

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 March 2006 (23.03.2006)

PCT

(10) International Publication Number
WO 2006/031204 A1

(51) International Patent Classification : **C12Q 1/68**,
C12N 15/00, 15/66

(21) International Application Number:
PCT/SG2005/000281

(22) International Filing Date: 17 August 2005 (17.08.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/939,592 13 September 2004 (13.09.2004) US

(71) Applicant (for all designated States except US):
AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH [SG/SG]; 20 Biopolis Way, #07-01 Centros, Singapore 138668 (SG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SUNG, Wing Kin Ken** [—/SG]; Kent Vale, Block B, Clementi Road 105, #12-03, Singapore 129789 (SG). **RUAN, Yijun** [US/SG]; 123 Verde Crescent, Singapore 688461 (SG).

(74) Agent: **AXIS INTELLECTUAL CAPITAL PTE LTD**;
21A Duxton Road, Singapore 089487 (SG).

(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, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, 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, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

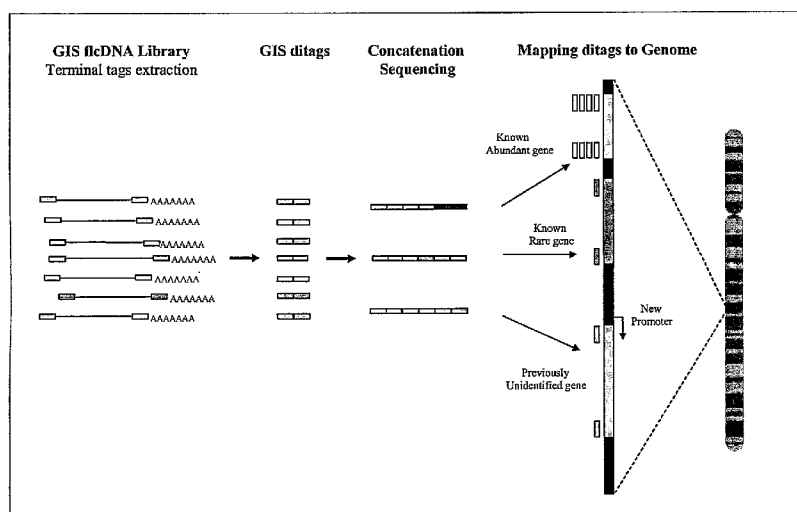
— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations

Published:

— with international search 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.

(54) Title: GENE IDENTIFICATION SIGNATURE (GIS) ANALYSIS FOR TRANSCRIPT MAPPING



(57) Abstract: A transcript mapping method according to an embodiment of the invention is described hereinafter and combines short tag based (SAGE and MPSS) efficiency with the accuracy of full-length cDNA (flcDNA) for comprehensive characterization of transcriptomes. This method is also referred to as Gene Identification Signature (GIS) analysis. In this method, the 5' and 3' ends of full-length cDNA clones are initially extracted into a ditag structure, with the ditag concatemers of the ditag being subsequently sequenced in an efficient manner, and finally mapped to the genome for defining the gene structure. As a GIS ditag represents the 5' and 3' ends of a transcript, it is more informative than SAGE and MPSS tags. Segment lengths between 5' and 3' tag pairs are obtainable including orientation, ordering and chromosome family for efficient transcript mapping and gene location identification. Furthermore, a compressed suffix array (CSA) is used for indexing the genome sequence for improve mapping speed and to reduce computational memory requirements.

WO 2006/031204 A1