Community Science Program 2026 Large Scale Proposal

Proposer's Name: Hao Chen, Jack Wang, Miaoiao Li, Zongliang Yue

Proposal Title: Linking Lignin Engineering to drought Stress Resilience: RNA-Seq and Metabolomic Insights in Lignin Biosynthetic Pathway Edited Populus

Proposal (WIP) ID: 511618

A) Brief description

Abstract: *Populus trichocarpa*, the first tree with a published reference genome, serves as an indispensable model for lignocellulosic bioenergy research. Leveraging a uniform wild-type background, we have generated 20 CRISPR/Cas9-edited lines targeting seven monolignol biosynthesis genes (*C3H3, CAD1, 4CL1, CCoAOMT1, COMT, PAL2, CAD2*), producing a validated panel of 21 genotypes with graded reductions in total lignin (up to 49 %) and altered S/G ratios (2.7 → 4.0) without compromising biomass. This project will dissect how precise lignin modifications influence drought resilience by integrating: (1) tissue-specific multi-omics—deeply sequenced transcriptomes (756 RNA-seq libraries), untargeted metabolomes (189 LC-MS profiles), high-resolution lignin compositional analyses (2D HSQC NMR), and physiological measurements (RWC, Ψ_leaf, gas exchange, hydraulics); (2) AI-driven regulatory network reconstruction—combining DESeq2, ArchR/Cicero, WGCNA, and graph-neural-network frameworks (MOGONET, LINGER, VCDN) to infer gene–metabolite–chromatin circuits in root, leaf, and xylem; and (3) predictive modeling—machine-learning (Random Forest, SVM, PLSR, SEM) to identify master regulators and biomarkers underlying drought tolerance and lignin properties. By subjecting clones to controlled desiccation and recovery in a greenhouse trial, we will generate a rigorously controlled, metadata-rich dataset. Our multidisciplinary team—spanning Auburn University, NC State, ORNL, and JGI—will deploy standardized pipelines for sample preparation, sequencing, and data integration. The resulting systems-level insights will not only elucidate fundamental mechanisms of drought adaptation and cell-wall regulation in a perennial woody species but also accelerate the development of *Populus* feedstocks that combine low-recalcitrance lignin with robust climate resilience, directly advancing sustainable bioenergy and forestry practices.

Scope of Work; **RNA-Sequencing (JGI):** We request transcriptome sequencing of 756 Populus trichocarpa samples (4 tissues × 3 treatments × 3 replicates × 21 genotypes), each as 150 bp stranded mRNA libraries at ~40 M reads/sample on NovaSeq. JGI will perform standard QC, alignment to the reference genome, and gene-count generation. **Metabolomics (JGI):** Following transcriptomic analyses, we will select 7 genotypes (six top-performing CRISPR-edited lines plus wild type) for untargeted LC-MS profiling of 189 samples (root, leaf, xylem as three tissue; three treatments × three replicates). Metabolic hypotheses driving these assays include altered abundance and flux through monolignol intermediates (e.g., coniferyl and sinapyl alcohols, ferulates), stress hormones (abscisic, salicylic, and jasmonic acids), and osmoprotectants (proline, glycine betaine, flavonoid glycosides). We ask JGI to perform dual-polarity LC-MS with internal–standard normalization and feature extraction via its standard metabolomics pipeline, delivering retention-time aligned feature tables and putative annotations. **Justification:** The scale and integration demands of this project—756 high-depth transcriptomes coupled with 189 untargeted metabolomes—exceed local core capacities. JGI's high-throughput sequencing infrastructure, standardized data-processing pipelines, and expertise in large-scale omics make it uniquely qualified to deliver uniformly processed, quality-controlled datasets. Similarly, JGI's metabolomics platform ensures rigorous QC (internal standards, retention-time alignment, TIC reproducibility) critical for cross-tissue and cross-genotype comparisons. While local and commercial services can generate individual datasets, only JGI can provide the end-to-end, integrated omics workflows and DOE-backed informatics support required to uncover robust, multi-layered insights into Populus drought resilience. Custom downstream analyses—including network reconstruction, machine-learning, and regulatory-element mapping—will be performed by the investigator team. Additionally, JGI could optimize the protocol based on the sample properties.

B) Background information

Technical Information: This proposal focuses on *Populus trichocarpa*, a perennial woody plant and foundational model system for lignocellulosic bioenergy crop research. The species was the first tree to have its genome sequenced and published (Tuskan et al., 2006), offering a high-quality reference genome of ~480 megabases with a GC content of approximately 33.8%. The genome architecture includes a complex repeat structure, comprising over 40% repetitive elements, predominantly long terminal repeat (LTR) retrotransposons, which pose both opportunities and challenges for transcriptome assembly and mapping. Despite this, the genome is well-annotated, with substantial transcriptomic and epigenomic resources available through platforms such as Phytozome and JGI's Genome Portal.

In terms of genetic diversity, *Populus trichocarpa* exhibits moderate polymorphism levels across natural populations, particularly within genes related to secondary cell wall biosynthesis and abiotic stress responses. The edited lines proposed for analysis in this project were derived from a uniform wild-type background using CRISPR/Cas9 genome editing. Targeted knockouts focused on key genes in the monolignol biosynthesis pathway—*C3H3, CAD1, 4CL1, CCoAOMT1, COMT, PAL2, and CAD2*—to generate a panel of hundreds genetically modified genotypes with diverse lignin content and composition profiles (Sulis et al., 2023). 21 genetically varieties (Table 1) were validated with altered lignin composition, structure, and content and enhanced pulp efficiency (Sulis et al., 2023) and biofuel production capacity (Bing et al., 2024), thereby enabling investigation of downstream effects on drought resilience in this study. Each variety was clonally propagated from a single-cell origin (protoplast based method), ensuring genetic homogeneity and stability. No background introgression or wild-type heterozygosity is expected to confound gene expression analyses, thus enabling precise attribution of molecular responses to specific gene edits.

Samples	Total sugars % of wt	Lignin % of wt	C/L ratio % of wt	Mass balance	Height (cm)	Height % of wt	Diameter (cm)	Diameter % of wt	Genotype
WT batch average	100.0	100.0	100.0	86.6	326.0	100.0	13.5	100.0	wild-type trees
K4-7-1	81.4	77.7	104.6	69.7	325.0	99.7	12.4	92.1	*C3H3 CAD1 C4H1 Knockout*
K4-4-3	86.3	69.4	124.3	70.9	285.0	87.4	11.1	82.4	*C3H3 AldOMT2 CAD1 CCoAOMT1 knockout*
K5-29-1	90.8	68.1	133.3	73.5	310.0	95.1	11.4	84.7	*C3H3 CAD1 C4H1 Knockout*
K4-7-2	85.6	84.7	100.9	73.9	335.0	102.8	12.6	93.6	*C3H3 CAD1 C4H1 CCoAOMT1 Knockout*
K6-14-1	91.0	77.7	117.0	75.8	300.0	92.0	10.6	78.7	*C3H3 CAD1 CCoAOMT1 CAD2 knockout*
K3-18-2	95.2	69.9	136.1	76.6	240.0	73.6	12.0	89.1	*C3H3 CAD1 AldOIMT1 Knockout-1*
K5-53-1	91.5	98.3	92.9	80.8	348.0	106.7	14.4	106.9	*C3H3 C4H1 CCoAOMT1 Knockout*
K5-71-2	96.7	85.2	113.4	81.2	325.0	99.7	12.5	92.8	*C3H3 C4H1 AldOIMT1 CCoAOMT1 Knockout*
K4-9-1	99.4	78.4	126.7	81.3	285.0	87.4	14.9	110.6	*C3H3 CAD1 C4H1 AldOIMT1 Knockout-1*
OC5KP-19-1	99.5	100.8	98.6	82.1	341.0	96.0	12.6	87.9	*PAL2PAL4PAL5 knockout*
K5-58-1	101.7	80.0	126.9	83.1	310.0	95.1	12.5	92.8	*C3H3 CAD1 C4H1 AldOIMT1 CCoAOMT1 Knockout*
OKCD1-17-1	98.1	114.7	85.4	83.9	280.0	91.3	15.9	106.2	*C3H3 AldOMT2 CAD1,2 CCoAOMT1,2 Knockout-1*
K6-21-3	113.5	52.4	216.2	84.2	235.0	72.1	10.5	78.0	*3H3 CAD1 AldOIMT1 CCoAOMT1, 2 CAD2 Knockout*
K4-2-3	114.2	52.4	217.6	84.7	262.0	80.4	10.3	76.5	*C3H3 CAD1 C4H1 AldOIMT1 Knockout-2*
K6-2-3	99.9	92.3	108.1	84.8	282.0	86.5	10.7	79.5	*C3H3 CAD1 AldOIMT1 CCoAOMT1 CAD2 Knockout-1*
K4-25-1	111.7	82.5	134.1	90.0	255.0	78.2	11.9	88.4	*C3H3 CAD1 AldOIMT1 C4H1 Knockout-2*
K5-50-3	115.0	75.2	152.8	90.4	295.0	90.5	12.3	91.3	*C4H1 CCoAOMT1 Knockout-1*
K6-48-1	112.6	84.8	132.6	91.1	245.0	75.2	9.2	68.3	*C3H3 CCoAOMT1/2 CAD2*
K5-19-3	114.0	84.1	134.1	91.9	290.0	89.0	12.1	89.9	*CAD1 C4H11 CCoAOMT1 Knockout-1*
K6-31-1	117.5	82.0	143.2	93.6	255.0	78.2	14.4	106.9	*AldOIMT1 CCoAOMT1 CAD1/2 Knockout*

Table 1. The lignin chemistry and morphological characteristics of wild-type and 20 CRISPR-based lignin-modified poplar varieties selected for this project analyses. The table presents data on total sugar content, lignin content, C/L ratio, mass balance, height, and diameter for different edited poplar varieties.

Throughout the duration of the study, all genotypes will be maintained under standardized and tightly controlled environmental conditions in greenhouse facilities at Auburn University. Plants will be grown in 3-gallon containers with a uniform soil substrate, under

a 16-hour photoperiod and controlled temperature and humidity conditions. A progressive soil desiccation method will be employed to impose drought stress uniformly across genotypes, a method previously validated to simulate natural water-deficit conditions (Galvez et al., 2011). Parallel cohorts of each genotype will be maintained under well-watered control conditions, allowing for robust paired comparisons. All tissues will be harvested at the defined peak stress timepoint, immediately flash-frozen in liquid nitrogen, and stored at –80°C to preserve molecular integrity for RNA and metabolite analyses.

Available Resources: The proposed work will leverage extensive existing infrastructure and funding support. Complementary resources have been secured through a USDA National Institute of Food and Agriculture (NIFA) grant and institutional start-up packages. These funds support the cultivation of edited poplar lines, implementation of controlled drought experiments, and the preparation and quality control of RNA and metabolite extracts. Initial bioinformatics processing will be conducted at Auburn University, while transcriptomic sequencing and metabolomic analyses will be supported by JGI. Key personnel across multiple institutions bring critical domain expertise. Dr. Hao Chen (Auburn University) will oversee plant growth, sample collection, and initial network analysis after RNA-seq results generate. Dr. Jack Wang (NC State University) will lead interpretation of lignin biosynthesis gene function and metabolomic analysis. Dr. Miaomiao Li (ORNL) will support regulatory network inference, while Dr. Zongliang Yue (Auburn University) will contribute machine learning-based feature selection and modeling. Collectively, the team is well-positioned to deliver high-quality, interpretable multi-omics datasets.

Technical Challenges and Mitigation Strategies: **One potential bottleneck is the extraction of high-quality RNA from lignified tissues such as xylem**. These tissues are notoriously challenging due to secondary wall deposition and co-extracted polyphenolic compounds. To mitigate this, the team will employ the Qiagen RNeasy Plant Mini Kit and Zymo RNA concentrator kit, supplemented with PVPP and optimized buffer conditions tailored for high-phenolic content samples. RNA integrity will be assessed using the Agilent 2100 Bioanalyzer (RIN ≥7.5 required), and quantification will be performed via Qubit fluorometry. **To address challenges associated with the metabolomics pipeline,** especially with respect to the broad dynamic range of small molecules in stressed woody tissues, we will use a dual-polarity LC-MS approach. Samples will undergo methanol:water:chloroform extraction followed by nitrogen drying and solvent reconstitution. Rigorous QC (e.g., internal standards, retention time alignment, TIC reproducibility) will ensure analytical robustness. **Secondly, high sample throughput and batch effects may be another potential pitfall.** The proposed experimental design involves profiling 20 CRISPR-edited *Populus* genotypes and wild-type controls under drought and well-watered conditions, across 4 tissue types, with three replicates generating several hundred RNA-seq libraries and a following selected set of metabolomic samples. This high sample throughput introduces risks of batch effects in both nucleic acid extraction and LC-MS workflows, which could obscure true biological signals. **To mitigate this**, all RNA and metabolite extractions will be conducted in randomized and balanced blocks to prevent alignment of batch with genotype or treatment. Further, JGI's standardized transcriptomic processing pipeline will enhance consistency across libraries, while downstream statistical modeling will include batch covariates to account for residual noise. **Thirdly, gene expression variation could be affected by environmental heterogeneity.** Even in climate-controlled greenhouse settings, subtle environmental

gradients—such as light intensity, air circulation, or differential soil moisture retention—can introduce spatial heterogeneity that impacts gene expression and metabolite accumulation. These microenvironmental effects pose a challenge for dissecting drought-specific responses. **To control this**, all plants will be randomized spatially within the greenhouse and rotated weekly to ensure even exposure to environmental variables. Soil moisture levels will be continuously monitored using gravimetric tracking and in-pot soil sensors to ensure uniform drought stress across all genotypes. Additionally, physiological stress indicators, such as predawn leaf water potential and stomatal conductance, will be collected to quantify actual drought exposure. **Last bottleneck could be the complexity of multi-omics data integration**. The integration of high-dimensional transcriptomic and metabolomic data across multiple genotypes and environmental treatments presents a significant analytical challenge, particularly with respect to identifying robust gene–metabolite associations and minimizing false discovery. Biological variation, technical noise, and complex interactions between pathways require rigorous statistical frameworks to extract meaningful insights. To address this, machine learning algorithms such as Random Forests, Support Vector Machines (SVM), and Elastic Net regression will be employed with embedded cross-validation and regularization to reduce overfitting and prioritize biologically meaningful features. Network-based methods including weighted gene co-expression network analysis (WGCNA) for transcriptomes and correlation-based clustering for metabolites will be used to define modules and identify hub features.

Starting Materials: All biological materials—**four** tissues (root, leaves, xylem, and apical bud) from 20 CRISPR-edited lines and wild-type controls (**21** genetical varieties)—will be harvested under identical developmental and treatment conditions. **Three** biological replicates across **three** conditions: No drought treatment, drought treatment, and post-drought recovery. Total **756** (21 varieties ×4 tissues×3 conditions×3 biological replicates) tissues will be flash-frozen at the time of collection and stored at −80°C until processing. RNA will be extracted as detailed above, meeting or exceeding JGI's requirements for RNA quantity (≥2 μg), purity (A260/A280 ≈ 2.0), and integrity (RIN ≥7.5). Each could be sequenced to ~40 million 150 bp single-end reads. Sample preparation will be completed by October 2025, with all materials ready for shipment to JGI in November 2025. Coordination for metadata submission and logistics will begin one month prior. Based on the RNA-seq results analyzed, we will select 6 CRIPSR-edited lines and wild-type control (7 genetic varieties) out for metabolite analyses. For metabolite analyses, **three** tissues (root, leaves, xylem) will be proposed due to the limited amount of apical bud. Therefore, for total **189** (7 varieties ×3 tissues×3 conditions×3 biological replicates) tissues, their metabolites will be extracted from lyophilized, cryo-ground tissues under strict QC procedures. No whole-genome sequencing or synthetic DNA work is proposed under this submission. However, genomic DNA can be prepared from the same genotypes using high molecular weight extraction protocols if needed for future long-read or structural analyses.

C) Project Description

i. Introduction

Lignin, which comprises roughly 20 % of woody biomass, plays a dual role in forest productivity and industrial processing: it provides mechanical strength and a hydrophobic barrier that safeguards against water loss, yet its rigid, recalcitrant structure is a major impediment to deconstructing biomass into renewable fuels and pulp. In the southeastern United States—where forestry contributes over $300 billion annually and supports millions

of livelihoods—increasingly severe droughts driven by climate change have already cut tree growth by up to 30 % and reduced pine sawtimber yields by 12 %, translating to nearly $300 million in annual losses and undermining a carbon-sequestration service valued at $35 billion per year (Krug et al., 2015; USDA Sustainability Plan, 2022) . Conducting field trials of CRISPR-edited Populus varieties under these real-world conditions is therefore critical to validate both drought resilience and downstream processing advantages.

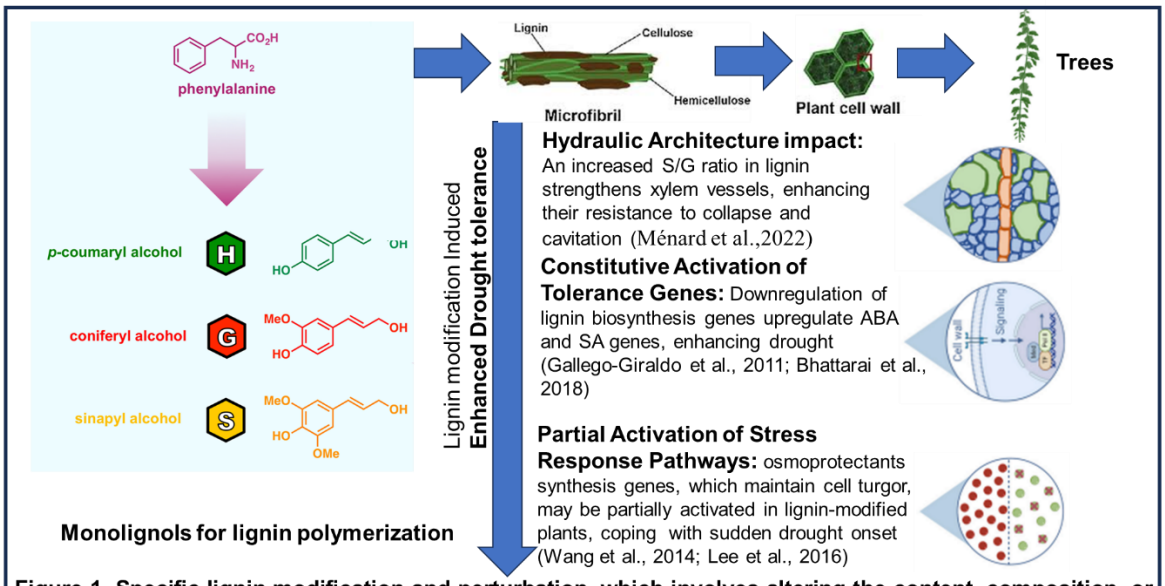

Figure 1. Specific lignin modification and perturbation, which involves altering the content, composition, or structure within plant cell walls, could enhance plant's drought tolerance and their related mechanisms. Lignin modifications, such as increasing the S/G ratio, enhance the structural integrity of xylem vessels, making them more resistant to collapse or cavitation, thus maintaining hydraulic conductivity under drought conditions. This structural reinforcement is coupled with the constitutive activation of stress tolerance genes, including those involved in ABA and SA pathways, priming plants for rapid and robust responses to drought and pathogen attacks. Additionally, lignin perturbation can partially activate stress response pathways, preparing the plant to quickly escalate its defense when stress occurs. These modifications lead to synergistic effects, improving water use efficiency and defense readiness. However, trade-offs exist, as excessive lignin reduction may weaken cell walls, increasing vulnerability to mechanical damage or pathogens, with the overall impact being based on environmental conditions. Figure is created by Biorender.

Genetic modifications that reduce total lignin content and increase the syringyl/guaiacyl (S/G) ratio have been shown to improve pulp yield and lower carbon emissions during processing (Wang et al., 2018; Sulis et al., 2023). Intriguingly, reduced-lignin plants often display enhanced drought tolerance through multiple mechanisms: heightened abscisic acid (ABA) sensitivity and stomatal control, release of bioactive wall fragments that prime defense pathways, improved water-use efficiency, and alterations in root architecture (Gallego-Giraldo et al., 2011; Cesarino et al., 2019; Choi et al., 2023; Figure 1). Yet these benefits are sometimes offset by unintended biomass or growth penalties—observed in alfalfa, Arabidopsis, and sorghum bmr mutants—which could negate overall gains in forestry applications (Muro-Villanueva et al., 2019; De Meester et al., 2022). Leveraging our AI-aided multiplex editing approach, we generated over 69,000 predicted lignin-pathway edits and identified 20 elite *Populus trichocarpa* genotypes that maintain ≥80 % of wild-type biomass in greenhouse trials while exhibiting up to 49 % reductions in total lignin and S/G ratios elevated from 2.7 to 4.0 (Sulis et al., 2023). To confirm that these engineered lines combine low recalcitrance with robust field performance, we will evaluate 20 of these CRISPR lines alongside the wild type in both controlled greenhouse and open-field plantations under short-rotation forestry regimes in the Southern U.S.—a management system that supports frequent harvest cycles for pulpwood and bioenergy (Wang et al., 2013).

Our multi-faceted experimental design will integrate molecular, biochemical, physiological, and eco-physiological assessments (Figure 2). Specifically: 1. Transcriptomics: **Four tissues (root, leaf, xylem, apical bud) from 21 genotypes will be sampled under well-watered, drought, and recovery treatments (3 replicates each), generating 756 RNA-seq libraries sequenced to ~40 M single-end 150 bp reads per sample. 2. Physiological monitoring:** In situ measurements of photosynthetic rate, stomatal conductance, hydraulic conductivity, and root architecture will be recorded throughout drought trials to link molecular changes to whole-plant performance. We will do these experiments at Auburn University before the tissue harvests, and along with drought treatment experiments. 3. Metabolomics: **Based on RNA-seq results, nine top-performing edited lines plus wild type (7 genotypes) will be selected for untargeted LC-MS profiling of root, leaf, and xylem (189 samples). We will quantify key metabolite classes—monolignol intermediates (coniferyl/sinapyl alcohols, ferulic acid), phytohormones (ABA, salicylic and jasmonic acids), and osmoprotectants/antioxidants (proline, glycine betaine, flavonoid glycosides)—under strict QC protocols.**

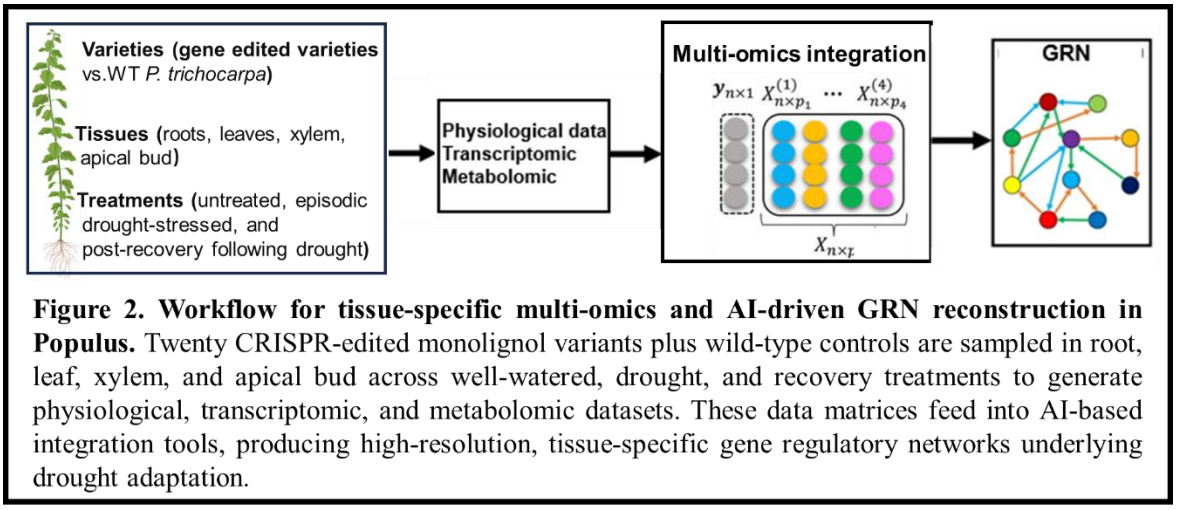

Figure 2. Workflow for tissue-specific multi-omics and AI-driven GRN reconstruction in Populus. Twenty CRISPR-edited monolignol variants plus wild-type controls are sampled in root, leaf, xylem, and apical bud across well-watered, drought, and recovery treatments to generate physiological, transcriptomic, and metabolomic datasets. These data matrices feed into AI-based integration tools, producing high-resolution, tissue-specific gene regulatory networks underlying drought adaptation.

This comprehensive dataset will allow us to answer three fundamental questions: 1. **How do precise lignin-pathway perturbations rewire drought-responsive transcriptomic and metabolic networks across contrasting tissues? 2. Which gene–metabolite signatures predict enhanced drought resilience and recovery in lignin-engineered lines? 3.Can we validate that low-lignin, high–S/G poplars maintain growth, yield, and carbon-sequestration performance under drought stress without biomass penalties?**

By integrating mechanistic greenhouse studies with rigorous field-trial validation under drought conditions in the Southern U.S., we will generate a comprehensive, multi-omic dataset—including 756 deeply sequenced transcriptomes and 189 untargeted metabolomes—using JGI's cutting-edge sequencing platforms, high-throughput data management systems, and bioinformatics expertise. This effort will leverage JGI's mission to deliver advanced genomic capabilities and large-scale data to the global research community, responsibly managing resources while leading genomic innovation for a sustainable bioeconomy. In doing so, our project not only furnishes a systems-level roadmap for engineering Populus feedstocks that unite low lignin recalcitrance with climate resilience, but also directly supports JGI BER priorities in abiotic-stress physiology and plant developmental processes and aligns with broader JGI BER goals in plant

production, bioenergy, natural resources, and the environment—ultimately advancing sustainable, economically viable forestry practices.

ii. Project narrative

a. Preliminary data and On-going related project

Here, we present preliminary data and on-going awarded project content that demonstrates our commitment and capacity to achieving the work outlined in JGI proposal.

Generating Genetic Diversity in Poplar using Artificial Intelligence (AI)-based Prediction and Multiplex CRISPR

We integrated transcriptomic, metabolic, proteomic, and phenomic data from various lignin biosynthesis mutants in *P. trichocarpa* into a mathematical model predicting poplar wood properties following the engineering of one or multiple lignin pathway genes (Wang et al., 2018). Utilized a predictive model for monolignol biosynthesis, we identified gene targets for multiplex genome editing, exploring over 69,000 sets of multigenic editing strategies to modify wood properties (Sulis et al., 2023). Multigenic editing expanded the range of achievable phenotypic variations compared to single-gene edits, resulting in wood with significantly lower lignin content and a higher S/G ratio, which combination is ideal for fiber production. Out of the strategies tested, only 347 met the criteria for lignin content, carbohydrate-to-lignin (C/L) ratio, S/G ratio, and tree growth (Sulis et al., 2023). We also took no penalty of plant growth and biomass accumulation into consideration for modeling out the strategies. Seven selected strategies were selected out with the prediction of substantial improvements in fiber traits, including reduced lignin content and increased S/G and C/L ratios, without compromising tree growth. To assess the impact of seven wood property-modifying strategies for fiber production, the multiplex CRISPR constructs were created and delivered into *P. trichocarpa* trees. A total of 174 independently edited *P. trichocarpa* lines were generated, targeting the concurrent editing of three to six monolignol genes (Figure 3A and B). These edited lines exhibited varying degrees of loss-of-function mutations in the target genes, with complete biallelic loss-of-function editing achieved for strategies targeting three or four genes (Figure 3B). Substantial editing of most target genes was observed in specific lines (Figure 3C). The editing mainly involved small insertions and deletions near the cleavage sites (Figure 3C and D), with no detectable off-target edits, indicating high specificity in *P. trichocarpa* genome editing. The broad variation in target

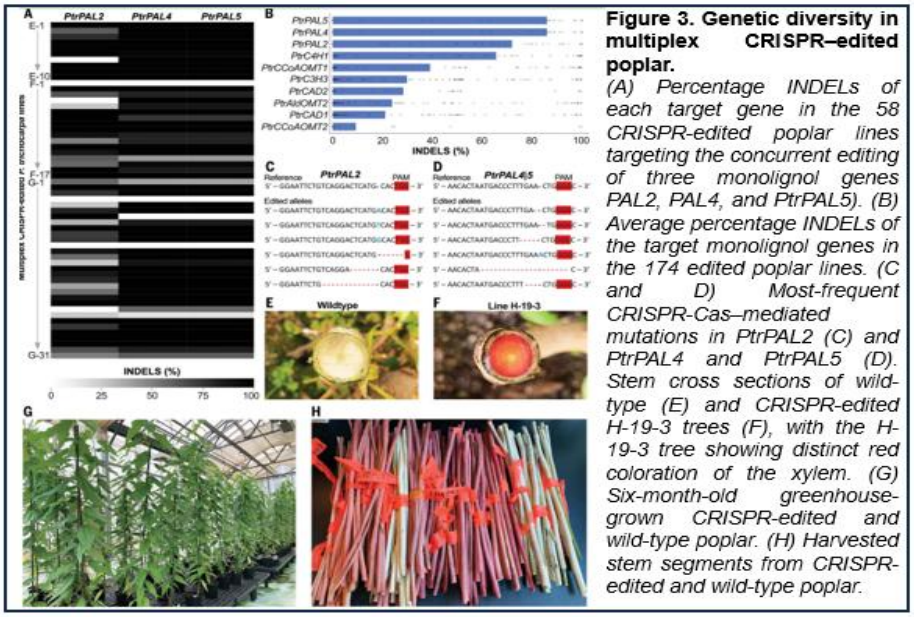

Figure 3. Genetic diversity in multiplex CRISPR–edited poplar. *(A) Percentage INDELs of each target gene in the 58 CRISPR-edited poplar lines targeting the concurrent editing of three monolignol genes PAL2, PAL4, and PtrPAL5). (B) Average percentage INDELs of the target monolignol genes in the 174 edited poplar lines. (C and D) Most-frequent CRISPR-Cas–mediated mutations in PtrPAL2 (C) and PtrPAL4 and PtrPAL5 (D). Stem cross sections of wild-type (E) and CRISPR-edited H-19-3 trees (F), with the H-19-3 tree showing distinct red coloration of the xylem. (G) Six-month-old greenhouse-grown CRISPR-edited and wild-type poplar. (H) Harvested stem segments from CRISPR-edited and wild-type poplar.*

gene editing profiles (Figure 3E to H) successfully achieve our seven model-predicted gene-editing strategy with introduced genetic diversity into monolignol biosynthesis, which provide us diverse clonal line for drought tolerance analyses and screening out the drought tolerant variety.

CRISPR-based Multiplex Editing Create Variabilities of Lignin Quality and Quantity but without Growth' Penalty in Some of Six-month-old Poplar Trees

Multiplex editing of poplar monolignol genes had a significant impact on the chemical and physical properties of the wood, as evidenced by various data These changes were in line with predictions made by the lignin model. In contrast to the stem wood of 6-month-old wild-type poplar which contains 20.9% lignin and has a C/L ratio of 3.0, the multiplex-edited wood showed reductions in lignin content of up to 49.1% and an increase in the C/L ratio by up to 228% when averaged across 65 lines (Figure 4). Additionally, two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) revealed alterations in lignin composition, with the S/G ratio increasing from 2.7 in the wild type to as high as 4.0 in the edited lines (Figure 4 and 5). The most substantial reductions in lignin content were observed in edited trees with four to six gene edits, though strategies targeting three genes also showed significant reductions of up to 32% (Figure 4).

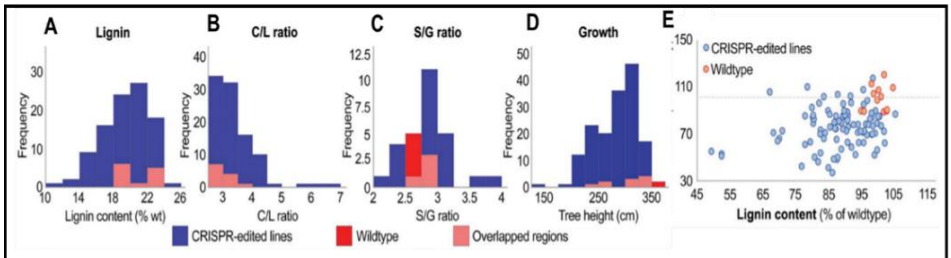

Figure 4. Phenotypic variation in CRISPR-edited and wild-type poplar. *(A to D) Distribution of lignin content (A), C/L ratio (B), S/G ratio (C), and tree growth (height) (D) in CRISPR-edited and wild-type poplar. Blue and red colors stand for CRISPR-edited lines and wild-type control measurements, respectively. Pink represents the overlapped regions of the blue (CRISPR-edited lines) and red (wild-type controls) bars. (E) Quantitative relationship between lignin content and wood volume in CRISPR-edited and wild-type poplar.*

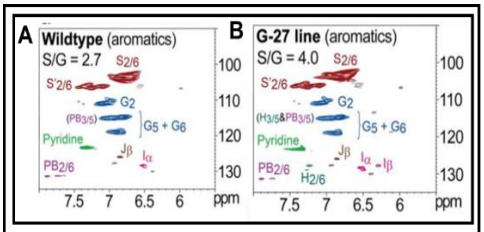

Figure 5. 2D NMR 13C-1 H (heteronuclear single quantum coherence) correlation spectra (side-chain and aromatic regions) of wild type (A) and the G-27 edited line (B). *Contours in these regions were used to estimate the distribution of lignin interunit linkages and lignin composition, namely S/G/H ratios, as well as p-hydroxybenzoate (PB). ppm, parts per million.*

Although some edited lines displayed reduced wood elasticity, overall, wood elasticity remained largely unchanged between wild-type and genome-edited poplars. There were no significant differences in wood density greater than 15.1% from the wild-type level in the edited poplars. Importantly, certain CRISPR-edited lines with significant reductions in lignin content also exhibited a reduction in tree growth. These findings suggest that some multiplex genome editing strategies may mitigate the adverse effects often associated with typical single-gene knockout approaches, although long-term field trials are needed to assess their impact on tree phenotypic properties and industrial relevance. **Given that we will screen the drought-tolerant poplar elite varieties with economic potential, we only use poplar varieties with favorable growth (>80% biomass compared to wild-type trees; Table 1) into drought tolerance test in this project, 20 CRIPSR edited varieties will finally be selected for the drought tolerance test.**

USDA AFRI CARE Field Trials Propel Drought Tolerance Evaluation in Elite CRISPR-Edited Poplar Varieties for This Project

The USDA AFRI CARE project, titled "Debottlenecking Fiber Production: Field Evaluation of Low-Lignin Poplar Varieties for Pulp Industries," is our ongoing research focused on establishing a short-rotation plantation of CRISPR-edited poplar varieties in Southeastern US. Led by Auburn University and NCSU, this project encompasses an extensive field trial of over 1,300 poplar trees that have undergone lignin modifications through CRISPR-based genetic editing. These trials provide an invaluable opportunity to assess how lignin alterations influence drought tolerance and related traits in real-world conditions, particularly in the southeastern U.S.—a region where drought presents a significant challenge to the forest industry (Mitchell et al., 2014). Leveraging the USDA AFRI CARE-funded field trials, this research benefits in several critical ways: **Validation of Drought Resilience:** The field trials enable a comprehensive, long-term assessment of drought tolerance in genetically edited poplar varieties. By monitoring growth performance, survival rates, and physiological responses, the trials provide key insights into how lignin modifications contribute to drought resilience. This real-world data is essential for validating findings from laboratory and greenhouse studies related to drought tolerance. **Evaluation of Lignin Traits:** The USDA AFRI CARE field trials offer a detailed examination of how lignin content and composition, such as S/G and C/L ratios, impact the physiological and structural responses of trees to drought. Together with our proposed physiological analyses in this project, these trials help identify specific lignin traits that enhance water use efficiency, hydraulic architecture, and overall tree resilience under drought conditions, laying the groundwork for a deeper understanding of the genetic and physiological mechanisms involved. **Optimization for Pulping and Saccharification Efficiency:** The field trials also yield critical data on the effects of lignin modifications on the pulping and **Saccharification** process. By evaluating the cell wall properties, fiber quality, and pulp and biofuel yield of field- grown poplars, the research identifies lignin traits that could be evaluated together with drought response related physiological data for analyzing the mechanisms how lignin affects drought physiological response, tolerance, and response of these varieties. This dual optimization is crucial for maintaining the economic viability of the forest industry while addressing the challenges of climate change. **Long-Term Ecological Viability:** The field trials facilitate an in-depth examination of the long-term ecological viability of lignin-modified poplars, assessing their contributions to carbon storage, resistance to drought-exacerbated stressors, and influence on local ecosystems. These findings are critical to ensuring that the developed poplar varieties are both environmentally sustainable and resilient in the face of climate change.

b. Experimental plan for sample and metadata collection and data generation

Hardwood cuttings from 20 CRISPR-edited monolignol poplar lines and the wild-type control (21 genotypes total) will be established in a climate-controlled greenhouse (25 °C/18 °C day/night; 16 h photoperiod; 70 % RH). Each cutting will be surface-sterilized, inserted into a peat–perlite (1:1) rooting medium, and monitored daily for root emergence. Upon achieving ≥ 90 % rooting, plantlets will be potted into 5 L containers filled with a homogenized mix of loam:peat:perlite (2:1:1 v/v/v). Twelve saplings per genotype (360 total) will be arranged in a randomized complete-block design across 12 greenhouse benches; each bench will host one replicate of every genotype to minimize microclimate effects. Soil volumetric moisture sensors (Decagon EC-5) and ambient

sensors (HOBO U23) will log substrate moisture and greenhouse conditions every 15 minutes. All plants will receive biweekly Hoagland nutrient solution and routine pest inspections; growth metrics (height, node count, leaf number) will be recorded weekly in our electronic lab notebook.

At six months, saplings will be divided into drought (n = 6/genotype) and control (n = 6/genotype) groups. The drought cohort will undergo a stepwise desiccation protocol adapted from Galvez et al. (2011), reducing soil moisture from 80 % to 30 % field capacity over 14 days before holding at 30 % for 28 days, with volumetric water content confirmed by sensor readings. Control plants remain at 80–90 % field capacity. Sampling will occur at three time points—pre-drought (day 0), drought peak (day 14+28), and post-recovery (14 days after rehydration)—for root, leaf, xylem, and apical bud tissues in three biological replicates per genotype per treatment (21 × 4 × 3 × 3 = 756 samples) (Figure 6). Each tissue sample (~100 mg fresh weight) will be excised, labeled with a barcoded ID, flash-frozen

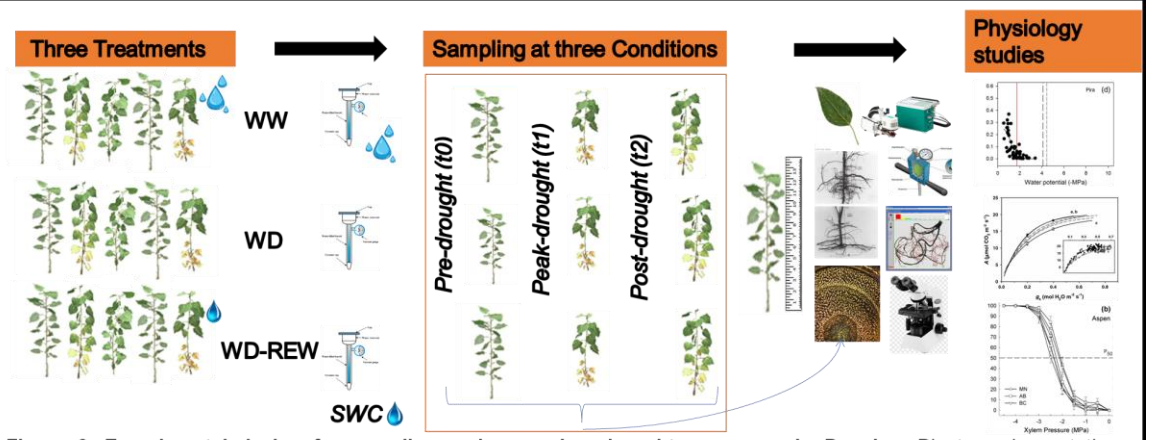

Figure 6. Experimental design for sampling and assessing drought responses in Populus. Plants underwent three treatments: well-watered (WW, 80% REW), drought-stressed (WD, 40% REW), and post-drought recovery (WD-REW, rehydrated to field capacity). Sampling occurred at pre-drought (t0), peak drought (t1), and post-recovery (t2) stages, with leaf, root, and xylem tissues collected for multi-omics and physiological analyses. Soil moisture was monitored to ensure controlled drought progression, while physiological assessments included leaf water potential, gas exchange, xylem hydraulics, and root architecture.

in liquid nitrogen within 30 s, and logged into a LIMS with metadata fields (genotype, treatment, time point, bench location, sensor-recorded soil moisture, harvest timestamp). Parallel to sampling, physiological measurements will be conducted on sister plants: leaf relative water content via gravimetric method (FW, TW, DW), leaf water potential with Scholander chamber (PMS Instrument Co.), in situ gas exchange (photosynthetic rate A and stomatal conductance g_s) on a LI-6400XT (1,000 µmol m^{-2} s^{-1} PPFD; 400 ppm CO_2; 25 °C; 60 % RH), weekly stem height/diameter, and xylem hydraulic conductivity and drought-induced cavitation curves on 10 cm stem segments using the Sperry apparatus. Concurrently, stem wood and fine-root samples from each harvest will be processed for lignin analysis: ~2 g of air-dried tissue per sample will be milled to 40–60 mesh, ball-milled under P_2O_5 with ZrO_2 beads at 600 rpm (15 min milling/30 min rest cycles; 6 h total), and stored under vacuum until analysis. Approximately 40 mg of milled material will be dissolved in 500 µL DMSO-d_6/pyridine-d_5 (4:1 v/v), sonicated to a homogenous gel, and transferred to 5 mm NMR tubes. Two-dimensional HSQC spectra will be acquired on a Bruker Avance NEO 700 MHz with the hsqcetgpsisp.2 pulse program (2,048 × 512 points; 125 ms F2/6.6 ms F1; D1 = 1 s; 32 scans). Monomer S/G/H ratios and β-O-4, β-5, β-β, β-1 linkage abundances will be quantified by contour integration in TopSpin 4.1.1, with results normalized to 100 % total monomers and reported per Wang et al. (2018). **The physiological data and lignin data will be analyzed by Auburn University and NCSU,**

and the obtained data will integrate together with following transcriptomic and metabolomics data for further analyses.

All 756 tissues will have undergone RNA extraction (Qiagen RNeasy; ≥ 2 μg; A_{260}/A_{280} = 1.9–2.1; RIN ≥ 7.5 measured on Agilent 2100), and stranded mRNA libraries (Illumina TruSeq Stranded mRNA, 150 bp single-end targeting 40 M reads/sample) will be constructed by JGI based on JGI's guideline. Following transcriptomic data return, nine CRISPR lines plus wild type (10 genotypes) will be chosen for targeted metabolomic profiling: root, leaf, and xylem samples (7 × 3 × 3 × 3 = 189) will be lyophilized, cryo-ground, and extracted (methanol–water–chloroform; 2.5:1:1 v/v/v with isotopic internal standards). Extracts will be clarified by centrifugation (16,000 × g, 10 min), dried under nitrogen, and reconstituted in 50 μL 50 % methanol for LC-MS, with blank, pooled QC, and standard injections interspersed for drift correction — all under the same rigorous metadata and chain-of-custody protocols. We will submit the samples for **transcriptomic and metabolomics analysis, and the detailed experiment plan will be discussed between JGI and us.**

c. Data analysis plan for all analyses

All raw data—including physiological measurements, lignin chemistry metrics, RNA-seq count matrices, and metabolomic feature tables—will be consolidated into R and Python environments for integrated analysis. Initial preprocessing will involve systematic outlier detection, imputation of missing values where appropriate, and normalization of each data type (for example, log-transformation of skewed physiological or metabolite distributions and variance stabilization of RNA-seq counts) to ensure comparability across datasets.

For **Objective 1 To Investigate the Impact of Alterations in Lignin Content and Composition (e.g., S/G, C/L Ratios) on the Drought Tolerance of Genetically Edited Poplar Varieties**, we will first examine pairwise associations between key drought-response traits (leaf relative water content, leaf water potential, photosynthetic rate, stomatal conductance, growth increments, hydraulic conductivity, and cavitation resistance) and lignin characteristics (total lignin content, S/G/H monomer ratios, and interunit linkage abundances) using both Pearson and Spearman correlation matrices. To quantify how variations in lignin composition drive quantitative drought-tolerance indices, we will employ Partial Least Squares Regression (PLSR) with leave-one-genotype-out cross-validation, thereby assessing predictive performance while guarding against overfitting. Recognizing that biological relationships may be nonlinear, we will train Random Forest regressors (2,000 trees, grid-searched mtry) and Support Vector Machine models (radial basis kernel, hyperparameters C and γ optimized via 10-fold cross-validation), using feature-importance rankings from the Random Forest to prioritize the most influential lignin variables. Finally, Structural Equation Modeling (SEM) implemented in the lavaan package will be used to disentangle direct and indirect causal pathways linking lignin modifications, hydraulic function, and biomass accumulation under drought, with model fit evaluated by χ^2 tests, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). SEM models will be validated against an independent subset of CRISPR lines to confirm robustness.

For **Objective 2 To elucidate the underlying genetic and physiological mechanisms that contribute to drought resilience in poplars with altered lignin traits**, transcriptomic and metabolomic datasets will be analyzed in concert. RNA-seq counts will

be processed through DESeq2 to identify genes differentially expressed ($|\log_2$ fold-change$| > 1$, FDR < 0.05) between drought and control treatments in root and xylem tissues. Metabolomic data will be subjected to principal component analysis (PCA) for global pattern recognition, followed by partial least squares-discriminant analysis (PLS-DA; 1,000 permutations, VIP > 1, adjusted $p < 0.05$) to pinpoint metabolites most strongly associated with drought. We will construct weighted gene co-expression networks (WGCNA) to detect modules of tightly co-regulated genes and correlate each module's eigengene with physiological and lignin traits ($|r| > 0.7$, $p < 0.01$), followed by Gene Ontology and KEGG pathway enrichment analyses (Fisher's exact test $p < 0.01$) to interpret biological functions. Metabolite identities will be mapped to KEGG and MapMan pathways for enrichment, and their abundances will be correlated with co-expression module eigengenes to infer gene–metabolite regulatory networks (Banerjee et al., 2022). These networks will be visualized in Cytoscape, and pathway-level integration will be facilitated by MetaboAnalyst. To develop predictive biomarkers of drought tolerance, we will combine omics and trait data into machine-learning pipelines, training Random Forest and SVM classifiers (80/20 train/test split, 10-fold cross-validation) and applying Recursive Feature Elimination and LASSO regression to select minimal, high-performance marker panels. Model performance will be assessed via accuracy, ROC AUC, precision, and recall, and final models will be validated on hold-out genotypes to ensure real-world applicability.

For Objective 3**: Tissue-Specific Multi-Omics and AI-Driven Gene Network Modeling for Populus Drought Adaptation.** To capture how different organs coordinate drought responses, we will generate parallel transcriptomic, chromatin-accessibility, and metabolomic profiles from root, leaf, xylem, and apical bud tissues at pre-drought, drought peak, and recovery stages (Objective 1 and 2). Bulk RNA-seq data will be processed in DESeq2 to identify tissue-specific differentially expressed genes (DEGs; $|\log_2 FC| > 1$, FDR < 0.05) under drought and recovery. Complementary ATAC-seq on the same tissues will be analyzed with ArchR and Cicero to map differentially accessible regions (DARs) in each organ and link peaks to putative target genes via co-accessibility. Metabolite extracts from each tissue will be profiled by LC-MS, and principal component analysis (PCA) followed by PLS-DA will uncover metabolites that discriminate drought responses in roots, leaves, and xylem. We will then annotate tissue-specific DEGs and DARs with HOMER and JASPAR to predict transcription factor (TF) binding motifs enriched under stress in each organ. Building on these datasets, we will reconstruct high-resolution, tissue-specific gene regulatory networks (GRNs) using AI-driven multi-omics integration. First, MOGONET will jointly model RNA-seq from each tissue to infer co-regulatory modules, while VCDN will uncover cross-omics correlations linking gene expression and chromatin accessibility patterns. Next, LINGER's pretrained graph neural networks will predict TF–target interactions within each organ, refining our understanding of the regulatory circuits that drive lignin biosynthesis, water-transport functions, and stomatal control under drought. To ensure comparability across batches and platforms, we will apply Harmony for ComBat for RNA-seq and metabolomics, and we will perform rigorous QC to exclude low-complexity or contaminated samples. Finally, we will validate key regulatory relationships by overlaying physiological and lignin-chemistry measurements—such as hydraulic conductivity and S/G ratios—onto the tissue-specific GRNs, thereby pinpointing master regulators whose modulation could enhance drought resilience in poplar.

d. Roles and responsibilities of the project team.

The complemental work is supported by a USDA NIFA grant (PI: Dr. Hao Chen; Co-PI: Dr. Jack Wang) and Auburn University start-up funds, which cover greenhouse operations, drought treatment implementation, sample collection, and quality control prior to submission to JGI. Funding also underwrites the salaries of technical staff and graduate students tasked with sample preparation, data analysis, and manuscript drafting. Transcriptomic and metabolomic data analyses will leverage institutional high-performance computing clusters and publicly available platforms such as the DOE Systems Biology Knowledgebase (KBase), Cytoscape for network visualization, and MetaboAnalyst for pathway integration.

The interdisciplinary project team is led by Dr. Hao Chen (Auburn University), who oversees greenhouse experiments, physiological assays, sample collection, and transcriptomic data analysis. Co-PI Dr. Jack P. Wang (North Carolina State University) provides expertise in lignin biosynthetic gene function and guides transcriptomic interpretation and network modeling. Dr. Miaomiao Li (Oak Ridge National Laboratory) contributes to the design and interpretation of regulatory network inference from RNA-seq data, supporting integrative systems biology approaches. Dr. Zongliang Yue (Auburn University) leads statistical, and machine-learning analyses and assists in manuscript preparation. All JGI-generated data will be jointly analyzed by this multidisciplinary team to maintain scientific rigor and ensure timely project progress.

e. Anticipated schedule of project milestones

Over the **first year (Months 0–12)**, we will focus on **deciphering tissue-type-specific regulation of drought resilience** by integrating multi-omics profiling (transcriptomics, metabolomics), physiological measurements (RWC, Ψ_leaf, gas exchange, hydraulics), and lignin trait analyses across root, leaf, and xylem. During this period, all sampling protocols, data-collection pipelines, and preliminary quality controls will be established and refined to ensure reproducibility. Beginning in **Month 10 and continuing through Month 24**, we will implement **AI-driven gene regulatory network (GRN) reconstruction** using our multi-omics datasets. Computational frameworks such as MOGONET, VCDN, and LINGER will be trained on the tissue-specific RNA-seq and ATAC-seq features to infer high-resolution GRNs. This phase will overlap with the tail end of our initial analyses, allowing iterative improvement of network models based on emerging biological insights. In **Months 18–30**, we will carry out **experimental validation of predicted transcription factor–cis-regulatory element interactions**, applying ChIP-seq and reporter assays to confirm key regulatory links in each tissue. From **Month 24 to Month 32**, we will **link these validated GRNs and CREs to lignin biosynthesis pathways and physiological drought-resilience phenotypes**, testing whether modulation of identified regulators alters S/G ratios, hydraulic function, and biomass allocation under water stress. Finally, during the **last four months (Months 32–36)**, we will **integrate our tissue-specific GRNs with CRISPR-edited lignin traits** to design and test enhanced poplar genotypes for improved drought tolerance and bioenergy production. Predictive machine-learning models—trained on the full suite of omics, physiological, and lignin-chemistry data—will guide the final experimental validations and prepare the groundwork for future field deployment.

iii. the size and nature of the larger research community that will use the data.

Academic Community: The comprehensive, tissue-specific multi-omics dataset—combining time-series transcriptomes, metabolomes, physiological traits, and lignin compositional profiles under well-watered, drought, and recovery conditions—will become a foundational resource for plant scientists. Molecular biologists and geneticists can mine the data to unveil transcriptional regulatory networks driving lignin biosynthesis and drought tolerance, guiding targeted functional studies and marker-assisted breeding. Systems biologists and bioinformaticians will leverage the integrated TRN maps to develop and benchmark new data-integration and network-inference algorithms, while ecologists and climate modelers can incorporate quantitative physiological and metabolic measurements into predictive models of forest carbon cycling and ecosystem resilience under water stress. Forestry researchers will apply these insights to engineer Populus and related species with optimized wood properties and drought resilience, informing sustainable management and conservation strategies. Industrial Community: By linking regulatory modules and metabolic signatures to wood quality and water-use efficiency, this multi-omics framework accelerates the development of superior poplar cultivars for bioenergy, pulp & paper, and timber applications. Bioenergy companies can exploit identified gene–metabolite markers to breed feedstocks with enhanced cellulose yield and reduced pretreatment requirements. Pulp and paper manufacturers will use lignin compositional data to optimize pulping processes and fiber recovery. Timber and construction industries can harness drought-adaptive regulatory circuits to produce wood with improved mechanical strength and stability under water-limited conditions. Agricultural biotechnology firms will translate key drought-response networks into commercial crops, while environmental consultants will adopt physiological and metabolomic biomarkers for precision reforestation and habitat restoration. Precision agriculture providers can integrate these markers into sensor-driven decision-support tools, enabling real-time irrigation management and climate-adaptive cropping systems.

The relevance of the project to the DOE mission

The proposed research focusing on the drought tolerance- and wood formation-transcriptional regulatory networks in *P. trichocarpa* is strategically designed to align with the DOE's JGI and the Office of Biological and Environmental Research (BER) missions, targeting sustainable bioenergy production and climate resilience. By exploring the genetic mechanisms that enhance drought tolerance and facilitate efficient wood formation, this project addresses critical elements of climate adaptation, particularly as extreme weather events and prolonged drought periods become more frequent due to climate change. Investigating *P. trichocarpa*, a model bioenergy crop known for its significant carbon sequestration capability, the study aims to develop renewable and sustainable biofuels and bioproducts from plant biomass. This aligns with the DOE's priority to use terrestrial plants as biofuel feedstocks, ensuring economic viability while enhancing the species' resilience to environmental variability. The research will delve into the transcriptional networks that govern these adaptive traits, potentially pioneering ways to genetically engineer trees to not only sequester more carbon but also thrive under drought stress without a growth penalty. This is crucial for maintaining high productivity and ensuring economic returns from bioenergy crops. Through multiomic studies and in-depth genetic analysis of lignin pathway gene, the project will enhance our capacity to predict and improve tree performance under various environmental stresses, which is essential for managing bioenergy resources and forest ecosystems sustainably. Additionally, the proposed work includes a comparative analysis of drought response in poplar across various plant tissue

types, which will enhance our understanding of plant biology and contribute to cross-disciplinary approaches that align with the DOE's environmental and energy objectives. Overall, this thorough exploration of *P. trichocarpa's* genetic and transcriptional profiles is poised to drive significant advancements in bioenergy and carbon cycling, enhancing our capacity to address national energy and environmental challenges effectively. Furthermore, Principal Investigator Hao Chen and Co-Principal Investigator Jack Wang have over a decade of experience in researching lignin biosynthesis and its regulation in the bioenergy woody crop, poplar. Our efforts are directly aligned with the DOE's mission of developing renewable and sustainable sources of biofuels and bioproducts from plant biomass. The proposed project aims to extend our research to maximize the benefits for both climate-smart forestry and bioenergy resource development.

D) References:

Adams, H. D., Guardiola-Claramonte, M., Barron-Gafford, G. A., Villegas, J. C., Breshears, D. D., Zou, C. B., … Huxman, T. E. (2009). Temperature sensitivity of drought-induced tree mortality portends increased regional die-off under global-change-type drought. Proceedings of the National Academy of Sciences, 106(17), 7063–7066.

Bing, Ryan G., et al. "Beyond low lignin: Identifying the primary barrier to plant biomass conversion by fermentative bacteria." *Science Advances* 10.42 (2024): eadq4941.

Cesarino, I., De Meester, B., & Boerjan, W. (2019). Selective lignin down-regulation leads to constitutive defense response expression in alfalfa (Medicago sativa L.). New Phytologist, 190(3), 627–639.

Choi, S. J., Kim, J.-H., & Dong, T. (2023). Modulation of lignin biosynthesis for drought tolerance in plants. Frontiers in Plant Science, 14, 1116426.

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., … Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature Protocols, 6(7), 1060–1083.

Gallego-Giraldo, L., Escamilla-Treviño, L., Jackson, L. A., & Dixon, R. A. (2011). Salicylic acid mediates the reduced growth of lignin-down-regulated plants. Proceedings of the National Academy of Sciences, 108(52), 21210–21215.

Galvez, D. A., Tyree, M. T., & Santiago, L. S. (2011). Whole-plant hydraulic conductance and stomatal conductance responses to light acclimation in four tropical forest species. Tree Physiology, 31(8), 883–893.

Muro-Villanueva, A., Ko, J.-H., & Han, K.-H. (2019). Growth penalties associated with lignin down-regulation in field-grown alfalfa. Plant Biotechnology Journal, 17(4), 740–752.

Sulis, D. B., Smith, J. P., Nguyen, E., Oliveira, L., Wang, J. P., … Chen, H. (2023). Multiplex CRISPR editing of wood for sustainable fiber production. Science, 381(6654), 216–221.

Tuskan, G. A., et al. (2006). The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science, 313(5793), 1596–1604.

U.S. Department of Agriculture. (2022). USDA 2022 Sustainability Plan. https://www.sustainability.gov/pdfs/usda-2022-sustainability-plan.pdf

Wang, D., LeBauer, D., & Dietze, M. C. (2013). Predicting yields of short-rotation hybrid poplar (Populus spp.) for the United States through model–data synthesis. Ecological Applications, 23(4), 944–958.

Wang, J. P., Bräutigam, K., & Bonawitz, N. D. (2018). Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis. Nature Communications, 9(1), 1579.

E) CVs of lead investigators: (3 page limit for each investigator)

NAME: Chen, Hao

ORCID iD: 0000-0002-6415-993X

POSITION TITLE: Assistant Professor

1. Professional qualifications

INSTITUTION AND LOCATION	DEGREE (if applicable)	END DATE MM/YYYY	FIELD OF STUDY
North Carolina State University	MS	2011-2013	Forestry and Environmental Resources
North Carolina State University	PhD	2011-2017	Forestry and Environmental Resources
North Carolina State University	Postdoc	2017-2023	Genetic and Genomics
Auburn University	Assistant professor	2023-	Forestry genomic and biotechnology

2. Expertise: Lignin biosynthetic, Transcriptional regulatory network, Plant cell wall formation, Tree genetic and improvement， Wood formation

3. Award grants:

Source of Support	Project Role	Title
AAES (Alabama Agriculture Experiment Station)	PI	Transcriptional regulation of stem cell self-renewal and differentiation for wood formation
USDA NIFA The Agriculture and Food Research Initiative (AFRI)	PI	Debottlenecking Fiber Production: Field Evaluation of Low-Lignin Poplar
USDA US Forest Service Forest Health Protection	PI	Safeguarding Endangered Southeastern US Oaks through Micropropagation
USDA US Forest Service	Co-PI	Enhancing Sustainability: Developing Fuel-Reducing Wood Biomass for Biochar Byproducts to Mitigate Fire Risks

4. Academic and professional positions

2023 – **Present** **Assistant Professor in Forestry genomics and Biotechnology**
Auburn University, Alabama, US

2017 – **2023** **Postdoctoral Research Scholar in Alonso-Stepanova Laboratory**
North Carolina State University, Raleigh, US Mentors: Prof. Jose Alonso and Prof. Anna Stepanova

2012 – **2017** **Research Assistant in Forest Biotechnology Group**
North Carolina State University, Raleigh, US Mentors: Prof. Vincent Chiang and Prof. Ronald R. Sederoff.

5. Publications

Selected paper lists related with this project (* as first author or co-first author; bold as the applicant; † as corresponding author)

1. **Chen, H**.*; Neubauer, M*.; Wang, J. Enhancing HR Frequency in Plants: Solidifying a Foundation for Precise Genome. *Front. Plant Sci.* **2022**
2. Tong, H. *****, **Chen, H. *†,** and Williams, C. M†. (2022). Identification of Transcription Factors Regulating SARS-CoV-2 Tropism Factor Expression by Inferring Cell-Type-Specific Transcriptional Regulatory Networks in Human Lungs. *Viruses*, 14(4), 837.
3. **Chen, H. *,** Alonso, J. M., & Stepanova, A. N. † (2022). A Ribo-Seq Method to Study Genome-Wide Translational Regulation in Plants. *Methods in molecular biology* (Clifton, NJ), 2494, 61-98.
4. **Chen, H.*;** Bullock, D.A., Jr.; Alonso, J.; Stepanova, A. To fight or to grow: the balancing role of ethylene in plant abiotic stress responses. *Plants* **2022**, *11*, 33. (Cover story of this issue in MDPI plants).
5. Tong, H*.; **Chen, H. *;** Williams, C.M. † Gene Regulatory Network of Secondary Cell Wall Biosynthesis during VND7 Induced de novo Xylem Formation. *IJBBB Gene* **2021**, *11*.

6. **Chen, H. *;** Wang, J.P*.; Liu, H.; Li, H.; Lin, Y.J.; Shi, R.; Yang, C.; Gao, J.; Zhou, C.; Li, Q; Sederoff RR; Vincent L. Chiang†. Hierarchical transcription factor and chromatin binding network for wood formation in *Populus trichocarpa*. *Plant Cell* **2019**, *31*, 602-626. (Cover story in Plant Cell) Also featured on *ScienceDaily* and *US Department of Energy* homepage
7. Lin, Y.J. *; **Chen, H***.; Li, Q.; Li, W.; Wang, J.P.; Shi, R.; Tunlaya-Anukit, S.; Shuai, P.; Wang, Z.; Ma, H.; Li, H.; Sun, Y.H.; Sederoff, R.R.; Chiang, V.L†. Reciprocal cross-regulation of VND and SND multigene TF families for wood formation in *Populus trichocarpa*. *Proc Natl Acad Sci U S A* **2017**, *114*, E9722-E9729
8. Xu, X.H. *; **Chen, H. *;** Sang, Y.L.; Wang, F.; Ma, J.P.; Gao, X.; Zhang, X. S†. Identification of genes specifically or preferentially expressed in maize silk reveals similarity and diversity in transcript abundance of different dry stigmas. *BMC Genomics* **2012**, *13*, 1-17.

b. Selected Papers related with this project that the applicant served as co-authors

9. Mu, D., Ding, C., Chen, H., Li, Y., & Raley, E. M. 2023. Developing tree improvement strategies for challenging environmental stresses under global climate change: a review from traditional tree breeding to genomics of adaptive traits for the quaking aspen. In The Poplar Genome. In: Porth, I., Klapste, J., Mckown, A. (eds) The Poplar genome. Springer, New York, NY. doi: 10.1007/978-3-031-50787-8
10. 11. Wang, Z.; Mao, Y.; Guo, Y.; Gao, J.; Liu, X.; Li, S.; Lin, Y.J.; Chen, H.; Wang, J.P.; Chiang, V.L. MYB transcription factor161 mediates feedback regulation of secondary wall-associated NAC-Domain1 family genes for wood formation. Plant Physiol 2020, 184, 1389-1406.
11. Yan, X.; Liu, J.; Kim, H.; Liu, B.; Huang, X.; Yang, Z.; Lin, Y.J.; Chen, H.; Yang, C.; Wang, J.P. CAD 1 and CCR 2 protein complex formation in monolignol biosynthesis in Populus trichocarpa. New Phytol 2019, 222, 244-260.
12. Wang, J.P.; Matthews, M.L.; Williams, C.M.; Shi, R.; Yang, C.; Tunlaya-Anukit, S.; Chen, H.; Li, Q.; Liu, J.; Lin, C. Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis. Nature communications 2018, 9, 1-16.
13. Wang, J.P.; Chuang, L.; Loziuk, P.L.; Chen, H.; Lin, Y.C.; Shi, R.; Qu, G.Z.; Muddiman, D.C.; Sederoff, R.R.; Chiang, V.L. Phosphorylation is an on/off switch for 5-hydroxyconiferaldehyde O-methyltransferase activity in poplar monolignol biosynthesis. Proc Natl Acad Sci U S A 2015, 112, 8481-8486.
14. Lin, Y.; Li, W.; Chen, H.; Li, Q.; Sun, Y.; Shi, R.; Lin, C.; Wang, J.P.; Chen, H.; Chuang, L. A simple improved-throughput xylem protoplast system for studying wood formation. Nature protocols 2014, 9, 2194-2205.
15. Li, W.; Lin, Y.; Li, Q.; Shi, R.; Lin, C.; Chen, H.; Chuang, L.; Qu, G.; Sederoff, R.R.; Chiang, V.L. A robust chromatin immunoprecipitation protocol for studying transcription factor–DNA interactions and histone modifications in wood-forming tissue. Nature protocols 2014, 9, 2180-2193.
16. Li, Q.; Lin, Y.C.; Sun, Y.H.; Song, J.; Chen, H.; Zhang, X.H.; Sederoff, R.R.; Chiang, V.L. Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in Populus trichocarpa. Proc Natl Acad Sci U S A 2012, 109, 14699-14704.
17. Sang, Y.L.; Xu, M.; Ma, F.F.; Chen, H.; Xu, X.H.; Gao, X.; Zhang, X.S. Comparative proteomic analysis reveals similar and distinct features of proteins in dry and wet stigmas. Proteomics 2012, 12, 1983-1998.

6. Scholarly Contributions and Community Engagement:

Dr. Hao Chen has contributed significantly to the scientific community through peer reviewing for top journals such as *New Phytologist* and *Plant Cell*, and serving as a guest editor for *Frontiers in Plant Science* and *Forests*. His outreach includes engaging with children through the *Plants4Kids* program and collaborating with the *CATALYST* program to introduce plant bioengineering to students with disabilities. Dr. Chen has enhanced ribo-seq technology, developed systems for poplar wood formation research, and contributed to the Arabidopsis ribosomal studies. He has received several prestigious awards, including the IUFRO (2024) and ICAR (2023) Travel Awards, and multiple research assistant scholarships. His presentations span topics such as CRISPR technology, plant resilience, and wood formation, delivered at international conferences and meetings from 2017 to 2024.

IDENTIFYING INFORMATION:

NAME: Wang, Jack

ORCID iD: https://orcid.org/0000-0002-5392-0076

POSITION TITLE: Associate Professor and Director

PRIMARY ORGANIZATION AND LOCATION: NC State University, Forest Biotechnology Group, Raleigh, North Carolina, United States

Professional Preparation:

ORGANIZATION AND LOCATION	DEGREE (if applicable)	RECEIPT DATE	FIELD OF STUDY
North Carolina State University, Raleigh, NC, USA	Postdoctoral Fellow	08/2012 - 07/2017	Forest Biotechnology
North Carolina State University, Raleigh, NC, USA	PHD	07/2012	Forest Biotechnology
Waikato University, Hamilton, Not Applicable, N/A, New Zealand	BS	12/2006	Chemistry and Biology

Appointments and Positions

2024 - present Associate Professor and Director, NC State University, Forest Biotechnology Group, Raleigh, North Carolina, United States

2019 - present Co-Founder and CSO, TreeCo Inc., Raleigh, NC, United States

2018 - 2024 Assistant Professor and Director, North Carolina State University, Forest Biotechnology Program, Raleigh, NC, United States

Products

Products Most Closely Related to the Proposed Project

1. Marques B, Sulis D, Suarez B, Yang C, Cofre-Vega C, Thomas R, Whitehill J, Whetten R, Barrangou R, Wang J. A Protoplast System for CRISPR-Cas Ribonucleoprotein Delivery in Pinus taeda and Abies fraseri. Plants. 2025 March 22; 14(7):996-. Available from: https://www.mdpi.com/2223-7747/14/7/996 DOI: 10.3390/plants14070996
2. Sulis D, Lavoine N, Sederoff H, Jiang X, Marques B, Lan K, Cofre-Vega C, Barrangou R, Wang J. Advances in lignocellulosic feedstocks for bioenergy and bioproducts. Nature Communications. 2025 February 01; 16(1):-. Available from: https://www.nature.com/articles/s41467-025-56472-y DOI: 10.1038/s41467-025-56472-y
3. Li W, Lin Y, Chen Y, Zhou C, Li S, De Ridder N, Oliveira D, Zhang L, Zhang B, Wang J, Xu C, Fu X, Luo K, Wu A, Demura T, Lu M, Zhou Y, Li L, Umezawa T, Boerjan W, Chiang V. Woody plant cell walls: Fundamentals and utilization. Molecular Plant. 2024 January; 17(1):112-140. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1674205223004021 DOI: 10.1016/j.molp.2023.12.008
4. Sulis D, Jiang X, Yang C, Marques B, Matthews M, Miller Z, Lan K, Cofre-Vega C, Liu B, Sun R, Sederoff H, Bing R, Sun X, Williams C, Jameel H, Phillips R, Chang H, Peszlen I, Huang Y,

Li W, Kelly R, Sederoff R, Chiang V, Barrangou R, Wang J. Multiplex CRISPR editing of wood for sustainable fiber production. Science. 2023 July 14; 381(6654):216-221. Available from: https://www.science.org/doi/10.1126/science.add4514 DOI: 10.1126/science.add4514

5. Bing R, Sulis D, Carey M, Manesh M, Ford K, Straub C, Laemthong T, Alexander B, Willard D, Jiang X, Yang C, Wang J, Adams M, Kelly R. Beyond low lignin: Identifying the primary barrier to plant biomass conversion by fermentative bacteria. Science Advances. 2024 October 18; 10(42):-. Available from: https://www.science.org/doi/10.1126/sciadv.adq4941 DOI: 10.1126/sciadv.adq4941

Other Significant Products, Whether or Not Related to the Proposed Project

1. Yu J, Zhou C, Li D, Li S, Jimmy Lin YC, Wang JP, Chiang VL, Li W. A PtrLBD39-mediated transcriptional network regulates tension wood formation in Populus trichocarpa. Plant Commun. 2022 Jan 10;3(1):100250. PubMed Central PMCID: PMC8760142.
2. Lin CY, Sun Y, Song J, Chen HC, Shi R, Yang C, Liu J, Tunlaya-Anukit S, Liu B, Loziuk PL, Williams CM, Muddiman DC, Lin YJ, Sederoff RR, Wang JP, Chiang VL. Enzyme Complexes of Ptr4CL and PtrHCT Modulate Co-enzyme A Ligation of Hydroxycinnamic Acids for Monolignol Biosynthesis in Populus trichocarpa. Front Plant Sci. 2021;12:727932. PubMed Central PMCID: PMC8527181.
3. Liu B, Liu J, Yu J, Wang Z, Sun Y, Li S, Lin YJ, Chiang VL, Li W, Wang JP. Transcriptional reprogramming of xylem cell wall biosynthesis in tension wood. Plant Physiol. 2021 May 27;186(1):250-269. PubMed Central PMCID: PMC8154086.
4. Matthews ML, Wang JP, Sederoff R, Chiang VL, Williams CM. A multiscale model of lignin biosynthesis for predicting bioenergy traits in Populus trichocarpa. Comput Struct Biotechnol J. 2021;19:168-182. PubMed Central PMCID: PMC7773863.
5. Wang JP, Matthews ML, Williams CM, Shi R, Yang C, Tunlaya-Anukit S, Chen HC, Li Q, Liu J, Lin CY, Naik P, Sun YH, Loziuk PL, Yeh TF, Kim H, Gjersing E, Shollenberger T, Shuford CM, Song J, Miller Z, Huang YY, Edmunds CW, Liu B, Sun Y, Lin YJ, Li W, Chen H, Peszlen I, Ducoste JJ, Ralph J, Chang HM, Muddiman DC, Davis MF, Smith C, Isik F, Sederoff R, Chiang VL. Improving wood properties for wood utilization through multi-omics integration in lignin biosynthesis. Nat Commun. 2018 Apr 20;9(1):1579. PubMed Central PMCID: PMC5910405.

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to information related to domestic and foreign appointments and positions.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Wang, Jack in SciENcv on 2025-06-10 16:48:39

IDENTIFYING INFORMATION:

NAME: Li, Miaomiao

ORCID iD: https://orcid.org/0000-0003-2132-6168

POSITION TITLE: R&D Associate Staff

PRIMARY ORGANIZATION AND LOCATION: Oak Ridge National Laboratory, OAK RIDGE, Tennessee, United States

Professional Preparation:

ORGANIZATION AND LOCATION	DEGREE (if applicable)	RECEIPT DATE	FIELD OF STUDY
Chinese Academy of Sciences, Beijing, Beijing, China	PHD	01/2018	Genome, genetic and development biology
Capital Normal University, Beijing, Beijing, China	MS	07/2012	Cell Biology
Huaibei Normal University, Huaibei, Huaibei, China	BS	06/2009	Biology

Appointments and Positions

2024 - present R&D Associate Staff, Oak Ridge National Laboratory, OAK RIDGE, Tennessee, United States

2023 - 2024 Research Scientist, New York University, New York, New York, United States

2018 - 2023 Postdoctoral associate, New York University, New York, New York, United States

Products

Products Most Closely Related to the Proposed Project

1. Li M, Yao T, Galli M, Lin W, Zhou Y, Chen J, Gallavotti A, Huang S. Diversification and conservation of DNA binding specificities of SPL family of transcription factors. [Preprint]. 2024 September 16. DOI: 10.1101/2024.09.13.612952
2. Galli M, Chen Z, Ghandour T, Chaudhry A, Gregory J, Li M, Zhang X, Dong Y, Song G, Walley J, Chuck G, Whipple C, Kaeppler H, Huang S, Gallavotti A. Transcription factor binding site divergence across maize inbred lines drives transcriptional and phenotypic variation. [Preprint]. 2024 June 03. DOI: 10.1101/2024.05.31.596834
3. Li M, Yao T, Lin W, Hinckley W, Galli M, Muchero W, Gallavotti A, Chen J, Huang S. Double DAP-seq uncovered synergistic DNA binding of interacting bZIP transcription factors. Nature Communications. 2023 May 05; 14(1):-. Available from: https://www.nature.com/articles/s41467-023-38096-2 DOI: 10.1038/s41467-023-38096-2
4. Li M, Zhang D, Gao Q, Luo Y, Zhang H, Ma B, Chen C, Whibley A, Zhang Y, Cao Y, Li Q, Guo H, Li J, Song Y, Zhang Y, Copsey L, Li Y, Li X, Qi M, Wang J, Chen Y, Wang D, Zhao J, Liu G, Wu B, Yu L, Xu C, Li J, Zhao S, Zhang Y, Hu S, Liang C, Yin Y, Coen E, Xue Y. Genome structure and evolution of Antirrhinum majus L. Nature Plants. 2019 January 28; 5(2):174-183. Available from: https://www.nature.com/articles/s41477-018-0349-9 DOI:

10.1038/s41477-018-0349-9

5. Bradley D, Xu P, Mohorianu II, Whibley A, Field D, Tavares H, Couchman M, Copsey L, Carpenter R, Li M, Li Q, Xue Y, Dalmay T, Coen E. Evolution of flower color pattern through selection on regulatory small RNAs. Science. 2017 November. issn: 0036-8075

Other Significant Products, Whether or Not Related to the Proposed Project

Certification:

I certify that the information provided is current, accurate, and complete. This includes but is not limited to information related to domestic and foreign appointments and positions.

I also certify that, at the time of submission, I am not a party to a malign foreign talent recruitment program.

Misrepresentations and/or omissions may be subject to prosecution and liability pursuant to, but not limited to, 18 U.S.C. §§ 287, 1001, 1031 and 31 U.S.C. §§ 3729-3733 and 3802.

Certified by Li, Miaomiao in SciENcv on 2025-06-10 12:55:48

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: **Zongliang Yue**

eRA COMMONS USER NAME (credential, e.g., agency login): ZONGYUE

POSITION TITLE: **Assistant Research Professor; Research committee board at AI@AU**

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Start Date MM/YYYY	Completion Date MM/YYYY	FIELD OF STUDY
Capital Normal University, Beijing, China	B.S.	08/2008	06/2013	Life Science
Indiana University, IN	M.S.	01/2014	06/2016	Bioinformatics
University of Alabama at Birmingham, AL	Ph.D.	06/2016	10/2020	Biomedical Science
University of Alabama at Birmingham, AL	Postdoctoral Fellow	12/2020	02/2023	Bioinformatics

A. Personal Statement

I am an assistant research professor in the **Health Outcomes Research and Policy department** of Harrison College of Pharmacy at Auburn University and a research committee board member at **AI@AU**. My expertise is in **healthcare data analysis, predictive AI, and systems biology modeling**, focusing on **genomic and clinical big data** to enhance drug discovery and healthcare outcomes. Over the past **13 years**, I have co-authored **42 peer-reviewed publications** and developed multiple bioinformatics tools, including the PAGER suite (PAGER, PAGER Web APP, PAGER-CoV, PAGER-scFGA), BEERE, and DEMA. My work has contributed to understanding **cancer biology**, **biomarker discovery**, and **drug repositioning**, particularly through **gene-set, network, and pathway analysis (GNPA)**. Leveraging single-cell omics data, my work extends to developing AI-driven models for **patient subgroup identification**, **disease progression monitoring**, and **personalized medicine**. At Harrison College of Pharmacy, I engage in **machine learning research** for precision interventions and treatment optimization, advancing **patient stratification**, **computational modeling**, and **AI applications** in healthcare. Given the proven impact of my work, I am confident in my contributions to this project.

Completed Research Support

- UAB CCTS/NCAT pilot award (UM1TR004771)
  Zongliang Yue (Co-PI)
  Super-PAGs: Mapping Multi-tier Network Modules to Characterize Genomics Data of Polygenic Diseases
  Amount: $60,000; Project Period: 05/01/2023 to 04/30/2024

Active Support

- University of Alabama at Birmingham SPARC center (UC267384)
  Zongliang Yue (contract PI)
  Development of PAGER 3.0 in Advancing Functional Genomics Analysis Project
  Amount: $30,000; Project Period: 06/2025 to 06/2026; Person Months: 1.5 months/year.
- DOE JGI
  Zongliang Yue (co-PI)
  Unraveling the Crosstalk in Poplar's Transcriptional Regulatory Network for Drought Tolerance and Wood Formation using DAP-seq Technology
  Amount: sequencing & analysis cost; Project Period: 05/2025 to 05/2027; Person Months: in-kind 0.6 month/year.
- National Academies
  Zongliang Yue (PI)
  Early-Career Research Fellowship Human Health & Community Resilience Track

Amount: $75,000; Project Period: 09/2025 to 08/2027; Person Months: 2.5 months/year.

B. Positions, Scientific Appointments

Positions and Scientific Appointments

2023 – Now Assistant Research Professor, Auburn University
2020 – 2023 Postdoctoral Fellow, University of Alabama at Birmingham

Other Experience and Professional Memberships

Selected Professional Committees and Review Panels

2025 – Present Special Issue Editor, Biology, MDPI
2023 – Present Editor board, JMIR Bioinformatics and Biotechnology
2022 – Present Guest Associate Editor, Frontiers in Artificial intelligence
2020 – Present Editor board, DNA and Cell Biology Reports

As Academic Meeting Organizers

2020 – Present Session chair, 16th - 20th MidSouth Conference on Computational Biology and Bioinformatics (MCBIOS)
2025 – 2028 Board member of MCBIOS association.

Honors

2022 Scientific Excellence Award, UAB Center for Clinical and Translational Science
2022 The third place, the Omics Hackathon, UAB
2021 MCBIOS Interdisciplinary Team Science Award, the AI Against Cancer Data Science Hackathon, UAB
2020 The Second place, the COVID-19 Data Science Hackathon, the University of Alabama at Birmingham
2019 Data Blitz Award, the UAB Aging Symposium
2019 The 4th place oral presentation, MCBIOS 2019
2014 11th MCBIOS Best Paper Award

C. Contributions to Science

1. **Development of Bioinformatics Tools for Functional Genomics and Molecular Network Analysis.** We have developed computational tools for complex diseases to enhance gene-set, network, and pathway analysis (GNPA). PAGER and its variants (PAGER-CoV, PAGER Web APP, and PAGER-scFGA) have been instrumental in unraveling disease mechanisms, including tumor progression in myeloid-derived suppressor cells (MDSC). We continue to innovate by creating databases and web-based applications to support functional genomics research. With experience developing bioinformatics databases and web servers, I am confident in designing databases and web-based application prototypes.
    a. Huang F, Welner RS, Chen JY*, **Yue Z***, (2024) PAGER-scFGA: Unveiling Cell Functions and Molecular Mechanisms in Cell Trajectories through Single-Cell Functional Genomics Analysis, *Front. in Bioinformatics*. https://doi.org/10.3389/fbinf.2024.1336135.
    b. **Yue Z**, Slominski R, Bharti S, Chen JY*, (2022) PAGER Web APP: An interactive, online gene set and network interpretation tool for functional genomics, *Front. in Genetics*. https://doi.org/10.3389/fgene.2022.820361.
    c. **Yue Z**#, Zhang E#, Xu C, Khurana S, Batra N, Dang S, and Chen JY* (2021) PAGER-CoV: A Pathway, Annotated-list and Gene-signature Electronic Repository for Coronavirus Diseases Studies. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkaa1094.
    d. **Yue Z**, Zheng Q, Neylon MT, Yoo M, Shin J, Zhao Z, Tan AC, and Chen JY§ (2017) PAGER 2.0: an update to the pathway, annotated-list and gene-signature electronic repository for Human Network Biology. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkx1040.

2. **Development of Bioinformatics Algorithms to Advance Genomics and Molecular Network Analysis.** We have designed innovative algorithms to implement molecular network analysis, such as the Enrichment of Gene-Gene Co-expression Correlation (E-GGCC) analysis, Distance-bounded Energy-field Minimization Algorithm (DEMA), Biomedical Entity Expansion, Ranking, and Explorations, etc. We've introduced several computational pipelines with new algorithms to enhance visualization and accelerate the disease's molecular mechanism discovery. With experience developing bioinformatics algorithms, I am confident in developing new bioinformatics algorithms to advance genomics analysis.
    a. Weng Z#, **Yue Z#**, Zhu Y*, Chen JY*, (2022) DEMA: a distance-bounded energy-field minimization algorithm to model and layout bio-molecular networks with quantitative features, *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btac261.
    b. **Yue Z**, Willey CD, Hjelmeland AB, and Chen JY*, (2019) BEERE: a Web Server for Biomedical Entity Expansion, Ranking, and Explorations. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz428.

c. **Yue Z**, Neylon MT, Nguyen T, Ratliff T, and Chen JY*, (2018) “Super gene set” causal relationship discovery from functional genomics data, *IEEE Transactions on Computational Biology and Bioinformatics*. https://doi.org/10.1109/TCBB.2018.2858755.
d. Chen JY*, **Yue Z**, Neylon MT, Nguyen T, Bulsara N, Arora I, and Ratliff T, (2016) “Towards constructing ‘Super Gene Sets’ regulatory networks”, *IEEE International Conference on Bioinformatics and Biomedicine*. https://doi.org/10.1109/BIBM.2016.7822534.

3. **Actionable Gene Module Discovery and Drug Repositioning.** We have developed a silico drug-protein map database called “Drug Directionality Map” or “DMAP”. We’ve found that DMAP uniquely identifies several new drug indications. Built on top of DMAP, we develop a gene module-based drug repositioning pipeline designed to address the challenges of insufficient outcomes typically associated with targeting a single gene in polygenic diseases. We applied the pipeline to Parkinson’s disease (PD), and the pipeline identified critical network modules in the PD pathway and several families of drug candidates. With experience in network mining and drug repositioning, I am confident in developing bioinformatics analytics workflows to discover actional gene modules.
    a. **Yue Z**, Yan D, Guo G, and Chen JY, (2023) Biological Network Mining. *Biostatistics Research*. https://doi.org/10.37256/bsr.1120231921.
    b. **Yue Z**, Yan D, Guo G, and Chen JY, (2021) Biological Network Mining. In: MUKHTAR S. (eds) Modeling Transcriptional Regulation. *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-0716-1534-8_8.
    c. **Yue Z**, Arora I, Zhang EY, Laufer V, Bridges SL, Chen JY§ (2017) Repositioning drugs by targeting network modules: a Parkinson’s disease case study, *BMC bioinformatics*. https://doi.org/0.1186/s12859-017-1889-0.
    d. Huang H, Nguyen T, Ibrahim S, Shantharam S, **Yue Z**, and Chen JY*, (2015) DMAP: a Connectivity Map Database to Enable Identification of Novel Drug Repositioning Candidates, *BMC bioinformatics*. https://doi.org/10.1186/1471-2105-16-S13-S4.

4. **Population Health and AI in Healthcare.** We’ve applied machine learning models to (1) discover the population disparities in vaccination decisions; (2) uncover cognitive impairment among Medicare beneficiaries; and (3) Screen polypharmacy and reveal drug pairs associated with quality-of-life decline. The ML models enhanced our understanding of population heterogeneity, addressing a fundamental need for safe and efficient treatments or tailored interventions. With experience in developing novel methodologies and machine learning (ML) architectures, I am confident in developing new models.
    a. **Yue Z**, McCormick NP, Ezeala OM, Durham SH, Westrick SC. (2024) EMSIG: Uncovering Factors Influencing COVID-19 Vaccination Across Different Subgroups Charactered by Embedding-based Spatial Information Gain Vaccines. *Vaccines.* https://doi.org/10.3390/vaccines12111253.
    b. **Yue Z**, Jaradat S, Qian J. (2024) Prediction of cognitive impairment among Medicare beneficiaries using a machine learning approach. *Archives of Gerontology and Geriatrics*. https://doi.org/10.1016/j.archger.2024.105623.
    c. **Yue Z**, Xue X, Qian J, The association between polypharmacy and health-related quality of life among older adults with prostate cancer, *Journal of Geriatric Oncology*. https://doi.org/10.1016/j.jgo.2024.101772.

5. **Biomarker Discovery using Systems Biology.** We have conducted several comprehensive analyses to reveal biomarkers and their associations with complex diseases, including pan-cancer studies, glioblastoma, colorectal cancer, leukemia, etc. To enhance biomarker discovery, we incorporated bioinformatics tools for biomarker screening regarding cancer diagnostics, metabolic profiling, and therapeutic target identification. This study demonstrated that I could develop robust pipelines by combining network-based approaches and computational modeling, ultimately facilitating molecular signaling detection and advancing biomarker-driven research.
    a. Huang F, Xu P, **Yue Z**, Song Y, Hu K, Zhao X, Gao M, Chong Z. (2024) Body Weight Correlates with Molecular Variances in Patients with Cancer. *Cancer Res.* Mar 4;84(5):757-770. https://doi.org/10.1158/0008-5472.
    b. Stackhouse CT, Anderson JC, **Yue Z**, Nguyen T, Eustace NJ, Lanov L, Langford CP, Wang J, Rowland JR, Xing C, Mikhail FM, Yang ES, Hjelmeland A, Miller CR, Chen JY, Gillespie GY, and Willey CD, (2022) A novel in vivo model of Glioblastoma radiation resistance identifies long non-coding RNAs and targetable kinases, *JCI Insight*. https://doi.org/10.1172/jci.insight.148717.
    c. Zindl CL#, Witte SJ#, Laufer VA#, Gao M, **Yue Z**, Janowski KM, Cai B, Frey BF, Silberger DJ, Harbour SN, Singer JR, Turner H, Lund FE, Vallance BA, Rosenberg AF, Schoeb TR, Chen JY, Hatton RD, and Weaver CT, (2022) A nonredundant role for T cell-derived interleukin 22 in antibacterial defense of colonic crypts, *Immunity*. https://doi.org/10.1016/j.immuni.2022.02.003.
    d. Patel SB, Nemkov T, Stefanoni D, Benavides GA, Bassal MA, Crown BL, Matkins VR, Camacho V, Kuznetsova V, Hoang AT, Tenen DE, Wolock SL, Park J, Ying L, **Yue Z**, Chen JY, Yang H, Tenen DG, Ferrell PB, Lu R, Darley-Usmar V, Alessandro AD, Bhatia R and Welner RS, (2021) Metabolic alterations mediated by STAT3 promotes drug persistence in CML, *Leukemia*. https://doi.org/10.1038/s41375-021-01315-0.

Full publication list: https://www.ncbi.nlm.nih.gov/myncbi/zongliang.yue.1/bibliography/public/