(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2004-508019
(P2004-508019A)
(43)公表日　平成16年3月18日(2004.3.18)

(51)Int.Cl.[7]	F I		テーマコード（参考）
C12N 15/09	C12N 15/00	A	4B024
C12M 1/00	C12M 1/00	A	4B029
C12Q 1/68	C12Q 1/68	A	4B063
G01N 33/53	G01N 33/53	M	
G01N 37/00	G01N 37/00	102	

審査請求 未請求　予備審査請求 有　（全 78 頁）　最終頁に続く

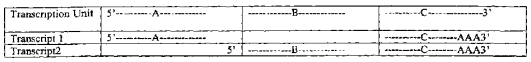
(21)出願番号　特願2002-516365(P2002-516365)
(86)(22)出願日　平成13年7月20日(2001.7.20)
(85)翻訳文提出日　平成15年1月27日(2003.1.27)
(86)国際出願番号　PCT/IB2001/001903
(87)国際公開番号　WO2002/010449
(87)国際公開日　平成14年2月7日(2002.2.7)
(31)優先権主張番号　60/221,607
(32)優先日　平成12年7月28日(2000.7.28)
(33)優先権主張国　米国(US)
(31)優先権主張番号　60/287,724
(32)優先日　平成13年5月2日(2001.5.2)
(33)優先権主張国　米国(US)

(71)出願人　503036623
コンピュジェン　インコーポレイテッド
アメリカ合衆国，　ニュージャージー州　０８８３１，　ジェイムズバーク，　センター　ドライヴ　７　スイート　９
(74)代理人　100103816
弁理士　風早　信昭
(74)代理人　100120927
弁理士　浅野　典子
(72)発明者　ショシャン，　アヴィ
イスラエル，　８２０００　キリャト　ガト，　ラハヴァト　ハー　シナイ　２／４

最終頁に続く

(54)【発明の名称】トランスクリプトームの中に場所を占めるＲＮＡ転写物及びスプライス変異体を検出するためのオリゴヌクレオチドライブラリー

(57)【要約】
本発明はトランスクリプトームの中に場所を占めるＲＮＡ転写物及びＲＮＡスプライス変異体であって対応するゲノム中に場所を占める遺伝子又は転写単位から転写されるＲＮＡ転写物及びＲＮＡ転写物及びＲＮＡスプライス変異体を検出することができるオリゴヌクレオチドライブラリーを提供する。本発明はオリゴヌクレオチドライブラリーから生成されるオリゴヌクレオチドアレイ及び様々なオリゴヌクレオチド検出システム及び発現プロファイリング研究にオリゴヌクレオチドライブラリーを用いる方法をも提供する。オリゴヌクレオチドの一種であり、ここで開示するオリゴヌクレオチドに基づくアンチセンス分子及び二本鎖干渉ＲＮＡも提供される。



【特許請求の範囲】
【請求項１】
トランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームがゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項２】
前記トランスクリプトームがヒトのトランスクリプトームである請求項１記載のオリゴヌクレオチドライブラリー。
【請求項３】
前記トランスクリプトームがラットのトランスクリプトームである請求項１記載のオリゴヌクレオチドライブラリー。
【請求項４】
前記トランスクリプトームがマウスのトランスクリプトームである請求項１記載のオリゴヌクレオチドライブラリー。
【請求項５】
組織起源のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが組織起源のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項６】
病理学的組織起源のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが病理学的組織起源のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項７】
前記病理学的組織起源ががん組織である請求項６記載のオリゴヌクレオチドライブラリー。
【請求項８】
発生段階のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが発生段階のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項９】
病気に罹患している患者のトランスクリプトームの中に場所を占めるメッセンジャーＲＮ

Ａを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームが病気に罹患している患者のゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。

【請求項１０】

前記病気ががんである請求項９記載のオリゴヌクレオチドライブラリー。

【請求項１１】

上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項１記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。

【請求項１２】

上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項５記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。

【請求項１３】

上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項６記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。

【請求項１４】

上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項８記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。

【請求項１５】

上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項９記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。

【請求項１６】

以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：

請求項１のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。

【請求項１７】

前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項１６記載の方法。

【請求項１８】

前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項１６記載の方法。

【請求項１９】

以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：

請求項５のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのス

プライス変異体の総レベルを決定する。
【請求項２０】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項１９記載の方法。
【請求項２１】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項１９記載の方法。
【請求項２２】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項６のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項２３】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項２２記載の方法。
【請求項２４】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項２２記載の方法。
【請求項２５】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項８のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項２６】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項２５記載の方法。
【請求項２７】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項２５記載の方法。
【請求項２８】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項９のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項２９】
トランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームがゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が

一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項３０】
前記トランスクリプトームがヒトのトランスクリプトームである請求項２９記載のオリゴヌクレオチドライブラリー。
【請求項３１】
前記トランスクリプトームがラットのトランスクリプトームである請求項２９記載のオリゴヌクレオチドライブラリー。
【請求項３２】
前記トランスクリプトームがマウスのトランスクリプトームである請求項２９記載のオリゴヌクレオチドライブラリー。
【請求項３３】
組織起源のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが組織起源のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項３４】
病理学的組織起源のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが病理学的組織起源のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項３５】
前記病理学的組織起源ががん組織である請求項３４記載のオリゴヌクレオチドライブラリー。
【請求項３６】
発生段階のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが発生段階のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。
【請求項３７】
病気に罹患している患者のトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームが病気に罹患している患者のゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリー。

【請求項３８】
前記病気ががんである請求項３７記載のオリゴヌクレオチドライブラリー。
【請求項３９】
上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項２９記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。
【請求項４０】
上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項３３記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。
【請求項４１】
上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項３４記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。
【請求項４２】
上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項３６記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。
【請求項４３】
上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が請求項３７記載のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられているＤＮＡマイクロアレイ。
【請求項４４】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項２９のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項４５】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項４４記載の方法。
【請求項４６】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項４４記載の方法。
【請求項４７】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項３３のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項４８】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項４７記載の方法。
【請求項４９】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項４７記載の方法。

【請求項５０】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項３４のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項５１】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項５０記載の方法。
【請求項５２】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項５０記載の方法。
【請求項５３】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項３６のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項５４】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項５３記載の方法。
【請求項５５】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項５３記載の方法。
【請求項５６】
以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法：
請求項３７のオリゴヌクレオチドライブラリー又はそのサブセットによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、一セットのスプライス変異体の総レベルを決定する。
【請求項５７】
前記ハイブリダイゼーションシグナルはヌクレオチドチップから得られる請求項５６記載の方法。
【請求項５８】
前記ハイブリダイゼーションシグナルは電気泳動ゲルから得られる請求項５６記載の方法。
【請求項５９】
トランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーから選択されたオリゴヌクレオチドに基づく二本鎖ＲＮＡ分子であって、トランスクリプトームがゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変

異体をコードし、二本鎖ＲＮＡ分子が３０以下の塩基対を含み、二本鎖ＲＮＡ分子がｍＲＮＡの翻訳と干渉することができる二本鎖ＲＮＡ分子。
【請求項６０】
トランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーから選択されたオリゴヌクレオチドに基づくアンチセンス分子であって、トランスクリプトームがゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードし、アンチセンス分子が３０以下の塩基対を含み、アンチセンス分子がｍＲＮＡの翻訳と干渉することができるアンチセンス分子。
【発明の詳細な説明】
【０００１】
関連出願に関するクロス参照
参照として以下の文献をここに組み入れる：２００１年５月２日に出願されたＵ．Ｓ．シリアルＮｏ．６０／２８７７２４；２０００年７月２８日に出願されたＵ．Ｓ．仮出願シリアルＮｏ．６０／２２１６０７；及び１９９８年８月１３日に出願されたＵ．Ｓ．出願シリアルＮｏ．０９／１３３９８７。この出願は２００１年５月２日に出願されたＵ．Ｓ．シリアルＮｏ．６０／２８７７２４及び２０００年７月２８日に出願されたＵ．Ｓ．仮出願シリアルＮｏ．６０／２２１６０７を優先権主張する。
【０００２】
発明の背景
発明の分野
本発明は生物学的サンプルからのメッセンジャーＲＮＡを検出するのに有用なオリゴヌクレオチドライブラリーを提供する。より特別には、本発明はトランスクリプトームの中に場所を占め、対応するゲノムの中に場所を占める遺伝子又は転写単位から転写されるＲＮＡスプライス変異体を含むＲＮＡ転写物を検出することができるオリゴヌクレオチドライブラリーを提供する。本発明はオリゴヌクレオチドライブラリーから生成されるオリゴヌクレオチドアレイ及び様々なオリゴヌクレオチド検出システム及び発現プロファイリング研究にオリゴヌクレオチドライブラリーを用いる方法をも提供する。
【０００３】
分野の記述
細胞、組織、器官における正常又は異常な生物学的機能及び過程の理解は関心のある生物学的場所において作用する遺伝子、これらの遺伝子から転写されるメッセンジャーＲＮＡ、及び究極的にはＲＮＡ転写物から生産される一セットのタンパク質の知識しだいで決まる。種レベルでは、すべてのＤＮＡ分子のレパートリーがその種のゲノムを作り上げている；これらのＤＮＡ分子の中で、ＲＮＡ分子に転写されるものは遺伝子又は転写単位と称される。対応して、ゲノム中で転写単位から転写されるすべてのメッセンジャーＲＮＡ分子（「ＲＮＡ転写物」又は「転写物」）はトランスクリプトームを作り上げている；そして、メッセンジャーＲＮＡ分子から翻訳されるすべてのタンパク質のレパートリーはプロテオームを作り上げている。
【０００４】
ある種のトランスクリプトームについての研究はその種の対応するゲノム及びプロテオームの研究者の理解に光を投げかける。従って、ＲＮＡ転写物を同定するための及び生物学的サンプル中のそれらの豊富さのレベルを決定するための様々な技術は様々な種のトランスクリプトームを解読する手段であり、それによって対応するゲノム及びプロテオームに関する発見を容易にしてきた。これらの技術はゲルベースの手順（Ｎｏｒｔｈｅｒｎ Ｂｌｏｔ ａｎａｌｙｓｉｓ，Ｄｏｔ Ｂｌｏｔ ａｎａｌｙｓｉｓ，Ｐｒｉｍｅｒ Ｅｘｔｅｎｔｉｏｎ ａｎａｌｙｓｉｓ，Ｓｕｂｓｔｒａｔｅ Ｅｎｒｉｃｈｍｅｎｔ

Ｈｙｂｒｉｄｉｚａｔｉｏｎ Ａｎａｌｙｓｉｓ，Ｄｉｆｆｅｒｅｎｔｉａｌ Ｄｉｓｐｌａｙ ａｎａｌｙｓｉｓ，Ｐｏｌｙｍｅｒａｓｅ Ｃｈａｉｎ Ｒｅａｃｔｉｏｎ（ＰＣＲ）ベースの分析の如き）及びもっとも最近はチップベースの手順（ＤＮＡマイクロアレイ分析の如き）の両方を含む。

【０００５】

多数の種についての全ゲノムの配列決定の完了はこれらのゲノム及びそれらの対応するトランスクリプトーム及びプロテオームの研究におけるよりよい効率、信頼性、及び理解を約束する。ヒトゲノム中の発現される遺伝子又は転写単位の総数は約３００００～４００００であると見積もられている（Ｖｅｎｔｅｒ Ｇ．ｅｔ ａｌ．，Ｓｃｉｅｎｃｅ ２０００，Ｖｏｌ，２９１：１３０４－１３５１；Ｔｈｅ Ｇｅｎｏｍｅ Ｉｎｔｅｒｔｎａｔｉｏｎａｌ Ｓｅｑｕｅｎｃｉｎｇ Ｃｏｎｓｏｒｔｉｕｍ，Ｎａｔｕｒｅ ２０００，Ｖｏｌ．４０９：８６０－９２１）。しかし、ヒトのプロテオームにコードされているタンパク質の数はこの見積もりを有意に超過すると予測されている。なぜなら、多くの場合一つ以上のＲＮＡスプライス変異体が転写単位又は遺伝子から転写されるからである。見積もりは変化する。しかし、研究者の中にはヒトのトランスクリプトームは５０００００までのＲＮＡ転写物を含むかもしれず、ヒトゲノム中の遺伝子又は転写単位の３０％以上がいくつかのＲＮＡスプライス変異体を生成すると信じているものもいる（Ｍｉｒｏｎｏｖ ｅｔ ａｌ．１９９９，Ｇｅｎｏｍｅ Ｒｅｓｅａｒｃｈ ９：１２８８－１２９３）。これらの数は控えめであると他人によって見なされている。選択的スプライシングもラット及びマウスではたとえば同様の頻度で、及びＤｒｏｓｏｐｈｉｌａ ｍｅｌａｎｏｇａｓｔｅｒ 及びＣａｅｎｏｒｈａｂｄｉｔｉｓ ｅｌｅｇａｎｓの如き下等生物でも起こる。選択的スプライシングの特別な事例は選択的ポリアデニル化であること、及び有意な数の遺伝子又は転写単位は選択的ポリアデニル化部位を有していることに注意すべきである。

【０００６】

異なるＲＮＡスプライス変異体から翻訳されるタンパク質はかなり異なる生物学的機能を有しているかもしれない。異なるＲＮＡスプライス変異体は異なる組織で、異なる発生段階で、異なる病気状況で発現されるかもしれない。転写単位からのＲＮＡ転写物及びＲＮＡスプライス変異体の検出、及び（ｉ）転写単位からのスプライス変異体を含むすべてのＲＮＡ転写物の豊富さのレベル及び（ｉｉ）サブセットのスプライス変異体又は一つのスプライス変異体の豊富さのレベルの決定は従ってトランスクリプトーム、ひいては対応するプロテオームの状況を正確に捉えることにおいて望ましい。そしてなお、ＲＮＡスプライス変異体の定性的及び定量的検出は分子生物学の分野においては不適当な挑戦のままである。特に、トランスクリプトーム規模ではそうである。

【０００７】

オリゴヌクレオチドライブラリーはゲルベースのシステム及びチップベースのシステムの両方においてＲＮＡ分子を検出してそれらの豊富さのレベルを測定するために用いられてきた。しかし、これらのライブラリーのオリゴヌクレオチド（オリゴ）は標的ＲＮＡにハイブリダイズするためのオリゴプローブとして用いられるが、スプライス変異体を正確に検出することは通常不可能である。つまり、それらは一つ以上の特定スプライス変異体にハイブリダイズすることができる（従って認識することができる）オリゴを含まず、それ故、選択的にスプライスされたＲＮＡ転写物を効率的に同定及び／又は区別することができず、又は異なるスプライス変異体の豊富さのレベルを決定することができない。

【０００８】

発明の概要

それ故、本発明の目的はゲノム中の転写単位から転写されるＲＮＡ転写物及びＲＮＡスプライス変異体を検出することができ、それによって対応するトランスクリプトームを定性的に及び定量的に特性決定することができるオリゴヌクレオチドライブラリーを提供することである。本発明の他の目的はオリゴヌクレオチドライブラリーから生成されるオリゴヌクレオチドアレイ及びオリゴヌクレオチドライブラリーを用いて関心のあるトランスク

リプトームを研究する方法、及びオリゴヌクレオチドライブラリーを発現プロファイリング研究に用いる方法を提供することである。

【０００９】

本発明によれば、一実施態様において、トランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームがゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができる、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００１０】

本発明の他の実施態様によれば、オリゴヌクレオチドライブラリーはヒトのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。他の実施態様によれば、本発明のオリゴヌクレオチドライブラリーはラットのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。さらに他の実施態様によれば、本発明のオリゴヌクレオチドライブラリーはマウスのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。さらなる実施態様によれば、本発明のオリゴヌクレオチドライブラリーはアラバドプシスのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。なおさらなる実施態様によれば、本発明のオリゴヌクレオチドライブラリーはドロソフィラ・メラノガスターのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。

【００１１】

本発明によれば、他の実施態様において、組織起源のサブトランスクリプトーム（特定の生物学的起源、構造、又は機能特性（たとえば組織特異性、病気特異性、発生特異性など）を担持するトランスクリプトームの一部）の中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが組織起源のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００１２】

本発明によれば、オリゴヌクレオチドライブラリーの組織起源は腎臓、脳、心臓、肺、骨、肝臓又は様々な実施態様における関心のある他の組織であることができる。

【００１３】

本発明によれば、さらに他の実施態様において、病理学的組織起源のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが病理学的組織起源のサブゲノム（特定の生物学的起源、構造、又は機能特性（たとえば組織特異性、病気特異性、発生特異性など）を担持するゲノムの一部）の中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００１４】

本発明によれば、病理学的組織起源は様々な実施態様における結腸がん組織、乳がん組織、肺がん組織、骨がん組織、前立腺がん組織、脳がん組織、又は肝臓がん組織のごときが

ん組織であることができる。本発明によれば、病理学的組織起源はたとえば様々な実施態様における異常心臓組織、異常ニューロン組織、異常肝臓組織、又は異常腎臓組織であることができる。

【００１５】

本発明によれば、さらに他の実施態様において、発生段階のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが発生段階のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００１６】

本発明によれば、発生段階はヒトの神経誘導、マウスの中胚葉の誘導、ヒトの赤血球誘導、ヒト又はマウスの幹細胞発生、又は様々な種における他の特異的発生生物学的段階であることができる。

【００１７】

本発明によれば、さらなる実施態様において、病気に罹患している患者のトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームが病気に罹患している患者のゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットのメッセンジャーＲＮＡ（一つ以上のＲＮＡ転写物）に対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００１８】

本発明によれば、病気は様々な実施態様における結腸がん、乳がん、肺がん、骨がん、前立腺がん、脳がん、又は肝臓がんのごときがんであることができる。本発明によれば、病気はたとえば様々な実施態様におけるアルツハイマー病、パーキンソン病、骨粗しょう症、糖尿病、慢性関節リウマチ、又は関心のある他の病気であることができる。

【００１９】

本発明によれば、さらなる実施態様において、トランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームがゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００２０】

本発明の一実施態様によれば、オリゴヌクレオチドライブラリーはヒトのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。他の実施態様によれば、本発明のオリゴヌクレオチドライブラリーはラットのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。さらに他の実施態様によれば、本発明のオリゴヌクレオチドライブラリーはマウスのトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出することができる。

【００２１】

本発明によれば、他の実施態様において、組織起源のサブトランスクリプトームの中に場

所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが組織起源のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００２２】

本発明によれば、オリゴヌクレオチドライブラリーの組織起源は腎臓、脳、心臓、肺、骨、肝臓又は様々な実施態様における関心のある他の組織であることができる。

【００２３】

本発明によれば、なお他の実施態様において、病理学的組織起源のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが病理学的組織起源のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００２４】

本発明によれば、病理学的組織起源は様々な実施態様における結腸がん組織、乳がん組織、肺がん組織、骨がん組織、前立腺がん組織、脳がん組織、又は肝臓がん組織のごときがん組織であることができる。本発明によれば、病理学的組織起源はたとえば様々な実施態様における異常心臓組織、異常ニューロン組織、異常肝臓組織、又は異常腎臓組織であることができる。

【００２５】

本発明によれば、さらに他の実施態様において、発生段階のサブトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、サブトランスクリプトームが発生段階のサブゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがサブゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、サブゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００２６】

本発明によれば、発生段階はヒトの神経誘導、マウスの中胚葉の誘導、ヒトの赤血球誘導、又は様々な種における他の特異的発生生物学的段階であることができる。

【００２７】

本発明によれば、さらなる実施態様において、病気に罹患している患者のトランスクリプトームの中に場所を占めるメッセンジャーＲＮＡを検出するためのオリゴヌクレオチドライブラリーであって、トランスクリプトームが病気に罹患している患者のゲノムの中に場所を占める極めて多数の転写単位から転写されるメッセンジャーＲＮＡを含み、ライブラリーが複数のオリゴヌクレオチドを含み、複数のオリゴヌクレオチド中の各オリゴヌクレオチドがゲノムの所定の転写単位から転写される一セットの又はサブセットのメッセンジャーＲＮＡに対して選択的にハイブリダイズすることができ、ゲノムの少なくとも一つの転写単位が一以上のメッセンジャーＲＮＡスプライス変異体をコードするオリゴヌクレオチドライブラリーが提供される。

【００２８】
本発明によれば、病気は様々な実施態様における結腸がん、乳がん、肺がん、骨がん、前立腺がん、脳がん、又は肝臓がんのごときがんであることができる。本発明によれば、病気はたとえば様々な実施態様におけるアルツハイマー病、パーキンソン病、骨粗しょう症、糖尿病、慢性関節リウマチ、又は関心のある他の病気であることができる。
【００２９】
本発明のオリゴヌクレオチドライブラリーは様々な実施態様において少なくとも９５，２００，４００，６００，８００、又は１０００のオリゴヌクレオチドを有することができる。本発明のライブラリーにおけるオリゴヌクレオチドは、本発明の様々な実施態様において７５，６５，６０，５５，５０，４５，４０，３５，３０、又は２０塩基の長さ、又はそれらの周りの又はそれらの間のいかなる整数であることができる。しかし、本発明によればより小さいか又はより長いオリゴヌクレオチドも用いられることができる。
【００３０】
本発明によれば、なおさらなる実施態様において、様々な実施態様における本発明の上述のオリゴヌクレオチドライブラリーの各オリゴヌクレオチドは、固体表面への取り付けを可能にする改変を含む。取り付けはたとえば共有結合的、又は静電的であることができる。
【００３１】
本発明によれば、他の実施態様において、上に複数のオリゴヌクレオチド配列をスポッティングされたＤＮＡマイクロアレイであって、前記複数のオリゴヌクレオチド配列が様々な実施態様における本発明の上述のオリゴヌクレオチドライブラリーによって与えられているＤＮＡマイクロアレイが提供される。
【００３２】
本発明によれば、なお他の実施態様において、以下のステップを含む様々な豊富さの二つ以上のＲＮＡを含む細胞又は組織サンプルを発現プロファイリングする方法が提供される：
様々な実施態様における本発明の上述のオリゴヌクレオチドライブラリーによって与えられる複数のオリゴヌクレオチド配列に対して前記サンプルのハイブリダイゼーションシグナルを測定し、それによって前記サンプル中の前記二つ以上のＲＮＡのレベルを決定し、そしてＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合、（ｉ）一セットのスプライス変異体の総レベル，（ｉｉ）そのサブセットの総レベル、又は（ｉｉｉ）その一つのスプライス変異体のレベルを決定する。
【００３３】
本発明によれば、発現プロファイリングする方法は様々な実施態様においてヌクレオチドチップ、膜又はフィルター、そこからインプリントされた電気泳動ゲル及びフィルター又は膜、又は上でハイブリダイゼーションが行われる他の同様のプラットフォームを用いることができる。
【００３４】
本発明なお他の側面によれば、本発明のオリゴヌクレオチド又はオリゴヌクレオチドに基づいたＤＮＡ構築物をもちいて翻訳を阻害又は防止する方法が提供され、そこでは構築物は細胞レベルでアンチセンスＲＮＡを生産することができる。
【００３５】
本発明さらに他の側面によれば、ライブラリーのオリゴヌクレオチドは一本鎖、二本鎖、又は部分的二本鎖であることができる。
【００３６】
本発明さらに他の側面によれば、特定遺伝子の翻訳後サイレンシングを行うことができる二本鎖オリゴヌクレオチドが提供される。一実施態様においては、これらのオリゴヌクレオチドは短い二本鎖ＲＮＡである。
【００３７】
図面の簡単な説明

図１．これは３つのエクソン及びそこから転写される２つの選択的にスプライスされた転写物を有する転写単位を示す図である。

【００３８】

図２．これは３つのエクソン及びそこから転写される２つの異なるスプライス変異体を有する転写単位を示す図である。

【００３９】

図３．これは５つのエクソン及びそこから転写される４つの選択的にスプライスされた転写物を有する転写単位を示す図である。

【００４０】

好ましい実施態様の詳細な説明

ライブラリー、オリゴヌクレオチド及びその改変

トランスクリプトーム又はその一部分を特性決定するためには、メッセンジャーＲＮＡサンプルへのハイブリダイゼーションをテストするためのプローブとして用いられるオリゴヌクレオチドは、ＲＮＡ転写物の総集団又は関心のあるその一部分を代表しなければならない。この種のオリゴヌクレオチド情報を生成するためには、Ｕ．Ｓ．特許出願シリアルＮｏ．０９／１３３９８７（‘９８７出願）で概説されているアプローチを用いることができる。

【００４１】

公知のヌクレオチド配列、配列断片又は発現された配列タグ（“ＥＳＴ”）の大きなプール、例えばＧｅｎＢａｎｋ，ＥＭＢＬ，ＤＢｅｓｔ配列コレクション、は配列クラスター及び長いコンティグ（連続配列、即ち短いＥＳＴから集成された長い配列）を得るために厳格なクラスタリング及び集成手順に供される。これらの手順はＥＳＴ及び他の配列断片の大きなプールと比較して遺伝子又は転写単位のより良い代表である配列を生み出す。

【００４２】

所定の遺伝子又は転写単位についての代表的配列が選択されて、この配列から一以上のオリゴヌクレオチド（又はオリゴ、オリゴプローブ）が誘導される。代表的配列は標的遺伝子又は転写単位に対して高い特異性を有する（即ち、それは他の遺伝子又は転写単位からの配列に対して少ない相同性を共有し、それ故、非特異的結合の頻度が最少化される）；配列の質は良好である（即ち、遺伝子配列情報は正確である）。オリゴヌクレオチドは高い感度を有する。即ち、標的遺伝子又は転写単位に対するプローブのハイブリダイゼーション親和性は高い。プローブは標的遺伝子又は転写単位の３′末端に近い配列を認識することが好ましく、かくして標的遺伝子サンプルが調製される特定の手順においてはその収率は最大化される。オリゴヌクレオチドは二次構造（例、ヘアピン）を欠くか又は二次構造を形成する傾向があるので、ハイブリダイゼーションは損なわれない。

【００４３】

より重要なことには、標的遺伝子及び転写単位の選択的スプライシングを考慮に入れて代表的配列は選択され、一以上のオリゴヌクレオチドは設計される。このように、オリゴヌクレオチドは（ｉ）特定遺伝子又は転写単位のすべてのスプライス変異体、（ｉｉ）すべてのスプライス変異体のサブセット、又は（ｉｉｉ）遺伝子又は転写単位のスプライス変異体の一つのいずれかにハイブリダイズして検出するために用いられることができる。

【００４４】

図１を参照すると、転写単位又は遺伝子は例えば三つのエクソン：Ａ，Ｂ及びＣを有し、転写において選択的にスプライスされて二つの変異体：転写物１（ＡＣ）及び転写物２（ＢＣ）を生ずる。配列Ａに対して相補的なオリゴ又はその断片はそれ故、転写物１のみを検出し、転写物２は検出しないであろう。配列Ｂに対して相補的なオリゴ又はその断片は対照的に転写物２のみを検出し、転写物１は検出しないであろう。

【００４５】

図２に示されているシナリオは若干異なる。転写単位は三つのエクソン：Ａ，Ｂ及びＣを有し、これらは２つのスプライス変異体：転写物１（ＡＣ）及び転写物２（ＡＢＣ）へと転写される。配列Ａに対して相補的なオリゴ又はその断片は転写物１及び転写物２の両者

を検出することができ、両者の豊富さのレベルを集合的に測定することができるであろう。対照的に、配列Ｂに対して相補的なオリゴ又はその断片は転写物２を検出するが転写物１を検出することができず、それによって転写物２の豊富さについての直接的な情報を与えるであろう。

【００４６】

図３は５つのエクソン：Ａ，Ｂ，Ｃ，Ｄ及びＥを有する転写単位又は遺伝子を示す。４つの選択的にスプライスされた転写物がある：転写物１（ＡＣＤ）、２（ＡＣＥ）、３（ＢＣＤ）及び４（ＢＣＥ）である。これは一層複雑なスキームであるが、原理は同じままである：配列Ａに対して相補的なオリゴ又はその断片は転写物１及び２のみを検出することができ（従って、“１＋２＝Ａ”）；配列Ｂに対して相補的なオリゴは転写物３及び４のみを検出することができ（従って“３＋４＝Ｂ”）；配列Ｄに対して相補的なオリゴは転写物１及び３のみを検出することができ（従って“１＋３＝Ｄ”）；そして配列Ｅに対して相補的なオリゴは転写物２及び４のみを検出することができるであろう（従って“２＋４＝Ｅ”）。それ故、我々はサブセットのスプライス変異体を特異的に検出することができる多数のオリゴを有する。加えて、上述の４つの多変数方程式を解けば各スプライス変異体（Ａ，Ｂ，Ｄ及びＥ）の豊富さのレベルが明らかとなるであろう。更に、配列Ｃに対して相補的なオリゴ又はその断片は４つの転写物すべてを検出することができ、かくしてこれらのスプライス変異体の豊富さの総レベルを測定することができるであろう。

【００４７】

同様に、本発明によれば、代表的配列及び一以上のオリゴヌクレオチドプローブが選択的ポリアデニル化部位を有する特定遺伝子又は転写単位について生成されるように、選択的ポリアデニル化が考慮に入れられる。これらのプローブは（ｉ）すべての選択的ポリアデニル化転写物、（ｉｉ）これらの選択的ポリアデニル化転写物のサブセット、又は（ｉｉｉ）標的遺伝子又は転写物のポリアデニル化部位の一つのいずれかにハイブリダイズして検出するために用いられることができる。

【００４８】

本発明のオリゴヌクレオチドはゲルベースのシステムにおけるそれらの適用又はＲＮＡ検出のためのチップ又はアレイベースのシステム上への取り付けを容易にするために末端で改変されることができる。例えば一つの実施態様においては、オリゴヌクレオチドは５´末端で改変されている：Ａ５´Ｃ６－アミノ変性はガラスアレイ表面上へのオリゴヌクレオチドの共有結合取り付けを可能にするために行われる。炭素はスペーサーとして機能し、反応性アミン基はガラス表面に共有結合的に取り付けられているアルデヒドと相互作用する。予め処理されたガラスは多数の会社、例えばＡｒｒａｙＩｔ．ｃｏｍ（ｈｔｔｐ：／／ａｒｒａｙｉｔ．ｃｏｍ），Ｃｏｒｎｉｎｇ（ｈｔｔｐ：／／ｗｗｗ．ｃｏｒｎｉｎｇ．ｃｏｍ），Ｃｌｏｎｔｅｃｈ（ｈｔｔｐ：／／ｗｗｗ．ｃｌｏｎｔｅｃｈ．ｃｏｍ）、から商業的に入手することができる。これらのガラス及び他の類似物はオリゴヌクレオチドの結合のために好適であり、それにより本発明によるオリゴヌクレオチドライブラリーを用いたオリゴアレイを製造するために好適である。取り付けは例えば共有結合的又は静電的であることができる。一例はポリ－Ｌ－リシンガラス上へのオリゴの結合であり、これは静電的である：ポリ－Ｌ－リシンは正に帯電しており、一方、ＤＮＡのリン酸骨格は負に帯電している。他の例はアルデヒドガラス上へのオリゴの結合であり、これは共有結合的である。

【００４９】

ゲルベースの及びアレイベースのプラットフォームと共に用いられるオリゴヌクレオチド上述の通り、本発明によればトランスクリプトーム中のＲＮＡ種を測定するために様々な検出プラットフォームが用いられることができる。一実施態様においては、ゲルベースのプラットフォームが用いられる；他の実施態様においては、チップ又はアレイベースのプラットフォームが用いられる。ゲルベースのプラットフォームが用いられる場合、Ｎｏｒｔｈｅｒｎ　Ｂｌｏｔ　分析、Ｄｏｔ　Ｂｌｏｔ分析、Ｐｒｉｍｅｒ　Ｅｘｔｅｎｓｉｏｎ　分析、Ｓｕｂｓｔｒａｔｅ　Ｅｎｒｉｃｈｉｍｅｎｔ　Ｈｙｂｒｉｄｉｚａｔｉｏｎ

分析、Ｄｉｆｆｅｒｅｎｔｉａｌ Ｄｉｓｐｌａｙ 分析、Ｐｏｌｙｍｅｒａｓｅ Ｃｈａｉｎ Ｒｅａｃｔｉｏｎ（ＰＣＲ）ベースの分析又は他の同様の手順の如き技術が用いられることができる。アレイはニトロセルロースの如き基質上に直接スポッテイングすることによって作られることもできる。本質的に、本発明のオリゴヌクレオチドはラベリングされ（放射能活性ラベリング、蛍光ラベリングを介して、又は他の好適なラベリング方法及び技術分野において知られている捕獲成分を用いて）、そしてＲＮＡ又はｃＤＮＡサンプルが泳動されたゲルからインプリントされたフィルター又は膜にハイブリダイズされる。ハイブリダイゼーションシグナルはオリゴヌクレオチドプローブが表すＲＮＡ転写物（及びそのスプライス変異体）の同一性及び豊富さを示す。これらのゲルベースのヌクレオチドハイブリダイゼーション及び検出手順は熟練した分子生物学者には周知である。一般的に、Ｓａｍｂｒｏｏｋ Ｊ．ａｎｄ Ｒｕｓｓｅｌｌ，２００１，Ｍｏｌｅｃｕｌａｒ Ｃｌｏｎｉｎｇ，ａ Ｌａｂｏｒａｔｏｒｙ Ｍａｎｕａｌ，３ｒｄ Ｅｄ．を参照されたい。ヌクレオチドの分離が含まれないいくつかの場合においては、ヌクレオチドプローブ及びサンプルは電気泳動ゲルを使用しないでハイブリダイゼーションのために膜又はフィルターに直接適用されることができる。

【００５０】

特定のチップ又はアレイベースのプラットフォームが用いられる場合、本発明のオリゴヌクレオチドは固体表面、例えばガラススライド上に、上述のような様々な取り付け化学を介して又は他の好適な取り付け手順を介してスポッティング又はプリントされる。スライドは次に典型的なゲルベースの分離及びハイブリダイゼーション実験におけるフィルター又は膜と類似の様式で処理され、続いて、関心のある組織又は細胞系の如き生物学的サンプルから由来するラベリングされた（放射能活性ラベリング、蛍光ラベリングを介して、又は他の好適なラベリング方法を用いて）ｃＤＮＡ又はＲＮＡ分子を用いてプロービングされる。いくつかの場合には、ラベリングされたクローン又は総ゲノムＤＮＡが代わりに用いられることができる。ハイブリダイゼーションに続いて、スライドは洗浄され、スキャニングされてハイブリダイゼーションシグナルが読み取られて記録される。これらのシグナルはオリゴヌクレオチドプローブが対応しておりかつ検出されるＲＮＡ転写物（及びそのスプライス変異体）の同一性及び豊富さを表す。特定の生物学的サンプル中のＲＮＡ転写物及びそれらのスプライス変異体のかかる測定はそれ故、特定の生物学的状況 － 正常、病気、又は治療された － におけるトランスクリプトームのスナップ写真を与え、従って様々な生物学的状況における特性及び相互作用を解明するのを助ける。

【００５１】

本発明によれば、ライブラリーの全てのオリゴヌクレオチド又はその一部は単一のマイクロアレイ上にスポッティングされることができ、それ故、単一のハイブリダイゼーション手順により多数の転写物又はそれらの対応する多数の遺伝子をテストすることができる。変形として、他の実施態様においてはオリゴヌクレオチドはアレイの表面上で直接合成されることができる。固体表面上でのオリゴの合成方法は技術分野では知られている。例えばＵ．Ｓ．特許第５５９３８３９号を参照されたい。この実施態様においては、ライブラリー中のオリゴヌクレオチドは約２０～２５塩基の長さであることが好ましい。何故なら、長い配列のｉｎ ｓｉｔｕ合成はしばしばエラーを生ずる傾向があるからである。アレイ上に直接合成される本発明のライブラリーはかくして特定の生物学的状況下でサブトランスクリプトームの転写物を検出するために設計された特異的ライブラリー又はミニライブラリーとして有用であるかもしれない。

【００５２】

オリゴヌクレオチドライブラリーのための使用

ゲルベースの、アレイベースの又は他の好適な検出システムが本発明により用いられている場合、オリゴヌクレオチドライブラリーはヒトのトランスクリプトーム、マウスのトランスクリプトーム又はラットのトランスクリプトームの如き所定の種のトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出するために構築されることができる。ライブラリーは特定の生物学的又は病理学的起源の、又は特定の生物学的又は病理学的状況

におけるサブトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出するために構築されることもできる。

【００５３】

例えば、本発明の一実施態様においては、サブトランスクリプトームは特定の組織起源のものであることができる。つまり、それは腎臓のサブトランスクリプトーム、脳組織のサブトランスクリプトーム、心臓組織のサブトランスクリプトーム、肺組織のサブトランスクリプトーム、骨組織のサブトランスクリプトーム、肝臓組織のサブトランスクリプトームなどであることができる。従って、本発明のライブラリーは関心のある組織のサブトランスクリプトームの中に場所を占めるＲＮＡ転写物及びスプライス変異体を特異的に検出するために用いられることができる。特定遺伝子（即ち、あるサブトランスクリプトーム中にのみ存在し、他のサブトランスクリプトーム中には存在しない特定のＲＮＡ転写物又はスプライス変異体）の組織特異的発現又は組織特異的様式（即ち、特定のＲＮＡ転写物又はスプライス変異体があるサブトランスクリプトーム中では他のサブトランスクリプトーム中よりも多かれ少なかれ豊富である）での特定遺伝子のディファレンシャル発現はそれ故、本発明のオリゴヌクレオチドライブラリーを用いて検出され監視されることができる。転写物又は特定スプライス変異体の存在又は不在、及びそれらの豊富さの相違は統計学的有意さ水準下で評価されることができる。好適な統計学的評価は技術分野では知られている。

【００５４】

他の実施態様においては、サブトランスクリプトームは特定の病理学的組織起源のものである。つまり、それはがん組織のサブトランスクリプトーム、例えば結腸がん組織のサブトランスクリプトーム、乳がん組織のサブトランスクリプトーム、肺がん組織のサブトランスクリプトーム、骨がん組織のサブトランスクリプトーム、前立腺がん組織のサブトランスクリプトーム、脳がん組織のサブトランスクリプトーム、又は肝臓がん組織のサブトランスクリプトーム、であることができる。それは異常心臓組織のサブトランスクリプトーム、異常ニューロン組織のサブトランスクリプトーム、異常肝臓組織のサブトランスクリプトーム、又は異常腎臓組織のサブトランスクリプトームなどであることもできる。

【００５５】

それ故、本発明のライブラリーは関心のある病理学的組織起源のサブトランスクリプトームの中に場所を占めるＲＮＡ転写物及びスプライス変異体を特異的に検出するために用いることができる。従って、本願は組織及び病気特異的遺伝子、即ち特定組織においてのみ及び特定の病理学的状況下でのみ発現される遺伝子の検出を可能にする。何故なら、対応する転写物及びスプライス変異体は特定のサブトランスクリプトームにのみ見出されることができるからである。加えて、本願は組織及び病気特異的様式でディファレンシャルに発現される遺伝子、即ち、他の組織よりも又は正常又は他の状況下よりも特定組織及び特定の病理学的状況下で異なるレベルで発現される遺伝子、の検出を更に可能にする。何故なら、対応する転写物及びスプライス変異体は特定のサブトランスクリプトームにおいて多かれ少なかれ豊富であるからである。転写物又はスプライス変異体の存在又は不在、及びそれらの豊富さの相違は統計学的有意さ水準下で評価されることができる。

【００５６】

さらに他の実施態様においては、サブトランスクリプトームは特定の発生段階のものである。つまり、それはヒトの神経誘導、マウスの中胚葉の誘導、ヒトの赤血球誘導、又は様々な種における他の特異的発生生物学的段階からのサブトランスクリプトームであることができる。それ故、本発明のライブラリーは関心のある発生的生物学的段階のサブトランスクリプトームの中に場所を占めるＲＮＡ転写物及びスプライス変異体を特異的に検出するために用いることができる。従って、本願は発生段階特異的遺伝子、即ち特定の発生段階においてのみ発現される遺伝子の検出を可能にする。何故なら、対応する転写物及びスプライス変異体は特定のサブトランスクリプトームにのみ見出されることができるからである。加えて、本願は発生段階に依存する様式でディファレンシャルに発現される遺伝子、即ち、他の状況よりも特定発生段階で異なるレベルで発現される遺伝子、の検出を更に

可能にする。何故なら、対応する転写物及びスプライス変異体は特定のサブトランスクリプトームにおいて多かれ少なかれ豊富であるからである。転写物又はスプライス変異体の存在又は不在、及びそれらの豊富さの相違は統計学的有意さ水準下で評価されることができる。

【００５７】

更なる実施態様においては、オリゴヌクレオチドライブラリーは特定の病気に罹患している患者のトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出するために構築される。つまり、それは結腸がん患者、乳がん患者、肺がん患者、骨がん患者、前立腺がん患者、脳がん患者、又は肝臓がん患者のごときがん患者のトランスクリプトームであることができる。それはたとえばアルツハイマー病患者、パーキンソン病患者、骨粗しょう症患者、糖尿病患者、慢性関節リウマチ患者、又は関心のある他の病気の患者のトランスクリプトームであることもできる。

【００５８】

本発明は病気特異的遺伝子、即ち特定の病気に罹患している患者においてのみ発現される遺伝子の検出を可能にする。何故なら、対応する転写物及びスプライス変異体はかかる患者のトランスクリプトームにのみ見出されることができるからである。加えて、本発明は特定の病気の患者でディファレンシャルに発現される遺伝子、即ち、正常又は健康な個体におけるよりも特定の病気の患者において異なるレベルで発現される遺伝子、の検出を更に可能にする。何故なら、対応する転写物及びスプライス変異体はかかる患者のトランスクリプトームにおいて多かれ少なかれ豊富であるからである。転写物又はスプライス変異体の存在又は不在、及びそれらの豊富さの相違は例えば特定のｐスコアを要求する統計学的有意さ水準下で評価されることができる。更に、本願は特定の病気に罹患している患者の発現パターン又はプロファイルの個人化された特性決定（所定の患者の転写物及びスプライス変異体の同一性及び豊富さのコンピレーション）を可能にし、従って個人化された治療方法の開発の基礎を提供する。

【００５９】

まとめると、本発明のオリゴヌクレオチドライブラリーはＲＮＡ転写物及びスプライス変異体を検出してそれにより関心のあるトランスクリプトーム又はサブトランスクリプトームを特定決定するための様々な機能的関連において用いられることができる。この関連において、サブトランスクリプトームの範囲はタンパク質の機能によって制限されることもできる。例えば、オリゴヌクレオチドライブラリーは細胞表面抗原に対応するサブトランスクリプトームのＲＮＡ転写物及びスプライス変異体を検出するために用いられることができる。以下に規定するような４０以上のタンパク質機能的グループのリスト（Ｕ．Ｓ．仮出願シリアルＮｏ．６０／２２１６０７も参照されたい）は、例えばサブプロテオーム又は特定の機能のその一つのタンパク質及びその異常型を生ずるサブトランスクリプトームを特定決定することができるオリゴヌクレオチドライブラリーを構築するために本発明において有用である。

【００６０】

グループ１。アダプター結合タンパク質。これらは他の細胞成分と関連し　－　それらと結合又は相互作用し　－　それらの構造的統合性を維持し、その機能的活性を発揮させる。

【００６１】

グループ２。接着分子。これらは隣接細胞間の接着の調節に関与している。

【００６２】

グループ３。アポリポタンパク質。これらはリポタンパク質粒子の一部であるタンパク質であり、これらの粒子の結合及び内面化における細胞シグナリング作用を有する。アポリポタンパク質欠失は異常に高レベル又は低レベルのリポタンパク質及びコレステロールに関与する病気、並びに形成又は動脈硬化に関与する状況において見出される。

【００６３】

グループ４。アポトーシスに関連するタンパク質。これらは阻害又は刺激様式でアポトー

シス経路に関与するタンパク質及び酵素である。これらのタンパク質の異常は退行病（例えば神経退行病）の如き細胞の未成熟死に関与する病気、又は老化に関連する状況、又は代わりに要求されるアポトーシスが起こらない病気を引き起こす。かかる病気の例はがん及び心筋梗塞後の心臓機能の欠如である。

【００６４】

グループ５。がんに関連するタンパク質。これらはＤＮＡ修復タンパク質、腫瘍マーカー及び抗原、腫瘍サプレッサー、及び腫瘍形成に参加するメッセンジャー分子などを含み、様々な種類のがん及び対応する転移状況の治療及び検出に関与する。

【００６５】

グループ６。カルボキシラーゼ。これらのタンパク質の異常は他の成分からCO_2基を除去する酵素反応の異常制御を引き起こす。

【００６６】

グループ７。細胞表面抗原。これらは細胞の表面で発現されるタンパク質である。これらのタンパク質の異常は例えばＡＩＤＳ及び細胞表面抗原に関与するがんの如き自己免疫疾患において見られることができる。

【００６７】

グループ８。細胞増殖を制御するタンパク質。これらのタンパク質は退行病（低い増殖）又はがん（未制御の増殖）における欠陥構造又は機能を有する。

【００６８】

グループ９。凝集に関連するタンパク質。これらのタンパク質の異常は例えば血友病、又は発作及び血管のブロックを引き起こす。

【００６９】

グループ１０。変換酵素。これらの酵素はプレカーサータンパク質の特異的開裂によって一つのタンパク質を他のタンパク質へと変換する。

【００７０】

グループ１１。サイクラーゼ酵素。これらの酵素はトリホスフェートをサイクリックモノホスフェートへと変換する。これらの酵素の異常はトリホスフェートのサイクリックモノホスフェートへの不十分な又は過剰の変換を引き起こし、それによって細胞シグナリングに影響を与える。

【００７１】

グループ１２。タンパク質の分解に関与するタンパク質。これらのタンパク質の異常は他のタンパク質の異常な分解を引き起こし、それは様々なタンパク質性生産物の細胞中への異常な蓄積をもたらすことがある。

【００７２】

グループ１３。発生に関与するタンパク質。これらのタンパク質の異常は例えば胎児の異常発生に関与する遺伝病において表れる。

【００７３】

グループ１４。ドメインタンパク質。これらのタンパク質はタンパク質とタンパク質の相互作用に関与している。

【００７４】

グループ１５。エステラーゼ。これらのタンパク質はエステル結合を開裂する。

【００７５】

グループ１６。成長因子。これらのタンパク質の例はサイトカイン、インターロイキン、インターフェロン、及びリンポカインなどを含む。これらのタンパク質は少し例を挙げれば自己免疫疾患、炎症関連の病気、移植片対ホスト病、感染性病原体により生ずる病気、及びがんに関与している。

【００７６】

グループ１７。ホルモン及びポイエチンタンパク質。これらのタンパク質の異常は様々な内分泌病において見られる。

【００７７】

グループ１８。ハウスキーピングタンパク質。これらの例はホメオボックスタンパク質、ヒートショックタンパク質、及びシャペロニンなどである。
【００７８】
グループ１９。ヒドロラーゼ。これらの酵素はヒドロキシル基を変性する；例としてヒドロゲナーゼ、デヒドロゲナーゼ、ヒドロラーゼ、及びヒドロキシラーゼがある。
【００７９】
グループ２０。免疫関連のタンパク質。これらは免疫系に関与するタンパク質であり、抗原、抗体及びそれらの関連タンパク質を含む。これらのタンパク質は炎病、自己免疫病、感染性病及びがんのプロセスの如き病理学的状況において重要である。
【００８０】
グループ２１。阻害剤。これらのタンパク質は細胞過程において他のタンパク質の機能を阻害する。
【００８１】
グループ２２。キナーゼ。これらのタンパク質の異常は不完全な細胞シグナリングに導く。
【００８２】
グループ２３。リパーゼ、ホスホリパーゼ、及びライソホスホリパーゼ。これらのタンパク質は異常脂質代謝に関与している。
【００８３】
グループ２４。細胞マトリックス及び細胞骨格タンパク質。
【００８４】
グループ２５。変性酵素。これらのタンパク質はパラオキソナーゼ、ＧＴＰａｓｅ、ＡＴＰａｓｅ、及びアンヒドラーゼの如き多方面にわたる多数の酵素を含む。これらの酵素の機能不完は様々な細胞過程に関与している。
【００８５】
グループ２６。ムターゼ及びスーパーオキシドジスムターゼ。これらのタンパク質はがん及び老化に関する様々な他の病理学的過程に関与している。
【００８６】
グループ２７。神経に関連するタンパク質。これらのタンパク質は様々な種類の痴呆、神経退行病、てんかん、精神障害などを含む中央神経系の病気に関与している。
【００８７】
グループ２８。オキシダーゼ及びペルオキシダーゼ。これらの酵素の異常は例えば過酸化物を含む代謝問題を引き起こすかもしれない。
【００８８】
グループ２９。オキシゲナーゼ（モノ及びジオキシゲナーゼ）。
【００８９】
グループ３０。ホスファターゼ及びホスホリラーゼ。
【００９０】
グループ３１。リンタンパク質、及びリン脂質。
【００９１】
グループ３２。プロテアーゼ、ペプチダーゼ、及びプロテイナーゼ。
【００９２】
グループ３３。レセプター。
【００９３】
グループ３４。レダクターゼ。
【００９４】
グループ３５。分泌タンパク質。これらのタンパク質はホルモン、神経伝達物質、及び細胞によって細胞外環境へと分泌される様々な他のタンパク質を含む。
【００９５】
グループ３６。シグナル伝達タンパク質。Ａ　Ｇタンパク質はこのグループのタンパク質

の一例である。
【００９６】
グループ３７。サブ細胞タンパク質。これらのタンパク質の例はリボゾームタンパク質を含む。
【００９７】
グループ３８。シンターゼ及びシンテターゼ。
【００９８】
グループ３９。ヌクレオチドの相互作用に関与するタンパク質。これらは転写因子、ＲＮＡ及びＤＮＡ結合タンパク質、ジンクフィンガー、ヘリカーゼ、イソメラーゼ、ヒストン、ヌクレアーゼを含む。
【００９９】
グループ４０。トランスフェラーゼ。これらはタンパク質の官能基の転移に関与する。
【０１００】
グループ４１。翻訳因子。これらは伸長及び開始因子の如き翻訳過程に関与するタンパク質及び酵素である。これらのタンパク質の異常は細胞のタンパク質生産を損なうかもしれない。
【０１０１】
グループ４２。トランスポーター。これらのタンパク質はチャネル、エクスチェンジャー、及びポンプを含む分子及びマクロ分子の輸送を媒介する。
【０１０２】
オリゴヌクレオチドライブラリーを用いた発現プロファイリング
本発明によるオリゴヌクレオチドライブラリーの一つの適用は発現プロファイリングである。これはこれまで様々な場所で言及されているが、いくつかの追加の議論を必要とする。発現プロファイリングは本発明で用いられるように遺伝子発現プロファイリングとも称され、生物学的サンプル中のＲＮＡ転写物及びスプライス変異体の定性的及び定量的決定を意味する。生物学的サンプルは組織又は細胞系サンプル、正常な又は病気の組織サンプル、病気の組織サンプル又は治療のいくらかの期間の後の組織サンプル、異なる発生の生物学的段階から採取されたサンプル、医薬又は治療方法の適用の過程の様々な時点での患者から採取された生物学的サンプル又は液体サンプルであることができる。発現プロファイリングは、特定の生物学的源又は特定の生理学的又は病理学的状況についてのＲＮＡ転写物及びスプライス変異体の同一性及び豊富さのコンピレーションである発現プロファイル又は発現パターンを確立する。
【０１０３】
発現プロファイリングは上述の通り、本発明によりゲル（及び／又は膜）ベースの又はアレイベースの系を介して行われることができる。本質的に、様々な豊富さの二以上のＲＮＡを含む生物学的サンプルの発現プロファイリングは複数のオリゴヌクレオチド配列に対するサンプルのハイブリダイゼーションシグナルを測定してそれによってサンプル中のＲＮＡ転写物のレベルを決定することによって、及びＲＮＡが一セットのスプライス変異体を有する転写単位から転写される場合には、（ｉ）一セットのスプライス変異体の総レベル；（ｉｉ）そのサブセットの総レベル、又は（ｉｉｉ）その一つのスプライス変異体のレベルを決定することによって行われる。複数のオリゴヌクレオチドが上述の様々な実施態様において本発明のオリゴヌクレオチドライブラリーによって与えられる。
【０１０４】
従って、本発明はオリゴヌクレオチドライブラリー及びそのサブライブラリー、カスタマイズされた又は改変されたオリゴヌクレオチド、ライブラリーからのオリゴヌクレオチドを上にスポッティングされたオリゴヌクレオチドアレイ、様々なヌクレオチド検出システムでオリゴヌクレオチドを用いる方法、及びオリゴヌクレオチドライブラリーを用いた発現プロファイリング方法を包含する。
【０１０５】
アンチセンスの用途及びＲＮＡの干渉

本発明のオリゴヌクレオチドはトランスクリプトームの一以上のスプライス変異体に結合することができるので、これらのオリゴヌクレオチドはｉｎ ｓｉｔｕ及びｉｎ ｖｉｖｏアンチセンスの関連で用途を有する。例えば、一本鎖アンチセンスＤＮＡオリゴヌクレオチド自体、又は変性された二本鎖ＤＮＡオリゴヌクレオチドは哺乳類細胞（ヒト細胞を含む）の如き細胞に、及び哺乳類の体（ヒトを含む）の如き体に注入されることができ、それらはそこでｍＲＮＡ転写物に結合すると予測され、それはｍＲＮＡのタンパク質への翻訳を阻害又は防止するであろう。更に、二本鎖オリゴヌクレオチドのセンス鎖又は一本鎖センスオリゴヌクレオチドはアンチセンスＲＮＡの生成のための最初の設計鋳型として用いられることができる。アンチセンスＲＮＡ分子のｉｎ ｖｉｖｏ発現についての当該技術分野では知られているアプローチを用いると、本発明のオリゴヌクレオチドに基づくアンチセンスＲＮＡはｍＲＮＡの翻訳を細胞レベルで阻害又は防止するために用いられることができる。

【０１０６】

ライブラリーのオリゴヌクレオチドはＲＮＡの干渉において用いられることもできる。ＲＮＡの干渉現象は Ｂａｓｓ，Ｎａｔｕｒｅ ４１１：４２８－２９（２００１）；Ｅｌｂａｈｉｒ ｅｔ ａｌ．，Ｎａｔｕｒｅ ４１１：４９４－９８（２００１）；及び Ｆｉｒｅ ｅｔ ａｌ．，Ｎａｔｕｒｅ ３９１：８０６－１１（１９９８）において議論されており、そこでは干渉するＲＮＡの作製方法も議論されている。ここで開示される配列に基づく二本鎖ＲＮＡは長さが１００塩基対（“ｂｐｓ”）未満であり、連続性は約３０ｂｐｓ以下であることが好ましく、相補的ＤＮＡ鎖又は合成アプローチの使用を含む当該技術分野では知られているアプローチを用いて作製されることができる。干渉を引き起こすことができるＲＮＡは小さな干渉ＲＮＡ（“ｓｉＲＮＡ”）と称されることができ、哺乳類細胞（ヒト細胞を含む）の如き細胞中の、及び哺乳類の体（ヒトを含む）の如き体中の特定遺伝子の転写後サイレンシングを引き起こすことができる。本発明による例示的ｓｉＲＮＡは２９ｂｐｓ，２５ｂｐｓ，２２ｂｐｓ，２１ｂｐｓ，２０ｂｐｓ，１５ｂｐｓ，１０ｂｐｓまで、又はそれらの周り又はそれらの間のいかなる数も有することができる。

【０１０７】

実施例１．選択的にスプライスされた転写物すべてを検出するためのオリゴヌクレオチドライブラリー：ヒト、ラット及びマウスのライブラリー。

６０塩基または６５塩基の長さのオリゴヌクレオチドが合成された。それらの各々はアルデヒドガラス表面への続く共有結合的取り付けを可能にするために５′末端でＣ６アミノ付加により変性された。各オリゴヌクレオチドは特定の転写単位又は遺伝子から転写された存在するＲＮＡスプライス変異体の間で共通するＲＮＡ転写物の一部分に対して相補的である。各オリゴヌクレオチドはＲＮＡ転写物及び／又はＲＮＡスプライス変異体に対して特異的又はユニークでもあり、それらのＲＮＡ転写物及び／又はＲＮＡスプライス変異体に対してそれは相補的である。即ち、他の転写単位又は遺伝子からのＲＮＡ転写物又はスプライス変異体に対する各オリゴヌクレオチドの結合は無視しうる。換言すれば、各遺伝子又は転写単位に対して一つのオリゴが選択される。各オリゴは各遺伝子又は転写単位について知られている又は予測されている事実上最大数のスプライス変異体に対して共通である配列セグメントから由来される。加えて、オリゴは高精度の配列の質が配列決定エラーを回避するために維持されるような、二次構造が効果的なハイブリダイゼーションを促進するために回避されるような、及び融解温度がライブラリー中のオリゴコレクション全体にわたって一致するために標準化されるような態様で設計されて合成される。

【０１０８】

いくつかの代表的実施態様においては、ヒトのオリゴライブラリーは１８８６１個のヒトオリゴヌクレオチドを含み、それぞれは６０塩基の長さである；ラットのオリゴライブラリーは４８５４個のラットオリゴヌクレオチドを含み、それぞれは６５塩基の長さである；マウスのオリゴライブラリーは７５２４個のオリゴヌクレオチドを含み、それぞれは６５塩基の長さである。１８８６１個のヒトオリゴ、４８５４個のラットオリゴ、及び７５

２４個のマウスオリゴの配列はＳｅｃ．８０１（ａ）下で提出された配列表中に含まれている（出願人名はＣｏｍｐｕｇｅｎ　Ｌｔｄ．（ＣＲＦ及びコピー）であり、ＣＤ－Ｒメディアで本願に沿って提出されたＰａｔｅｎｔｉｎ　３．０中の３２３３７個の配列を含む。ファイルサイズ５．７８ＭＢ、ファイル名Ｓｈｏｓｈａｎ、代理人整理番号３６６８８－０００６。この全体は参照文献としてここに組み入れられる。これはＣＤ－Ｒメディアで本願に沿って提出されたＰａｔｅｎｔｉｎ３．０中の３２３３７個の配列を含む配列表（ＣＲＦ及びコピー）中に規定される配列を含む。ファイルサイズ５．７８ＭＢ、ファイル名Ｓｈｏｓｈａｎ、代理人整理番号３６６８８－０００４。この全体は参照文献としてここに組み入れられる。これは２００１年５月２日に出願されたＵ．Ｓ．シリアルＮｏ．６０／２８７７２４において提出されている。）。配列表中で用いられている３２３３７個の配列もここで与えられる。これらはＲａｗ　Ｓｅｑｕｅｎｃｅｓというラベルを貼られたＣＤ－Ｒディスク上のＭＳ　Ｗｏｒｄ　フォーマット中の、Ｓｅｃ．８０１（ａ）下で提出された配列表であり、代理人整理番号３６６８８－０００６を有する５．７８ＭＢの配列表についてのＰＣ／ＭＳ－ＤＯＳ　ＰＡＴＥＮＴＩＮ３．０であり、その全体は参照文献としてここに組み入れられる。これらのライブラリー中の各オリゴは特定の機能的タンパク質をコードする転写単位又は遺伝子から転写されるＲＮＡ転写物及び／又は全ての存在する又は予測されるＲＮＡスプライス変異体を表す。多くの場合、オリゴ配列についての対応する遺伝子又は転写単位の名称、ＧｅｎＢａｎｋアセッション番号、及び他の注釈情報は特定されている。上記配列の数は絶対的ではなく、それ故、異なる数の配列を含む他のライブラリーも出願人の教示に従って作製されることができる。

【０１０９】

ライブラリーのオリゴはプレート上に、例えば３８４又は９６ウェルフォーマットで配置されることができる。それらはスポッティングされるか又はすでにプリントされており、かくしてマイクロアレイ分析において便利に適用されることができる。

【０１１０】

他の種、例えばウシ、ブタ又はアラバドプシスのためのオリゴヌクレオチドライブラリーも本発明に従って同様に構築されることができる；これらのライブラリーはマイクロアレイ分析において、及び当該技術分野で知られている他のＲＮＡ又はｃＤＮＡ検出システムにおいて直ちに用いられることができる。

【０１１１】

実施例２．ＲＮＡ転写物全て又は個別のスプライス変異体を検出するためのミニオリゴヌクレオチドライブラリー

ヒトのオリゴミニライブラリーはヒトのがん／アポトーシス遺伝子、肥満症／糖尿病遺伝子、及び毒素遺伝子について構築された。６０塩基の長さのオリゴヌクレオチドががん／アポトーシス、オベシティ／糖尿病、及び毒素に関与する遺伝子についてＵ．Ｓ．出願シリアルＮｏ．０９／１３３９８７に外略が示されているアプローチを用いて、また選択的スプライシング現象を考慮に入れながらそれぞれ設計又は選択された。これらのオリゴは次に上記実施例１で議論したのと同様のことを考慮に入れて合成された。

【０１１２】

ヒトのがん／アポトーシス遺伝子の個別のスプライス変異体及び／又はそれらの転写物を検出することができるヒトのオリゴミニライブラリー中のオリゴヌクレオチドは２０１個の配列を含む。ヒトのがん／アポトーシス遺伝子からの存在する又は予測されるスプライス変異体を全て検出することができるヒトのオリゴミニライブラリー中のオリゴヌクレオチドは２８０個の配列を含む。毒性に関与する遺伝子の個別のスプライス変異体を検出することができるヒトのオリゴミニライブラリー中のオリゴヌクレオチドは２３４個の配列を含み、そして、ヒトの毒性遺伝子からの存在する又は予測されるスプライス変異体を全て検出することができるヒトのオリゴミニライブラリー中のオリゴヌクレオチドは９６個の配列を含む。ヒトの肥満症／糖尿病遺伝子の個別のスプライス変異体を検出することができるヒトのオリゴミニライブラリー中のオリゴヌクレオチドは１９５個の配列を含み、そしてヒトの肥満症／糖尿病遺伝子からの存在する又は予測されるスプライス変異体を全

て検出することができるヒトのオリゴミニライブラリー中のオリゴヌクレオチドは９２個の配列を含む。再び、上記配列の数は絶対的ではなく、それ故、異なる数の配列を含む他のライブラリーも出願人の教示に従って作製されることができる。

【０１１３】

６つのミニライブラリー中の全てのオリゴの配列は本願に沿って提出された配列表中に含まれている。これらのオリゴはプレート上に、例えば３８４又は９６ウェルフォーマットで配置されることができる。それらはすでにプリントされており、かくしてマイクロアレイ分析において便利に適用されることができる。それ故、病気特異的に又はディファレンシャルに発現される遺伝子はヒトの毒性、糖尿病／肥満症、及びがん／アポトーシス状況についてのこれらのオリゴライブラリー及びマイクロアレイをそれぞれ用いて検出又はモニターされることができる。

【０１１４】

本発明は教示の明確さの目的で例を用いて詳細に説明されてきたが、以上の説明は本発明の範囲を限定することを意図されていないということは理解されるべきである。本発明の教示に鑑みて当業者には明らかな他の側面、利点及び変形は特許請求の範囲の範囲内にある。

【図面の簡単な説明】

【図１】

三つのエクソン及びそこから転写される２つの選択的にスプライスされた転写物を有する転写単位を示す図である。

【図２】

三つのエクソン及びそこから転写される２つの異なるスプライス変異体を有する転写単位を示す図である。

【図３】

五つのエクソン及びそこから転写される４つの選択的にスプライスされた転写物を有する転写単位を示す図である。

【図１】

Fig. 1



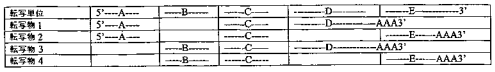
【図２】

Fig. 2



【図３】

Fig. 3



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10449 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/IB01/01903

(22) International Filing Date: 20 July 2001 (20.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/221,607 28 July 2000 (28.07.2000) US
60/287,724 2 May 2001 (02.05.2001) US

(71) Applicant *(for all designated States except US)*: **COMPUGEN INC.** [US/US]; 9700 Great Seneca Highway, Rockville, MD 20850 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SHOSHAN, Avi** [IL/US]; 37 West 75th Street, New York, NY 10023 (US). **WASSERMAN, Alon** [IL/US]; Apartment 3C, 328 West 19th Street, New York, NY 10011 (US). **MINTZ, Eli** [IL/US]; 16 Nancy Street, Kendall Park, NJ 08824 (US). **MINTZ, Liat** [IL/US]; 16 Nancy Street, Kendall Park, NJ 08824 (US). **FAIGLER, Simchon** [IL/US]; 2605 Wildberry Court, Edison, NJ 08817 (US).

(74) Agent: **ISACSON, John, P.**; Heller, Ehrman, White & Mcauliffe, Suite 300, 1666 K Street, N.W., Washington, DC 20006 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE VARIANTS THAT POPULATE A TRANSCRIPTOME

Transcription Unit	5'---------A----------	----------B----------	---------C----------3'
Transcript 1	5'---------A----------		---------C-------AAA3'
Transcript2	5'	----------B----------	---------C-------AAA3'

(57) Abstract: The present invention provides oligonucleotide libraries capable of detecting RNA transcripts and RNA splice variants which populate a transcriptome and which are transcribed from genes or transcription units that populate the corresponding genome. The present invention also provides oligonucleotide arrays generated from the oligonucleotide libraries and methods of using the oligonucleotide libraries in various oligonucleotide detection systems and expression profiling studies. Antisense molecules and double-stranded interfering RNAs, which are types of oligonucleotides, based on the oligonucleotides disclosed herein also are provided.

WO 02/10449 A2

OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE VARIANTS THAT POPULATE A TRANSCRIPTOME

Cross Reference to Related Applications

Incorporated by reference are U.S. Serial No. 60/287,724, filed May 2, 2001; U.S. Provisional Application Ser. No. 60/221,607, filed on July 28, 2000; and U.S. Application Ser. No. 09/133,987, filed on August 13, 1998. This application claims priority to U.S. Serial No. 60/287,724, filed May 2, 2001 and U.S. Provisional Application Ser. No. 60/221,607, filed on July 28, 2000.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention provides oligonucleotide libraries that are useful for detecting messenger RNAs from a biological sample. More specifically, the present invention provides oligonucleotide libraries capable of detecting RNA transcripts, including RNA splice variants, which populate a transcriptome and which are transcribed from genes or transcription units that populate the corresponding genome. The present invention also provides oligonucleotide arrays generated from the oligonucleotide libraries and methods of using the oligonucleotide libraries in various oligonucleotide detection systems and expression profiling studies.

DESCRIPTION OF THE FIELD

Understanding biological functions and processes, normal or abnormal, in a cell, a tissue, an organ, and an organism hinges upon the knowledge of genes operative in the biological locale of interest, the messenger RNAs transcribed from these genes, and ultimately the set of proteins produced from the RNA transcripts. At the species level, the repertoire of all DNA molecules makes up the genome of that species; among these DNA molecules, the ones that are transcribed into RNA molecules are referred to as genes or transcription units. Correspondingly, the repertoire of all messenger RNA molecules transcribed from transcription units (hence "RNA transcripts" or "transcripts") in a genome makes up a transcriptome; and, the repertoire of all proteins translated from messenger RNA molecules makes up a proteome.

Studies on a transcriptome of a species shed light on researchers' understanding of the corresponding genome and proteome of that species. Various techniques for identifying RNA transcripts and determining their levels of abundance in a biological sample have thus been instrumental in deciphering transcriptomes of a variety of species and thereby facilitating discoveries relating to the corresponding genomes and proteomes. These techniques include both gel-based procedures such as Northern Blot analysis, Dot Blot analysis, Primer Extension analysis, Substrate Enrichment Hybridization Analysis, Differential Display analysis, Polymerase Chain Reaction (PCR) based analysis and, most recently, chip-based procedures such as DNA microarray analysis.

The completion of whole genome sequencing for a number of species promises better efficiency, reliability, and comprehension in studying these genomes and their corresponding transcriptomes and proteomes. It is estimated that the total number of expressed genes or transcription units in a human

genome is around 30,000 – 40,000 (Venter G. et al., Science 2000, Vol, 291: 1304-1351; The Genome International Sequencing Consortium, Nature 2000, Vol. 409: 860-921). However, the number of proteins encoded in a human proteome is expected to significantly exceed this estimate because, in many cases, more than one RNA splice variant is transcribed from a transcription unit or a gene. Estimates vary, but some researchers believe that a human transcriptome may contain up to 500,000 RNA transcripts and that, more than 30% of genes or transcription units in the human genome produce several RNA splice variants. (Mironov et al. 1999, Genome Research 9:1288-1293). These numbers are considered by others to be conservative. Alternative splicing also occurs in rat and mouse with similar frequencies, for example, and in lower organisms, such as Drosophila melanogaster and Caenorhabditis elegans. It should be noted that a special case of alternative splicing is alternative polyadenylation and that a significant number of genes or transcription units have alternative polyadenylation sites.

Proteins translated from different RNA splice variants may have significantly different biological functions. Different splice variants may be expressed in different tissues, different developmental stages, and different disease states. The detection of RNA transcripts and RNA splice variants from a transcription unit and the determination of (i) level of abundance of all RNA transcripts, including the splice variants from the transcription unit and (ii) levels of abundance of a subset of the splice variants or one splice variant are, therefore, desirable in accurately capturing the state of a transcriptome and hence, the corresponding proteome. And yet, the qualitative and quantitative detection of RNA splice variants remains an unmet challenge in the field of molecular biology, particularly on the scale of a transcriptome.

Oligonucleotide libraries have been used in both gel based and chip based systems to detect RNA molecules and measure their levels of abundance.

However, oligonucleotides ("oligos") of these libraries, which are used as oligo probes for hybridizing to target RNAs, are typically incapable of accurately detecting splice variants. That is, they typically do not contain oligos capable of hybridizing (thus recognizing) one or more specific splice variants, and therefore cannot effectively identify and/or distinguish alternatively spliced RNA transcripts or determine the levels of abundance of different splice variants.

SUMMARY OF THE INVENTION

It is therefore an object of this invention to provide oligonucleotide libraries that are capable of detecting RNA transcripts and RNA splice variants transcribed from transcription units in a genome, thereby qualitatively and quantitatively characterizing the corresponding transcriptome. It is another object of this invention to provide oligonucleotide arrays that are generated from the oligonucleotide libraries and methods of using the oligonucleotide libraries to study a transcriptome of interest and, methods of using the oligonucleotide libraries in expression profiling studies.

In accordance with the present invention, there is provided, in an embodiment, an oligonucleotide library for detecting messenger RNAs that populate a transcriptome, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

According to another embodiment of the invention, the oligonucleotide library is capable of detecting RNA transcripts and splice variants of a human transcriptome. According to another embodiment, the oligonucleotide library of the present invention is capable of detecting RNA transcripts and splice variants of a rat transcriptome. According to yet another embodiment, the oligonucleotide library of this invention is capable of detecting RNA transcripts and splice variants of a mouse transcriptome. According to a further embodiment, the oligonucleotide library of this invention is capable of detecting RNA transcripts and splice variants of a arabadopsis transcriptome. According to a still further embodiment, the oligonucleotide library of this invention is capable of detecting RNA transcripts and splice variants of a drosophila melanogaster transcriptome.

In accordance with the present invention, there is provided, in another embodiment, an oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome (a portion of a transcriptome that bears certain biological origin, structure, or function traits, e.g., tissue specificity, disease specificity, developmental specificity, etc.) of a tissue origin, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

According to the invention, the tissue origin of the oligonucleotide library may be kidney, brain, heart, lung, bone, liver or other tissue of interest in various embodiments.

In accordance with the present invention, there is provided, in yet another embodiment, an oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a pathological tissue origin, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome (a portion of a genome that bears certain biological origin, structure, or function traits, e.g., tissue specificity, disease specificity, developmental specificity, etc.) of the pathological tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

According to the invention, the pathological tissue origin may be cancer tissue, such as colon cancer tissue, breast cancer tissue, lung cancer tissue, bone cancer tissue, prostate cancer tissue, brain cancer tissue, or liver cancer tissue, in various embodiments. The pathological tissue origin, according to the invention, may also be abnormal heart tissue, abnormal neuronal tissue, abnormal liver tissue or abnormal kidney tissue in various embodiments, for example.

In accordance with the present invention, there is provided, in still another embodiment, an oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a developmental stage, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the developmental stage, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the sub

genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

According to the invention, the developmental stage may be human neural induction, mouse mesoderm induction, human erythrocyte differentiation, human or rat stem cell development, or other specific developmental biology states in various species.

In accordance with the present invention, there is provided, in a further embodiment, an oligonucleotide library for detecting messenger RNAs that populate a transcriptome of patients suffering from a disorder, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome of patients suffering from the disorder, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs (one or more RNA transcripts) transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

According to the invention, the disorder may be cancer, such as colon cancer, breast cancer, lung cancer, bone cancer, prostate cancer, brain cancer, or liver cancer, in various embodiments. The disorder, according to the invention, may also be Alzheimer's disease, Parkinson's disease, osteoporosis, diabetes, rheumatoid arthritis, or other disease of interest, in various embodiments.

In accordance with the present invention, there is provided, in a still further embodiment, an oligonucleotide library for detecting messenger RNAs that populate a transcriptome, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a

genome, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

According to one embodiment of the invention, the oligonucleotide library is capable of detecting RNA transcripts and splice variants of a human transcriptome. According to another embodiment, the oligonucleotide library of the present invention is capable of detecting RNA transcripts and splice variants of a rat transcriptome. According to yet another embodiment, the oligonucleotide library of this invention is capable of detecting RNA transcripts and splice variants of a mouse transcriptome.

In accordance with the present invention, there is provided, in another embodiment, an oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a tissue origin, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

According to the invention, the tissue origin of the oligonucleotide library may be kidney, brain, heart, lung, bone, liver or other tissue of interest in various embodiments.

In accordance with the present invention, there is provided, in yet another embodiment, an oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a pathological tissue origin, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the pathological tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

According to the invention, the pathological tissue origin may be cancer tissue, such as colon cancer tissue, breast cancer tissue, lung cancer tissue, bone cancer tissue, prostate cancer tissue, brain cancer tissue, or liver cancer tissue, in various embodiments. The pathological tissue origin, according to the invention, may also be abnormal heart tissue, abnormal neuronal tissue, abnormal liver tissue or abnormal kidney tissue in various embodiments.

In accordance with the present invention, there is provided, in still another embodiment, an oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a developmental stage, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the developmental stage, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

According to the invention, the developmental stage may be human neural induction, mouse mesoderm induction, human erythrocyte differentiation, or other specific developmental states in various species.

In accordance with the present invention, there is provided, in a further embodiment, an oligonucleotide library for detecting messenger RNAs that populate a transcriptome of patients suffering from a disorder, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome of patients suffering from the disorder, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

According to the invention, the disorder may be cancer, such as colon cancer, breast cancer, lung cancer, bone cancer, prostate cancer, brain cancer, or liver cancer, in various embodiments. The disorder, according to the invention, may also be Alzheimer's disease, Parkinson's disease, osteoporosis, diabetes, rheumatoid arthritis, or other disease of interest, in various embodiments.

The oligonucleotide libraries of the present invention may have at least 95, 200, 400, 600, 800, or 1000 oligonucleotides in various embodiments. The oligonucleotides in the library of the present invention may be 75, 65, 60, 55, 50, 45, 40, 35, 30, or 20 bases in length, or any integer thereabout or therebetween, in various embodiments of the invention, although smaller and larger oligonucleotides may be employed according to the invention.

In accordance with the present invention, in a still further embodiment, each oligonucleotide of the aforementioned oligonucleotide libraries of the invention in various embodiments contains a modification that enables attachment to a solid surface. The attachment may be covalent or electrostatic, for example.

In accordance with the present invention, there is provided, in another embodiment, a DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the aforementioned oligonucleotide libraries of the invention in various embodiments.

In accordance with the present invention, there is provided, in yet another embodiment, a method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining (i) the total level of the set of splice variants, (ii) the total level of a subset thereof, or (iii) the level of one splice variant thereof, wherein said plurality is provided by the aforementioned oligonucleotide library the invention in various embodiments.

According to the invention, the method for expression profiling may, in various embodiments, employ a nucleotide chip, a membrane or filter, an electrophoresis gel and a filter or membrane imprinted therefrom, or other similar platforms, on which hybridization is carried out.

In accordance with yet another aspect of the invention, there are provided methods of inhibiting or preventing translation using the

oligonucleotides of the present invention or DNA constructs based on the oligonucleotides, wherein the constructs can yield antisense RNA at the cellular level.

In accordance with still another aspect of the invention, the oligonucleotides of the libraries can be single stranded, double stranded or partially double stranded.

In accordance with a still further aspect of the invention, there are provided double stranded oligonucleotides that can cause post-transcriptional silencing of specific genes. In one embodiment, these oligonucleotides are short double-stranded RNAs

Brief Description of Drawings

Fig. 1. This is a diagram demonstrating a transcription unit with three exons and, transcribed therefrom, two alternatively spliced transcripts.

Fig. 2. This is a diagram demonstrating a transcription unit with three exons and, transcribed therefrom, two different splice variants.

Fig. 3. This is a diagram demonstrating a transcription unit with five exons and, transcribed therefrom, four alternatively spliced transcripts.

DETAIL DESCRIPTIONS OF PREFERRED EMBODIMENTS

Libraries, Oligonucleotides, and Modifications Thereof

To characterize a transcriptome or a portion thereof, oligonucleotides used as probes to test hybridization to messenger RNA samples must be

representative of the total population of RNA transcripts or a portion thereof that is of interest. To generate this type of oligonucleotide information, the approach outlined in U.S. Patent Application Ser, No. 09/133,987 ("the '987 application") can be employed.

The large pool of publicly known nucleotide sequences, sequence fragments, or expressed sequence tags ("EST"), e.g., GenBank, EMBL, DBest sequence collections, are subject to rigorous clustering and assembly procedures to obtain sequence clusters and longer contigs (contiguous sequences, i.e., longer sequences assembled from shorter ESTs). These procedures give rise to sequences that are better representations of genes or transcription units compared to the large pool of ESTs and other sequence fragments.

A representative sequence for a given gene or transcription unit is chosen and from this sequence one or more oligonucleotide (or oligo, oligo probe) is derived. The representative sequence has high specificity to the target gene or transcription unit (i.e., it shares little homology to sequences from other genes or transcription unit therefore the frequency of non-specific binding is minimized); the sequence quality is good (i.e., the gene sequence information is accurate). The oligonucleotide probe has high sensitivity, i.e., the affinity of hybridization of the probe to the target gene or transcription unit is high. The probe preferably recognizes a sequence close to the 3' terminus of the target gene or transcription unit such that in certain procedures where target gene sample is prepared the yield of which is maximized. Additionally, the oligonucleotide lacks secondary structure (e.g., hairpin) or tendency to form the same, such that hybridization is not hampered.

Most importantly, a representative sequence is chosen and one or more oligonucleotides are designed taking into account alternative splicing of the

target gene and transcription unit. As such, the oligonucleotide may be used to hybridize, and hence detect, either (i) all the splice variants of a particular gene or transcription unit, (ii) a subset of all the splice variants, or (iii) one of the splice variant of the gene or transcription unit.

Referring to Fig. 1, a transcription unit or gene has, for example, three exons: A, B, and C, and is alternatively spliced in transcription giving rise to two variants: transcript 1 (AC) and transcript 2 (BC). An oligo complimentary to sequence A, or a fragment thereof, would therefore only detect transcript 1, not transcript 2. An oligo complimentary to sequence B or a fragment thereof, by contrast, would only detect transcript 2, not transcript 1.

The scenario shown in Fig. 2 is slightly different. The transcription unit has three exons: A, B, and C, which are transcribed into two splice variants, transcript 1 (AC) and transcript 2 (ABC). An oligo complementary to sequence A or a fragment thereof would be able to detect both transcript 1 and transcript 2 and, collectively, measure the level of abundances of both. In contrast, the oligo complementary to sequence B or a fragment thereof would be able to detect transcript 2 not transcript 1, thereby providing direct information on the abundance of transcript 2.

Fig. 4 demonstrates a transcription unit or a gene that has 5 exons: A, B, C, D, and E. There are four alternatively spliced transcripts: transcript 1 (ACD), 2 (ACE), 3 (BCD), and 4 (BCE). This is a more complex scheme but the principle remains the same: An oligo complementary to sequence A or a fragment thereof would be able to only detect transcripts 1 and 2 (hence "1+2=A"); an oligo complementary to sequence B would be able to only detect transcripts 3 and 4 (hence "3+4=B"); an oligo complementary to sequence D would be able to only detect transcripts 1 and 3 (hence "1+3=D"); and an oligo complimentary to sequence E would be able to only detect transcripts 2 and 4

(hence "2+4=E") Therefore, we have a number of oligos that can specifically detect a subset of the splice variants. And, additionally, resolving the four polyvariance equations above would reveal the level of abundance of each splice variant (A, B, D, and E). Moreover, an oligo complementary to sequence C or a fragment thereof would be able to detect all four transcripts and thus measure the total level of abundances of these splice variants.

Analogously, according to the present invention, alternative polyadenylation is taken into account such that a representative sequence and one or more oligonucleotide probe are generated for a particular gene or transcription unit that has alternative polyadenylation sites. These probes may be used to hybridize, and hence detect, either (i) all the alternative polyadenylated transcripts, (ii) a subset of these alternative polyadenylated transcripts, or (iii) one of the polyadenylated sites of the target gene or transcript.

The oligonucleotides of the present invention may be modified at the termini to facilitate their application in a gel-based system or attachment on a chip or array-based system for RNA detection. For example, in one embodiment, the oligonucleotides are modified at the 5' terminus: A 5' C6-amino modification is performed to enable covalent attachment of oligonucleotides on a glass array surface. The carbons function as a spacer and the reactive amine group interacts with aldehydes that are covalently attached to a glass surface. Pre-processed glasses are commercially available from a number of vendors, e.g., ArrayIt.com (http://arrayit.com), Corning (http://www.corning.com), Clontech (http://www.clontech.com). These glasses, and other similar ones, are suitable for oligonucleotide binding and thereby for making oligo arrays using the oligonucleotide libraries according to this invention. The attachment may be covalent or electrostatic, for example. One example is binding of oligos on poly-L-lysine glass, which is electrostatic:

poly-L-lysine is positively charged whereas phosphate backbone of DNA is negatively charged. Another example is binding of oligos on aldehyde glass, which is covalent.

Oligonucleotides Used with Gel-Based and Array-Based Platforms

As mentioned above, various detection platforms may be used according to the present invention to measure RNA species in a transcriptome. In one embodiment, a gel-based platform is used; in another embodiment, a chip or array-based platform is used. When a gel-based platform is used, techniques such as Northern Blot analysis, Dot Blot analysis, Primer Extension analysis, Substrate Enrichment Hybridization Analysis, Differential Display analysis, Polymerase Chain Reaction (PCR) based analysis, or other similar procedures may be employed. Arrays also can be made by spotting directly on a substrate like nitrocellulose. Essentially, oligonucleotides of this invention are labeled (via radioactive-labeling, fluorescent-labeling, or using other suitable labeling methods and capture moieties known in the art) and are hybridized to a filter or membrane imprinted from the gel in which RNA or cDNA samples have been run. Hybridization signals indicate the identity and abundance of the RNA transcripts (and splice variant(s) thereof) that the oligonucleotide probes represent. These gel-based nucleotide hybridization and detection procedures are well known to a skilled molecular biologist. See generally, Sambrook J. and Russell, 2001, Molecular Cloning, a Laboratory Manual, 3rd Ed. In some cases where separation of nucleotides is not involved, nucleotide probes and samples may be directly applied to a membrane or filter for hybridization without the use of an electrophoresis gel.

When certain chip or array-based platforms are used, the oligonucleotides of the present invention are spotted or printed on a solid

surface, e.g., a glass slide, via various attachment chemistries as discussed above or via other suitable attachment procedures. The slide is then treated in a manner similar to a filter or membrane in a typical gel-based separation and hybridization experiment and, subsequently, probed with labeled (via radioactive-labeling, fluorescent-labeling, or by other suitable labeling methods) cDNA or RNA molecules that are derived from a biological sample such as a tissue or cell line of interest. In some cases, labeled clones or total genomic DNA may be used instead. Following hybridization, the slide is washed and scanned to read and record the hybridization signals. These signals represent the identity and abundance of the RNA transcripts (and splice variant(s) thereof) that the oligonucleotide probes correspond to and detect. Such measurement of RNA transcripts and their splice variants in a particular biological sample therefore provides a snapshot of the transcriptome under a particular biological state – normal, pathological, or treated – and hence helps to elucidate the characteristics and interactions in different biological states.

According to the invention all oligonucleotides of a library, or a portion thereof, may be spotted on a single microarray and, therefore, a single hybridization procedure may test a multiplicity of transcripts or their corresponding genes. As a variation, in another embodiment, oligonucleotides may be directly synthesized on the surface of an array. Methods of oligo synthesis on a solid surface are known in the art. See, e.g., U.S. Patent No. 5,593,839. In this embodiment, the oligonucleotides in the library preferably are around 20 or 25 bases in length, since *in situ* synthesis of longer sequences is often error prone. The libraries of the present invention that are directly synthesized on an array thus may be useful as specialized or mini libraries designed to detect transcripts of a sub-transcriptome under a particular biological state.

Uses for Oligonucleotide Libraries

Whether gel-based, array-based, or other suitable detection systems are employed, according to the present invention, the oligonucleotide library may be constructed to detect RNA transcripts and splice variants of the transcriptome of a given species, such as a human transcriptome, a mouse transcriptome, or a rat transcriptome. The library also may be constructed to detect RNA transcripts and splice variants of a sub-transcriptome that is of a specific biological or pathological origin or in a specific biological or pathological state.

For example, in one embodiment of this invention, the sub-transcriptome is of a specific tissue origin. That is, it may be a kidney sub-transcriptome, a brain tissue sub-transcriptome, a heart tissue sub-transcriptome, a lung tissue sub-transcriptome, a bone tissue sub-transcriptome, a liver tissue sub-transcriptome, etc. The libraries of the invention can thus be used to specifically detect RNA transcripts and splice variants that populate the sub-transcriptome of a tissue of interest. Tissue-specific expression of certain genes (i.e., certain RNA transcripts or splice variants that exist only in some sub-transcriptome(s) but not in others) or differential expression of certain genes in a tissue-specific manner (i.e., certain RNA transcripts or splice variants are more – or less – abundant in some sub-transcriptome(s) than in another) therefore can be detected and monitored using the oligonucleotide library of the present invention. The presence or absence of a transcript or specific splice variant and the differences in their abundance may be evaluated under a statistical significance standard. Suitable statistical evaluations are known in the art.

In another embodiment, the sub-transcriptome is of a particular pathological origin. That is, it may be a cancer tissue sub-transcriptome, e.g., a

colon cancer tissue sub-transcriptome, a breast cancer tissue sub-transcriptome, a lung cancer tissue sub-transcriptome, a bone cancer tissue sub-transcriptome, a prostate cancer tissue sub-transcriptome, a brain cancer tissue sub-transcriptome, or a liver cancer tissue sub-transcriptome. It may also be an abnormal heart tissue sub-transcriptome, an abnormal neuronal tissue sub-transcriptome, an abnormal liver tissue sub-transcriptome, or an abnormal kidney tissue sub-transcriptome, etc.

The libraries of the invention can therefore be used to specifically detect RNA transcripts and splice variants that populate the sub-transcriptome of a pathological origin of interest. This application, therefore, allows detection of tissue- and pathology- specific genes, i.e., genes that are only expressed in a specific tissue and under a specific pathological condition, since the corresponding transcripts and splice variants can only be found in the specific sub-transcriptome. Additionally, this application further permits detection of genes that are differentially expressed in a tissue- and pathology-specific manner, i.e., genes that are expressed at a different level in a specific tissue and under a specific pathological condition than in other tissues or under normal or other conditions, because the corresponding transcripts and splice variants are more – or less – abundant in the specific sub-transcriptome. The presence or absence of a transcript or splice variant and the differences in their abundance may be evaluated under a statistical significance standard.

In yet another embodiment, the sub-transcriptome is of a certain developmental stage. That is, it may be a sub-transcriptome from human neural induction, mouse mesoderm induction, human erythrocyte differentiation, or other specific developmental biology states in various species. The libraries of the invention can therefore be used to specifically detect RNA transcripts and splice variants that populate the sub-transcriptome of a developmental biology stage of interest. This application allows detection

of developmental specific genes, i.e., genes that are only expressed in a specific developmental state, since the corresponding transcripts and splice variants can only be found in the specific sub-transcriptome. Additionally, this application also allows detection of genes that are differentially expressed in a developmental-dependent manner, i.e., genes that are expressed at a different level in a specific developmental state than in other state, because the corresponding transcripts and splice variants are more – or less – abundant in the specific sub-transcriptome. The presence and absence of a transcript or splice variant and the differences in their abundance may be evaluated under a statistical significance standard.

In a further embodiment, the oligonucleotide library is constructed to detect RNA transcripts and splice variants of a transcriptome of a patient suffering from a particular disorder. That is, it may be the transcriptome of a cancer patient, e.g., a colon cancer patient, a breast cancer patient, a lung cancer patient, a bone cancer patient, a prostate cancer patient, a brain cancer patient, or a liver cancer patient. It may also be the transcriptome of an Alzheimer patient, a Parkinson's patient, an osteoporosis patient, a diabetes patient, a rheumatoid arthritis, or a patient of other disease of interest.

The present invention allows detection of disease specific genes, i.e., genes that are only expressed in patients afflicted with a specific disease, since the corresponding transcripts and splice variants can only be found in the transcriptome of such patients. Additionally, the present invention allows detection of genes that are differentially expressed in patients of a specific disease, i.e., genes that are expressed at a different level in patients of a specific disease than in normal or healthy individuals, because the corresponding transcripts and splice variants are more – or less – abundant in the transcriptome of such patients. The presence and absence of a transcript or splice variant and the differences in their abundance may be evaluated under a

statistical significance standard, e.g., requiring a specific p score. Furthermore, this application allows individualized characterization of expression patterns or profiles (the compilation of the identity and abundance of transcripts and splice variants of a given patient) of patients suffering from a particular disease and therefore provides a basis for development of individualized therapeutics.

To summarize, the oligonucleotide libraries of the present invention may be used in a variety of functional contexts to detect RNA transcripts and splice variants and thereby characterize a transcriptome or sub-transcriptome of interest. In this connection, the scope of the sub-transcriptome may also be delimited by protein functions. For example, an oligonucleotide library may be used to detect RNA transcripts and splice variants of a sub-transcriptome that corresponds to cell surface antigens. A list of more than 40 protein functional groups, as set forth below (see also, U.S. Provisional Application Ser. No. 60/221,607), is useful in the present invention, for example, for constructing oligonucleotide libraries capable of characterizing sub-transcriptomes that give rise to sub-proteomes or proteins of specific functions and abnormality thereof.

Group 1. Adaptor-binding proteins. These are proteins that are associated to other cell components – binding to or interacting with – in maintaining their structural integrity and performing its functional activities.

Group 2. Adhesion molecules. These are proteins involved in modulation of adhesion between adjoining cells.

Group 3. Apolipoproteins. These are proteins that are part of lipoprotein particles and that function in cellular signaling in binding and internalization of these particles. Apolipoprotein defects are found in diseases which involve abnormally high or low levels of lipoprotein and cholesterol, as well as conditions involved in the formation or arteriosclerosis.

Group 4. Apoptosis related proteins. These are proteins and enzymes that are involved in the apoptosis pathway, in an inhibitory or stimulatory manner. Abnormalities of these proteins cause diseases which are involved in premature death of cells, such as degenerative diseases, for example neurodegenerative diseases or conditions associated with aging, or alternatively, diseases wherein required apoptosis does not take place. Example of such diseases are cancerous diseases and loss of cardiac function after myocardial infarction.

Group 5. Cancer related proteins. These proteins, including DNA repair proteins, tumor markers and antigens, tumor suppressors, and messenger molecules participating tumorigenesis, etc., are involved in various kind of cancers and the treatment and detection of the corresponding metastasis states.

Group 6. Carboxylases. Abnormality of these proteins causes mal-regulation of enzymatic reactions that remove CO_2 groups from other moieties.

Group 7. Cell surface antigens. Proteins expressed on the surface of cells. Abnormalities of these proteins can be seen in autoimmune disease, e.g., AIDS and cancers that involve cell surface antigens, for example.

Group 8. Proteins controlling cell growth. These proteins have defective structure or function in degenerative diseases (low growth) or cancerous diseases (uncontrolled growth).

Group 9. Coagulation – related proteins. Abnormality of these proteins may cause hemophilia or stroke and blockage of blood vessels, for example.

Group 10. Converting enzymes. These enzymes convert one protein to another by specific cleavage of the precursor protein.

Group 11. Cyclase enzymes. These enzymes convert triphosphate to cyclic monophosphate. Abnormality of these enzymes causes insufficient or excessive conversion of triphosphate to cyclic monophosphate thereby affecting cellular signaling.

Group 12. Proteins involved in protein degradation. Abnormalities of these proteins may cause abnormal degradation of other proteins, which may lead to abnormal accumulation of various proteinaceous product in cells.

Group 13. Proteins involved in development. Abnormalities of these proteins are manifested in genetic diseases involving abnormal development of a fetus, for example.

Group 14. Domain proteins. These are proteins that are involved in protein-protein interactions.

Group 15. Esterase. These proteins cleave an ester bond.

Group 16. Growth factors. Examples of these proteins include cytokines, interleukins, interferon, and lymphokines, etc. These proteins are implicated in autoimmune diseases, inflammation related disease, Graft vs. Host disease, diseases caused by infectious agents, and cancer, to name a few.

Group 17. Hormones and poietin proteins. Abnormality of these proteins are seen in various endocrine disorders.

Group 18. Housekeeping proteins. Examples are homeobox proteins, heat shock proteins, and chaperonins, etc.

Group 19. Hydrolases. These enzymes modify hydroxyl groups; examples are hydrogenase, dehydrogenase, hydrolase, and hydroxylase.

Group 20. Immunology-related genes. These are proteins that are involved in the immune system, including antigens, antibodies, and their associated proteins. These proteins are important in pathological conditions such as inflammation, autoimmune diseases, infectious diseases, and cancerous processes.

Group 21. Inhibitors. These proteins inhibit the function of other proteins in cellular processes.

Group 22. Kinases. Abnormality of these proteins may lead to defective cellular signaling.

Group 23. Lipases, phospholipases, and lysophospholipases. These proteins are implicated in abnormal lipid metabolism.

Group 24. Cell matrix and cytoskeleton proteins.

Group 25. Modifying enzymes. These proteins include a number of miscellaneous enzymes such as paraoxonase, GTPase, ATPase, and anhydrase. Malfunctions of these enzymes are implicated in a variety of cellular processes.

Group 26. Mutases and superoxidedismutases. These proteins are implicated in cancer and various other pathological processes involved in aging.

Group 27. Neurology related proteins. These proteins are involved in central nervous system disorders, including various types of dementia, neurodegenerative diseases, epilepsy, psychiatric disorders, etc.

Group 28. Oxidases and peroxidases. Abnormalities of these enzymes may cause metabolic problems involving peroxide, for example.

Group 29. Oxygenases, mono- and di-oxygenases.

Group 30. Phosphatases and phosphorylases.

Group 31. Phosphoproteins and phospholipids.

Group 32. Proteases, peptidases, and proteinases.

Group 33. Receptors.

Group 34. Reductases.

Group 35. Secreted proteins. These proteins include hormones, neurotransmitters, and various other proteins secreted by cells to the extracellular environment.

Group 36. Signal-transduction proteins. A G protein is one example of this group of proteins.

Group 37. Sub-cellular proteins. Examples of these proteins include ribosomal proteins.

Group 38. Synthases and synthetases.

Group 39. Proteins involved in nucleotide interactions. These include transcription factors, RNA and DNA binding proteins, zinc fingers, helicase, isomerase, histones, nucleases.

Group 40. Transferases. These are enzymes involved in transfer of protein functional groups.

Group 41. Translational-factors. These are proteins and enzymes involved in the translational process, such as elongation and initiation factors. Abnormalities in these proteins may impair cellular protein production.

Group 42. Transporters. These proteins mediate the transport of molecules and macromolecules, including channels, exchangers, and pumps.

Expression Profiling Using Oligonucleotide Libraries

One application of the oligonucleotide libraries according to this invention is expression profiling, which is mentioned supra in various places but which warrants some additional discussion. Expression profiling, also termed gene expression profiling, as used in this invention, means quantitative and qualitative determination of RNA transcripts and splice variants in a biological sample. The biological sample may be a tissue or a cell line sample, a normal or diseased tissue sample, a diseased tissue sample or a tissue sample after some period of treatment, a sample taken from different developmental biology state, a biological tissue or fluid sample taken from a patient at various time points in the course of application of a drug or a treatment regime. Expression profiling establishes expression profiles or expression patterns, which are compilations of identity and abundance of RNA transcripts and slice variants, for a particular biological source or a particular physiological or pathological state.

Expression profiling may be performed via a gel (and/or membrane)-based or array-based system according to the present invention, as discussed supra. Essentially, expression profiling of a biological sample that contains two or more RNAs of various abundances is preformed by measuring hybridization signals of the sample to a plurality of oligonucleotide sequences, thereby determining the levels of the RNA transcripts in the sample and, where an RNA is transcribed from a transcription unit that has a set of splice variants, determining (i) the total level of the set of splice variants, (ii) the total level of a subset thereof, or (iii) the level of one splice variant thereof. The plurality of

oligonucleotides is provided by the oligonucleotide library of the present invention in various embodiments discussed above.

Accordingly, the present invention encompasses oligonucleotide libraries and sub-libraries thereof, custom or modified oligonucleotides, oligonucleotide arrays having spotted thereon the oligonucleotides from the libraries, methods of using oligonucleotides in various nucleotide detection systems, and methods of expression profiling using the oligonucleotide libraries.

ANTISENSE USES AND RNA INTERFERENCE

Because the oligonucleotides of the present invention are able to bind to one or more splice variants of the transcriptome, these oligonucleotides have uses in *in situ* and in *vivo* antisense contexts. For example, single-stranded antisense DNA oligonucleotides themselves, or double stranded DNA oligonucleotides in denaturing contexts, can be injected into a cell, such as a mammalian cell (including human cells) and the body, such as a mammalian body (including humans), and be expected to bind to mRNA transcripts, which would inhibit or prevent translation of the mRNA into protein. Moreover, the sense strand of an double-stranded oligonucleotide or a single-stranded sense oligonucleotide can be used as an initial design template for creation of antisense RNA. Using approaches known in the art for *in vivo* expression of antisense RNA molecules, an antisense RNA based upon the oligonucleotides of the present invention can be employed to inhibit or prevent translation of an mRNA at the cellular level.

The oligonucleotides of the library also can be employed in an RNA interference context. The phenomenon of RNA interference is discussed in Bass, *Nature* 411: 428-29 (2001); Elbahir *et al., Nature* 411: 494-98 (2001);

and Fire *et al.*, *Nature* 391: 806-11 (1998), where methods of making interfering RNA also are discussed. The double-stranded RNA based upon a sequence disclosed herein is less than 100 base pairs ("bps") in length and constituency and preferably is about 30 bps or shorter, and can be made be approaches known in the art, including the use of complementary DNA strands or synthetic approaches. The RNAs that are capable of causing interference can be referred to a small interfering RNAs ("siRNA"), and can cause cause post-transcriptional silencing of specific genes in cells, such as mammalian cells (including human cells) and in the body, such as mammalian bodies (including humans). Exemplary siRNAs according to the invention could have up to 29 bps, 25 bps, 22 bps, 21 bps, 20bps, 15 bps, 10 bps or any number thereabout or therebetween.

Example 1. Oligonucleotide Libraries For Detecting All Alternatively Spliced Transcripts: Human, Rat, and Mouse Libraries

Oligonucleotides either 60 bases or 65 bases in length are synthesized, each of which is modified at the 5' terminus with a C6-amino addition to enable subsequent covalent attachment to aldehyde glass surface. Each oligonucleotide is complementary to a portion of a RNA transcript that is common among the existing RNA splice variants transcribed from a particular transcription unit or gene. Each oligonucleotide is also specific or unique to the RNA transcripts and/or the RNA splice variants to which it is complementary, i.e., its binding to RNA transcripts or splice variants from other transcription units or genes is negligible. In other words, a single oligo is selected for each gene or transcription unit. Each oligo is derived from a sequence segment that is common to a practically maximal number of splice variants known or predicted for each gene or transcription unit. Additionally, the oligos are designed and synthesized in such a manner that high accuracy of

sequence quality is maintained by avoiding sequencing errors, that the secondary structure is avoided to promote effective hybridization, and that the melting temperature is normalized to be consistent across the entire oligo collection in the library.

In some representative embodiments, the human oligo library comprises 18,861 human oligonucleotides, each 60 bases in length; the rat oligo library comprises 4,854 rat oligonucleotides, each 65 bases in length; and the mouse oligo library comprises 7,524 oligonucleotides, each 65 bases in length. The sequences of the 18,861 human oligos, the 4,854 rat oligos, and the 7,524 mouse oligos are contained in Sequence Listing Filed Under Sec. 801(a), applicant name Compugen Ltd. (CRF and copies) comprising 32,337 sequences in Patentin 3.0 filed along with this application on CD-R media, file size 5.78 MB, file name Shoshan, with Attorney Docket No. 36688-0006, the entirety of which is hereby incorporated by reference, which contains the sequences set forth in the Sequence Listing (CRF and copies) comprising 32,337 sequences in Patentin 3.0 filed along with this application on CD-R media, file size 5.78 MB, file name Shoshan, with Attorney Docket No. 36688-0004, the entirety of which is hereby incorporated by reference, which was filed in U.S. Serial No. 60/287,724, filed May 2, 2001. Also provided herewith are the 32,337 sequences used in the sequence listing, which are in a MS Word format on a CD-R disk labeled Raw Sequences, Sequence Listing Filed Under Sec. 801(a), PC/MS-DOS PATENTIN 3.0 for the 5.78 MB sequence listing, with Attorney Docket No. 36688-0006, the entirety of which is hereby incorporated by reference. Each oligo in these libraries represents a RNA transcript and/or all existing or predicted RNA splice variants that are transcribed from a transcription unit or gene encoding a certain functional protein. The name, GenBank accession number, and other annotative information of the corresponding gene or transcription unit for the oligo

sequences in many cases are specified. The number of sequences above are not absolute, and therefore other libraries can be made according to applicants' teachings that contain different numbers of sequences.

The oligos of the library may be arranged on plates, e.g., in 384 or 96 well format. They are made spot or print-ready, and thus can conveniently be applied in microarray analysis.

Oligonucleotide libraries for other species, such as bovine, porcine, or arabadopsis may be similarly constructed according to the present invention; those libraries may be readily used in microarray analysis and, in other RNA or cDNA detection systems known in the art.

Example 2. Mini Oligonucleotide Library for Detecting All RNA Transcripts or Distinct Splice Variants

Human oligo mini-libraries are constructed for human cancer/apoptosis genes, obesity/diabetes genes, and toxicology genes. Oligonucleotides 60 bases in length are designed or selected for genes implicated in cancer/apoptosis, obesity/diabetes, and toxicology, respectively, using approaches outlined in U.S. Application Ser. No. 09/133,987, also taking into account the phenomenon of alternative splicing. These oligos are then synthesized taking into account the similar considerations as discussed in Example 1 supra.

The oligonucleotides in the human oligo mini-library capable of detecting distinct splice variants of human cancer/apoptosis genes and/or their transcripts comprise 201 sequences. The oligonucleotides in the human oligo mini-library capable of detecting all existing or predicted splice variants from human cancer/apoptosis genes comprise 280 sequences. The oligonucleotides

in the human oligo mini-library capable of detecting distinct splice variants of genes implicated in toxicology comprise 234 sequences; and the oligonucleotides in the human oligo mini-library capable of detecting all existing or predicted splice variants from human toxicology genes comprise 96 sequences. The oligonucleotides in the human oligo mini-library capable of detecting distinct splice variants of human obesity/diabetes genes comprise 195 sequences; and the oligonucleotides in the human oligo mini-library capable of detecting all existing or predicted splice variants from human obesity/diabetes genes comprise 92 sequences. Again, the number of sequences above are not absolute, and therefore other libraries can be made according to applicants' teachings that contain different numbers of sequences.

The sequences of all the oligos in the six mini-libraries are enclosed in the Sequence Listing filed along with this application. These oligos may be arranged on plates, e.g., in 384 or 96 well format. They are made print-ready, and thus can conveniently applied in microarray analysis. Disease specific or differentially expressed genes can therefore be detected or monitored using these oligo libraries and microarrays for human toxicology, diabetes/obesity, and cancer/apoptosis conditions, respectively.

It is to be understood that while the invention has been described in detail by way of example and illustration for the purpose of clarity of teaching, the foregoing description is not intended to limit the scope of the invention. Other aspects, advantages, and modifications that are apparent to one of skill in the art in light of the teachings of this invention are within the scope of the following claims.

What is claimed is

1. An oligonucleotide library for detecting messenger RNAs that populate a transcriptome, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

2. The oligonucleotide library of claim 1, wherein said transcriptome is a human transcriptome.

3. The oligonucleotide library of claim 1, wherein said transcriptome is a rat transcriptome.

4. The oligonucleotide library of claim 1, wherein said transcriptome is a mouse transcriptome.

5. An oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a tissue origin, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

6. An oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a pathological tissue origin, wherein the sub

transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the pathological tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

7. The oligonucleotide library of claim 6, wherein said pathological tissue origin is cancer tissue.

8. An oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a developmental stage, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the developmental stage, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

9. An oligonucleotide library for detecting messenger RNAs that populate a transcriptome of patients suffering from a disorder, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome of patients suffering from the disorder, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

10. The oligonucleotide library of claim 9, wherein said disorder is cancer.

11. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 1 or a subset thereof.

12. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 5 or a subset thereof.

13. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 6 or a subset thereof.

14. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 8 or a subset thereof.

15. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 9 or a subset thereof.

16. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 1 or a subset thereof.

17. The method of claim 16, wherein said hybridization signals are obtained from a nucleotide chip.

18. The method of claim 16, wherein said hybridization signals are obtained from an electrophoresis gel.

19. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 5 or a subset thereof.

20. The method of claim 19, wherein said hybridization signals are obtained from a nucleotide chip.

21. The method of claim 19, wherein said hybridization signals are obtained from an electrophoresis gel.

22. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 6 or a subset thereof.

23. The method of claim 22, wherein said hybridization signals are obtained from a nucleotide chip.

24. The method of claim 22, wherein said hybridization signals are obtained from an electrophoresis gel.

25. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 8 or a subset thereof.

26. The method of claim 25, wherein said hybridization signals are obtained from a nucleotide chip.

27. The method of claim 25, wherein said hybridization signals are obtained from an electrophoresis gel.

28. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 9 or a subset thereof.

29. An oligonucleotide library for detecting messenger RNAs that populate a transcriptome, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

30. The oligonucleotide library of claim 29, wherein said transcriptome is a human transcriptome.

31. The oligonucleotide library of claim 29, wherein said transcriptome is rat transcriptome.

32. The oligonucleotide library of claim 29, wherein said transcriptome is a mouse transcriptome.

33. An oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a tissue origin, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

34. An oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a pathological tissue origin, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of

transcription units that populate a sub genome of the pathological tissue origin, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

35. The oligonucleotide library of claim 34, wherein said pathological tissue origin is cancer tissue.

36. An oligonucleotide library for detecting messenger RNAs that populate a sub transcriptome of a developmental stage, wherein the sub transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a sub genome of the developmental stage, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the sub genome, wherein at least one transcription unit of the sub genome encodes one or more messenger RNA splice variants.

37. An oligonucleotide library for detecting messenger RNAs that populate a transcriptome of patients suffering from a disorder, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome of patients suffering from the disorder, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to one or a subset of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants.

38. The oligonucleotide library of claim 37, wherein said disorder is cancer.

39. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 29 or a subset thereof.

40. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 33 or a subset thereof.

41. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 34 or a subset thereof.

42. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 36 or a subset thereof.

43. A DNA microarray having spotted thereon a plurality of oligonucleotide sequences, wherein said plurality is provided by the oligonucleotide library of claim 37 or a subset thereof.

44. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 29 or a subset thereof.

45. The method of claim 44, wherein said hybridization signals are obtained from a nucleotide chip.

46. The method of claim 44, wherein said hybridization signals are obtained from an electrophoresis gel.

47. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 33 or a subset thereof.

48. The method of claim 47, wherein said hybridization signals are obtained from a nucleotide chip.

49. The method of claim 47, wherein said hybridization signals are obtained from an electrophoresis gel.

50. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 34 or a subset thereof.

51. The method of claim 50, wherein said hybridization signals are obtained from a nucleotide chip.

52. The method of claim 50, wherein said hybridization signals are obtained from an electrophoresis gel.

53. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 36 or a subset thereof.

54. The method of claim 53, wherein said hybridization signals are obtained from a nucleotide chip.

55. The method of claim 53, wherein said hybridization signals are obtained from an electrophoresis gel.

56. A method for expression profiling a cell or tissue sample that contains two or more RNAs of various abundances, comprising measuring hybridization signals of said sample to a plurality of oligonucleotide sequences, thereby determining the levels of said two or more RNAs in said sample and, where a RNA is transcribed from a transcription unit that has a set of splice variants, determining the total level of the set of splice variants, wherein said plurality is provided by the oligonucleotide library of claim 37 or a subset thereof.

57. The method of claim 56, wherein said hybridization signals are obtained from a nucleotide chip.

58. The method of claim 56, wherein said hybridization signals are obtained from an electrophoresis gel.

59. A double stranded RNA molecule based on an oligonucleotide selected from an oligonucleotide library for detecting messenger RNAs that populate a transcriptome, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants, wherein

the double-stranded RNA molecule comprises no more than 30 basepairs, wherein the double-stranded RNA molecule can interfere with translation of an mRNA.

60. An antisense molecule based on an oligonucleotide selected from an oligonucleotide library for detecting messenger RNAs that populate a transcriptome, wherein the transcriptome comprises messenger RNAs transcribed from a multiplicity of transcription units that populate a genome, wherein the library comprises a plurality of oligonucleotides, wherein each oligonucleotide in the plurality is capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, wherein at least one transcription unit of the genome encodes one or more messenger RNA splice variants, wherein

the antisense molecule comprises no more than 30 bases, wherein the double-stranded RNA molecule can interfere with translation of an mRNA.

Fig. 1

Transcription Unit	5'--------A----------	----------B----------	--------C----------3'
Transcript 1	5'--------A----------		--------C--------AAA3'
Transcript2	5'	----------B----------	--------C--------AAA3'

Fig. 2

Transcription Unit	5'----------A----------|	|----------B----------|	|----------C----------3'
Transcript 1	5'----------A----------|	|----------B----------|	|----------C----------AAA3'
Transcript2	5'----------A----------|	|----------B----------|	|----------C----------AAA3'

Fig. 3

Transcription Unit	5'----A----	-----B------	------C------	---------D---------	-------E----------3'
Transcript1	5'----A----		------C------	---------D----------AAA3'	
Transcript2	5'----A----		------C------		-------E------AAA3'
Transcript3		-----B------	------C------	---------D----------AAA3'	
Transcript4		-----B------	------C------		-------E------AAA3'

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/010449 A3

(51) International Patent Classification⁷: C12Q 1/68

(21) International Application Number: PCT/IB01/01903

(22) International Filing Date: 20 July 2001 (20.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/221,607 28 July 2000 (28.07.2000) US
60/287,724 2 May 2001 (02.05.2001) US

(71) Applicant *(for all designated States except US)*: **COMPUGEN INC.** [US/US]; 9700 Great Seneca Highway, Rockville, MD 20850 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SHOSHAN, Avi** [IL/US]; 37 West 75th Street, New York, NY 10023 (US). **WASSERMAN, Alon** [IL/US]; Apartment 3C, 328 West 19th Street, New York, NY 10011 (US). **MINTZ, Eli** [IL/US]; 16 Nancy Street, Kendall Park, NJ 08824 (US). **MINTZ, Liat** [IL/US]; 16 Nancy Street, Kendall Park, NJ 08824 (US). **FAIGLER, Simchon** [IL/US]; 2605 Wildberry Court, Edison, NJ 08817 (US).

(74) Agent: **ISACSON, John, P.**; Heller, Ehrman, White & Mcauliffe, Suite 300, 1666 K Street, N.W., Washington, DC 20006 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report:
5 December 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE VARIANTS THAT POPULATE A TRANSCRIPTOME

WO 02/010449 A3

Transcription Unit	5'---------A----------	----------B----------	----------C----------3'
Transcript 1	5'---------A----------		----------C----------AAA3'
Transcript2	5'	----------B----------	----------C----------AAA3'

(57) Abstract: The present invention provides oligonucleotide libraries capable of detecting RNA transcripts and RNA splice variants which populate a transcriptome and which are transcribed from genes or transcription units that populate the corresponding genome. The present invention also provides oligonucleotide arrays generated from the oligonucleotide libraries and methods of using the oligonucleotide libraries in various oligonucleotide detection systems and expression profiling studies. Antisense molecules and double-stranded interfering RNAs, which are types of oligonucleotides, based on the oligonucleotides disclosed herein also are provided.

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 01/01903

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, BIOSIS, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 53319 A (KINZLER KENNETH W ;VOGELSTEIN BERT (US); UNIV JOHNS HOPKINS (US)) 26 November 1998 (1998-11-26) the whole document - in particular, e.g., abstract and p.16, ll. 19-25 the whole document	1-58
X	LOCKHART D J ET AL: "EXPRESSION MONITORING BY HYBRIDIZATION TO HIGH-DENSITY OLIGONUCLEOTIDE ARRAYS" BIO/TECHNOLOGY, NATURE PUBLISHING CO. NEW YORK, US, vol. 14, no. 13, 1 December 1996 (1996-12-01), pages 1675-1680, XP002022521 ISSN: 0733-222X the whole document	1-58

-/--

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 August 2002	11/09/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Madlener, M

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 01/01903

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	SCHENA M ET AL: "QUANTITATIVE MONITORING OF GENE EXPRESSION PATTERNS WITH A COMPLEMENTARY DNA MICROARRAY" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 270, no. 5235, 20 October 1995 (1995-10-20), pages 467-470, XP000644675 ISSN: 0036-8075 the whole document	1-58
X	CROFT LARRY ET AL: "ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome." NATURE GENETICS, vol. 24, no. 4, April 2000 (2000-04), pages 340-341, XP002208599 ISSN: 1061-4036 the whole document - in particular p. 341, middle and right columns	1-58
X	ZAMORE PHILLIP D ET AL: "RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals." CELL, vol. 101, no. 1, 31 March 2000 (2000-03-31), pages 25-33, XP002208683 ISSN: 0092-8674 abstract page 27, right-hand column, paragraph 2 -page 28, left-hand column, paragraph 2 page 30, right-hand column, paragraph 2 -page 31, right-hand column, paragraph 1 figure 7	59
X	VARGA LIA V ET AL: "Antisense strategies: Functions and applications in immunology." IMMUNOLOGY LETTERS, vol. 69, no. 2, 3 August 1999 (1999-08-03), pages 217-224, XP002208684 ISSN: 0165-2478 page 217, right-hand column	60

-/---

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 01/01903

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	BLACK DOUGLAS L: "Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology." CELL, vol. 103, no. 3, 27 October 2000 (2000-10-27), pages 367-370, XP002208600 ISSN: 0092-8674 the whole document - in particular p. 369, right column, last paragraph	1-58
P,X	ELBASHIR SAYDA M ET AL: "RNA interference is mediated by 21- and 22-nucleotide RNAs" GENES AND DEVELOPMENT, COLD SPRING HARBOR LABORATORY PRESS, NEW YORK, US, vol. 15, no. 2, 15 January 2001 (2001-01-15), pages 188-200, XP002204651 ISSN: 0890-9369 the whole document	59
A	MIRONOV ANDREY A ET AL: "Frequent alternative splicing of human genes" GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, US, vol. 9, no. 12, December 1999 (1999-12), pages 1288-1293, XP002205853 ISSN: 1088-9051 cited in the application the whole document	1-58
A	GAUTHERET DANIEL ET AL: "Alternate polyadenylation in human mRNAs: A large-scale analysis by EST clustering." GENOME RESEARCH, vol. 8, no. 5, May 1998 (1998-05), pages 524-530, XP002208602 ISSN: 1088-9051 the whole document	1-58
E	WO 01 57252 A (AEOMICA INC) 9 August 2001 (2001-08-09) page 8, line 9 - line 14 page 8, line 24 -page 9, line 8 page 35, line 24 -page 62, line 26 page 75, line 5 - line 25 page 76, line 5 -page 86, line 2 figures 5,11-13 examples 1-4 claims 1-14,16-30,35-38	1-58

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

tional Application No
PCT/IB 01/01903

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9853319 A	26-11-1998	AU 7499198 A	11-12-1998
		WO 9853319 A2	26-11-1998
WO 0157252 A	09-08-2001	AU 3087801 A	14-08-2001
		AU 3087901 A	14-08-2001
		AU 3088001 A	14-08-2001
		AU 3088101 A	14-08-2001
		AU 3088201 A	14-08-2001
		AU 3088301 A	14-08-2001
		AU 3275701 A	14-08-2001
		AU 3275801 A	20-11-2001
		AU 3275901 A	14-08-2001
		AU 3276001 A	14-08-2001
		AU 3311401 A	14-08-2001
		AU 3658901 A	14-08-2001
		AU 6343201 A	11-12-2001
		GB 2361238 A ,B	17-10-2001
		WO 0157270 A2	09-08-2001
		WO 0157271 A2	09-08-2001
		WO 0157272 A2	09-08-2001
		WO 0157273 A2	09-08-2001
		WO 0186003 A2	15-11-2001
		WO 0157274 A2	09-08-2001
		WO 0157275 A2	09-08-2001
		WO 0157276 A2	09-08-2001
		WO 0157277 A2	09-08-2001
		WO 0157278 A2	09-08-2001
		WO 0157251 A2	09-08-2001
		WO 0157252 A2	09-08-2001
		WO 0192524 A2	06-12-2001
		US 2002081590 A1	27-06-2002
		US 2002048763 A1	25-04-2002
		US 2002048800 A1	25-04-2002
		GB 2360284 A ,B	19-09-2001
		EP 1158049 A1	28-11-2001
		US 2002102252 A1	01-08-2002
		WO 0226818 A2	04-04-2002

Form PCT/ISA/210 (patent family annex) (July 1992)

フロントページの続き

(51)Int.Cl.[7]	FI	テーマコード（参考）
	Ｃ１２Ｎ 15/00	Ｆ

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72)発明者 ワッサーマン, アロン
アメリカ合衆国, ニューヨーク州 １００１１, ニューヨーク, ウエスト ナインティース ストリート ３２８, アパートメント ３シー
(72)発明者 ミンツ, エリ
アメリカ合衆国, ニュージャージー州 ０８８２４, ケンダール パーク, ナンシー ストリート １６
(72)発明者 ミンツ, リアット
アメリカ合衆国, ニュージャージー州 ０８８２４, ケンダール パーク, ナンシー ストリート １６
(72)発明者 フェイグラー, シムチョン
アメリカ合衆国, ニュージャージー州 ０８８１７, エディソン, ワイルドベリー コート ２６０５

Ｆターム(参考) 4B024 AA12 AA19 CA01 CA12 HA12
4B029 AA07 AA23 BB20 FA15
4B063 QA19 QA20 QQ08 QQ53 QR08 QR42 QR56 QS25 QS34 QS39
QX02