**Gerstein lab experience in minor splicing analysis**

The Gerstein lab has worked closely with the research groups led by Drs. Mark Rubin and Rahul Kanadia to study systematic differences in gene expression associated with minor-intron-containing genes (MIGs). This work was published in *Molecular Cell* (Augspach et al, 2023). In the context of this study, we used silhouette scores to rigorously demonstrate that MIGs exhibit stronger differential gene expression across cancer types and stages of cancer development, relative to genes that do not contain minor introns (that is, relative to non-MIG genes).

**Gerstein lab experience in GENCODE and annotating pseudogenes throughout the genome**

The Gerstein lab has extensive experience in annotating pseudogenes throughout the genome. Along these lines, we are one of the major participants of the GENCODE project. As part of our 15-year involvement in GENCODE, our focus has been on pseudogenes (Sisu et al, 2014; Pei et al, 2012) and genome annotation (Frankish et al, 2019). We have published several papers on pseudogene annotation, analysis, and regulatory annotation throughout the genome (Lam et al, 2009; Liu et al, 2009; Balasubramanian et al, 2009; Balasubramanian et al, 2011; Abyzov et al, 2013; Khurana et al, 2010; Zheng et al, 2007). In particular, we developed the PseudoPipe tool (Fig. 1) to find the pseudogenes (Zhang et al, 2006), and it is one of the major tools used to annotate pseudogenes in GENCODE (Frankish et al, 2019).

A black background with white text

Description automatically generated

**Fig. 1: Workflow of PseudoPipe.**

**References**

Augspach, Anke, et al. "Minor intron splicing is critical for survival of lethal prostate cancer." Molecular cell 83.12 (2023): 1983-2002.

Abyzov, Alexej, Rebecca Iskow, Omer Gokcumen, David W. Radke, Suganthi Balasubramanian, Baikang Pei, Lukas Habegger, Charles Lee, Mark Gerstein, and 1000 Genomes Project Consortium. "Analysis of variable retroduplications in human populations suggests coupling of retrotransposition to cell division." Genome research 23, no. 12 (2013): 2042-2052.

Balasubramanian, Suganthi, Deyou Zheng, Yuen-Jong Liu, Gang Fang, Adam Frankish, Nicholas Carriero, Rebecca Robilotto, Philip Cayting, and Mark Gerstein. "Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes." Genome biology 10 (2009): 1-10.

Balasubramanian, Suganthi, Lukas Habegger, Adam Frankish, Daniel G. MacArthur, Rachel Harte, Chris Tyler-Smith, Jennifer Harrow, and Mark Gerstein. "Gene inactivation and its implications for annotation in the era of personal genomics." Genes & development 25, no. 1 (2011): 1-10.

Frankish, Adam, Mark Diekhans, Anne-Maud Ferreira, Rory Johnson, Irwin Jungreis, Jane Loveland, Jonathan M. Mudge et al. "GENCODE reference annotation for the human and mouse genomes." Nucleic acids research 47, no. D1 (2019): D766-D773.

Khurana, Ekta, Hugo YK Lam, Chao Cheng, Nicholas Carriero, Philip Cayting, and Mark B. Gerstein. "Segmental duplications in the human genome reveal details of pseudogene formation." Nucleic acids research 38, no. 20 (2010): 6997-7007.

Lam, Hugo YK, Ekta Khurana, Gang Fang, Philip Cayting, Nicholas Carriero, Kei-Hoi Cheung, and Mark B. Gerstein. "Pseudofam: the pseudogene families database." Nucleic acids research 37, no. suppl\_1 (2009): D738-D743.

Liu, Yuen-Jong, Deyou Zheng, Suganthi Balasubramanian, Nicholas Carriero, Ekta Khurana, Rebecca Robilotto, and Mark B. Gerstein. "Comprehensive analysis of the pseudogenes of glycolytic enzymes in vertebrates: the anomalously high number of GAPDH pseudogenes highlights a recent burst of retrotrans-positional activity." BMC genomics 10 (2009): 1-12.

Pei, Baikang, Cristina Sisu, Adam Frankish, Cédric Howald, Lukas Habegger, Xinmeng Jasmine Mu, Rachel Harte et al. "The GENCODE pseudogene resource." Genome biology 13 (2012): 1-26.

Sisu, Cristina, Baikang Pei, Jing Leng, Adam Frankish, Yan Zhang, Suganthi Balasubramanian, Rachel Harte et al. "Comparative analysis of pseudogenes across three phyla." Proceedings of the National Academy of Sciences 111, no. 37 (2014): 13361-13366.

Zhang, Zhaolei, Nicholas Carriero, Deyou Zheng, John Karro, Paul M. Harrison, and Mark Gerstein. "PseudoPipe: an automated pseudogene identification pipeline." Bioinformatics 22, no. 12 (2006): 1437-1439.

Zheng, Deyou, Adam Frankish, Robert Baertsch, Philipp Kapranov, Alexandre Reymond, Siew Woh Choo, Yontao Lu et al. "Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution." Genome research 17, no. 6 (2007): 839-851.