Introduction: Unraveling the genotype-phenotype map (GPM) of complex traits has been the goal of quantitative geneticists for decades. Thus far, genome-wide association studies (GWAS) have predominately been utilized to dissect the genetic basis of complex traits. Epistasis and pleiotropy are typically unaccounted for in GWAS and as such prevent adequately describing the GPM[1]. Therefore, more advanced tools that provide causal and mechanistic insight into the GPM are needed to model epistasis and pleiotropy. Although GWAS of simply inherited phenotypes have provided us with valuable information, we must integrate GWAS with transcriptomic and metabolomic data, and most importantly the computational modeling of systems biology[2,3] to better connect genotype to phenotype.

Improved knowledge of the GPM is particularly important for nutritional crop improvement. Kale (*Brassica oleracea* var. acephela) is an economically valuable vegetable crop with vast genomic resources and high secondary metabolite diversity making it an ideal organism to use for this study[4,5]. In particular, kale contains the anti-carcinogenic glucosinolates and nutrients commonly deficient in human diets (provitamin A and other carotenoids, flavonoids)[6]. Kale's global popularity, nutritional importance, and highly variable nutrient profile also makes it an ideal candidate for biofortification —the enhancement of crop nutritional quality through plant breeding.

Objective: To produce a more accurate GPM for the glucosinolate, carotenoid, and flavonoid pathways of kale.

Aim: I will integrate GWAS with mechanistic metabolic models to further unravel the GPM of kale nutrient profiles. The results of this experiment will shed light on whether insights into metabolic function and its genetic basis can be gained through both statistical associations between genotypic and phenotypic variation and the mathematical analysis of mechanistic metabolic networks.

Methods: This project will be conducted with [redacted] as part of a Ph.D. in Plant Breeding and Genetics. **Genome-wide association:** My project will capitalize on [redacted] lab's established methodology and experience with integrating transcriptomic and metabolomic data into enhanced GWAS[7]. I will grow a 300-member kale panel from the USDA germplasm in a replicated complete block design at [redacted] and use an established genotyping-by-sequencing (GBS), RNAseq, and high performance liquid chromatography (HPLC) pipeline on leaf tissue. These analyses will generate genome-wide SNP, RNAseq, and metabolomic data sets for downstream metabolic models and GWAS. I will perform a preliminary GWAS using a mixed-linear model to control for population structure and relatedness to generate a list of genes associated with nutrient profiles. **Generating metabolic models*:*** To complement the GWAS results, I will construct, parameterize, and validate metabolic network models for glucosinolate, carotenoid, and flavonoid biosynthetic pathways for each member of the diversity panel using SloppyCell[8]. SloppyCell is a Python-based computational systems biology program created by Drs. [redacted] (both at [redacted]) capable of fitting model parameters to experimental data. These models are created by exploiting "sloppiness" in mechanistic models, where system behaviors are robust to variation in certain parameters but highly sensitive to other parameter changes. Sloppiness allows for greater emphasis to be placed on prediction rather than parameters and allows for the powerful predictive capability of the created metabolic network models. **Integrating systems biology and quantitative genetic data sets:** I will integrate data sets with two additional genome-wide associations using the same linear-mixed model controlling for population structure and relatedness. I will (1) perform a GWAS using parameter ensembles from SloppyCell as the dependent variable to associate parameter variation

with genetic variation. (2) Perform another GWAS using RNAseq transcript abundance (from Cufflinks) as the independent variable to associate transcript abundance with nutrient profiles as well as parameter ensembles[9].

Anticipated Results: The preliminary GWAS should produce a list of candidate genes and alleles related to different metabolic profiles and nutrient levels. It is expected that some of the candidate genes will be similar to GWAS results for provitamin A in maize and glucosinolates in *Brassica rapa* (unpublished data, Bird). The results of the SloppyCell modeling will be a metabolic model identifying possible control points in the biosynthetic pathways and sensitive parameters that can predict the phenotypic effect of changes in parameter ensembles. The parameter ensemble GWAS holds the most promising results. A previous study indicates that a GWAS on model parameters could provide more causal variants and create a better foundation for prediction methods than a GWAS on the physical phenotype in question[10] The transcript abundance GWAS will provide loci connected to nutritional variation and parameter ensemble variation missed by traditional GWAS due to minimal SNP diversity. These insights should also help identify epistatic and pleiotropic genes missed by traditional GWAS.

Intellectual Merit: This project would be one of the first studies to combine mechanistic metabolic models, transcriptomics, and metabolomics in a GWAS and the first ever done on plants. It has the potential to improve upon the traditional quantitative genetic methods and provide a new methodology for dissecting the GPM. SloppyCell's ability to produce hundreds of models from minimal background knowledge of parameters could usher in a new era of systems biology integration with plant breeding and genetic studies. Additionally this project will dissect the genetic basis of glucosinolates, carotenoids, and flavonoids with unprecedented accuracy and resolution, making an accurate and robust genotype-phenotype map closer than ever before.

Broader Impacts: I plan on building off of Dr. [redacted] relationship with HarvestPlus, an organization focused enhancing the biofortification infrastructure and broadly disseminating research to breeders in Africa. I will work with members of HarvestPlus to push for adoption of kale as a crop to biofortify based on my results. Kale is commonly eaten in many countries in Eastern and Southern Africa including Congo, Tanzania, and Kenya. The identification of accessions with greater nutrient density would be beneficial to sustainably fight anemia, blindness, and premature deaths caused by micronutrient deficiency. Additionally, I plan to partner with the Gates foundation's [redacted] program at [redacted]. [redacted] aims to promote bioengineered crops in appropriate contexts to politicians and the public. The program creates a global network to promote the teaching, training, and learning of the knowledge and tools of [redacted] to local communities. I will help educate and train international and local farmers and speak to the general public about bioengineered crops to provide a balanced understanding of the benefits and risks of bioengineered crops. The outcome of this program is a greater acceptance and use of bioengineering crops that can improve quality of life where agriculture is crucial to community and personal livelihood. My passion for science communication and genetics, make these programs a natural fit.

References: **1.** Mackay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). Nat Rev Gen, 565-577 **2.** Civelek, M., & Lusis, A. J. (2014). Nat Rev Gen, 15(1), 34–48. **3.** Nadeau J.H. and Dudley A.M. Science (2011). 331 (6020), 1015-1016 **4.** (http://faostat.fao.org/). **5**. Liu, S., et. al (2014). Nat Comm, 5(3930), 1-11 **6.** Becerra-Moreno, A., et al., (2014). CyTA - Journal of Food, 12(3), 298–303 **7.** Diepenbrock, C. H., & Gore, M. A. (2015). Crop Sci, 55(4), 1-12 **8.** Myers, C.R., Gutenkunst, R.N., and J.P. Sethna. (2007). Comput. Sci. Eng., 9(3), 34-37: **9.** Hirsch, C. N. et. al (2014). Plant Cell, 26, 121–35. **10.** Wang, Y.et al. (2012). PLoS Comput Biol, 8(4). e1002459