Proximity of sQTLs and target genes in 3D Genome

1 Introduction

Single nucleotide polymorphisms (SNPs) are the most frequent genetic variants in humans and represent an invaluable resource to understand the genetic basis of diseases [10]. Genome-wide association studies (GWAS) have found abundant SNPs associated with various traits and diseases, but most of risk loci lack clear molecular mechanisms [1]. Expression quantitative trait locus (eQTL) studies have been employed to identify SNPs that may influence the expression levels of genes, thereby contributing to the phenotype outcome [5, 2, 4]. However, only a moderate proportion of GWAS-identified loci are strong eQTLs [], which might be partly due to the small sample sizes, the tissues studied, and a focus on overall gene level expression measurements without consideration of transcript isoforms [].

Alternative splicing (AS) is a widespread process that increases structural transcript variation and proteome diversity. The invention of RNA sequencing greatly facilitated the identification of AS at the genomic level [6]. Aberrant splicing patterns are frequently observed in cancer initiation, progress, prognosis and therapy. Increasing evidence has demonstrated that AS events could undergo modulation by genetic variants. The identification of splicing quantitative trait loci (sQTLs), genetic variants that affect AS events, might represent an important step toward fully understanding the contribution of genetic variants in disease development. AS process is often altered in cancer cells to produce aberrant proteins that drive the progression of cancer. A better understanding of the misregulation of alternative splicing will shed light on the development of novel targets for pharmacological interventions of cancer.

Genome-wide sQTL mapping has been achieved for multiple species ranging from human to Arabidopsis Thaliana [8].

1.1 Related Work

Joint analyses of multi-tissue Hi-C and eQTL data demonstrate close spatial proximity between eQTLs and their target genes by [14] and [3] discuss work similar to ours.

2 Experiments and Datasets

We obtain sQTLs for cancer cells from CancerSplicingQTL database [12] `http://www.cancersplicingqtl-hust.com/#/sqtls`, which compiles genome-wide sQTL data from publications []. We intersect gene names in this database with gene names or IDs in the Ensembl database [13] and select genes that are associated with a unique range in the Ensembl database to produce a collection S of sQTLs.

- Prostate cancer: A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118629` In this paper, normal prostate (RWPE1) and prostate cancer cells (C42B and 22Rv1). sQTL is in CancerSplicingQTL database under PRAD (Prostate Adenocarcinoma) [12].

- LAML (Acute Myeloid Leukemia): is of K562 cell type. Hi-C data for K562 cell is obtained in [9], sQTL is in CancerSplicingQTL database [12].

- CML (Chronic Myeloid Leukemia): is of KBM7 cell type. Hi-C data for K562 cell is obtained in [9], sQTL is in CancerSplicingQTL database [12] ??.

- Multiple Myeloma (MM): 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. Could not find sQTL database.

- Breast Cancer (BRCA): *Information will be added!*

- Glioblastoma (GBM): *Information will be added!*

- Dorsolateral Prefrontal Cortex (DLPFC): sQTL is available in Supplementary files of Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci paper [11]

- Hippocampus: *Information will be added!*

- Some drugs also target 3D genome organization as in [7] on HT1080 cells (fibrosarcoma(SARC)). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3304

Cool extension files are interaction matrices. sQTLs are available under xxx in CancerSplicingQTL

- Does sQTLs appear more frequently in TAD boundaries?
- Median survival-QTL

References

[1] Jiang Chang, Jianbo Tian, Yang Yang, Rong Zhong, Jiaoyuan Li, Kan Zhai, Juntao Ke, Jiao Lou, Wei Chen, Beibei Zhu, Na Shen, Yi Zhang, Yajie Gong, Ying Zhu, Danyi Zou, Xiating Peng, Kun Huang, and Xiaoping Miao. A rare missense variant in tcf7l2 associates with colorectal cancer risk by interacting with a gwas-identified regulatory variant in the myc enhancer. *Cancer Research*, 78(17):5164–5172, 2018.

[2] Zou Danyi, Jiao Lou, Juntao Ke, Shufang Mei, Jiaoyuan Li, Yajie Gong, Yang Yang, Ying Zhu, Jianbo Tian, Jiang Chang, Rong Zhong, Jing Gong, and Xiaoping Miao. Integrative expression quantitative trait locus-based analysis of colorectal cancer identified a functional polymorphism regulating slc22a5 expression. *European journal of cancer (Oxford, England : 1990)*, 93:1–9, 02 2018.

[3] Geet Duggal, Hao Wang, and Carl Kingsford. Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Research*, 42(1):87–96, 10 2013.

[4] Jing Gong, J Tian, J Lou, X Wang, J Ke, J Li, Y Yang, Y Gong, Y Zhu, D Zou, X Peng, N Yang, S Mei, R Zhong, Jiang Chang, and X Miao. A polymorphic myc response element in kbtbd11 influences colorectal cancer risk, especially in interaction with a myc regulated snp rs6983267. *Annals of oncology : official journal of the European Society for Medical Oncology*, 29, 12 2017.

[5] Xingyi Guo, Weiqiang Lin, Qiuyin Cai, Jiajun Shi, Yaqiong Sun, Xiao Pan, Mi-Ryung Han, Wanqing Wen, Bingshan Li, Jirong Long, Jianghua Chen, and Wei Zheng. Abstract 1441: A comprehensive cis-eqtl analysis revealed target genes in breast cancer susceptibility loci identified in genome-wide association studies. *Cancer Research*, 77:1441–1441, 07 2017.

[6] Daejin Hyung, Jihyun Kim, Soo Young Cho, and Charny Park. ASpedia: a comprehensive encyclopedia of human alternative splicing. *Nucleic Acids Research*, 46(D1):D58–D63, 11 2017.

[7] Omar Kantidze, Artem Luzhin, Ekaterina Nizovtseva, Alfiya Safina, Maria Valieva, Arkadiy Golov, Artem Velichko, Alexander Lyubitelev, Alexey Feofanov, Katerina Gurova, Vasily Studitsky, and Sergey Razin. The anti-cancer drugs curaxins target spatial genome organization. *Nature Communications*, 10:1441, 03 2019.

[8] Waqas Khokhar, Musa A. Hassan, Anireddy S. N. Reddy, Saurabh Chaudhary, Ibtissam Jabre, Lee J. Byrne, and Naeem H. Syed. Genome-wide identification of splicing quantitative trait loci (sqtls) in diverse ecotypes of arabidopsis thaliana. *Frontiers in Plant Science*, 10:1160, 2019.

[9] S.S.P. Rao, M.H. Huntley, Neva Durand, E.K. Stamenova, Ivan Bochkov, J.T. Robinson, A.L. Sanborn, I. Machol, Arina Omer, E.S. Lander, and E.L. Aiden. A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 162:687–688, 08 2015.

[10] Barkur S Shastry. Snps: impact on gene function and phenotype. *Methods in molecular biology (Clifton, N.J.)*, 578:3—22, 2009.

[11] Atsushi Takata, Naomichi Matsumoto, and Tadafumi Kato. Genome-wide identification of splicing qtls in the human brain and their enrichment among schizophrenia-associated loci. *Nature Communications*, 8:14519, 02 2017.

[12] Jianbo Tian, Zhihua Wang, Shufang Mei, Nan Yang, Yang Yang, Juntao Ke, Ying Zhu, Yajie Gong, Danyi Zou, Xiating Peng, Xiaoyang Wang, Hao Wan, Rong Zhong, Jiang Chang, Jing Gong, Leng Han, and Xiaoping Miao. CancerSplicingQTL: a database for genome-wide identification of splicing QTLs in human cancer. *Nucleic Acids Research*, 47(D1):D909–D916, 10 2018.

[13] Andrew D Yates, Premanand Achuthan, Wasiu Akanni, James Allen, Jamie Allen, Jorge Alvarez-Jarreta, M Ridwan Amode, Irina M Armean, Andrey G Azov, Ruth Bennett, Jyothish Bhai, Konstantinos Billis, Sanjay Boddu, José Carlos Marugán, Carla Cummins, Claire Davidson, Kamalkumar Dodiya, Reham Fatima, Astrid Gall, Carlos Garcia Giron, Laurent Gil, Tiago Grego, Leanne Haggerty, Erin Haskell,

Thibaut Hourlier, Osagie G Izuogu, Sophie H Janacek, Thomas Juettemann, Mike Kay, Ilias Lavidas, Tuan Le, Diana Lemos, Jose Gonzalez Martinez, Thomas Maurel, Mark McDowall, Aoife McMahon, Shamika Mohanan, Benjamin Moore, Michael Nuhn, Denye N Oheh, Anne Parker, Andrew Parton, Mateus Patricio, Manoj Pandian Sakthivel, Ahamed Imran Abdul Salam, Bianca M Schmitt, Helen Schuilenburg, Dan Sheppard, Mira Sycheva, Marek Szuba, Kieron Taylor, Anja Thormann, Glen Threadgold, Alessandro Vullo, Brandon Walts, Andrea Winterbottom, Amonida Zadissa, Marc Chakiachvili, Bethany Flint, Adam Frankish, Sarah E Hunt, Garth IIsley, Myrto Kostadima, Nick Langridge, Jane E Loveland, Fergal J Martin, Joannella Morales, Jonathan M Mudge, Matthieu Muffato, Emily Perry, Magali Ruffier, Stephen J Trevanion, Fiona Cunningham, Kevin L Howe, Daniel R Zerbino, and Paul Flicek. Ensembl 2020. *Nucleic Acids Research*, 48(D1):D682–D688, 11 2019.

[14] Jingting Yu, Ming Hu, and Chun Li. Joint analyses of multi-tissue hi-c and eqtl data demonstrate close spatial proximity between eqtls and their target genes. *BMC Genetics*, 20, 12 2019.