BMC Bioinformatics

Proceedings of the Fourteenth Annual UT- KBRIN Bioinformatics Summit 2015

Eric C Rouchka[1,2*], Julia H Chariker[2,3], Benjamin J Harrison[2,4]

The University of Tennessee (UT) and the Kentucky Biomedical Research Infrastructure Network (KBRIN) have collaborated over the past fourteen years to share research and educational expertise in bioinformatics. One result is an annual regional summit for researchers, educators and students. The Fourteenth Annual UT-KBRIN Bioinformatics Summit was held at Paris Landing State Park in Buchanan, Tennessee from March 20-22, 2015. A total of 200 participants pre-registered, with 115 from Tennessee, 69 from Kentucky, and the remainder from various states and international locales. Among the registrants were 70 faculty, 65 students, 40 staff, and 23 postdocs. The conference program consisted of a workshop on R and three days of presentations broken into plenary sessions: Where are They Now, Medical Informatics, and Epigenetics. In addition, a poster session with 42 posters was held on Saturday evening.

Friday workshops

Danny Arends (Humbolt University in Berlin) began the Summit with the workshop "Using R/qtl MQM to improve QTL mapping in inbred crosses." This workshop covered the use of the free R package R/qtl for mapping and exploring quantitative trait loci (QTL). Included were topics related to Multiple QTL Mapping (MQM), such as handling missing data, using permutations for determining significance thresholds for QTL and QTL hotspots, and visualizations for cis-trans and QTL interaction effects [1,2].

Following this workshop, Dr. Arends presented "CTL Mapping in Inbred Strains." He covered the topic of correlated traits loci (CTL), a novel method used to correlate differences in expression to regions of the genome [3,4]. CTL detects a number of regions not found with traditional QTL mapping, and therefore increases the number of candidate regions in QTL and genome-wide association studies (GWAS).

Session I: where are they now

Dan Goldowitz (UBC), one of the original organizers of the Bioinformatics Summit, opened the scientific session with a Bioinformatics Summit Retrospective. As part of this presentation, Dr. Goldowitz discussed the beginnings of the Bioinformatics Summit as part of the University of Tennessee Health Science Center's (UTHSC) Center for Integrative and Translational Genomics (CITG). Throughout this presentation, discussions were made about where former keynote and plenary speakers as well as attendees were, and the current directions of their research.

Ramin Homayouni (University of Memphis) followed with an engaging presentation "From Junior Faculty to Center Director and Entrepreneur" which he renamed "Hindsight is 20/20". During his presentation, Dr. Homayouni progressed from where he began as a biologist studying the role of individual genes in the nervous system [5-7] to part of the self-proclaimed "Dirty Dozen" at UTHSC which formed the basis of the CITG from which the Bioinformatics Summit originated. Ramin discussed how the Summit has been critical in forming collaborations, leading to the founding of his company, ComputableGenomix, which developed the GeneIndexer software based largely upon text-mining approaches [8-12].

Session II: medical informatics

Hauke Bartsch (UCSD) kicked off the plenary speaker session with "PING: Pediatric Imaging, Neurocognition, and Genetics." The PING Data Resource involves 1,400 students aged 3-21 years during childhood and adolescent

* Correspondence: eric.rouchka@louisville.edu
[1]Department of Computer Engineering and Computer Science, University of Louisville, Duthie Center for Engineering, Louisville, KY 40292, USA
Full list of author information is available at the end of the article

development [13]. The goal of PING is to research what a normal child looks like during development, and measures information including demographic, developing mental and emotional functions, family history, cognitive abilities (attention and memory), multimodal neuroimaging, and genotype information. PING is a longitudinal study, taking measurements as part of a summer camp experience. The goal is to track "normal" development, including changes in cortical thickness over time, hippocampal volume, and fiber tracts. Specific types of data gathered include age, gender, handedness, height, education, family history, body mass index (height and weight), mini mental state (30 questions), genetic "Europeanness" (based on 550,000 single nucleotide polymorphisms), cortical information (surface area, thickness, and volume), sub-cortical volume, white matter volume, and diffusivity (40 tracts). Among the results of PING include a gyrus-based brain atlas [14], a genetically derived brain atlas [15], and a web portal for exploring and visualizing the PING data, which allows for a "Big Data" hypothesis space [16] by including datasets from a total of 21 different projects.

Joshua Denny (Vanderbilt University) followed with "Harnessing routine clinical care to accelerate genomic discovery ... or entering the era of Precision Medicine using EHRs." This presentation began with an overview of the history of personalized medicine, which goes back to Hippocrates [17], and more recently has been touted by President Obama with the announcement of the new Precision Medicine Initiative [18]. Dr. Denny discussed personalized medicine in the context of Vanderbilt's BioVU [19,20], an opt-out DNA Biobank started in 2007 which contains a cohort of over 200,000 samples and over 2 million components, including, among others, clinical notes, clinical messaging, medications, orders, ICD9 codes, tumor registry, and death index information. In terms of genomics, BioVU contains nearly 200,000 subjects with DNA samples, including dense, GWAS-level genotyping. He discussed the use of BioVU for phenome-wide association studies (PheWAS) to discover all phenotypes associated with a particular genotype, a reverse-engineering approach to GWAS studies [21-23]. As a result of large-scale biorepositories, the Electronic Medical Records and Genomics (eMERGE) network began at 10 sites with a goal to perform GWAS using electronic medical record (EMR) derived phenotypes and to initiate implementation of actionable variants into the EMR [24].

Santosh Kumar (University of Memphis) closed out the Medical Informatics Session with "Designing Sensor-triggered Just-in-Time Mobile Health Interventions – Challenges and Opportunities." This presentation centered on the National Institutes of Health (NIH) Big Data to Knowledge (BD2K) initiative. Dr. Kumar serves as the director of the MD2K Center of Excellence – NIH Big Data to Knowledge Mobile Data to Knowledge – Mobile Measurement of Exposure, Risk, and Outcomes at the University of Memphis. During his presentation, Dr. Kumar discussed how recent advances in wearable sensors and mobile technology have allowed for the measurement of dynamic changes in an individual's health state as well as physical, biological, behavioral, social, and environmental factors that contribute to health and disease risk [25,26]. He discussed opportunities in detection, prediction, and adaptation, including data measurement in sensor and platform design; prediction of adverse events on a personalized basis; and treatment options in terms of engaging individuals and suggesting non-adverse adaptations. Technologies that can trigger just-in-time mHealth interventions include smartwatches and chestbands which help to measure electrocardiograms, GSR, respiration, and temperature; smartphones which can measure location via GPS, acceleration, and allow for self-reporting; and other Easy-Sense (contactless) sensors such as eyeglass and autosense sensors. The two key contributions of MD2K are in the area of data science research for the development of sensors, and in developing approaches for knowledge discovery.

Session III: epigenetics

Amelie Baud (European Bioinformatics Institute) led off the final scientific session on Sunday morning with "Leveraging genetic variation in the social partners to study effects of the social environment on phenotypes of biomedical relevance." In this presentation, Dr. Baud expanded upon work that shows phenotypes of individuals are socially affected by the genetics of the individuals in which they interact. The example used considered aggressive behavior and growth rate of pigs relative to the genetics of their pen-mates [27-30]. Based on this study, Dr. Baud's group has studied whether or not the behavior of a lab mouse or rat is affected by the genotypes of its cage-mates. The results indicate that some phenotypes, including anxiety, growth, and wound healing are dependent on cage-mate genotype, but others including general locomotor activity or helplessness, are not. Their research suggests indirect genetic effects on gene expression likely lead to the differences in phenotype expression.

Alicia Smith (Emory University) followed with "Leveraging epigenome-wide data for human studies of complex traits." Dr. Smith opened with an introduction of how epigenetic changes are measured, showing there are many experimental design issues to consider in methylation studies, including identification and control of sources of biological and technical variation that can influence interpretation. Some of the work presented included use of umbilical cord blood samples from the CANDLE project to show that DNA methylation varies with gestational age [31]; an illustration that methylation

changes across the lifespan [32], differs by race [33], is influenced by medication exposure, has sex-specific effects (such as hormonal influence), varies by cell type, and varies by genotype [34]. As a result of all of these influences, Dr. Smith discussed the use of proxy tissues to create a controlled epigenetic environment [35].

Genetic-epigenetic correlations were described in detail, as part of a research effort involving analysis of 222,888 common SNPS and 20,093 CpG sites, excluding methylation probes that contained SNPs [36]. From this study, 2000 meQTLs in each of 7 cohorts were enriched in GWAS studies and miRNA binding sites. Interestingly, 40-70% of meQTLs overlap between African American and Caucasian cohorts, while 44-51% of meQTLs overlap at birth and in adulthood; 19-37% of meQTLs are detected in both blood and brain cohorts; and only 36-72% are common between different brain tissues in the same person.

Michael Kobor (University of British Columbia) closed the scientific session with "Epigenetic variation in human health and disease." He built upon the research presented by Dr. Smith, showing that epigenetic imprints vary widely, and they are largely based on diet, environment, and quality of maternal care. Examples of issues with DNA methylation include a higher correlation within than between tissues [37], showing that DNA methylation is tissue specific. Dr. Kobor pointed out that individual tissue type is most important for determining methylation status, followed by cell type. Dr. Kobor also showed the difficulty in linking epigenetic and disease factors, since DNA methylation and gene expression are not tightly linked across individuals [38] and sites with high variability in methylation are largely not correlated with gene expression. Most of the research Dr. Kobor presented was based on "Early Life Experiences Getting Under the Skin" where he discussed the effect of early life socio-economic status and parental stress level on childhood temperament, behavior, and trajectories [38-40], and the effect of maternal warmth on resilience to a pro-inflammatory phenotype [41].

Future plans

The 2016 UT KBRIN Bioinformatics Summit is scheduled for April 8-10 at Lake Barkley State Park in Cadiz, Kentucky and the 2017 Summit is scheduled for April 21-23 at Montgomery Bell State Park in Tennessee. Sessions will focus more on bioinformatics research in Kentucky and Tennessee to help forge more collaborative efforts. Plans are to include more Tennessee and Kentucky speakers, in terms of intermediate length (20-30 minute) presentations and short (10 minute) presentations that complement plenary session topics.

Authors' details
[1]Department of Computer Engineering and Computer Science, University of Louisville, Duthie Center for Engineering, Louisville, KY 40292, USA. [2]Kentucky Biomedical Research Infrastructure (KBRIN) Bioinformatics Core, 522 East Gray Street, Louisville, KY 40292, USA. [3]Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY 40292, USA. [4]Department of Anatomical Sciences and Neurobiology, University of Louisville, Louisville, KY 40292, USA.

References
1. Arends D, Prins P, Jansen RC, Broman KW: **R/qtl: high-throughput multiple QTL mapping.** *Bioinformatics* 2010, **26(23)**:2990-2992.
2. Durrant C, Swertz MA, Alberts R, Arends D, Moller S, Mott R, Prins P, van der Velde KJ, Jansen RC, Schughart K: **Bioinformatics tools and database resources for systems genetics analysis in mice–a short review and an evaluation of future needs.** *Brief Bioinform* 2012, **13(2)**:135-142.
3. Arends D, van der Velde KJ, Prins P, Broman KW, Moller S, Jansen RC, Swertz MA: **xQTL workbench: a scalable web environment for multi-level QTL analysis.** *Bioinformatics* 2012, **28(7)**:1042-1044.
4. Zych K, Li Y, van der Velde JK, Joosen RV, Ligterink W, Jansen RC, Arends D: **Pheno2Geno - High-throughput generation of genetic markers and maps from molecular phenotypes for crosses between inbred strains.** *BMC Bioinformatics* 2015, **16**:51.
5. Homayouni R, Byrne JH, Eskin A: **Dynamics of protein phosphorylation in sensory neurons of Aplysia.** *J Neurosci* 1995, **15(1 Pt 1)**:429-438.
6. Homayouni R, Nunez-Regueiro M, Byrne JH, Eskin A: **Identification of two phosphoproteins affected by serotonin in Aplysia sensory neurons.** *Brain Res* 1997, **750(1-2)**:87-94.
7. Homayouni R, Rice DS, Sheldon M, Curran T: **Disabled-1 binds to the cytoplasmic domain of amyloid precursor-like protein 1.** *J Neurosci* 1999, **19(17)**:7507-7515.
8. Heinrich KE, Berry MW, Homayouni R: **Gene tree labeling using nonnegative matrix factorization on biomedical literature.** *Comput Intell Neurosci* 2008, 276535.
9. Homayouni R, Heinrich K, Wei L, Berry MW: **Gene clustering by latent semantic indexing of MEDLINE abstracts.** *Bioinformatics* 2005, **21(1)**:104-115.
10. Roy S, Heinrich K, Phan V, Berry MW, Homayouni R: **Latent Semantic Indexing of PubMed abstracts for identification of transcription factor candidates from microarray derived gene sets.** *BMC Bioinformatics* 2011, **12(Suppl 10)**:S19.
11. Tjioe E, Berry MW, Homayouni R: **Discovering gene functional relationships using FAUN (Feature Annotation Using Nonnegative matrix factorization).** *BMC Bioinformatics* 2010, **11(Suppl 6)**:S14.
12. Xu L, Furlotte N, Lin Y, Heinrich K, Berry MW, George EO, Homayouni R: **Functional cohesion of gene sets determined by latent semantic indexing of PubMed abstracts.** *PLoS One* 2011, **6(4)**:e18851.
13. Jernigan TL, Brown TT, Hagler DJ Jr, Akshoomoff N, Bartsch H, Newman E, Thompson WK, Bloss CS, Murray SS, Schork N, *et al*: **The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository.** *Neuroimage* 2015.
14. Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, *et al*: **An automated labeling**

system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 2006, **31**(3):968-980.

15. Chen CH, Gutierrez ED, Thompson W, Panizzon MS, Jernigan TL, Eyler LT, Fennema-Notestine C, Jak AJ, Neale MC, Franz CE, *et al*: Hierarchical genetic organization of human cortical surface area. *Science* 2012, **335**(6076):1634-1636.

16. Bartsch H, Thompson WK, Jernigan TL, Dale AM: A web-portal for interactive data exploration, visualization, and hypothesis testing. *Front Neuroinform* 2014, **8**:25.

17. Karagiannis TC: The timeless influence of Hippocratic ideals on diet, salicytates and personalized medicine. *Hell J Nucl Med* 2014, **17**(1):2-6.

18. Collins FS, Varmus H: A new initiative on precision medicine. *N Engl J Med* 2015, **372**(9):793-795.

19. McGregor TL, Van Driest SL, Brothers KB, Bowton EA, Muglia LJ, Roden DM: Inclusion of pediatric samples in an opt-out biorepository linking DNA to de-identified medical records: pediatric BioVU. *Clin Pharmacol Ther* 2013, **93**(2):204-211.

20. Pendergrass S, Dudek SM, Roden DM, Crawford DC, Ritchie MD: Visual integration of results from a large DNA biobank (BioVU) using synthesis-view. *Pac Symp Biocomput* 2011, 265-275.

21. Carroll RJ, Bastarache L, Denny JC: R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* 2014, **30**(16):2375-2376.

22. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, *et al*: Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol* 2013, **31**(12):1102-1110.

23. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC: PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 2010, **26**(9):1205-1210.

24. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, Li R, Masys DR, Ritchie MD, Roden DM, *et al*: The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics* 2011, **4**:13.

25. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, Riley WT, Shar A, Spring B, Spruijt-Metz D, *et al*: Mobile health technology evaluation: the mHealth evidence workshop. *Am J Prev Med* 2013, **45**(2):228-236.

26. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, Pavel M, Atienza AA: Advancing the science of mHealth. *J Health Commun* 2012, **17**(Suppl 1):5-10.

27. Bergsma R, Kanis E, Knol EF, Bijma P: The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). *Genetics* 2008, **178**(3):1559-1570.

28. Bouwman AC, Bergsma R, Duijvesteijn N, Bijma P: Maternal and social genetic effects on average daily gain of piglets from birth until weaning. *J Anim Sci* 2010, **88**(9):2883-2892.

29. Camerlink I, Turner SP, Bijma P, Bolhuis JE: Indirect genetic effects and housing conditions in relation to aggressive behaviour in pigs. *PLoS One* 2013, **8**(6):e65136.

30. Camerlink I, Ursinus WW, Bijma P, Kemp B, Bolhuis JE: Indirect genetic effects for growth rate in domestic pigs alter aggressive and manipulative biting behaviour. *Behav Genet* 2015, **45**(1):117-126.

31. Schroeder JW, Conneely KN, Cubells JC, Kilaru V, Newport DJ, Knight BT, Stowe ZN, Brennan PA, Krushkal J, Tylavsky FA, *et al*: Neonatal DNA methylation patterns associate with gestational age. *Epigenetics* 2011, **6**(12):1498-1504.

32. Alisch RS, Barwick BG, Chopra P, Myrick LK, Satten GA, Conneely KN, Warren ST: Age-associated DNA methylation in pediatric populations. *Genome Res* 2012, **22**(4):623-632.

33. Barfield RT, Almli LM, Kilaru V, Smith AK, Mercer KB, Duncan R, Klengel T, Mehta D, Binder EB, Epstein MP, *et al*: Accounting for population stratification in DNA methylation studies. *Genet Epidemiol* 2014, **38**(3):231-241.

34. Klengel T, Mehta D, Anacker C, Rex-Haffner M, Pruessner JC, Pariante CM, Pace TW, Mercer KB, Mayberg HS, Bradley B, *et al*: Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nat Neurosci* 2013, **16**(1):33-41.

35. Smith AK, Kilaru V, Klengel T, Mercer KB, Bradley B, Conneely KN, Ressler KJ, Binder EB: DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain. *Am J Med Genet B Neuropsychiatr Genet* 2015, **168B**(1):36-44.

36. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN: Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics* 2014, **15**:145.

37. Jiang R, Jones MJ, Chen E, Neumann SM, Fraser HB, Miller GE, Kobor MS: Discordance of DNA methylation variance between two accessible human tissues. *Sci Rep* 2015, **5**:8257.

38. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS: Factors underlying variable DNA methylation in a human community cohort. *Proc Natl Acad Sci U S A* 2012, **109**(Suppl 2):17253-17260.

39. Essex MJ, Boyce WT, Hertzman C, Lam LL, Armstrong JM, Neumann SM, Kobor MS: Epigenetic vestiges of early developmental adversity: childhood stress exposure and DNA methylation in adolescence. *Child Dev* 2013, **84**(1):58-75.

40. Miller GE, Chen E, Fok AK, Walker H, Lim A, Nicholls EF, Cole S, Kobor MS: Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc Natl Acad Sci U S A* 2009, **106**(34):14716-14721.

41. Chen E, Miller GE, Kobor MS, Cole SW: Maternal warmth buffers the effects of low early-life socioeconomic status on pro-inflammatory signaling in adulthood. *Mol Psychiatry* 2011, **16**(7):729-737.