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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 April 2005 (21.04.2005)

PCT

(10) International Publication Number
WO 2005/035792 A2

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

(21) International Application Number:
PCT/US2004/016060

(22) International Filing Date: 21 May 2004 (21.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/472,843 23 May 2003 (23.05.2003) US
60/472,845 23 May 2003 (23.05.2003) US

(71) Applicant *(for all designated States except US)*: **COLD SPRING HARBOR LABORATORY** [US/US]; 1 Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WIGLER, Michael, H.** [US/US]; 52 Snake Hill Road, Cold Spring Harbor, NY 11724 (US). **HEALY, John** [US/US]; 32 Daly Road, East Northport, NY 11731 (US). **LUCITO, Robert** [US/US]; 1477 Eric Lane, East Meadow, NY 11554 (US).

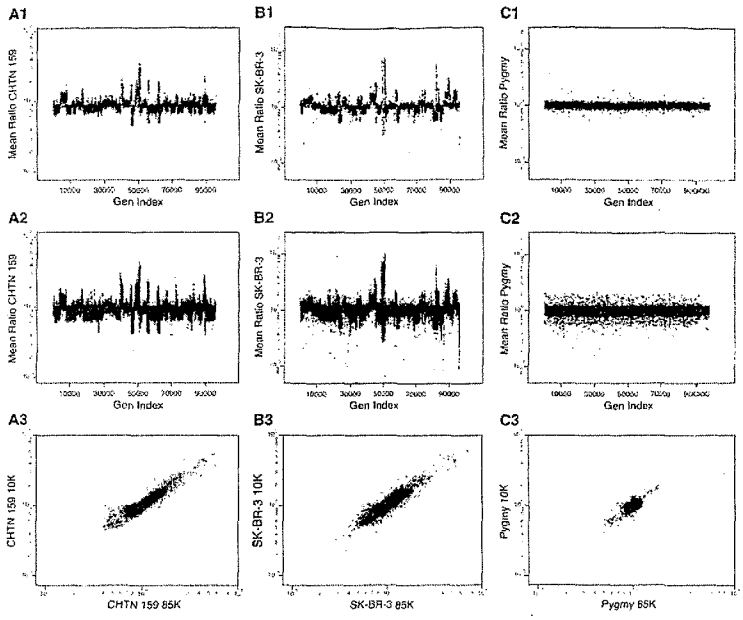
(74) Agents: **HALEY, James, F., Jr.** et al.; Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, 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, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,

[Continued on next page]

(54) Title: VIRTUAL REPRESENTATIONS OF NUCLEOTIDE SEQUENCES



(57) Abstract: The invention provides oligonucleotide probes that can be used to hybridize to a representation of nucleic acid sequences. Compositions containing the probes such as microarrays are also provided. The invention also provides methods of using these probes and compositions in therapeutic, diagnostic, and research applications. Systems and methods for using a word counting algorithm that can quickly and accurately count the number of times a particular string of characters (i.e., nucleotides) appears in a nucleotide sequence (e.g., a genome) are provided. This algorithm can be used to identify the oligonucleotide probes of the invention. The algorithm uses a transform of a genome and an auxiliary data structure to count the number of times a particular word occurs in the genome.

WO 2005/035792 A2

SK, TR), OAPI (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*

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.