Role of polycomb repressive complex 2 in regulation of human transcription factor gene expression

Jay C. Brown
Department of Microbiology, Immunology and Cancer Biology
University of Virginia School of Medicine
Charlottesville, Virginia, 22908

Corresponding author:
Dr. Jay C. Brown
Department of Microbiology, Immunology and Cancer Biology
University of Virginia School of Medicine Box 800734
Charlottesville, Virginia 22908
USA

Email: jcb2g@virginia.edu
Phone: 01-434-924-1814

ABSTRACT

Control of gene expression is now recognized as a central issue in the field of molecular biology. We now know the sequences of many genomes including that of the human genome, and we know the nature of many pathways involved in control of gene expression. It remains difficult, however, to look at the DNA sequences surrounding a particular gene and tell which methods of regulatory control are in use. I have been pursuing the idea that progress might be made by comparing the regulatory regions of paired gene populations in which one population is strongly expressed and the other weakly. Here I report the results obtained with human genes encoding transcription factors (TF). In this population, broadly expressed genes are strongly expressed while tissue targeted TF expression is suppressed in most tissues. The results demonstrated that the promoter region of broadly expressed TF genes is enriched in binding sites for POLR2A, a component of RNA polymerase II while promoters of tissue targeted genes are enriched in EZH2, a subunit of polycomb repressive complex 2 (PRC2). It was rare to observe promoters with binding sites for both POLR2A and EZH2. The findings are interpreted to indicate that strong expression of broadly expressed TF genes is due to the presence of RNA polymerase II at the promoter while weak expression of tissue targeted promoters results from the presence of PRC2. Finally, transcription factor families were compared in the proportion of broadly expressed and tissue targeted genes they contain. The results demonstrated that most families possess both broadly expressed and tissue targeted members. For instance, this was the case with 16 of 20 TF families examined. The results are interpreted to indicate that while individual TFs such as EZH2 may be specific for broadly expressed or tissue targeted genes, this is not a property of most TF families.

Keywords: tissue targeted gene; transcription factor, polycomb repressive complex 2, RNA-seq, RNA polymerase II

Highlights

1. Human transcription factor (TF) genes were noted to be broadly expressed or tissue targeted depending on their breadth of tissue expression.
2. The results of ChIP-seq experiments were used to characterize the promoter regions of broadly expressed and tissue targeted TF genes.
3. The results demonstrated that the promoters of broadly expressed TF genes are enriched in binding sites for RNA polymerase II (pol II) but lack sites for polycomb repressive complex 2 (PRC2) while tissue targeted promoters are the reverse. These have prominent binding sites for PRC2 but are depleted in pol II sites.
4. The results are interpreted to indicate that PRC2 complexes suppress TF expression in tissues where expression is not required.

1. Introduction

Studies on the control of gene expression have benefitted from recent advances in DNA sequencing and from identification of biochemical systems that result in up- or down-regulation of transcription. Whole genome sequences are now available for a wide variety of organisms including humans, and well-characterized elements of regulatory control include promoters, epigenetic signaling, CpG islands, structured chromosome domains, mRNA splicing and many others [1-4]. In view of the variety of regulatory pathways found in higher vertebrates, it is unreasonable to assume that there is an overall regulatory system that applies to all genes. More likely is the idea that distinct regulatory regimes apply to distinct gene populations with the effects of individual regulatory regimes integrated to create the overall developmental program observed.

Considering the environment of regulatory studies described above, I have adopted the strategy of comparing the regulatory elements of two populations of related genes that differ in their level of expression. The idea is that such a comparison might lead to information about the identity of the regulatory features involved.

Here I describe the results of a study focused on the genes encoding human transcription factors (TF). Beginning with a database of nearly all TF genes, the genes were divided into two pools depending on whether they are broadly expressed, or tissue targeted. This distinction is relevant to the present study because nearly all broadly expressed TF genes are transcribed while most tissue targeted TF genes are not. Unexpressed tissue targeted TF genes are those present in tissues not targeted by the

TF gene (e.g. brain specific TF genes in the liver). The study described here compares the transcription factor binding sites present in the promoter region of broadly expressed and tissue targeted human TF genes.

The results are interpreted to support the view that polycomb repressive complex 2 (PRC2) plays a central role in regulating the expression of human TF genes. PRC are multi-subunit molecular complexes able to bind the promoter region of a target gene and suppress the gene's expression [5-6]. First discovered in Drosophila in 1947 [7], PRC functions have been actively studied ever since. PRC or components of PRC are found in all metazoans and in some more primitive species including fungi. Two forms of PRC complex are known, PRC1 and PRC2, and both can introduce covalent modifications into the histone components of chromatin. PRC1 ubiquitinates H2 at lysine119 [8] while PRC2 methylates H3 at lysine27 [9-12]. PRC1 and PRC2 are thought to function together to repress gene expression and/or to report on the expression due to other control factors.

2. Materials and methods

2.1 Gene databases

2.1.1 All human transcription factor genes (Supplementary Table 1; 1020 genes)

This database was derived from the list of Vaquerizas et al. [13] with a modification involving zinc finger transcription factor genes. ZNF genes were examined individually using GeneCards and NCBI Gene. A ZNF gene was retained in the database only if its role as a transcription factor was supported by an experimental study.

131

132 2.1.2 Sample of all human genes (SupplementaryTable 2, 183 genes)

133 A sample of all human protein coding genes was accumulated to serve as a control for
134 analysis of transcription factor genes. Included genes began with ACO1 and proceeded
135 rightward on chromosome 9 for a total of 183 genes.

136

137 2.1.3 Sample of testis-specific genes (Supplementary Table 3, 204 genes)

138 This database was curated by Brown [14].

139

140 2.1.4 Sample of all human, broadly expressed genes (Supplementary Table 4, 1125
141 genes)

142 This database contains genes that are broadly expressed (i.e. “housekeeping” genes) in
143 both human and mouse genomes as curated by Hounkpe et al. [15].

144

145 2.1.5 All database transcription factor genes also present in the broadly expressed gene
146 database (Supplementary Table 5, 36 genes)

147 This list contains the names of the genes shown in aggregated form in the Results
148 section.

149

150 2.1.6. All database transcription factor genes also present in the tissue targeted gene
151 database (Supplementary Table 6, 112 genes)

152 This list contains the names of the genes shown in aggregated form in the Results
153 section.

2.1.7 Database of all human transcription factor genes showing predominant structural features present in the TF protein (Supplementary Table 7, 1020 genes)
Structural element annotations were obtained from GeneCards and NCBI Gene.

2.1.8 Database of 2413 tissue targeted human genes. This database was used unaltered from Brown [16].

2.2 RNA-seq and ChIP-seq results for transcription factor and control genes
Transcription levels for all genes reported here were obtained from the NIH GTEx RNA-seq results as reported in the UCSC Genome Browser (version hg38 (https://genome.ucsc.edu/)). A gene was annotated as “broadly expressed” (B) if it was found to be expressed in 90% or more of the 52 tissues reported by GTEx. Otherwise, the gene was annotated as “tissue targeted” (T). Borderline genes were rare. They are indicated by B/T or T/B depending on the category they most resemble.

Transcription factor binding sites in the promoter region of transcription factor genes were derived from ChIP-seq studies reported in the ENCODE project (3 November 2018 version by way of the UCSC Genome Browser). Only binding sites in the promoter region are reported here; sites in non-promoter regions were ignored. Special attention was paid to binding sites for POLR2A and for EZH2 as most transcription factor promoters were found to be positive for one or the other but not both (see Results section). A promoter was annotated to be positive for POLR2A if a POLR2A site was

identified in 50% or more of the experimental studies reported by ENCODE. The same criterion was employed for EZH2 sites.

2.3 Transcription factor families and protein structural features

These were derived from GeneCards (https://www.genecards.org) and from NCBI Gene (https://www.ncbi.nlm.nih.gov).

2.4 Data handling

Data were manipulated with Excel and rendered graphically with SigmaPlot v14.5 or Adobe Illustrator.

3. Results

3.1 Broadly expressed and tissue targeted human TF genes

The project was carried out beginning with a database of 1020 human transcription factor genes (Supplementary Table 1). The database was that of Vaquerizas et al. [13] with the modifications described in Materials and Methods. Each gene was classified as either broadly expressed or tissue targeted using information about its tissue expression in the RNA-seq results reported by GTEx via the UCSC Genome Browser (https://genome.ucsc.edu/). Of the 1020 database genes, 588 (58%) were classified as broadly expressed and 423 (41%) as tissue targeted (Fig. 1). Nine genes could not be classified. For comparison with the TF gene classification, the same classification was performed with 183 unselected human protein coding genes (Supplementary Table 2). Here the results showed 77% broadly expressed genes and 23% tissue targeted (Fig.

1) demonstrating that the TF genes have a higher proportion of tissue targeted than the unselected gene population (41% vs 23%). This result is interpreted to reflect the fact that as more transcription factors affect specific tissues compared to an unselected gene population, more tissue targeted transcription factors might be expected.

3.2 TF binding sites in the promoter region of TF genes

TF factor binding sites were compared in broadly expressed and tissue targeted TF genes since it is expected that the two populations differ in their expression. Broadly expressed TF genes are expected to be expressed in all tissues while tissue targeted TF genes are expressed in their targeted tissue(s) but suppressed in all others. It was hoped that comparing TF binding sites in the two populations might yield information about the cause of the distinct expression levels.

TF binding sites were identified using ChIP-seq results available from ENCODE. Identified sites in graphical form were examined visually in the UCSC Genome Browser. Images revealed that binding sites for POLR2A and EZH2 were asymmetrically distributed in broadly expressed compared to tissue targeted TF gene promoters. For instance, among 594 broadly expressed TF genes, 538 (91%) had one or more POLR2A binding sites in the promoter while among 426 tissue targeted TF genes, 303 (71%) had binding sites for EZH2 (Fig. 2a). It was rare to find EZH2 binding sites in broadly expressed gene promoters (8%) or POLR2A in tissue targeted genes (19%; see Fig. 2a). It was also rare to find TF gene promoters that lacked both POLR2A and

EZH2. Of the 1020 genes examined, only 53 (5%) lacked binding sites for both (Fig. 2a).

3.3 Control: POLR2A and EZH2 in the promoters of unselected human protein coding genes

As a control, an experiment was performed to test whether the asymmetry observed between POLR2A and EZH2 in the promoter region of TF genes would also be observed in an unselected population of human protein coding genes. Is asymmetry observed only in TF genes, or does it exist more widely? To test the idea, the experiment described above was repeated beginning with an unselected population of human protein coding genes rather than with TF genes only (Supplementary Table 2). The results showed that among 141 unselected, broadly expressed genes, 128 (91%) had POLR2A in the promoter while 8 (6%) had EZH2, a result comparable to the TF gene population. In contrast, in the tissue targeted population the results were 45% and 47% for POLR2A and EZH2, respectively, a result quite different from the TF genes (Fig. 2b). The results are interpreted to support the view that the POLR2A/EZH2 asymmetry is found in TF genes only and is not found in an unselected gene population.

3.4 Control: POLR2A and EZH2 in testis-specific genes

A further control study was carried out to examine POLR2A and EZH2 binding in the promoters of a population of tissue targeted genes, a population of 204 testis-specific genes (Supplementary Table 3). ChIP-seq results, as described above, were examined in the promoter regions of the testis-specific genes. It was expected that EZH2 binding

sites would be observed in a high proportion of genes if testis-specific genes were regulated in the same way found for TF. Quite a different outcome was detected (Fig. 2c).

Promoters of testis-specific genes were not found to be enriched in either POLR2A or EZH2. Instead, the promoters had a mixture of transcription factor binding sites in which no predominant species could be identified (see Supplementary Table 3). For instance, of 195 testis-specific genes that yielded a result, 135 (69%) had neither a POLR2A nor an EZH2 site. POLR2A and EZH2 were found in 22% and 9%, respectively (Fig. 2c). The result is interpreted to indicate that while EZH2 binding sites are prevalent in the promoters of tissue targeted TF genes, the same is not found in other populations of tissue targeted genes.

3.5 Control: begin with tissue targeted rather than TF genes

A second control was done to test the prevalence of EZH2 selectively in the promoter of tissue targeted TF genes. Instead of beginning with a database of transcription factor genes, I started with a list of tissue targeted genes (Supplementary Table 3). Genes on the tissue targeted list were then compared with those on the TF list, with the expectation that genes on both lists would be TF genes with EZH2 in the promoter region. The results showed that of 112 tissue targeted TF genes, 76 (68%) had one or more EZH2 binding sites in the promoter while 36 (32%) did not (Fig. 3; Supplementary Table 6). Conversely, when I started with a database of broadly expressed human genes (Supplementary Table 4) and compared that with the TF genes, the expected result was also obtained. Of 35 broadly expressed TF genes, all were found to have

POLR2A in the promoter (Fig. 3; Supplementary Table 5). The results are interpreted to support the view that most tissue targeted TF genes have EZH2 binding sites in the promoter while broadly expressed TF genes have POLR2A.

3.6 Transcription factor families

The information above indicating that promoter binding sites for EZH2 are located preferentially in tissue targeted TF genes raises the question of whether there may be families of TFs whose members have a selectivity for controlling tissue targeted TF genes only, or broadly expressed TF genes only. Could binding to tissue targeted promoters be a property of a TF family and not only of an individual TF? To explore the above question, I have identified TF families in the database employed here (Supplementary Table 1) and characterized them according to the number of broadly expressed compared to tissue targeted family genes. All results are shown in Supplementary Table 7 with a sample of 20 families shown graphically in Fig. 4.

Of the 20 families the greatest number (16) contain both broadly expressed and tissue targeted members. The remaining four families have members with only broadly expressed (ATF and CREB) and only tissue targeted (NKX and PAX) genes. Members of the largest family (ZNF; 109 genes) are predominantly broadly expressed genes (101 broadly expressed vs. 8 tissue targeted genes). Forkhead (FOX) and homeobox (HOX) families have the highest number of tissue targeted genes (31 and 37, respectively; see Fig. 4). The results support the conclusion that most human TF families contain both broadly expressed and tissue targeted member genes.

A similar study was carried out with TF genes grouped according to a prominent protein structural feature. Each TF database gene was assigned to a structural group if one could be identified, and the results were then summed for each group (Supplementary Table 7). The outcome for 8 prominent structural features is shown in Fig. 5. Overall, the results resemble those obtained with TF families. All 8 structure groups were found to have members that bind the promoters of broadly expressed TF genes and others that bind the promoters of tissue targeted TF genes (Fig. 5). No TF structural feature could be associated with broadly expressed or tissue targeted expression only.

4. Discussion

4.1 Interpretation of POLR2A and EZH2 binding sites

The central observation reported here has to do with asymmetry observed in the promoters of broadly expressed compared to tissue targeted human TF genes. Binding sites for POLR2A were found in the promoters of broadly expressed TF genes while EZH2 sites were found in tissue targeted ones. Little overlap was observed between the two types of promoters. Most broadly expressed TF gene promoters had prominent POLR2A binding sites, but few had sites for EZH2. In contrast, most tissue targeted TF gene promoters had binding sites for EZH2, but few had comparable sites for POLR2A (see Fig. 2a). The reported distribution appears to be selective for promoters of TF genes. No comparable asymmetry was observed in the promoters of unselected human protein coding genes nor in the promoters of human genes targeted specifically for expression in testis (Figs. 2b and c).

As POLR2A (Rpb1) is a prominent component of RNA polymerase II [17, 18], the results described above are interpreted to indicate that RNA polymerase II is bound at the promoter region of broadly expressed TF genes. Similarly, since EZH2 is a prominent core component of polycomb repressive complex 2 (PRC2) the results indicate that PRC2 is bound at the promoter region of tissue targeted TF genes. The functions of RNA polymerase II (pol II) and PRC2 are consistent with the expression properties of the gene populations they are proposed to control. As pol II is expected to be permissive for transcription of all broadly expressed TF genes [5, 6], it is reasonable to find pol II binding sites in the promoters. As PRC2 is associated with repression of gene expression, it is reasonable to find PRC2 binding sites in the promoters of a repressed gene population. The above interpretation is shown graphically in Fig. 6.

4.2 Comparison to the role of PRC2 in human X-chromosome inactivation

PRC2 is well known for its role in suppressing all gene expression from one X chromosome in human females. Targeted to one of the two X chromosomes by an untranslated RNA (Xist), PRC complexes bind along the entire chromosome to silence the genes permanently, creating the Barr body [19-21]. I suggest that permanent gene repression by PRC2 complexes may be relevant to their role in suppression of tissue targeted TF in the non-targeted tissues. Permanent rather than temporary repression of non-targeted TF may well be the rule for most tissues [22].

PRC2 deposition to create the Barr body differs in one significant way compared to deposition at promoter regions as described here. Whereas PRC complexes coat an entire X chromosome to form a Barr body, deposition sites are widely distributed in the genome to suppress tissue targeted TF gene expression. Most tissue targeted TF genes are expected to be separate sites of PRC2 deposition. This situation creates a more formidable problem for targeting complexes at TF genes than it does with an entire X chromosome, and the problem will need to be addressed in the future.

4.3 Experimental strategy

The experimental strategy employed here suggests itself as a way to make further progress in understanding the control of gene expression. The idea is to identify two populations of related genes that differ in their level of expression. The results of ChIP-seq experiments as described here can then be used to compare TF binding sites in the promoter regions of genes in the two populations with the goal of identifying sites that affect regulation. Analysis could begin with any gene having tissue specific expression as each such gene would have activating elements driving specific expression and also suppressing elements for non-targeted tissues.

4.4 Families of human TF

Evidence that a TF can act specifically on a tissue targeted or a broadly expressed gene as described here raises the possibility that such specificity might apply to TF groups as well as to individual TFs. Could all members of a TF family, for instance, be specialized to target only broadly expressed genes? Studies to address this issue

generally supported the view that most TF groups contain members able to recognize broadly expressed genes and others that recognize tissue targeted ones. Little evidence was observed for broadly expressed/tissue targeted specificity among either TF families or among TFs with the same protein structural motifs (Figs 4 and 5; Supplementary Table 7). Despite the lack of overall specificity observed, some examples of selectivity were noted. For instance, zinc finger TFs exhibited a strong preference for broadly expressed TF genes while homeodomain TFs showed a preference for tissue targeted genes (Fig. 4).

I suggest that future studies of gene regulatory control might benefit from a focus on TF groups such as NKX and PAX in which all members have tissue targeted expression. It is expected that the promoter regions of these genes will have binding sites for both activating elements able to activate genes in the targeted tissue and repressive elements for non-targeted tissues. Examination of the promoters in such genes might facilitate the identification of the features that influence gene expression.

Acknowledgements

I gratefully acknowledge Karsten Siller for advice on computational aspects of this project.

References

1. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;172(4):650-65.

https://doi.org/10.1016/j.cell.2018.01.029

2. Schubeler D. Function and information content of DNA methylation. Nature. 2015;517(7534):321-6.

https://doi.org/10.1038/nature14192

3. Szalaj P, Plewczynski D. Three-dimensional organization and dynamics of the genome. Cell Biol Toxicol. 2018;34(5):381-404.

https://doi.org/10.1007/s10565-018-9428-y

4. Cavalli G, Heard E. Advances in epigenetics link genetics to the environment and disease. Nature. 2019;571(7766):489-99.

https://doi.org/10.1038/s41586-019-1411-0

5. Blackledge NP, Rose NR, Klose RJ. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nat Rev Mol Cell Biol. 2015;16(11):643-9.

https://doi.org/10.1038/nrm4067

6. Schuettengruber B, Bourbon HM, Di Croce L, Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell. 2017;171(1):34-57.

https://doi.org/10.1016/j.cell.2017.08.002

7. Lewis EB. A gene complex controlling segmentation in Drosophila. Nature. 1978;276(5688):565-70.

https://doi.org/10.1038/276565a0

8. Cao R, Tsukada Y, Zhang Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol Cell. 2005;20(6):845-54.

https://doi.org/10.1016/j.molcel.2005.12.002

9. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science. 2002;298(5595):1039-43.
DOI: 10.1126/science.1076997

10. Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell. 2002;111(2):185-96.
https://doi.org/10.1016/S0092-8674(02)00975-3

11. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 2002;16(22):2893-905.
Doi: 10.1101/gad.1035902

12. Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell. 2002;111(2):197-208.
https://doi.org/10.1016/S0092-8674(02)00976-5

13. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009;10(4):252-63.
https://doi.org/10.1038/nrg2538

14. Brown JC. Control of human testis-specific gene expression. PLoS One. 2019;14(9):e0215184.
https://doi.org/10.1371/journal.pone.0215184

15. Hounkpe BW, Chenou F, de Lima F, De Paula EV. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Res. 2021;49(D1):D947-D55. https://doi.org/10.1093/nar/gkaa609

16. Brown JC. Role of Gene Length in Control of Human Gene Expression: Chromosome-Specific and Tissue-Specific Effects. Int J Genomics. 2021;2021:8902428. https://doi.org/10.1155/2021/8902428

17. Coulombe B, Langelier MF. Functional dissection of the catalytic mechanism of mammalian RNA polymerase II. Biochem Cell Biol. 2005;83(4):497-504. https://doi.org/10.1139/o05-061

18. Woychik NA, Young RA. RNA polymerase II: subunit structure and function. Trends Biochem Sci. 1990;15(9):347-51. https://doi.org/10.1016/0968-0004(90)90074-L

19. Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T, Wutz A. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. PLoS Biol. 2004;2(7):E171. https://doi.org/10.1371/journal.pbio.0020171

20. Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, et al. Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell. 2003;4(4):481-95. https://doi.org/10.1016/S1534-5807(03)00068-6

21. Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, et al. Role of histone H3 lysine 27 methylation in X inactivation. Science. 2003;300(5616):131-5.
Doi: 10.1126/science.1084274

22. Okamoto I, Otte AP, Allis CD, Reinberg D, Heard E. Epigenetic dynamics of imprinted X inactivation during early mouse development. Science. 2004;303(5658):644-9.
DOI: 10.1126/science.1092727

Figure Legends

Figure 1. Broadly expressed and tissue targeted genes among: database human transcription factor genes (right); and a sample of all human protein coding genes (left). Note that the proportion of tissue targeted genes is higher in the transcription factor population.

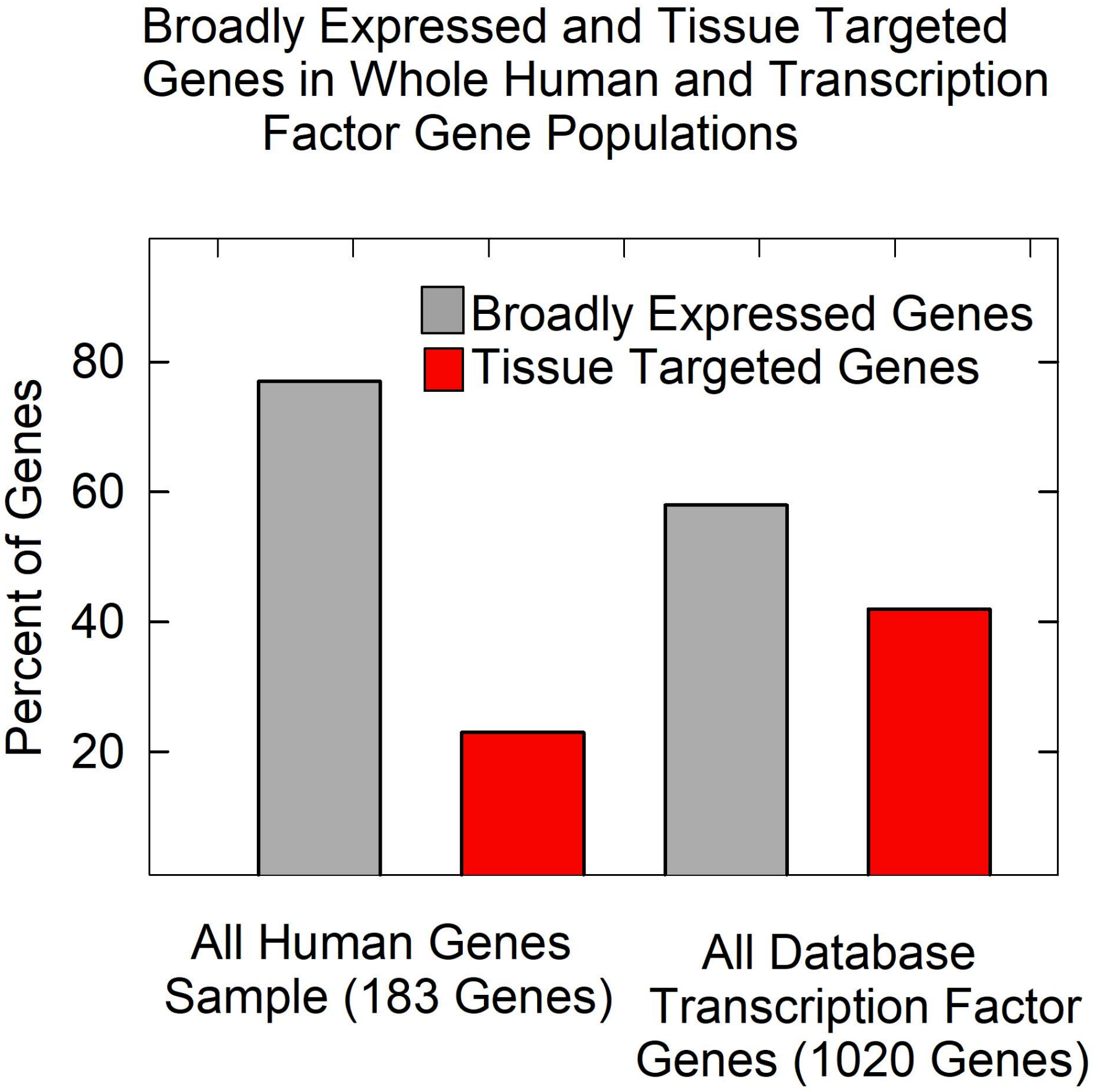
Figure 2. POLR2A and EZH2 binding sites in the promoter regions of broadly expressed and tissue targeted genes. Results are shown for (a) all database human transcription factor genes; (b) an unselected population of human protein coding genes; and (c) a sample of human testis-specific genes. Note that EZH2 binding sites (indicative of PRC2 binding) are prominent in tissue targeted TF genes, but not in broadly expressed TF genes (a), a mixed population of human protein coding genes (b) or in testis-specific human genes (c).

Figure 3. Results of control experiment to examine the presence of POLR2A binding sites in the promoter region of TF found in a database of broadly expressed genes (left). On the right is shown the results of a similar control in which the presence of EZH2 binding sites were noted in the promoter region of tissue targeted TF genes found in a database of tissue targeted genes. Note that POLR2A binding sites (indicative of RNA polymerase II binding) were observed selectively in broadly expressed TF genes while EZH2 sites (indicative of PRC2 binding) were predominant in tissue targeted TF genes.

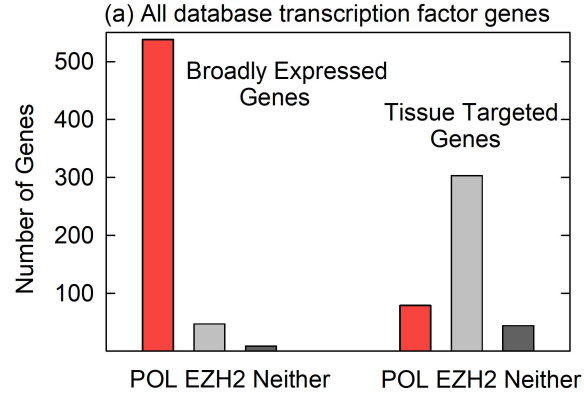
Figure 4. Number of broadly expressed and tissue targeted TFs in selected TF families. See Supplementary Table 7 for all results. Note that most families contain both broadly expressed and tissue targeted TF genes. Families shown: ATF, activating TF; CREB, cAMP response element binding TF; E2F, E2 factors; ETV, FGF-regulated TFs; FOX, forkhead box; HOX, homeodomain; IRF, interferon-regulatory factor; KLF, Krüppel-like TF's; LHX, LIM homeobox; NKX, NKX homeodomain box; NR, nuclear receptor TF's; PAX, paired homeobox; POU, Pit-Oct-Unc TF's; SP, SRY-related HMG box; SP, Sp; TBX, T-box; TCF, TCF/LEF; ZBTB, ZBTB family; ZNF, zinc finger TF's; ZSCAN, zinc finger and scan domain TF's.

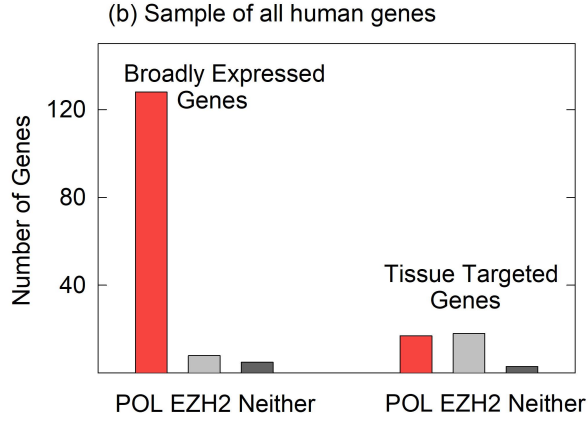
Figure 5. Number of broadly expressed and tissue targeted TFs in selected groups sharing the indicated protein structural feature(s). See Supplementary Table 7 for all results. Note that all structure groups contain both broadly expressed and tissue targeted TF genes. Structural groups shown: bHLH, basic helix-loop-helix; bZIP, basic leucine zipper domain; ETS, erythroblast transformation specific domain; Forkhead,

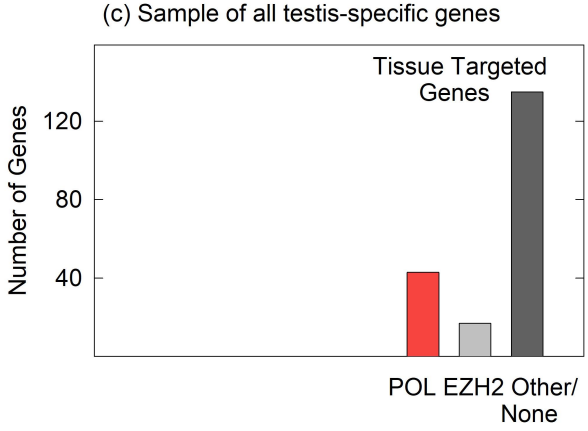
497 forkhead box; Homeobox, homeobox sequence; NR, multiple structural domains; ZNF,
498 zinc finger domain; ZNF-BTB, zinc finger plus bric-à-brac motifs.
499
500 **Figure 6.** Graphical representation of the proposed regulation of human transcription
501 factor gene expression. PRC2 is suggested to repress TF expression in tissue targeted
502 TF's selectively in all non-targeted tissues.
503


Broadly Expressed and Tissue Targeted Genes in Whole Human and Transcription Factor Gene Populations
Broadly Expressed Genes
Tissue Targeted Genes
Percent of Genes
80
60
40
20
All Human Genes Sample (183 Genes)
All Database Transcription Factor Genes (1020 Genes)

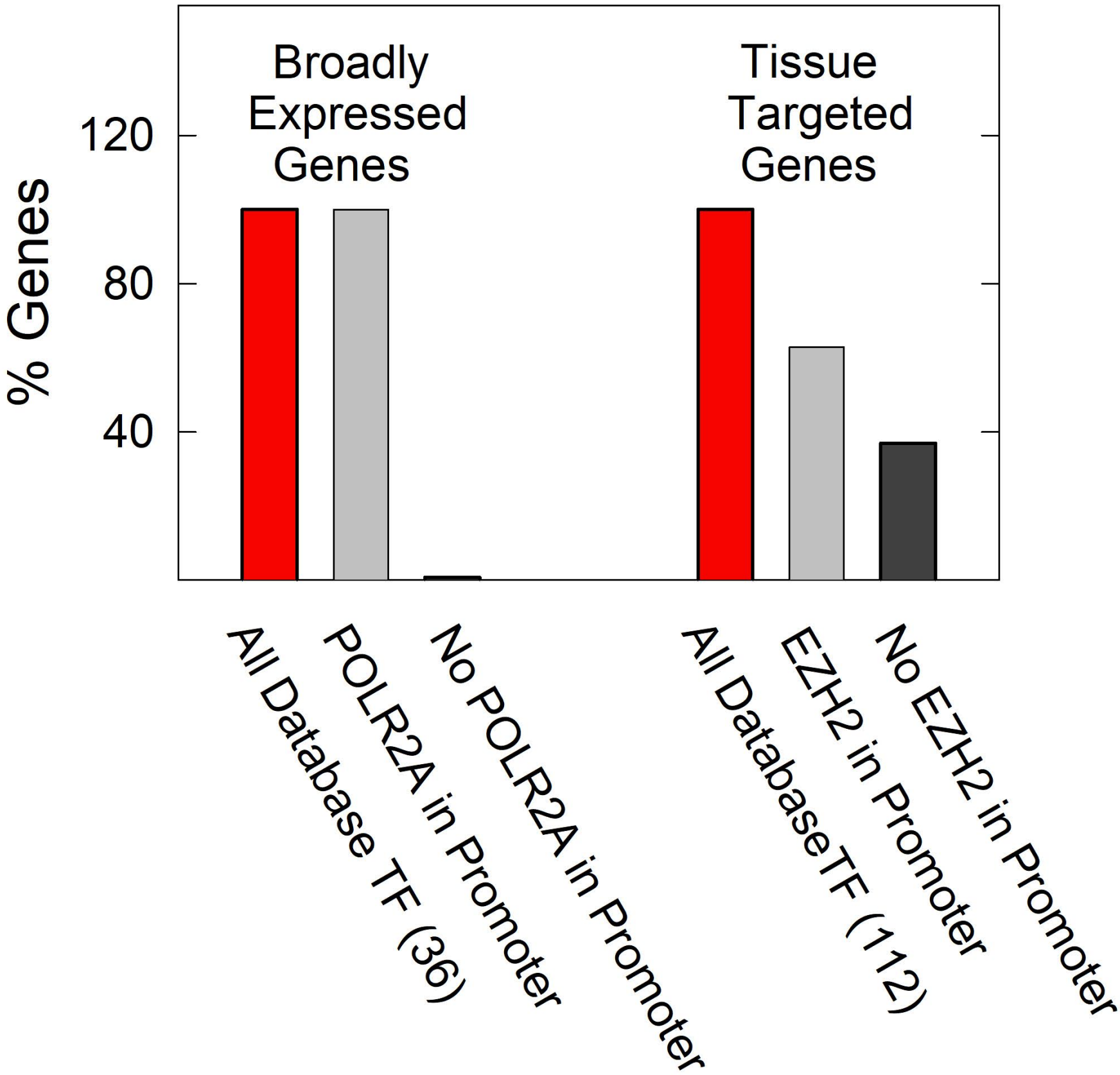
High Proportion of EZH2 Binding Sites in Promoters of Tissue Targeted Human Transcription Factor Genes

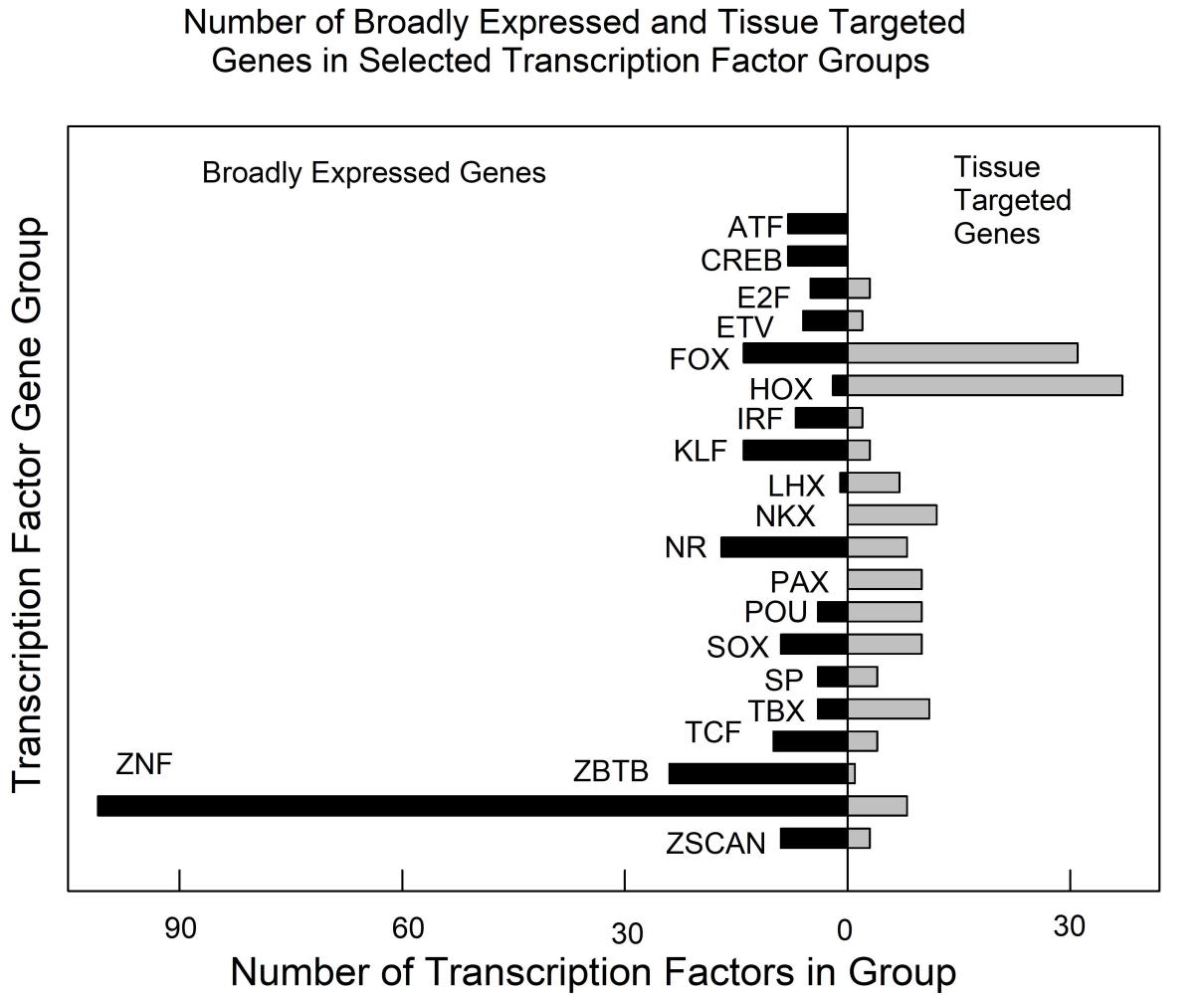


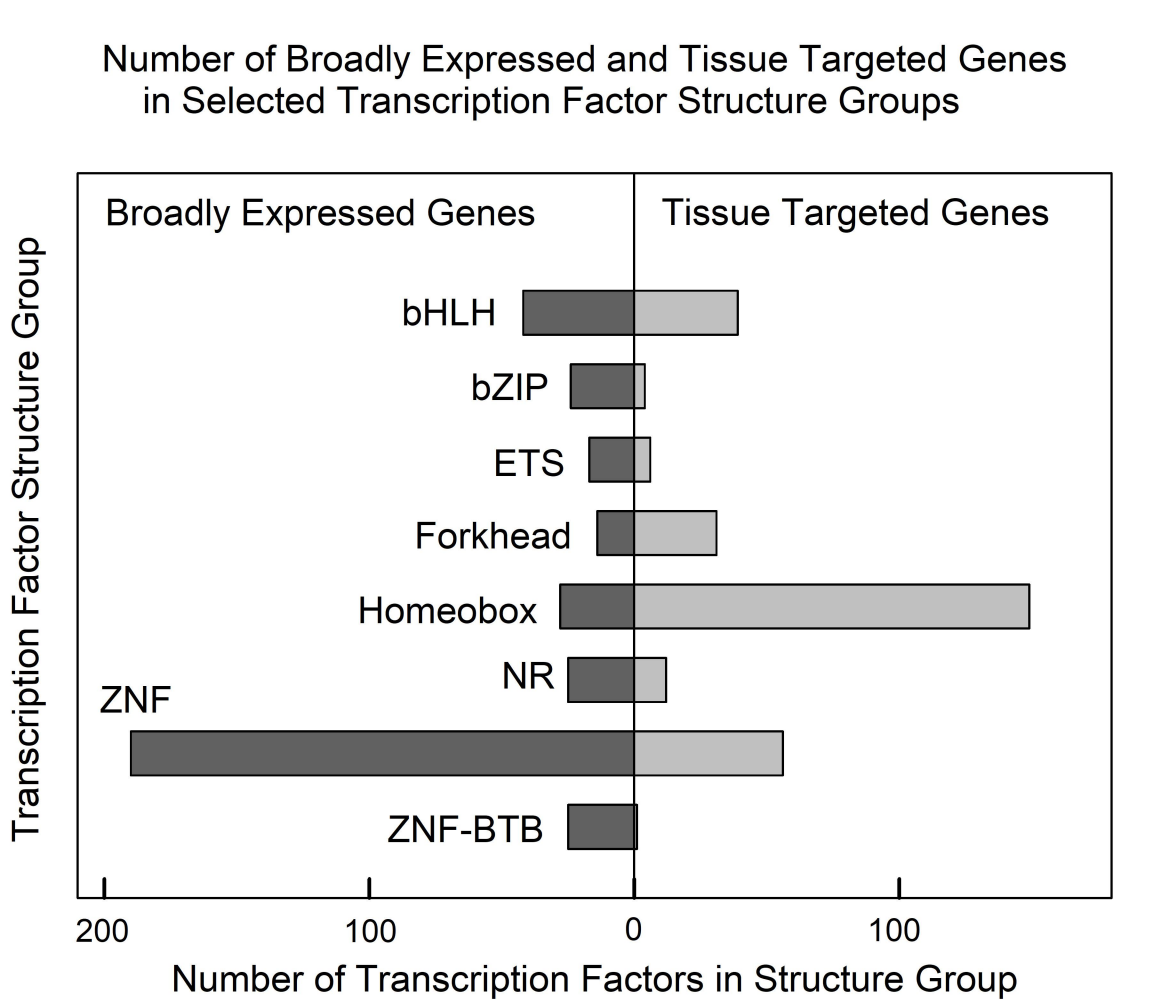


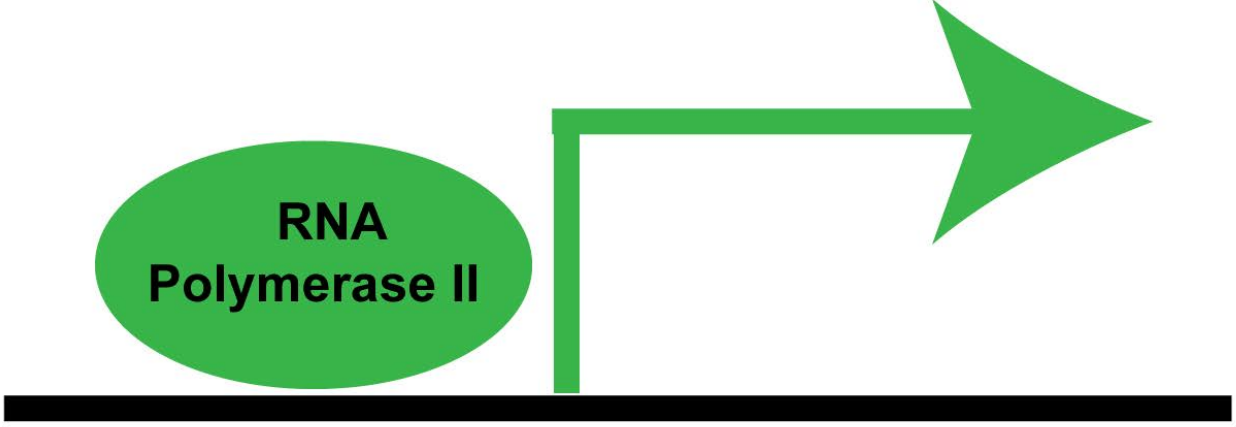


Control: POLR2A in TF from Broadly Expressed Gene Database; EZH2 in TF Tissue Targeted Gene Database



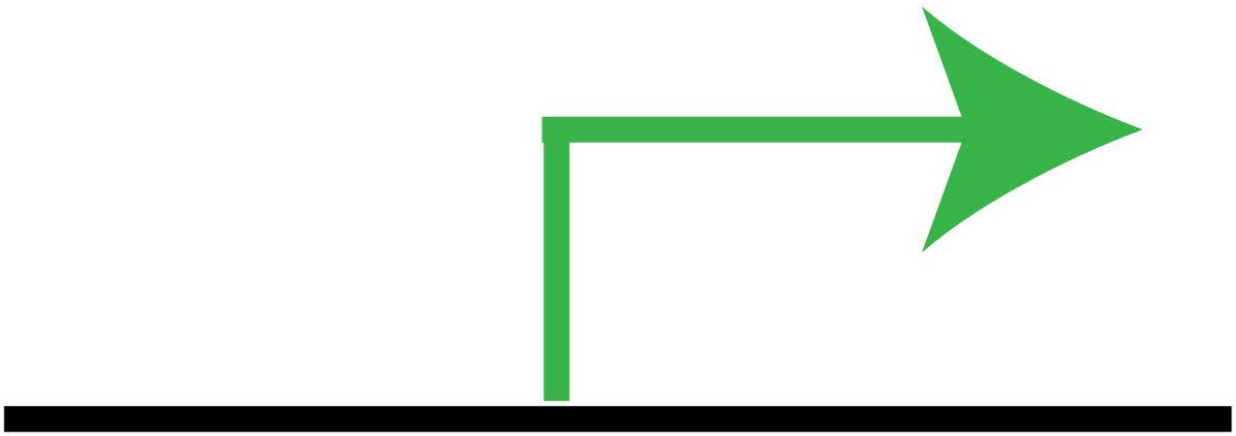

Number of Broadly Expressed and Tissue Targeted Genes in Selected Transcription Factor Groups
Broadly Expressed Genes
Tissue Targeted Genes
Transcription Factor Gene Group
ATF
CREB
E2F
ETV
FOX
HOX
IRF
KLF
LHX
NKX
NR
PAX
POU
SOX
SP
TBX
TCF
ZNF
ZBTB
ZSCAN
90
60
30
0
30
Number of Transcription Factors in Group


Number of Broadly Expressed and Tissue Targeted Genes
in Selected Transcription Factor Structure Groups
Broadly Expressed Genes
Tissue Targeted Genes
bHLH
bZIP
ETS
Forkhead
Homeobox
NR
ZNF
ZNF-BTB
Transcription Factor Structure Group
200
100
0
100
Number of Transcription Factors in Structure Group



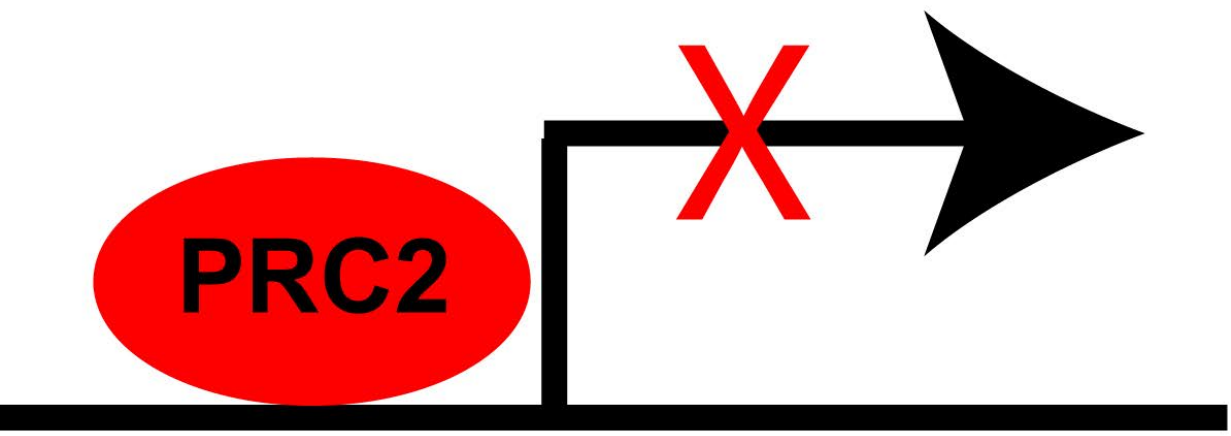
Broadly expressed transcription factor gene

All tissues



Tissue targeted transcription factor gene

Targeted tissue



Tissue targeted transcription factor gene

All untargeted tissues