**Aim 2. Nutriomics: integration of genome-wide responses to NPK combinations.**

***Rationale:*** In this Aim, we begin to explore the molecular basis for the NPK nutrient-enhancement effect(s) on of biomass quantified in Aim 1, using transcriptome analysis. We propose to uncover the transcriptional responses to NPK nutrient/signals (e.g. the “nutriome”) in both young seedlings and during later developmental stages. Our goal is to identify early gene markers for the NPK effect, which correlate with early morphological traits associated with biomass. On a practical level, the discovery of such molecular predictors of biomass at the seedling stage will represent an invaluable tool for anticipating biomass production in genetic screens and field studies to accelerate the isolation of high yield crops (or to estimate the yield of a crop). We will also analyze transcriptome data using Gene Ontology Analysis to identify the underlying metabolic and cellular processes whose expression correlates with biomass. These studies will identify the target genes and pathways that will be investigated in our time-series analysis in Aim 3, where we aim to predict the TF networks controlling the early gene markers and metabolic pathways associated with biomass.

***Approach:*** To generate the NPK “nutriome” datasets, we will perform transcriptome analysis on Arabidopsis seedlings transiently treated with the complete NPK matrix from Aim 1A, to acquire a genome-wide view of the plant adaptation to each NPK nutritional regime. For the later, developmental series, we will restrict our transcriptome analysis to the 3 NPK:phenotype states selected in Aim 1B and will determine how and when the the “nutriome” landscapes reach equilibrium. In order to integrate these nutriome datasets with morphometric parameters (from Aim 1), parametric and non-parametric correlation will allow us to identify early molecular sentinels of and metabolic/cellular pathways that correlated with changes in biomass. The results of this analysis will show how the execution of a specific genetic program can shape a number of cellular machines - metabolic or signaling pathways - to support growth in relation to N availability.

**Aim 2A. The Seedling Nutriome: Identifying early molecular predictors of biomass.** In this subaim, we will measure genome-wide expression that arise in seedlings transiently treated with all the combinations of the NPK matrix. This dataset will provide a transcriptional baseline to identify the genes that are highly induced or repressed in each NPK combination. Using these data, we will proceed to establish an association between gene expression and seedling traits that are early markers for biomass (Aim 1B). To gather transcriptomic data, Arabidopsis seedlings will be grown and transiently starved for NPK, as determined in Aim 1. Following starvation, seedlings will be treated with all NPK combinations from the nutrient matrix and harvested after a 2 hours treatment to extract mRNA for transcriptome analysis. As controls, seedlings that were nutrient starved but not resupplied with any NPK combination will be used. Trancriptome analysis will be performed using either ATH1 chips, or deep-sequencing. Our lab currently uses both methods, and as costs decrease, we anticipate fully switching to deep-sequencing.

Dennis- please, check this section in yellow and fill in.

We will integrate transcriptome analysis with the root and shoot traits measured across treatments, using parametric and non-parametric correlation using correction for multiple testing [Statistical significance for genomewide studies"by John Storey and Robert Tibshirani PNAS August 5, 2003 9440-9445] This analysis will identify genes that are significantly correlated or anti-correlated (>0.8 or <-0.8) with the morphometric root and shoot parameters that were indicated as early predictor of biomass in Aim 1C.

This analysis will reveal genetic markers that are associated with morphometric root and shoot traits that are in turn associated with high or low biomass. The analysis will therefore also generate genetic markers for biomass itself.

**Aim 2B. The Developmental Nutriome: Identifying metabolic and cellular pathways associated with selected NPK:biomass states.**  To identify the metabolic and cellular processes that correlate with the NPK effect on biomass, we will generate a developmental series of root and shoot transcriptome for selected NPK:phenotype states that correlate as follows: 1. High-N:High biomass, 2) Low-N:Low biomass, and 3) LowN:High (See Aim 1). mRNA samples will be extracted from 1-, 2-, 3-, 4-, 5- and 6-week old plants, grown hydroponically on the three specific NPK combinations associated with the above- mentioned states to be analyzed by ATH1 chips or RNA-seq. Using correlation with correction for multiple testing as above above, we will find genetic markers for biomass at each developmental time point. We expect a large overlap in both positive genetic markers (i.e. genes that are induced for plants having high biomass) and negative ones, but some genes may turn on or off later in development. We will use the genes that are consistently positive or negative markers as sentinel genes in aim 3.

Gloria, Dennis-

* what if these are not the same genes found in Aim 2A that correlate with the genes that correlate with the early predictors of biomass from Aim 1C? Instead, should we say we are going to investigate the correlation between the early and the late molecular markers using these datasets and the statistics?
* What will we used as sentinels in Aim 3? I would say the genes that are found to be markers in the earliest possible time points and are consistent throughout

**Aim 2D. Identification of biological process that are correlated with biomass production.**

We next aim to identify the metabolic and cellular pathways that correlate with the three states (1. High-N:High biomass, 2) Low-N:Low biomass, and 3) LowN:High biomass. Toward this goal, we will exploit GO term analysis tools (e.g. BioMaps and Sungear) in VirtualPlant (www.virtualplant.org) (Katari *et al.*, 2010) to identify “GO markers” for high biomass. Analogously to our correlation analysis for genetic markers, we will identify over-represented and under-represented GO terms that are associated with high biomass over developmental time. Over-representation will be determined by a probability measure based on a hypergeometric distribution.