

①⑨ RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
PARIS

①① N° de publication :
(à n'utiliser que pour les
commandes de reproduction)

2 856 410

②① N° d'enregistrement national : **03 07508**

⑤① Int Cl⁷ : C 12 Q 1/68, G 01 N 33/53, G 06 F 19/00 // G 06 F 159:00

①②

DEMANDE DE BREVET D'INVENTION

A1

②② Date de dépôt : 20.06.03.

③① Priorité :

④③ Date de mise à la disposition du public de la demande : 24.12.04 Bulletin 04/52.

⑤⑥ Liste des documents cités dans le rapport de recherche préliminaire : *Se reporter à la fin du présent fascicule*

⑥① Références à d'autres documents nationaux apparentés :

⑦① Demandeur(s) : *INTEGRAGEN Société anonyme* — FR.

⑦② Inventeur(s) : LINDENBAUM PIERRE et GES-NOUIN PHILIPPE.

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : CABINET BECKER ET ASSOCIES.

⑤④ METHODES DE PREPARATION DE PUCES GENETIQUES ET UTILISATIONS.

⑤⑦ La présente demande concerne des méthodes et compositions pour la production de biopuces, ainsi que l'utilisation de ces biopuces dans des domaines variés, de la génomique fonctionnelle au diagnostic par exemple, notamment en recherche ou dans le domaine médical. Elle concerne également des outils et procédés de sélection de polynucléotides, permettant la production de Biopuces améliorées. Les biopuces de l'invention comprennent notamment des clones BAC.

FR 2 856 410 - A1



Méthodes de Préparation de Puces Génétiques et Utilisations

Introduction

5 La présente demande concerne des méthodes et compositions pour la production ou la fabrication de biopuces, ainsi que l'utilisation de ces biopuces dans des domaines variés, de la génomique fonctionnelle au diagnostic par exemple, notamment en recherche ou dans le domaine médical. Elle concerne également des outils et procédés de sélection de polynucléotides, permettant la production de
10 Biopuces améliorées.

Avec le développement de la génomique et des techniques de miniaturisation, de nouvelles stratégies d'identification de gènes, d'analyse ou de diagnostic ont vu le jour. Ces stratégies reposent notamment sur des réactions d'hybridation entre un
15 échantillon test, dont on souhaite analyser le contenu, et une banque de polynucléotides immobilisés sur un support. De telles approches sont ou seront utilisées en diagnostic, en recherche, en pharmacogénomique, etc., pour analyser une population d'acides nucléiques ou un échantillon biologique.

20 Différents types de polynucléotides peuvent ainsi être ainsi immobilisés sur des supports, selon le type d'application recherché. Ainsi, des oligonucléotides, des fragments PCR, des BAC, des gènes ou fragments de gènes particuliers, des ARNs, etc., ont été immobilisés sur des supports. En outre, il peut s'agir de polynucléotides de séquence pré-définie ou connue, ou de séquence aléatoire, ou
25 d'une combinaison. Différentes stratégies de production de tels supports ont également été rapportées, qui se classent en deux catégories principales : la synthèse in situ ou le couplage. Dans les méthodes de synthèse in situ, les polynucléotides sont synthétisés par élongation directe sur le support, par exemple par photolithographie (voir par exemple le brevet US5,510,270). Cette
30 technique est essentiellement limitée à la synthèse de biopuces à oligonucléotides.

- Dans les techniques de couplage, les polynucléotides sont immobilisés par dépôt sur le support, après synthèse et/ou sélection. Différentes techniques ont été décrites, comprenant un couplage direct sur le support, ou une interaction avec un oligonucléotide complémentaire, ou le couplage par l'intermédiaire d'un bras espaceur, etc. La préparation par couplage permet d'élargir le champ des biopuces à tout type de polynucléotide, comme indiqué ci-avant. Par ailleurs, différents types de supports ont également été proposés, tels que des supports en (ou à base de) verre, plastic, polymère, métal, matériaux biologiques, silices, nylon, etc.
- 10 Dans le cadre de la présente demande, on désignera par le terme générique « biopuce » tout support sur lequel sont immobilisés des polynucléotides. Les polynucléotides sont généralement immobilisés sur la surface du support ou sur une aire de celle-ci, de manière à être accessible pour une réaction d'hybridation. L'immobilisation peut être covalente ou non, ordonnée ou non, dense ou non, etc.
- 15 De préférence, elle est covalente et ordonnée.

Les applications des biopuces sont variées, allant de la recherche au diagnostic. Ainsi, des puces sont utilisées pour rechercher des différences d'expression de gènes, des altérations génétiques, pour comparer des échantillons, pour localiser des marqueurs, en séquençage, pour comparer des nombres de copies de gènes, etc.

20

Pour ces différentes applications, un échantillon à analyser est mis en contact avec la biopuce et un signal d'hybridation est détecté. Généralement, les acides nucléiques de l'échantillon sont marqués au préalable, pour faciliter la détection. Selon l'amplitude du signal détecté, la position du signal, etc., il est possible de déterminer la présence, dans l'échantillon, d'un acide nucléique particulier, d'une forme altérée d'un gène ou messenger, un niveau d'expression, etc. De multiples approches ont été proposées pour le marquage des échantillons, la mise en contact

25

des échantillons et de la puce, la lecture des signaux d'hybridation, l'analyse des résultats, etc.

Toutefois, il existe actuellement un besoin pour des méthodes améliorées de
5 production de biopuces, dont la composition et la constitution soient mieux contrôlées, et permettant des applications et des lectures plus fiables.

Résumé de l'Invention

- 10 La présente demande décrit maintenant des méthodes et outils pour la production de biopuces particulières. Elle décrit également l'utilisation de ces biopuces dans des domaines variés, de la génomique fonctionnelle au diagnostic par exemple, notamment en recherche ou dans le domaine médical. Elle concerne également des outils et procédés de sélection de polynucléotides, permettant la production de
15 biopuces améliorées. Les biopuces selon l'invention sont notamment caractérisées par le fait qu'elles comportent une pluralité de polynucléotides formant un ensemble (ou une collection) représentatif du génome d'un organisme considéré (e.g., de sa séquence, de son organisation, etc.). Le génome considéré est préférentiellement un génome humain. De telles biopuces sont particulièrement
20 adaptées pour localiser, positionner ou cartographier tout acide nucléique d'intérêt, ou pour des applications diagnostiques ou en pharmacogénomique, pour évaluer la présence ou les niveaux d'expression (absolus ou relatifs) de gènes, chez des sujets ou dans des cellules en culture, par exemple.
- 25 Un premier objet de l'invention réside plus particulièrement dans un procédé de production d'une biopuce comprenant un support sur lequel est immobilisé un ensemble de polynucléotides, caractérisé en ce qu'il comprend :
- (ii) la sélection, parmi une pluralité de clones BAC comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un génome, de
30 préférence d'un génome humain, d'un ensemble de clones BAC comprenant un

insert unique dans ledit génome, les clones BAC de l'ensemble sélectionné comprenant des inserts nucléiques répartis de manière substantiellement uniforme sur le génome et, de préférence, espacés les uns des autres par un intervalle de l'ordre de 1Mb environ, et

- 5 (iii) le dépôt, sur un support, de clones ainsi sélectionnés, ou des inserts nucléiques qu'ils contiennent, ou d'une partie d'entre eux, dans des conditions permettant aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire.
- 10 Un objet plus particulier de l'invention réside notamment dans un procédé de production d'une biopuce comprenant un support sur lequel est immobilisé un ensemble de polynucléotides marqueurs d'un génome, notamment du génome humain, caractérisé en ce qu'il comprend :
- (i) l'obtention de clones BAC comprenant un insert nucléique correspondant à ou
- 15 spécifique d'une portion d'un génome, de préférence humain,
- (ii) la sélection, parmi les clones BAC ainsi obtenus, d'un ensemble de clones BAC comprenant un insert unique dans ledit génome, les clones BAC de l'ensemble sélectionné comprenant des inserts nucléiques répartis de manière substantiellement uniforme sur le génome et espacés les uns des autres par un
- 20 intervalle de l'ordre de 1Mb environ, et
- (iii) le dépôt, sur un support, de clones ainsi sélectionnés, ou des inserts nucléiques qu'ils contiennent, ou d'une partie d'entre eux, dans des conditions permettant aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire.

25

Un autre aspect de l'invention concerne une biopuce, caractérisée en ce qu'elle comprend un support sur lequel est immobilisé un ensemble de clones BAC comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un génome, de préférence humain, chaque clone comprenant un insert unique dans

- 30 ledit génome, les clones BAC de l'ensemble comprenant des inserts nucléiques

répartis de manière substantiellement uniforme sur ledit génome et avantageusement espacés les uns des autres par un intervalle régulier de l'ordre de 1Mb environ. Préférentiellement, les clones BACs sont agencés de manière déterminée sur le support. Dans une variante préférée, le support est un lame de verre.

Un autre aspect de l'invention réside dans l'utilisation d'une biopuce telle que définie ci-avant pour l'analyse génétique, notamment pour l'identification de gènes, pour la cartographie génétique, pour le diagnostic, en pharmacogénomique, etc.

Un autre aspect de l'invention concerne un procédé d'identification ou de localisation d'un acide nucléique sur le génome humain, comprenant la mise en contact d'un acide nucléique test avec une biopuce telle que définie ci-avant dans des conditions permettant une hybridation entre séquences complémentaires, la détection d'un signal d'hybridation, et l'identification de la position de l'acide nucléique sur le génome par identification des clones engagés dans les hybridations.

Un autre aspect de l'invention concerne un procédé de détection de la présence ou de l'abondance (e.g., niveaux d'expression absolus ou relatifs) d'un gène dans un échantillon biologique, comprenant la mise en contact de l'échantillon avec une biopuce telle que définie ci-avant dans des conditions permettant une hybridation entre séquences complémentaires, et la détection d'un signal d'hybridation, ledit signal étant indicatif de la présence ou de l'abondance d'un gène dans l'échantillon. Avantageusement, l'échantillon biologique est d'origine humaine et comprend des acides nucléiques (biopsie, culture cellulaire, lysat cellulaire, fluide biologique, tissu, organe, etc.). L'échantillon peut être traité préalablement, afin de rendre accessible (ou de favoriser l'accès) des acides nucléiques pour une réaction d'hybridation.

L'invention fournit ainsi de nouveaux procédés et outils pour la production de biopuces améliorées, comportant une pluralité de clones BAC formant un ensemble (ou une collection) représentatif du génome humain (de sa séquence, de son organisation, etc.). La présente demande décrit des méthodes de sélection, préparation, dépôt (e.g., « spotting ») sur support et hybridation des supports ainsi obtenus avec des échantillons biologiques, permettant de cartographier, localiser et identifier des gènes d'intérêt.

10

Description détaillée de l'Invention

Le terme clone BAC ou « Bacterial Artificial Chromosome » désigne un clone bactérien comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un génome, de préférence d'un génome humain. Le terme clone BAC désigne soit le clone bactérien comportant l'insert nucléique, soit un vecteur extrait (ou isolé à partir) du clone bactérien, et comprenant l'insert nucléique, soit l'insert nucléique lui-même, ou une partie de celui-ci, soit les acides nucléiques totaux de la bactérie. Les BAC sont des vecteurs adaptés au clonage de fragments d'ADN de longueur importante, et sont utilisés pour la constitution de banques. D'après les données publiées, on sait qu'il existe actuellement plusieurs milliers de clones BAC différents disponibles dans des collections, comportant chacun un insert nucléique distinct représentatif d'un segment de génome humain.

Pour la mise en œuvre de la présente invention, les clones BAC peuvent être obtenus, collectés ou rassemblés à partir de sources multiples, comme des bases de données, des données de séquençage, des collections d'échantillons, etc. Le séquençage du génome humain a permis de révéler et de mettre à disposition de multiples marqueurs ou clones, correspondant à des régions du génome humain. Ces marqueurs ou clones recouvrent maintenant la totalité de la séquence du génome humain, mais ne sont pas ordonnés ou classés de façon suffisamment

- complète et précise pour être utilisables efficacement. Ainsi, ces multiples clones comportent des inserts chevauchant, redondants, non-spécifiques, mal positionnés, parfois non caractérisés, etc. En raison de leur pluralité, complexité et diversité, ces clones n'ont pu être exploités de manière satisfaisante jusqu'à
- 5 présent pour la production de produits diagnostic ou analytiques validés. La présente demande propose maintenant de produire des biopuces à partir de clones BAC. Elle propose avantageusement de nouveaux outils et procédés permettant de sélectionner des ensembles de clones validés.
- 10 Dans les procédés de l'invention, des clones BAC sont obtenus afin de fournir une collection de clones susceptibles de recouvrir l'intégralité de la séquence du génome considéré, de préférence d'un génome humain. Des sources publiques de clones BAC sont notamment *BACPAC resources* (chori.org), *Research Genetics* (resgen.com) ou le centre Sanger (*sanger.ac.uk*, *CloneRequest*). Ces collections
- 15 sont accessibles au public, par exemple sur internet, et sont bien connues de l'homme du métier. Dans ces collections, les clones se présentent généralement sous forme de clones de bactéries comprenant un vecteur BAC contenant l'insert nucléique. Les clones BAC obtenus à partir de telles sources sont donc préférentiellement sous forme de culture bactérienne, qui peuvent être stockées,
- 20 analysées, répliquées, etc. Le vecteur BAC portant l'insert peut également être isolé et analysé, amplifié, sous-cloné, etc. Les clones ainsi obtenus sont typiquement stockés dans des boîtes de culture, ou dans tout dispositif approprié (tube, fiole, flasque, etc.). Ils peuvent être lyophilisés, congelés, etc.
- 25 De préférence, on rassemble suffisamment de clones BAC pour obtenir une collection susceptible de recouvrir l'intégralité de la séquence du génome considéré, de préférence d'un génome humain. On préfère les clones BAC pour lesquels certaines informations structurales et/ou fonctionnelles sont disponibles (e.g., données d'hybridation in situ (« FISH », séquence partielle, etc.). La
- 30 sélection des clones BAC constitue ensuite une caractéristique très importante de

l'invention. Elle permet en effet de fournir, à partir de multiples clones, un ensemble de clones validé, cohérent et utilisable pour la production de biopuces adaptées à des expériences de cartographie ou d'identification fiables.

- 5 La sélection est avantageusement réalisée par élimination(s), le cas échéant selon plusieurs cycles successifs au cours desquels la sélection est de plus en plus poussée et la qualité des clones augmentée.

La sélection des clones BAC d'intérêt peut être réalisée avantageusement au
10 moyen d'un programme informatique, ou en utilisant, dans certaines étapes, des règles de décision informatiques. L'invention est tout particulièrement adaptée à la production de puces pour l'analyse du génome humain.

Dans un mode de mise en œuvre préféré, la sélection des clones comprend :

- 15 (a) l'élimination des clones non uniques ;
(b) l'élimination des clones partageant un même STS ;
(c) l'élimination des STS qui sont marqués en deux endroits différents au moins sur le génome considéré, notamment humain ;
(d) le classement des clones en fonction de leur position sur le génome,
20 définissant ainsi des clones voisins ; et
(e) l'élimination supplémentaire de clones, par l'application d'un procédé itératif, pour obtenir notamment une répartition substantiellement uniforme sur le génome considéré, notamment humain des clones finalement sélectionnés.

25

L'ordre des étapes (a) à (e) peut être interchangé. En outre, certaines de ces étapes peuvent être réalisées simultanément. On notera cependant que l'étape (e) ne peut pas être réalisée avant l'étape (d).

Avantageusement, les étapes de sélection (a) à (e) ou une partie d'entre-elles sont répétées (un ou plusieurs cycles de sélection peuvent ainsi être réalisés), jusqu'à obtention d'un ensemble de clones tel que défini ci-avant. Typiquement, deux clones sont d'abord analysés, puis des clones supplémentaires sont progressivement introduits dans l'analyse, jusqu'à balayer la majeure partie, de préférence l'intégralité du génome.

L'étape (a) comprend donc une élimination des clones non uniques, c'est-à-dire des clones qui sont marqués à deux endroits différents au moins sur le génome considéré, de préférence humain. Le terme "marqués" indique que l'insert nucléique que ces clones contiennent est présent en plusieurs positions dans un génome ou spécifique (ou complémentaire) de plusieurs régions dans le génome. Dans la mesure où de tels clones peuvent hybrider avec des régions distinctes du génome considéré, ils ne peuvent pas permettre une localisation efficace d'un fragment d'acide nucléique et sont donc avantageusement éliminés.

La mise en évidence de la non unicité d'un clone peut être réalisée de différentes manières, comme par exemple par compilation ou analyse des informations connues pour un clone (emplacement, marqueur, séquence, etc.), par analyse informatique de la séquence de l'insert nucléique qu'il contient (si sa séquence est composée de motifs répétés ou consensus, il n'aura a priori pas un caractère unique), par des expériences biologiques (e.g., hybridation in situ, etc.).

L'étape (b) comprend l'élimination des clones partageant un même STS "Sequence Tagged Site". Le STS est le site sur le génome "étiqueté" par une séquence, c'est-à-dire, dans le cas présent, le site cible d'un clone BAC sur le génome humain. Lorsque plusieurs clones partagent un même STS, ces clones sont redondants et rendent la biopuce plus complexe. La présente demande propose donc une sélection de clones comprenant l'élimination des clones partageant un même STS augmentant la spécificité de la biopuce. L'identification

du STS d'un clone peut être réalisée par des techniques connues en soi, comme par exemple à partir des informations disponibles pour chacun des clones, de la séquence de l'insert nucléaire, de données d'hybridation in situ, etc.

- 5 Les STS des clones sont ensuite comparés et lorsque deux clones partagent un même STS, l'un est éliminé.

L'étape (c) comprend l'élimination des STS qui sont marqués en deux endroits différents au moins sur le génome considéré.

10

L'étape (d) comprend le classement des clones BAC en fonction de leur position sur le génome. Ceci peut être réalisé par analyse des marques connues, ou par toute autre méthode connue en soi par l'homme du métier. Cette étape conduit à la définition de clones « voisins », c'est-à-dire de clones immédiatement
15 adjacents sur le génome considéré.

Pour réaliser ce classement, on repère la position de l'insert nucléaire de chaque clone BAC sur le génome considéré, en représentant par exemple ce dernier sous la forme d'une échelle graduée de 0 à 100. Chaque insert nucléaire peut alors être représenté par un segment du génome, de coordonnées x_i et y_i , avec $y_i > x_i$, sur
20 l'échelle graduée de 0 à 100.

Les clones BAC notés r_i peuvent ainsi être classés sur cette échelle graduée par ordre croissant de leurs coordonnées x_i , ou bien, en variante, par ordre croissant de leurs coordonnées y_i . On obtient ainsi le classement suivant, dans le cas où on compare les coordonnées x_i :

25

$$0 < x_1 < \dots < x_i < \dots < x_n < 100.$$

L'étape (e) est une étape d'élimination de clones BAC, parmi l'ensemble des clones précédemment classés $\{r_1, \dots, r_n\}$, pour obtenir un ensemble E de clones finalement sélectionnés.

Cette étape débute par une première sous-étape (e_1) d'extraction d'un sous-ensemble E' de clones BAC susceptibles d'être éliminés de l'ensemble E de clones finalement sélectionnés, cette sous-étape d'extraction étant réalisée par l'application d'un premier critère de rejet.

- 5 Ce premier critère de rejet est basé sur le calcul d'un écart algébrique, désigné dans la suite "écart" entre clones BAC.

L'écart entre deux clones BAC $r_i(x_i, y_i)$ et $r_j(x_j, y_j)$, avec $j > i$, est défini par la relation suivante :

$$d(r_i, r_j) = x_j - y_i.$$

- 10 On notera que l'écart entre deux clones BAC prend une valeur négative si les deux clones BAC se chevauchent, une valeur nulle si le second clone BAC se trouve dans le prolongement du premier, et une valeur positive si ces deux clones BAC n'ont pas de partie commune.

- 15 Il est important que les clones BAC finalement sélectionnés de l'ensemble E comprennent des inserts nucléiques répartis de manière substantiellement uniforme sur le génome humain, c'est-à-dire qu'il n'y ait pas un écart trop important entre deux clones BAC voisins successifs sélectionnés. Pour ce faire, le premier critère de rejet est défini comme un seuil S dont la valeur correspond à
- 20 l'écart maximal toléré entre deux clones BAC voisins dans l'ensemble E des clones finalement sélectionnés. Par exemple, $S = 1,5$ Mb, ce qui se traduit par une valeur s d'écart sur l'échelle graduée de 0 à 100 du génome humain.

- Pour tout i compris entre 2 et n , pour déterminer si un clone BAC r_i doit appartenir au sous-ensemble E' des clones susceptibles d'être éliminés, on calcule
- 25 l'écart entre le clone r_{i-1} et le clone r_{i+1} . Pour le clone BAC r_1 , on calcule l'écart $d(0, r_2) = x_2$. Enfin, pour le clone r_n , on calcule l'écart $d(r_{n-1}, 100) = 100 - y_{n-1}$. Si l'écart calculé est inférieur à la valeur s , le clone r_i correspondant appartient à E' .

- Lors d'une seconde sous-étape (e_2), un second critère est appliqué aux éléments du sous-ensemble E' des clones susceptibles d'être éliminés, pour déterminer
- 30 l'unique clone de ce sous-ensemble qui sera effectivement éliminé.

Pour ce faire, chaque clone BAC r_i est associé à une liste de propriétés à partir de laquelle est calculée par exemple une fonction coût $f(r_i)$ pour chaque clone BAC. Ceci permet de sélectionner dans le sous-ensemble E' , le clone qui maximise cette fonction coût. Ce clone est alors éliminé. De façon classique, la fonction
5 coût est à valeurs positives et est d'autant plus élevée que le clone correspondant est susceptible d'être éliminé.

En variante, on peut remplacer le calcul d'une fonction coût par un système à base de règles qui permet de comparer les listes de propriétés de deux clones et de
10 prendre une décision d'élimination de l'un des deux clones comparés.

La liste des propriétés attachées à un clone BAC peut par exemple comporter un paramètre de disponibilité, un paramètre définissant la collection d'origine du clone BAC, un paramètre concernant la validation par hybridation in-situ.

15

On peut également imaginer d'autres propriétés telles qu'un paramètre lié à un recouvrement d'un clone BAC par un autre, permettant d'éliminer de préférence les clones BAC se recouvrant, ou un paramètre incitant à éliminer de préférence des clones BAC qui génèrent des écarts entre le clone précédent et le clone
20 suivant supérieurs à un seuil S' , choisi en fonction du seuil S , de sorte que les inserts nucléiques sélectionnés soient espacés les uns des autres par un intervalle à peu près régulier. En choisissant une valeur $S = 1,5$ Mb et $S' = 0,7$ Mb, on obtient un ensemble de clones BAC sélectionnés dont les inserts nucléiques sont espacés les uns des autres par intervalles de l'ordre de 1 Mb environ, sur le
25 génome humain. De même que pour S , la valeur S' se traduit par une valeur s' d'écart sur l'échelle graduée de 0 à 100 du génome humain.

La succession des deux sous-étapes (e_1) et (e_2) décrites précédemment est réitérée jusqu'à ce que l'on ne puisse plus éliminer un clone BAC, sans créer un écart
30 supérieur au seuil S sur le génome humain, c'est-à-dire lorsque le sous-ensemble

E' de clones BAC susceptibles d'être éliminés obtenu à l'étape (e₁) est l'ensemble vide.

La première sous-étape (e₁) d'extraction du sous-ensemble E' et la seconde sous-étape (e₂) d'élimination d'un clone de ce sous-ensemble E' peuvent également s'articuler de la façon suivante (en prenant à titre d'exemple le cas du calcul d'une fonction coût) :

- étape 1 - initialisation d'un index k à la valeur nulle et d'un index F à la valeur nulle ;
- 10 - étape 2 - pour tout i allant de 1 à n :
 - (e1) on détermine si r_i appartient au sous-ensemble E' ;
 - (e2) si "oui", on teste si $f(r_i)$ est supérieur ou égal à F et si c'est le cas, on affecte à k la valeur i et à F la valeur $f(r_i)$;
- 15 - étape 3 - si k est nul, cela signifie que le sous-ensemble E' est l'ensemble vide et dans ce cas on passe à l'étape 6 ;
- étape 4 - sinon on élimine le clone r_k et on réordonne les clones restants en décrémentant n d'une unité ;
- étape 5 - on passe à l'étape 1 ;
- étape 6 - fin du procédé.

20

Comme indiqué ci-avant, l'ordre des étapes (a) à (e) peut être modifié, notamment l'ordre des étapes (a) à (c). En outre certaines étapes peuvent être réalisées de manière simultanée. La sélection est avantageusement réalisée en utilisant un programme informatique particulier capable d'analyser et de comparer les données de chaque clone BAC. A cet égard, la présente demande propose également de nouveaux outils facilitant la réalisation des étapes (a) à (c) à partir d'un nombre très élevé de clones BAC de départ, notamment des outils informatiques.

Ainsi, la présente invention décrit l'outil CloneTrek, qui est une suite logicielle ayant pour but de sélectionner des sous-ensembles d'objets (clones BACs) en fonction de leurs propriétés (validation par hybridation in situ, collection d'origine, disponibilité) et de leur localisation sur un axe 1-D (le génome humain).

5 Par ailleurs, CloneTrek offre la possibilité de représenter graphiquement les cartes obtenues, de compléter une carte avec des données supplémentaires, de comparer des cartes, de générer des fichiers d'entrée pour certains types de robots de repiquage de plaques, de calculer des statistiques sur les BACs d'une collection, etc.

10 L'ensemble des programmes de CloneTrek utilisent et échangent des données formatées en XML, ("eXtended Markup Language") selon des DTD ("Document Type Definition ") propriétaires (XMLMap et XMLPlateHandler). Des programmes d'importation de données à partir des ressources internes et publiques ont été développés afin de traduire ces données dans ces formats.

15

L'algorithme de baliseclones et les données utilisées sont décrits ci-après :

Objet : sélectionner des BACs (clones bactériens comportant un insert humain donné) afin de les utiliser comme balise de position sur une puce à ADN.

Données de départ :

- 20 - STS_aliases : liste des STS référencés, données associées dont FISH, fournies par le NCBI ;
- Clones_BAC : liste des BACs référencés, données associées (dont FISH) collectées (NCBI, etc.).

Position : fichiers fournis par le NCBI et le Golden Path listant par position
25 sur le génome les informations qui y sont attachées, en particulier les clones et les STS. Pour les dernières versions, nous utilisons essentiellement les données du NCBI.

Collection des fournisseurs : les clones BACs doivent être commandés chez un fournisseur qui les a disponibles sous forme de collections.

30

Même si les biopuces préférées de l'invention permettent de couvrir l'ensemble d'un génome humain, il est également possible de produire des biopuces couvrant seulement une portion d'un génome (par exemple un ou plusieurs chromosomes).

5 Le procédé décrit précédemment comportant les étapes (a) à (e) comprend, dans ce cas, les étapes suivantes :

- d'abord sont éliminés les clones et les STS qui sont marqués à deux endroits différents sur la position de génome considéré, conformément aux étapes (a) à (c) précédemment décrites ; .

10 - ensuite, conformément aux étapes (d) et (e), pour chaque chromosome :

- les clones sont ordonnées sur le chromosome, en fonction de leur position ou de la position des STS auxquels ils se réfèrent ; et
 - l'élimination supplémentaire de clones, par l'application d'un procédé itératif, pour obtenir notamment une répartition substantiellement
- 15 uniforme sur le génome humain des clones finalement sélectionnés.

Après sélection des clones BAC, ceux ci (ou les inserts nucléiques qu'ils contiennent, ou une partie de ces clones ou inserts), sont déposés sur un support,

20 dans des conditions permettant aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire. Différentes techniques peuvent être mises en œuvre pour le dépôt des clones ou inserts, tels que le couplage direct sur le support, ou une interaction avec un oligonucléotide complémentaire, ou le couplage par l'intermédiaire d'un bras espaceur, d'agents

25 bi-fonctionnels, etc. D'une manière générale, pour permettre aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire, il est préférable que les clones ou inserts soient liés au support par une de leurs extrémités. Différentes méthodes sont possibles, comme l'utilisation d'une molécule espaceur (WO99/51773), d'un support revêtu de groupes fonctionnels

30 tels le polyéthylèneimine (GB2,197,720) ou l'avidine (WO97/18226), ou de

bras arborescents (EP 647 719, WO99/61662, WO99/10362). D'autres techniques d'immobilisation sont décrites par exemple dans les brevets US4,925,785 et EP 373 203. Par ailleurs, différents types de supports peuvent être utilisés, comme des supports plans ou non, rigides ou non, à base de différents matériaux tels le verre, plastic, polymère, métal, des matériaux biologiques, silices, nylon, etc. On peut citer à titre illustratif des membranes de nylon, des lames de verre, des plaques de silice, etc.

Dans un mode préféré de mise en œuvre, le support est une lame de verre. Le dépôt sur lame de verre peut être réalisé par dépôt direct des échantillons sur la lame, ou après pré-traitement de la lame pour favoriser l'interaction avec les acides nucléiques. Des lames utilisables sont par exemple des lames de verre recouvertes d'amino-silane (par exemple lames GAPS II, Corning).

De manière particulièrement préférée, le dépôt est réalisée de manière ordonnée, c'est-à-dire selon un arrangement et/ou une densité (pré-)déterminés.

Avantageusement, chaque clone ou insert est positionné sur une zone du support identifiable (une cellule). Le dépôt peut être fait avantageusement au moyen d'un robot. La densité des clones ou inserts sur le support peut varier, en fonction du nombre de clones ou inserts distincts et de la surface du support. Généralement, moins de 1000 clones ou inserts distincts sont déposés sur une surface de 1 cm². Bien entendu, chaque clone est généralement présent en plusieurs copies, de manière à augmenter la sensibilité de la biopuce.

Préalablement au dépôt sur le support, les clones de l'ensemble sélectionné peuvent être repiqués, amplifiés, caractérisés, stockés, etc. De cette façon, il est possible et aisé de reproduire des biopuces de l'invention. A cet égard, une variante de mise en œuvre de l'invention réside dans un procédé de production

d'une biopuce comprenant un support sur lequel est immobilisé un ensemble de polynucléotides marqueurs du génome humain, caractérisé en ce qu'il comprend :

- (ii) la sélection, parmi une pluralité de clones BAC comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un génome humain, d'un
5 ensemble de clones BAC comprenant un insert unique dans le génome humain, les clones BAC de l'ensemble sélectionné comprenant des inserts nucléiques répartis de manière substantiellement uniforme sur le génome humain et espacés les uns des autres par un intervalle de l'ordre de 1Mb environ,
- (iii) l'amplification des clones BAC de l'ensemble sélectionné et/ou des inserts
10 nucléiques qu'ils contiennent, et
- (iv) le dépôt de manière ordonnée, sur un support, de clones ainsi sélectionnés, ou des inserts nucléiques qu'ils contiennent, ou d'une partie d'entre eux, dans des conditions permettant aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire.

15

Un objet plus spécifique de l'invention réside dans un procédé de production d'une biopuce comprenant un support sur lequel est immobilisé un ensemble de polynucléotides, caractérisé en ce qu'il comprend :

- (i) l'obtention de clones BAC comprenant un insert nucléique correspondant à ou
20 spécifique d'une portion d'un génome humain,
- (ii) la sélection, parmi les clones BAC ainsi obtenus, d'un ensemble de clones BAC comprenant un insert unique dans le génome humain, les clones BAC de l'ensemble sélectionné comprenant des inserts nucléiques répartis de manière substantiellement uniforme sur le génome humain et espacés les uns des autres
25 par un intervalle de l'ordre de 1Mb environ,
- (iii) l'amplification des clones BAC de l'ensemble sélectionné et/ou des inserts nucléiques qu'ils contiennent, et
- (iv) le dépôt de manière ordonnée, sur un support, de clones ainsi sélectionnés, ou des inserts nucléiques qu'ils contiennent, ou d'une partie d'entre eux, dans des

conditions permettant aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire.

Un autre aspect de l'invention concerne une biopuce, caractérisée en ce qu'elle
5 comprend un support sur lequel est immobilisé un ensemble de clones BAC
comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un
génome humain, chaque clone comprenant un insert unique dans le génome
humain, les clones BAC de l'ensemble comprenant des inserts nucléiques répartis
10 des autres par un intervalle régulier de l'ordre de 1Mb environ.
Préférentiellement, les clones BACs sont agencés de manière déterminée sur le
support. Dans une variante préférée, le support est une lame de verre. Il est entendu
que la biopuce de l'invention peut comprendre en outre d'autres polynucléotides,
qui peuvent être ou non des clones BAC. Ainsi, la biopuce peut comprendre des
15 polynucléotides contrôles, d'origine, nature et taille variées.

Un autre objet de l'invention réside dans l'utilisation d'une biopuce telle que
définie ci-avant pour l'identification de gènes, pour la cartographie génétique,
pour le diagnostic, en pharmacogénomique, etc. Les biopuces de l'invention sont
20 utilisables en recherche, pour l'identification de gènes, le clonage, l'analyse de
différences d'expression entre des cellules ou tissus, etc. Elles sont également
utilisables dans des procédés de diagnostic ou de pharmacogénomique, pour
détecter des altérations génomiques ou génétiques, pour détecter des différences
d'expression de gènes, etc.

25

Des applications particulières sont notamment :

- la détection du changement de nombre de copie d'un chromosome ou d'une portion d'un chromosome associé à une pathologie,

- des études toxicologiques : identification des composés toxiques susceptibles d'induire des modifications sur le nombre de copie d'une portion du génome (toxicogénomique),
- la détection de translocation
- 5 - la caractérisation de panel d'hybrides d'irradiations pour différentes espèces,
- l'étude non-quantitative de l'expression de gènes, en particulier ceux non présents dans les puces dites d'expression,
- l'étude à l'échelle de l'ensemble du génome des interactions ADN-protéines,
- toute application basée sur la cartographie génomique et/ou le changement de
- 10 ratio,
- etc.

A cet égard, un objet particulier de l'invention réside dans un procédé d'identification ou de localisation d'un acide nucléique sur le génome humain,

15 comprenant la mise en contact d'un acide nucléique avec une biopuce telle que définie ci-avant dans des conditions permettant une hybridation entre séquences complémentaires, la détection d'un signal d'hybridation, et l'identification de la position de l'acide nucléique sur le génome par identification des clones engagés dans les hybridations.

20

L'acide nucléique testé peut être de nature, forme et origine variées. Il peut s'agir d'un ARN ou, plus préférentiellement d'un ADN. Le procédé est utilisable pour analyser un acide nucléique isolé, ou pour tester une composition ou un échantillon complexe comprenant une pluralité d'acides nucléiques non

25 caractérisés individuellement. La présence d'une hybridation peut être mise en évidence de différentes manières. Généralement, l'acide nucléique test est préalablement marqué, et la formation d'un hybride est détectée par mise en évidence du marquage sur la biopuce. Le marquage peut être radioactif, fluorescent, enzymatique, luminescent, chimique, etc. D'autres techniques de

30 détection utilisent des sondes de révélation, des détecteurs électriques, etc.

Un autre aspect de l'invention réside dans un procédé d'identification de gènes associés à un trait de caractère donné, comprenant (i) l'identification de fragments d'acides nucléiques identiques entre deux échantillons provenant de sujets
5 présentant un trait de caractère commun, et (ii) l'hybridation des fragments ainsi identifiés sur une biopuce telle que définie ci-avant. La détection d'un signal d'hybridation permet de localiser le(s) fragment(s) sur le génome humain, et ainsi d'identifier un ou des gènes présents dans celui-ci, associés au dit trait de caractère. Le trait de caractère peut être une pathologie (e.g., maladie génétique
10 monogénique ou complexe), un phénotype donné (e.g., réponse à un traitement), etc. L'étape (i) peut être réalisée par différentes techniques, telles que celles décrites dans la demande WO00/53802 ou dans le brevet US5,376,526.

D'autres aspects et avantages de la présente invention apparaîtront à la lecture des
15 exemples qui suivent, qui doivent être considérés comme illustratifs et non limitatifs.

Légende des Figures

20 Figure 1 : Représentation synthétique de la couverture du génome humain (Build 33) par les 2263 clones positionnés.

Figure 2 : Courbe de représentation pour la détermination de la valeur seuil.

25 Figure 3 : Analyse de prédiction d'IBD

Figure 4 : Analyse génétique (A) Analyse sur l'ensemble du génome. Les signaux correspondant aux autosomes donnent comme attendu un signal comparable pour les 2 individus. Le signal Cy5 mâle apparaît clairement réduit pour le

chromosome X en adéquation avec la différence d'un facteur 2 entre mâle et femelle. (B) Délétion du gène de la dystrophine sur le chromosome X.

Exemples

5

Exemple 1 : Sélection des clones

Pour la sélection des clones, les données de départ utilisées étaient les suivantes :

- collections de clones (*Clone Registry @ NCBI*)
- 10 - plaques contenant des clones (*fournisseur*)
- cartographie de clones sur le génome (*Golden Path* , *NCBI*)
- caractéristiques des clones (STS, FISH, ACN, phase.. – *Golden Path* , *NCBI*).

- 15 L'ensemble de ces données a été inséré dans une base de données relationnelle locale.

Un fichier au format XML a été généré pour servir d'entrée au programme Clonotrek, qui décrit de façon synthétique l'ensemble des propriétés des clones

- 20 BACs prises en compte pour la sélection.

La sélection initiale a été effectuée à partir des clones FISH (c'est-à-dire des clones positionnés par hybridation in-situ) disponibles. Nous avons ainsi un total de 76 plaques :

25

- VGC: 47 plaques (3294 clones, 2708 positionnés)
- CSMC: 15 plaques (860 clones, 735 positionnés)
- CCAP: 14 plaques (719 clones, 662 positionnés)

sur le draft du Golden Path (GP) du 12/12/2001.

30

Certaines de ces plaques étant contaminées, elles ont été éliminées. Au total, 41 plaques ont été utilisées dans cette étude, représentant 2460 clones.

5 Après balisage par Clonetrek (-en paramètres distance maximale 900000 et distance minimale 200000) l'algorithme a retenu 2041 clones avec un espacement moyen de 1.8 Mb. Nous avons comblé les espaces avec des clones supplémentaires 'non-FISH', soit commandés individuellement auprès de la société Research Genetics, soit sélectionnés auprès de banques.

10 Nous avons ainsi effectué une sélection à partir d'un ensemble d'environ 292000 clones (dont 90000 positionnés sur le GP du 28/06/2002) plus une librairie dite ONCOBAC (6 plaques soit 579 clones dont 108 positionnés). Nous avons répété l'étape précédente en incluant toute la librairie ONCOBAC ainsi que les clones RP11 placés sur le GP.

15

Après balisage, nous avons obtenu 2264 clones (avec un espacement moyen (« gap ») de 1.2Mb), auxquels nous avons ajouté les oncobacs non positionnés mais utilisés sur la puce. Après re-arrangement de ces plaques (« cherry-picking ») l'identité de ces BACs a été vérifiée par séquençage terminal (439 non
20 séquencés / 2779 clones):

Chaque paire de séquences a été alignée sur la séquence du draft du génome humain afin de confirmer ou positionner le clone correspondant. Nous avons utilisé Blast dans sa version à jour à la date du calcul. Ainsi, sur le Build 30 du
25 28/06/2002 nous pouvions confirmer 1787 clones, sur le build 33 d'avril 2003, 2263 clones étaient positionnés (Figure 1). Les gaps les plus longs sont les centromères des chromosomes 1, 2, 3,4, 5, 6, 7, 10, 11, 12, 16, 17, 18, 19, 20 et 23. Les chromosomes 13, 14, 15, 21 et 22 sont acrocentriques ce qui explique l'absence de clones à leur début.

30

Exemple 2. Fabrication d'une puce

2a. Préparation de la matrice BAC

- 5 Après identification des BACs, chaque clone a été isolé et une mini préparation de quelques μg d'ADN obtenue par des méthodes classiques décrites dans la littérature (ex utilisation des kits développés par Qiagen).

- 10 Un aliquot d'une centaine de ng d'ADN ainsi extrait a ensuite été amplifié. De nombreuses méthodes d'amplification telles que la Rolling Circle amplification (développé par Amersham) ou encore la DOP-PCR (Degenerate Oligonucleotide Primed-PCR) peuvent être employées utilisant des enzymes telles que la templi Phi 29 ou la taq polymerase. Ces ADNs amplifiés sont ensuite purifiés, par exemple par une précipitation à l'éthanol ou QIAGEN.

- 15 Les produits finaux sont complémentés avec les composants d'une solution adaptée à l'impression sur des lames. Ces solutions peuvent être 3xSSC ou 50% DMSO.

20 2b- Impression des lames

- De nombreux types de lames existent dans le commerce. Dans cet exemple, nous avons opté pour des lames recouvertes d'une couche d'amino-silane, distribuées par la société Corning (lames dénommées GAPSII).

- 25 Tout comme les lames, il existe de nombreuses machines de spotting. Dans cette étude, nous avons effectué le spotting sur un Microgrid II fabriqué par la société Biorobotics. Les ADNs, resuspendus en solution de DMSO, ont été spottés à l'aide d'aiguilles creuses à réservoir (microspot pins 2500, distribué par Biorobotics) d'un diamètre de 100 μm .
- 30

Les conditions de spotting pour l'ensemble de notre collection de BAC qui représentait 2600 BACs était les suivantes :

- Température 20°C, Humidité 50%
- 5 • Chaque BAC est spotté en quadruplicat sur 2 zones de la lame (1 duplicat par zone)
- Le nombre de prespot est de 20 avec un écart de 400 μm entre chaque prespot.
- L'écart entre chaque spot est de 275 μm .
- 10 • Rechargement de la matrice tous les 140 dépôts (prespot ou spot)

Après spotting des lames, l'ADN déposé sur les lames est fixé par traitement aux UV. Typiquement cela est réalisé dans un cross linker, en exposant les lames à une énergie Ultra Violet de 70 mJ.

15

Exemple 3. Hybridation

Chaque lame subit ensuite une étape dite de pré-fixation, les lames sont traitées par une des nombreuses méthodes existantes qui puisse fixer de l'ADN de façon non spécifique.

20

Elle peut comprendre un blocage chimique des groupements amines ou bien une incubation avec de la sérum-albumine bovine et du sodium docécylsulphate. Dans cette étude, nous avons utilisé cette dernière méthode.

25

Pour chaque lame, on mélange les ADN marqués en Cy3 et en Cy5 dans une solution tamponnée contenant de la formamide et d'autres composants pour réduire l'hybridation non-spécifique (tRNA, ADN de sperme de saumon, Cot1 subscript 0, etc.). Après incubation à 70°C, les sondes sont laissées au minimum 1h à 37°C. La sonde ainsi préparée est ensuite mise en contact sur la lame

30

contenant les BACs spottés puis la lame est recouverte d'un cover slip. Ce montage est alors transféré dans une chambre d'hybridation de type Corning, où l'on dépose quelques μL d' H_2O dans chacun des deux puits présents aux extrémités, permettant d'y maintenir un certain taux d'humidité. Après fermeture
5 hermétique de la chambre, celle ci est immergée dans un bain marie à 42°C .

Les temps d'hybridation varient de 16 heures à 3 jours selon le type de sondes hybridées. Il existe des systèmes type « machine à hybrider » qu'il est aussi possible d'utiliser dans ce contexte.

10

Après l'incubation, les lames hybridées sont lavées dans des solutions de lavage classiques afin d'éliminer les sondes non hybridées et diverses hybridations a-spécifiques.

15 **Exemple 4. Lecture des lames**

Il existe de nombreux lecteurs de fluorescence pour Biopuces. Nous avons opté pour un appareil distribué par Agilent qui présent l'avantage d'avoir un carrousel permettant de lire à la suite 48 lames.

20

Dans le cas présent, nous avons travaillé avec les réglages suivants :

- Surface 60 x 21.6mm
- Résolution 10 μm / pixel
- Coefficient photo-multiplicateur (PMT) de 100% pour les 2 canaux.

25

Les images sont capturées par lot dans le carrousel de l'appareil (maximum 40), Chaque image peut-être ré-orientée (flip & rotation) et chaque longueur d'onde stockée dans un fichier séparé par souci de compatibilité avec les logiciels de traitement utilisés en aval.

30

Le traitement de l'image consiste en la transformation d'une donnée brute sous forme d'images en données qualitatives et quantitatives pour chaque spot. A chaque type de puce est associé un fichier d'information qui indique sa topologie et pour chaque point son identité. Cette identité permet ultérieurement de lier les résultats à une base de données et, pour chaque clone, d'obtenir sa localisation sur le génome.

L'analyse d'images se compose d'une première étape de segmentation (identification des blocs et des spots) et d'une seconde étape de quantification qui calcule un ensemble de données numériques pour chaque spot (coordonnées, surface, intensité et bruit de fond pour chaque longueur d'onde, ratios, etc.).

Nous utilisons selon le contexte Geneprix Pro (version 4.1) ou bien Recife, un programme développé en interne pour collecter ces informations.

15

Exemple 5. Exemples de résultats obtenus

5a. Application à l'identification de fragments identiques par descendance (« IBD »)

20

Dans le but de valider nos puces, nous avons effectué des expériences sur des paires d'individus de statut connus.

a. Données de départ:

25 20 fois la paire : famille CEPH (Centre d'Etude du Polymorphisme Humain) 1331 individus 7 et 9

20 fois la paire : famille CEPH 1347 individus 3 et 6

20 fois la paire : famille CEPH 1362 individus 3 et 8

→ full sibs

30

8 fois la paire : famille CEPH 1347 individus 12 et 8

8 fois la paire : famille CEPH 1347 individus 13 et 8

→ Grand-parent – petit-enfant

5 La sélection des paires de germains a été faite selon les critères suivant:

→ optimisation du nombre de clones dont le statut IBD est connu
(par génotypage de marqueurs microsatellites)

→ optimisation du nombre de clones IBD pour lesquels chacun des
trois statut 0, 1, 2 est présent.

10

-> Possibilité de confrontation des IBD observés versus les IBD connues.

b. Analyse des data :

10 images par paires de germains ont été analysées (total =30), ainsi que 4 images

15 par paires de grand-parent – petit-enfant. (total 8)

i – analyse d’images:

20 → **lecture** : Les analyses d’image ont été effectuées en utilisant le
programme Genepix, version 4.1.

→ **Filtre** : sur chaque image ont été appliqués plusieurs filtres afin
d’écarter les spots non-analysables ou non-informatifs.

Sont conservés les points qui remplissent les conditions suivantes :

25 o $Cy5 < 40\,000$ (saturation physique des spots)

o $Cy3 < 40\,000$ (saturation physique des spots)

o $Cy3 - BGCy3 > \text{Moyen } BGCy3$

o Nombre de réplicat par clones ≥ 3

o variance entre les réplicat d’un même clone $< 2 * \text{variance (}$
variance de l’ensemble des clones).

30 (BG = background = bruit de fond)

→ **Normalisation :**

o Normalisation du signal Cy3 par signal Cy3 de tous les spots.

o Normalisation du signal Cy5 par signal Cy5 de tous les spots

→ **Annotation:**

- 5 Positionnement des clones sur le génome (position en paires de bases
ainsi que par position cytogénétique obtenue par FISH)

Un fichier de sortie par paires est produit à l'issue de l'ensemble de la procédure
informatique. Seuls sont conservés dans le fichier les noms et positions des clones
10 ainsi que la médiane du ratio des signaux entre les réplicats:

$$15 \quad \text{Ratio} = \frac{\left[\sum_{i=1}^N (\text{Cy5}_i - \text{BGcy5}_i) \right] / N}{\left[\sum_{i=1}^N (\text{Cy3}_i - \text{BGcy3}_i) \right] / N}$$

N = nombre de pixels par spot.

25

ii Prédiction d'IBD:

a -) Moyenne glissante :

Pour chaque clone une moyenne des ratios est calculée en prenant en compte les
30 clones voisins sur chaque chromosome indépendant:

Le ratio lissé nous permet de déterminer la valeur seuil pour laquelle nous
pouvons distinguer l'IBD du non-IBD.

35 **b -) Détermination de la valeur seuil (Figure 2).**

Pour déterminer la valeur seuil pour laquelle on pourra distinguer le statut IBD du non-IBD on utilise deux facteurs :

- La distribution gaussienne du ratio
- 5 → la probabilité entre deux germains d'être IBD = 0.75

donc le seuil est la valeur pour laquelle on a 75 % de clones d'un côté et 25 % de clones de l'autre (Figure 2).

10 c -) Prédiction IBD :

Chaque valeur du ratio lissé est alors comparée à la valeur seuil:

- Ratio lisse > Seuil : IBD = 1
- Ratio lisse ≤ Seuil : IBD = 0

15 Un filtre est ensuite appliqué sur une recherche de motif :

- clone IBD seul au milieu de la région non-IBD : motif : 0 0 1 0 0
recodage en 0 0 0 0 0
- clone non-IBD seul dans région IBD motif : 1 1 0 1 1
re-codage en 1 1 1 1 1

20

iii Résultat (Figure 3):

L'analyse de prédiction d'IBD a été effectuée sur l'ensemble des paires CEPH disponible. Les résultats obtenus sont présentés sur la Figure 3 et illustrent la fiabilité des puces de l'invention. Ces résultats ont été comparés au statut IBD attendu pour les clones pour lesquels on dispose de l'information IBD: Trois pourcentages ont été calculés:

- IBD observé = IBD attendu (% de succès)
- IBD observé ≠ IBD attendu :
 - 30 - observé = 0 attendu = 1 : faux négatif

- observé = 1 attendu = 0 : faux positif

5b. Application à la détection d'une différence du nombre de copies d'un fragment du génome incluant la délétion d'un gène impliqué dans une maladie.

Avec un échantillon qui comprend une délétion qui couvre 0,2% du génome, la plate-forme d'Integrage a permis la détection d'une portion de la délétion qui ne représente que 1/20.000ème du génome.

10

Un ADN humain femelle issu d'un individu normal a été marqué en Cy3 et un ADN humain male issu d'un individu atteint de myopathie de Duchesne en Cy5. Cette affection est accompagnée d'une délétion de 5 à 10 Mb dans la partie 5' du gène de la dystrophine et les régions adjacentes.

15

Le patient est également atteint de granulomatose chronique (CGD), de rétinite pigmentaire et de retard mental (Collection Coriell GM07947). Une quantité d'ADN comparable à ce qu'il est possible d'obtenir par biopsie a été amplifiée de façon générique, et hybridée à une puce telle que décrite dans l'exemple 2. Les résultats obtenus sont présentés sur la figure 4.

20

Ils montrent qu'une région de plus de 5 Mb a été mise en évidence par des analyses cytogénétiques dans les bandes Xp21.3 à Xp21.1. Le clone représenté par un carré est situé dans la délétion. Le clone représenté par un cercle est proche de la limite cytogénétique, son signal relatif est nettement entre les valeurs du cluster du chromosome X observé lors des hybridations mâle/femelle et la zone de délétion. De ce fait, il peut être conclu qu'il contient le « breakpoint » pour la délétion.

25

Au vu de ces observations, on peut affirmer qu'une puce selon l'invention comprenant un contig de BACs recouvrants dans cette région permettrait une détermination plus précise des points de cassure de la délétion et une connaissance plus en avant des gènes voisins de la dystrophine jouant un rôle

5 dans l'apparition de traits phénotypiques complexes.

REVENDICATIONS

1. Procédé de production d'une biopuce comprenant un support sur lequel est immobilisé un ensemble de polynucléotides, caractérisé en ce qu'il comprend :
 - 5 (i) l'obtention de clones BAC comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un génome humain,
 - (ii) la sélection, parmi les clones BAC ainsi obtenus, d'un ensemble de clones BAC comprenant un insert nucléique unique dans le génome humain, les clones BAC de l'ensemble sélectionné comprenant des inserts nucléiques
 - 10 répartis de manière substantiellement uniforme sur le génome humain ; et
 - (iii) le dépôt, sur un support, de clones ainsi sélectionnés, ou des inserts nucléiques qu'ils contiennent, ou d'une partie d'entre eux, dans des conditions permettant aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire.
- 15 2. Procédé selon la revendication 1, caractérisé en ce que l'étape (i) comprend l'obtention d'une collection de clones BAC comprenant des inserts nucléiques susceptibles de recouvrir l'intégralité de la séquence du génome humain.
- 20 3. Procédé selon la revendication 1 ou 2, caractérisé en ce que les clones BAC de l'ensemble sélectionné comprennent des inserts nucléiques espacés les uns des autres par un intervalle de l'ordre de 1 Mb environ.
4. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce que
- 25 l'étape (ii) de sélection des clones comprend :
 - a) l'élimination des clones non uniques ;
 - b) l'élimination des clones partageant un même STS ;
 - c) l'élimination des STS qui sont marqués en deux endroits différents au moins sur le génome humain ;

d) le classement des clones en fonction de leur position sur le génome, définissant ainsi des clones voisins ; et

e) l'élimination supplémentaire de clones, par l'application d'un procédé itératif, pour obtenir notamment une répartition substantiellement uniforme sur le
5 génome humain des clones finalement sélectionnés.

5. Procédé selon la revendication 4, caractérisé en ce que les étapes de sélection a) à e), ou une partie d'entre elles, sont répétées au moins une fois jusqu'à obtention d'un ensemble de clones BAC comprenant un insert unique dans le
10 génome humain, les clones BAC de l'ensemble sélectionné comprenant en outre des inserts nucléiques répartis de manière substantiellement uniforme sur le génome humain, et recouvrant l'ensemble du génome humain.

6. Procédé selon la revendication 4 ou 5, caractérisé en ce que l'étape a) est
15 réalisée par compilation ou analyse des informations connues pour un clone (emplacement, marqueur, séquence, etc...), par analyse informatique de la séquence de l'insert nucléique qu'ils contiennent et/ou par des expériences biologiques.

20 7. Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que l'étape e) comporte une première sous-étape e₁) d'extraction à l'aide d'un premier critère d'un sous-ensemble de clones BAC susceptibles d'être éliminés de l'ensemble de clones finalement sélectionnés, et une seconde sous-étape e₂), lors de laquelle un second critère est appliqué aux éléments du sous-ensemble des
25 clones susceptibles d'être éliminés, pour déterminer l'unique clone de ce sous-ensemble qui sera effectivement éliminé.

8. Procédé selon la revendication 7, caractérisé en ce que la première sous-étape e₁) d'extraction du sous-ensemble de clones BAC susceptibles d'être éliminés est
30 réalisée par l'application d'un premier critère de rejet basé sur le calcul d'un écart

entre clones BAC voisins, notamment un critère définissant une distance maximale autorisée entre deux clones voisins de l'ensemble des clones finalement sélectionnés.

- 5 9. Procédé selon la revendication 7 ou 8, caractérisé en ce que le second critère appliqué lors de la seconde sous-étape e_2) comprend l'exigence d'une distance minimale entre deux clones voisins de l'ensemble des clones finalement sélectionnés.
- 10 10. Procédé selon la revendication 8 ou 9, caractérisé en ce que le premier critère tel que défini dans la revendication 8 est basé sur une distance maximale de 1,5 Mb et le second critère tel que défini dans la revendication 9 comprend l'exigence d'une distance minimale de 0,7 Mb.
- 15 11. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que l'étape (ii) de sélection des clones BAC est réalisée au moyen d'un programme informatique.
12. Procédé selon l'une quelconque des revendications précédentes, caractérisé
20 en ce que, dans l'étape (iii), les clones BAC ou les inserts nucléiques qu'ils contiennent, ou une partie de ces clones ou inserts, sont déposés sur un support par couplage direct sur le support, ou par une interaction avec un oligonucléotide complémentaire, ou par l'intermédiaire d'un bras espaceur.
- 25 13. Procédé selon la revendication 12, caractérisé en ce que le support est plan et/ou rigide.
14. Procédé selon la revendication 12 ou 13, caractérisé en ce que le support est réalisé à base de matériaux choisis parmi le verre, plastic, polymère, métal,
30 matériaux biologiques, silices et nylon.

15. Procédé selon la revendication 14, caractérisé en ce que le support est une lame de verre.

16. Procédé selon l'une quelconque des revendication précédentes, caractérisé en ce que, lors de l'étape (iii), le dépôt est réalisé selon un arrangement et/ou une densité (pré-)déterminés.

17. Procédé selon l'une quelconque des revendication précédentes, caractérisé en ce que, préalablement à l'étape (iii), les clones de l'ensemble sélectionné sont repiqués, amplifiés, caractérisés et/ou stockés.

18. Procédé de production d'une biopuce comprenant un support sur lequel a été immobilisé un ensemble de polynucléotides, caractérisé en ce qu'il comprend :

- (i) l'obtention de clones BAC comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un génome humain ;
- (ii) la sélection, parmi les clones BAC ainsi obtenus, d'un ensemble de clones BAC comprenant un insert unique dans le génome humain, des clones BAC de l'ensemble sélectionné comprenant des inserts nucléiques répartis de manière substantiellement uniformes sur le génome humain ;
- (iii) l'amplification des clones BAC de l'ensemble sélectionné et/ou des inserts nucléiques qu'ils contiennent ; et
- (iv) le dépôt, sur un support, de clones ainsi sélectionnés ou des inserts nucléiques qu'ils contiennent, ou d'une partie d'entre eux, dans des conditions permettant aux clones ou inserts déposés d'hybrider avec un acide nucléique ayant une séquence complémentaire.

19. Biopuce, caractérisée en ce qu'elle comprend un support sur lequel est immobilisé un ensemble de clones BAC comprenant un insert nucléique correspondant à ou spécifique d'une portion d'un génome humain, chaque clone comprenant un insert unique dans le génome humain, les clones BAC de

l'ensemble comprenant des inserts nucléiques répartis de manière substantiellement uniforme sur le génome humain.

20. Utilisation d'une biopuce selon la revendication 19 pour l'identification de
5 gènes, pour la cartographie génétique, pour le diagnostic ou en pharmacogénomique.

21. Procédé d'identification ou de localisation d'un acide nucléique sur le génome humain, comprenant la mise en contact d'un acide nucléique avec une biopuce
10 selon la revendication 19, dans des conditions permettant une hybridation entre séquences complémentaires, la détection d'un signal d'hybridation et, l'identification de la position de l'acide nucléique sur le génome par identification des clones engagés dans les hybridations.

15 22. Procédé d'identification de gènes associés à un trait de caractère donné, comprenant (i) l'identification de fragments d'acides nucléiques identiques entre au moins deux échantillons provenant de sujets présentant un trait de caractère commun, et (ii) l'hybridation des fragments ainsi identifiés sur une biopuce selon la revendication 19 et, (iii) la détection d'un signal d'hybridation, permettant de
20 localiser le(s) fragment(s) sur le génome humain et ainsi d'identifier un ou des gènes présents dans celui-ci, associés au dit trait de caractère.

1/4

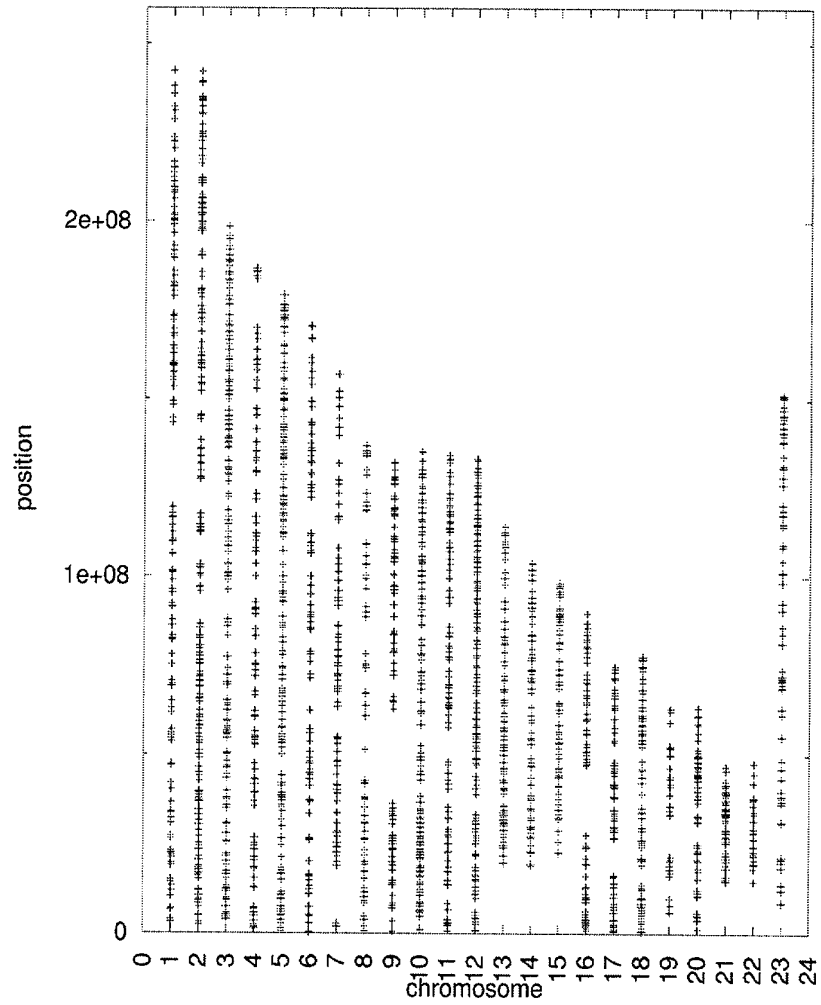
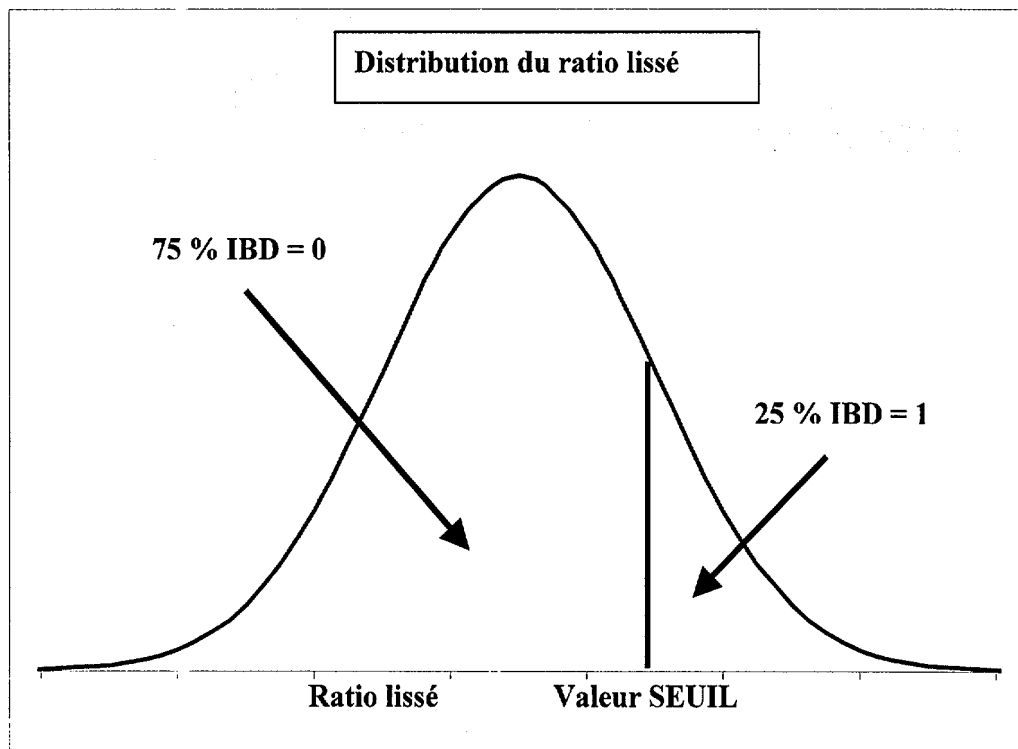


Figure 1

2/4

**Figure 2**

3/4

Fig 3A

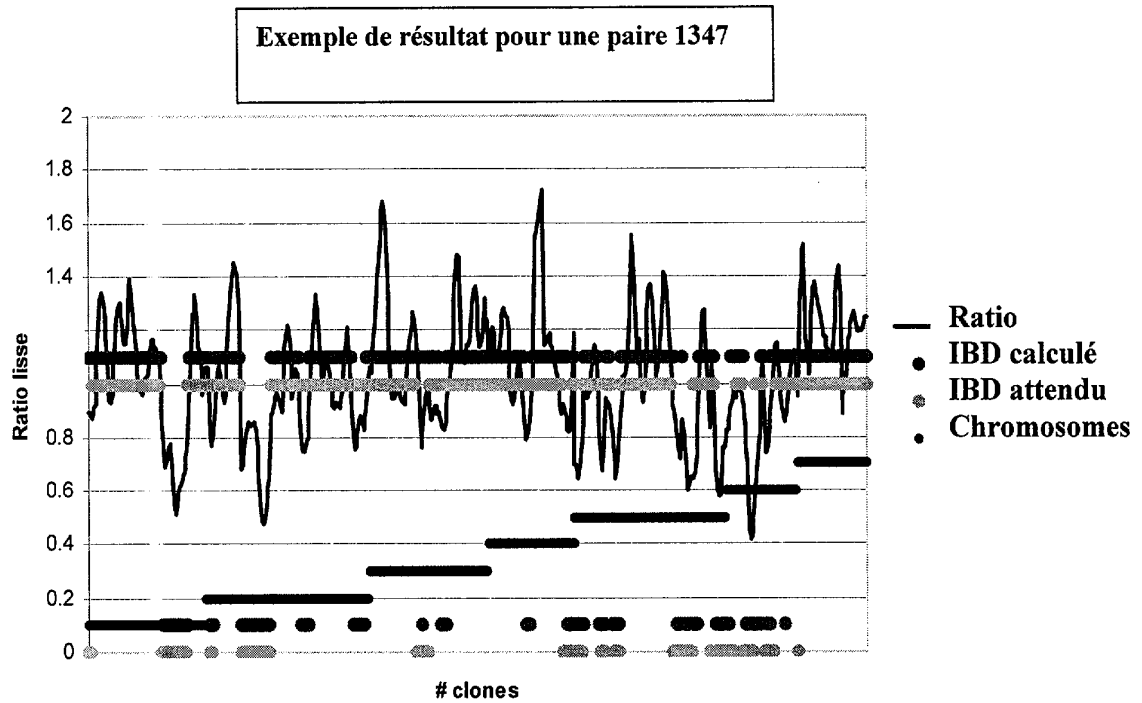


Fig 3b

Suite

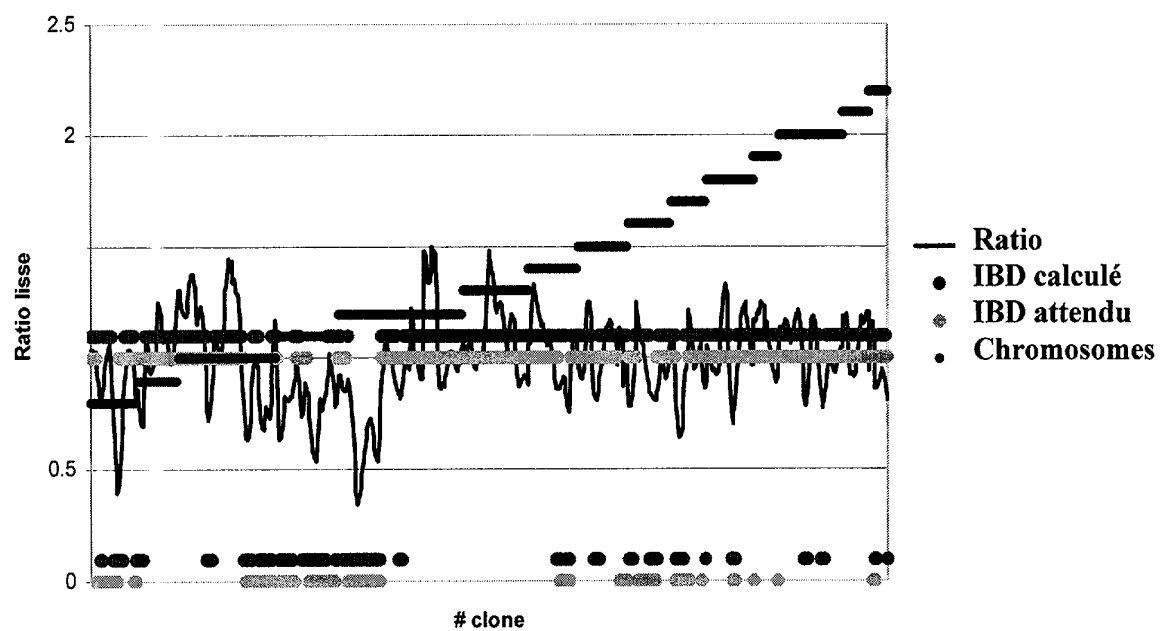


Figure 3

4/4

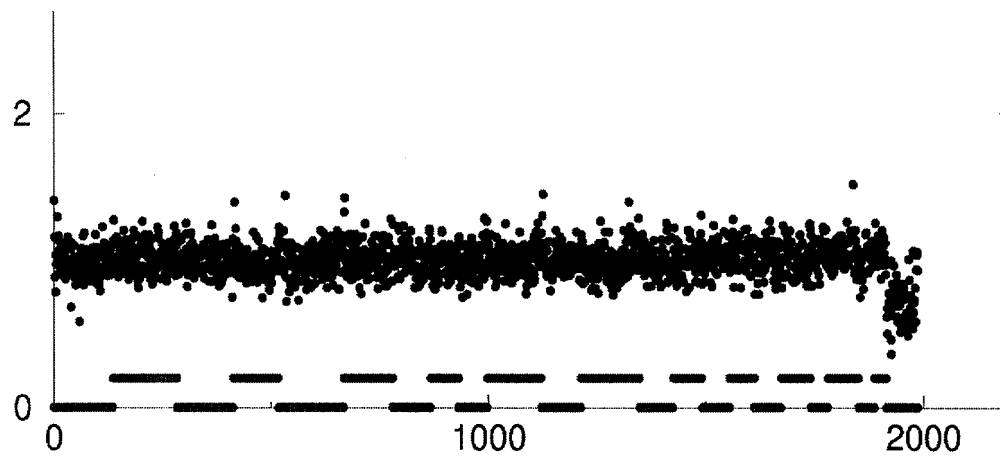


Fig. 4A

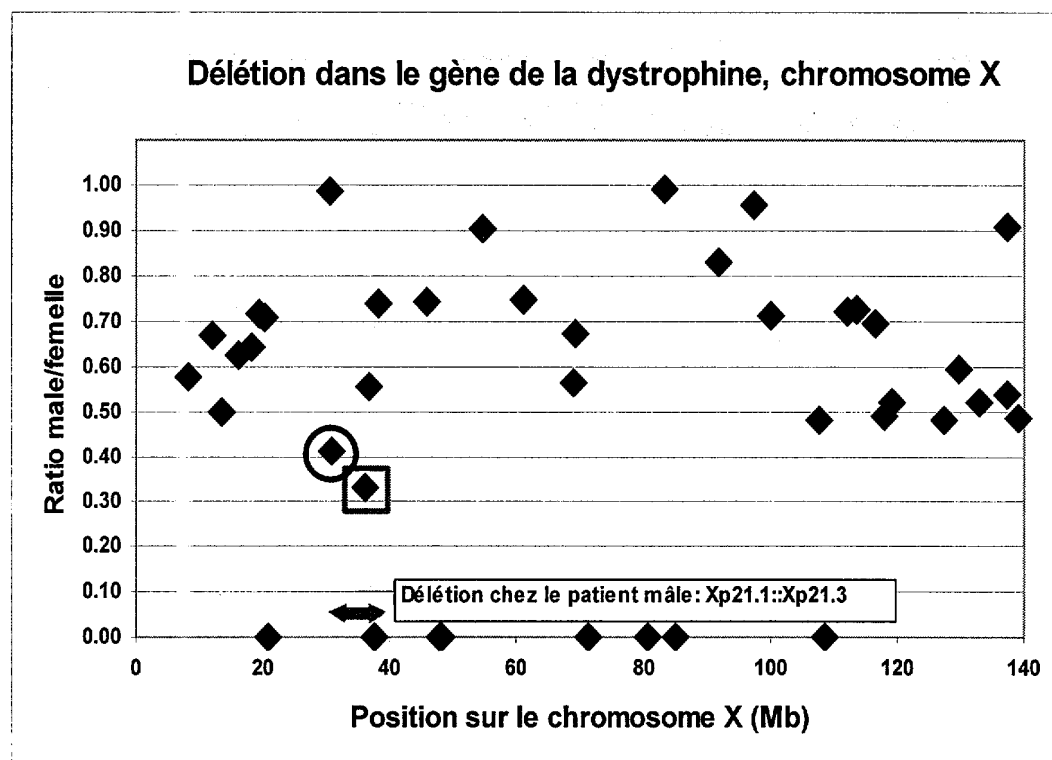


Fig. 4B

Figure 4



RAPPORT DE RECHERCHE PRÉLIMINAIRE

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 636941
FR 0307508

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	SNIJDERS A M ET AL: "Assembly of microarrays for genome-wide measurement of DNA copy number" NATURE GENETICS 2001 UNITED STATES, vol. 29, no. 3, 2001, pages 263-264, XP002273931 ISSN: 1061-4036	1-3, 12-21	C12Q1/68 G01N33/53 G06F19/00
Y	* le document en entier * ---	4-11,22	
X	KORENBERG JULIE R ET AL: "Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine" GENOME RESEARCH, vol. 9, no. 10, octobre 1999 (1999-10), pages 994-1001, XP001180262 ISSN: 1088-9051 * abrégé * * page 997 * ---	1	
X	COWELL J K ET AL: "High-Resolution Analysis of Genetic Events in Cancer Cells Using Bacterial Artificial Chromosome Arrays and Comparative Genome Hybridization" ADVANCES IN CANCER RESEARCH 2003 UNITED STATES, vol. 90, 2003, pages 91-125, XP009027840 ISSN: 0065-230X * abrégé * * page 96, ligne 11 - page 100, ligne 26 * --- -/--	1-22	DOMAINES TECHNIQUES RECHERCHÉS (Int.CL.7) C12Q
Date d'achèvement de la recherche		Examineur	
19 mars 2004		Luzzatto, E	
CATÉGORIE DES DOCUMENTS CITÉS X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire		T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant	

3

EPO FORM 1503 12.99 (P04C14)



RAPPORT DE RECHERCHE PRÉLIMINAIRE

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 636941
FR 0307508

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	CHEUNG V G ET AL: "A resource of mapped human bacterial artificial chromosome clones." GENOME RESEARCH. UNITED STATES OCT 1999, vol. 9, no. 10, octobre 1999 (1999-10), pages 989-993, XP002273933 ISSN: 1088-9051 * le document en entier *	1	
X	WO 03 027311 A (BRADLEY ALLAN ;BAYLOR COLLEGE MEDICINE (US); CAI WEI-WEN (US); GIB) 3 avril 2003 (2003-04-03) * revendications *	1-3	
Y	----- * revendications *	4-11,22	
D,A	WO 00 53802 A (HAGER JOERG) 14 septembre 2000 (2000-09-14) * revendications *	22	
			DOMAINES TECHNIQUES RECHERCHÉS (Int.CL.7)
Date d'achèvement de la recherche		Examineur	
19 mars 2004		Luzzatto, E	
CATÉGORIE DES DOCUMENTS CITÉS X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intercalaire T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons & : membre de la même famille, document correspondant			

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE**RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 0307508 FA 636941**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.

Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du **19-03-2004**

Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 03027311 A	03-04-2003	US 2003092007 A1 WO 03027311 A2	15-05-2003 03-04-2003
WO 0053802 A	14-09-2000	DE 19911130 A1 CA 2366107 A1 WO 0053802 A1 EP 1179088 A1 JP 2002537855 A	21-09-2000 14-09-2000 14-09-2000 13-02-2002 12-11-2002