

RAMANA V. DAVULURI

420 W 12th Avenue, TMRF 524
1@medctr.osu.edu

Email: Davuluri-

Human Cancer Genetics Program
Department of Mol. Virology, Immunology and Medical Genetics
College of Medicine and Public Health
The Ohio State University
Columbus, OH 43210

Tel: 614-688-3088
Fax: 614-247-6842

Experience

- | | |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2001-present | Assistant Professor, Bioinformatics & Computational Biology, Human Cancer Genetics Program, Department of Mol. Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH 43210, USA |
| 2001 | Research Investigator, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA |
| 1999–2001 | Computational Post Doctoral Fellow, Zhang Lab, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA |
| 1998-1999 | Special Project Scientist (Bioinformatics), International Crops Research Institute for the Semi Arid Tropics, (ICRISAT), Patancheru, AP, INDIA and Visiting Scientist, Avesthagengraine Technologies, Bangalore, INDIA |
| 1998 | Short term Post Doctoral Fellow (Bioinformatics), Dept. of Plant Genetics, VIB, University of Ghent, Gent, BELGIUM. |
| 1993-1998 | Asst. Professor of Statistics, ANGR Agricultural University, Hyderabad, INDIA |

Education

- | | |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| February 1996 | PhD, Indian Agricultural Statistics Research Institute, IARI, New Delhi, India <ul style="list-style-type: none">• <u>Major</u>: Agricultural Statistics and <u>Minor</u>: Computer Applications Advisor: V.K.Gupta, PhD |
| 1991 | MSc, Indian Agricultural Statistics Research Institute, IARI, New Delhi, India <ul style="list-style-type: none">• <u>Major</u>: Agricultural Statistics and <u>Minor</u>: Computer Applications Advisor: V.K.Gupta, PhD |
| 1988 | BSc (Mathematics), Nagarjuna University, Guntur, India <ul style="list-style-type: none">• Major: Mathematics; Minors: Physics and Chemistry |

Honors and Awards

- | | |
|---------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1997 | Young Scientist Award (Statistics) merit certificate- 1996-97 by <i>Indian Science Congress Association</i> (ISCA), during the 84 th annual session at Delhi, India. |
| 1991-93 | IASRI (Indian Council of Agricultural Research) Senior research fellowship |
| 1989-91 | IASRI (Indian Council of Agricultural Research) Junior research fellowship |
| 1993 | One of the three finalists for ISAS Young Scientist Award - Indian Society of Agricultural Statistics (ISAS) for the year 1993. |
| 1991-93 | VVR Murthy award for the best student of MS (Ag.Stat.) 1989-91 batch |
| 1989 | Sri CVS Narasimha Rao endowment scholarship awarded by SRR & CVR Government College, Vijayawada (India) |
| 1983-85 | State special merit scholarship awarded by AP state government, India |
| 1983 | Honorary medal awarded by Board of Secondary Education, AP, India, for securing state 7 th rank in Secondary School Certificate examination. |

Bioinformatics software developed

FirssEF: Program to predict First Exons (transcription start site to first donor site) and promoter regions in the human genome

JTEF: Program to predict 3' terminal exons and polyA signals in the human DNA

Publications

Tabaska, E. Jack, Ramana V. Davuluri and Michael Q. Zhang (2001). A novel 3'-Terminal exon recognition program for Human DNA Sequences. *Bioinformatics* **17**: 602-607.

Ramana V. Davuluri, Yutaka Suzuki, Sumio Sugano and Michael Q. Zhang (2000). CART classification of human 5'UTR sequences, *Genome Research* **10**: 1817-1827.

Nathali Pavy, Stephan Rombauts, Patrice Dehais, Catherine Mathe, Davuluri V. Ramana, Philippe Leroy and Pierre Rouze (1999). Evaluation of gene prediction software using a genomic dataset: Application to Arabidopsis Thaliana sequences. *Bioinformatics* **15**: 887-899.

Gupta, V.K., Ramana, D.V.V. and Parsad, R. (1999). Weighted A-efficiency of block designs for making treatment-control and treatment-treatment comparisons. *J. of Statistical Planning and Inference*, **77**:301-319.

Balakrishna Murthy, K. Krishnaiah, P.V. Thirumala Rao, K. and Ramana, D.V.V. (1996) Efficacy and economics of insecticide management strategies against fruit borers of bhendi (*Abelmoschus esculentus* L. moench), *New Agriculturist*, **7(1)**.

Gupta, V.K., Ramana, D.V.V. (1993) Robustness of designs against a single missing observation. *Int. J. Math. Statist. Sci.* **2**: 45-56.

Papers Under Review/Accepted

Ramana V. Davuluri, Ivo Grosse and Michael Q. Zhang. Computational identification of first exons and promoters in human genome. Accepted (Nature Genetics).

Invited Talks

Computational identification of first exons and promoters in human genome. *Genome Sequencing & Biology Meeting*, May 9-13, 2001, Cold Spring Harbor, NY.

Weighted A-optimality and Efficiency of block designs for comparisons between two sets of treatments. 84th *Indian Science Congress*, Delhi University, Delhi, January 3-8, 1997 for the ISCA Young Scientist Award.

Robustness of designs against a single missing observation. 45th *Annual Conference of the Indian Society of Agricultural Statistics* for the ISAS Young Scientist Award, 1993.

Published Abstracts

CART classification of human 5'UTR sequences. *Computational Genomics Conference* organized by The Institute of Genomic Research at Baltimore, MD, November 16-19, 2000.

Statistical characterization of human 5'UTR sequences. *Genome Sequencing & Biology Meeting*, Cold Spring Harbor, NY, May 10-14, 2000.

Research Projects (ongoing/completed during the last 3 years) – A summary

Computational identification of first exons and promoters in human genome – 2000-2001:

- Identification of promoters and first exons has been the most difficult problem in gene-finding. We developed a set of discriminant functions that can recognize structural and compositional features such as CpG islands, promoter regions, and first donor sites. We explain the implementation of the discriminant functions into a decision tree that constitutes a new program called FirstEF. We used different models to predict CpG-related and non-CpG-related first exons. We showed, by cross-validation, that the program could predict 86% of the first exons with 17%

false positives. We also demonstrated its prediction accuracy on genome level by applying it to the finished sequences of human chromosomes 21 and 22, and by comparing the predictions with the experimentally verified first exons. Finally, we annotated all the first exons of chromosomes 21 and 22 as a pilot project towards the annotation of the whole human genome.

- During the course of this project, I developed different Perl programs to BLAST thousands of sequences and parse the output automatically. The BLAST output parser program reformats the BLAST output into an easily readable form, which will be used by other Perl programs to identify exact genomic matches for exons and introns. I also developed different Perl programs to calculate various sequence scoring functions that are being used as feature variables in building the discriminant models. I developed S (Splus) functions to build different discriminant models and clustering procedures.

Classification of human promoter sequences – 2000-present:

- Taking the promoter regions of the ~1500 human genes (from the first-exon database) I intend to classify them into tissue-specific or functional categories, by using the micro-array expression data. I am trying to discover and model putative transcriptional regulatory modules. This classification will also facilitate the identification of composite modules of transcription factor binding-sites that occur together more frequently than others. This helps us to understand and discover the gene regulation code.
- During this project, I developed Perl programs and UNIX shell scripts that can identify most frequent hexamers (with one mis-match) that are over-represented in human promoter regions (500 nt upstream region of the transcription start site). Comparing the observed frequency with the expected frequency of the hexamers these programs identify the over-represented hexamers. The expected frequencies were calculated based on a random sample of 247,648 sequences, each of length 500 nt. This random sample was retrieved from the human genome working draft (<http://genome.ucsc.edu>) by systematic sampling procedure.

Statistical characterization of 5'UTRs of Human – 1999-2000:

- A non-redundant database of 2312 full-length human 5'UTRs was carefully prepared. A comprehensive computational analysis of this data was conducted for characterizing the 5'UTR features. CART analysis was used to classify the data into three distinct classes. Class I consists of mRNAs that are believed to be poorly translated with long 5'UTRs filled with potential inhibitory features. Class II consists of TOP (terminal oligopyrimidine tract) mRNAs that are regulated in growth dependent manner and Class III consists of mRNAs with favorable 5'UTR features that may help efficient translation. The most accurate tree we found has 92.5% classification accuracy as estimated by cross validation. The classification model included presence of TOP, secondary structure, 5'UTR length and presence of upstream AUGs (uAUGs) as the most relevant variables. The present classification and characterization of the 5'UTRs provide precious information for better understanding the translational regulation of human mRNAs. Furthermore, this database and classification can help people to build better computational models for predicting the 5'terminal-exon and separating the 5'UTR from the coding region.

Developing a human 3'-Terminal exon prediction program – 1999-2000:

- A new program, called JTEF, for finding 3' terminal exons in human DNA sequences was developed. This program was based on quadratic discriminant analysis, a standard non-linear statistical pattern recognition method. The quadratic discriminant functions used for building the algorithm were trained on a set of 3' terminal exons of type *3tuexon* (those containing the true STOP codon). We showed that the average predictive accuracy of JTEF is higher than the presently available best programs (GenScan and Genemark.hmm) based on a test set of 65 human DNA sequences with 121 genes. In particular JTEF performs well on larger genomic contigs containing multiple genes and significant amounts of intergenic DNA. It will become a valuable tool for genome annotation and gene functional studies.

Evaluation of Gene prediction programs for Arabidopsis Thaliana – 1998-1999:

- While working on this collaborative project, I developed statistical measures and programmed in C on SOLARIS to compare the accuracy of different gene-prediction programs GENSCAN, GRAIL, GENMARK, and MZEF etc.

Teaching Experience – A summary

Conducted Training in Computers and Bioinformatics:

- Organized informal courses and delivered seminars to train scientists and research associates in the Molecular Biology Unit of ICRISAT on use of genome databases. Trained Statistics Unit staff to enhance their understanding of new statistical techniques for analysis of molecular data.
- Conducted a training program to the faculty of Agricultural College, Bapatla, Acharya N.G. Ranga Agricultural University in “Computer Literacy” (DOS, WINDOWS, Spread Sheets, Databases, Word Processors and Statistical Packages).

University Teaching: 5 Years: 1993-1998:

- Taught undergraduate and graduate courses in Applied Statistics and Computer Applications to students of Agriculture, Engineering and Home Science at Agricultural College, ANGR Agricultural University, Bapatla, India. Courses taught include *Statistics, Introduction to Statistics, Statistical Methods, Sampling Techniques, Data Processing and Programming Languages – BASIC & FORTRAN and Introduction to Computer Applications.*

Training courses attended

“Computational Genomics” at Cold Spring Harbor Laboratory, Nov. 4-9, 1999

“Overview of the Wisconsin Package” by The Genetics Computer Group, Oct. 27-29, 1999

“Pedagogy Training Program” by Extension Education Institute, Hyderabad of ANGR Agricultural University under AHRD program, Mar. 4-12, 1998

Committees

1994-1998 Member, Graduate Student Advisory Committee, ANGR Agrl University

Professional Memberships

Member of the *International Society for Computational Biology*

Member of the *American Statistical Association*

Life Member of the *Indian Science Congress Association*

Member of the *Indian Society of Agricultural Statistics*