Tomas Babak, Ph.D.
Assistant Professor

Dept. of Biology, Queen's University, Kingston, ON
tomasbabak@gmail.com | 343-333-6131

SUMMARY

Computational Biologist with 13 years experience inventing and testing novel applications for emerging technologies and large scale biological datasets. Driven by high-impact scientific discovery with a proven research track record in both industry and academia and a broadly established network of collaborators. Adept statistical geneticist, integrative biologist, and molecular biologist with 5 years management experience.

PROFESSIONAL EXPERIENCE

Queen's University, Dept. of Biology — **Assistant Professor**
Kingston, ON, Canada — July 2014-present

- Multiple funded research programs:
  - Develop cheap long-read (>5 kb) sequencing technology for next-gen sequencers
  - Decipher the causes of schizophrenia and bipolar disorders using via genetics and transcriptomics
  - Developing CRISPR/Cas9 single-cell synthetic lethal screening technology
  - Globally characterize *cis*-regulatory drivers of breast and ovarian cancers using TCGA/GTEx data
  - Build miRNA regulatory network on basis of allele-specific expression in GTEx
  - Establishing evolutionary history of genomic imprinting in mammals
- >300 hours formal bioinformatics, cell and systems biology teaching experience
- Collaborating on >10 research projects

Stanford University, Dept. of Biology — **Research Fellow**
Stanford, CA, 94305 — Jan. 2012-June 2014

- Developed approach for systematically discriminating genetic from epigenetic influence on gene expression in humans; generated comprehensive body atlas of genomic imprinting in human and mouse; involved integrative analysis of 7 data types across thousands of samples; in press at **Nature Genetics**
- Developed empirically derived statistical tool for quantifying allele-specific expression (manuscript in prep)
- Developed molecular method for capturing genomic DNA using mRNA as bait (manuscript in prep)
- Supervised 4 graduate students for one year while PI was on sabbatical

NuGEN Technologies, — **Consultant**
San Carlos, CA, 94070 — Feb. 2012-July 2012

- Advised on development of 6 independent molecular next-gen methods for genome sequencing, mapping methylation, and discovery of genetic variants from nucleic-acid limiting samples; 4 of these are now commercially available as kits

Merck Research Labs, Next-Gen Sequencing Methods Development Lab — **Head**
Boston, MA, 02115 — Nov. 2009-Dec. 2011

- Worked closely with ~40 of scientists in Oncology, Neuroscience, Cardiovascular Health, Diabetes, Biologics, Vaccine Development to develop custom next-gen sequencing solutions incl. both molecular

methods and analysis strategy (22 research projects in total, >10 novel methods developed, e.g. PolyA-Seq published in **Genome Research**)
- Supervised team of 1 Ph.D. Research Fellow, 1 Senior Lab Technician, 4 co-op graduate students

Merck Research Labs, Dept. of Informatics and Analysis — **Senior Research Scientist**
Boston, MA, 02115 — Oct. 2008-Oct. 2009

- Built next-gen sequencing analysis pipeline that has evolved and continues to be widely utilized; modular design enables rapid customization with emerging software; parallelized by design to scale with need; handles all popular formats, common quality control, and analysis outputs
- Designed and implemented sequencing analysis workflows with BGI where the bulk of Merck's sequencing was outsourced
- Performed custom analyses for the Oncology franchise, e.g. integrating RNAi+small molecule screening data with exome-discovered SNVs to predict drug response

Rosetta Inpharmatics, Dept. of Molecular Informatics — **Postdoctoral Fellow (with Lee P. Lim)**
Seattle, WA, 98101 — Feb. 2007-Sept. 2008

- Developed and applied NSR for genome-wide discovery of genomic imprinting in mouse; recapitulated 25 years of work in a single experiment; published in **Current Biology** and **Nat Methods**

EDUCATION

University of Toronto, Toronto, ON — **Ph.D. – Molecular Genetics (2002-2007)**
Queen's University, Kingston, ON — **B.Sc. – Biology, minor in Physics (1998-2002)**

TECHNICAL SKILLS

Informatics

- Programming: Perl, Python, Matlab, HTML, SQL, Bootstrap
- Systems: Linux/Unix (incl. PBS, Torque, LSF, or SGE), Windows, MacOS X
- Software: next-gen sequencing analysis tools (e.g. bwa, SOAP, Bowtie suite, Novocraft utilities, BEDTools, SAMtools, GATK, SRA Toolkit, Bismark, Oasis, Velvet + others incl. dozens of custom tools), SHAPEIT, MACH, Impute2, Bioconductor (R), Gist SVM, Matlab, MS Office (Excel, Word, Powerpoint), Adobe Illustrator+Photoshop, Sequencher
- Genomics: sequence analysis, functional inference from expression data, integrating any type of systematically collected structured data, machine learning classification, ROC analysis
- Genetics: SNP phasing and imputation, eQTL analysis, allele-specific expression, GWAS (using plink), FDR/Power estimates, probabilistic integration with genomics data

Molecular Biology

- Next-gen sequencing: library construction (e.g. (un)stranded-mRNA-Seq, Targeted capture/Exome, WGS, ChIP-Seq), >10 custom approaches (e.g. NSR, PolyA-Seq)
- Tissue culture: CRISPR/Cas9 engineering, RNAi, transformation/transfection, FACS
- Basic methods: cloning, (q/RT)-PCR, northerns/southerns/westerns, Sanger seq, IVT
- Specialized technology: MiSeq, Illumina GAIIx, Fluidigm, microarrays, Biomek liquid handling, Agilent bioanalyzer, Covaris

SELECTED PUBLICATIONS

(of 30 total with 2481 citations)

1. **Babak T**, et. al. (2014) Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. **Nature Genetics** (*in press; scheduled for publication March 9*).
2. DeVeale B, van der Kooy D, **Babak T**. (2012) Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. **PLoS Genetics** Mar;8(3):e1002600. (**95 citations**)

   Featured In: • Imprinted Genes … and the Number Is? **PLoS Genetics,** Kelsey, G. Bartolomei M. (2012).
   • Fewer imprinted genes at re-count. **Nature** 484: 145, Shen H. (2012).
   • RNA studies under fire. **Nature** 484: 428, Hayden EC. (2012).
3. Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, **Babak T**. (2012) A quantitative atlas of polyadenylation in five mammals. **Genome Research** Jun;22(6):1173-83. (**84 citations**)

   Featured In: Sequencing the end. **Nature Methods** 9, 534 (2012).
4. **Babak T,** et. al. (2010) Genetic validation of whole-transcriptome sequencing for mapping expression affected by cis-regulatory variation. **BMC Genomics** 13;11:473 (**24 citations**)
5. Armour CD, Castle JC Chen R, **Babak T,** Loerch P, Jackson S, Shah JK, Dey J, Rohl C, Johnson JM, Raymond CK. (2009) Digital profiling of whole transcriptomes using selective hexamer priming for cDNA synthesis. **Nature Methods** 6(9):647-9 (**115 citations**)
6. **Babak T,*** DeVeale B,* et. al. (2008) Global survey of genomic imprinting by transcriptome sequencing. **Current Biology**, 18: 1735-1741 (**108 citations**)
7. Huang JC*, **Babak T***, Corson TW, Chua G, Khan S, Gallie BL, Hughes TR, Blencowe BJ, Frey BJ, Morris QD. (2007) Using expression profiling data to identify human miroRNA targets. **Nature Methods** 4(12):1045-9 (**259 citations**)
8. **Babak T**, Blencowe BJ, Hughes TR. (2007) Considerations in the identification of functional RNA structural elements in genomic alignments. **BMC Bioinformatics** 8:33 (**55 citations**)
9. **Babak T**, Blencowe BJ, Hughes TR. (2005) A systematic search for new mammalian noncoding RNAs indicates little conserved intergenic transcription. **BMC Genomics** 6:104 (**59 citations**)
10. **Babak T**, Zhang W, Morris Q, Blencowe BJ, Hughes TR. (2004) Probing microRNAs with microarrays: tissue specificity and functional inference. **RNA** 10(11):1813-9 (**330 citations**)

SELECTED INVITED PRESENTATIONS (of 12)

Institute of Molecular Life Sciences. *Elucidating evolutionary origins of genomic imprinting through a cross-species transcriptome comparison.* University of Zurich, Switzerland. Oct. 15, 2013.

Next Generation Sequencing and Bioinformatics Summit. *Integrating RNA-Seq features to understand schizophrenia and bipolar disease.* Berlin, Germany. Oct. 9, 2013.

Next Generation Sequencing: GTC-Bio Meeting. *Global mapping of allele-specific expression to understand complex disease.* San Francisco, CA. Jun 21, 2013.

University of Pittsburgh, Computational and Systems Biology. *Integrating RNA-Seq features to understand non-mendelian phenotypes.* Pittsburgh, PA. Feb. 20, 2013.