レポートまたはテーブルなどの追加的な中間生成物も生成してもよい。実施形態の例では、外部アルゴリズムは、Ｊａｖａ言語、またはＪａｖａから呼び出され得る任意の他の言語で実装され得る。外部アルゴリズムに追加するために、開発者は、ワークスペース内に新たな「集団ノード」を作成するためにＦｌｏｗＪｏ製品によって使用されるＪａｖａインタフェースを実装することとなり、それは、グラフ及び統計値を作成するためにＦｌｏｗＪｏの幾何学的にゲーティングされた集団ノードのように操作され得る。

【００２２】

図１に示すように、取得コンピュータは、取得機器によって取得されたデータの解析に有用な複数のソフトウェアプログラム及びアルゴリズムを記憶し、実行してもよい。例えば、解析ソフトウェアは、ＦｌｏｗＪｏなどの単一細胞解析プログラムを含んでもよい。サードパーティアルゴリズムは、自動集団同定プログラムまたは上述した外部サーバ関数などであるがこれらに限定されない、解析ソフトウェアを補完する処理を実行してもよい。取得コンピュータは、解析ソフトウェアの指示において外部アルゴリズムを実行してもよい。いくつかの実施形態では、取得コンピュータは、外部アルゴリズムを実行してもよく、別の実施形態では、図１に示すサーバなどのリモートコンピュータが、外部アルゴリズムを実行し、ネットワークを介して外部アルゴリズムの処理結果を取得コンピュータに提供してもよい。

【００２３】

図２は、解析ソフトウェアの処理セッション内で、外部アルゴリズム、または解析結果のプレ／ポスト処理を呼び出すための例示的なフレームワーク及びインタフェースを示す。本明細書で説明されるフレームワークは、既存の科学データ解析ソフトウェアに基づき構築され得る。例えば、解析ソフトウェアが、フローサイトメトリデータを解析するために生成されたソフトウェアである場合、フレームワークは、フローサイトメータによって集められたデータ内で細胞集団を同定するように外部アルゴリズムに要求してもよい。外部サーバ及び外部アルゴリズムと対話するためのフレームワークは、データ解析ソフトウェア内に含まれてもよい。

【００２４】

例えば、フレームワークは、Ｊａｖａなどのプログラミング言語によって定義されるような、クラス及びそれらの命令シーケンスの協働セットを含んでもよい。Ｊａｖａは、プログラミング言語の一例として与えられているが、任意の数のプログラミング言語のうちの１つが、本明細書で説明されるプロセス及びフレームワークを実行するプログラミング言語としての役割をしてもよい。複数のプログラミング言語が、本明細書で説明されるシステム及び方法を実現し得るが、Ｊａｖａは、他のプログラミング言語よりも望ましいものとなるある利点、即ち、Ｃ、Ｒ、またはウェブベースの計算エンジン言語などの他のプログラミング言語に呼び出しをかけるＪａｖａの能力を有する。科学機器によって収集されるデータの統計解析を実行する多くの外部アルゴリズムは、Ｒ言語で書かれている。よって、Ｒに呼び出しをかけるＪａｖａの能力は、Ｒで書かれた外部アルゴリズムに解析ソフトウェアを橋渡しする。当然ながら、外部アルゴリズムがＲで書かれていない場合、Ｊａｖａは、外部アルゴリズムのプログラミング言語にも呼び出しをかけてもよい。

【００２５】

フレームワークは、現在及び将来のデータ解析アルゴリズムが呼び出される機構にデー

タ値の入力セットを、解析結果の後続の処理と同様に、イベントクラスタ値、数式、視覚的図形、または幾何学的に定義された境界定義の形式で提供する。言い換えると、フレームワークは、入力データのセットを生成し、入力データを外部アルゴリズムまたは外部サーバに伝達するように２つのインタフェースのうちの１つに要求する。外部アルゴリズムの処理後、フレームワークは、解析結果を外部アルゴリズムまたはサーバから受信し、アルゴリズムの呼び出しまたはプレ／ポスト処理が表され、ファイルに保存される機構を提供する。ファイルに保存された解析結果は、ダウンストリーム統計計算、結果のグラフ化、または他のアルゴリズム（追加的な外部アルゴリズム、後続のプレ／ポスト処理、もしくは解析ソフトウェア内に含まれるアルゴリズムなど）の呼び出しのために解析ソフトウェアと統合され得る。

【００２６】

フレームワークは、また、データ解析ソフトウェア自体の外部のアルゴリズムである、統合されたアルゴリズムの呼び出しを管理する。解析ソフトウェアは、生物学者がそれを通してこれらのアルゴリズムと対話し得るインタフェースを提供する。解析ソフトウェアは、生物学者（例えば、解析が実行されるべき特定の集団を選択する）とプラグイン開発者（例えば、入力（例えば、生物学者が選択した集団のデータ値に対応するＣＳＶファイル）としてアルゴリズムが必要とするデータに対する要件、ならびに、解析に続き、プラグインインタフェースがユーザに提示するのに利用可能な出力の場所及びタイプを指定する）の双方によって提供される命令に基づく。インタフェースは、また、解析が常に階層的に正しく、生物学的に関連があり続けるように、解析における更新がそれを通して伝達されるエージェントとしての役割をする。より具体的には、フレームワークは、解析が最初に実行される際に統合済みアルゴリズムを呼び出すだけでなく、データ値の入力セットが変化するといつでも、フレームワークは統合済みアルゴリズムの再実行も行う。したがって、科学者は、データ入力の複数セットに対して迅速に解析を実行することができ、フレームワークは、入力データ値が変化し、またはユーザが実験パラメータを変更したときにはいつでも、ユーザ対話なしに統合済みアルゴリズムを呼び出し、再実行することとなる。例えば、いくつかのデータパラメータを変更することによって、統合済みアルゴリズムによって集団がどのように同定されるかが変更されてもよい。データ入力の変化に気付くと、フレームワークは、統合済みアルゴリズムを呼び出して集団を再同定し、フレームワークは、統合済みアルゴリズムによって生成された解析結果を使用する。統合済みアルゴリズムから解析結果を受信すると、フレームワークは、解析ソフトウェアによって理解されるデータフォーマットで、結果を解析ソフトウェアに提供してもよく、解析ソフトウェアは、統計解析、グラフ化、またはレポート化などの結果に対するダウンストリーム解析を実行してもよい。

【００２７】

ワークスペースがさらなる解析のために保存され再度開かれ得るように、フレームワークは、アルゴリズム統合がワークスペースとして保存されることを可能にする。

【００２８】

フレームワークは、リモートコンピュータシステムと通信するためのインタフェース、及び外部アルゴリズムと通信するためのインタフェースを含む。それぞれのインタフェースは、外部サーバ上に記憶された外部アルゴリズムまたは関数が、ユーザ対話なしに呼び出され得る手段を提供する。実際には、グラフィカルユーザインタフェースを通してデータ処理を見るユーザには、統計、グラフ、または解析ソフトウェアによって生成される他のレポートなどを通して、外部アルゴリズムによって実行される解析結果のみがユーザに示され得るため、外部アルゴリズムの呼び出しは目に見えない。

【００２９】

概して、統合されたアルゴリズムの呼び出しのためのインタフェースは、データ値の入力ファイル、出力先フォルダ、１つまたは複数の実験からのデータセットのＸＭＬ記述を含むが、これらに限定されない。このＸＭＬ記述は、生データへのポインタ、プラグイン駆動型解析を含んで実行される全ての解析、データについてのメタ情報、及びロジクル、

双指数関数、ハイパーログ、及び双極逆正弦などの、データを処理及び視覚化するために任意選択的に使用されるデータ変換を含んでもよい。ＸＭＬ記述は、解析及び関連する結果に生データを階層的にリンク付けするマークアップによってこの情報を指定するＸＭＬ文書の形式をとってもよい。図２Ｂは、ワークスペースのＸＭＬ記述の例を示し、図２Ｃは、プラグインのＸＭＬ記述の例を示す。同一データ及び解析アーキテクチャを記憶し得る独自仕様のバイナリファイルなど、ＸＭＬ以外の形式が使用されてもよいことを理解すべきである。さらに、ＸＭＬまたは別のフォーマットのいずれにせよ、データセットの記述は、任意のプラグインベースの解析用の入力パラメータに関するメタデータ及び外部アルゴリズムによって生成される任意の派生データへのポインタを含んでもよい。ＸＭＬメタ情報が外部アルゴリズムによって使用されるかどうかは、呼び出されるアルゴリズム次第である。外部アルゴリズムインタフェースは、保存され、フレームワークによって後で復元されるアルゴリズム呼び出しについてのステップも定義する。インタフェースは、グラフィック、導出パラメータ、表形式データ、ゲーティングデータ（ゲーティングＭＬフォーマットなど）、分類結果（ＣＬＲ）ファイル、ＸＭＬデータ、またはコンマ区切り値（ＣＳＶ）ファイルの形式で統合済みアルゴリズムから解析結果を受信することも可能である。言い換えると、インタフェースは、統合済みアルゴリズムによって生成される中間生成物を管理するように構成される。

【００３０】

インタフェースは、外部アルゴリズム及びサーバ関数が、外部アルゴリズムを解析ソフトウェアにプラグインするために遵守しなければならない取り決めを定義している。外部アルゴリズムインタフェース及びプレ／ポスト処理インタフェースのそれぞれが、外部サーバ上でのプレ／ポスト処理とインタフェースするため、または外部アルゴリズムとインタフェースするための取り決めを定義する。別のインタフェース実装ステップが、図３及び図４においてより詳細に示されている。

【００３１】

図３を参照すると、リモートコンピュータとインタフェースするための実装ステップが示されている。方法は、インタフェースがワークスペースを開くことで開始する。ワークスペースを開くことは、プロセッサがワークスペースのＸＭＬ及びプレ／ポスト処理インタフェースのＸＭＬを読み出すことを含む。ワークスペースＸＭＬは、各サンプルに関連付けられたメタデータ（取得されたデータ、機器タイプ、パラメータ名など）、及び取得後に追加された、任意のユーザ定義されたサンプル固有のメタデータを含み、一方、プラグインインタフェースに固有のＸＭＬは、プラグインモジュールの実行／更新のために必要な変数、例えば、データベースまたはサーバのＵＲＩを保持する。ワークスペースを読み出し、ＵＲＩを受信した結果、プロセッサは、サーバまたはそこに記憶されているデータストア（例えば、データベース）への接続を確立して、後述のように認証を開始し、クエリを実行し、データベースからのデータ及びワークスペースＸＭＬの修正を取り出す。ワークスペースを開くステップは、プロセッサによって実行されるプレ／ポスト処理インタフェースをさらに含み、データベース（例えば、指定されたデータファイルについてのメタデータ及び解析命令を含む、サンプル追跡用の実験室情報管理システム（ＬＩＭＳ））からの検索に基づいて、ＸＭＬワークスペース（解析ソフトウェアへの入力）を増強または修正する。さらに、ＸＭＬ入力は、ゲート、統計値、サンプル名、またはワークスペースＸＭＬに含まれ得る全てのものを追加するように修正されてもよい。入力が、解析ソフトウェアによって定義される定義済みスキーマに従う限り、これらの追加は、解析ソフトウェアにおいて計算及び表示を呼び出し得る。入力の検証及び十分なエラー報告が、インタフェースを通して扱われ、テスト入力に適した検証が、展開において実行される。承認も実行されてもよく、それは、解析ソフトウェアがサーバへのアクセスを有することを確認し、外部サーバがオンラインかどうかを判断し、資格証明書を交換し、または任意の他の承認ステップの形式でもたらされてもよい。ＸＭＬ増強は、プレ／ポスト処理ステップがリモートサーバによって実行されるべきということを反映するようにプロセッサがメタデータを生成することまたは変更することを含んでもよい。

【００３２】
次に、方法は、解析ソフトウェア内にワークスペースを保存する。保存ステップは、プロセッサがワークスペース及びプレ／ポスト処理インタフェースの状態を保存することを含む。プラグインは、ワークスペース内のそれ自体のＸＭＬ表現を更新してその「状態」を保持することとなり、及び／または、ＸＭＬをトラバースして、データを抽出し、及び、例えば、データベースを指定された統計値で更新するなどのアクションを実行してもよい。このステップの間、プレ／ポスト処理インタフェースは、とられた解析アクションのＳＱＬ出力またはログなどの、追加的な中間生成物を生成してもよく、プレ／ポスト処理インタフェースは、外部システムと通信する。この通信中に、インタフェースは、入力データを外部システムに提供し、収集されＭＩＦｌｏｗＣｙｔ規格に従って注釈がつけられたデータをダウンロードする、臨床トラッカーから追跡識別番号を受信する、または任意の他のプレ／ポスト処理ステップなど、データを外部システムから受信する。プレ／ポスト処理インタフェースは、サーバＵＲＬを参照してこの通信を行ってもよい。
【００３３】
外部サーバとの通信が完了した後、プロセッサは、セッションを終了し、プレ／ポスト処理インタフェースは、データベース接続などのコンピュータリソースを解放する。
【００３４】
図４を参照すると、外部アルゴリズムとインタフェースするための実装ステップが示されている。方法は、外部集団ノードを作成することによって開始し、それは、外部アルゴリズムインタフェースのためのラッパーとして定義され得る。この作成ステップ中に、プロセッサは、パラメータの設定、動作変数の設定、ファイルの命名など、外部アルゴリズムに関連するオプションでユーザにプロンプトし得るが、このユーザプロンプトステップは、任意選択であり、呼び出される外部アルゴリズムに依存してもよい。
【００３５】
次に、プロセッサは、外部アルゴリズムによって実行される計算を呼び出すために、ＸＭＬ表現を生成することによってエンジン要求を構成する。ＸＭＬ表現は、どのアルゴリズムを実行するかまたはどの視覚化を生成するか、及び例えば、ファイルパス、パラメータ数、クラスタ数、次元削減用の変数、色選択、視覚化の種類、保存用画像タイプなどの関連する入力及び引数を表す。
【００３６】
要求を構成した後、プロセッサは、外部アルゴリズムを呼び出す。外部アルゴリズムを呼び出すことは、外部アルゴリズムにＦＣＳファイル、ＦＣＳファイルと共に含まれるＸＭＬ（イベントの数、サンプルファイル名、及び集団名を含む）、及び外部アルゴリズムがその結果を保存すべき出力フォルダを提供することを含む。それに応じて、外部アルゴリズムは、その処理及び計算を実行する。外部アルゴリズムが、要求された処理及び計算を実行した後、解析ソフトウェアインタフェースは、結果を受信し、それらを解析ソフトウェアに統合する。これらの結果は、ＣＳＶファイル、ＣＬＲファイル、ゲーティングＭＬファイル、またはＦＣＳファイルの形式でもたらされてもよい。ＣＳＶまたはＣＬＲファイルをインポートするとき、ＣＳＶまたはＣＬＲの各行が、ＦＣＳファイル内のイベントに対応し、列番号が、クラスタ番号に対応する。さらに、外部アルゴリズムインタフェースは、導出パラメータを作成し、解析ソフトウェアは、導出パラメータに基づいて自動的にゲーティングして亜集団を作成する。結果を受信後、プロセッサは、アルゴリズムへの入力を修正してもよい。一実施形態では、プロセッサは、所与の出力ファイルに記憶されたデータを参照することによって、外部アルゴリズムの結果を受信する。
【００３７】
外部アルゴリズムから結果を受信した後、プロセッサは、ファイルシステムにワークスペースを保存し、解析ソフトウェアワークスペースを復元する。プロセッサは、次いで、解析ソフトウェアの指示で、追加のダウンストリーム解析を実行してもよい。
【００３８】
このようにして、外部サーバ上に記憶された外部アルゴリズム及び関数は、解析ソフト

ウェアに完全に統合されなくとも、解析ソフトウェアに利用可能である。解析ソフトウェアのユーザは、主要なワークフローの改変またはコマンドラインの知識なしに、無数のより多くの解析オプション及び機能性を得る。その代わりに、ユーザは、解析ソフトウェアのグラフィカルユーザインタフェースを使用して、サーバ上に記憶された外部アルゴリズムまたは外部関数をシームレスに呼び出し得る。

【００３９】

データ発見ノードアーキテクチャ及びプロセス

解析ソフトウェア内で、「ノード」は、幾何学ベースのツールを用いて幾何学的クラスタを定義するステップ、または取得機器によって取得されるデータに統計解析を適用するステップなど、解析ステップ全体を表す。このような「ノード」は、処理ステップ、または入力、フルセットもしくはサブセットもしくはイベントレベルの生データ、及び、細胞サブセットの幾何学的定義、または数学的モデル（例えば、細胞周期における細胞のパーセンテージ）などの出力を用いた計算を表す。言い換えると、ノードは、解析ソフトウェアが、集団同定、統計計算、数学関数、幾何学的ゲーティング、結果提示、結果増強などの解析計算を実行するように解析ソフトウェアに命令することによって作成されるデータ構造である。さらに、ノードデータ構造は、解析機能に入力するデータの仕様、及びＣＳＶファイル、ゲーティングＭＬファイルなど、結果を提示する方法を含む。データ構造は、さらに、データ入力のタイプ次第であってもよい。

【００４０】

ユーザが、データ解析アプリケーション内の「データ発見ノード」（ＤＤＮ）フレームワークを通してデータセットに対するデータ解析を指定し、実行することができるように、本明細書で説明される技術は、上述したノードの概念を拡張する。ここで、ＤＤＮフレームワークは、データ解析ソフトウェア自体に既に存在し得るいかなるインテリジェンスも越えて、幅広いナレッジベースへのアクセスをデータ解析に提供する。例えば、ＤＤＮは、また、上記で開示されたプラグインインタフェース及びフレームワークを用いて解析ソフトウェアにプラグインされる外部アルゴリズムから行われ得る決定をカプセル化することができる。アルゴリズムベースの決定は、主観的なバイアスを有する個々の解析者からデータ駆動型アルゴリズムに意思決定をシフトすることによって、解析の主観性を除去する。本明細書で説明されるデータ発見ノードアーキテクチャ及びプロセスは、また、一方向性ノードを少なくとも以下の４つの目標を達成するアクティブノードに変換する。１）アクティブノードは、繰り返される再現可能な解析が、サンプル、グループ、及び研究間の比較を提供することを可能にする（即ち、解析者の主観的バイアスに影響されない）。２）アクティブノードは、ドラックアンドドロップ機構を通して、複雑な解析及びレポート化への障壁を下げる。３）アクティブノードは、入力データ変更を更新するために生きた状態を保つ。４）ノードが、解析においてスタックされ、コマンドラインモードで実行され得るため、アクティブノードは、自動化を容易にする。

【００４１】

図５Ａは、アクティブノードアーキテクチャを表す。実施形態の例によるＤＤＮフレームワークは、以下のコンポーネント、ソフトウェア、データ構造、アルゴリズム、及びネットワークを介してアクセス可能なデータベースを含む。上述したように、データ解析アプリケーション内のＤＤＮフレームワークは、本来、ノードの各アプリケーションに基づき構築されたナレッジベースにユーザがアクセスする方法である。よって、ユーザは、ユーザインタフェースを介してジェスチャを行って、完了されるべき解析についてのＤＤＮを作成し、どのタイプの解析が完了されるべきかを指示する。ＤＤＮ（ワークスペース内で「ノード」として物理的に表される）のエンドユーザインスタンスは、ナレッジを含まないが、むしろ、それによってユーザは、自分が解析しているものより大きなコンテンツ（例えば、高齢者における参照ＣＤ３＋パーセンテージ）にプラグインすることが可能となる。例えば、

ａ．ユーザは、ゲートノードとしてＴ細胞集団を同定するためにＣＤ３＋ゲートを作成する。

ｂ．ノードは、２つの結論を有するユーザインタフェースにおいて、ユーザによりＤＤＮとして割り当てられる（ローカルクライアントにおいて、即ち、「これをＤＤＮにする」）。

ｉ．以下の集団及びサンプル情報が、ナレッジベースの物理メモリに書き込まれる。

１．「サンプル情報」

ａ．ＦＣＳ（生）ファイルに含まれるメタデータ、例えば、どの機器で、どの取得ソフトウェアによって

ｂ．サンプルコンテキスト（細胞タイプ、種）が、ＤＤＮナレッジに送信する。

２．ＤＤＮ実行パラメータ。以下にその概略を示す。

３．生物学的結果情報－解析の統計値及び数値結果

ｉｉ．ＤＤＮが、反復数ｎ＞１の場合、ＤＤＮは、「自分のデータに基づいて、このＣＤ３＋頻度は、前の観測より低い２つの標準偏差である」などの、任意のフラグをユーザに返す。

ｃ．よって、ナレッジベースは、参照を提供し、ＤＤＮは、当面の解析者と、ＤＤＮパラメータの前述の例（図５Ｂを参照）によって確立された現在のパターンに合致する、全ての以前の解析者のデータとの間に双方向の対話を提供する。ＤＤＮは、この交換を「駆動する」コンポーネント、即ち、（「生」データファイルにアクセスするプロセッサとは別々の）ホストされたネットワークを有効にするユーザ対向ノードである。

【００４２】

図５Ａ及び５Ｂに示されるように、入力データが、ＤＤＮに提供され、ＤＤＮは、結果を生成する解析ステップを実行する。ＤＤＮによって生成される、結果となるデータは、ＤＤＮにフィードバックされてもよく、または結果となるデータは、入力データを枝刈りすること、入力データからノイズを除去すること、または入力データのパラメータを変更することなどによって、入力データを変更する。結果となるデータが、多少なりとも入力データに影響を及ぼすとき、ＤＤＮは、新たなデータセットで同一の解析ステップを適用してもよく、または、ＤＤＮは、新たなデータセットに基づいて異なる解析ステップを適用してもよい。このようにして、ＤＤＮは、最初の反復の後、「データ駆動型」と考えられてもよい。

【００４３】

さらに、結果となるデータは、ダウンストリーム処理とさらなる関係を有してもよい。例えば、ＤＤＮは、集団同定アルゴリズムを表してもよく、結果となるデータは、決定的でない、または望ましくない結果を生じ得る。ノードは、結果となるデータを解析することができ、結果となるデータの解析に基づいて、ＤＤＮは、集団同定アルゴリズムのパラメータを変更して入力データ内の集団をより良好に同定することができる。別の例では、結果となるデータは、同定される表現型（例えば、ＣＤ８＋）が、罹患率または治療効果とは全く相関性を有しないと判断してもよい。結果となるデータによって、罹患率または治療効果との相関性が見つからない場合、ＤＤＮまたはＤＤＮを訓練する科学者は、今後の解析についてこの表現型を無視するようにＤＤＮに命令してもよい。このようにして、ＤＤＮは、参照される集団同定アルゴリズムを用いて、集団を最も正確に同定するように最適化する。上記の実施例によって分かるように、データ及びアルゴリズム駆動型の決定は、ＤＤＮによって行われる。ＤＤＮが受信するデータが多くなればなるほど、かつＤＤＮが処理するデータが多くなればなるほど、ＤＤＮはより多くを学習する。このデータ駆動型の方法は、以下でより詳細に説明される。

【００４４】

フレームワークが攻撃にあまり影響されないように、実施者は、ＤＤＮフレームワークにセキュリティまたはキュレーションレイヤを含めるように選択してもよいことにも留意すべきである。これによって、悪い、または訓練されていない行為者がナレッジベースを汚染すること（例えば、１００人の人が１％の誤ったＣＤ３＋頻度をゲーティングすること、及びその悪いデータを提出すること）を防止するのを助け得る。

【００４５】

図６は、ＤＤＮによって実行されるデータ解析フローについてのライフサイクルを示す。図６によって示されるプロセスでは、四角で示される方法ステップは、プロセッサによって実行されるアクションステップを表し、ひし形で示される方法ステップは、プロセッサによって実行されるテストステップまたは判断ステップを表し、楕円形で表される方法ステップは、テストステップの可能性のある結果を表す。

【００４６】

高レベルで、図６に表される方法は、３つのフェーズ、第１のＤＤＮサイクル、次のｎ回ＤＤＮサイクルフェーズ、及び完了フェーズを含む。第１のＤＤＮサイクルフェーズは、１回だけ実行されるが、次のｎ回ＤＤＮサイクルは、充足基準が満たされるまで反復し続けてもよい。充足基準が満たされた後にのみ、方法は、完了フェーズに入ることとなる。

【００４７】

ＤＤＮ関数を定義及び制御するデータオブジェクトの種類は、ここでは、図６で示される方法がどのように動作するかをより良く理解するために説明されるものとする。これらのデータオブジェクトは、動作変数、一時オブジェクト、ポインタ、メタデータ、及び生のリストモードデータを含む。

【００４８】

第１に、ＤＤＮは、動作変数データオブジェクトを含む。動作変数は、ユーザまたは解析ソフトウェアのいずれかによって設定される変数であり、１）充足変数閾値、２）メタデータルール、及び３）解析ソフトウェアアルゴリズムまたは指定データについて実行される動作の仕様を含む。充足変数は、ＤＤＮサイクルが完了すると考えるために充足されるべき、使用によって設定される閾値であってもよい。メタデータルールは、入力によって充足されるべき基準を定義する。例えば、メタデータルールは、入力データが生データのメタデータ内にＣＤ４パラメータを示すということを指定してもよい。解析ソフトウェアアルゴリズムまたは指定された動作は、外部アルゴリズム、解析ソフトウェア内に含まれる数学関数、またはＦｌｏｗＪｏの多変量グラフ化、ＦｌｏｗＪｏのレポート生成、幾何平均生成、集団同定、または解析ソフトウェアもしくはプラグインされた外部アルゴリズムによって提供される任意の他の関数などの、解析ソフトウェア内に含まれる任意の他の関数を含んでもよい。

【００４９】

図７は、ＤＤＮを作成し、動作変数を設定及び定義するために使用されるユーザインタフェースを示す。最初に、ユーザは、ファイルを選択し、発見するためのジェスチャをする。ファイルは、取得機器から収集されたデータのセットであってもよく、取得コンピュータ内のディスクドライブに保存されていてもよい。このジェスチャは、ユーザが選択されたファイルにＤＤＮを適用したいということを解析ソフトウェアに通知する。ジェスチャは、ユーザがファイルを右クリックすること、キーボードのショートカットを使用すること、グラフィカルユーザインタフェース内のアイコンをクリックすること、またはプロセッサによって理解される任意の他のジェスチャを含んでもよい。発見のためのジェスチャの後、ユーザは、新規のＤＤＮを訓練するか、またはデータベースもしくは他のファイル記憶コンテナに保存されているＤＤＮを適用するかのいずれかを選択し得る。ユーザが、データベースからＤＤＮを呼び出すことを選択する場合、取得コンピュータは、ＤＤＮデータ構造を記憶しているデータベースに呼び出しをかけ、保存済みのＤＤＮのリストを提示し、ユーザがＤＤＮのうちの１つを解析用に選択できるようにする（図示せず）。ユーザが、新規のＤＤＮを訓練することを選択する場合、取得コンピュータは、グラフィカルユーザインタフェースを通して、ＤＤＮを定義することとなる動作変数のリストを提示する。

【００５０】

図７は、選択用の例示的な動作変数のセットを示しているが、本開示は、図７に示される動作変数に限定されない。動作変数は、パラメータ、特徴、反復変数、及び範囲変数などのセットにグループ化されてもよいが、より多くの動作変数のグループが定義され、ユ

ーザインタフェース内に提示されてもよい。例えば、ユーザは、前方散乱光（ＦＳＣ）、側方散乱光（ＳＳＣ）、蛍光１（ｆｌ１）、蛍光２（ｆｌ２）、蛍光３（ｆｌ３）、蛍光ｎなどであるがこれらに限定されないパラメータから選択してもよい。パラメータ選択は、単一細胞解析において重要な役割を果たし、ＤＤＮは、それが適用されるデータの種類、即ち「実行パラメータ」に加えて、それ自体の動作についてのメタデータを含む。選択されるパラメータの例は、以下のものを含んでもよい。

ａ．細胞表現型が定義されたパラメータ。例としてフローサイトメトリを用いると、散乱パラメータは、サイズ及び粒度の相対的な測定値であり、例えば血液中の主要な細胞サブセットを同定するのに有用であり、一方、蛍光パラメータは、生体分子の測定値である。このように、パラメータは、本来、相互交換不可能であり、ＤＤＮについて選択されたレベルで使用されるパラメータ及びその階層は、解析の再現性を容易にする生物学的に関連のある情報である。

ｂ．集団を同定するために使用される任意のアルゴリズムについての種類及び入力変数に関するパラメータ、例えば、ゲーティング及び解析情報（頂点、隣接集団の位置、ゲートタイプ、集団特性（凸状、希少など）、集団名、ゲートが描かれたパラメータ、親ゲート（オントロジー）、集団を同定するために使用されるアルゴリズム）。

ｃ．一連のアルゴリズム及び計算についての階層型解析の種類の数（及び、したがって、動作の次数）

【００５１】

このようにして、研究者は、ＣＤ８＋Ｔ細胞の集団を指定し、それは、ＣＤ３＋の子であり、生きている、及びリンパ球の幾何学的に定義されたゲートである、ＣＤ８蛍光パラメータと対比するＣＤ４に対して実行されるｋ－ｍｅａｎｓクラスタリングアルゴリズム（ここで、ｋ＝３が入力変数であった）によって同定される。ＤＤＮは、この情報をナレッジベースへ及びナレッジベースから送信することを可能にする。

【００５２】

ユーザインタフェースは、同様にこれらのパラメータの名前を変更する能力をユーザに与える。ユーザは、また、ＤＤＮによって処理されるべきデータの量を制限するために、これらのサイトメータプリセットパラメータのうちのいずれかを除外してもよい。ＤＤＮは、解析するためのパラメータの選択、解析するための特徴（ピーク、谷、または範囲など）、反復するかどうか、及びどの範囲で解析するかを受信する。これらの、及び潜在的に他の動作変数を選択した後、コンピュータは、新規ＤＤＮを作成し、それはまた、ＤＤＮデータベースに保存されることとなる。作成されたＤＤＮは、データを解析し、結果、または解析ソフトウェア内に含まれるか、もしくはプラグインインタフェース及びフレームワークを通して解析ソフトウェアにアクセス可能な任意の他の関数を生成する準備ができる。

【００５３】

ＤＤＮをセットアップするために、プロセッサは、入力データの選択を受信し、それは、イベントのセット、または等価性のいくらかの暗黙の感覚を有するファイルのセット（例えば、複数の時点にわたって取り込まれたＣＤ３測定値）である。入力データは、単一のサンプルまたはサンプルのグループであってもよい。入力データの選択後、プロセッサは、入力データに応じて利用可能な解析の種類を判断してもよい。一旦ＤＤＮデータベースがセットアップされると、第１のステップは、「専門家」に、参照セットを作成するためにサンプル情報及び実行パラメータの両方でナレッジベースにシード値を与えさせることである。上記の例を続けると、高齢者の患者からのＣＤ３＋データが、専門家によって定義される。非専門家は、「新規」サンプルに対してＤＤＮを作成し、ＤＤＮは、サンプル及び実行パラメータを比較して、それが、専門家主導の解析を再作成できるかどうかを検査する。合致が存在すると、それは、生物学的な結果情報－現在の測定値対ナレッジベースの比較を行う。ナレッジベース内に情報を構築することによってＤＤＮの「訓練」が、使用とともに生じ、したがって、ナレッジベースのＣＤ３＋部分への各クエリが、新たな生物学的結果情報を既知の範囲のプールへ保管する。この２段階アプローチは、（１）

解析が、適用され実行され得ること、及び（２）参照データのナレッジベースと比較され得ることを検証する。

【００５４】

言い換えると、ＤＤＮが計算及び実行し得るものは、入力データに依存する。一例では、ＣＤ４イベントがロードされたメタデータ内に存在するかどうかを、プロセッサが判断して、プロセスが、選択されたデータに対してＣＤ４集団同定アルゴリズムを実行し得るかどうかを判断してもよい。

【００５５】

図７は、パラメータとして蛍光１、ピーク特徴、固定値の反復変数２、及び範囲変数として２～９８のパーセンタイルという、例示的なユーザ選択を示す。ユーザが動作変数を設定した後、ユーザインタフェースは、選択されたファイル下の作成済みデータ発見ノードを表示する。ユーザは、将来の参照用にデータ発見ノードの名前を変更してもよいが、例示のために、図７では、単に「Ｄｉｓｃｏｖｅｒｙ　Ｎｏｄｅ」という名前の、作成済みデータ発見ノードを示す。データ発見ノードについてのこれらの例示的な選択は、ゲーティングツリーと同等であり、ゲーティングツリーもまた、図７の右下の角に示されている。したがって、図７に示される例示的な動作変数の選択は、ゲーティングツリーと同等である。

・Ｃｏｍｐ－ＡＰＣ－Ａｘ７００－Ａ　ｓｕｂｓｅｔ　これは、通常手動で定義されるサブセットである。この例では、ＤＤＮは、上記で概説したそのパラメータによって、ナレッジベースからの情報をアルゴリズム的に使用してこの集団を同定し、ピーク検出（集団同定のための別のアルゴリズム的方法）を実行し、次いで、その順序で、子の亜集団に対して統計値の計算を呼び出す。

○ピーク１

■幾何平均：ＣＤ３（Ｃｏｍｐ－ＡＰＣ－Ａｘ７００－Ａ　ｓｕｂｓｅｔ）　ユーザは、解析アプリケーションツールを用いてＣｏｍｐ－ＡＰＣ－Ａｘ７００－Ａサブセット集団の幾何平均を計算している。図７の右下にある図は、この解析の階層及びユーザへの表示を示している。

■中央値ＣＤ３（Ｃｏｍｐ－ＡＰＣ－Ａｘ７００－Ａ　ｓｕｂｓｅｔ）　上記幾何平均についてと同様であるが、この場合は中央値についてである。

○ピーク２

■幾何平均：ＣＤ３（Ｃｏｍｐ－ＡＰＣ－Ａｘ７００－Ａ　ｓｕｂｓｅｔ）

■中央値ＣＤ３（Ｃｏｍｐ－ＡＰＣ－Ａｘ７００－Ａ　ｓｕｂｓｅｔ）

【００５６】

再び図６を参照すると、動作変数に加えて、ＤＤＮは、最初の計算後に一時データオブジェクトを生成する。一時データオブジェクトは、ＤＤＮ進行オブジェクトを表す。一時データオブジェクトは、計算の反復数及び充足変数を少なくとも含み得る。計算の反復数は、ＤＤＮによって実行される追加計算毎にインクリメントし、充足変数は、次のＮ回ＤＤＮサイクルフェーズのサイクル中の充足変数の状態を示す。例えば、充足変数は、充足変数閾値が満たされたか、または充足変数閾値を超えたかを示してもよい。充足変数閾値と各反復において作成されるＤＤＮ作成の一時データオブジェクトとの比較を通して、これらのデータオブジェクトは、ＤＤＮがステートフルに保持されることを可能にする。

【００５７】

ポインタは、一意な識別子であり、ＤＤＮがそのシーケンスについてアクセスするワークスペース内の１つまたは複数のノードを指し示す。それについては以下でさらに説明される。ポインタは、メタデータ及び生のリストモードデータを含むファイルの位置を指し示し、それは、ＤＤＮの動作にとっても重要である。

【００５８】

ＤＤＮにとって重要なメタデータは、２つの異なるタイプの参照ノートからもたらされる。第１に、メタデータは、データの特定のサブセットを得るために専門家によって行われる決定によってもたらされてもよく、それは、概して、専門家によって定義されるゲー

トの形式である。データのサブセットは、階層型ゲートによってもたらされてもよい。特定の例では、前回のゲートのＸＭＬ階層は、ＤＤＮデータ構造により使用するためのメタデータにおいて表されるコンテキスト情報を提供する。専門家の決定の代わりに、メタデータは、生物学的に有意義である染色名（「ＣＤ３－ＦＩＴＣ」）についてのパラメータを含む、親ＦＣＳファイルからのキーワードメタデータを含んでもよい。メタデータは、生データに関連付けられ、生データに関連付けられたメタデータは、解析されるべき生データのソースであるＦＣＳファイルのヘッダ及びノード名も含んでもよい。

【００５９】

最後に、生のリストモードデータは、イベント／細胞毎に収集されたｎ個のパラメータについての生のイベント／細胞レベルデータを含む。

【００６０】

図６に示される方法は、上述された全てのデータオブジェクトを使用する。ＤＤＮ法／ライフサイクルは、第１のＤＤＮサイクルフェーズによって開始する。第１のＤＤＮサイクルフェーズでは、プロセッサは、動作変数をメモリにロードする。続いて、プロセッサは、上述したメタデータをロードする。プロセッサがＤＤＮフローを通して解析されるべきファイルをロードする前に、プロセッサは、メタデータ及び動作変数をロードし、動作変数は、テスト用のルール及び変数を定義する。

【００６１】

動作変数及びメタデータのロード後、プロセッサは、メタデータルール動作変数（複数可）に対してメタデータをテストして、メタデータが、ＤＤＮの基準を満たすかどうかを判断する。例えば、メタデータルール動作変数は、ユーザによって設定されるキーワードメタデータ、ＦＣＳファイルによって設定される表現型メタデータ、染色識別メタデータ、または取得機器によって生成されるファイル内に含まれる任意の他のメタデータのいずれかを通して、ＣＤ４細胞が存在することを示すメタデータパラメータを指定する。

【００６２】

動作値に対するメタデータのテストは、緩やかなモード、中程度モード、及び厳密モードなどの複数のモードを有してもよい。

【００６３】

緩やかなモードは、メタデータ要件を全く有しなくてもよい。緩やかなモードでは、ＤＤＮは、メタデータの値に関わらず実行されることになる。例えば、緩やかなモードでは、ＤＤＮは、提供されるリストモード生データ中の２点間の極小値を計算し、次いでＤＤＮは、生データをメモリ内にロードさせ、計算を呼び出し、ユーザに表示されるべきワークスペースに統計値を追加することによって完了することとなる。

【００６４】

中程度モードでは、合致するメタデータの閾値がユーザによって設定され、例えば、ＤＤＮについての６個のパラメータ中の３個が設定される場合、それがデータスペース内の細胞集団を同定するのに十分なパラメータを有しているため、ＤＤＮを実行する。

【００６５】

厳密モードでは、ＤＤＮの実行を開始するために、全てのメタデータ要件が満たされなければならず、プロセッサは、生データをメモリ内にロードせず、ＤＤＮ方法が停止し、それ以上の計算は実行されない。

【００６６】

メタデータは、メタデータルール動作値の基準を満たすか、動作値によって設定される基準を満たさないかのいずれかである。メタデータが動作値の基準を満たさない場合、プロセッサは、生データをメモリ内にロードせず、ＤＤＮ方法が停止し、それ以上の計算は実行されない。メタデータが動作値の基準を満たす場合、プロセッサは、生データをメモリ内にロードする。メモリ内にロードされる生データは、生取得データ、別のノードからのデータ、１つもしくは複数のゲートからのデータ、または解析ソフトウェアにアクセス可能な任意の他の生データの形式でもたらされてもよい。

【００６７】

生データのロード後、プロセッサは、動作変数によって指定される計算またはアルゴリズムを実行する。例えば、プロセッサは、本明細書で説明されるプラグインアーキテクチャ及びフレームワークを用いて外部アルゴリズムを実行して、生データ内の１つまたは複数の集団を同定してもよい。さらに、プロセッサは、上述したＤＤＮ一時オブジェクトを作成する。ＤＤＮ一時オブジェクトを作成することは、プロセッサが反復変数を開始番号に設定すること、及び実行された計算またはアルゴリズムの結果に基づいて充足値を定義することを伴う。ＤＤＮ一時オブジェクトの作成後、第１のＤＤＮサイクルフェーズは完了し、プロセッサは、次のｎ回ＤＤＮサイクルフェーズの実行を開始する。

【００６８】

次のｎ回ＤＤＮサイクルフェーズにおいて、フェーズは、ＤＤＮ一時オブジェクトをロードすること、及びＤＤＮ一時オブジェクトの充足値が充足閾値もしくは動作変数によって設定される充足基準を満たすか、または超えるかを判断することによって開始する。ＤＤＮ一時オブジェクトを充足閾値と比較することは、プロセッサが、反復変数をＤＤＮの充足変数と比較することを含んでもよい。例えば、充足変数が、５回反復するようにＤＤＮに命令し、一時オブジェクトの反復変数が５より小さい場合、充足変数は満たされておらず、ＤＤＮは再び反復することとなる。別の例として、プロセッサは、ＤＤＮ一時オブジェクトまたは任意の他の動作変数が次の計算のための「指示」を指定したかどうかを判断してもよい。例えば、ＤＤＮ一時オブジェクトによって指定される指示は、メモリ内の生データのサブセットのみが次の反復において使用されるべきであることを示してもよい。別の例として、充足値は、カテゴリ内のイベントのパーセンテージを定義することなどによって、精度を示す値を含んでもよく、プロセッサは、精度の数値を充足基準と比較してもよい。精度の数値の例は、散乱ゲートの推定純度及び回収率の３カラーフローの解析を含んでもよい。ここでは、純度及び回収率の最良の組み合わせに到達するまで、散乱ゲートが再定義され得る。純度効果及び回収率効果の値が９０％を超えるまで、最適化ループは、全てのサンプルに適用されるゲートを縮小及び拡大する。

【００６９】

ＤＤＮ一時オブジェクトの充足変数が、充足閾値もしくは充足基準を満たすか、または超える場合、プロセッサは、完了フェーズを実行する。

【００７０】

ＤＤＮ一時オブジェクトの充足変数が、充足閾値もしくは充足基準を満たさないか、または超えない場合、プロセッサは、メモリにロードされた生データのサブセット、またはメモリにロードされた生データのフルセットを、一時オブジェクトが次の反復用に必要とするかどうかを判断する。上記を振り返って、動作変数は、データのサブセットまたはデータのフルセットに対して計算またはアルゴリズムを実行するかどうかを示してもよい。例えば、動作変数は、データのセットが、外部アルゴリズムを使用してゲーティングされるべきであり、ダウンストリーム数学計算が、ゲーティングされたデータに対してのみ実行されるべきであることを示していてもよい。データまたは生のリストモードデータのメタデータを解析してどの計算またはアルゴリズムを適用するかを判断するように、メタデータがプロセッサに命令してもよいことに留意すべきである。メタデータは、指定された計算またはアルゴリズムの実行前にプロセッサによって実行されるべき分岐または決定木を呼び出してもよい。例えば、生データがＣＤ４イベントを示唆するようにプロセッサが生データを解析する場合、プロセッサは、ＣＤ４集団同定アルゴリズムを適用してもよいが、生データがＣＤ８イベントを示唆するようにプロセッサが生データを解析する場合、プロセッサは、ＣＤ８集団同定アルゴリズムを適用してもよい。

【００７１】

動作変数が、データのフルセットを指定する場合、プロセッサは、指定された計算またはアルゴリズムを生データのフルセットに対して実行し、プロセッサは、反復変数をインクリメントすること、及びデータのフルセットに対して実行された計算またはアルゴリズムの結果に基づいて充足値を再定義することによって、一時オブジェクトを更新する。これらのフェーズの間、データのフルセットがメモリ内に残っていてもよい。一時オブジェ

クトの更新後、プロセッサは、新たな一時オブジェクト値に基づいて次のｎ回ＤＤＮサイクルを繰り返す。

【００７２】

動作変数が、データのサブセットを指定する場合、プロセッサは、指定された計算またはアルゴリズムを生データの指定されたサブセットに対して実行し、プロセッサは、反復変数をインクリメントすること、及びデータのサブセットに対して実行された計算またはアルゴリズムの結果に基づいて充足値を再定義することによって一時オブジェクトを更新する。データの指定されたサブセット内に含まれないデータは、メモリから解放され、その他の場所に記憶されてもよい。一時オブジェクトの更新後、プロセッサは、新たな一時オブジェクト値に基づいて次のｎ回ＤＤＮサイクルを繰り返す。

【００７３】

次のｎ回ＤＤＮサイクルフェーズは、充足閾値もしくは基準を満たすかまたは超えるまで続く。一旦満たすかまたはと、プロセッサは、完了フェーズへと続き、完了フェーズでは、プロセッサは、動作変数によって指定される出力タイプを判断する。反復オプションでは、ユーザは、ＤＤＮ実行パラメータとして記憶される、反復数を設定してもよい。この判断に基づいて、プロセッサは、ワークスペースにおけるアクションをとり、ＤＤＮフローの結果をワークスペースファイルに書き込む。例えば、プロセッサは、結果及びとられるアクションに応じて複数の視覚化のうちの１つを提示してもよく、またはプロセッサは、ワークスペースファイル内の新たな集団もしくは統計値を定義してもよい。

【００７４】

完了フェーズで発生する、ワークスペースでとられるアクションは、新たな入力データでＤＤＮを再呼び出しすることを伴ってもよい。例えば、完了フェーズ中に生成される出力は、新たな入力データセットであってもよい。入力データセットが変化すると、ＤＤＮは、再度処理を呼び出し実行してもよい。したがって、入力データセットが変化したときにはいつでも、ＤＤＮは、その必要な処理を実行してもよい。

【００７５】

図８をここで参照すると、任意の実験、臨床試験、調査、研究プロジェクトなどにおいて、専門家の数は制限される。即ち、ある人が研究分野、トピック、細胞表現型、科学的特性などについてより多くを知っているほど、それらの専門家の存在はより少なく、専門家の時間は制限される。一方、高度な技能と知識を有し得るが、専門家が所有する豊富な知識に欠ける解析者は、より一般的かつ豊富である。専門家の不足及び解析者の豊富さに起因して、専門家は、概して、実験の運用などのいくつかのタスクを解析者に委任し、専門家は、解析者の作業成果物を監督する。一方、専門家は、自分がレビューするあらゆる実験解析から全ての解析ステップをレビューする時間が単純に不足しているため、従来の方法では、専門家が、幾何学的ゲートがどのように適用されるかなど、実験及び解析のそれぞれの個々のステップを見ることができなかった。

【００７６】

専門家活用の従来の方法とは対照的に、図８は、解析者が、専門的に訓練された解析フローを呼び出し、取得されたデータのセットに展開し得るように、専門家によりＤＤＮを訓練するプロセスを示している。上述のように、専門家は、ＤＤＮの動作データ構造を設定することによって、及び保存された一時オブジェクトを通してＤＤＮにより得られるナレッジを使用することによって、訓練をＤＤＮデータ構造にもたらし得る。階層型ゲーティングの形式でのような専門家の決定は、ＤＤＮのメタデータ内に保存され、表され得る。図８は、専門家が自分自身の専門技術及び経験を用いてデータ発見ノードを訓練することを示す。訓練プロセスは、図６に示されるステップのうちのいくつか、または全てを含んでもよい。専門的に訓練されたＤＤＮは、解析フローの一部、または解析フロー全体を表してもよい。例えば、専門的に訓練されたＤＤＮは、専門家の知識に基づく的確な幾何学的ゲーティング技術を適用してもよい。代替的に、ＤＤＮは、集団同定用の外部発見アルゴリズムに呼び出しをかける解析ステップを含んでもよく、専門的に訓練されたＤＤＮは、専門家によって提供される発見プロセスのために特定のパラメータを提供してもよい

。専門家がＤＤＮを訓練し、フローステップを指定し、集団同定についての制限を指定し、任意の数学的モデルを指定しているため、ＤＤＮは、解析者が解析にもたらし得るいかなるバイアスも除去する。本明細書で論じるＤＤＮフレームワーク及びプロセスでは、解析者のバイアスは除去され、ＤＤＮを用いて行われる全ての実験は、同じように行われることとなり、それによって均一な結果がもたらされる。

【００７７】

図８Ｂは、専門家がどのようにＤＤＮを訓練し得るかについての例を示している。この例では、専門家は、ＣＤ４ゲートが幅広いほど、良好な解析結果を生じることに気が付いていてもよい。専門家は、コンピュータ上のユーザインタフェースを用いて、自身のＤＤＮにおいてＣＤ４ゲートの定義を広げてもよく、それは、専門家においてＣＤ４集団を検査すること、解析においてより多くのＣＤ４＋細胞を含むように範囲ゲートを編集することによって行われる。ＤＤＮの調整後、調整済みのＤＤＮは、データベース内の保存されたものを取得する。解析者は、ＤＤＮが異なるＣＤ４ゲート定義を有することを知ることなく、調整済みのＤＤＮを呼び出してもよい。調整済みのＤＤＮを呼び出すことによって、調整済みＤＤＮによって定義される解析フロー全体が、解析ソフトウェアの単一セッション内で発生することとなる。調整済みＤＤＮは、調整済みの方法に従って結果を生成してもよい。この方法の多くの利益のうち、実質的な利益は、解析者が異なるアクションを全く行わなくても、調整済み解析方法が専門家によって完全に検証されているということが分かることである。

【００７８】

別の利益として、ＤＤＮは、グループまたは個人間で共有されてもよい。Ｔ細胞の専門家が、ＮＫ細胞の専門家によって作成され最適化されたＤＤＮを取り出して、ＮＫ細胞についての解析を実行してもよい。このように、専門的技術が、専門家の間で共有されてもよく、実験が、多数の表現型に対して効率的に実行されてもよい。

【００７９】

アクティブノードの主な利益のうちの１つは、ノードが、特定のデータセットから分離され、データ駆動型であるということである。データが解析フローを駆動するため、利用可能になる解析のタイプは、入力データの選択に応じて異なることとなる。言い換えると、ＤＤＮが計算及び実行し得るものは、入力データ次第である。概して、入力データは、科学データを表すイベントのセット、または等価性の暗黙の感覚を有するファイルのセットである。例えば、入力データは、複数時点にわたって取り込まれたＣＤ３測定値であってもよい。別の例として、入力データは、取得機器によって取り込まれた生データであってもよい。さらに別の例では、入力データは、解析ソフトウェアまたは外部アルゴリズムによって生成される結果となるデータであってもよい。

【００８０】

ＤＤＮのメタデータは、また、制約、分岐、決定木、自己最適化、またはリアルタイムでの反復を適用するかどうかを指定してもよく、それは、ユーザによって指定され、ＤＤＮ実行パラメータとして記憶される。入力解析ステップは、ＤＤＮが解析ステップ後にフィードバックされる結果データを受信するときに何度も発生してもよい。分岐し、制約を適用し、決定木を適用するかどうかなどは、ＤＤＮのメタデータまたは充足変数内で設定されてもよい。

【００８１】

ＤＤＮが制約を適用すると、ＤＤＮは、データの範囲を絞る。例えば、絞られるべき入力データが、単一のパラメータ分布であった場合、制約は、１から１００の範囲に及ぶイベントなどの範囲であり得る。範囲を絞ることによって、ＤＤＮは、デブリであるか、または著しいノイズを付加し得る、極端なビン内の細胞を除外することができる。ＤＤＮのコンテキストにおける別の制約の適用は、低白血球数またはＨＩＶ　Ｔ細胞の逆位などの、サブセットの頻度または２つの表現型の比率を計算するためにノイズを除去することであり、患者のＴ細胞型の比率が「反転」する。例えば、制約は、制約されたデータのサブセットに対してのみ計算を実行するように動作変数を設定することによって適用されても

よい。

【００８２】

ＤＤＮが分岐を適用すると、結果が後続の実行ステップに影響を及ぼすワークフローにおいて、ＤＤＮはポイントを生成する。簡単な例として、ＤＤＮがＣＤ３＋サブセットを見つけようとしているが、ＤＤＮがＣＤ３＋イベントが存在しないと判断する場合、その情報は、プロセス内で使用され、したがって適応的にダウンストリーム解析をリダイレクトし得る。この例では、ＤＤＮは、ＣＤ３＋細胞を探索するために集団同定アルゴリズムを適用してもよい。ＤＤＮは、ＣＤ３＋細胞が全く見つからなかったと識別するクラスタ集団結果を受信してもよい。ＤＤＮは、図５のフィードバックループを表す集団同定アルゴリズムの結果を解析してもよく、ＣＤ３＋細胞についてのレポートを生成するステップが無用であると判断してもよい。したがって、ＤＤＮは、その代わりに、新たな集団を同定するように集団同定アルゴリズムに要求してもよい。上述のＨＩＶ逆位の例では、ＤＤＮが、ＤＤＮによってロードされたメタデータを使用してＨＩＶ逆位の状況を検出する場合、ＤＤＮは、より綿密なＴ細胞数のレポートまたはＴ細胞数が正常範囲内であったというレポートを行うように、解析ソフトウェアに命令してもよい。分岐ステートメントの使用は、セッション中の処理を変更し、それによって、適応的実行及びインメモリデータの両方を活用することが可能となる。動作変数は、完了フェーズ中でこの種の分岐を指定してもよい。代替的に、メタデータは、データのフルセットまたはサブセットのいずれかに適用される、指定された計算またはアルゴリズムを変更する固有の分岐を含んでもよい。

【００８３】

ＤＤＮは、特定の成果を見つけるための処理フロー全体の表現である、決定木を適用してもよい。例えば、図９は、特定の集団を見つけるための決定木の例を示しており、これはイベントまたは次数削減を伴う。図９に関するいくつかのコンテキストについて、サンプルデータセット内の特定の種類の異常を検出するためのいくつかの専門的なパネルが、既に存在している。これらのパネルが表す特定表現型の組み合わせは、マーカーを含むデータから取り出され得る。ＤＤＮは、このロジックのためのコンテナとして供給され得る。この構成は、パネルが主として不必要となる場所にポイントが到達するように、パネルベースの解析からアルゴリズム型解析への飛躍を可能にする。複数のパネルをテストするよりもむしろ、１つが、１つのチューブ内への関連するマーカーの全てを含んでもよく、それは、このようなテストから生じる多数のデータパラメータを通してナビゲートするために、ＤＤＮの高度な処理能力が使用され得ることを意味する。

【００８４】

図９を参照すると、入力データは、ゲートまたはファイルの集合であってもよい。図９から分かるように、ＤＤＮは、ＣＤ４５＋ＳＳＣデータが利用可能かどうかを判断してもよい。ｙｅｓの場合、ＤＤＮは、データを解析して、突き出した「不鮮明な」ＣＤ４５のピークがあるかどうかを判断する。順番に、ＤＤＮは、以下の比較及び解析を実行する。

ａ．図９における２つの分岐は、第１に解析が完了され得るかどうかを検査するために、ＤＤＮが実行する検証のプロセスを示す。（ＤＤＮ実行パラメータとの比較、この場合、サンプルがＳＳＣ及びＣＤ４５パラメータを含むか？）

ｂ．そうであれば、ナレッジベースからの専門家のゲートが、ＳＳＣ及びＣＤ４５パラメータによって同定される集団に適用される。

ｃ．ピーク検出（集団同定）アルゴリズムは、（既にゲーティングされたＣＤ４５＋集団に対して）ＣＤ４５の不鮮明なピークがあるかどうかを見るためだけに、ＣＤ４５パラメータを検査して実行される。

ｉ．ピークが存在する場合、この場合は急性単球白血病（ＡＭＬ）の芽細胞を同定するように、別の専門家の一連の階層型ゲートが適用される。

ｄ．ともかく、ＣＤ１９＋細胞は、リンパ球上のＣＤ１９＋頻度が異常に高い（ＤＤＮ作成後に専門家によって定義されるような２つの標準偏差よりも大きい）かどうか検査するためにナレッジベースと比較されるＤＤＮ適用された集団定義によって同定される。Ｃ

Ｄ１９頻度が、リンパ球上で異常に高い場合、ＤＤＮは、Ｋａｐｐａ／Ｌａｍｂｄａ、ＣＤ１０、ＣＤ２２などのようなＢ細胞固有情報についての解析を適用することとなる。

【００８５】

あらゆる場合に、ＤＤＮは、検証を実行し（実行される解析であり得る）、解析フェーズを実行する（例えば、これらの専門家が定義した幾何学的ゲートを適用する、またはピーク検出を実行する）、生物学的結果と比較し、繰り返してもよい。このようにして、ＤＤＮは、解析を指示するためにその３つの情報タイプを活用する。

【００８６】

図９における非限定的な例から見られ得るように、ＤＤＮは、決定木内の各点おける判断結果に基づいて、処理を変更し得る。処理の変更は、ユーザ呼び出し後、例えば、検証基準が満たされないときに、ＤＤＮのメタデータ及び動作変数によって表されてもよい。さらに従来の方法とは対照的に、ＤＤＮは、解析ソフトウェアの単一セッション内で全てのこれらの決定及び結果を処理するため、図９に表される決定木は、人間による主観的なバイアスを除去する。

【００８７】

ＤＤＮは、また、最適化技術を用いていくつかの解析「パス」にわたって結果を精密化してもよい。最適化の一例が、解析ソフトウェアが細胞分裂周期のフェーズ内の細胞数の推定を計算する、細胞周期適合解析である。分裂周期における正確な細胞数は、計算において見つかった数を反復して精密化するために最も良く見つけられる。精密化及び最適化は、複数のパスを必要とし、ＤＤＮは、正確な結果を計算するのに必要な「パス」の数についての制限をユーザが設定することを可能にする。制限は、反復数であってもよく、閾値デルタを使用してもよい。それによって、計算における正確性の改善が、改善閾値を超えなければならず、さもなければプロセスは終了する。細胞周期適合解析は、同定技術がもはや改善閾値デルタを超えるまで同定アルゴリズムが反復して表現型を決定し得る、集団同定に拡大され得る。プロセッサは、最適化技術に基づいてＤＤＮメタデータを変更してもよい。

【００８８】

さらに、ＤＤＮは、反復を用いて、各ステップ後に次数またはパラメータ範囲を削減させながら処理を繰り返してもよい。例えば、ＤＤＮは、データ範囲の最小値または最大値から始まって解析することによってデータの分布における全てのピーク（最大値）を見つけてもよい。第１のピークが見つかると、ＤＤＮが２番目以降の最大ピークなど、より多くのピークを見つけることができるように、ＤＤＮはデータセットからピークを除去する。最終的に１つのピークのみが残ることとなり、ＤＤＮが最後のピークを見つけた後、反復が停止する。反復は、充足変数内に含まれる反復変数によって定義されてもよい。

【００８９】

最後に、ＤＤＮは、他の類似のＤＤＮから学習される訓練及びナレッジを活用してもよい。ＤＤＮが、専門家によって作成されると、ＤＤＮは、類似のＤＤＮデータ構造についてのデータベースに問い合わせするように構成される。ＤＤＮは、そのメタデータ内の類似の名前または類似の項目を探索することによってこの問い合わせを行ってもよい。例えば、ＤＤＮが、それをＣＤ４同定ノードとして同定するメタ情報を有する場合、ＤＤＮは、類似のまたは同一のメタデータを有するＤＤＮデータベースに保存されている他のＤＤＮを探索してもよい。ＤＤＮは、任意の意味論的方法を通して類似のＤＤＮを見つけてもよい。類似のＤＤＮを見つけると、新たに訓練されたＤＤＮは、以前に作成されたＤＤＮによって得られるナレッジ及び訓練をＤＤＮが受信することを可能にする、データベースに保存された類似のＤＤＮから情報を得てもよい。例えば、類似のＤＤＮが、幾何学的ゲート、即ちゲートの最小／最大範囲、ゲートについてのパーセンタイル、または臨床的に有意義な結果を生成することを助ける数学的関係を専門的に定義していることを、新たに作成されたＤＤＮが見つけてもよい。各ＤＤＮは、データに適用された回数を他のＤＤＮデータ構造に伝達してもよい。上述したように、取得されたデータにＤＤＮがより多く適用されるほど、ＤＤＮが生成する結果はより良好である。したがって、より多くのデータ

に適用されているＤＤＮは、範囲、パーセンタイル、ゲート、数学的関係、パラメータの枝刈り、または任意の他の重要なナレッジを、他の類似のデータ構造に伝達してもよく、その結果、類似のデータ構造は、「より古い」ＤＤＮの訓練を活用し得る。ＤＤＮは、呼び出しを通して、かつ、データベース内の他の類似のＤＤＮデータ構造との通信をも通して学習し、したがって、最適な、例えば、集団同定をもたらすために、専門家のネットワーク及び反復型実験を活用する。さらに別の例では、ＤＤＮは、やり方を変更し、または取得機器によってデータを収集するやり方への変更を提案してもよい。

【００９０】

ＤＤＮは、コンピュータのメモリ内で、及びメモリに記憶された入力データに対して動作する。ユーザがＤＤＮを使用するジェスチャをすると、ＤＤＮは、必要な入力データをメモリに集め、メモリ内の入力データに対してデータ処理を実行する。ＤＤＮを反復し、制約を適用し、決定を行い、分岐または最適化を行うにつれて、データは削減され枝刈りされ得る。ＤＤＮが多くのインテリジェンスを得るにつれて、ＤＤＮは、入力データに対して最初のプレ処理を行い、その結果、メモリに記憶されるデータ量が最小化される。メタデータレベルで発生する、データのプレ処理を行うことによって、ＤＤＮが訓練され続けるにつれてコンピュータの性能が向上する。さらに、主観的にバイアスされた手動の幾何学的ゲーティングのステップを除去することによって、以前の実験方法よりも高速にユーザに結果が提示される。取得コンピュータ、解析コンピュータ、またはサーバは、追加的な処理を実行してＤＤＮの特徴全てを実行してもよいが、ＤＤＮの使用で効率は向上する。

【００９１】

ＤＤＮは、また、ユーザに結果を提示するための、解析ソフトウェア内に含まれるテーブルエディタまたはレイアウトエディタを活用してもよい。いくつかのコンテキストにおいて、ＤＤＮは、解析フロー全体をカプセル化してもよく、その結果、任意の他のステップが解析ソフトウェアを通して実験結果とともに提示されることなく、解析者などのユーザが、単にＤＤＮを呼び出し得る。このようにして、ＤＤＮは、実験全体を含み得る。

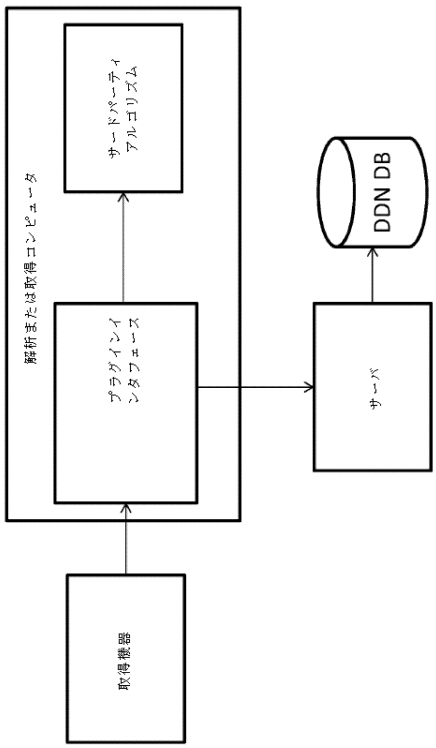
【００９２】

前述の観点から、本発明の複数の利点が実現され達成されることが分かる。

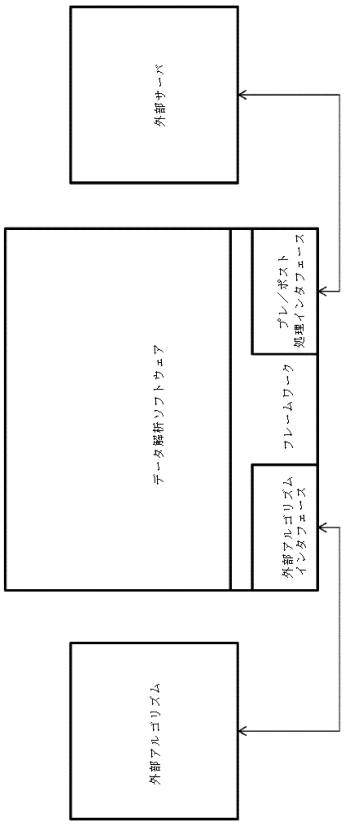
【００９３】

実施形態は、本発明の原理、及び多様な実施形態において、かつ考えられる特定の用途に適している多様な修正を用いて、それにより当業者が本発明を最もよく利用可能にするためのその実際の適用を最もよく説明するために選択され説明された。多様な修正は、本発明の範囲から逸脱することなく本明細書で説明され例示される構築及び方法において行われ得るため、前述の説明に含まれ、または添付図面に示される全ての事項は、限定ではなく、例示として解釈されるものとすることを意図している。

【図１】



【図２Ａ】



【図２Ｂ】

ワークスペースＸＭＬの例

カプセル化したデータソース、解析、結果、視覚化、及びプラグイン入力

```
<Workspace>
    <Groups>
        … analysis
    </Groups>
    <SampleList>
        <Sample>
            … data source reference
            <Population>
                … analysis results
                <Plugin>
                    … plugin inputs
                </Plugin>
            </Population>
        </Sample>
    </SampleList>
    <TableEditor>
        … tabular report generation
    </TableEditor>
    <LayoutEditor>
        … visualizations and batch generation specifications
    </LayoutEditor>
</Workspace>
```

【図２Ｃ】

プラグインＸＭＬの例

プラグインがその作業を実行する必要がある全ての入力を保存する

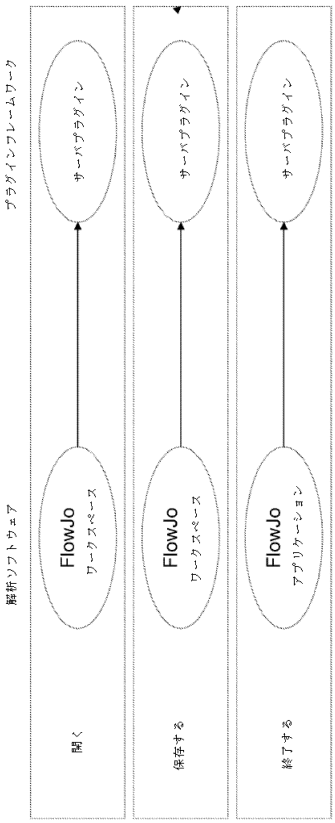
E.g. DownSample

```
<Sample>
    <SampleNode>
        <Subpopulation>
            <ExternalPopNode>
                <DownSample numEvents="8000" version="1.0"/>
```

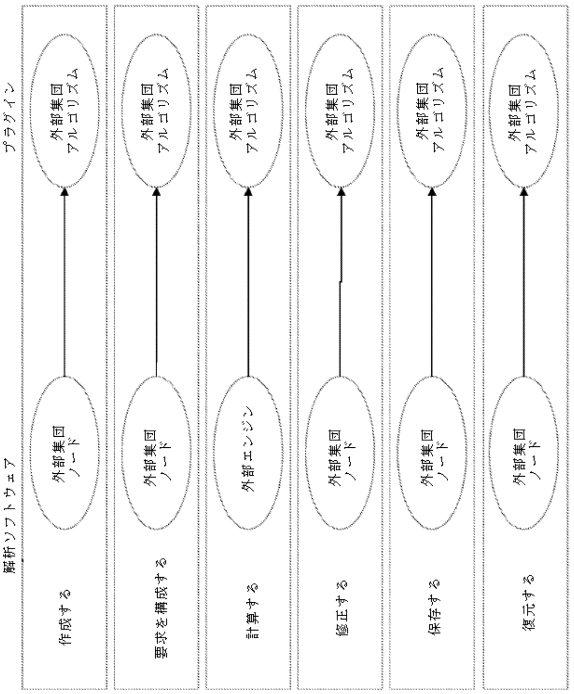
E.g. TSNE

```
<Sample>
    <SampleNode>
        <Subpopulation>
            <ExternalPopNode>
                <TSne version="1.0"/>
                    <Option NumberOfIterations="1000"/>
                    <Option Perplexity="20"/>
                    …
                    <Parameter name="Rh103di"/>
                    <Parameter name="Xe131Di"/>
                    …
```

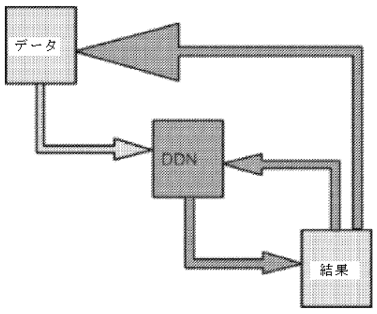
【図３】


プラグインフレームワーク
解析ソフトウェア
FlowJo
ワークスペース
サーバプラグイン
FlowJo
ワークスペース
サーバプラグイン
FlowJo
アプリケーション
サーバプラグイン
開く
保存する
終了する

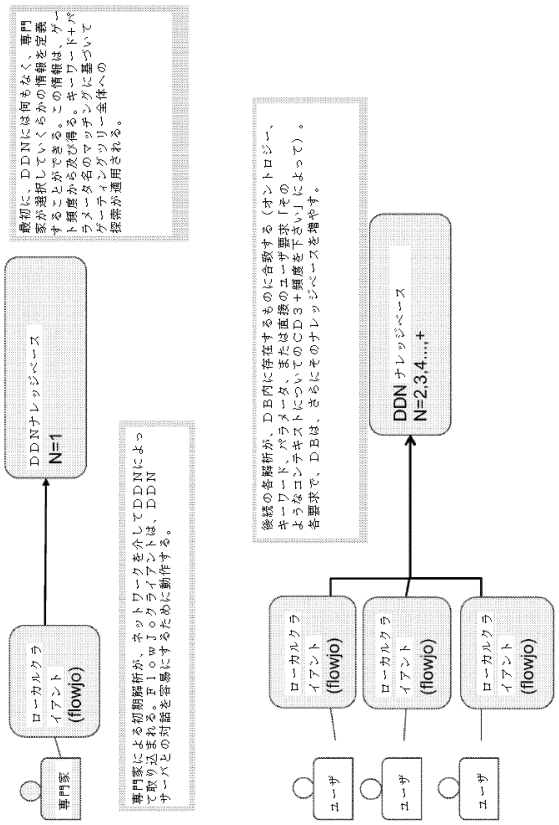
【図４】


プラグイン
解析ソフトウェア
外部集団
ノード
外部集団
アルゴリズム
作成する
要求を構成する
外部エンジン
計算する
修正する
保存する
復元する

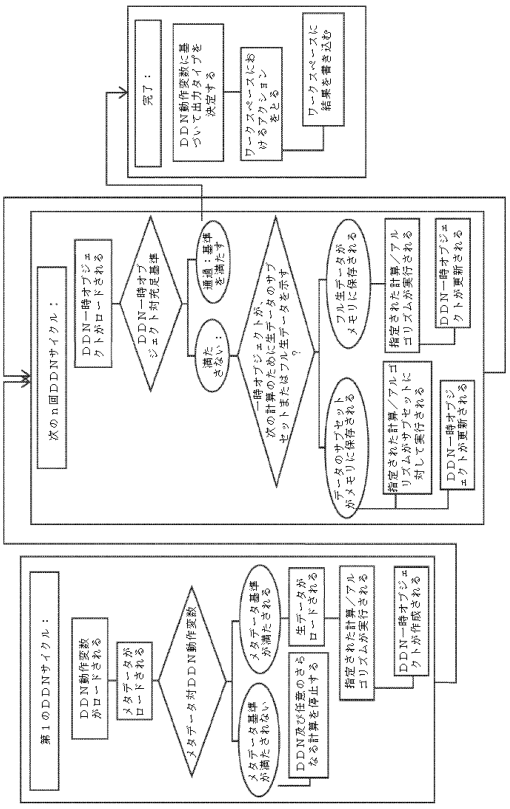
【図５Ａ】


データ
DDN
結果

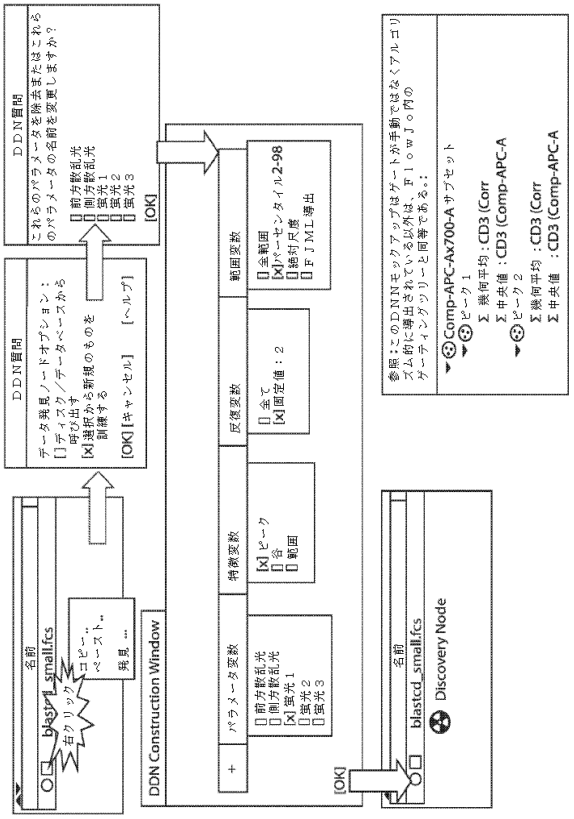
【図５Ｂ】


最初に、ＤＤＮには何もなく、専門家が選択していくらかの情報を定義することができる。この情報は、ゲート頻度から及び得る。キーワード＋パラメータ名のマッチングに基づいてゲーティングツリー全体への探索が適用される。
専門家による初期解析が、ネットワークを介してＤＤＮによって取り込まれる。ＦｌｏｗＪｏライアントは、ＤＤＮサーバとの対話を容易にするために動作する。
ＤＤＮナレッジベース
Ｎ=1
ローカルクライアント
(flowjo)
専門家
後続の各解析が、ＤＢ内に存在するものに合致する（オントロジー、キーワード、パラメータ、または直接のユーザ要求「そのようなコンテキストについてのＣＤ３＋頻度を下さい」によって）。各要求で、ＤＢは、さらにそのナレッジベースを増やす。
ＤＤＮ ナレッジベース
Ｎ=2,3,4…,+
ローカルクライアント
(flowjo)
ローカルクライアント
(flowjo)
ローカルクライアント
(flowjo)
ユーザ
ユーザ
ユーザ

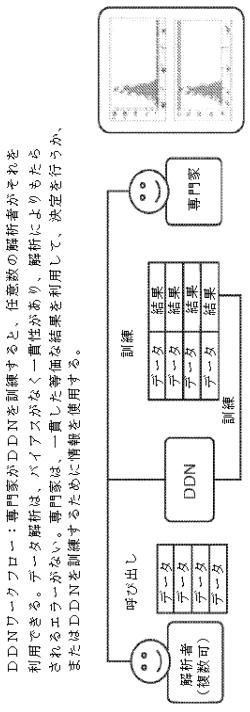
【図６】


第１のＤＤＮサイクル：
ＤＤＮ動作変数がロードされる
メタデータがロードされる
メタデータ対ＤＤＮ動作変数
メタデータ基準が満たされない
メタデータ基準が満たされる
ＤＤＮ及び任意のさらなる計算を停止する
生データがロードされる
指定された計算／アルゴリズムが実行される
ＤＤＮ一時オブジェクトが作成される
次のｎ回ＤＤＮサイクル：
ＤＤＮ一時オブジェクトがロードされる
ＤＤＮ一時オブジェクト対充足基準
満たさない：
通過：基準を満たす
一時オブジェクトが、次の計算のために生データのサブセットまたはフル生データを示す？
データのサブセットがメモリに保存される
フル生データがメモリに保存される
指定された計算／アルゴリズムがサブセットに対して実行される
指定された計算／アルゴリズムが実行される
ＤＤＮ一時オブジェクトが更新される
ＤＤＮ一時オブジェクトが更新される
完了：
ＤＤＮ動作変数に基づいて出力タイプを決定する
ワークスペースにおけるアクションをとる
ワークスペースに結果を書き込む

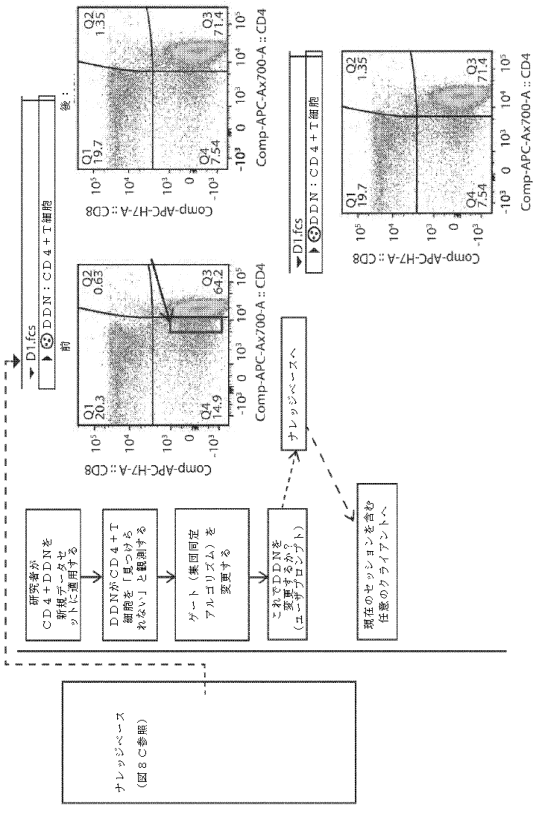
【図７】


DDN質問
データ発見ノードオプション：
[ ] ディスク／データベースから呼び出す
[x] 選択から新規のものを訓練する
[OK] [キャンセル]
[ヘルプ]
DDN質問
これらのパラメータを除去またはこれらのパラメータの名前を変更しますか？
前方散乱光
側方散乱光
蛍光１
蛍光２
蛍光３
[OK]
名前
blastcd_small.fcs
右クリック
コピー…
ペースト…
発見…
DDN Construction Window
パラメータ変数
前方散乱光
側方散乱光
蛍光１
蛍光２
蛍光３
特徴変数
ピーク
谷
範囲
反復変数
全て
固定値：２
範囲変数
全範囲
パーセンタイル2-98
絶対尺度
ＦＪＭＬ導出
[OK]
名前
blastcd_small.fcs
Discovery Node
参照：このＤＮＮモックアップはゲートが手動ではなくアルゴリズム的に導出されている以外は、ＦｌｏｗＪｏ内のゲーティングツリーと同等である。：
Comp-APC-Ax700-A サブセット
ピーク１
Σ 幾何平均：CD3 (Corr
Σ 中央値：CD3 (Comp-APC-A
ピーク２
Σ 幾何平均：CD3 (Corr
Σ 中央値：CD3 (Comp-APC-A

【図８Ａ】


ＤＤＮワークフロー：専門家がＤＤＮを訓練すると、任意数の解析者がそれを利用できる。データ解析は、バイアスがなく一貫性があり、解析によりもたらされるエラーがない。専門家は、一貫した等価な結果を利用して、決定を行うか、またはＤＤＮを訓練するために情報を使用する。
解析者（複数可）
呼び出し
データ
データ
データ
データ
DDN
訓練
データ
結果
データ
結果
データ
結果
データ
結果
訓練
専門家

【図８Ｂ】


D1.fcs
DDN：CD４＋T細胞
前
後：
Comp-APC-H7-A :: CD8
Comp-APC-Ax700-A :: CD4
Q1 20.3
Q2 0.63
Q3 64.2
Q4 14.9
Q1 19.7
Q2 1.35
Q3 71.4
Q4 7.54
研究者がCD４＋DDNを新規データセットに適用する
DDNがCD４＋T細胞を「見つけられない」と観測する
ゲート（集団同定アルゴリズム）を変更する
これでDDNを変更するか？（ユーザプロンプト）
ナレッジベースへ
現在のセッションを含む任意のクライアントへ
ナレッジベース（図８Ｃ参照）

【図８Ｃ】

ナレッジベースの内容：

１．サンプル情報

サンプルメタデータ

日付：10-OCT-2008

システム :Windows XP 5.1

サイトメータ :LSRII

ファイル:B1 .fcs

ファイル：URI:file:/Users/mikestadnisky/Deskto

$BEGINDATA: 3584

$BEGINSTEXT: 0

$BTIM: 11:52:08

$BYTEORD: 4,3,2,1

$CYT: LSRII

$DATATYPE: F

$DATE: 10-OCT-2008

$ENDDATA: 2231503

$ENDSTEXT: 0

$ETIM: 11:52:32

$FIL: B1 .fcs

パラメータを含む：

FSC-A
FSC-H
FSC-W
SSC-A
SSC-H
SSC-W
Time
Comp-FITC_Alexa 488-A :: CD56
Comp-PE-A :: CD11c
Comp-PE-Cy55-A :: CD3
Comp-PE-Cy7-A :: CD45RO
Comp-CD3 CD14 Live_Dead-A :: 7AAD
Comp-APC_Ax 647-A :: 2H2
Comp-APC-Ax700-A :: CD4
Comp-APC-H7-A :: CD8

２．ＤＤＮ実行パラメータ

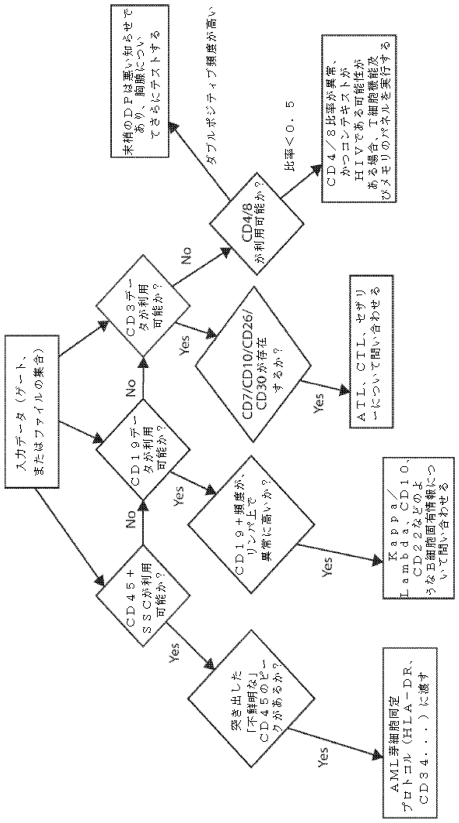
解析の階層（次数）、解析アルゴリズム、パラメータ

名前	統計値	#セル	グラフ仕様	ゲート座標	ゲーティングパラメータ	実換設定
B1.fcs		37132	Pseudocolor of FSC-A() vs. FSC-W(null):to(TNL			<Transformation>...
Singlets	88.7	32950	Histogram* of Comp-PE-Cy55-A(null):to(TNL	[36122, 87710] [251...	FSC-A vs. FSC-W	
CD3+T cells	71.3	23479	Pseudocolor of Comp-APC-Ax700-A() vs. Comp-APC	[1448, 0.55] [97848,...	Comp-PE-Cy55-A	
Q1: CD4-, CD8+	17.6	4125	Pseudocolor of Comp-APC-Ax700-A() vs. Comp-APC	2[-, -] [3814, 2293]	Comp-APC-Ax700-A...	
Q2: CD4+, CD8+	1.75	412	Pseudocolor of Comp-APC-Ax700-A() vs. Comp-APC	3[3814, -] [-, 2293]	Comp-APC-Ax700-A...	
Q3: CD4+, CD8-	74.8	17555	Pseudocolor of Comp-APC-Ax700-A() vs. Comp-APC	1[3814, 2293] [-, -]	Comp-APC-Ax700-A...	
Q4: CD4-, CD8-	5.91	1387	Pseudocolor of Comp-APC-Ax700-A() vs. Comp-APC	0[-, 2293] [3814, -]	Comp-APC-Ax700-A...	

３．生物学的結果情報－参照統計値

系統サブセット統計値	親のシングレット頻度	親のシングレットCD3＋T細胞頻度	親のシングレット／CD3＋T細胞Q1：CD4－，CD8＋頻度	親のシングレット／CD3＋T細胞Q2：CD4＋，CD8＋頻度	親のシングレット／CD3＋T細胞Q3：CD4＋，CD8－頻度	親のシングレット／CD3＋T細胞Q4：CD4－，CD8－頻度
B1.fcs	88.7	71.3	17.6	1.75	74.8	

【図９】



フロントページの続き

(74)代理人 100067013
弁理士 大塚 文昭
(74)代理人 100086771
弁理士 西島 孝喜
(74)代理人 100109070
弁理士 須田 洋之
(74)代理人 100109335
弁理士 上杉 浩
(74)代理人 100120525
弁理士 近藤 直樹
(74)代理人 100139712
弁理士 那須 威夫
(74)代理人 100196612
弁理士 鎌田 慎也
(72)発明者 マチェイ シム
アメリカ合衆国 オレゴン アッシュランド リット ウェイ ４３２
(72)発明者 ジェイ アルマロード
アメリカ合衆国 オレゴン レバノン ピー．オー．ボックス ２６９２
(72)発明者 マイケル ディ． スタドニスキー
アメリカ合衆国 オレゴン アッシュランド モラダ レーン ２３９７

審査官 吉田 誠

(56)参考文献 特開２０１２－１４１８４７（ＪＰ，Ａ）
特表２０１６－５３１３６７（ＪＰ，Ａ）
特開２００６－２７７４２４（ＪＰ，Ａ）
特表２０１１－５１３８４１（ＪＰ，Ａ）
米国特許出願公開第２０１２／０２１５４８１（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １６／００－１６／９５８
Ｇ０１Ｎ ３３／４８
Ｇ０６Ｆ １６／２８
Ｇ０６Ｑ ５０／２２