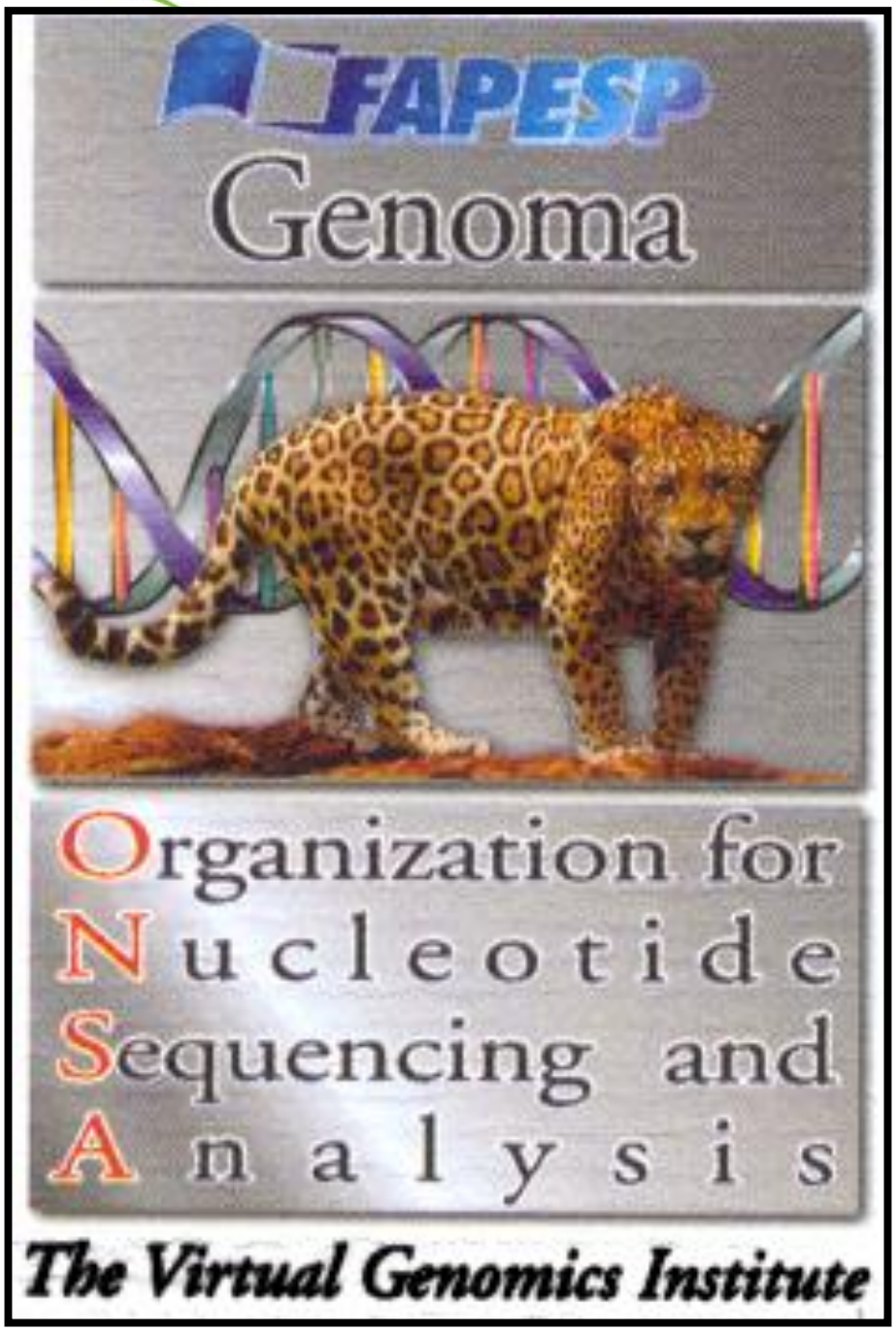

FAPESP
Genoma
Organization for
Nucleotide
Sequencing and
Analysis
The Virtual Genomics Institute

13 July 2000

International weekly journal of science

nature

$10.00

www.nature.com

Citrus pathogen sequenced

Isotope geology
Strange sulphates

AIDS
Mbeki responds to critics

Molecular logic
Chemistry meets computing

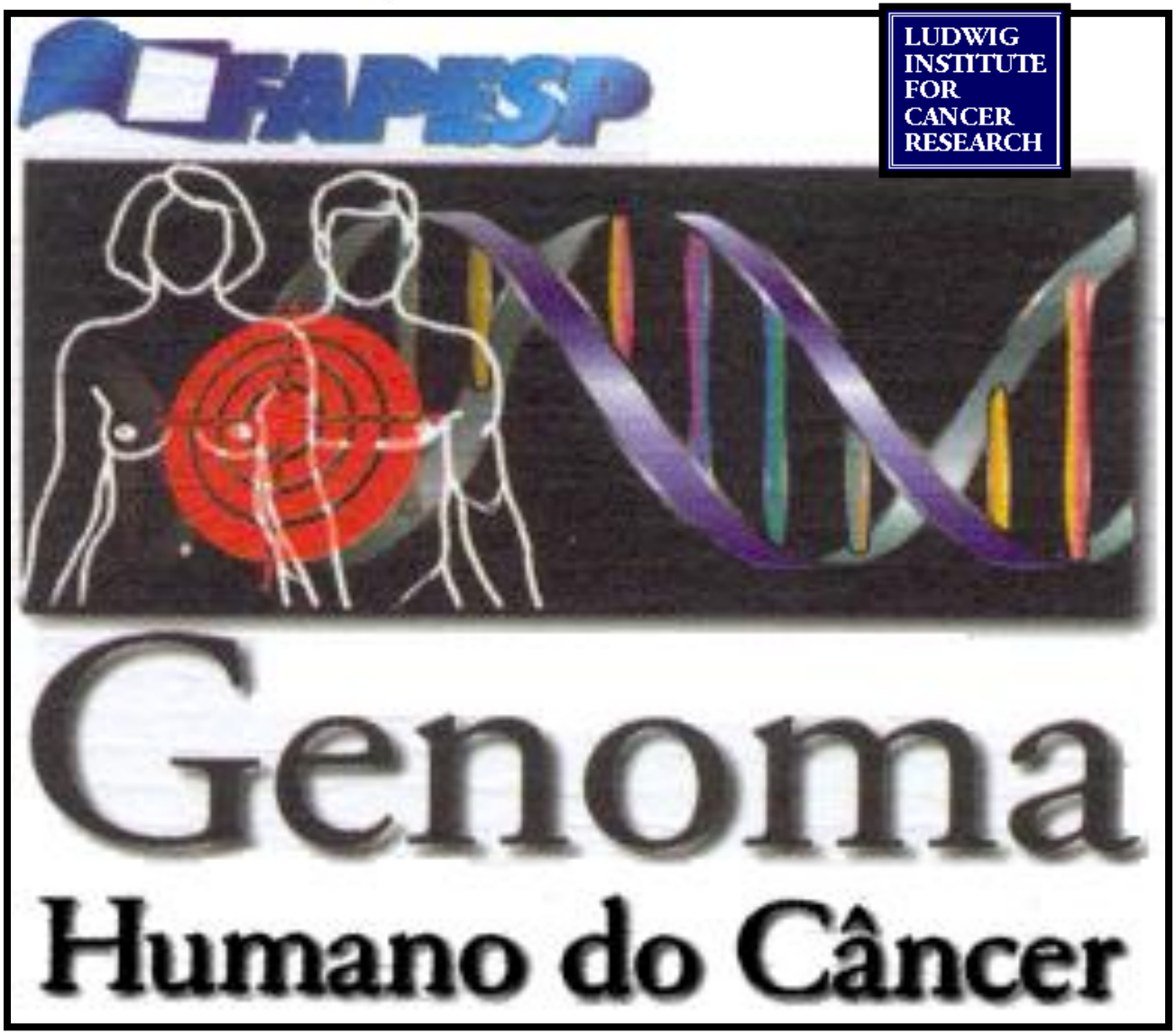
nature jobs
focus on biochemistry

nature

Brazilian scientists team up for cancer genome project

[SÃO PAULO] Brazilian researchers have entered the competitive field of human genome sequencing with the signing of an agreement between the state funding agency of São Paulo (FAPESP) and the US-based Ludwig Institute for Cancer Research.

Each will contribute US$5 million to a two-year Human Cancer Genome Project. According to FAPESP, the programme is "aimed at providing sequences from genes expressed in tumours that are important within the context of public health in the state of São Paulo".



The FAPESP/LICR-Human Cancer Genome Project

RESEARCH NOTE

The Schistosome Genome Project: RNA Arbitrarily Primed- PCR Allows the Accelerated Generation of Expressed Sequence Tags

Emmanuel Dias Neto+, Richard Harrop **, Rodrigo Corrêa-Oliveira, Sérgio DJ Pena *, R Alan Wilson **, Andrew JG Simpson***

Centro de Pesquisas René Rachou - FIOCRUZ, Av. Augusto de Lima 1715, 30190-002 Belo Horizonte, MG Brasil *Departamento de Bioquímica e Imunologia, ICB-UFMG, Caixa Postal 486, 30161-970 Belo Horizonte, MG, Brasil **Department of Biology, University of York, Heslington, York, Y015DD, UK ***Laboratório de Genética de Câncer, Instituto Ludwig de Pesquisas Contra o Câncer, R. Prof. Antônio Prudente 109/ 4º andar, 01509-010 São Paulo, SP, Brasil

the databases before the project was initiated. The adoption of such a strategy proved to be extremely valuable in the analysis of abundant gene transcripts as well as in the discovery of new genes. However, there are some problems inherent to this method, such as the sequencing of vectors without insert and more importantly, the high level of redundancy. For example, of the 429 clones sequenced, 46 (10.7%) represented vectors without insert, and of 202 identified ESTs, homology with only 77 different genes was found, indicating a redundancy of 62% (Franco et al. 1995a *loc. cit.*). Whilst the former problem can be overcome, there is no simple way to solve the latter. These factors reduce the number of useful sequences obtained and thus increase the costs and the time required to tag all of the genes expressed by the parasite. In addition, to tag cDNAs derived from either rare, tissue or stage-specific mRNAs requires complex pre-processing of libraries, such as normalization, subtraction or differential hybridization (C Hoog 1991 *Nucl Acids Res 19*: 6123-6127).

With the aim of increasing the overall efficiency of EST generation, we have applied a technique of RNA arbitrarily primed PCR (RAP-PCR) to this process. RAP-PCR and its close relative "Differential Display", were first described by J Welsh et al. (1992 *Nucl Acids Res 20*: 7213-7218), and P

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

PNAS 97, 12690-12693, 2000

Emmanuel Dias Neto[a], Ricardo Garcia Correa[a], Sergio Verjovski-Almeida[b], Marcelo R. S. Briones[c], Maria Aparecida Nagai[d], Wilson da Silva, Jr.[e], Marco Antonio Zago[e], Silvana Bordin[f], Fernando Ferreira Costa[f], Gustavo Henrique Goldman[g], Alex F. Carvalho[a], Adriana Matsukuma[b], Gilson S. Baia[b], David H. Simpson[h], Adriana Brunstein[a], Paulo S. L. de Oliveira[a], Philipp Bucher[i], C. Victor Jongeneel[j], Michael J. O'Hare[k], Fernando Soares[l],

Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags

PNAS 97, 3491-3496, 2000

Sandro J. de Souza[a], Anamaria A. Camargo[a], Marcelo R. S. Briones[b], Fernando F. (
Sergio Verjovski-Almeida[e], Marco A. Zago[f], Luis Eduardo C. Andrade[g], Helaine Ca
Enilza M. Espreafico[i], Angelita Habr-Gama[j], Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m],
Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p], Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s],
Maria Luisa Paçó-Larson[i], Maria Inês M. C. Pardini[t], Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w],
Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y], Mari C. Sogayar[e], Maria de Fátima Sonati[z], Eloiza H. Tajara[aa],
ttiP,
ura[z],
. A. Melo[b],
laa

The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome

PNAS 98, 12103-12108, 2001

Anamaria A. Camargo[a], Helena P. B. Samaia[a], Emmanuel Dias-Neto[a], Daniel F. Sin
Marcelo R. S. Briones[b], Fernando F. Costa[c], Maria Aparecida Nagai[d], Sergio Verjov
Luis Eduardo C. Andrade[e], Helaine Carrer[h], Hamza F. A. El-Dorry[a], Enilza M. Espreafico[i], Angelita Habr-Gama[j],
Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m], Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p],
Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s], Maria Luisa Paçó-Larson[t], Maria Inês M. C. Pardini[t],
Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w], Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y],
a,cc
son[s],
. Carraro[h],
ra[z],
ria Valente[j],
valho[a],

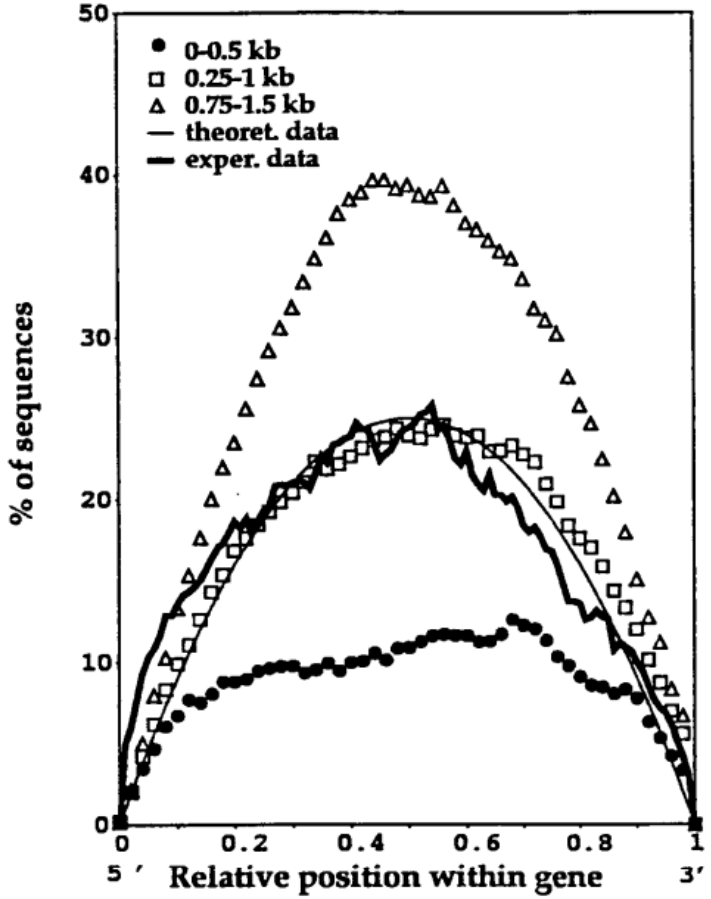
The generation and utilization of a cancer-oriented representation of the human transcriptome, by using expressed sequence tags

PNAS 100, 13418-13423, 2003

Helena Brentani[a], Otávia L. Caballero[a], Anamaria A. Camargo[a], Aline M. da Silva[b], Wilson Araújo da Silva, Jr.[c], Emmanuel Dias Neto[d], Marco Grivet[e], Arthur Gruber[f], Pedro Edson Moreira Guimaraes[d], Winston Hide[g], Christian Iseli[h], C. Victor Jongeneel[h], Janet Kelso[g], Maria Aparecida Nagai[i], Elida Paula Benquique Ojopi[d], Elisson C. Osorio[a], Eduardo M. R. Reis[b], Gregory J. Riggins[l], Andrew John George Simpson[a,k], Sandro de Souza[a], Brian J. Stevenson[h], Robert L. Strausberg[l], Eloiza H. Tajara[m], Sergio Verjovski-Almeida[b],
The Human Cancer Genome Project/Cancer Genome Anatomy Project Annotation Consortium*,
and The Human Cancer Genome Project Sequencing Consortium[†]

FAPESP
Genoma
Humano do Câncer

ORESTES Distribution

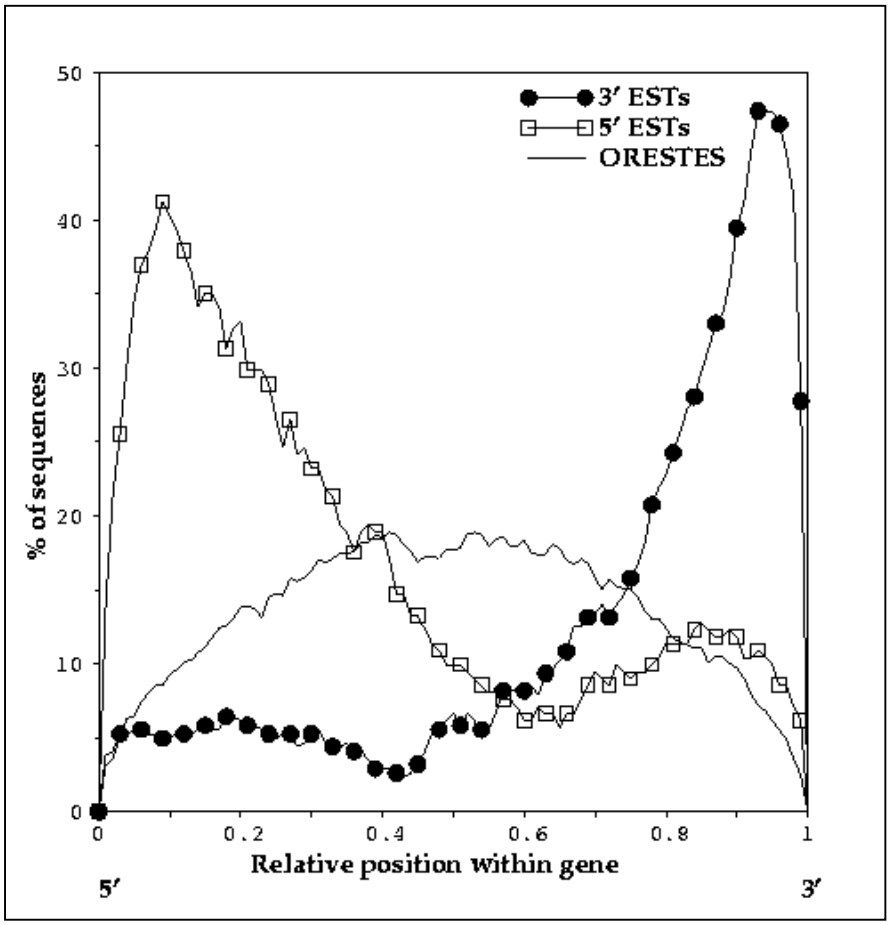


PNAS vol 97, 3491-3496, 2000

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Emmanuel Dias Neto[a], Ricardo Garcia Correa[a], Sergio Verjovski-Almeida[b], Marcelo R. S. Briones[c], Maria Aparecida Nagai[d], Wilson da Silva, Jr.[e], Marco Antonio Zago[e], Silvana Bordin[f], Fernando Ferreira Costa[f], Gustavo Henrique Goldman[g], Alex F. Carvalho[a], Adriana Matsukuma[b], Gilson S. Baia[b], David H. Simpson[h], Adriana Brunstein[a], Paulo S. L. de Oliveira[a], Philipp Bucher[i], C. Victor Jongeneel[j], Michael J. O'Hare[k], Fernando Soares[l], Ricardo R. Brentani[a], Luis F. L. Reis[a], Sandro J. de Souza[a], and Andrew J. G. Simpson[a,m]

ORESTES Distribution

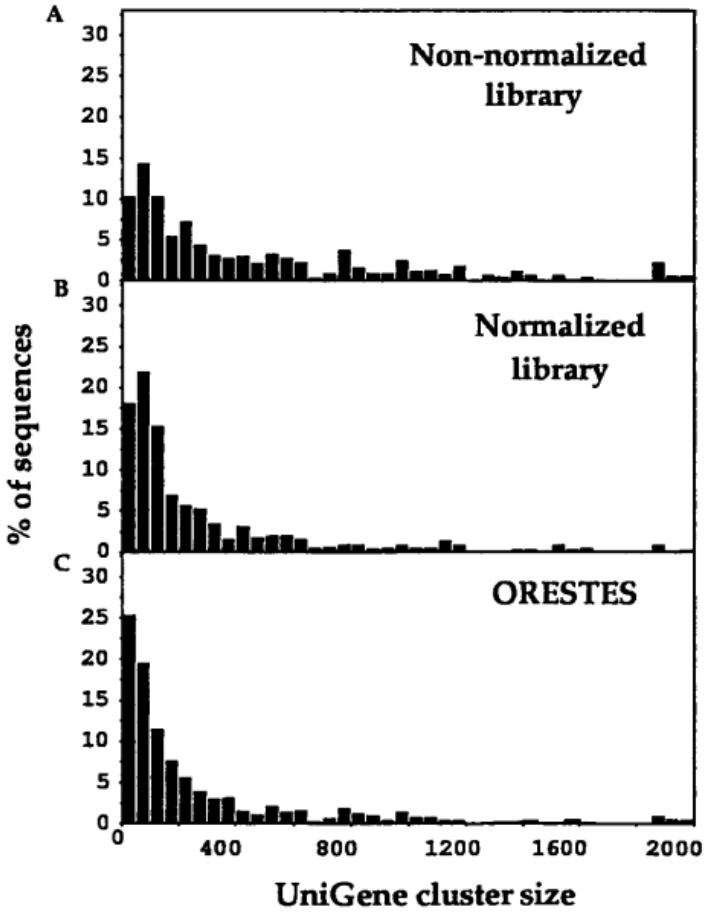


PNAS vol 97, 3491-3496, 2000

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Emmanuel Dias Neto[a], Ricardo Garcia Correa[a], Sergio Verjovski-Almeida[b], Marcelo R. S. Briones[c], Maria Aparecida Nagai[d], Wilson da Silva, Jr.[e], Marco Antonio Zago[e], Silvana Bordin[f], Fernando Ferreira Costa[f], Gustavo Henrique Goldman[g], Alex F. Carvalho[a], Adriana Matsukuma[b], Gilson S. Baia[b], David H. Simpson[h], Adriana Brunstein[a], Paulo S. L. de Oliveira[a], Philipp Bucher[i], C. Victor Jongeneel[j], Michael J. O'Hare[k], Fernando Soares[l], Ricardo R. Brentani[a], Luis F. L. Reis[a], Sandro J. de Souza[a], and Andrew J. G. Simpson[a,m]

ORESTES Distribution



PNAS vol 97, 3491-3496, 2000

Median cluster sizes 317, 138, and 125,

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Emmanuel Dias Neto[a], Ricardo Garcia Correa[a], Sergio Verjovski-Almeida[b], Marcelo R. S. Briones[c], Maria Aparecida Nagai[d], Wilson da Silva, Jr.[e], Marco Antonio Zago[e], Silvana Bordin[f], Fernando Ferreira Costa[f], Gustavo Henrique Goldman[g], Alex F. Carvalho[a], Adriana Matsukuma[b], Gilson S. Baia[b], David H. Simpson[h], Adriana Brunstein[a], Paulo S. L. de Oliveira[a], Philipp Bucher[i], C. Victor Jongeneel[j], Michael J. O'Hare[k], Fernando Soares[l], Ricardo R. Brentani[a], Luis F. L. Reis[a], Sandro J. de Souza[a], and Andrew J. G. Simpson[a,m]

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

PNAS 97, 12690-12693, 2000

Emmanuel Dias Neto[a], Ricardo Garcia Correa[a], Sergio Verjovski-Almeida[b], Marcelo R. S. Briones[c], Maria Aparecida Nagai[d], Wilson da Silva, Jr.[e], Marco Antonio Zago[e], Silvana Bordin[f], Fernando Ferreira Costa[f], Gustavo Henrique Goldman[g], Alex F. Carvalho[a], Adriana Matsukuma[b], Gilson S. Baia[b], David H. Simpson[h], Adriana Brunstein[a], Paulo S. L. de Oliveira[a], Philipp Bucher[i], C. Victor Jongeneel[j], Michael J. O'Hare[k], Fernando Soares[l],

Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags

PNAS 97, 3491-3496, 2000

Sandro J. de Souza[a], Anamaria A. Camargo[a], Marcelo R. S. Briones[b], Fernando F. (
Sergio Verjovski-Almeida[e], Marco A. Zago[f], Luis Eduardo C. Andrade[g], Helaine Ca
Enilza M. Espreafico[i], Angelita Habr-Gama[j], Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m],
Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p], Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s],
Maria Luisa Paçó-Larson[i], Maria Inês M. C. Pardini[t], Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w],
Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y], Mari C. Sogayar[e], Maria de Fátima Sonati[z], Eloiza H. Tajara[aa],
ttiP,
ura[z],
. A. Melo[b],
laa

The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome

PNAS 98, 12103-12108, 2001

Anamaria A. Camargo[a], Helena P. B. Samaia[a], Emmanuel Dias-Neto[a], Daniel F. Sin
Marcelo R. S. Briones[b], Fernando F. Costa[c], Maria Aparecida Nagai[d], Sergio Verjov
Luis Eduardo C. Andrade[e], Helaine Carrer[h], Hamza F. A. El-Dorry[a], Enilza M. Espreafico[i], Angelita Habr-Gama[j],
Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m], Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p],
Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s], Maria Luisa Paçó-Larson[t], Maria Inês M. C. Pardini[t],
Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w], Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y],
a,cc
son[s],
. Carraro[h],
ra[z],
ria Valente[j],
valho[a],

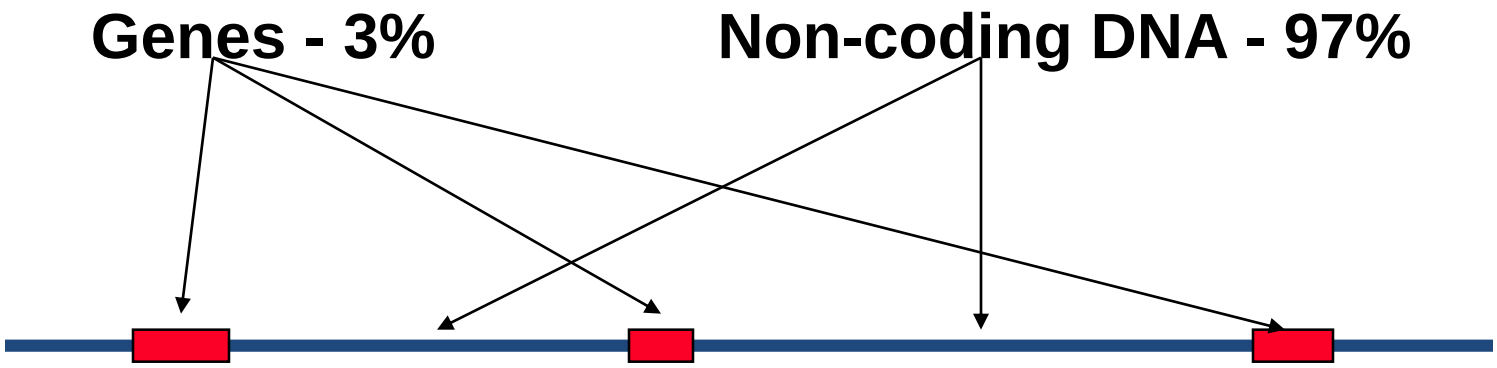
The generation and utilization of a cancer-oriented representation of the human transcriptome, by using expressed sequence tags

PNAS 100, 13418-13423, 2003

Helena Brentani[a], Otávia L. Caballero[a], Anamaria A. Camargo[a], Aline M. da Silva[b], Wilson Araújo da Silva, Jr.[c], Emmanuel Dias Neto[d], Marco Grivet[e], Arthur Gruber[f], Pedro Edson Moreira Guimaraes[d], Winston Hide[g], Christian Iseli[h], C. Victor Jongeneel[h], Janet Kelso[g], Maria Aparecida Nagai[i], Elida Paula Benquique Ojopi[d], Elisson C. Osorio[a], Eduardo M. R. Reis[b], Gregory J. Riggins[l], Andrew John George Simpson[a,k], Sandro de Souza[a], Brian J. Stevenson[h], Robert L. Strausberg[l], Eloiza H. Tajara[m], Sergio Verjovski-Almeida[b],
The Human Cancer Genome Project/Cancer Genome Anatomy Project Annotation Consortium*,
and The Human Cancer Genome Project Sequencing Consortium[†]

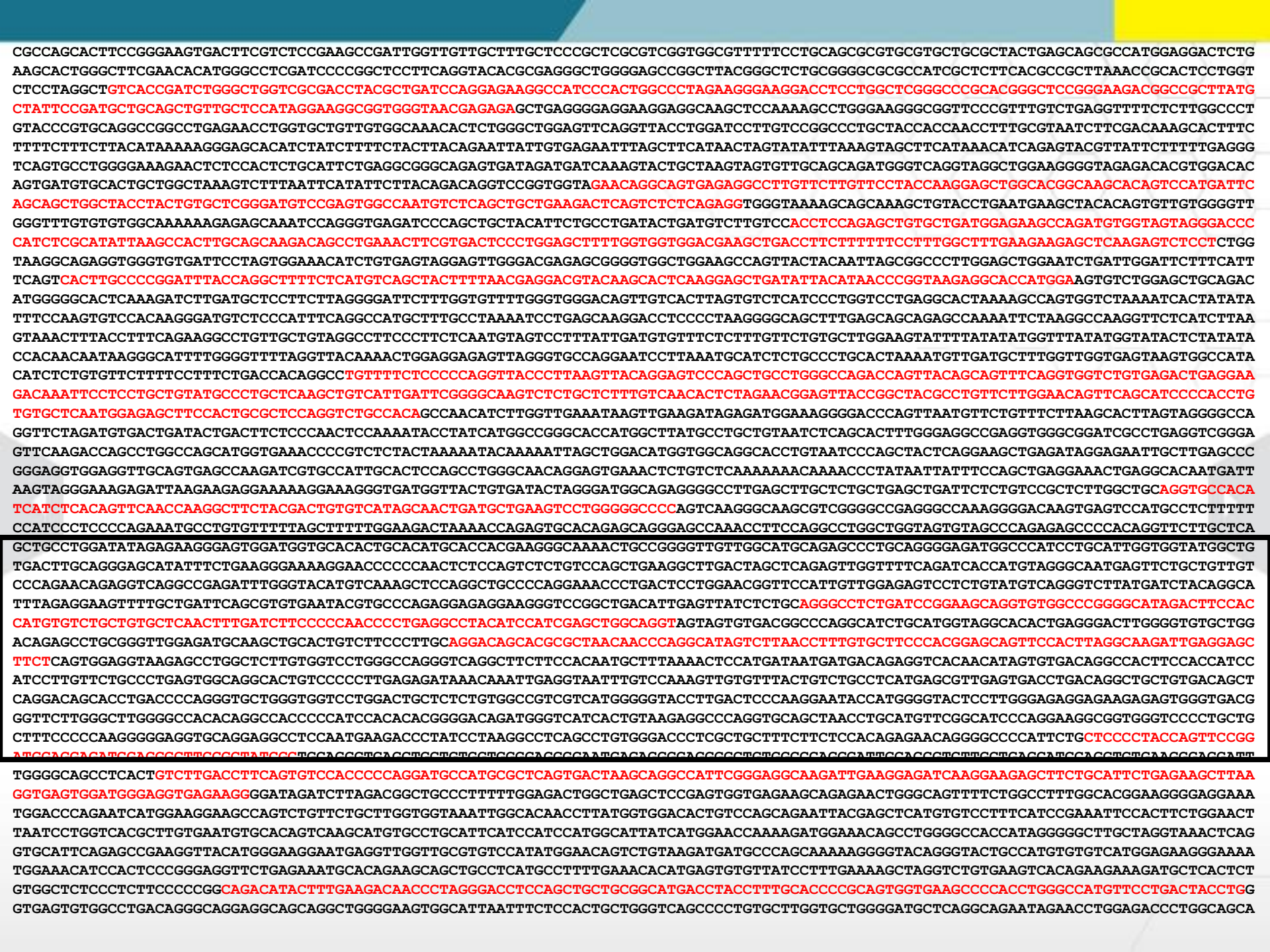
FAPESP
Genoma
Humano do Câncer

Human Genome Sequencing



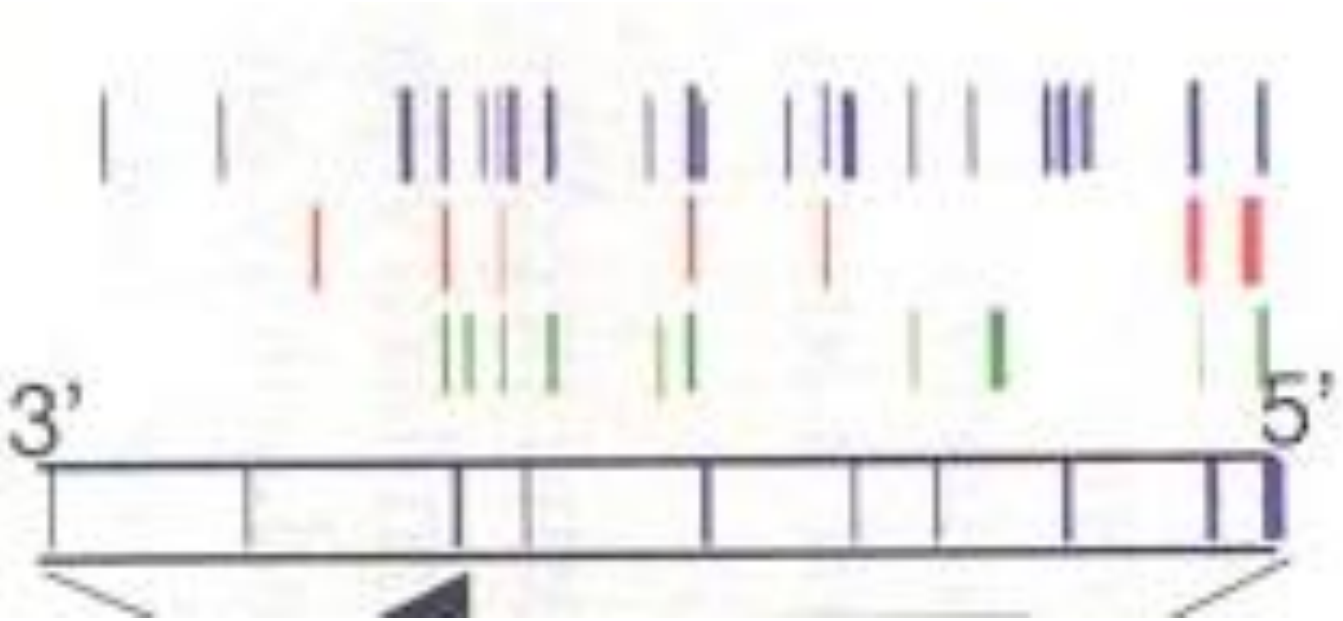
HGP

Celera



AGCTGCCTGGATATAGAGAAGGGAGTGGATGGTGCACACTGCACATGCACCACGAAGGGCAAAAC
TGCCGGGGTTGTTGGCATGCAGAGCCCTGCAGGGGAGATGGCCCATCCTGCATTGGTGGTATGGC
TTGACTTGCAGGGAGCATATTTCTGAAGGGAAAAGGAACCCCCCAACTCTCCAGTCTCTGTCCAG
CTGAAGGCTTGACTAGCTCAGAGTTGGTTTTCAGATCACCATGTAGGGCAATGAGTTCTGCTGTT
GTCCCAGAACAGAGGTCAGGCCGAGATTTGGGTACATGTCAAAGCTCCAGGCTGCCCCAGGAAAC
CCTGACTCCTGGAACGGTTCCATTGTTGGAGAGTCCTCTGTATGTCAGGGTCTTATGATCTACAG
GCATTTAGAGGAAGTTTTGCTGATTCAGCGTGTGAATACGTGCCCAGAGGAGAGGAAGGGTCCGG
CTGACATTGAGTTATCTCTGCAGGGCCTCTGATCCGGAAGCAGGTGTGGCCCGGGGCATAGACTT
CCACCATGTGTCTGCTGTGCTCAACTTTGATCTTCCCCCAACCCCTGAGGCCTACATCCATCGAG
CTGGCAGGTAGTAGTGTGACGGCCCAGGCATCTGCATGGTAGGCACACTGAGGGACTTGGGGTGT
GCTGGACAGAGCCTGCGGGTTGGAGATGCAAGCTGCACTGTCTTCCCTTGCAGGACAGCACGCGC
TAACAACCCAGGCATAGTCTTAACCTTTGTGCTTCCCACGGAGCAGTTCCACTTAGGCAAGATTG
AGGAGCTTCTCAGTGGAGGTAAGAGCCTGGCTCTTGTGGTCCTGGGCCAGGGTCAGGCTTCTTCC
ACAATGCTTTAAAACTCCATGATAATGATGACAGAGGTCACAACATAGTGTGACAGGCCACTTCC
ACCATCCATCCTTGTTCTGCCCTGAGTGGCAGGCACTGTCCCCCTTGAGAGATAAACAAATTGAG
GTAATTTGTCCAAAGTTGTGTTTACTGTCTGCCTCATGAGCGTTGAGTGACCTGACAGGCTGCTG
TGACAGCTCAGGACAGCACCTGACCCCAGGGTGCTGGGTGGTCCTGGACTGCTCTCTGTGGCCGT
CGTCATGGGGGTACCTTGACTCCCAAGGAATACCATGGGGTACTCCTTGGGAGAGGAGAAGAGAG
TGGGTGACGGGTTCTTGGGCTTGGGGCCACACAGGCCACCCCCATCCACACACGGGGACAGATGG
GTCATCACTGTAAGAGGCCCAGGTGCAGCTAACCTGCATGTTCGGCATCCCAGGAAGGCGGTGGG
TCCCCTGCTGCTTTCCCCCAAGGGGAGGTGCAGGAGGCCTCCAATGAAGACCCTATCCTAAGGC
CTCAGCCTGTGGGACCCTCGCTGCTTTCTTCTCCACAGAGAACAGGGGCCCCATTCTGCTCCCCT
ACCAGTTCCGGATGGAGGAGATCGAGGGCTTCCGCTATCGCTGCAGGGTGAGCTGCTGTGGTGGG

Exon Prediction



MZEF

Genescan

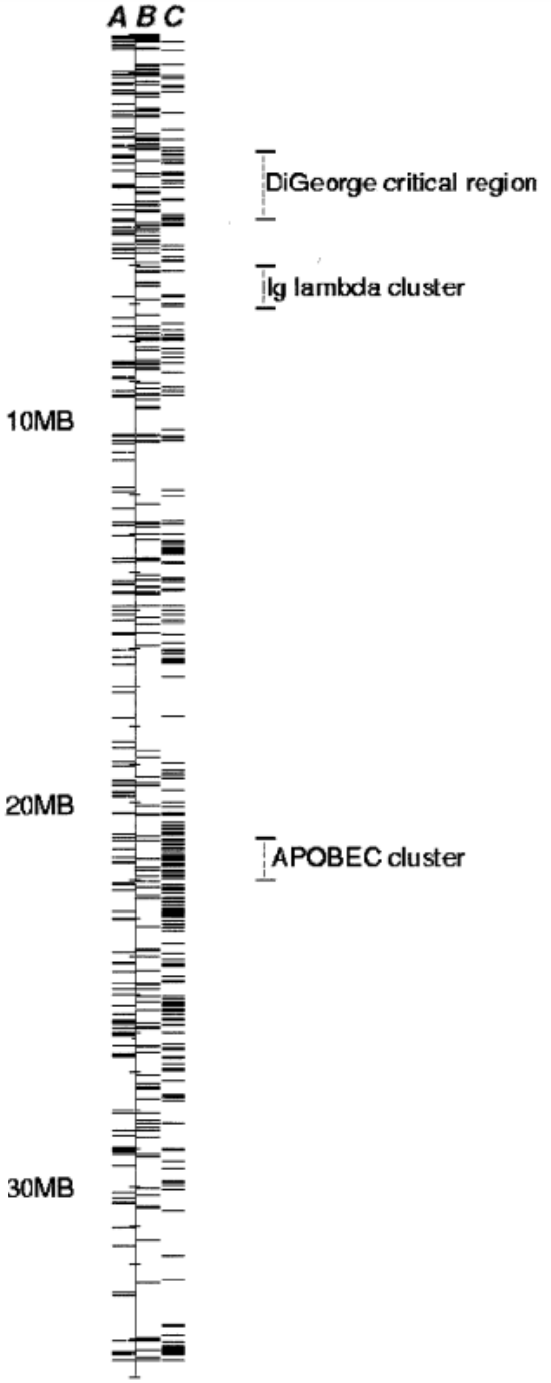
Grail

Experimentally defined exons

Exon prediction in an A+T rich region of chromosome 21

- 250, 000 ORESTES
- 81,428 contigs
- 1,181 mapped to chromosome 22 (65.6% of 247 known genes*)
- 219 unanotated transcribed sequences of which 171 actually present in GenBank as cDNAs, 48 totally new

A = ORESTES predicted transcribed sequences
B = Known genes *
C = Genes predicted from other ESTs *



*

articles

The DNA sequence of human chromosome 22

I. Dunham, N. Shimizu, B. A. Roe, S. Chissoe *et al.*†

† *A full list of authors appears at the end of this paper*

NATURE | VOL 402 | 2 DECEMBER 1999 | www.nature.com

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

PNAS 97, 12690-12693, 2000

Emmanuel Dias Neto[a], Ricardo Garcia Correa[a], Sergio Verjovski-Almeida[b], Marcelo R. S. Briones[c], Maria Aparecida Nagai[d], Wilson da Silva, Jr.[e], Marco Antonio Zago[e], Silvana Bordin[f], Fernando Ferreira Costa[f], Gustavo Henrique Goldman[g], Alex F. Carvalho[a], Adriana Matsukuma[b], Gilson S. Baia[b], David H. Simpson[h], Adriana Brunstein[a], Paulo S. L. de Oliveira[a], Philipp Bucher[i], C. Victor Jongeneel[j], Michael J. O'Hare[k], Fernando Soares[l],

Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags

PNAS 97, 3491-3496, 2000

Sandro J. de Souza[a], Anamaria A. Camargo[a], Marcelo R. S. Briones[b], Fernando F. (
Sergio Verjovski-Almeida[e], Marco A. Zago[f], Luis Eduardo C. Andrade[g], Helaine Ca
Enilza M. Espreafico[i], Angelita Habr-Gama[j], Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m],
Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p], Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s],
Maria Luisa Paçó-Larson[i], Maria Inês M. C. Pardini[t], Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w],
Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y], Mari C. Sogayar[e], Maria de Fátima Sonati[z], Eloiza H. Tajara[aa],
ttiP,
ura[z],
. A. Melo[b],
laa

The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome

PNAS 98, 12103-12108, 2001

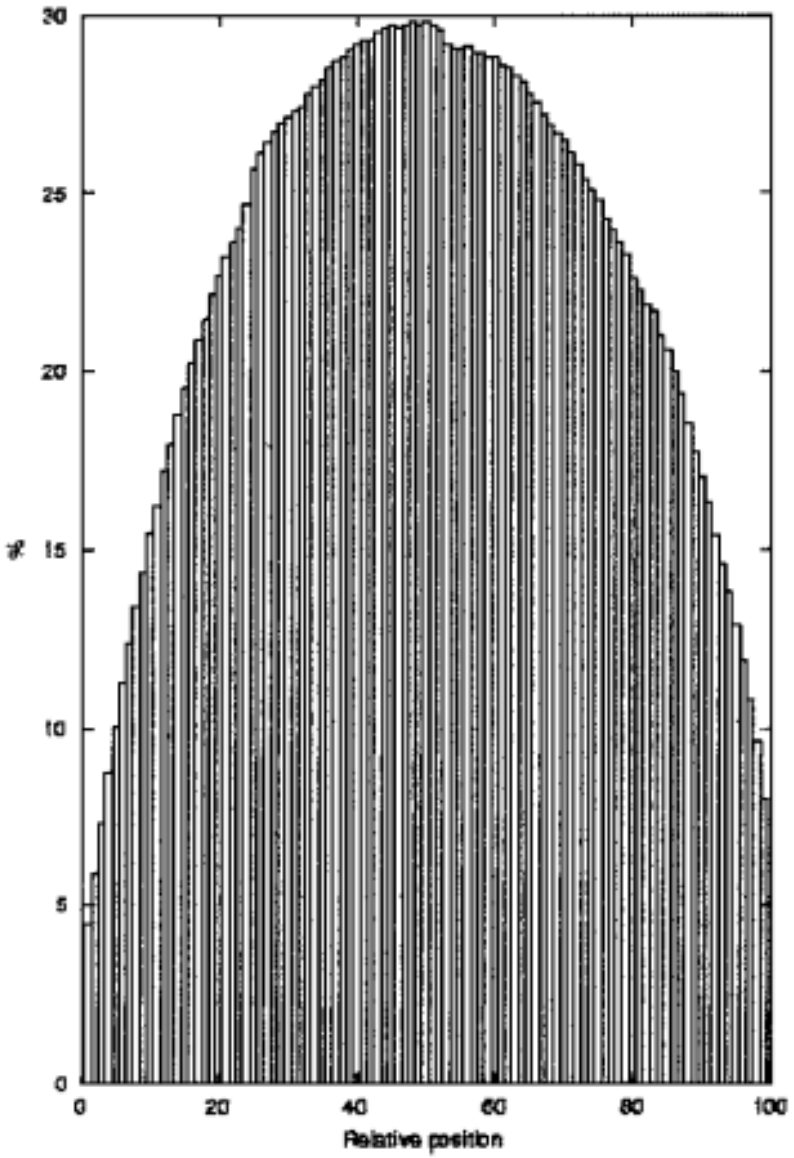
Anamaria A. Camargo[a], Helena P. B. Samaia[a], Emmanuel Dias-Neto[a], Daniel F. Sin
Marcelo R. S. Briones[b], Fernando F. Costa[c], Maria Aparecida Nagai[d], Sergio Verjov
Luis Eduardo C. Andrade[e], Helaine Carrer[h], Hamza F. A. El-Dorry[a], Enilza M. Espreafico[i], Angelita Habr-Gama[j],
Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m], Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p],
Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s], Maria Luisa Paçó-Larson[t], Maria Inês M. C. Pardini[t],
Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w], Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y],
a,cc
son[s],
. Carraro[h],
ra[z],
ria Valente[j],
valho[a],

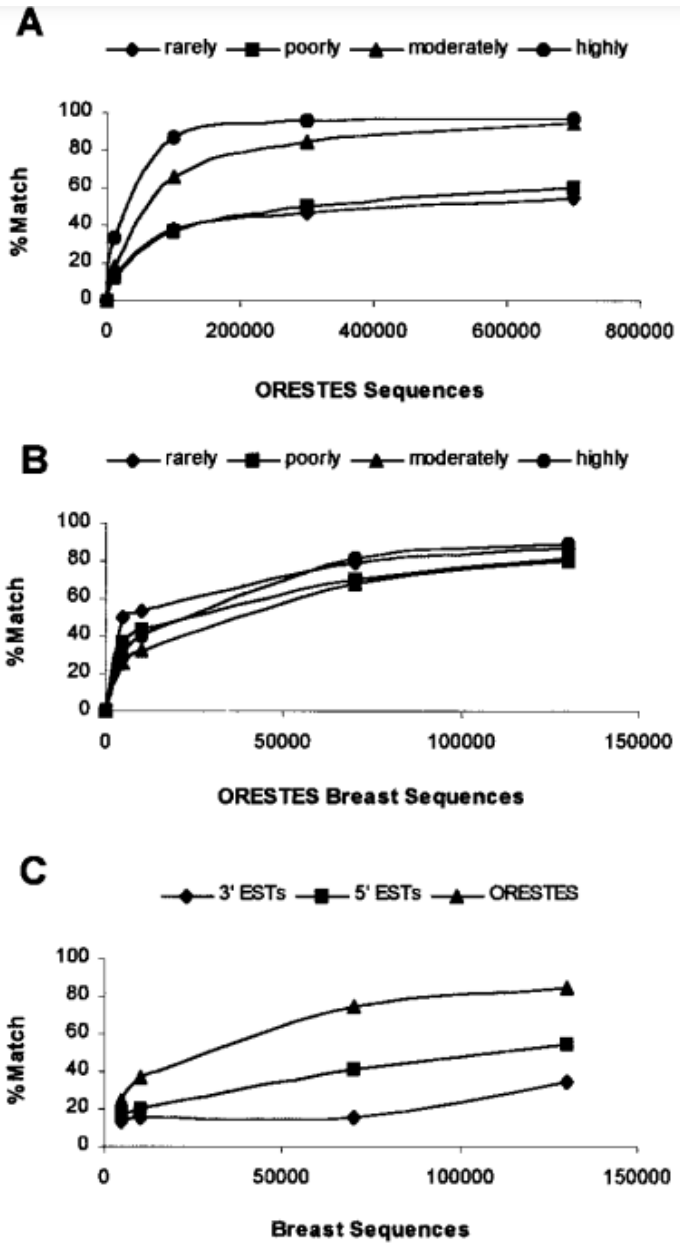
The generation and utilization of a cancer-oriented representation of the human transcriptome, by using expressed sequence tags

PNAS 100, 13418-13423, 2003

Helena Brentani[a], Otávia L. Caballero[a], Anamaria A. Camargo[a], Aline M. da Silva[b], Wilson Araújo da Silva, Jr.[c], Emmanuel Dias Neto[d], Marco Grivet[e], Arthur Gruber[f], Pedro Edson Moreira Guimaraes[d], Winston Hide[g], Christian Iseli[h], C. Victor Jongeneel[h], Janet Kelso[g], Maria Aparecida Nagai[i], Elida Paula Benquique Ojopi[d], Elisson C. Osorio[a], Eduardo M. R. Reis[b], Gregory J. Riggins[l], Andrew John George Simpson[a,k], Sandro de Souza[a], Brian J. Stevenson[h], Robert L. Strausberg[l], Eloiza H. Tajara[m], Sergio Verjovski-Almeida[b],
The Human Cancer Genome Project/Cancer Genome Anatomy Project Annotation Consortium*,
and The Human Cancer Genome Project Sequencing Consortium[†]

FAPESP
Genoma
Humano do Câncer


30
25
20
15
10
5
0
0
20
40
60
80
100
Relative position


A
rarely
poorly
moderately
highly
% Match
100
80
60
40
20
0
0
200000
400000
600000
800000
ORESTES Sequences
B
rarely
poorly
moderately
highly
% Match
100
80
60
40
20
0
0
50000
100000
150000
ORESTES Breast Sequences
C
3' ESTs
5' ESTs
ORESTES
% Match
100
80
60
40
20
0
0
50000
100000
150000
Breast Sequences

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

PNAS 97, 12690-12693, 2000

Emmanuel Dias Neto[a], Ricardo Garcia Correa[a], Sergio Verjovski-Almeida[b], Marcelo R. S. Briones[c], Maria Aparecida Nagai[d], Wilson da Silva, Jr.[e], Marco Antonio Zago[e], Silvana Bordin[f], Fernando Ferreira Costa[f], Gustavo Henrique Goldman[g], Alex F. Carvalho[a], Adriana Matsukuma[b], Gilson S. Baia[b], David H. Simpson[h], Adriana Brunstein[a], Paulo S. L. de Oliveira[a], Philipp Bucher[i], C. Victor Jongeneel[j], Michael J. O'Hare[k], Fernando Soares[l],

Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags

PNAS 97, 3491-3496, 2000

Sandro J. de Souza[a], Anamaria A. Camargo[a], Marcelo R. S. Briones[b], Fernando F. (
Sergio Verjovski-Almeida[e], Marco A. Zago[f], Luis Eduardo C. Andrade[g], Helaine Ca
Enilza M. Espreafico[i], Angelita Habr-Gama[j], Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m],
Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p], Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s],
Maria Luisa Paçó-Larson[i], Maria Inês M. C. Pardini[t], Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w],
Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y], Mari C. Sogayar[e], Maria de Fátima Sonati[z], Eloiza H. Tajara[aa],
ttiP,
ura[z],
. A. Melo[b],
laa

The contribution of 700,000 ORF sequence tags to the definition of the human transcriptome

PNAS 98, 12103-12108, 2001

Anamaria A. Camargo[a], Helena P. B. Samaia[a], Emmanuel Dias-Neto[a], Daniel F. Sin
Marcelo R. S. Briones[b], Fernando F. Costa[c], Maria Aparecida Nagai[d], Sergio Verjov
Luis Eduardo C. Andrade[e], Helaine Carrer[h], Hamza F. A. El-Dorry[a], Enilza M. Espreafico[i], Angelita Habr-Gama[j],
Daniel Giannella-Neto[k], Gustavo H. Goldman[l], Arthur Gruber[m], Christine Hackel[n], Edna T. Kimura[o], Rui M. B. Maciel[p],
Suely K. N. Marie[q], Elizabeth A. L. Martins[r], Marina P. Nóbrega[s], Maria Luisa Paçó-Larson[t], Maria Inês M. C. Pardini[t],
Gonçalo G. Pereira[u], João Bosco Pesquero[v], Vanderlei Rodrigues[w], Silvia R. Rogatto[x], Ismael D. C. G. da Silva[y],
a,cc
son[s],
. Carraro[h],
ra[z],
ria Valente[j],
valho[a],

The generation and utilization of a cancer-oriented representation of the human transcriptome, by using expressed sequence tags

PNAS 100, 13418-13423, 2003

Helena Brentani[a], Otávia L. Caballero[a], Anamaria A. Camargo[a], Aline M. da Silva[b], Wilson Araújo da Silva, Jr.[c], Emmanuel Dias Neto[d], Marco Grivet[e], Arthur Gruber[f], Pedro Edson Moreira Guimaraes[d], Winston Hide[g], Christian Iseli[h], C. Victor Jongeneel[h], Janet Kelso[g], Maria Aparecida Nagai[i], Elida Paula Benquique Ojopi[d], Elisson C. Osorio[a], Eduardo M. R. Reis[b], Gregory J. Riggins[l], Andrew John George Simpson[a,k], Sandro de Souza[a], Brian J. Stevenson[h], Robert L. Strausberg[l], Eloiza H. Tajara[m], Sergio Verjovski-Almeida[b],
The Human Cancer Genome Project/Cancer Genome Anatomy Project Annotation Consortium*,
and The Human Cancer Genome Project Sequencing Consortium[†]

FAPESP
Genoma
Humano do Câncer

In two large projects, extensive EST sequencing of human tumor tissues has been undertaken: the Cancer Genome Anatomy Project (CGAP) (3) and the Fundaçao de Amparo à Pesquisa do Estado de Sao Paulo/Ludwig Institute for Cancer Research–Human Cancer Genome Project (HCGP) (4, 5). CGAP, launched in 1997 by the National Cancer Institute, has used single-pass sequencing from the 5′ and/or 3′ extremities of cDNA clones for sequence generation (6). The HCGP project adopted an alternative EST-based strategy, termed ORESTES, which generates sequences biased toward the central coding regions of transcripts (7). The data gathered by these two projects are thus complementary and have been combined into an International Database of Cancer Gene Expression (5), available at http://cgap.nci.nih.gov. They also constitute the basis of the Human Cancer Index at TIGR (www.tigr.org/tdb/tgi/hcgi).

Comparative Study > Nat Genet. 2000 Jun;25(2):232-4. doi: 10.1038/76115.

Analysis of expressed sequence tags indicates 35,000 human genes

B Ewing [1], P Green

Affiliations + expand

PMID: 10835644 DOI: 10.1038/76115

NATURE | VOL 409 | 15 FEBRUARY 2001 | www.nature.com

articles

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium*

* *A partial list of authors appears on the opposite page. Affiliations are listed at the end of the paper.*

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

The rediscovery of Mendel's laws of heredity in the opening weeks of the 20th century[1–3] sparked a scientific quest to understand the nature and content of genetic information that has propelled biology for the last hundred years. The scientific progress made falls naturally into four main phases, corresponding roughly to the four quarters of the century. The first established the cellular basis of heredity: the chromosomes. The second defined the molecular basis of heredity: the DNA double helix. The third unlocked the informational basis of heredity, with the discovery of the biological mechanism by which cells read the information contained in genes and with the invention of the recombinant DNA technologies of cloning and sequencing by which scientists can do the same.

The last quarter of a century has been marked by a relentless drive

coordinate regulation of the genes in the clusters.

- There appear to be about 30,000–40,000 protein-coding genes in the human genome—only about twice as many as in worm or fly. However, the genes are more complex, with more alternative splicing generating a larger number of protein products.
- The full set of proteins (the 'proteome') encoded by the human genome is more complex than those of invertebrates. This is due in part to the presence of vertebrate-specific protein domains and motifs (an estimated 7% of the total), but more to the fact that vertebrates appear to have arranged pre-existing components into a richer collection of domain architectures.
- Hundreds of human genes appear likely to have resulted from horizontal transfer from bacteria at some point in the vertebrate

Table 1. HCGP and CGAP transcript sequence generation and clustering

Form of gene representation	Number of sequences, clusters, or genes
ORESTES submitted to GenBank	823,121 sequences
CGAP EST submitted to GenBank	1,214,358 sequences
TOTAL EST submitted to GenBank	2,037,479 sequences
Total clusters	32,129 clusters
Total clusters with known genes	22,152 clusters
Clusters without known genes	9,977 clusters
Clusters without known genes but with coding potential	1,285 clusters
Estimated total genes based on HCGP and CGAP data	23,437 genes

number of human genes - Googl

google.com/search?q=number+of+human+genes&rlz=1C1CHZN_pt-BRBR916BR916&oq=number+of+human+genes&aqs=chrome..69i57j0i5...

Google

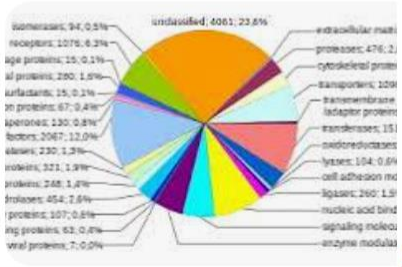
number of human genes

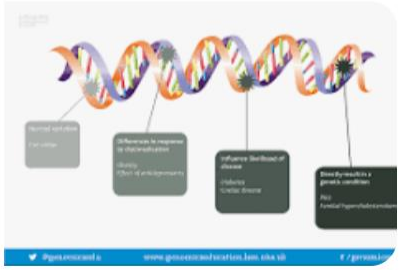
Sign in

All Images News Videos Books More Tools

About 780,000,000 results (0.57 seconds)

An international research effort called the Human Genome Project, which worked to determine the sequence of the human genome and identify the genes that it contains, estimated that humans have **between 20,000 and 25,000** genes. Every person has two copies of each gene, one inherited from each parent. Mar 22, 2021





https://medlineplus.gov › ... › Cells and DNA

What is a gene?: MedlinePlus Genetics

About featured snippets • Feedback

A Transcript Finishing Initiative for Closing Gaps in the Human Transcriptome

The Ludwig–FAPESP Transcript Finishing Initiative,[1] Mari Cleide Sogayar,[2] and Anamaria A. Camargo[2]

14:1413–1423 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

Complete List of Authors

Coordination Group Ludwig Institute

Fabiana Bettoni,[3] Dirce Maria Carraro,[3] Lilian C. Pires,[3] Raphael B. Parmigiani,[3] Elisa N. Ferreira,[3] Eloísa de Sá Moreira,[3,32] Maria do Rosário D. de O. Latorre,[4] Andrew J.G. Simpson,[3] and Anamaria A. Camargo[3]

Coordination Group University of São Paulo Chemistry Institute

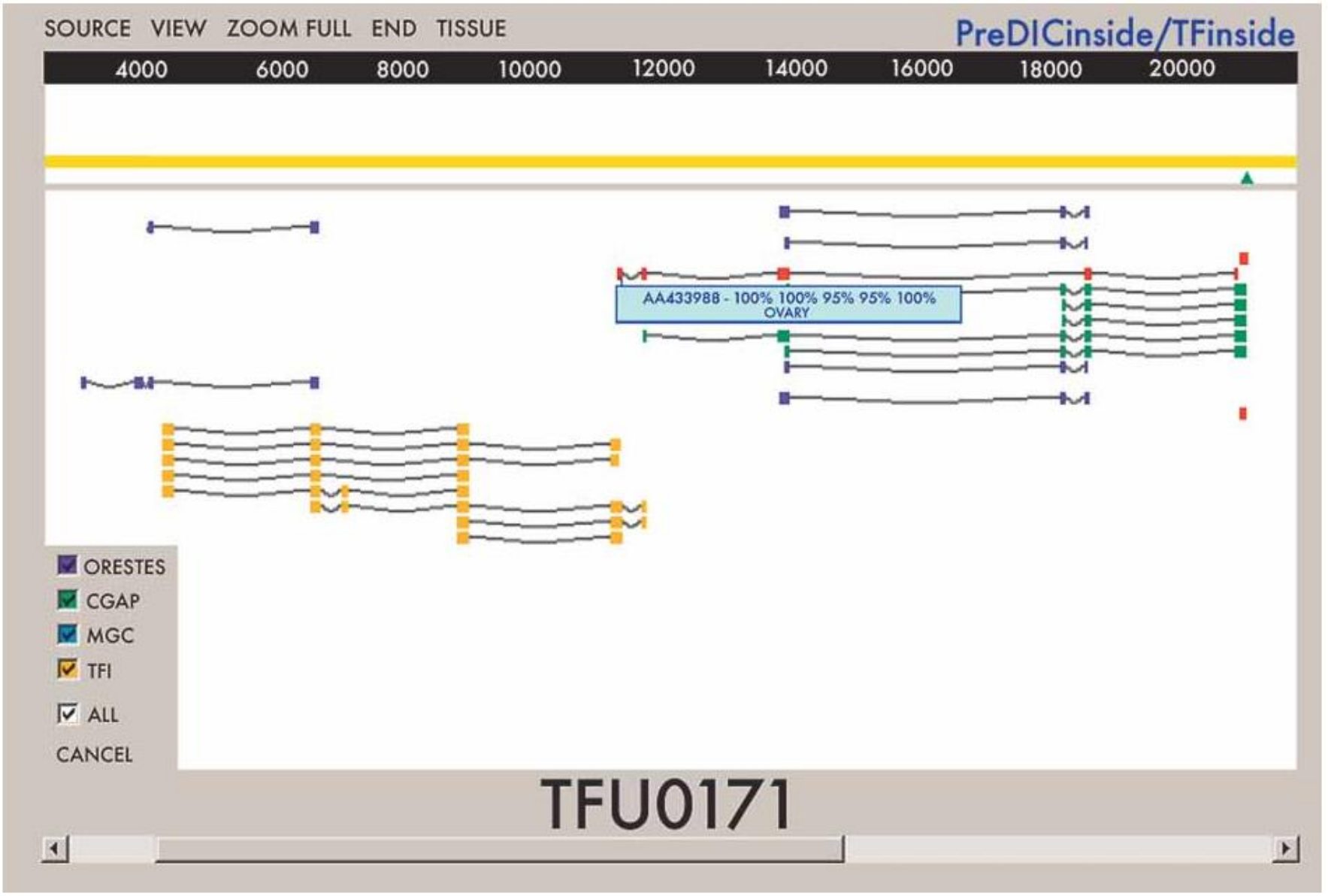
Luciana Oliveira Cruz,[5] Theri Leica Degaki,[5] Fernanda Festa,[5] Katlin B. Massirer,[5] and Mari C. Sogayar[5]

Bioinformatics Groups

Fernando Camargo Filho,[6] Luiz Paulo Camargo,[6] Marco A.V. Cunha,[7] Sandro J. De Souza,[8] Milton Faria Junior,[6] Silvana Giuliatti,[6] Leonardo Kopp,[9] Paulo S.L. de Oliveira,[9] Paulo B. Paiva,[10] Anderson A. Pereira,[6] Daniel G. Pinheiro,[7] Renato D. Puga,[6] and Jorge Estefano S. de Souza[8]

Validation Groups

Dulcineia M. Albuquerque,[11] Luís E.C. Andrade,[12] Gilson S. Baia,[13] Marcelo R.S. Briones,[14] Ana M.S. Cavaleiro–Luna,[15] Janete M. Cerutti,[16] Fernando F. Costa,[11] Eugenia Costanzi-Strauss,[17] Enilza M. Espreafico,[18] Adriana C. Ferrasi,[19] Emer S. Ferro,[13] Maria A.H.Z. Fortes,[15] Joelma R.F. Furchi,[20] Daniel Giannella-Neto,[15] Gustavo H. Goldman,[21] Maria H.S. Goldman,[22] Arthur Gruber,[23] Gustavo S. Guimarães,[16] Christine Hackel,[24] Flavio Henrique-Silva,[20] Edna T. Kimura,[13] Suzana G. Leoni,[11] Cláudia Macedo,[25] Bettina Malnic,[26] Carina V. Manzini B.,[26] Suely K.N. Marie,[27] Nilce M. Martinez-Rossi,[25] Marcelo Menossi,[28,29] Elisabete C. Miracca,[30] Maria A. Nagai,[30] Francisco G. Nobrega,[31] Marina P. Nobrega,[31] Sueli M. Oba-Shinjo,[27] Márika K. Oliveira,[18] Guilherme M. Orabona,[32] Audrey Y. Otsuka,[33] Maria L. Paço-Larson,[18] Beatriz M.C. Paixão,[7] Jose R.C. Pandolfi,[34] Maria I.M.C. Pardini,[19] Maria R. Passos Bueno,[32] Geraldo A.S. Passos,[35] Joao B. Pesquero,[36] Juliana G. Pessoa,[36] Paula Rahal,[37] Cláudia A. Rainho,[38] Caroline P. Reis,[28] Tatiana I. Ricca,[14] Vanderlei Rodrigues,[39] Silvia R. Rogatto,[38] Camila M. Romano,[23] Janaína G. Romeiro,[37] Antonio Rossi,[39] Renata G. Sá,[39] Magaly M. Sales,[19] Simone C. Sant'Anna.[24] Patrícia L. Santarosa.[40] Fernando Segato,[25] Wilson A. Silva Jr.,[7,25] Ismael D.C.G. Silva,[33] Neusa P. Silva,[12] Andrea Soares-Costa,[20] Maria F. Sonati,[41] Bryan E. Strauss,[42] Eloiza H. Tajara,[37] Sandro R. Valentini,[34] Fabiola E. Villanova,[33] Laura S. Ward,[40] and Dalila L. Zanette[7]


SOURCE VIEW ZOOM FULL END TISSUE
PreDICinside/TFinside
4000
6000
8000
10000
12000
14000
16000
18000
20000
AA433988 - 100% 100% 95% 95% 100%
OVARY
ORESTES
CGAP
MGC
TFI
ALL
CANCEL
TFU0171

We report the results of a transcript finishing initiative, undertaken for the purpose of identifying and characterizing novel human transcripts, in which RT-PCR was used to bridge gaps between paired EST clusters, mapped against the genomic sequence. Each pair of EST clusters selected for experimental validation was designated a transcript finishing unit (TFU). A total of 489 TFUs were selected for validation, and an overall efficiency of 43.1% was achieved. We generated a total of 59,975 bp of transcribed sequences organized into 432 exons, contributing to the definition of the structure of 211 human transcripts. The structure of several transcripts reported here was confirmed during the course of this project, through the generation of their corresponding full-length cDNA sequences. Nevertheless, for 21% of the validated TFUs, a full-length cDNA sequence is not yet available in public databases, and the structure of 69.2% of these TFUs was not correctly predicted by computer programs. The TF strategy provides a significant contribution to the definition of the complete catalog of human genes and transcripts, because it appears to be particularly useful for identification of low abundance transcripts expressed in a restricted set of tissues as well as for the delineation of gene boundaries and alternatively spliced isoforms.

18066–18071 | PNAS | November 28, 2006 | vol. 103 | no. 48

Characterization of a cancer/testis (CT) antigen gene family capable of eliciting humoral response in cancer patients

Raphael B. Parmigiani*, Fabiana Bettoni*, Maria D. Vibranovski*†, Marilene H. Lopes*, Waleska K. Martins*, Isabela W. Cunha‡, Fernando A. Soares‡, Andrew J. G. Simpson§, Sandro J. de Souza*, and Anamaria A. Camargo*¶

*Ludwig Institute for Cancer Research, SP 01509-010, São Paulo, Brazil; †Biochemistry Department, University of São Paulo, SP 05508-900, São Paulo, Brazil; ‡Pathology Department, Hospital A. C. Camargo, São Paulo, SP 01509-010, Brazil; and §Ludwig Institute for Cancer Research, Memorial Sloan–Kettering Cancer Center, New York, NY 10158

Communicated by Lloyd J. Old, Ludwig Institute for Cancer Research, New York, NY, October 6, 2006 (received for review August 9, 2006)

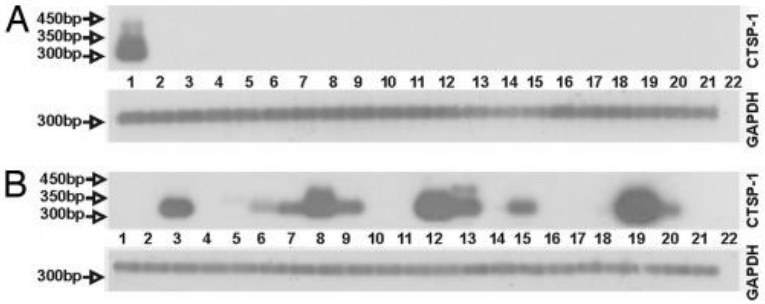


Fig. 4. CTSP-1 mRNA expression pattern in normal tissues and tumor cell lines. Southern blot of RT-PCR products amplified with CTSP-1-specific primers. (*A*) Normal cDNA samples used were as follows. Lanes: 1, testis; 2, lung; 3, prostate; 4, small intestine; 5, breast; 6, brain; 7, heart; 8, uterus; 9, bone marrow; 10, placenta; 11, colon; 12, fetal brain; 13, liver; 14, fetal liver; 15, thymus; 16, salivary gland; 17, spinal cord; 18, kidney; 19, spleen; 20, skeletal

Table 1. Frequency of mRNA expression of CTSP family members in tumor samples

Tumor	CTSP-1 (%)		CTSP-2 (%)		CTSP-4 (%)	
Breast	9/25	(36.0)	1/6	(16.5)	0/6	(0.0)
Colon	7/18	(39.0)	0/14	(0.0)	0/14	(0.0)
Esophagus	2/5	(40.0)	2/4	(50.0)	0/4	(0.0)
Glioblastoma	6/13	(46.0)	Not done		Not done	
Lung	8/14	(57.0)	2/11	(18.8)	0/11	(0.0)
Melanoma	10/17	(59.0)	1/9	(11.0)	0/9	(0.0)
Prostate	14/24	(58.0)	8/17	(47.0)	0/17	(0.0)
Stomach	4/9	(44.0)	1/3	(33.3)	0/3	(0.0)
Thyroid	7/24	(29.0)	0/6	(0.0)	0/6	(0.0)
Uterus	8/20	(40.0)	4/14	(28.5)	0/14	(0.0)
Total	75/169	(44.4)	19/84	(22.6)	0/84	(0.0)

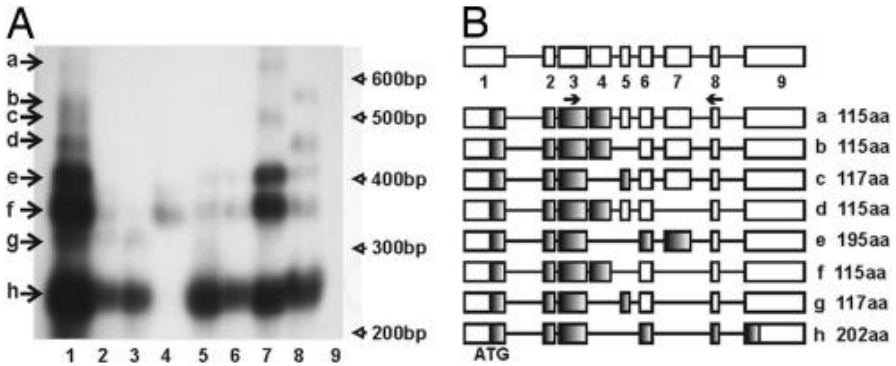


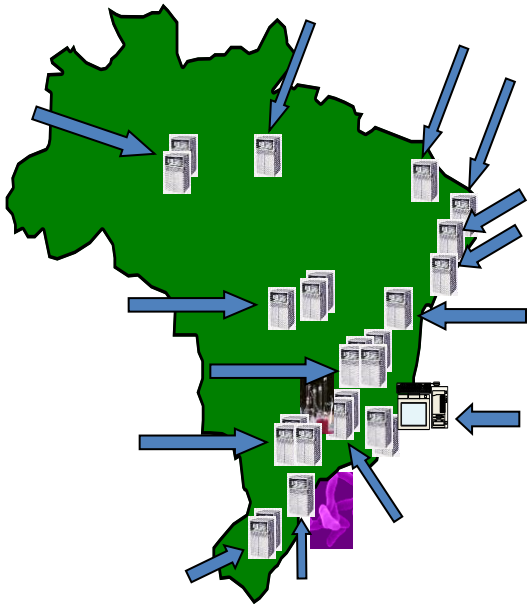
Table 2. Frequency of anti-CTSP-1 antibodies in plasma samples from cancer patients

Tumor	Antibody response (%)	
Breast	3/18	(16.6)
Colon	0/20	(0.0)
Esophagus	0/4	(0.0)
Lung	1/13	(8.0)
Melanoma	1/22	(4.5)
Prostate	5/24	(20.8)
Stomach	1/8	(12.5)
Thyroid	2/10	(20.0)
Uterus	1/22	(4.5)
Total	14/141	(10.0)

The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability

PNAS 100, 11660-11665, 2003

Brazilian National Genome Project Consortium*



*Brazilian National Genome Project Consortium: Ana Tereza Ribeiro de Vasconcelos[a], Darcy F. de Almeida[b], Mariangela Hungria[c], Claudia Teixeira Guimarães[d], Regina Vasconcellos Antônio[e], Francisco Cunha Almeida[f], Luiz G. P. de Almeida[g], Rosana de Almeida[a], José Antonio Alves-Gomes[h], Elizabeth Mazoni Andrade[a], Julia Araripe[i], Magnólia Fernandes Florêncio de Araújo[k], Spartaco Astolfi-Filho[l], Vasco Azevedo[c], Alessandra Jorge Baptista[m], Luiz Artur Mendes Batista[n], Jacqueline da Silva Batista[b], André Belé[o], Cássio van den Berg[p], Maurício Bogo[q], Sandro Bonatto[r], Juliano Bordignon[s], Marcelo Macedo Brigido[m], Cristiana Alves Brito[t], Marcelo Brocchi[u], Hélio Almeida Burity[v], Anamaria Aranha Camargo[w], Divina das Dores de Paula Cardoso[m], Newton Portilho Carneiro[d], Dirce Maria Carraro[x], Cláudia Márcia Benedetto Carvalho[r], Júlio César de Mattos Cascardo[y], Benildo Sousa Cavada[z], Ligia Maria O. Chaves[e], Tânia Beatriz Creczynski-Pasa[a], Nivaldo Costa da Cunha-Junior[c], Nelson Fagundes[r], Clarissa Lima Falcão[d], Fabiana Fantinatti[aa], Ioni Pires Farias[r], Maria Sueli Soares Felipe[m], Lilian Pereira Ferrari[g], Jesus Aparecido Ferro[bb], Maria Inês Tiraboschi Ferro[bb], Gloria Regina Franco[t], Nana Suzy Aguiar de Freitas[cc], Luiz Roberto Furlan[dd], Ricardo Tostes Gazzinelli[e], Eliane Aparecida Gomes[d], Pablo Rodrigues Gonçalves[r], Thalles Barbosa Grangeiro[e], Dario Grattapaglia[ee], Edmundo Carlos Grisard[s], Ebert Seixas Hanna[g], Silvia Neto Jardim[g], Jomar Laurino[r], Lélia Cristina Tenório Leoi[a], Lucymara Fassarella Agnez Lima[ff], Maria de Fátima Loureiro[c], Maria do Carmo Catanho Pereira de Lyra[gg], Humberto Maciel França Madeira[e], Gilson Paulo Manfio[hh], Andrea Queiroz Maranhão[ii], Wellington Santos Martins[a], Sônia Marli Zingaretti di Mauro[bb], Silvia Regina Batistuzzo de Medeiros[dd], Rosely de Vasconcellos Meissner[k], Miguel Angelo Martins Moreira[t], Fabrício Ferreira do Nascimento[t], Maria Fabiana Nicolás[c], Jaquelline Germano Oliveira[t], Sergio Costa Oliveira[t], Roger Ferreira Cury Paixão[a], Juliana Alves Parente[a], Fabio de Oliveira Pedrosa[jj], Sergio Danilo Junho Pena[t], José Odair Pereira[kk], Maristela Pereira[a], Luciana Santos Rodrigues Costa Pinto[a], Luciano da Silva Pinto[a], Jorge Ivan Rebelo Porto[s],


6G11+7C10+7A9+10G6+5A4
MP RU IC JJ IL/II
3H11+3E1+7B11+7E5+3C3+7B7+7G8+7C1
CB/EQ JJ CE RC IL/BF UV JE EZ
8G8+7H3+2D9+5E12+10B10+5B8
JJ/RC QR QR QR CB MR
2A11
EZ
9A12+ 5B3+ 7A11+7A2+8A4+6D6+11A7+11A8+11A9+7B4+7C7+11D3+6D10+9H12+11A4
QR JE MI CC PF MC MC IL CE AC BF CM QR QR AC
11A2 + 5F5+10H5+ 3H12+ 7C9
CC UT CC EP MP
7F2
IC
11B8+9C12+11H5
QV II EZ/EQ
4A4 +1G6 +9E10+ 9D10 + 7G2+ 11A5+ 7B1+ 7B12+ 7B5+ 11A10+ 1H9+ 4A9+ 1A1+ 3C12 +7C12
QR UT /CE BG/MC EQ/CB JE RF EQ CB II IL QR QV PM MC AG
9A10+9H11+5G2+ 11A3
QR JJ IU RF
1G4 + 7B9 + 7A12+ 3C11+ 8G12+ 9D11
BG PM MR UV UV AC/BF
11G6+ 7C6
EP/CM BG
1E1+ 2D3+ 7G7+ 2F10+ 6C10+ 7B9 + 8A2+
AG/IU IU IC JJ AG JJ JJ
7A7+ 7H4+ 3D3+ 7B3+ 7B10 + 6C6
UT QR QV QV RP RU/RP
9G12 + 7B2 +7C4 +6A8 +7A1 +6F5 +7A8
PF PF JE QH CC IL RC
7A10+2E6+7A4+10F3+3A12
QH QH UI EQ/CB IC
+8A3+ 9E9+5G3+8F1+2G4+ 8D10+ 7H2 + 6H3
QV JJ JE JJ EZ MI QS QV/UT
2G12+1A3+19H11
QS/QH UV JJ

Research Report

PCR-Assisted Contig Extension: Stepwise Strategy for Bacterial Genome Closure

BioTechniques 34:626-632 (March 2003)

D.M. Carraro, A.A. Camargo, A.C.M. Salim, M. Grivet[1], A.T. Vasconcelos[2], and A.J.G. Simpson

Ludwig Institute for Cancer Research, SP, [1]CETUC - PUC, RJ, and [2]LNCC/MCT, RJ, Brazil

INTRODUCTION

Bacterial genome projects typically accrue most sequence data through high-throughput shotgun sequencing. The resulting shotgun draft is then converted into a complete representation of the organism's DNA through a more laborious completion phase. This phase is devoted mainly to the generation of additional sequence data to close gaps

that could, in principle, have been exploited for such a strategy. The majority of these have been utilized for the isolation of 5′ flanking genomic regions of cDNA clones to clone insertion-tagged genes and to isolate the extremities of the inserts of large clones such as those in YACs (6–13). However, none of these methods have been previously evaluated as a strategy for bacterial genome closure. The majority

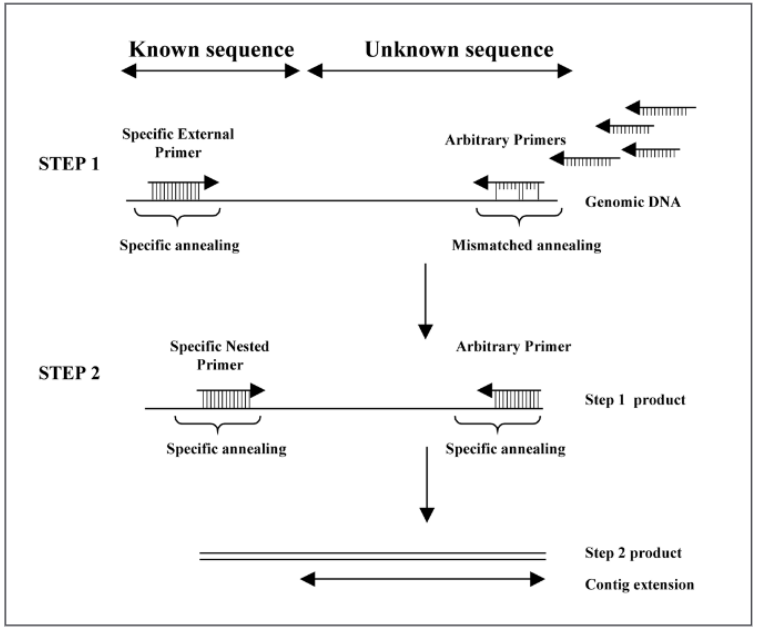


Figure 1. A schematic outline of the PACE reaction.

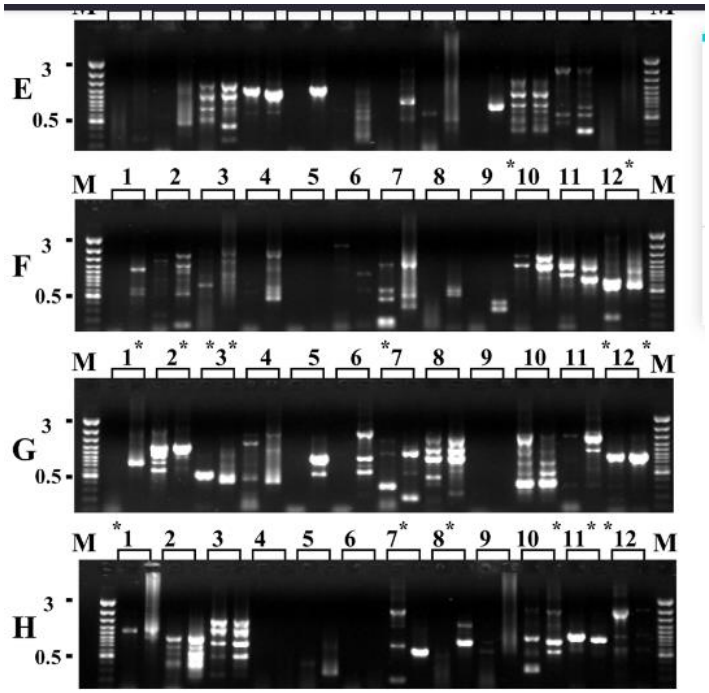


Table 1. Summary of Results Obtained with PACE in Extending *C. violaceum* Shotgun Contigs

	Amplification[b]		Single Products[c]		Sequences[d]		Specific Products[e]		Largest Single
Contig[a]	step 1	step 2	step 1	step 2	step 1	step 2	step 1	step 2	Product[f] (kb)
1	24	48	6	10	1	4	1	4	2.1
2	20	72	8	29	3	17	3	17	2.5
3	27	69	6	12	0	2	0	2	1.5
4	34	76	12	27	0	4	0	4	2.5
5	23	40	1	7	0	3	0	3	1.5
6	43	71	16	25	4	17	4	17	2
7	14	58	2	15	0	8	0	7	1.6
8	30	46	7	19	0	4	0	4	1.7
9	32	54	11	19	4	3	0	2	1.5
10	32	72	15	21	2	7	0	7	1.3
11	40	56	3	17	0	7	0	7	3
12	33	45	10	22	3	8	3	8	1.5
13	55	59	14	20	1	8	0	8	1.8
14	39	56	10	11	2	0	0	0	2.1
15	26	44	9	23	0	13	0	13	1.9
16	46	55	6	14	0	4	0	4	1.3
17	53	49	11	9	0	3	0	3	2.6
18	24	49	8	13	0	1	0	1	1.5
19	15	56	1	15	0	3	0	3	1.7
20	36	53	6	11	0	8	0	7	1.8
21	28	38	3	10	0	4	0	4	1.4
22	14	34	6	17	1	1	0	1	3
Total	688	1200	171	366	21	129	11	126	

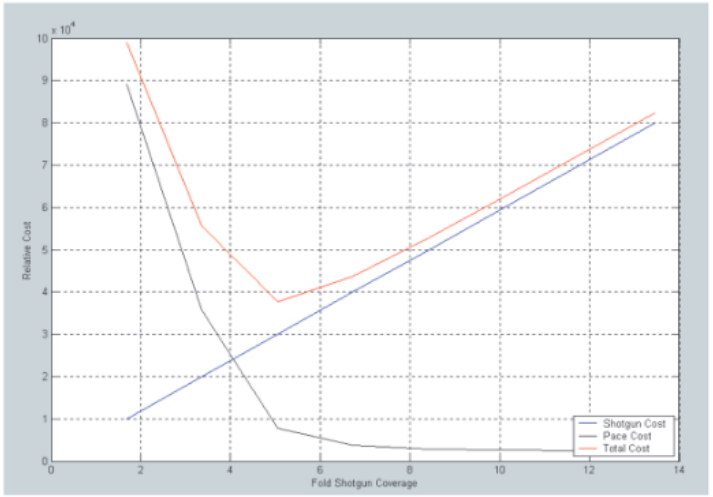


Figure 3. Relative overall cost of a complete 5-Mb genome sequence generated using PACE lowing different levels of shotgun coverage.

Published in final edited form as:
Nat Biotechnol. 2006 December ; 24(12): 1573–1580. doi:10.1038/nbt1256.

The genome and transcriptomes of the anti-tumor agent *Clostridium novyi-NT*

Chetan Bettegowda[1,3,4], Xin Huang[1,4], Jimmy Lin[1,4], Ian Cheong[1], Manu Kohli[1], Stephen A Szabo[1], Xiaosong Zhang[1], Luis A Diaz Jr[1], Victor E Velculescu[1], Giovanni Parmigiani[2], Kenneth W Kinzler[1], Bert Vogelstein[1], Shibin Zhou[1]

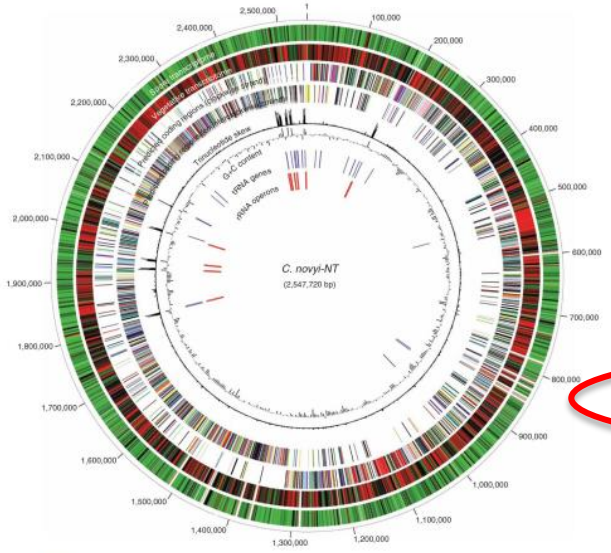


Figure 1.
Circular representation of *C. novyi-NT* genome and transcriptomes. Predicted CDS are color coded based on functional classification. CDS in the transcriptome circles are also color coded, with red representing the highest mRNA abundance, green the lowest and white the intergenic regions. The G+C content and trinucleotide skew were determined using 2,000-nt windows with 1,000-nt incremental shifts.

bacterium from a primary salt crystal. Nature 407, 897–900 (2000). [PubMed: 11057666]

41. Ewing B, Hillier L, Wendl MC & Green P Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8, 175–185 (1998). [PubMed: 9521921]
42. Ewing B & Green P Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194 (1998). [PubMed: 9521922]
43. Gordon D, Abajian C & Green P Consed: a graphical tool for sequence finishing. Genome Res. 8, 195–202 (1998). [PubMed: 9521923]
44. Carraro DM et al. PCR-assisted contig extension: stepwise strategy for bacterial genome closure. Biotechniques 34, 626–628, 630–622 (2003). [PubMed: 12661167]
45. Seshadri R et al. Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. Science 307, 105–108 (2005). [PubMed: 15637277]
46. El-Deiry WS et al. WAF1, a potential mediator of p53 tumor suppression. Cell 75, 817–825 (1993). [PubMed: 8242752]
47. Nuwaysir EF et al. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. Genome Res. 12, 1749–1755 (2002). [PubMed: 12421762]

nature

www.nature.com/nature

Vol 466 | Issue no. 7304 | 15 July 2010

Brazil's biotech boom

Ten years ago, Brazilian bioscience was transformed by a bold initiative. Scientists and the government must develop and extend the progress that has resulted.

In May 1997, a pair of Brazilian scientists spent a weekend in the country discussing a bold idea. José Fernando Perez, the science director at the São Paulo Research Foundation (FAPESP), a state-funded agency and one of Brazil's leading research sponsors, had been looking for game-changing research initiatives. Biologist Fernando Reinach, one of his advisers, had a sufficiently adventurous plan: kick-start biotechnology research throughout Brazil by sequencing a genome.

For many risk-averse scientists in the old guard, who were acutely aware of how far the country lagged behind the rest of the world in biotechnology, this plan seemed overly ambitious. But the duo pushed ahead to build the capacity for genomics and bioinformatics that Brazil lacked, quickly organizing a team to conduct the project and then settling on a bacterium to sequence. FAPESP invested the equivalent of US$12 million, largely dedicated to sequencers, computers and reagents, while the team brought together and trained researchers from a range of fields to develop a broad and long-lasting set of skills and knowledge.

On 13 July 2000 that effort paid off when the team, by then comprising more than 100 researchers in 35 Brazilian labs, published the genetic code for the citrus pathogen *Xylella fastidiosa* in an article featured on the cover of *Nature* (A. J. G. Simpson *et al. Nature* 406, 151–157; 2000). Ten years later, the fruits of that project keep coming.

Before its *Xylella* paper had even come out, for example, the network was busy sequencing another citrus pathogen while taking its first stab at the complex sugarcane genome and contributing to the

Brazilian biotechnology has matured to the point at which its scientists are players on the international stage. And FAPESP is still promoting big ideas, including a new programme to pump money into a broad portfolio of bioenergy research even as the Ministry of Science and Technology constructs a bioethanol research centre; both initiatives seek to build on Brazil's lead in this field. FAPESP is also working to overcome one of the biggest impediments to progress — a lack of doctoral researchers — by encouraging scientists to fill the gaps with young stars from the United States and Europe, part of a broader effort to internationalize Brazilian science.

All of this is good, but more efforts are needed in the same vein — more attitude, more risk-taking and more entrepreneurialism that puts public science into private practice, an area in which Brazil continues to lag. Universities and funding agencies must continue to advance technology-transfer programmes, and the government must streamline regulations that slow even simple activities such as purchasing scientific equipment from abroad. But if there is anything holding Brazil back, it is the same unjustified fear of failure that the country overcame ten years ago with *Xylella*. Although institutions can promote, fund and reward bold thinking, it is worth noting that *Xylella* was not simply a bricks-and-mortar research centre run by a foundation, but a science project. Ultimately the task of promoting Brazilian biotechnology comes down to the science, and it will be up to individual scientists to accept the challenge and expand their research horizons.

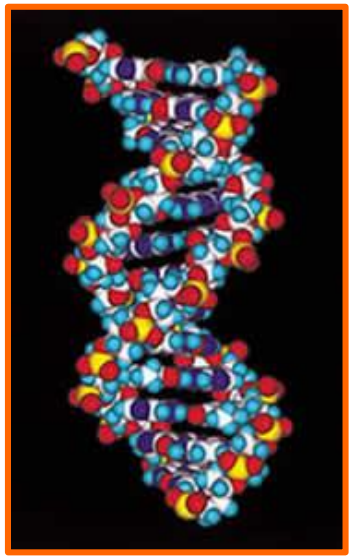
orygen®
biotecnologia

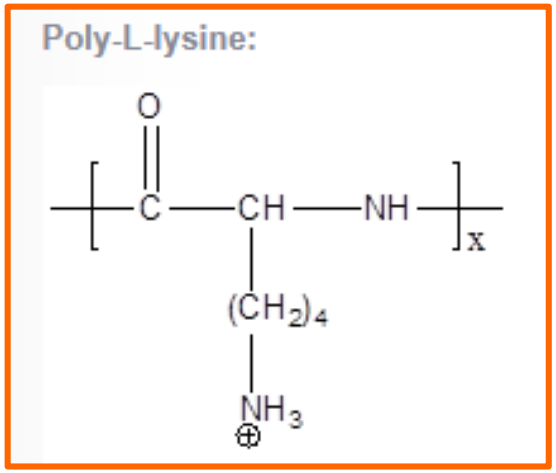
biolab
FARMACÊUTICA

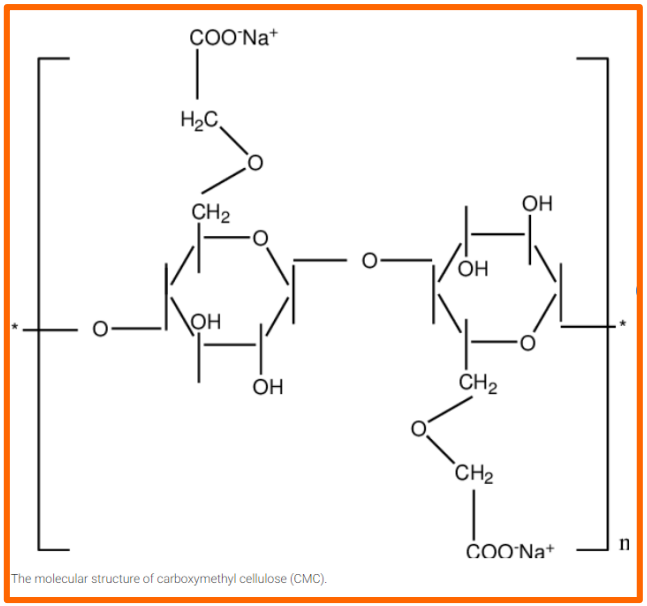
eurofarma
sua vida move a nossa

Poly-ICLC (Hiltonol®)

Polyriboinosinic acid – polyribocytidylic acid + poly-lysine + CMC







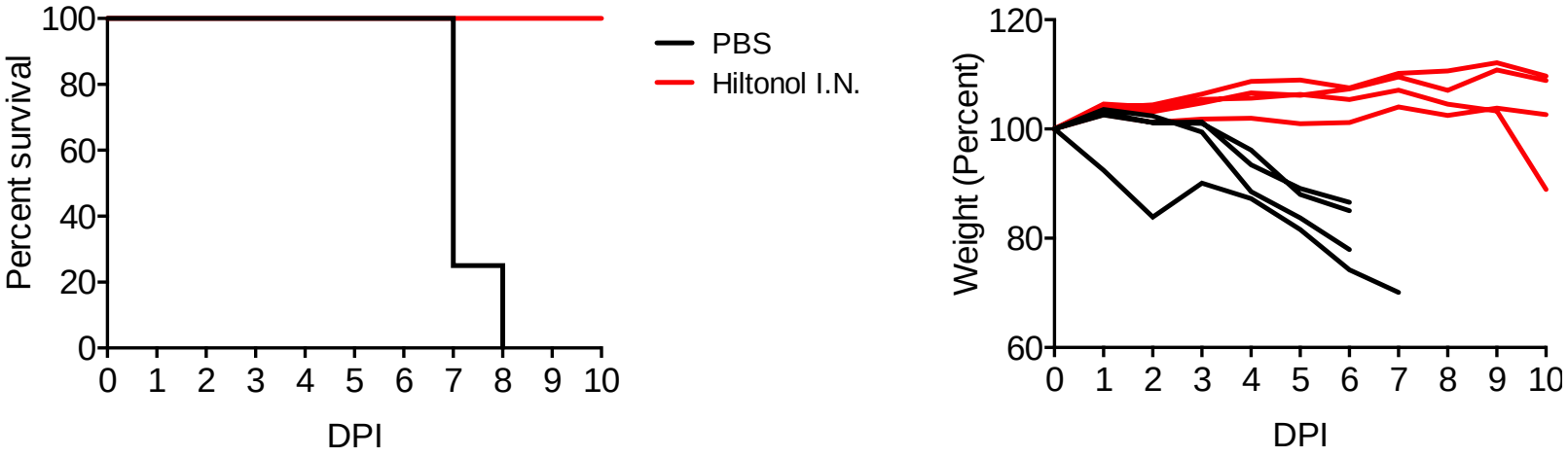
poly-IC + poly-lysine + carboxymethylcellulose

Poly-ICLC is a stabilized form of double stranded RNA that mimics a viral genome and acts as a potent and rapid activator of the imune response to be used either alone or with an antigen to direct the immune response to tumors or infectious agents being developed by **Oncovir.**

Supplied as a 1.8mg/ml sterile solution in saline in vials containing 1.2ml

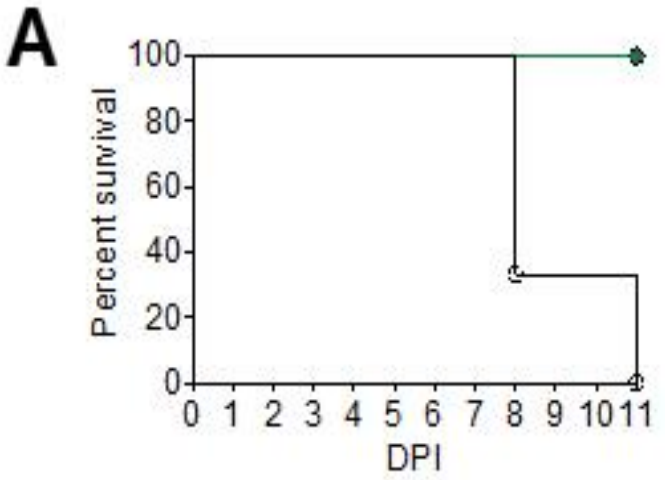
Proteção contra COVID

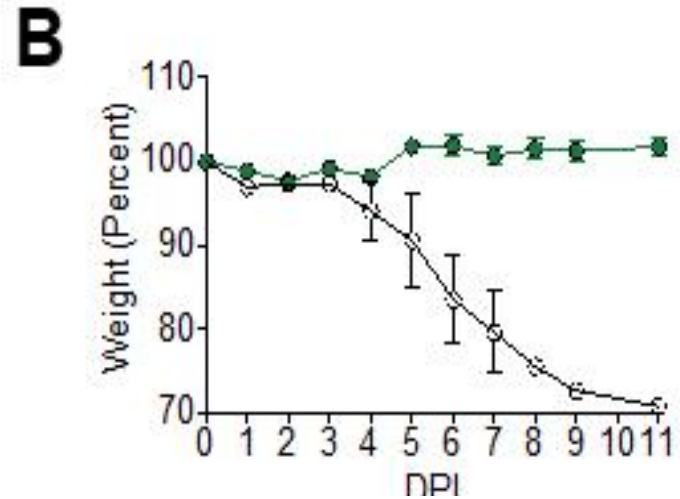
Mice: 50ug (28uL) Hiltonol i.n. (days -4 and -3)

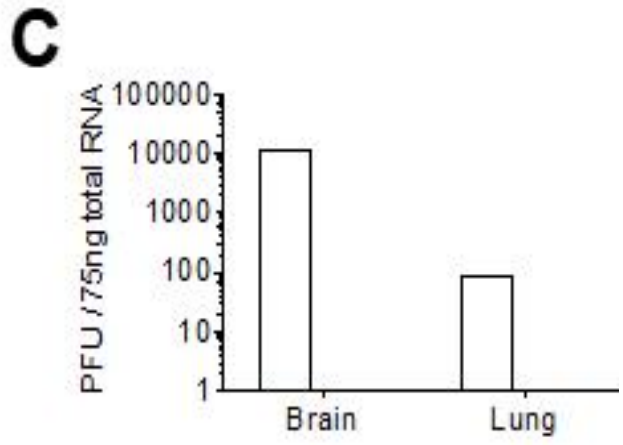


UFMG

Vacina contra COVID







nature communications

Article

https://doi.org/10.1038/s41467-022-32547-y

Promotion of neutralizing antibody-independent immunity to wild-type and SARS-CoV-2 variants of concern using an RBD-Nucleocapsid fusion protein

Received: 15 April 2022

Accepted: 5 August 2022

Published online: 17 August 2022

Check for updates

Julia T. Castro[1,2,3], Patrick Azevedo[1,2], Marcílio J. Fumagalli[3], Natalia S. Hojo-Souza[1,2], Natalia Salazar[1], Gregório G. Almeida[2], Livia I. Oliveira[2,4], Lídia Faustino[2], Lis R. Antonelli[2], Tomas G. Marçal[2], Marconi Augusto[4], Bruno Valiate[2], Alex Fiorini[1], Bruna Rattis[3], Simone G. Ramos[3], Mariela Piccin[3], Osvaldo Campos Nonato[3], Luciana Benevides[3], Rubens Magalhães[1], Bruno Cassaro[1], Gabriela Burle[1,2], Daniel Doro[1,2], Jorge Kalil[5], Edson Durigon[6], Andrés Salazar[7], Otávia Caballero[7], Helton Santiago[1,8], Alexandre Machado[1,2], João S. Silva[3], Flávio da Fonseca[1,8], Ana Paula Fernandes[1,9], Santuza R. Teixeira[1,8] & Ricardo T. Gazzinelli[1,2,3,10]

LUDWIG INSTITUTE FOR CANCER RESEARCH