

Supplementary Material Information

- 1. Supplementary Table 1:** A list of the 68 significant TRANSFAC positional weight matrices (PWMs) identified by TF-Spiker and their corresponding TFClass classifications based on the analysis of the breast cancer gene set [1].

PWM	TFClass	
V\$EGR1_01	2.3.1	
V\$EGR1_02	2.3.1	
V\$EGR2_Q6	2.3.1	
V\$EGR_Q6	2.3.1	
V\$FKLF_Q5	2.3.1	
V\$NGFIC_01	2.3.1	
V\$SP1_02	2.3.1	
V\$SP1_03	2.3.1	
V\$SP1_Q2_01	2.3.1	
V\$SP1_Q4_01	2.3.1	
V\$SP1_Q6_01	2.3.1	
V\$SP1_Q6	2.3.1	
V\$SP1SP3_Q4	2.3.1	
V\$SP2_Q3_01	2.3.1	
V\$SP4_Q5	2.3.1	
V\$GC_01	2.3.1	2.3.3
V\$CKROX_Q2	2.3.3	
V\$CTCF_01	2.3.3	
V\$MZF1_Q5	2.3.3	
V\$PLAG1_02	2.3.3	
V\$PLAGL1_03	2.3.3	
V\$WT1_Q6	2.3.3	
V\$ZBP89_Q4	2.3.3	
V\$ZFP281_04	2.3.3	
V\$E2F_03	3.3.2	
V\$E2F1DP1_01	3.3.2	
V\$E2F1_Q6	3.3.2	
V\$E2F3_Q6	3.3.2	
V\$E2F4DP2_01	3.3.2	
V\$E2F_Q2	3.3.2	
V\$E2F_Q4_01	3.3.2	
V\$E2F_Q4_02	3.3.2	
V\$ER81_01	3.5.2	
V\$ETV3_01	3.5.2	
V\$ETV3_02	3.5.2	
V\$FLI1_02	3.5.2	
V\$HIC1_02	2.3.4	
V\$MAZ_Q6_01	2.3.4	
V\$MAZ_Q6	2.3.4	
V\$FOXK1_04	3.3.1	
V\$FOX_Q2	3.3.1	
V\$HNF3_Q6_01	3.3.1	
V\$CMYC_02	1.2.6	
V\$MYCMAX_B	1.2.6	

V\$SMAD1_Q6	7.1.1	
V\$SMAD4_Q4	7.1.1	
V\$NFKAPPAB_Q1	6.1.1	
V\$P50RELAP65_Q5_Q1	6.1.1	
V\$AR_Q1	2.1.1	
V\$CACBINDINGPROTEIN_Q6	4.2.1	
V\$CBF_Q1	6.4.1	
V\$CDX_Q5	3.1.1	
V\$E2A_Q2	1.2.1	1.2.2
V\$MYOGENIN_Q6	1.2.2	
V\$EBF_Q6	6.1.5	
V\$ETF_Q6	3.6.1	
V\$LRH1_Q5	2.1.5	
V\$MEF2C_Q4	5.1.1	
V\$NKX61_Q3	3.1.2	
V\$PAX4_Q5	3.2.1	
V\$PBX1_Q4	3.1.4	
V\$PPARDR1_Q2	2.1.2	
V\$POU1F1_Q6	3.1.10	
V\$RFXDC2_Q4	3.3.3	
V\$TAL1_Q6_Q1	1.2.3	
V\$TCFAP2C_Q3	1.3.1	
V\$ZFP740_Q3	2.3.2	
V\$NFE4_Q5	unknown	

2. **Supplementary Table 2:** A list of the 65 significant TRANSFAC PWMs identified by TF-Spiker and their corresponding TFClass classifications based on the analysis of the NF- κ B gene set [2].

PWM	TFClass
V\$CKROX_Q2	2.3.3
V\$CTCF_Q1	2.3.3
V\$FPM315_Q1	2.3.3
V\$LRF_Q2	2.3.3
V\$MTF1_Q2	2.3.3
V\$MZF1_Q5	2.3.3
V\$PLAG1_Q1	2.3.3
V\$PLAG1_Q2	2.3.3
V\$WT1_Q6	2.3.3
V\$ZBP89_Q4	2.3.3
V\$ZFP161_Q4	2.3.3
V\$ZIC1_Q4	2.3.3
V\$EGR1_Q2	2.3.1
V\$EGR_Q6	2.3.1
V\$FKLF_Q5	2.3.1
V\$NGFIC_Q1	2.3.1
V\$SP1_Q2	2.3.1
V\$SP1_Q3	2.3.1
V\$SP1_Q2_Q1	2.3.1
V\$SP1_Q4_Q1	2.3.1
V\$SP1_Q6_Q1	2.3.1
V\$SP1_Q6	2.3.1
V\$E2F1DP1_Q1	3.3.2

V\$E2F1_Q3_01	3.3.2
V\$E2F1_Q3	3.3.2
V\$E2F1_Q4	3.3.2
V\$E2F3_Q6	3.3.2
V\$E2F4DP2_01	3.3.2
V\$E2F_Q2	3.3.2
V\$HIC1_02	2.3.4
V\$HIC1_06	2.3.4
V\$MAZ_Q6_01	2.3.4
V\$MAZ_Q6	2.3.4
V\$ZFP128_04	2.3.4
V\$NFKAPPAB_01	6.1.1
V\$NFKB_C	6.1.1
V\$NFKB_Q6	6.1.1
V\$P50RELAP65_Q5_01	6.1.1
V\$AP2GAMMA_Q5_01	1.3.1
V\$AP2_Q3	1.3.1
V\$AP2_Q6	1.3.1
V\$TCFAP2A_03	1.3.1
V\$AHRARNT_01	1.2.5
V\$AHR_Q5	1.2.5
V\$ASCL2_03	1.2.2
V\$E2A_Q2	1.2.2
V\$AR_01	2.1.1
V\$ESR1_01	2.1.1
V\$NKX61_03	3.1.2
V\$NKX62_Q2	3.1.2
V\$CACBINDINGPROTEIN_Q6	4.2.1
V\$E2A_Q2	1.2.1
V\$EBF1_01	6.1.5
V\$ER71_02	3.5.2
V\$ETF_Q6	3.6.1
V\$FOX_Q2	3.3.1
V\$LBP1_Q6	6.7.2
V\$MYCMAX_B	1.2.6
V\$PAX4_05	3.2.1
V\$SMAD4_04	7.1.1
V\$TAL1_Q6_01	1.2.3
V\$ZFP740_03	2.3.2
V\$CHCH_01	unknown
V\$MINI19_B	unknown
V\$NFE4_Q5	unknown

Please visit the TFClass classifications web page (<http://tfclass.bioinf.med.uni-goettingen.de/tfclass>) to retrieve the corresponding transcription factor family names for the classification schemes (e.g. *Jun-related factors* for '1.1.1').

3. **Supplementary Table 3:** GO Term Enrichment analysis for significant Reactome pathways via the geneXplain platform (<http://www.genexplain.com/>) based on the mapped transcription factors which belong the identified exclusive TFClass families (see Results section of the paper) for the breast cancer gene set.

ID	Title	Number of hits	Group size	Expected hits	Adjusted P-value	Hit names
1266741	REACT_111045: Developmental Biology	7	142	0.949544073	7.61750224758979E-5	"EBF1" "FOXH1" "KLF5" "NKX6-1" "NR5A2" "SMAD4" "TCF3"
744240	REACT_12627: Generic Transcription Pathway	4	47	0.314285714	5.73802055529988E-4	"MVC" "RUNX2" "SMAD4" "SP1"
2173790	REACT_121061: Transcriptional activity of SMAD2:SMAD3:SMAD4 heterotrimer	3	18	0.120364742	5.73802055529988E-4	"MVC" "SMAD4" "SP1"
1235025	REACT_111057: Signaling by NODAL	2	5	0.03343465	0.001005789	"FOXH1" "SMAD4"
1502539	REACT_150238: Signaling by Activin	2	6	0.040121581	0.001202516	"FOXH1" "SMAD4"
453316	REACT_6844: Signaling by TGF-beta Receptor Complex	3	30	0.200607903	0.001365352	"MVC" "SMAD4" "SP1"
633896	REACT_13698: Regulation of beta-cell development	2	22	0.147112462	0.012475661	"NKX6-1" "NR5A2"
500247	REACT_27161: Transcriptional Regulation of White Adipocyte Differentiation	2	26	0.173860182	0.01513523	"EBF1" "KLF5"
835437	REACT_71: Gene Expression	4	210	1.404253319	0.045738425	"MVC" "RUNX2" "SMAD4" "SP1"

4. Supplementary Table 4: GO Term Enrichment analysis for significant Reactome pathways via the geneXplain platform which belong the identified exclusive TFClass families for the NF- κ B gene set.

ID	Title	Number of hits	Group size	Expected hits	Adjusted P-value	Hit names
1234184	REACT_120815: Regulation of Hypoxia-inducible Factor (HIF) by Oxygen	3	9	0.076595745	2.53704850031564E-4	"EPAS1" "HIF1A" "HIF3A"
2262755	REACT_120956: Cellular responses to stress	3	9	0.076595745	2.53704850031564E-4	"EPAS1" "HIF1A" "HIF3A"
2262756	REACT_121311: Cellular response to hypoxia	3	9	0.076595745	2.53704850031564E-4	"EPAS1" "HIF1A" "HIF3A"
2173790	REACT_121081: Transcriptional activity of SMAD2/SMAD3/SMAD4 heterotrimer	3	18	0.153191489	0.001766379	"MYC" "SMAD4" "SP1"
1235025	REACT_111057: Signaling by NODAL	2	5	0.042553191	0.002520164	"FOXH1" "SMAD4"
1502539	REACT_150238: Signaling by Activin	2	6	0.05106383	0.003134874	"FOXH1" "SMAD4"
893504	REACT_24941: Circadian Clock	3	27	0.229787234	0.003457323	"ARNTL" "CLOCK" "PAS2"
453316	REACT_6844: Signaling by TGF-beta Receptor Complex	3	30	0.255319149	0.004135826	"MYC" "SMAD4" "SP1"
1286741	REACT_111045: Developmental Biology	5	142	1.208510338	0.00960636	"EBF1" "FOXH1" "KLF5" "KXK6-1" "SMAD4"
744240	REACT_12627: Generic Transcription Pathway	3	47	0.4	0.011942942	"MYC" "SMAD4" "SP1"
2032362	REACT_118859: Signaling by NOTCH1	2	16	0.136170213	0.011942942	"HIF1A" "MYC"
2823053	REACT_160214: Signaling by NOTCH1 in Cancer	2	16	0.136170213	0.011942942	"HIF1A" "MYC"
500247	REACT_27161: Transcriptional Regulation of White Adipocyte Differentiation	2	26	0.221276596	0.028382328	"EBF1" "KLF5"
419725	REACT_299: Signaling by NOTCH	2	27	0.229787234	0.028382328	"HIF1A" "MYC"

5. **Supplementary Figure 1:** A network for the transcription factor PBX1 based on the analysis of the breast cancer gene set. See figure caption for more information.

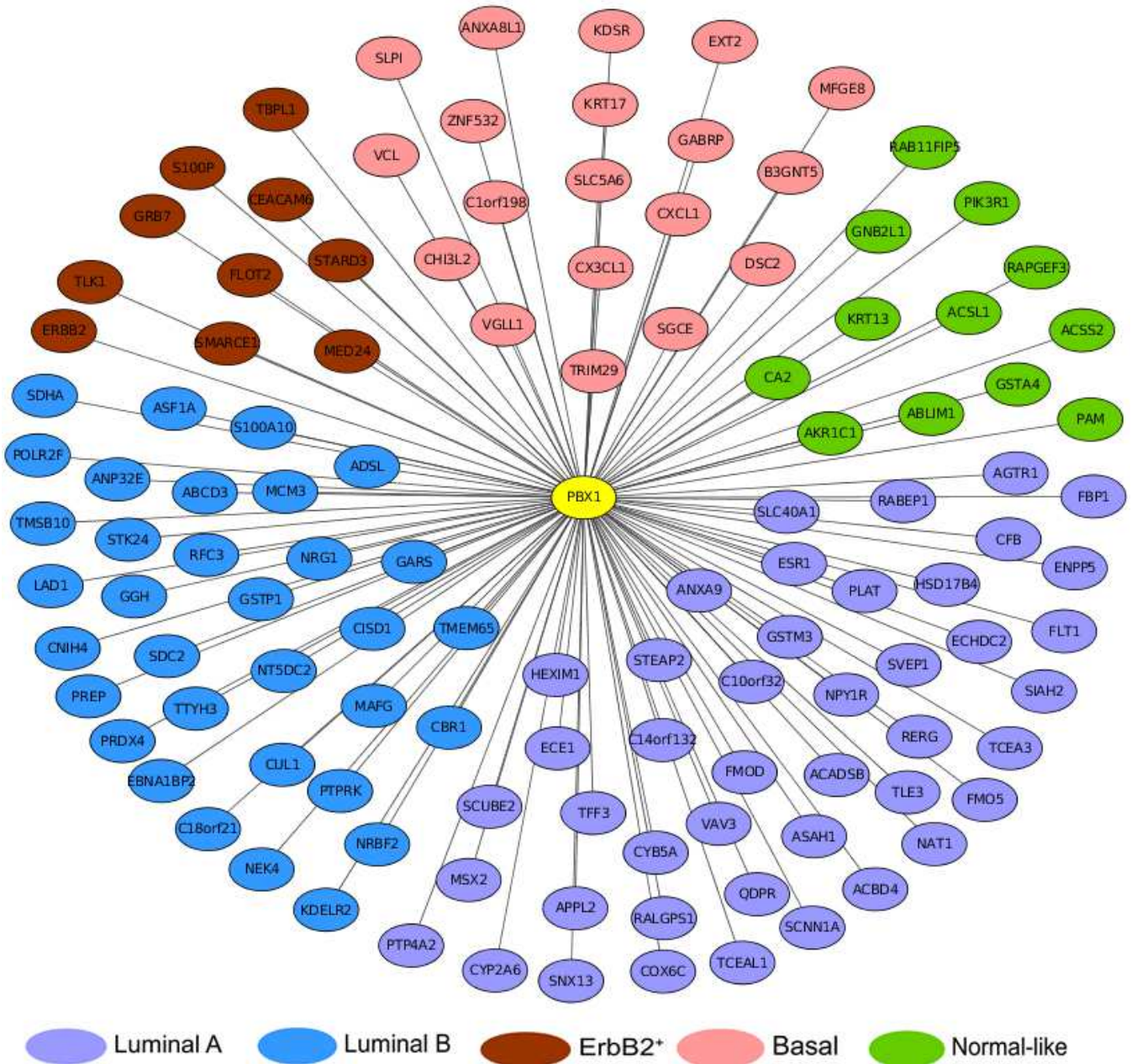


Figure 1: Predicted functional role of PBX1 based on TFBSs found in the promoter regions of target genes. The network shows predicted interactions of PBX1 with 118 genes for different molecular subtypes of breast cancer. 44 of them are associated with Luminal A, depicted as purple nodes. 33 out of 118 genes are associated with Luminal B, shown as blue nodes. 19 genes are related to Basal which are shown in the network with light red nodes. The remaining 12 and 10 genes in the network are associated with Normal-like and ErbB2+, where they are shown with green nodes and brown nodes, respectively.

References

1. Joshi, H., Nord, S., Frigessi, A., Borresen-Dale, A.-L., and Kristensen, V. (2012). Overrepresentation of transcription factor families in the genesets underlying breast cancer subtypes. *BMC Genomics*, 13(1), 199.
2. Ho Sui, S. J., Mortimer, J. R., Arenillas, D. J., Brumm, J., Walsh, C. J., Kennedy, B. P., and Wasserman, W. W. (2005). oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Research*, 33(10), 3154–3164.