

1. Objectives/aims. This project seeks to develop an information hub for training/ teaching agricultural researchers with an aim to facilitate application of single-cell functional genomics in quantitative trait loci (QTLs) mapping of agriculturally important traits. Integration of genetic variation data with cellular and molecular data has been used to map expression QTL (eQTL) or methylation QTL (mQTL) or chromatin accessible QTL (caQTL) linked to diverse phenotypes (Kumasaka *et al.*, 2016; Volkov *et al.*, 2016; Ciuculete *et al.*, 2017; Benaglio *et al.*, 2020a; Keele *et al.*, 2020; Zhao *et al.*, 2020). Such approaches have also been applied to map QTLs linked to traits of agricultural importance (Long *et al.*, 2011; Liu *et al.*, 2020; Kushanov *et al.*, 2021; Yuan *et al.*, 2021). However, these studies have been performed with functional genomics data derived from bulk tissues that cannot determine if the phenotype is influenced by specific cell types in the tissue. Recent spur in single-cell RNA sequencing (scRNA-seq) and single-cell Assay for Transposase-Accessible Chromatin sequencing (scATAC-seq) have generated new opportunities to integrate genetic variation with changes in gene expression and open chromatin profiles to identify single-cell eQTLs and caQTLs (Benaglio *et al.*, 2020b; van der Wijst *et al.*, 2020; Neavin *et al.*, 2021). These methods have great capabilities to untangle cellular and molecular links to major phenotypic traits such as crop yield, animal production and plant resistance to insect pests to sustain agricultural productivity (Cole *et al.*, 2021; Tripathi and Wilkins, 2021; Zhang *et al.*, 2021; Zhu *et al.*, 2021; Nyysölä *et al.*, 2022).

Towards an objective to promote single-cell QTL mapping in agriculture research, our current proposal has two specific aims. **Specific aim 1.** **A.** Develop an online resource to provide knowledge on cost-effective methods of QTL mapping and standards and best practices of single-cell eQTL and caQTL data analysis. **B.** Develop a computational workflow by implementing published methods for integrated analysis of genetic variants and scRNA-seq and scATAC-seq data. **Specific aim 2.** Collaborate with animal, plant, and entomology scientists, and the AG2PI community members, to help **A)** generate preliminary data, and **B)** acquire skills to perform integrated analysis of genetic variation and single-cell genomics data in QTL mapping studies.

We have performed preliminary work that will help achieve these specific aims. They include survey of literature on single-cell QTL mapping, standardization of protocol to prepare single nuclei from frozen samples, generation of single-cell functional genomics data, and data analysis. Earlier, we developed a computational workflow by implementing methods described by Butler and colleagues (Butler *et al.*, 2018) to identify cells and marker genes of self-renewal capacity of uterine organoids and uterine influence on fetoplacental communication (Fitzgerald *et al.*, 2019; Dhakal *et al.*, 2021). We have standardized a lab protocol to isolate high-quality single nuclei from frozen tissues (**Figure 1**). This method is suitable for processing large number of samples in single-cell QTL mapping projects in a fast and cost-effective manner. We have shown that the method can also be applied to insects (Cui *et al.*, 2022). Currently, the P.I. lab has been routinely performing single-cell RNA-seq and ATAC-seq in mice and pigs (**Figure 2**). An ongoing project in the lab focuses on mapping single-cell QTLs linked to litter size in pig. We have collected phenotype data and uterine samples from pigs producing either small or large litter size. Currently, we are generating scRNA-seq data from those samples. The lab of co-PI Decker has developed



Figure 1. Single nuclei prepared from frozen fetal brain of mouse by using our lab protocol.

innovative analytical methods to map signatures of polygenic selection. He has broad experience in mapping QTLs via genome-wide association studies (GWAS) in livestock animals. In addition, two collaborative labs have been identified with whom we will work towards developing QTL mapping studies of specific agronomic traits in rice plant and insect pest by leveraging single-cell genomics (letters attached).

Plans for specific aim 1: We will start this project by creating an information hub that will provide comprehensive knowledge to agricultural researchers about cost-effective methods of mapping single-cell QTLs. Cost of generating single-cell genomics data is a major limitation in large-scale mapping studies. Our plan in this project is to provide information on emerging strategies that employ cost-effective methods of Bulk-Segregant Analysis and pooled-GWAS (Craig *et al.*, 2009; Gaj *et al.*, 2012; Riaz *et al.*, 2016; Navarro-Escalante *et al.*, 2020; Xu *et al.*, 2020; Zhang and Panthee, 2020; Huang *et al.*, 2021). The information hub will provide educational resources relating to methods and principles of single-cell QTL mapping and other key aspects such as power analysis to determine sample size as well as the standards and best practices of analysis of single-cell QTL data. Within this specific aim, we will also make available a computational workflow to the research community for integrative analysis of genotype data and single-cell functional genomics data. The workflow will implement published methods of single-cell eQTL (Cuomo *et al.*, 2021) and single-cell caQTL (Kumasaka *et al.*, 2016) to develop a user-friendly pipeline using Python *Snakemake* framework. We will test the pipeline using the data described in Kumasaka *et al.* (Kumasaka *et al.*, 2016), and make it available for free download (upon registration with no cost) in the proposed website.

Plans for specific aim 2: In this specific aim, our plan is to assist agricultural researchers working on plants, animals or insect pests to promote single-cell QTL mapping projects of important economic traits. Our efforts will include i) help design project, and assist in generating preliminary data via single-cell genomics assays, and ii) provide bioinformatics training to integrate genetic variation data with single-cell genomics data. We will use zoom meetings for interactive sessions as well as YouTube videos for self-paced learning of the step-by-step methods of data analysis. The training will target researchers at all levels including graduate students, postdocs, scientists, and principal investigators. The ultimate aim of these collaborations is to provide researchers the knowledge and skills required to apply single-cell genomics in QTL mapping. We will demonstrate the feasibility of this specific aim by working with two collaborators who have shown interest to participate in the project to develop single-cell QTL mapping projects (see letters). Collaborator Dr. Nair will work towards mapping plant resistance to brown plant hopper in rice whereas collaborator Dr. Kang is interested to map host response to lipid transporter in fruit fly *Drosophila suzuki*. We will provide them necessary training and informational resources to help develop these projects in their respective labs.

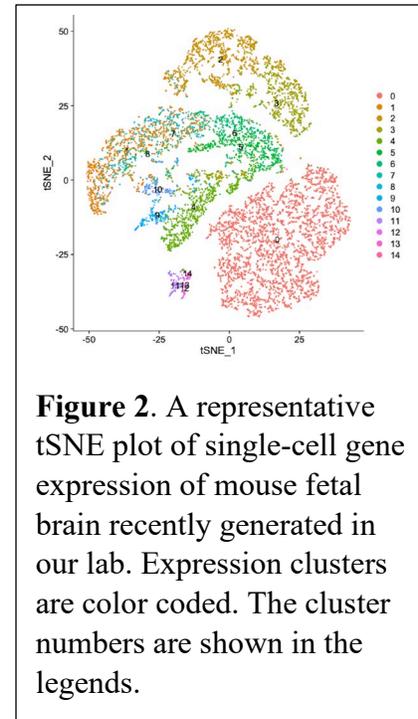


Figure 2. A representative tSNE plot of single-cell gene expression of mouse fetal brain recently generated in our lab. Expression clusters are color coded. The cluster numbers are shown in the legends.

2. Furthering the aims of the AG2PI. This project is expected to contribute to ongoing efforts to map agriculturally important traits in animals, plants, and insects. A great deal of advance has been made in QTL mapping in livestock animals as evident from the list of mapped QTLs curated in the Animal QTLdb (Hu *et al.*, 2022). Similar efforts have also been made in crop plants to map important agronomic traits (Zeng *et al.*, 2007; Ni *et al.*, 2009; Kim *et al.*, 2014; Singh *et al.*, 2021). However, efforts to map QTLs in agricultural insect pests are limited (Behura *et al.*, 2000, 2004; Zhao *et al.*, 2016). Our current effort will advance the goals of AG2PI to foster initiation steps that can help develop research needs and opportunities in the emerging areas of single-cell integration in QTL mapping. As outlined in our research plans, we will work in collaboration with national and international scientists to help advance this area of research. We have set three specific evaluation criteria to measure success of the proposed project: 1) Participation. We will evaluate success of our project based on interest and participation of agricultural researchers and AG2PI research community in our project, 2) Training. A key component of this project is to train researchers in performing single-cell data analysis. Number of participants (based on free registration) to our zoom sessions and YouTube learning materials will also be assessed to evaluate success of this project. 3) Collaboration. We will also use the number of labs that will benefit from this project to develop single-cell QTL mapping studies for evaluating success of our project.

3. Expected outcomes & deliverables. This project is expected to produce two major outcomes: 1) Disseminating information to facilitate application of single-cell genomics in QTL mapping in agricultural research, and 2) enabling national and international collaboration among animal, plant, and insect scientists to use single-cell genomics in trait mapping studies. The specific deliverables of this project will include- 1) launching a website to provide comprehensive informational resources relating to strategies and approaches of integrating single-cell data in BSA-seq and pool-GWAS methods, 2) developing a pipeline to perform bioinformatics data analysis of single-cell QTL mapping, 3) establishing collaboration with animal, plant and insect scientists to help develop preliminary data for single-cell QTL mapping, 4) providing online (zoom and YouTube) training to graduate students, postdocs and faculties interested to learn about single-cell genomics data analysis, and 5) facilitating collaborative efforts to help develop single-cell QTL mapping projects in agriculture. We have identified two collaborators within this project. We will assist them in study design, generating preliminary data and acquiring skills in bioinformatics analyses. Throughout this project, we will communicate with other researchers including the community members of AG2PI and the members of Functional Annotation of Farm Animal Genomes (FAANG) to join our efforts to expand this emerging area of genomics in agricultural research.

4. Qualifications of the project team. The P.I. is an Assistant Professor in Bioinformatics and Computational Biology in the Division of Animal Sciences. He obtained education in Mathematics, Biotechnology and Life Sciences. He has research experiences in gene and genome mapping, functional genomics and animal genomics. His research combines both experimental and computational work. He has published more than 90 papers. Besides research, he teaches a graduate level (4 credit hours) course on 'Functional Genomics of Mammals' that include a module on single-cell genomics. He is a Core Faculty member of the University of Missouri Institute of Data Science and Informatics (MUIDSI), and also a member of FAANG community.

The Co-P.I. is Wurdack Chair in Animal Genomics and Associate Professor in the Division of Animal Sciences. He is a faculty within the Genetics Area Program and also a Core Faculty member in the Institute of Data Science and Informatics. Decker's Computational Genomics Group combines population genomics and quantitative genomics to improve the sustainability of cattle and swine populations. Decker has been P.I. on over \$4 millions of competitive federal grant funding.

5. Proposal timeline (June 1, 2022 – May 31, 2023). This project requires 12-months' time to achieve the proposed specific aims. The tentative timelines to achieve the proposed specific aims are as follows:

Months 0-3: Development of resource material and launching of the website. The website will be launched by the end of the third month.

Months 3:6: Development and testing of computational pipeline. The pipeline will be available on the website for free downloads by the end of the sixth month.

Months 6:9: Development of collaboration. Within this period, we will assist two identified collaborators to help develop single-cell QTL mapping projects of their interest. In addition, we will also seek and work with other agricultural researchers interested in developing such mapping projects.

Months 9:12: Providing training. Within this period, we will provide online bioinformatics training in interactive and self-paced learning formats to students, postdocs or investigators interested to learn data analysis in single-cell QTL mapping.

6. Engaging AG2P scientific communities & underrepresented groups. This project will provide necessary informational resources as well as data analytics training to any member with the AG2P scientific community, and others, to learn about single-cell QTL mapping. In addition, this project will be a great opportunity for the University of Missouri initiatives on *Maximize Student Diversity (IMSD)* and *Maximizing Access to Research Careers (MARC)* to engage underrepresented students in research activities.

Bibliography/References cited

- Behura SK, Nair S, Sahu SC, Mohan M. An AFLP marker that differentiates biotypes of the Asian rice gall midge (*Orseolia oryzae*, Wood-Mason) is sex-linked and also linked to avirulence. *Mol Gen Genet* 2000;**263**:328–334.
- Behura SK, Valicente FH, Rider SD, Shun-Chen M, Jackson S, Stuart JJ. A physically anchored genetic map and linkage to avirulence reveals recombination suppression over the proximal region of Hessian fly chromosome A2. *Genetics* 2004;**167**:343–355.
- Benaglio P, Newsome J, Han JY, Chiou J, Aylward A, Corban S, Okino M-L, Kaur J, Gorkin DU, Gaulton KJ. Mapping genetic effects on cell type-specific chromatin accessibility and annotating complex trait variants using single nucleus ATAC-seq. 2020a;2020.12.03.387894. bioRxiv Available from: <https://www.biorxiv.org/content/10.1101/2020.12.03.387894v1>.
- Benaglio P, Newsome J, Han JY, Chiou J, Aylward A, Corban S, Okino M-L, Kaur J, Gorkin DU, Gaulton KJ. Mapping genetic effects on cell type-specific chromatin accessibility and annotating complex trait variants using single nucleus ATAC-seq. 2020b;2020.12.03.387894. bioRxiv Available from: <https://www.biorxiv.org/content/10.1101/2020.12.03.387894v1>.
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;**36**:411–420.
- Ciuculete DM, Boström AE, Voisin S, Philipps H, Titova OE, Bandstein M, Nikontovic L, Williams MJ, Mwinyi J, Schiöth HB. A methylome-wide mQTL analysis reveals associations of methylation sites with GAD1 and HDAC3 SNPs and a general psychiatric risk score. *Transl Psychiatry* 2017;**7**:e1002–e1002. Nature Publishing Group.
- Cole B, Bergmann D, Blaby-Haas CE, Blaby IK, Bouchard KE, Brady SM, Ciobanu D, Coleman-Derr D, Leiboff S, Mortimer JC, *et al.* Plant single-cell solutions for energy and the environment. *Commun Biol* 2021;**4**:1–12. Nature Publishing Group.
- Craig JE, Hewitt AW, McMellon AE, Henders AK, Ma L, Wallace L, Sharma S, Burdon KP, Visscher PM, Montgomery GW, *et al.* Rapid inexpensive genome-wide association using pooled whole blood. *Genome Res* 2009;**19**:2075–2080.
- Cui Y, Behura SK, Franz AWE. Cellular diversity and gene expression profiles in the male and female brain of *Aedes aegypti*. *BMC Genomics* 2022;**23**:119.
- Cuomo ASE, Alvari G, Azodi CB, McCarthy DJ, Bonder MJ, single-cell eQTLGen consortium. Optimizing expression quantitative trait locus mapping workflows for single-cell studies. *Genome Biology* 2021;**22**:188.
- Dhakal P, Strawn M, Samal A, Behura SK. Fetal Brain Elicits Sexually Conflicting Transcriptional Response to the Ablation of Uterine Forkhead Box A2 (*Foxa2*) in Mice. *International Journal of Molecular Sciences* 2021;**22**:9693. Multidisciplinary Digital Publishing Institute.
- Fitzgerald HC, Dhakal P, Behura SK, Schust DJ, Spencer TE. Self-renewing endometrial epithelial organoids of the human uterus. *Proc Natl Acad Sci U S A* 2019;**116**:23132–23142.

- Gaj P, Maryan N, Hennig EE, Ledwon JK, Paziewska A, Majewska A, Karczmarski J, Nesteruk M, Wolski J, Antoniewicz AA, *et al.* Pooled Sample-Based GWAS: A Cost-Effective Alternative for Identifying Colorectal and Prostate Cancer Risk Variants in the Polish Population. *PLOS ONE* 2012;**7**:e35307. Public Library of Science.
- Hu Z-L, Park CA, Reecy JM. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res* 2022;**50**:D956–D961.
- Huang L, Tang W, Wu W. Optimization of BSA-seq experiment for QTL mapping. *G3 (Bethesda)* 2021;**12**:jkab370.
- Keele GR, Quach BC, Israel JW, Chappell GA, Lewis L, Safi A, Simon JM, Cotney P, Crawford GE, Valdar W, *et al.* Integrative QTL analysis of gene expression and chromatin accessibility identifies multi-tissue patterns of genetic regulation. *PLOS Genetics* 2020;**16**:e1008537. Public Library of Science.
- Kim C-K, Seol Y-J, Lee D-J, Lee J-H, Lee T-H, Park D-S. RiceQTLPro: an integrated database for quantitative trait loci marker mapping in rice plant. *Bioinformatics* 2014;**10**:664–666.
- Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet* 2016;**48**:206–213.
- Kushanov FN, Turaev OS, Ernazarova DK, Gapparov BM, Oripova BB, Kudratova MK, Rafieva FU, Khalikov KK, Erjigitov DSh, Khidirov MT, *et al.* Genetic Diversity, QTL Mapping, and Marker-Assisted Selection Technology in Cotton (*Gossypium* spp.). *Frontiers in Plant Science* [Internet] 2021;**12**:.
- Liu Y, Liu X, Zheng Z, Ma T, Liu Y, Long H, Cheng H, Fang M, Gong J, Li X, *et al.* Genome-wide analysis of expression QTL (eQTL) and allele-specific expression (ASE) in pig muscle identifies candidate genes for meat quality traits. *Genet Sel Evol* 2020;**52**:59.
- Long Y, Xia W, Li R, Wang J, Shao M, Feng J, King GJ, Meng J. Epigenetic QTL Mapping in *Brassica napus*. *Genetics* 2011;**189**:1093–1102.
- Navarro-Escalante L, Zhao C, Shukle R, Stuart J. BSA-Seq Discovery and Functional Analysis of Candidate Hessian Fly (*Mayetiola destructor*) Avirulence Genes. *Frontiers in Plant Science* [Internet] 2020;**11**:.
- Neavin D, Nguyen Q, Daniszewski MS, Liang HH, Chiu HS, Wee YK, Senabouth A, Lukowski SW, Crombie DE, Lidgerwood GE, *et al.* Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biology* 2021;**22**:76.
- Ni J, Pujar A, Youens-Clark K, Yap I, Jaiswal P, Teclé I, Tung C-W, Ren L, Spooner W, Wei X, *et al.* Gramene QTL database: development, content and applications. *Database (Oxford)* 2009;**2009**:bap005.
- Nyysölä A, Suhonen A, Ritala A, Oksman-Caldentey K-M. The role of single cell protein in cellular agriculture. *Current Opinion in Biotechnology* 2022;**75**:102686.
- Riaz M, Lorés-Motta L, Richardson AJ, Lu Y, Montgomery G, Omar A, Koenekoop RK, Chen J, Muether P, Altay L, *et al.* GWAS study using DNA pooling strategy identifies association of

variant rs4910623 in OR52B4 gene with anti-VEGF treatment response in age-related macular degeneration. *Sci Rep* 2016;**6**:37924. Nature Publishing Group.

Singh K, Batra R, Sharma S, Saripalli G, Gautam T, Singh R, Pal S, Malik P, Kumar M, Jan I, *et al.* WheatQTLdb: a QTL database for wheat. *Mol Genet Genomics* 2021;**296**:1051–1056.

Tripathi RK, Wilkins O. Single cell gene regulatory networks in plants: Opportunities for enhancing climate change stress resilience. *Plant, Cell & Environment* 2021;**44**:2006–2017.

Volkov P, Olsson AH, Gillberg L, Jørgensen SW, Brøns C, Eriksson K-F, Groop L, Jansson P-A, Nilsson E, Rönn T, *et al.* A Genome-Wide mQTL Analysis in Human Adipose Tissue Identifies Genetic Variants Associated with DNA Methylation, Gene Expression and Metabolic Traits. *PLoS One* 2016;**11**:e0157776.

Wijst M van der, Vries D de, Groot H, Trynka G, Hon C, Bonder M, Stegle O, Nawijn M, Idaghdour Y, Harst P van der, *et al.* The single-cell eQTLGen consortium. In Pérez Valle H, Rodgers P, Montgomery SB, Fagny M, editors. *eLife* 2020;**9**:e52155. eLife Sciences Publications, Ltd.

Xu Y, Li J, Yang W, Tang X, Huang B, Liu J, Lin J, Zhang J, Yang W, Li S, *et al.* A Pooling Genome-Wide Association Study Identifies Susceptibility Loci and Signaling Pathways of Immune Thrombocytopenia in Chinese Han Population. *International Journal of Genomics* 2020;**2020**:e7531876. Hindawi.

Yuan Z, Sunduimijid B, Xiang R, Behrendt R, Knight MI, Mason BA, Reich CM, Prowse-Wilkins C, Vander Jagt CJ, Chamberlain AJ, *et al.* Expression quantitative trait loci in sheep liver and muscle contribute to variations in meat traits. *Genet Sel Evol* 2021;**53**:8.

Zeng H, Luo L, Zhang W, Zhou J, Li Z, Liu H, Zhu T, Feng X, Zhong Y. PlantQTL-GE: a database system for identifying candidate genes in rice and Arabidopsis by gene expression and QTL information. *Nucleic Acids Res* 2007;**35**:D879-882.

Zhang J, Panthee DR. PyBSASeq: a simple and effective algorithm for bulked segregant analysis with whole-genome sequencing data. *BMC Bioinformatics* 2020;**21**:99.

Zhang T-Q, Chen Y, Liu Y, Lin W-H, Wang J-W. Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat Commun* 2021;**12**:2053. Nature Publishing Group.

Zhao C, Shukle R, Navarro-Escalante L, Chen M, Richards S, Stuart JJ. Avirulence gene mapping in the Hessian fly (*Mayetiola destructor*) reveals a protein phosphatase 2C effector gene family. *J Insect Physiol* 2016;**84**:22–31.

Zhao Q, Dacre M, Nguyen T, Pjanic M, Liu B, Iyer D, Cheng P, Wirka R, Kim JB, Fraser HB, *et al.* Molecular mechanisms of coronary disease revealed using quantitative trait loci for TCF21 binding, chromatin accessibility, and chromosomal looping. *Genome Biology* 2020;**21**:135.

Zhu Y, Li J, Lin X, Huang X, Hoffmann MR. Single-Cell Phenotypic Analysis and Digital Molecular Detection Linkable by a Hydrogel Bead-Based Platform. *ACS Appl Bio Mater* 2021;**4**:2664–2674.