

INTRODUCTION

Open Access

Proceedings of the ninth annual UT-ORNL-KBRIN Bioinformatics Summit 2010

Eric C Rouchka^{1*}, Robert M Flight^{1,2}, Claire Rinehart³

From UT-ORNL-KBRIN Bioinformatics Summit 2010
Cadiz, KY, USA. 19-21 March 2010

The University of Tennessee (UT), the Oak Ridge National Laboratory (ORNL), and the Kentucky Biomedical Research Infrastructure Network (KBRIN), have collaborated over the past decade to share research and educational expertise in bioinformatics. One result of this collaboration is the joint sponsorship of an annual regional summit to bring together researchers, educators and students who are interested in bioinformatics from a variety of research and educational institutions. This summit provides unique opportunities for collaboration and forging links between members of the various institutions. This year, the Ninth Annual UT-ORNL-KBRIN Bioinformatics Summit was held at Lake Barkley State Park in western Kentucky, from March 19 to 21, 2010. A total of 269 participants pre-registered for the summit, with 150 from various Tennessee institutions and 93 from various Kentucky institutions. A number of additional participants came from universities and research institutions from other states and countries, e. g. Emory University, University of British Columbia, University of Cincinnati, Iowa State University, etc. Eighty-one registrants were faculty, with an additional 94 students, 63 staff, and 31 postdoctoral participants.

The conference program consisted of three days of presentations. The first day included a pre-summit of talks by researchers from Kentucky institutions, two workshops on the use of Bioconductor for DNA microarray data analysis, and a session on Next Generation Sequencing. The next two days were dedicated to scientific presentations divided into two primary sessions on Epigenetics and Medical Informatics.

Pre-Summit Kentucky bioinformatics session

Dr. Nigel Cooper, the project director of KBRIN, started the pre-summit Kentucky Bioinformatics session with a

summary of Bioinformatics in the state of Kentucky. Dr. Cooper reviewed the history, development and current state of bioinformatics programs in Kentucky institutions, as well as plans for the future. The presentation highlighted the many sources of funding providing support for the various bioinformatics initiatives, the undergraduate, graduate, and post-graduate programs currently underway, as well as physical resources available to bioinformatics researchers in Kentucky. Dr. Cooper also emphasized the need to have institutional support in the grant application process both for continuing grant support and in applying for larger grants to allow bioinformatics research to move forward in the future.

The remaining talks of the pre-summit session highlighted current bioinformatics research from faculty at various Kentucky institutions. The subjects covered included sequence analysis, next generation sequencing, transcriptomics, proteomics and metabolomics.

In the first research talk, Dr. Eric Rouchka of the University of Louisville, used sequence analysis of the 5' untranslated region (UTR) of 18000 mouse transcripts to determine those that may be translationally regulated by CPEB1 [1]. Regulation of gene expression at the translational level is important because it allows the cell to create a pool of mRNA transcript that can be quickly translated to protein without the need to start transcription. This work required a model of the CPEB1 binding site, and a scoring method that considered how many binding sites the 5' UTR contained in addition to how well the sites matched the consensus sequence. Of the 18000 transcripts examined, 1200 were found with a high score and three binding sites. Ontological analysis of those genes determined a subset that appears to be functionally related.

The next talk, by Dr. Arnold Stromberg from Kentucky University, highlighted a rather unusual situation

* Correspondence: eric.rouchka@louisville.edu

¹Department of Computer Engineering and Computer Science, University of Louisville, Duthie Center for Engineering, Louisville, KY 40292, USA

in DNA microarray analysis, having only one sample available for each treatment under investigation. In DNA microarray studies, with few exceptions, current practice is to have multiple samples for each treatment under investigation. However, in this study multiple samples were not possible. To answer the biological questions being considered, namely which samples were most similar and which genes show differential expression between two of the samples, Dr. Stromberg examined the various comparisons between pairs of samples, as well as the expression patterns across all four samples. Two important considerations emerged in this analysis, that of false positives, and the determination of genes that are truly “unexpressed”.

Dr. Ted Kalbfleisch from the University of Louisville presented his work on assessing the prevalence of the LINE-1 retrotransposon using Next Generation Sequence (NGS) data. The long interspersed nuclear element 1 (LINE-1) is a repeat sequence that has been found pervasively throughout the genome. The LINE-1 element is important because it has the ability to reinsert copies of itself into the genome, thereby causing changes in gene regulation, overall expression, or generating splice isoforms of a gene. This action has been previously documented as the causative agent in some diseases. However, the true prevalence of LINE-1 in the human genome is unknown, and until the advent of NGS was very difficult to assess. However, even with the availability of NGS data, determining the frequency of this element in the genome is difficult due to its repeatability, the very thing that is normally removed from consideration during sequence assembly. Using short portions of the known LINE-1 sequence, Dr. Kalbfleisch discovered at least 22 insertions that are currently being verified in the lab. Future work includes developing an assay to determine a comprehensive LINE-1 insertion profile, as well as the screening of large populations to fully characterize the variation present in the human population.

Dr. Susmita Datta presented her work regarding an improved automated method for detecting monoisotopic peaks in MS data. Automated peak detection in MS data analysis is important because of the continued use of various MS methods in both metabolomics and proteomics, however the amount of data generated in any given MS run is overwhelming to be able to analyze by hand, and current peak detection methods mistakenly identify peaks due to many different factors, including the presence of molecules with overlapping mass profiles. To allow detection of the peaks, Dr. Datta modelled the isotopic distribution of the peptides using a mixture of location shifted Poisson distributions, and an EM algorithm to determine the weighting parameter of the distributions. Statistical analysis and comparison

with the LIMPIC method showed improvement in the number of detected peaks and their identification.

Dr. Ryan Gill from the University of Louisville presented his work on the determination of differences in gene networks between two biological samples. Traditional DNA microarray analysis focuses on generating lists of differentially expressed genes; however there may be greater significance in identifying how gene association networks change between conditions. To determine if networks have changed between conditions, however; one must be able to statistically differentiate between two networks. Dr. Gill focused on the methods used to determine differential connectivity of a single gene or class of genes, and differences in modular structure [2]. Particularly important in this work was the determination of the sensitivity of the differential score to the network construction parameters, as these will influence the final networks obtained from the underlying data. Using data from an experiment designed to determine the underlying genetic changes between fat and lean mice, Dr. Gill demonstrated the results and effectiveness of the differential measure developed.

Dr. Xiang Zhang of the University of Louisville presented work on developing a metabolic pipeline geared towards solving three challenges in metabolomics bioinformatics: metabolite identification, metabolite quantification, and the visualization of metabolite networks. In contrast to many other methods, Dr. Zhang’s pipeline uses an *in-silico* method to predict the gas chromatographic retention times of the possible metabolites. This is followed by a peak alignment method that considers both dimensions of separation and the spectra of the molecules to allow peak alignment over different temperature ranges in the GC column. Using a combination of significance scores from different statistical tests, those molecules that are regulated by the process under investigation are then determined. The inter-molecular correlations can then be visualized using the SysNet [3] software. This workflow has provided Dr. Zhang and colleagues with many interesting insights into various biological problems.

The final talk of the pre-summit session was presented by Dr. Guy Brock on the biological impact of missing value imputation on the downstream analysis of DNA microarray data [4]. DNA microarrays frequently have missing expression values, and a common solution to this problem is to impute values using different methods. Although many different studies have examined the problem of missing value imputation, this study was unique in that it examined the ability of the imputation methods to recreate the original data, performance measures used to determine which method should be used, impact of the methods on the down-stream analysis using three different biological impact measures.

Bioconductor workshops

Dr. Guy Brock began the official summit program with two workshops on the use of R and Bioconductor [5] for bioinformatics data analysis. The first workshop started by providing a basic introduction to the R environment, as well as an overview of the philosophy and capabilities of R and the Bioconductor suite. This was followed by an introduction to the ExpressionSet data structure and methods used to interact with the underlying data. Using the ALL expression data set available in Bioconductor, attendees were walked through a typical data analysis workflow to determine which genes are differentially expressed in a DNA microarray experiment.

The second workshop continued by introducing some of the tools available to the researcher who has a list of differentially expressed genes and is searching for biological commonalities among them. These include examining gene annotations such as gene names and the Gene Ontology. Dr. Brock also reviewed the use of *clValid* [6], a recently released R package for cluster validation, which can be used to determine the validity of gene clustering results based on various measures. Finally, the use of Gene Set Enrichment Analysis (GSEA) to determine the presence of over-represented biological functions or pathways was reviewed.

Next generation sequencing

The Next Generation Sequencing (NGS) session was kicked off by Dr. Ted Kalbfleisch from the University of Louisville with a thorough introduction to this rapidly developing field. Dr. Kalbfleisch reviewed DNA sequencing methods from the early days of Maxam-Gilbert sequencing through the current slate of Next Generation sequencers such as the 454 and SoLiD platforms. He also reviewed the various software tools used to perform alignment and raw assembly with commentary on some of the pros and cons associated with each method, in addition to the many applications for which NGS is being used. His presentation ended with a discussion of a particular examination of genetic variation of the Line-1 retrotransposon from NGS data that is publicly available. This project required aligning known Line-1 insertion patterns to whole genome NGS data for identification of novel insertion sites.

Dr. David Sexton of Vanderbilt University followed up with an examination of the resources required to manage the massive amounts of data produced by NGS based upon his experience as the Director of the Computational Genomics Core in the Center for Human Genetics Research [7]. Dr. Sexton continually emphasized the need to be able to handle the incredible volumes of data that these sequencers generate, in terms of storage, backup, transport, and analysis. For every

consideration, he described the currently implemented solutions at Vanderbilt, giving attendees a glimpse into the day-to-day operations of the NGS center.

The last talk of the day was by Dr. Steven Jones from the University of British Columbia. Dr. Jones focussed on the use of NGS to characterize individual mutations in cancer cells, and thereby offer tailored therapies, or individualized genomics. Using NGS, they were able to determine which gene pathways were modified in individual cancers, and suggest appropriate treatments for the individual. This stands in marked contrast to the previously accepted method of isolating one particular strain of cancer cells and treating it as a representative for other cancers in the same bodily tissue [8]. One particularly interesting example involved one individual cancer where the mutations were characterized over time in response to treatment, with examination of the various gene pathways that were disrupted as the cancer underwent further mutation in response to the treatments.

Epigenetics

The second day of talks began with the session on Epigenetics. Dr. James Cheverud of Washington University presented his research on epigenetic sources of variation in mice obesity. This work involved the crossing of very large (overweight, LG) females with very small (lean, SM) males, and correlating the resultant phenotypes with various loci, and determining how the interactions between loci lead to a particular phenotype [9-11]. Particular emphasis was placed on examining combinations of gene-effects to provide more explanatory power than examining single gene effects in isolation.

Dr. Rosanna Weksberg from the Hospital for Sick Children at the University of Toronto presented research on the genetic imprinting effect in Beckwith-Wiedemann syndrome (BWS) [12,13] which has both genetic causes and epigenetic phenotypes. Two groups of genes have been shown to have changed methylation patterns in connection with BWS, leading to changes in their expression. Dr. Weksberg also discussed Prader Willi Syndrome, and Russell-Silver Syndrome, which are additional human imprinting disorders associated with growth.

Rounding out the epigenetics session was Dr. Robert Lane from the University of Utah Department of Pediatrics and his examination of genetic imprinting via histone methylation as a mechanism of adaptation. It has been known for some time that the methylation state of genes affects their expression, however there are still many questions regarding the causes and effects of particular histone methylation patterns. The primary system examined by Dr. Lane was intrauterine growth restriction

(IUGR), a condition of restricted growth in the embryo that leads to altered expression of insulin-like growth factor 1 (IGF-1) in the adult, with subsequent effects on growth and development [14,15]. A rat model of IUGR combined with examination of the methylation patterns of the histones found reproducible changes to the methylation of particular histone residues that were also gene position dependent. This resulted in a change of the ratio of particular IGF-1 transcripts, thereby resulting in altered phenotypes. Similar examination of pups of hyperglycemic rats (simulating diabetic pregnancies) revealed analogous changes in methylation of the histones. Incredibly, they were able to prevent the phenotypic changes and reprogram the methylation via supplementation with essential nutrients, thereby re-establishing IGF-1 expression to control levels.

Following the poster session and evening buffet, there was an overview of Computable Genomix's Gene Indexer program. Gene Indexer is an automated latent semantic indexing engine that extracts both explicit and implicit relationships from the literature. It is efficient and uses an investigator's questions, not predetermined pathways or ontologies, to interrogate the scientific literature and determines the functional cohesiveness of any groups of genes.

Dr. Robert Hettich of the Oak Ridge National Laboratory presented work on the use of -omics to examine human gut microbial communities and their involvement in Crohn's disease [16,17]. This work required examining the metaproteomes and metagenomes, or set of all proteins and genes from a heterogeneous population of bacteria in twins where one, both or neither suffered from Crohn's disease (CD) or Ulcerative Colitis (UC). Dr. Hettich's group specifically performed the proteomics work to identify as many expressed proteins as possible in the stool samples based on the gene sequences that had been previously determined. They were able to identify changes in the composition of the microbiome and the metaproteome that correlated with the presence or absence of disease, providing rich data sets that will require extensive follow-up.

Day two's keynote address was given by Dr. Mike Hawrylycz of the Allen Institute for Brain Science on the informatics of large scale digital aliasing in neuroscience [18,19]. As an example, Dr. Hawrylycz provided an in-depth examination of the information required to create the Developing Mouse Brain Atlas. This resource combines 2 D slices of brain tissue, 3 D models of the brain changing during development [20], with *in-situ* hybridization gene expression levels for ~2000 genes. This publicly available resource allows anyone to examine the expression levels of these genes over the developmental stages and determine those genes that have correlated expression either temporally or spatially. A variety of

analysis tools and metrics are built into the Atlas, providing a wealth of information regarding gene expression in the brain to the interested researcher.

Medical informatics

Only one talk was presented in the Medical Informatics session, by Dr. Myriam Fornage from the University of Texas at Houston, on the search for genes responsible for vascular diseases in the brain. Although studies have shown that by far the greatest indicators for vascular disease are age, sex, and high blood pressure, it may be that particular genetic factors predispose individuals to ischemic injuries in the brain. This has been borne out by previous studies that indicate many individuals suffer from pre-clinical lesions, and that true ischemic stroke tends to be "the tip of the iceberg", in that it is a clinical symptom of a long underlying process. Dr. Fornage reported on a series of genome wide association studies performed to try and determine genes that may make it more likely that an individual will suffer from brain vascular disease (BVD). A variety of SNP's have been identified and confirmed in follow up studies. Two genes in particular were found to be associated with a higher risk of BVD, *Ninj2* and *Wnk1* [21]. *Ninj2* is a cell surface adhesion molecule that is induced after nerve injury, and is theorized to affect how the brain tolerates/recovers from ischemic insults. *Wnk1* is a kinase that has been previously associated with hypertension, and is theorized to promote injury to the brain. Using a mouse ischemia model, the expression of both genes was followed over time following injury to the brain, and the expression of both was shown to be concordant with their known roles. Although GWAS analyses are identifying new genetic variants involved in BVD, the effect of many of them is small, thereby requiring large sample sizes to allow their detection. A question that does result though, is how small an effect is too small to be clinically relevant and not worthy of further investigation or drug development? In addition, there is the potential for more work on examining gene-gene interactions, as it is likely that variants found in conjunction will have a larger effect.

Posters & short talks

During the two-hour poster session on Saturday afternoon, fifty-two posters (eighteen more than the previous year) were presented. The abstracts were divided into the general groupings of Bioimaging, Bioinformatics of Health and Disease, Bioinformatics Infrastructure, Comparative Genomics, Databases, Functional Genomics, Gene Regulation, Genome Annotation, Genomics, Machine Learning/Algorithms, Microarrays, Neuroscience, Proteomics, Sequence Analysis, and Structure and Function Prediction.

Two sessions (one each on Saturday and Sunday) featured presentations from the poster abstracts. They included “Inferring Gene Coexpression Networks for Low Dose Ionizing Radiation using Graph Theoretical Algorithms and Systems Genetics” (Gary L. Rogers), “Integrating metagenomics with metaproteomics for characterization of the molecular activities of the human distal gut microbiome in healthy and Crohn’s Disease” (Alison Erickson), “Discovering Disease-specific Biomarker Genes for Cancer Diagnosis and Prognosis” (Zhongming Zhao), “Analysis of equine protein-coding gene structure and expression by RNA-sequencing” (Steven Coleman), “A systematic pathway-based analysis of GWAS data revealed susceptibility pathways to schizophrenia” (Peilin Jia), “Association of Genomewide Newborn DNA Methylation Patterns with Maternal Diet, Birth Weight and SNP Variation” (Ronald Adkins), “High-throughput sequencing of the DBA/2J mouse genome” (Rob Williams), and “Subsystems-based servers for rapid annotation of genomes and metagenomes” (Rami Aziz).

Future plans

The 2011 Bioinformatics summit will return to the state of Tennessee in the spring of 2011. Potential focus areas include current technological trends in molecular biology, applications of next-generation sequencing, and systems biology. The format will likely be expanded to include more local talks and hands-on workshops at introductory and advanced levels.

Acknowledgements

We would like to thank the additional Conference Program Committee members Nigel Cooper (University of Louisville), Dan Goldowitz (University of British Columbia), Ramin Homayouni (University of Memphis), Julia Krushkal (University of Tennessee-Memphis), Mike Langston (University of Tennessee-Knoxville), Terry Mark-Major (University of Tennessee-Memphis), Cynthia Peterson (University of Tennessee-Knoxville), Arnold Stromberg (University of Kentucky), Rob Williams (University of Tennessee-Memphis) and Zhongming Zhao (Vanderbilt University) for organizing an outstanding scientific program. In addition, we wish to thank Stephanie Dearing, Terry Mark-Major, Michelle Padgett, Donna Schulte and Jane Thornton for all of their efforts in dealing with the conference organization details. We would also like to thank Kerry Allen and the wonderful folks at Lake Barkley State Park, who helped make our stay as pleasant as possible. Funding for the UT-ORNL-KBRIN Summit is provided in part by the Kentucky Biomedical Research Infrastructure Network (KBRIN), University of Tennessee Center for Integrative and Translational Genomics, University of Tennessee Molecular Resource Center, UT-ORNL Science Alliance, and NIH grant P20RR16481.

Author details

¹Department of Computer Engineering and Computer Science, University of Louisville, Duthie Center for Engineering, Louisville, KY 40292, USA.

²Neuroscience Training Program, University of Louisville, Louisville, KY 40292, USA. ³Department of Biology, Western Kentucky University, Bowling Green, KY 42101, USA.

Published: 23 July 2010

References

1. Rouchka EC, Wang W, Graham JH, Cooper NGF: **Computational prediction of genes translationally regulated by cytoplasmic polyadenylation elements.** *Bioinformatics and Computational Biology Berlin/Heidelberg*: Springer 2009, 353-361.
2. Gill R, Datta S, Datta S: **A statistical framework for differential network analysis from microarray data.** *BMC Bioinformatics* 2010, **11**:95.
3. Zhang M, Ouyang Q, Stephenson A, Kane MD, Salt DE, Prabhakar S, Burgner J, Buck C, Zhang X: **Interactive analysis of systems biology molecular expression data.** *BMC Syst Biol* 2008, **2**:23.
4. Brock GN, Shaffer JR, Blakesley RE, Lotz MJ, Tseng GC: **Which missing value imputation method to use in expression profiles: a comparative study and two selection schemes.** *BMC Bioinformatics* 2008, **9**:12.
5. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**:R80.
6. Brock G, Datta S, Pihur V, Datta S: **clValid: An R package for cluster validation.** *Journal of Statistical Software* 2008, **25**:1-22.
7. Richter BG, Sexton DP: **Managing and analyzing next-generation sequence data.** *PLoS Comput Biol* 2009, **5**:e1000369.
8. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S: **Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution.** *Nature* 2009, **461**:809-813.
9. Fawcett GL, Jarvis JP, Roseman CC, Wang B, Wolf JB, Cheverud JM: **Fine-mapping of Obesity-related Quantitative Trait Loci in an F(9/10) Advanced Intercross Line.** *Obesity (Silver Spring)* 2009.
10. Wolf JB, Hager R, Cheverud JM: **Genomic imprinting effects on complex traits: a phenotype-based perspective.** *Epigenetics* 2008, **3**:295-299.
11. Wolf JB, Cheverud JM: **A framework for detecting and characterizing genetic background-dependent imprinting effects.** *Mamm Genome* 2009, **20**:681-698.
12. Weksberg R, Smith AC, Squire J, Sadowski P: **Beckwith-Wiedemann syndrome demonstrates a role for epigenetic control of normal development.** *Hum Mol Genet* 2003, **12**(Spec No 1):R61-R68.
13. Weksberg R, Shuman C, Beckwith JB: **Beckwith-Wiedemann syndrome.** *Eur J Hum Genet* 2010, **18**:8-14.
14. Aagaard-Tillery KM, Grove K, Bishop J, Ke X, Fu Q, McKnight R, Lane RH: **Developmental origins of disease and determinants of chromatin structure: maternal diet modifies the primate fetal epigenome.** *J Mol Endocrinol* 2008, **41**:91-102.
15. Fu Q, Yu X, Callaway CW, Lane RH, McKnight RA: **Epigenetics: intrauterine growth retardation (IGR) modifies the histone code along the rat hepatic IGF-1 gene.** *FASEB J* 2009, **23**:2438-2449.
16. Deneff VJ, Kalnejais LH, Mueller RS, Wilmes P, Baker BJ, Thomas BC, Verberkmoes NC, Hettich RL, Banfield JF: **Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities.** *Proc Natl Acad Sci USA* 2010, **107**:2383-2390.
17. Mahowald MA, Rey FE, Seedorf H, Turnbaugh PJ, Fulton RS, Wollam A, Shah N, Wang C, Magrini V, Wilson RK, Cantarel BL, Coutinho PM, Henrissat B, Crock LW, Russell A, Verberkmoes NC, Hettich RL, Gordon JI: **Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla.** *Proc Natl Acad Sci USA* 2009, **106**:5859-5864.
18. Lau C, Ng L, Thompson C, Pathak S, Kuan L, Jones A, Hawrylycz M: **Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain.** *BMC Bioinformatics* 2008, **9**:153.
19. Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Floss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawai R, Kidney JM, Knapiak RH, Kuan CL, Lake JH, Laramie AR, Larsen KD, Lau C,

- Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova NV, Svisay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN, Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR: **Genome-wide atlas of gene expression in the adult mouse brain.** *Nature* 2007, **445**:168-176.
20. Ng L, Pathak SD, Kuan C, Lau C, Dong H, Sodt A, Dang C, Avants B, Yushkevich P, Gee JC, Haynor D, Lein E, Jones A, Hawrylycz M: **Neuroinformatics for genome-wide 3 D gene expression mapping in the mouse brain.** *IEEE/ACM Trans Comput Biol Bioinform* 2007, **4**:382-393.
21. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, Debette S, Lumley T, Folsom AR, van den Herik EG, Bos MJ, Beiser A, Cushman M, Launer LJ, Shahar E, Struchalin M, Du Y, Glazer NL, Rosamond WD, Rivadeneira F, Kelly-Hayes M, Lopez OL, Coresh J, Hofman A, DeCarli C, Heckbert SR, Koudstaal PJ, Yang Q, Smith NL, Kase CS, Rice K, Haritunians T, Roks G, de Kort PL, Taylor KD, de Lau LM, Oostra BA, Uitterlinden AG, Rotter JI, Boerwinkle E, Psaty BM, Mosley TH, van Duijn CM, Breteler MM, Longstreth WT, Wolf PA: **Genomewide association studies of stroke.** *N Engl J Med* 2009, **360**:1718-1728.

doi:10.1186/1471-2105-11-S4-11

Cite this article as: Rouchka et al.: Proceedings of the ninth annual UT-ORNL-KBRIN Bioinformatics Summit 2010. *BMC Bioinformatics* 2010 **11** (Suppl 4):11.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

