* 1. Hotspot analysis
     1. Are we seeing cis Fx for Botrytis?
     2. What is the distribution of hotspots in Botrytis?
        1. Frequency?
        2. Location?
        3. Compared to random?
     3. What does the hotspot distribution look like for Arabidopsis?
        1. Frequency?
        2. Location?
        3. Compared to random?
     4. Overlap in hotspots, At to Bc?
        1. Still figuring out method for this…
  2. Figure T4. Frequency of distance between gene and top SNP: transcriptome-wide cis search
     + 1. From top 10 SNPs/ gene, lsm
       2. For Col0?
       3. From top 1/ 100 SNP/ gene
  3. Figure T5. Cis-diagonal plot for gene to top SNP: transcriptome-wide cis search
     + 1. From top 10 SNPs/gene, lsm
       2. Col0
       3. Top 1/ 100 SNP/gene
  4. Focus on cis effects within BoA/ Bot/ Net5
  5. Focus on Botcynic Acid cis effects
     1. Figure T1. Pvclust tree BoA
        1. Isolates cluster into 4 or 5 groups based on SNP diversity
        2. Fd1 and B05.10 group
        3. UKRazz, 1.03.04, and 2.04.09 subgroup
        4. 3 major groups
           1. ANOVA: no significant expression difference between these 3 groups
        5. Deletion group
     2. Figure T2. Excel gene deletions
        1. One sub-cluster defined by ~whole-cluster deletion. Deletion is polymorphic.
        2. Teleomeric loss
        3. Additional isolates outside cluster with deletions – independent events
           1. UKRazz, 1.03.04, and 2.04.09
     3. Figure T3. Violin plot of whole-network expression based on clustering
        + 1. UKRazz, 1.03.04, and 2.04.09 low expression with independent origin?

Using GEMMA, we associated these transcription profiles of XX genes with variable expression in B. cinerea on A. thaliana with genome-wide SNP variation in the B. cinerea genome. In these transcript-to-SNP associations, we find that cis- control of the loci is largely drowned out by patterns of trans-acting variation. We first focused on the single top SNP hit per transcript, with the highest probability (lowest p-value) of significant effect on expression in the gene of interest. If control of gene expression is localized to the gene itself or cis-acting loci, we would expect a strong linear association between the center of each gene and the genomic location of its top SNP hit. However, we find that few genes have a top SNP hit within the same chromosome, and even fewer within 1Mb (Figure X5).

Figure X5. Cis-diagonal plot correlating gene position to position of top associated SNP. We retained only the single SNP with highest probability (lowest p-value) of significant effect on expression of the transcript of interest. Each point represents a single transcript from our B. cinerea expression profile, with y axis of transcript center and x axis of top SNP location. Points are color-coded according to the chromosome on which the transcript is located. The opacity of each point represents proximity between the transcript and top SNP hit, from low (SNP is not located on the same chromosome as the transcript) to high (SNP within 1Mb of the transcript). This opacity setting should magnify any patterns of cis-effect variation, if present.

Methods

We focused further cis-effects analysis on three networks which were highly conserved across *B. cinerea* isolates {Zhang 2018}. We clustered isolates by SNP data within focal networks. Hierarchical clustering was computed using the R package pvclust based on mean linkage (UPGMA), with correlation distance and 1000 bootstrap replications {Suzuki 2015}. AU p-values are reported in red, BP values in green. Edges with high AU values are considered strongly supported by the data, and clustering is drawn according to these edges with AU > 95%.