Gerstein Lab Highlights in 2017

The Gerstein lab research

During 2017, the Gerstein lab was involved in several research projects in the areas of personal genome variation, genomics at the forefront of Data Science, analysis of diverse networks and next generation sequencing. These projects resulted in various publications in journals such as Nature and Nat. communications,

Core Publications (for full citations see publication list)

A large part of the lab's research is focused on disease genomics. The identification of coding and especially non-coding drivers remains challenging. In papillary kidney cancer, we performed a whole-genome analysis to find significant noncoding alterations(Li, Shuch, & Gerstein, 2017). For single nucleotide variants, we developed a tool called ALoFT which determines the impact of loss of function variants in protein-coding genes(Balasubramanian et al., 2017). Loss-of-function variants play a significant role in cancer, autism and Mendelian diseases. To detect highly mutated regions, we developed MOAT, a software which accounts for covariant factors that influence mutation rate(Lochovsky, Zhang, & Gerstein, 2017). Finally, we did a commentary in Nature about the search for drivers in cancer with particular attention to noncoding DNA sequences(Kumar & Gerstein, 2017).

While studying structural variants, we identified the landscape and variation of novel retroduplications in 26 human populations(Zhang, Li, Abyzov, & Gerstein, 2017). Retroduplications come from reverse transcription of mRNA and re-insertion into the genome. A lot of retroduplications are associated with diseases. For analyzing genome structure and organization, we developed MrTADFinder, a software that identifies Topological Associating Domains (TADs). Using HI-C sequencing technology and network modularity, we explored the intra-chromosomal contact map to identify and reveal TADs (Yan, Lou, & Gerstein, 2017; Yan, Yardlmcl, Yan, Noble, & Gerstein, 2017).

Collaborative publications

In addition to our lab's main publications, our lab participates in numerous consortia, collaborations and genomic projects. Amongst others, the Gerstein lab is part of the Encyclopedia of DNA Elements (ENCODE)(Consortium, 2013; Feingold et al., 2004), BRAINSPAN(Kang et al., 2011) and psychENCODE (ENCODE for noncoding regulatory elements and epigenetic modifications for brain and mental diseases)(Akbarian et al., 2015), and the Pancancer Analysis of Whole Genomes (PCAWG) research initiatives(Campbell et al., 2017). Within this framework, we explored the transcriptional landscape of proteins in post-natal human brains(Carlyle et al., 2017). Finally, in the past few years there is an increasing interest in extracellur RNAs (ex-RNA), associating selected microRNAs with various cardiovascular diseases. In a collaborative effort, we found 7 ex-RNAs associated with stroke prevalence or incidence(Mick et al., 2017).

REFERENCES

Akbarian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., Crawford, G. E.,

… Sestan, N. (2015). The PsychENCODE project. *Nature Neuroscience*. http://doi.org/10.1038/nn.4156

Balasubramanian, S., Fu, Y., Pawashe, M., McGillivray, P., Jin, M., Liu, J., … Gerstein, M. (2017). Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. *Nature Communications*, *8*(1). http://doi.org/10.1038/s41467-017-00443-5

Campbell, P. J., Getz, G., Stuart, J. M., Korbel, J. O., Stein, L. D., & Net, - ICGC/TCGA Pan-Cancer Analysis of Whole Genomes. (2017). Pan-cancer analysis of whole genomes. *bioRxiv*, 162784. http://doi.org/10.1101/162784

Carlyle, B. C., Kitchen, R. R., Kanyo, J. E., Voss, E. Z., Pletikos, M., Sousa, A. M. M., … Nairn, A. C. (2017). A multiregional proteomic survey of the postnatal human brain. *Nature Neuroscience*, *20*(12), 1787–1795. http://doi.org/10.1038/s41593-017-0011-2

Consortium, E. P. (2013). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. http://doi.org/10.1038/nature11247.An

Feingold, E. A., Good, P. J., Guyer, M. S., Kamholz, S., Liefer, L., Wetterstrand, K., … Harvey, S. C. (2004). The ENCODE (ENCyclopedia of DNA Elements) Project. *Science*. http://doi.org/10.1126/science.1105136

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., … Šestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature*, *478*(7370), 483–489. http://doi.org/10.1038/nature10523

Kumar, S., & Gerstein, M. (2017). Cancer genomics: Less is more in the hunt for driver mutations. *Nature*. http://doi.org/10.1038/nature23085

Li, S., Shuch, B. M., & Gerstein, M. B. (2017). Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations. *PLoS Genetics*, *13*(3). http://doi.org/10.1371/journal.pgen.1006685

Lochovsky, L., Zhang, J., & Gerstein, M. (2017). MOAT: efficient detection of highly mutated regions with the Mutations Overburdening Annotations Tool. *Bioinformatics*. http://doi.org/10.1093/bioinformatics/btx700

Mick, E., Shah, R., Tanriverdi, K., Murthy, V., Gerstein, M., Rozowsky, J., … Freedman, J. E. (2017). Stroke and Circulating Extracellular RNAs. *Stroke*, *48*(4), 828–834. http://doi.org/10.1161/STROKEAHA.116.015140

Yan, K. K., Lou, S., & Gerstein, M. (2017). MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions. *PLoS Computational Biology*, *13*(7). http://doi.org/10.1371/journal.pcbi.1005647

Yan, K. K., Yardlmcl, G. G., Yan, C., Noble, W. S., & Gerstein, M. (2017). HiC-spector: A matrix library for spectral and reproducibility analysis of Hi-C contact maps. In *Bioinformatics* (Vol. 33, pp. 2199–2201). http://doi.org/10.1093/bioinformatics/btx152

Zhang, Y., Li, S., Abyzov, A., & Gerstein, M. B. (2017). Landscape and variation of novel retroduplications in 26 human populations. *PLoS Computational Biology*, *13*(6). http://doi.org/10.1371/journal.pcbi.1005567