**Aim 1.** **The nutriome-to-phenome matrix: optimizing N, P and K combinations to support Arabidopsis shoot-root phenotypes.** We are going to explore the effect of NPK combinations on the allometry of young plants and biomass production of adult plants to develop computational models that uncover early markers of biomass.

* 1A. Nutriome matrix
* 1B. Morphometrics (seedlings) and biomass (mature plants).
* 1C. Computational models to find
1. early morphometrics markers; correlate morphometrics PCs and biomass measurements
2. identify the NPK combinations that determine the NPK:phenotype states (High-N:High biomass, Low-N:Low biomass and LowN:High biomass)

Dennis- surely there will be more than 1 combination per state and combinations that will give high shoot and other that will give high root biomass (although root biomass will never surpass shoot biomass in weight but it could be an interesting trait): how do we pick one?

Dennis responds: if we’re going to do root and shoot biomass separately, then we will need more treatments. I strongly urge that we just do total biomass. Then we have a coherent story. If there are several conditions that are equal or nearly equal in terms of biomass, then we always choose the one that require the least nutrients.

Dennis:

1. More treatments? We are already doing ALL combinations of NPK, why would doing shoot and root biomass separately require more treatments?
2. Shoot/root: It is very possible that this is a trade off…..e.g. more shoots/less roots or vice versa….so this is important also to correlate with the early seedling morphometrics of root and shoot. If we only do total biomass how can we compare this to the shoot/root morphometric data. Finally, if we measure shoot and root biomass separately, we can always combine those to determine total…so it is a win/win.
3. Transcriptomics will also be done separately on shoots and roots where we know gene expression varies drastically.

Dennis responds: we would need more treatments for the time series potentially, because now “high biomass” has no meaning. So, if we did separate analyses for shoot biomass, root biomass, and total biomass but then did the time series just based on total biomass, that works.

**Aim 2.** **Nutriomics: Integration of genome wide responses to macronutrient combinations that enhance N-use efficiency**. Our goal is to characterize early and late genetic changes that underly biomass production for NPK:phenotype states (selected from Aim 1) to identifying the genes and process that are involved in building this trait over extended periods of time using parametric and non-parametric correlation as well as regression with loss functions to correlate them with biomass.

* 2A. Transcriptomics of all NPK combinations at 30 min
	+ Dennis, we don’t understand how the 30 min data are going to be used since the developmental series will be done on 3 NPK combination only! Stochastic gradient descent? Well, I was thinking that getting all the combinations will enable us to zero in on the processes and genes that differ most in the low biomass cases from the high ones. The goal here is to identify gene and GO process markers.

Dennis-

1. All NPK matrix transcriptomics: OK, if we do ALL transcriptomics, 12 x 3 samples, what methods will you do to “zero” in on the Biomarkers? Correlation or “stochastic gradient descent? PLEASE give us a specific response to the computational method, as this grant is going to a computational panel! The Networks and regulation panel is NOT about physiology, but is about the interplay of computation and experimentation. Right now, the write up is heavy on experimentation and light on the computational approaches we will use and develop. I think we will start with correlation, because for the purpose of finding markers that’s all we need. If correlation doesn’t yield anything, then we will try to find Boolean combinations of a few marker genes (e.g. gene g1 or gene g2). We’ll have to be careful about multiple testing.
* 2B. Developmental series transcriptomics for selected NPK:phenotype states (Aim 1)
* 2C. Computational models
1. biological processes (GO terms); correlation to determine which GO Terms are positively associated with each NPK:phenotype state over developmental time. What method? Correlation (parametric and non-parametric)
2. early markers of biomass; find key genes by comparing early, developmental and biomass (Aim 1) what method? Correlation as well as regression with loss function.

**Aim 3.** **Generation of dynamic networks and predictive models of macronutrient signaling.** We will infer the genetic networks underpinning the immediate response to nutrient input to find early markers of biomass yield using expression data from plants transiently treated with the selected NPK combinations. We will identify TF to be tested in Aim 4.

* 3A.Transcriptomics of time series (3,6,9…..min) under selected NPK conditions
* 3B. Causal networks; state-space comparing selected NPK treatments
* 3C: In silico validations…. Consisting of prediction of the direction of expression in time points outside of a training set.

**Aim 4. Validation.**

* 4A. Mutants of most influential TFs
* 4B Overexpressessor of TFs
* 4C: Computational: Model refinement and new predictions?
	+ Dennis: How will you use genomic analysis of mutant/transgenics to validate, refine, and generate new predctions? The overexpression data will constitute a test of our edges, e.g. if we believe that g1 induces g2, but overexpressing g1 does NOT increase the expression of g2, then we have to refine our hypotheses. As we refine the network, we will identify new critical transcription factors to overexpress.

**Dennis-** We completely understand this….the question is how will we use transcripome data from the mutants and transgenics to come up with global regulatory models, not gene by gene models….? I don’t understand the question. A global regulatory model tells about the relationship or non-relationship between every pair of genes. [Please import the section on state space from our last year’s grant and I will edit from there]