

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
11 November 2004 (11.11.2004)

PCT

(10) International Publication Number
WO 2004/097577 A3

- (51) International Patent Classification⁷: **G01N 33/48**
- (21) International Application Number:
PCT/US2004/012921
- (22) International Filing Date: 23 April 2004 (23.04.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/464,983 24 April 2003 (24.04.2003) US
- (71) Applicant (for all designated States except US): **NEW YORK UNIVERSITY** [US/US]; 550 First Avenue, New York, NY 10016 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **CHEREPINSKY, Vera** [US/US]; 68 Hope Street, Unit 11, Stamford, CT 06906 (US). **FENG, Jia-Wu** [CN/US]; Box 92, 1230 York Avenue, New York, NY 10021 (US). **REJALI, Marc** [GB/GB]; 57 Danecroft Road, London, SE24 9PA (GB). **MISHRA, Bhubaneswar** [IN/US]; 16 Dunster Road, Great Neck, NY 11021 (US).
- (74) Agent: **ABELEV, Gary**; Dorsey & Whitney LLP, 250 Park Avenue, New York, NY 10177 (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, 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, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
- with international search report
 - before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
- (88) Date of publication of the international search report:
1 September 2005
- 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: METHODS, SOFTWARE ARRANGEMENTS, STORAGE MEDIA, AND SYSTEMS FOR PROVIDING A SHRINK-AGE-BASED SIMILARITY METRIC

(57) Abstract: The present invention relates to systems, methods, and software arrangements for determining associations between two or more datasets. The systems, methods, and software arrangements used to determine such associations include a determination of a correlation coefficient that incorporates both prior assumptions regarding such datasets and actual information regarding the datasets. The systems, methods, and software arrangements of the present invention can be useful in an analysis of microarray data, including gene expression arrays, to determine correlations between genotypes and phenotypes. Accordingly, the systems, methods, and software arrangements of the present invention may be utilized to determine a genetic basis of complex genetic disorder (e.g. those characterized by the involvement of more than one gene).

WO 2004/097577 A3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/12921

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : GO1N 33/48
 US CL : 702/19

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)
 U.S. : 702/19, 702/32

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 practicable, search terms used)
 Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	ANBAZHAGAN. R. Classification of Small Cell Lung Cancer and Pulmonary Carcinoid by Gene xpression Profiles, Cancer Research, October 1999, Vol. 59, pages 5119-5122, especially page 5120.	1, 2, and 7 ----- 10, 11, 15-22, 49, 50, 55, 58, 59,63-70, 73, 74, 79, 82, 83, and 87-94.
X --- Y	EISEN. M.B. Cluster Analysis and Display of Genome-wide Expression Patterns. PNAS, December 1998, Vol. 95, pages 14863-14868, especially page 14864.	1 and 7 ----- 10, 11, 15-22, 49, 50, 55, 58, 59,63-70, 73, 74, 79, 82, 83, and 87-94.
Y	HOFFMAN. K., Stein Estimation - A Review. Statistical Papers. 2000, Vol. 41, pages 127-158, especially pages 128-129, 136, 146-147.	10-13, 15-22, 58-61, 63-70, 82-85, and 87-94.

Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent published on or after the international filing date	"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
"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)	"&"	document member of the same patent family
"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		

Date of the actual completion of the international search

23 February 2005 (23.02.2005)

Date of mailing of the international search report

15 JUL 2005

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
 Commissioner of Patents
 P.O. Box 1450
 Alexandria, Virginia 22313-1450

Authorized officer

Marjorie Moran.

Janice Ford for

Telephone No. (571-272-1600)

Facsimile No. (703) 305-3230

INTERNATIONAL SEARCH REPORT

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, P	US 2003/0129630 A (AAKALU et al) 10 July 2003 (10.07.2003), pages 1 and 4.	17-22, 49-50, 55, 58-59, 63-70, 73-74, 79, 82-83, and 87-94.
Y	JAMES. W. et al. Estimation with Quadratic Loss. 1961, Vol. 1, pages 361-380, especially pages 361, 362-363.	3-7, 10-11, 51-55, 58-59, 75-79, and 82-83.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/12921

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claim Nos.: 25-48
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 25-48 are directed to a program and not required to be searched under PCT Rule 39.

2. Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
 2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
 3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

 4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: Please See Continuation Sheet
- Remark on Protest The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:

Specie A: the association between the first and second datasets in claims 7, 31, 55, and 79; or 8, 32, 56, and 80; or 9, 33, 57, and 81.

Specie B: estimation of a shrinkage parameter γ in claims 13, 37, 61, and 85; or 14, 38, 62, and 86.

Specie C: correlation between types of data in claims 18-22, 42-46, 66-70, and 90-94; or 23, 47, 71, and 95; or 24, 48, 72, and 96.

The claims are deemed to correspond to the listed species set forth above.

If the applicant elects claims drawn to genotype/phenotype correlation of Specie C, than the applicant is required to elect a particular phenotype recited in claims 22, 46, 70, and 94.

The following claim(s) are generic: 1-6, 10-12, 15-17, 25-30, 34-36, 39-41, 49-54, 58-60, 63-65, 73-78, 82-84, 87-89, and 92-94.

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons:

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons:

1. The associations between datasets of Specie A are divergent because different associations utilize parameters and algorithms independently recognized in the literature so that methods based on the algorithms are not dependent.
2. The estimations of a shrinkage parameter of Specie B are divergent because the estimations are base on independent algorithms and parameters so that correlation coefficient models utilizing the parameters are independent.
3. The correlations between different types of data of Specie C such as genotype/phenotype or financial information for financial instruments or user profiles for an e-commerce applications are separate arts and data for each correlation model are independent from each other.

Continuation of Box II Item 4:

1-7,10-13,15-22,25-31,34-37,39-46,49-55,58-61,63-70,73-79,82-85 and 87-94

INTERNATIONAL SEARCH REPORT

PCT/US04/12921

Continuation of B. FIELDS SEARCHED Item 3:

East, Biosis, Compuscience, Biotechds, Biotechno, Medline, Embase, WPI covering search terms: gene, similarity correlation coefficient, clustering, similarity score, genotype(ing), phenotype(ing), matrix, metric, analysis, shrinkage parameter, data or set or dataset, jamesstein.