# Harry J. Mangalam, PhD

1 Whistler Ct, Irvine, CA, USA, 92617

949 824 0084 (o), 949 285 4487 (c), [hjm@tacgi.com](mailto:hjm@tacgi.com)

## *Personal Information:*

Age: 52, Naturalized Canadian Citizen, Permanent Resident of the US, married to a US Citizen, 2 children, 1 dog.  Speaks and writes English well, French barely.

## *Employment History:*

| Period | Title |
|---|---|
| 5.2006-present | Research Computing Specialist/Programmer, Office of Information Technology, UC Irvine |
| 2.2005 -5.2006 | Research Associate, Earth System Science, UC Irvine |
| 7.2002-2.2005 | Principal, tacg Informatics LLC, a bioinformatics and software consulting firm, Irvine, California |
| 1.2002-7.2002 | Sr. Bioinformatics Scientist, Acero Inc |
| 8.2001-12.2001 | Principal, tacg Informatics LLC |
| 2000-7.2001 | Senior Research Scientist & Project Manager , Gene Expression, National Center for Genome Resources, Santa Fe, NM |
| 1999-2000 | Principal, tacg Informatics LLC |
| 1996-1998 | Senior Analyst for Research Computing, Developmental and Cell Biology, School of Biological Sciences, University of California, Irvine |
| 1993-1996 | Director of Biocomputing, College of Medicine and Associate Specialist, Microbiology and Molecular Genetics, University of California, Irvine |
| 1991-1993 | Sequence Analysis Specialist for the Dept of Biocomputing at the Salk Institute |

## *Education:*

**PostDoc:** Medical Research Council of Canada Post-Doctoral Fellowship, with John B. Thomas of the Molecular NeuroBiology Laboratory, Salk Institute, La Jolla, 1991.  Genetics and biochemistry of the Drosophila *singleminded* gene and its gene products.

**Ph.D.:** with Michael G. Rosenfeld, HHMI. Physiology and Pharmacology, University of California at San Diego, La Jolla, Dec. 1989. Transcriptional Regulation of the Growth Hormone and Prolactin genes; Cloning and characterization of Pit-1, one of the first POU-homeo genes.

**M.Sc.:** with David R. Jones, Zoology, University of British Columbia, Vancouver, May, 1985. Physiology and pharmacology involved in the diving response in mammals and birds.

**B.Sc.:** Zoology, University of British Columbia, Vancouver, May. 1980.  Honours thesis with  David R. Jones.

## *Work Experience:*

**OIT:** I work with researchers to help them accelerate their work in large scale visualization, bioinformatics, evolutionary biology, high-throughput sequencing, large scale data processing, compute cluster planning & implementation, system administration, data center planning, and grant preparation. Often this requires finding software to support their needs or knitting together multiple pieces of software into a pipeline or complex script, usually addressed with Perl, Python, and R/Bioconductor. I also assist in general infrastructure problems such as campus-wide software distribution, storage, backup, & network problems, outreach to faculty to identify and resolve bottlenecks, representation to campus and UC-wide discussion & planning groups.

**Earth System Science:** I worked with Charlie Zender of Earth System Science at UC Irvine on the tuning and analysis of the NetCDF Operators (NCOs), a suite of utilities for subsetting and manipulating climate model data in the form of NetCDF formatted files. These data sets are among the largest in the world and the NCO programs are responsible for a significant amount of the data reduction used in analysing them. Because of this dependence on the NCOs, their efficiency is of high interest by the climate-modeling community. In working to optimize these utilities, I analyzed the execution and data via a number of techniques, including oprofile (pure software), the HPCToolKit/PAPI (which access CPU hardware counters), and Valgrind, a synthetic CPU (see publication below). This includes evaluation and testing of storage and network infrastructures including file system types, hardware and software RAID implementations,. I was responsible for automating the configure, build, and test processes on a number of platforms both in serial and parallel modes. The now-merged NetCDF and HDF formats provide exquisitely compact and efficient data storage that bioinformatics applications have largely ignored, altho the [BioHDF project](#) may reverse that.

**tacg Informatics:** I have done bioinformatics contract work for the Epidemic Outbreak Surveillance taskforce (now part of the Homeland Security Department), GeneCodes, the CDC, Allergan, and Accelerys. (see also tacg, below)

**Acero:** I worked in the Science group on their Genomics Knowledge Platform (GKP), which provided both syntactic and semantic integration of biological information from a number of sources through their "Biological Object Model". I was involved with a number of projects including managing academic relationships, presentations to and consultations with customers (chiefly pharmaceutical IT heads), working with developers to implement requested features, reviewing external applications for integration with the GKP, competitive analysis of various 'Bio-IT' industry sectors, mapping data to the BOM, and helping to develop loading strategies for various biological data types.

**NCGR**: I worked on the GeneX gene expression database project, an Open Source Gene Expression database for storing and analyzing results from large scale gene expression projects. It is described in more detail in the IBM Systems Journal paper cited below. My contribution to the project was the original conception and implementation while at UC Irvine, and later scientific lead and project manager at NCGR. I also co-wrote the NSF grant that was funded to Dr. Jennifer Weller, now at UNC, Chapel Hill.

**tacg:** I wrote the Open Source sequence analysis application **tacg** and it's CGI Web interface **tacgi** to make a small, fast (~30X faster than GCG or EMBOSS), free, and capable molecular biology tool available for Linux/Unix. It has been incorporated into a number of other packages, including UCSD's Biologist's WorkBench, University of Cambridge Pathology server, Korean Rice Genome server, the Japanese Institute for Virus Reseach, EMBNet's SRSNET server, the Pasteur Institute, ANGIS (Australia) and University of Manitoba's BIRCH System, and Curagen's CuraTools, Entigen's

BioNavigator suite, Genops Ngene and others (See: http://tacg.sf.net and in the BMC Bioinformatics paper cited below.

These efforts reflect my interests in data flow and databases, algorithm development for sequence analysis, with a side interest in the sociology and implementation of Open Source software. I'm always interested in learning new approaches and try to find jobs that allow me to contribute meaningfully to good science while learning data and programming techniques from different fields. The interesting areas of a field are always at its borders.

## *Technical Skills*

I can program in ANSI C, Perl, Python, bash, Postscript and the statistical language R/BioConductor for genomic and expression analysis. I've worked with web technologies Joomla, Plone, MediaWiki, Turbogears, Trac, as well as implementing and administering them on Apache2. I use the GNU toolchain (autoconf/automake, gcc, gdb, and the ddd visual debugger GUI. Besides GDB, I've also used the dmalloc memory debugging tool to track recalcitrant memory problems, as well as Valgrind, oprofile, and HPCToolkit/PAPI (see nco benchmarking paper below). I've used both SWIG and numpy/f2py to 'Pythonize' Fortran and C.

Examples of my technical writing outside of journal publications, are available here:
  – The BDUC Cluster Users HOWTO
  – Manipulating Data on Linux
  – How to transfer large amounts of data via network
  – Mind your NegaBIT$
  – UCI OSS Backup Evaluation
  – How to Evaluate Open Source Software
  – The Storage Brick - Fast, Cheap, Reliable Terabytes

## *Databases/Applications*

I have experience in large scale gene expression systems, including some coding work with Affymetrix's expression formats. In my work with the GeneX project, I have examined and reviewed a number of data formats, including many pure XML (GeneXML, BIOML) and mixed XML/binary formats (XDF), as well as looking into the more standard (for large data sets) netCDF & HDF formats. I've installed and administered databases primarily using PostgreSQL, but also WAIS, Isite, Glimpse, MySQL, PostgreSQL, SQLite, and various applications using the biological databases Genbank, Flybase, Protein Data Bank, TFD, TRANSFAC, EPD, ACeDB, advanced user/sysadmin use of BLAST, BLOCKS, FASTA, and PROFILES. I've used many of the molecular biology and sequence analysis programs and have reviewed some of them for publication (see below).

## *Awards:*

Co-PI (with Drs. Carl Cotman, Ken Longmuir, Tatsuya Suda, David Walker, all of UC Irvine) of a Pacific Bell CalREN grant: Medical Informatics Use of Asynchronous Transfer Mode (ATM) service across the LA basin (June '94-September '97)

Co-PI (with Dr. Thomas Cesario) of Irvine Health Foundation Grant to implement ATM-based telemedicine applications using the infrastructure of the above CalREN grant (July, 1995 - September, 1997).

Medical Research Council of Canada Post-doctoral Fellowship to study Drosophila genetics with Dr. John Thomas at the Salk Institute (1990-1991).

## *Selected Reviewed Publications:*

Zender, C.S. and Mangalam, H. (2007) *Scaling Properties of Common Statistical Operators for Gridded Datasets*, Int'l J. High Perf. Comp Appl.; 21; 485 (online: http://hpc.sagepub.com/cgi/content/abstract/21/4/485 )

Mangalam, H.J. (2002) *tacg, a grep for DNA*.  BMC Bioinformatics: 3:8 (online: http://www.biomedcentral.com/1471-2105/3/8 ).

Mangalam, H.J., Stewart, J.E., Zhou, J., Schlauch, K.A, Waugh, M.E., Chen, G., Farmer, A.D., Colello, G.D., and Weller, J.W. (2001). *GeneX: an Open Source gene expression database and integrated toolset.* IBM Systems J. 40(2):552-569. (online: http://www.research.ibm.com/journals/sj/402/mangalam.html )

Long, A.D, Mangalam, H.J., Chan, R.Y.P., Tolleri, L, Hatfield, G.W., and Baldi, P. (2001). *Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework*. J. Biological Chem. 276(23):19937-44.

Kliche, S., Stitz, L., Mangalam, H., Shi, L., Binz, T., Niemann, H., Briese, T., Lipkin, I.  (1996). *Characterization of the Borna Disease Virus Phosphoprotein, p23*. J. Virology, 70(11):8133-8137.

Mangalam, H.J., V.R. Albert, H.A. Ingraham, M.S. Kapiloff, L. Wilson, C. Nelson, M.G. Rosenfeld. (1989).  *A pituitary POU domain protein, Pit-1, transcriptionally activates both growth hormone and prolactin promoters.*  Genes Dev, 3:946-958

Ingraham, H.A., R. Chen, H.J. Mangalam, H.P. Elsholtz, S.E. Flynn, C. Lin, D.M. Simmons, L. Swanson, M.G. Rosenfeld. (1988).  *A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype.*  Cell, 55:519-529.

Elsholtz, H.P., H.J. Mangalam, E. Potter, V.R. Albert, S. Supowit, R.M. Evans, and M.G. Rosenfeld. (1986).  *Two different cis-active elements transfer the transcriptional effects of both EGF and phorbol esters.*  Science, 234:1552-1557.

## *Reviews and Commentary*

Mangalam, H.J. (2002)  *The Bio\* Toolkits - a brief overview.*  Briefings in Bioinformatics. 3(3):296-302.
Stewart, J.E., Mangalam, H.J., and Zhou, J.  (2001) *Open Source Software meets Gene Expression.* Briefings in Bioinformatics. 2(4):319-328.
Mangalam, H.J. (1999) *Set my software (mostly) free; the pros and cons of free software.*

Biotechnology Software. 16(4):5-7.

Mangalam, H. (1996). *Sculpting proteins with a virtual chisel.* TIBS, 21(10):398-399.

Mangalam, H.J.. (1993). *Macintosh DNA Analysis Software: Stridin' the turf with the Gang of Four* (Commentary). TIBS, 18(5):187-188.

Ingraham, H.A., V.R. Albert, R.P. Chen, E.B. Crenshaw III, H.P. Elsholtz, X. He, M.S. Kapiloff, H.J. Mangalam, L.W. Swanson, M.N. Treacy, M.G. Rosenfeld. (1990). *A family of POU-domain and Pit-1 tissue-specific transcription factors in pituitary and neuroendocrine development.* Ann. Rev. Physiol. 52:773-791.

Rosenfeld, M.G., C.K. Glass, S. Adler, E.B. Crenshaw III, X. He, S.A. Lira, H.P. Elsholtz, H.J. Mangalam, J.M. Holloway, C. Nelson, V.R. Albert, and H.A. Ingraham. (1988). *Response and binding elements for ligand-dependent positive transcription factors intergrate positive and negative regulation of gene expression.* Cold Spring Harbour Symposia on Quantitative Biology.

Or by Google Scholar: <http://goo.gl/iKWj>

You're going to Google me, so how about trying these:
harry mangalam uci
harry mangalam tacg
harry mangalam nco
harry mangalam bioinformatics

## ***References available on request.***