**BRC1 downstream transcriptional network.**

**Data acquisition and pre-proccessing**

In order to select genes in *Arabidopsis thaliana* to be included in the construction of a transcriptional network downstream the regulation exerted by BRC1 a co-expression analysis was performed. Details on how it was performed by Carlos.

These analysis resulted in the identification of 931?? genes significantly co-expressed with BRC1 that can be organised into 9 different clusters (details on the clustering here). These clusters are enriched in specific GO terms suggesting that they are involved in distinct biological processes. Transcription factors were identified among these genes and the public freely available databases GEO (Gene Expression Omnibus) and SRA (Sequence Read Archive) were systematically searched for ChIP-seq and DAP-seq data generated for them in previously published studies.

Table ??? presents the 37 different transcription factors for which ChIP-seq or DAP-seq data are included in this analysis together with the ChIP-seq and RNA-seq data from BRC1 generated in this study.

|  |  |  |  |
| --- | --- | --- | --- |
| **TF AGI ID** | **TF Name** | **Accession Number** | **Data Set Type** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

The transcription factors potential targets were determined from ChIP-seq data

**Network construction and visualization**

A transcriptional network was constructed based on the potential targets of each transcription factor from table 1 determined as described previously. In our network nodes represent genes and an edge or arrow is drawn from genei to genej when genei codifies for a transcription factor that binds to the promoter of genej, that is genej is a potential target regulated transcriptionaly by the transcription factor codified by genei. The R package igraph for network construction and analysis was used in this part of our study. The network adjacency matrix of order 931x931; number of rows (genes) x number of columns (genes); was constructed by setting component (i,j) to 1 when genei codifies for a transcription factor that binds to the promoter of genej else it was set to 0. The function graph.adjacency with input the previously described adjacency matrix and mode equal to directed was used to generated our network comprising 931 nodes or genes and 7414 edges or potential direct transcriptional regulations, Figure 1.

A web app was developed using the R package shiny to analyse this network and to enable other researcher to explore it independently. This tool is freely available at ….

**Network topological analysis**

In order to identify the most relevant genes in our transcriptional network we computed the *closeness* of each node using the R package igraph. This topological parameter provides a measure of the centrality of a node in a network. The distance between two nodes in a network is defined as the length of the shortest path connecting them, that is, the minimun number of arrows between the two given nodes. According to this, the closeness centrality of a node is defined as the inverse of the sum of all distances between the given node and every other node in the network. For normalization purposes, typically, the previous value is multiply by the number nodes in the network:

The highest the closeness of a node the smallest the distance from it to the rest of nodes in the network. Therefore, the higher the closeness of a node the more central it is in the network. In transcriptional networks genes represented by nodes with high closeness tend to constitute master regulators of the entire network as any other gene can be regulated by them through short regulatory cascades, i.e., the corresponding nodes are connected through short paths.

The gene in our network with the highest closeness is *BRC1*. On the one hand, this indicates that gene selection and network construction have been performed using sound methods since from the beginning our goal was the construction of the gene network around BRC1. On the other hand, this constitutes evidence of the high relevance of BRC1 in the response it triggers since none of the transcription factors downstream whose potential targets have also been added to the network surpasses its centrality and BRC1 remains as the master regulator.

Next, we tested the distribution of the top 5% most central genes, nodes with closeness greater than the 95% quantile, over the network. This distribution is not random instead, according to Fisher’s exact test, cluster number 6 is significantly enriched in central genes with a FDR (BH adjusted p-value) of 9.49 ∙ 10-3 and an enrichment value of 2.58, see Table. Key regulators in the network downstream from *BRC1* such as *ABI5*, *DREB2A,* *NAC032* and *NAC102* are located in this cluster and exhibit a high centrality in the network.

In order to determine the less central or distant genes in our network we computed for each node its *eccentricity* using the R package igraph. This topological parameter is defined as the maximum distance between the given node and any other node in the network. Thus the greater the eccentricity of a node the less central it is in the network.

Similarly, we tested using Fisher’s exact test the significance of the distribution of the top 5% most eccentric genes over our network resulting in significant enrichments for clusters 7 and 8 with FDR 1.52 ∙ 10-9  and 3.9 ∙ 10-6 and enrichement values of 2.23 and 1.83 respectively.

Summing up, our topological analysis confirms the central role played by *BRC1* as the master regulator of the genes contained in our transcriptional network since it presents the highest closeness value. Additionally, cluster 6 together with *BRC1* can be considered central regulators of our network potentially constituted the first steps in the response triggered by BRC1 since this cluster is significantly enriched in central genes such as as *ABI5*, *DREB2A, NAC032* and *NAC102*. Finally, cluster 7 and 8 are constituted by genes located farthest away downstream the regulation of *BRC1* since they are significantly enriched in highly eccentric genes.

For details, see supplementary tables: intersections\_Closeness0.9.txt, intersections\_Closeness0.95.txt, intersections\_Eccentricity0.9.txt and intersections\_Eccentricity0.95.txt.

|  |  |  |  |
| --- | --- | --- | --- |
| **Closeness** | | | |
| **Cluster** | **FDR** | **Enrichment** | **Genes in the top 5%** |
| 6 | 9.49 ∙ 10-3 | 2.58 | AT1G01720 (ATAF1), AT1G27730 (STZ), AT1G36060 (TG),  AT1G77450 (NAC032), AT2G18550 (HB21),  AT2G36270 (ABI5), AT4G11660 (HSF7), AT4G36740 (HB40),  AT5G04340 (ZAT6), AT5G05410 (DREB2A),  AT5G63790 (NAC102), AT5G66700 (HB53) |
| **Eccentricity** | | | |
| **Cluster** | **FDR** | **Enrichment** | **Genes in the top 5%** |
| 7 | 1.52 ∙ 10-9 | 2.23 | AT1G01080, AT1G04420, AT1G18485, AT1G20020, AT1G48350, AT1G55490, AT1G63610, AT1G64510, AT1G69200, AT1G75350, AT2G02500, AT2G03420, AT2G21385, AT2G24060, AT2G33450, AT2G36990, AT2G44650, AT3G02450, AT3G04260, AT3G08740, AT3G09050, AT3G14330, AT3G16000, AT3G18110, AT3G18680, AT3G28760, AT3G44890, AT3G47450, AT3G47650, AT3G53130, AT3G54090, AT3G55330, AT3G57180, AT3G62910, AT3G63160, AT4G03150, AT4G04350, AT4G15510, AT4G18480, AT4G20130, AT4G24090, AT4G33470, AT4G34830, AT4G39620, AT5G02710, AT5G14320, AT5G38290, AT5G44650, AT5G47190, AT5G49030, AT5G64580 |
| 8 | 3.9 ∙ 10-6 | 1.83 | AT1G01370, AT1G03780, AT1G05440, AT1G07370, AT1G08560, AT1G10780, AT1G21740, AT1G23410, AT1G44900, AT1G47210, AT1G51060, AT1G55200, AT1G67180, AT1G67630, AT1G76310, AT1G78430, AT1G80370, AT2G16270, AT2G24490, AT2G29570, AT2G31270, AT2G33560, AT2G38810, AT3G03130, AT3G14190, AT3G14740, AT3G51280, AT3G52170, AT3G53650, AT3G56870, AT4G05190, AT4G15830, AT4G28310, AT4G29170, AT4G37210, AT4G37490, AT5G01910, AT5G10400, AT5G23910, AT5G25090, AT5G43080, AT5G44560, AT5G52950, AT5G55820, AT5G56580, AT5G65350, AT5G66230, AT5G67100, AT5G67270 |

Hemos repetido el análisis teniendo el cuenta el 10% de los genes más centrales y excéntricos pero los resultados no eran tan claros. De todas formas os añadimos las tablas a continuación por si considerais que esta opción es mejor que la del 5%.

|  |  |  |  |
| --- | --- | --- | --- |
| **Closeness (90% quantile)** | | | |
| Cluster | FDR | Enrichment | Genes in the top 10% |
| 1 | 1.43E-3 | 7.03 | AT1G09530 (PIF3); AT1G43700 (VIP1); AT1G69490 (NAP); AT4G01120 (GBF2); AT5G39610 (NAC6) |
| 2 | 4.42E-6 | 12.56 | AT1G22640 (MYB3); AT1G52880 (NAM); AT2G01570 (RGA1); AT2G22430 (HB6); AT2G22540 (SVP); AT2G43060 (IBH1); AT5G60850 (OBP4) |
| 3 | 1.13E-3 | 7.89 | AT1G06850 (bZIP52); AT1G44830 (ERF014); AT5G25190 (ESE3); AT5G47370 (HAT2); AT5G67060 (HEC1) |
| 6 | 2.93E-11 | 16.02 | AT1G01720 (ATAF1); AT1G27730 (STZ); AT1G36060 (TG); AT1G77450 (NAC032) ; AT2G18550 (HB21); AT2G36270 (ABI5); AT4G11660 (HSF7); AT4G36740 (HB40); AT5G04340 (ZAT6); AT5G05410 (DREB2A); AT5G63790 (ANAC102); AT5G66700 (HB53) |
| **Eccentricity (90% quantile)** | | | |
| Cluster | FDR | Enrichment | Genes in the top 10% |
| 5 | 7.49E-4 | 4.40 | AT1G22885, AT3G12500, AT3G14595, AT3G20510, AT4G22160, AT4G23400, AT4G35350, AT5G18970 |
| 6 | 3.03E-6 | 4.67 | AT1G09140, AT1G14200, AT1G53540, AT1G72800, AT2G19310, AT2G46610, AT3G48020, AT4G17550, AT4G27350, AT4G37370, AT5G10695, AT5G17780, AT5G47420, AT5G51440 |
| 7 | 2.71E-46 | 13.83 | AT1G01080, AT1G04420, AT1G18485, AT1G20020, AT1G48350, AT1G55490, AT1G63610, AT1G64510, AT1G69200, AT1G75350, AT2G02500, AT2G03420, AT2G21385, AT2G24060, AT2G33450, AT2G36990, AT2G44650, AT3G02450, AT3G04260, AT3G08740, AT3G09050, AT3G14330, AT3G16000, AT3G18110, AT3G18680, AT3G28760, AT3G44890, AT3G47450,  AT3G47650, AT3G53130, AT3G54090, AT3G55330, AT3G57180, AT3G62910, AT3G63160, AT4G03150, AT4G04350, AT4G15510, AT4G18480, AT4G20130, AT4G24090, AT4G33470, AT4G34830, AT4G39620, AT5G02710, AT5G14320, AT5G38290, AT5G44650,  AT5G47190, AT5G49030, AT5G64580 |
| 8 | 6.81E-40 | 11.39 | AT1G01370, AT1G03780, AT1G05440, AT1G07370, AT1G08560, AT1G10780, AT1G21740, AT1G23410, AT1G44900, AT1G47210, AT1G51060, AT1G55200, AT1G67180, AT1G67630, AT1G76310, AT1G78430, AT1G80370, AT2G16270, AT2G24490, AT2G29570, AT2G31270, AT2G33560, AT2G38810, AT3G03130, AT3G14190, AT3G14740, AT3G51280, AT3G52170, AT3G53650, AT3G56870, AT4G05190, AT4G15830, AT4G28310, AT4G29170, AT4G37210, AT4G37490, AT5G01910, AT5G10400, AT5G23910, AT5G25090, AT5G43080, AT5G44560, AT5G52950, AT5G55820, AT5G56580, AT5G65350, AT5G66230, AT5G67100, AT5G67270 |
| 9 | 4.14E-7 | 5.18 | AT1G06550, AT1G24030, AT1G74000, AT1G74210, AT2G44450, AT3G06035, AT3G09540, AT4G27860, AT4G39320, AT4G39510, AT5G08030, AT5G08640, AT5G17220, AT5G42800, AT5G47730 |

**Network motifs identification and analysis**

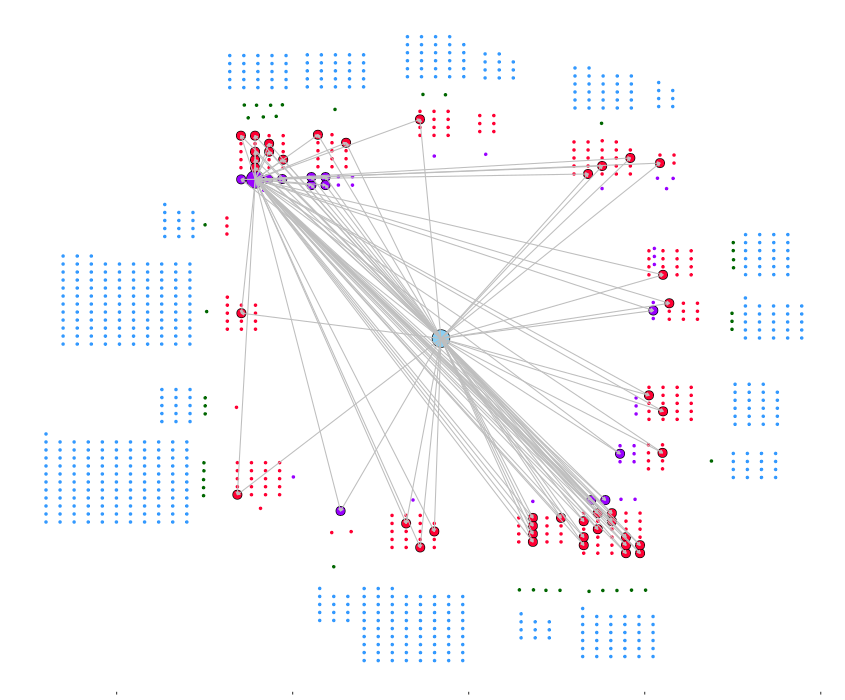
The high number of genes and transcriptional regulations in our network prevents the direct interpretation of its structure and potential dynamics. In order to identify *gene modules* constituted by two transcription factors that significantly regulate the same set of target genes in a coordinated manner we performed a network motif identification analysis. A network motif is defined as a subgraph whose occurrence is significantly greater in our network when compared to its occurrence in random networks with the same topological properties. This was performed using the R package igraph. Specifically, 1000 random networks were generated with the same number of nodes as our transcriptional network, namely 931 nodes. In order to capture the same topology as our network 37 nodes were chosen randomly that would play the role of the transcription factors in our network. In this respect, each one of these randomly chosen nodes was connected to the same number of nodes as one of the transcription factors in our network but these targets were chosen randomly. Following this process we generate random networks with the same number of nodes and edges as the original network and the same number of “*transcription factors”* but with randomly chosen targets. Several network motifs were found in our network, see supplementary table

Among the detected network motifs found in our network the *feed forward loop* (left) and the *feedback loop* (right) with multiple output are specially relevant to shed light on the dynamics that could emerge from the topology of our network.

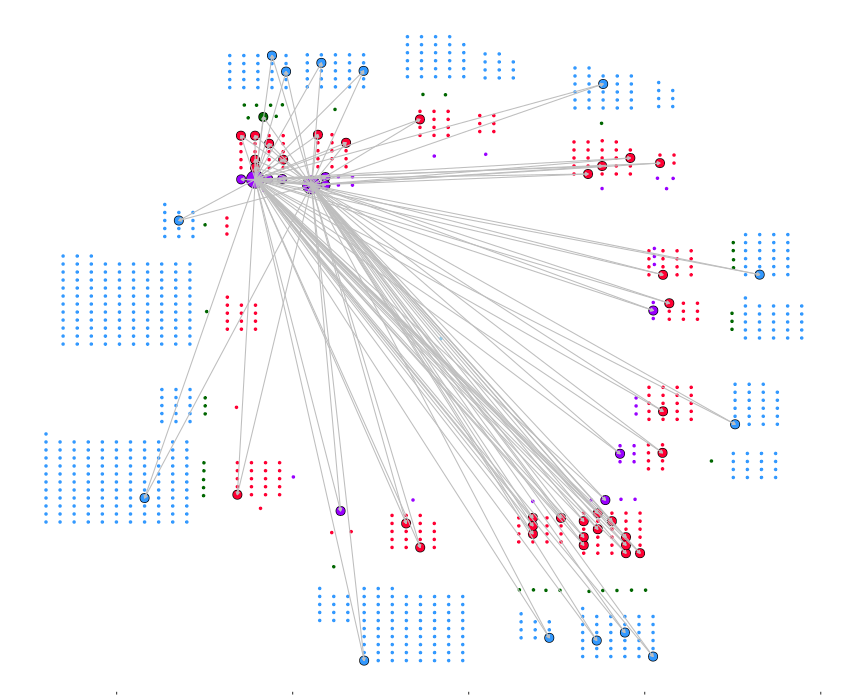


In the feed forward loop (FFL) with multiple output a gene considered the master regulator directly controls a set of target genes and at the same time it binds to an intermediary transcription factor that ultimately regulate in a coordinate manner the same set of target genes. The feed forward loop constitute one of the most common networks motifs identified in transcriptional networks that response to external stimuli in living organisms from bacteria to mammals. Depending on the logic gate of the join regulation exerted by the two transcription factors over their common target genes these network motifs produce relevant dynamical properties of the corresponding response such as noise filtering.

BRC1 appears as the master regulator in a large number of instances of the FFLs found in our network (supp table feedforward\_loops\_with\_multiple\_output.tsv) suggesting that BRC1, as the most central gene in the network, initiates a response to some stumili that is robust against noise and that can be reversible in the first steps. The gene targets of these FFLs are not randomly distributed over the network. For instance, the feedforward loop with multiple output constituted by BRC1 and DREB2A significantly targets genes in cluster 1 and cluster 6. See supplementary tables ffl\_outputs\_vs\_cluster\_(1,2,3,4,5,6).txt. Interestingly, no FFLs was identified significantly regulating the most eccentric clusters 7, 8 and 9 supporting their position farthest downstream away from the signalling response triggered by BRC1.



The feedback loop with multiple output was also significantly identified as a network motif in our network, see table feedback\_loops\_with\_multiple\_output.tsv. In this network motif two master regulators are embraced in a mutual regulation and jointly regulate a set of target genes. This network motif has been specifically identified in developmental networks since it confers the ability to make a transient signal into a stable response. When a transient signal activates one of the two master regulators it also activates the other one. In this way, both transcription factors embrace a mutual activation that would persist even when the initial signal disappears. BRC1 was not identified in any instance of a feedback loop in our network suggesting that the irreversible establishment of a response takes place way downstream from BRC1. Nevertheless, transcription factors directly regulated by BRC1 are the master regulators of feedback loops, for instance, DREB2A and NAC032. The targets of these feedback loops are not randomly distributed over the network instead they significantly regulate specific clusters, see tables feedback\_outputs\_vs\_cluster\_(1,2,3,4,5,6,7,8,9).txt. For example, the feedback loop constituted by DREB2A and NAC032 regulate significantly genes in cluster 1 and 6.



These network motifs do not appear isolated instead they are connected constituting transcriptional cascade that initialy are reversible and filter out noise using feedforward loops such as the one with BRC1 and DREB2A to subsequently activate feedback loops that irreversible fix the initial signal, i.e. the feedback loop constituted by DREB2A and NAC032.