Introduction

In multicellular organisms, the development of cell types, tissues, and organs is a central component in the regulation of metabolism. For diverse metabolites to exert their biological functions in facilitating growth and environmental interactions, their synthesis and accumulation must be controlled by development, such that metabolites are synthesized and stored at physiologically relevant organs, tissues, and cell types. *My research program aims to understand how plant genetics, epigenome, and development generate diverse metabolic profiles across cell types, organs, and populations.* Decoding how (epi)genome and development shape plant metabolism will lead to transformative discoveries pertinent to biotechnology and sustainable bioeconomy. For example, mechanistic insights of developmental regulation of metabolism will enable the redesign of both development and metabolism in multicellular organisms for sustainable biomanufacturing at scale.

Ongoing Research

1) Genome-to-pathway discovery pipeline using single cell omics. My postdoctoral research focuses on the medicinal plant *Catharanthus roseus* (Madagascar Periwinkle), which produces the chemotherapeutic agent vinblastine. Vinblastine belongs to a diverse class of specialized metabolites termed monoterpene indole alkaloids that includes several additional clinically relevant compounds. Decades of research on vinblastine biosynthesis has identified close to 40 biosynthetic steps. However, the final steps of vinblastine biosynthesis have remained a mystery. In the Buell laboratory, I led a project aiming to solve the final stage of vinblastine biosynthesis using a genome-to-pathway approach (Li et al., 2023, *Nature Chemical Biology,* Figure 1).

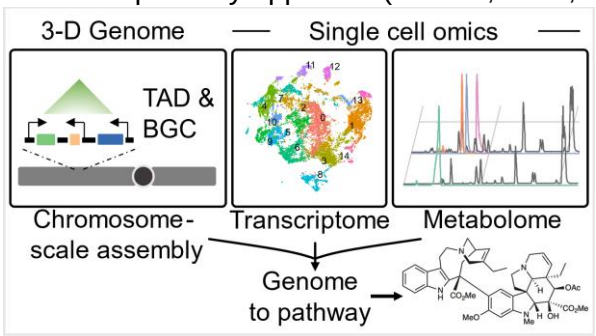

Figure 1: Graphical abstract for Li et al., (2023. *Nature Chem. Bio*). Long range chromosome interactions (e.g., topologically associated domain (TAD)) and biosynthetic gene clusters (BGC) identified candidate biosynthetic genes and transporters physically adjacent to known biosynthetic genes. Single cell omics led to the discovery of biosynthetic genes enriched in rare cell types associated with vinblastine biosynthesis.

After producing a gold standard genome assembly, I discovered rampant tandem duplications of biosynthetic genes, which are often organized into higher-order chromatin domains. By detecting long-range chromosome interactions linking novel genes to previously characterized biosynthetic genes, I discovered a new transporter for the pathway, as well as several candidate biosynthetic genes. Since specialized metabolism is often confined to rare and specialized cell types (to minimize auto-toxicity, for example), I applied single cell RNA-seq to investigate the expression pattern of biosynthetic genes. Such an approach prevents the dilution of rare cell types by more abundant cell types. I found that the entire vinblastine biosynthetic pathway is sequentially organized into three discrete cell types, and that the final known steps of the pathway are exclusively expressed in a rare cell type termed idioblast. Cell-type-specific accumulation of late-stage metabolic intermediates at the idioblast supported the inference that final steps of vinblastine biosynthesis occur exclusively in the idioblast. Integrating the above-mentioned datasets, I identified the missing enzymes that catalyze the formation of anhydrovinblastine, the direct precursor to vinblastine. This study represents a major advancement in the field of plant natural product research, showcasing the potential of single cell assisted genome-to-pathway discovery pipelines. Since the publication of this study, my lab and I have extended the single cell assisted genome-to-pathway discovery pipeline to other plant species that produce structurally related natural products (NSF: MCB-2309665 and NIH: R01AT012783) for which I am co-PI or Senior/Key personnel. Reconstitution of alkaloid biosynthesis across multiple species will enable combinatorial engineering of natural products, such as mix-and-matching biosynthetic genes from different species to generate new-to-nature products with enhanced or novel biological activities.

2) Cell-type-specific regulation of natural product biosynthesis. The mechanisms by which plants restrict the expression of biosynthetic genes to distinct cell types is not well understood. Currently, very little is known on how the final steps of vinblastine biosynthesis are regulated. To uncover how different cell types regulate biosynthetic genes, I applied single cell multiome (i.e., RNA-seq and ATAC-seq from the same nucleus) to profile both gene expression and accessible chromatin landscapes from the nuclei of *C. roseus* leaves (Li et al., 2024, *New Phytologist*). I discovered multiple cases in which the cis-

regulatory elements of cell-type-specific biosynthetic genes are exclusively accessible in the corresponding cell types. I catalogued a dictionary of *cis*-regulatory elements that may be involved in the regulation of vinblastine biosynthesis. Integrating overrepresented transcription factor (TF) binding motifs at accessible chromatin regions and cell type specific expression, I detected candidate TFs that may act as metabolic regulators in the idioblast, which I tested *in planta* by overexpression and RNA-seq. In particular, I discovered an idioblast specific MYB TF (*Idioblast MYB1*, *IDM1*) that controls cell type specific expression of late-stage vinblastine biosynthetic genes. Transporters and other putative biosynthetic genes that are co-regulated by this idioblast MYB TF are being investigated for their potential roles in natural product biosynthesis. This study is first-of-its-kind in applying chromatin accessibility and gene expression across single cells to discover new TFs relevant to plant specialized metabolism. Taking this approach, I discovered an idioblast specific TF that regulates vinblastine biosynthesis. While the exact biological function of such cell type specificity is unclear, regulatory components (TFs and cis-regulatory elements) described in this study provide new opportunities for studying how cell fates control specialized metabolism and how cell type specificity of metabolism contribute to stress tolerance (see below).

Future Research

Theme 1: The development of rare cell types underlying specialized metabolism. An emerging theme of specialized metabolism is that biosynthesis and metabolite accumulation is confined to rare cell types. However, the developmental biology (e.g., cell fate determining factors, patterning, differentiation) of these rare cell types is poorly understood. Given their critical roles in specialized metabolism, understanding and harnessing these elusive cell types is of both fundamental and practical importance.

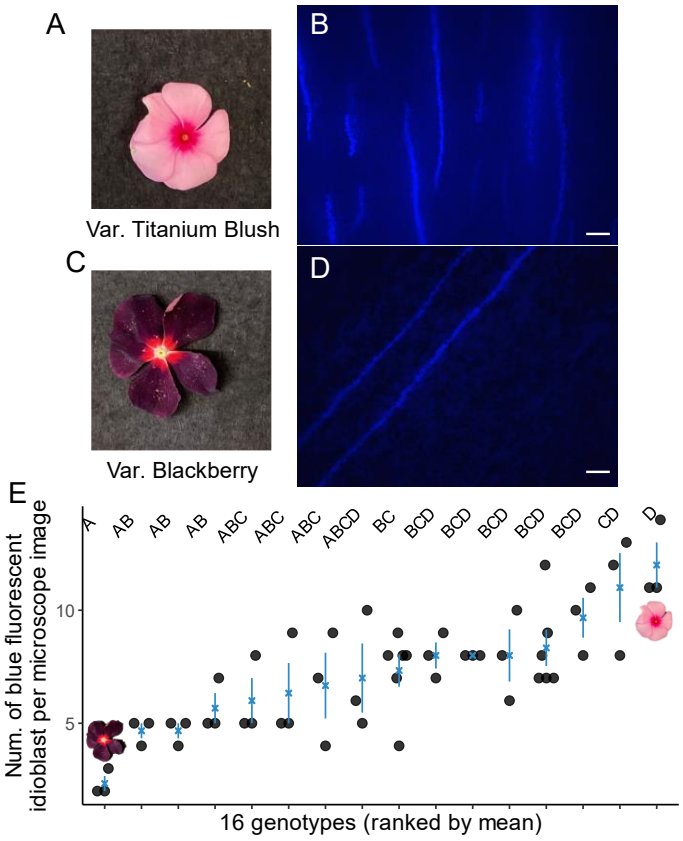

Figure 2: Diversity of idioblast densities. A: Titanium Blush is a high idioblast variety; B: Numerous idioblast cells (elongated, fluorescent cells) on Titanium Blush petals. C. Blackberry is a low idioblast variety. D: Very few idioblast (2 in this example) observed on Blackberry petals. E. Quantification of idioblast density across a sample of 16 genotypes (mean ± SE). Titanium Blush and Blackberry are marked. Bar in B and D: 40 μm.

Idioblast cells in *C. roseus* provide a unique and powerful model for probing metabolic regulation by development in rare cell types: 1) My colleagues and I have developed genomics and single cell omics resources for *C. roseus*, and an overview of idioblast transcriptome and accessible chromatin landscape is available. 2) Idioblast cells are easy to observe in leaves and petals due to their large size, elongated morphology, and blue autofluorescence (Figure 2). 3) There is substantial variation in idioblast density among *C. roseus* varieties, which can be coupled with genetic diversity to map loci underlying idioblast development (Figure 2). Comparing the genomes of three *C. roseus* varieties, I found 2.7-5 million SNPs (0.5%-1% of genome size) relative to the reference genome, suggestive of ample genetic diversity. 4) *C. roseus* petals are amenable to transient expression, permitting functional characterization via overexpression and gene silencing.

To investigate the developmental biology of idioblast and its implications in metabolism, my lab will take a four-pronged approach: 1) We will perform a genome-wide association study using a diversity panel of *C. roseus* (> 200 registered varieties available commercially and via USDA). 2) We will create an F2 mapping population between a high idioblast parent and a low idioblast parent to map loci underlying idioblast density. 3) We will perform a yeast-one-hybrid screen on accessible chromatin regions associated with *IDM1* and idioblast specific biosynthetic genes against a collection of TFs that are also idioblast specific.

4) We will perform promoter proximity labeling experiments by targeting the promiscuous biotin ligase TurboID to promoters of *IDM1* and idioblast specific biosynthetic genes using an enzymatically dead Cas9. Proteins in the TF assemblages will be biotinylated, which enables affinity enrichment and detection by proteomics.

We will validate hits using genetic manipulation (e.g., overexpression, gene silencing) followed by idioblast phenotyping and metabolite profiling. These experiments will reveal the genetic architecture of idioblast development and new regulators of monoterpene indole alkaloid biosynthesis. In addition, insights generated by this study will enable breeding or synthetic biology approaches for biomanufacturing, such as high alkaloid high idioblast genotypes and cell cultures with idioblast characteristics.

Theme 2: The interplay between cell fate, epigenome, and metabolism. Metabolic processes in multicellular organisms are highly partitioned across organs, tissues, and cell types, suggesting biosynthetic genes are under strong developmental control. However, very little is known about how different cell types control metabolic processes via the combination of gene activation and repression.

My lab will use two model systems that are uniquely suited for this research: *Arabidopsis thaliana* and *C. roseus* - two distantly related species with independently evolved natural products as foliar herbivory deterrent. Arabidopsis produces glucosinolates whose biosynthesis and regulation have been documented in specific organs and cell types. The ease of genetic engineering makes Arabidopsis the perfect system for generating an array of defined genotypes to test developmental regulation, either by overexpressing TFs that activates the pathway and/or mutating mechanisms that repress the pathway. *C. roseus* is a powerful model for metabolic regulation at the single cell level. Several *C. roseus* TFs that activate different stages of alkaloid biosynthesis have been characterized. Since *C. roseus* petals are amenable to transient expression, ectopic activation of different stages of alkaloid biosynthesis is possible. These independently evolved pathways, regulators, and associated biosynthetic cell types allow us to investigate what common features may exist among cell types recruited to host novel metabolic pathways, as well as any potential conserved mechanisms that exclude specialized metabolism from conserved 'housekeeping' cell types such as mesophyll.

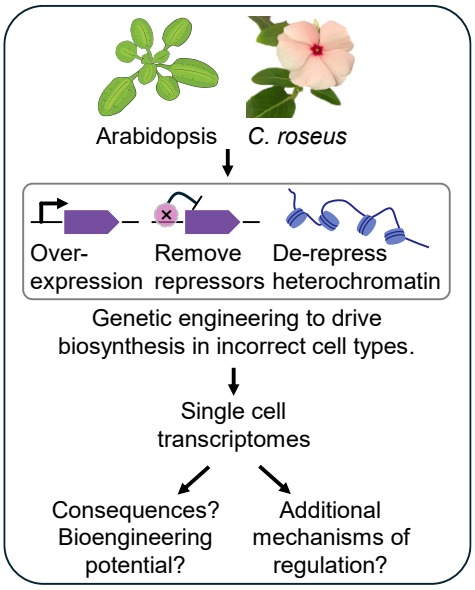

Figure 3: A project to investigate the interplay between cell fate, heterochromatin, and specialized metabolism using genetic engineering and single cell genomics.

For both species, we will generate single cell transcriptomes to reveal how different cell types accommodate or restrict the ectopic activation of biosynthetic pathways. To do so, we will engineer Arabidopsis and *C. roseus* to drive mis-expression of glucosinolate or alkaloid pathways in incorrect cell types and then characterize these tissues using single cell RNA-seq (Figure 3). This will also be done in the presence of mutants of known repressor TFs, defective Polycomb complexes, and/or constructs that lead to loss of Polycomb repression. This study will generate insights on the interplay between cell fate and metabolism, as well as whether perturbations of the epigenome permit ectopic biosynthesis at cell types where biosynthesis is normally prohibited. Understanding how development and epigenome restrict specialized metabolism will provide vital insights for bioengineering efforts across diverse multicellular organisms, including economically important crops and medicinal plants.

Summary. My research program consists of two themes. Both proposed themes build on my expertise in genomics and gene regulation and will expand in new areas such as quantitative genetics, synthetic biology, and proteomics. *Both themes will address how (epi)genome and development shape metabolism*. A deeper understanding on metabolic regulation by development will open opportunities for redesigning plant development and metabolism to create distinct plant morphotypes (plants with different architectures and cell type compositions) and chemotypes tailored to specific applications in the bioeconomy, in addition to generalizable principles of metabolic regulation under healthy and pathological conditions beyond plants.