

Supplementary Information For  
Regulatory Modules in Dunn Model  
(CisView Database)

March 24, 2018

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>List of Regulatory Modules (CRM) Found With CisView Database</b> | <b>3</b>  |
| <b>2</b> | <b>Transcription Factor Binding Sites (TFBSs)</b>                   | <b>4</b>  |
| <b>3</b> | <b>Interactions in CRN</b>  | <b>14</b> |
| <b>4</b> | <b>Results</b>  | <b>15</b> |
| 4.1      | Statistics . . . . .  | 15        |
| 4.2      | Figures . . . . .   | 15        |

# 1 List of Regulatory Modules (CRM) Found With CisView Database

All sequence positions refer to the Mouse Genome (mm9) of UCSC[2]. Only "high-quality" CRMs, as assessed by the database[3], present in a range of 2kb upstream or 250bp downstream of the gene Transcription Start Site, are listed.

| Gene   | Position                                   | CisView Region Annotations | CRM names & positions   |
|--------|--|----------------------------|---|
| Tcf3   | (U032010)<br>chr10:79871857-79896393       | R00045128<br>&R00045129    | (CM10015455)chr10:79,883,498-79,883,607<br>(CM10015456)chr10:79,884,495-79,884,795<br>(CM10015460)chr10:79,895,598-79,897,177 |
| Mapk1  | (U016859)<br>chr16:16,983,474-17,047,546   | R00041933                  | (CM16003054)chr16:16,983,139-16,983,719   |
| Stat3  | (U033438)<br>chr11:100,746,411-100,800,825 | R00010634                  | (CM11023938)chr11:100,799,556-100,799,935<br>(CM11023939)chr11:100,800,029-100,801,438  |
| Pou5f1 | (U017906)<br>chr17:35,642,981-35,647,722   | R00007223                  | (CM17006875)chr17:35,640,848-35,641,062<br>(CM17006876)chr17:35,641,859-35,642,269<br>(CM17006877)chr17:35,642,623-35,643,056 |
| Klf4   | (U025080)<br>chr4:55,540,014-55,545,347    | R00019128                  | (CM04010396)chr4:55,545,349-55,545,622  |
| Sall4  | (U023697)<br>chr2:168,573,831-168,592,701  | R00003873                  | (CM02038711)chr2:168,590,642-168,592,411  |

Table 1: List of CRMs associated with each gene present in the model (1)

| Gene    | Position                                  | CisView Region Annotations | CRM names & positions  |
|---------|---|----------------------------|--|
| Sox2    | (U003145)<br>chr3:34,548,916-34,551,383   | R00026811                  | (CM03005876)chr3:34,547,223-34,547,800<br>(CM03005877)chr3:34,547,917-34,549,345   |
| Tbx3    | (U006096)<br>chr5:120,120,677-120,134,613 | R00038510                  | (CM05023198):120,120,481-120,121,715   |
| Klf2    | (U105663)<br>chr8:74,842,931-74,845,553   | R00016555                  | (CM08013093)chr8:74,842,442-74,843,013   |
| Nanog   | (U007377)<br>chr6:122,657,506-122,664,651 | R00031035                  | (CM06025863)chr6:122,657,411-122,657,711   |
| Esrrb   | (U014218)<br>chr12:87,702,066-87,862,575  | R00000592<br>&R00000593    | (CM12018771)chr12:87,701,856-87,701,941<br>(CM12018772)chr12:87,702,191-87,702,892<br>(CM12018795)chr12:87,761,297-87,761,804<br>(CM12018796)chr12:87,762,393-87,763,111 |
| Tfcp2l1 | (U000714)<br>chr1:120,524,521-120,581,745 | R00023187                  | (CM01022019)chr1:120,524,222-120,524,617   |
| Gbx2    | (U021580)<br>chr1:91,824,530-91,829,865   | R00013675                  | (CM01018070)chr1:91,828,352-91,828,690<br>(CM01018071)chr1:91,828,976-91,829,144<br>(CM01018072)chr1:91,829,928-91,830,423   |

Table 2: List of CRMs associated with each gene present in the model (2)

## 2 Transcription Factor Binding Sites (TFBSs)

Only TFBSs which had a conservation score  $\geq 0.5$ , and a mismatch score  $\leq 0.2$ , were listed. CRMs are classified in two main types in CisView: either Promoter, or either distal (Distal), e.g. enhancer, insulator, silencer. TFs in bold are those which are present in the original model.

| Gene  | CRM name   | TFBSs & Strand Orientation   | Type      |
|-------|------------|--|-----------|
| Tcf3  | CM10015455 | GR (-), AR (-), MYB (+), CP2 (+), ELK (-), ETS (-), RFX1 (+), IRF (+)  | Distal    |
| Tcf3  | CM10015456 | ∅  | Distal    |
| Tcf3  | CM10015460 | SREBP (-), ZF5 (+), Tax/CREB1 (-), SREBP1 (+), Egr1 (+), MyoD1 (-), SP1 (-), Pax4 (-), <b>KLF</b> (-), GATA (+)  | Promoter. |
| Mapk1 | CM16003054 | SP1 (-), MTF1B (+), MAZ (-), WHN (-), SMAD (-)   | Promoter. |
| Stat3 | CM11023938 | MYB (+), AREB6 (+), IRF (-), GCNF (+), SF1A (+), LRH (+), <b>SOX</b> (+/-), LEF (+), HMG (+), Sox18 (-), SRY (-), FOXO (+/-), NF1 (+), Pax4 (-), CEBP (-), NKX61 (-), HOXA4 (+), MSX (+), HOXA5 (+), ZID (+), POU6F (+), FOXD1 (-), CDP (+), EVI (-), RP58 (-), MyoD (+), Tbx5 (+), FOXD (-), Sox9 (+) | Distal.   |

Table 3: Bindings in each detected CRM (1)

| Gene          | CRM name   | TFBSs & Strand Orientation   | Type      |
|---------------|------------|--|-----------|
| Stat3         | CM11023939 | MEIS (-), Pax4 (-), <b>KLF</b> (-), SREBP1 (+), MyoD (+/-), LF1 (-), ELK (+), ETS (+), TST (+), NFY (+/-), BRN (+), SRF (+), CHX (+/-), LHX (-), SP1 (-), BRACH (-), FXRDR3 (-), MTF1B (+), MAZ (+/-), AIRE (-), AP2 (-), AP2A (+), HEN1 (-), AP4 (+), AHR (+), Pax8 (+), Pax3 (+), ATF1 (+), CREB (+), E4F1 (+), ATF (+), MSX (+), LRH (-), ARNT (+), MYC (+), Tbx5 (+), RFX1 (+), MEF3 (+), ATF1 (+), ATF6 (-), NFAT (+), Egr1 (+) | Promoter. |
| Pou5f1 (Oct4) | CM17006875 | MyoD (-), <b>Oct4/Sox2</b> (-), HMG (+), Sox18 (+), LEF (-), AP1 (-), AP2A (+), YY1 (-), AIRE (+), MAZ (+)   | Distal.   |

Table 4: Bindings in each detected CRM (2)

| Gene             | CRM name   | TFBSs & Strand Orientation   | Type      |
|------------------|------------|--|-----------|
| Pou5f1<br>(Oct4) | CM17006876 | MAZ (+/-), PU1 (+), LEF (+), ETS (+/-), TEL (+), LF1 (+), SRF (+), ARP1 (-), NKX3A (-), FOXO (+/-), COMP (+), GCNF (+), SF1A (+), LRH (+), <b>NANOG1</b> (+), FXRIR1 (+), <b>Sox</b> (+), HNF4 (+), LEF (+), HMG (-), HSF (-), EVI (-), Pax4 (-), <b>KLF</b> (+), MAZR (+) | Distal.   |
| Pou5f1<br>(Oct4) | CM17006877 | Pax4 (+/-), <b>KLF</b> (-), SP1 (+), COUP (-), PPAR (+), HNF4 (-), SF1B (+), SF1 (-), GCNF (+), LRH (+), MAZ (+)   | Promoter. |

Table 5: Bindings in each detected CRM (3)

| Gene  | CRM name   | TFBSs & Strand Orientation   | Type      |
|-------|------------|--|-----------|
| Sall4 | CM02038711 | ZF5 (-), MTF1B (+/-), SREBP (+), RORA (+), SF1B (+), SF1 (-), MAZ (+/-), PU1 (+), FXR (-), AIRE (-), MyoD (+), LEF (-), ARP1 (+), MEF2 (+/-), RSRFC4 (-), GFI (+), HMG (+), YY1 (-), <b>KLF</b> (+), Pax4 (+), RFX1 (+), AHR (+)   | Promoter. |
| Sox2  | CM03005876 | FOXD (+), FOXO (+/-), FXRIR1 (+), <b>SOX</b> (-), <b>NANOG1</b> (+/-), Egr1 (+), MAZ (+), HELIOS (+), FOXD1 (+/-), FOXA (-), HNF1 (-), NKX3A (+), YY1 (-), HMX (+), LEF (+/-), HMG (-), Sox18 (-), SRY (+/-), MEF2 (-), XVENT (+), ATF6 (-), CREB (+), SP1 (-), MyoD (-) | Distal.   |

Table 6: Bindings in each detected CRM (4)



| Gene | CRM name   | TFBSs & Strand Orientation  | Type      |
|------|------------|---|-----------|
| Sox2 | CM03005877 | ETS (+/-), TEL (+), FOXP3 (-), NFY (+/-), FOXO (-), GABP (-), ER (+), SP1 (+/-), MAZR (+), HMG (+), LEF (-), Pax8 (-), COMP (+), ELK (+), ETS (+), Pax2 (+/-), MEIS (+), SREBP1 (+), MYC (+), HNF1 (-), <b>NANOG</b> (+), NKX3A (+), TGIF (+), EVI (+), HNF3 (+), HELIOS (+), NKX25 (+), AP2A (+), NFAT (-), MYB (-), GCM (+) | Promoter. |
| Klf4 | CM04010396 | CTCFA (-), SMAD (+), ZF5 (-), NKX3A (-), MAZ (+/-), SP1 (+/-), EVI (-), RFX (-), RFX1 (-), FOXD1 (+), FOXD (+), FOXO (+/-), FOXA (-), FOXJ1 (+), SRY (+), ARNT (+), MYC (+/-), Pax3 (+), CREB (+), <sup>9</sup> E4F1 (+), AP4 (+), MyoD (+), ZID (-), Tax/CREB1 (+)   | Distal.   |

Table 7: Bindings in each detected CRM (5)

| Gene  | CRM name   | TFBSs & Strand Orientation   | Type                    |
|-------|------------|--|-------------------------|
| Tbx3  | CM05023198 | MAZ (-), CTCFB (-), SP1 (+/-), MEF3 (-), TBX5 (-), TATA (-), ZID (+), NFY (+/-), WHN (-) E2F (-), COMP (-), LEF (-), PBX (+), AP2A (-), Pax3 (-), GFI (-), LHX (+), NKX61 (+), FOXJ2 (+), MEF2 (+/-), RSRFC4 (-), ARNT (-), Pax3 (+), PU1 (-), <b>Sox</b> (-), HMG (+), Sox18 (+), LEF (-) | Promoter. CpG Isl.: 109 |
| Klf2  | CM08013093 | SP1 (+), AIRE (-), FOXD1 (-), CP2 (-), GR (-), AR (-)  | Promoter. CpG Isl.: 151 |
| Nanog | CM06025863 | <b>Oct4/Sox2</b> (-), HMG (+), MyoD (+), PITX (+)  | Promoter.               |
| Esrrb | CM12018771 | SRF (-), TCF11 (+), IRF (+), <b>STAT</b> (+), NFAT (+), HMX (-), NKX25 (-), GCNF (-), Tax/Creb1 (+), ATF1 (+), Sox9 (-)  | Distal.                 |

Table 8: Bindings in each detected CRM (6)

| Gene  | CRM name   | TFBSs & Strand Orientation   | Type      |
|-------|------------|--|-----------|
| Esrrb | CM12018772 | FAC1 (-), EVI (-), MEIS (-), MyoD (-), AP4 (+), HLF (+), CDP (+), PBX (+), MEF2 (-), RSRFC4 (+), TATA (-), AP1 (+/-), LEF (-), RFX (+), TBX5 (-), BRACH (-), EF1D (+), NKX3A (+), HNF3 (+), <b>Oct4/Sox2</b> (+), NKX61 (+), <b>Nanog</b> (-), CDP (+), IPF1 (-), MEF2 (+) | Promoter. |
| Esrrb | CM12018795 | GCNF (+), REB6 (+), FOXJ2 (+), MEF2 (+), GATA (+), <b>NANOG1</b> (+), <b>Sox</b> (+), CDC5 (+), NFKB (+), FOXD1 (+), HOXA5 (-), CART (+), CDP (+), PBX (-), HELIOS (-), AREB6 (+), MyoD (+), IK (+), Pax4 (+), EF1D (-), TBX5 (+)  | Distal.   |

Table 9: Bindings in each detected CRM (7)

| Gene    | CRM name   | TFBSs & Strand Orientation   | Type      |
|---------|------------|--|-----------|
| Esrrb   | CM12018796 | KLF (+/-), HMG (-), Pax4 (-), MEIS (-), NFY (+), ZF5 (+), Pax8 (+), GFI (+), MYB (+), IK (+), Oct4A (+), ATF6 (-), AIRE (-), PU1 (-), TEF (-), TEF1 (+), TCF11 (+), HMX (-), IRF (+), HELIOS (-), E2F (+), Pax6 (+), GCM (+)   | Promoter. |
| Tfcp2l1 | CM01022019 | TATA (+)   | Promoter. |
| Gbx2    | CM01018070 | LXR2 (-), BRN(+/-), Pax2 (-), TEF1 (+), AP1 (+), MEF2 (-), CART (-), CHX (+/-), OCT (+), GATA (+/-), EVI (-), ZID (+), YY1 (+), MAZ (+), PU1 (+), CEBP (+), HLF (-), ATF6 (+), ARNT (+/-), CREB (-), CDX (+), HNF3 (+), BARBIE (+), POLYA (+), E4BP4 (+), MEF2 (+), HOXA5 (-), IRF (+), LHX (+/-), CHX (+), RFX1 (+), MZF (-), CP2 (-) | Distal.   |

Table 10: Bindings in each detected CRM (8)

| Gene | CRM name   | TFBSs & Strand Orientation   | Type    |
|------|------------|--|---------|
| Gbx2 | CM01018071 | Sox9 (-), HMG (-), IPF1 (+), GFI (+), GABP (+), ELK (+), ETS (+), ATF6 (-), E4BP4 (+), LHX (+), CHX (-), NKX61 (+), HOXA4 (-), LYF (+), GATA (+), CDX (+), <b>Stat</b> (-), LEF (+)  | Distal. |
| Gbx2 | CM01018072 | MEF2 (+), NKX3A (-), IRF (-), HMX (-), Pax4 (-), NKX25 (+), NCX (+), XVENT (+), CDX (-), <b>Oct4/Sox2</b> (+/-), Pou3f (+), <b>Oct</b> (+), <b>Oct4A</b> (+), POLYA (-), FOXJ2 (-), CDP (-), MRF2 (+), TST (+), FOXD1 (-), CART (-), HNF1 (-), LHX (-), HNF3 (+), FAC1 (+/-), ATF1 (+), ATF6 (-), CREB (+), COMP (+), PBX (-), EVI (+), LXR2 (+), HNF1 (+), CDP (-), AREB6 (+) | Distal. |

Table 11: Bindings in each detected CRM (9)

### 3 Interactions in CRN

| Gene  | TF Bindings to $\geq 1$ CRM | Regulators in Original Model[1]       | Number of Matches Overall TEs |
|-------|-----------------------------|---------------------------------------|-------------------------------|
| Tcf3  | KLF                         | NANOG, Stat3, Tbx3, Oct4, Klf2, Mapk1 | $\frac{1}{6}$                 |
| Mapk1 | $\emptyset$                 | $\emptyset$                           | 1                             |
| Stat3 | KLF                         | Esrrb, Tcf3, Sox2                     | $\frac{0}{3}$                 |

Table 12: Interactions in the resulting CRN (1)

| Gene        | TF Bindings to $\geq 1$ CRM | Regulators in Original Model[1]         | Number of Matches Overall TEs |
|-------------|-----------------------------|---|-------------------------------|
| Oct4/Pou5f1 | Oct4, Sox2, KLF, NANOG1     | Tcf3, Gbx2, Tfcp2l1, NANOG, Klf2, ESRRB | $\frac{2}{6}$                 |
| Klf4        | $\emptyset$                 | Gbx2, Tfcp2l1, Tbx3, Klf2, ESRRB, Stat3 | $\frac{0}{6}$                 |
| Sall4       | KLF                         | Klf2, Sox2, Tfcp2l1, Tbx3               | $\frac{1}{4}$                 |

Table 13: Interactions in the resulting CRN (2)

| Gene | TF Bindings to $\geq 1$ CRM | Regulators in Original Model[1]                            | Number of Matches Overall TEs |
|------|-----------------------------|--|-------------------------------|
| Sox2 | SOX, NANOG                  | Sall4, Stat3, Tbx3, NANOG                                  | $\frac{1}{4}$                 |
| Tbx3 | Sox                         | Sox2, Tcf3, Sall4, Gbx2, Tfcp2l1                           | $\frac{1}{5}$                 |
| Klf2 | $\emptyset$                 | Tcf3, Sall4, Gbx2, Tfcp2l1, Tbx3, Oct4, NANOG, Klf4, ESRRB | $\frac{0}{9}$                 |

Table 14: Interactions in the resulting CRN (3)

| Gene   | TF Bindings to $\geq 1$ CRM  | Regulators in Original Model[1]                          | Number of Matches Overall TFs |
|--------|------------------------------|--|-------------------------------|
| NANOG  | Oct4, Sox2                   | Klf2, ESRRB, Mapk1, Tcf3, Sox2, Tfc2l1, Oct4             | $\frac{2}{7}$                 |
| Esrrb  | STAT, Oct4, Sox2, Nanog, KLF | Gbx2, Stat3, Tfc2l1, Tbx3, Oct4, Nanog, Klf4, Klf2, Tcf3 | $\frac{4}{9}$                 |
| Tfc2l1 | $\emptyset$                  | Sall4, Tbx3, Oct4, Nanog, Klf4, Klf2, Esrrb, Tcf3, Stat3 | $\frac{0}{9}$                 |
| Gbx2   | OCT, Stat, Sox2              | Tbx3, Oct4, Klf4, Klf2, ESRRB, Stat3                     | $\frac{2}{6}$                 |

Table 15: Interactions in the resulting CRN (4)

## 4 Results

### 4.1 Statistics

- Number of TF matches overall all TF bindings in the original model: 18.9%
- Average percentage of TF matches overall TF bindings for each gene: 25.1%

### 4.2 Figures

See script *rm\_analysis.R* in the code section.

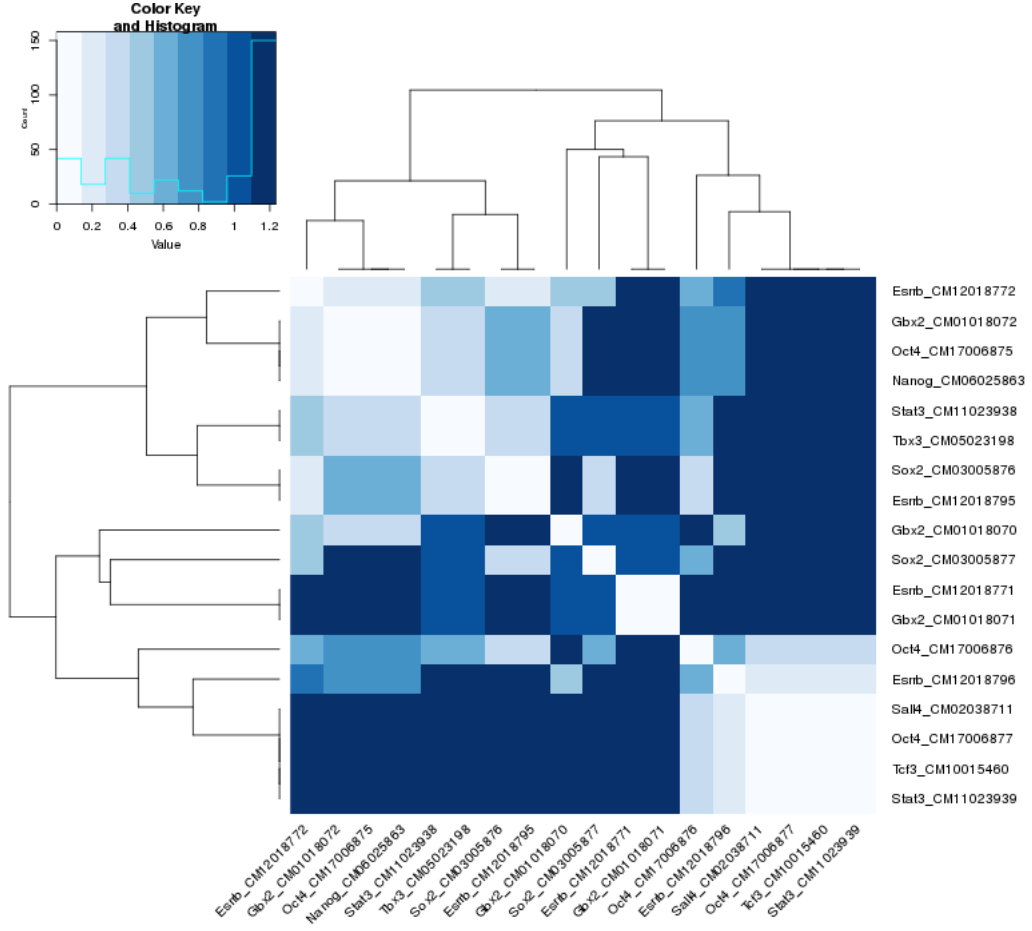


Figure 1: Spearman's  $\rho$  correlation matrix between all the CRMs (for TFs involved in the original model). Original matrix  $M$  is such as  $M_{i,j} = \delta_{TF_j \in CRM_i}$  with  $\delta$  being the Kronecker symbol (associated function: *heatmap\_c*).



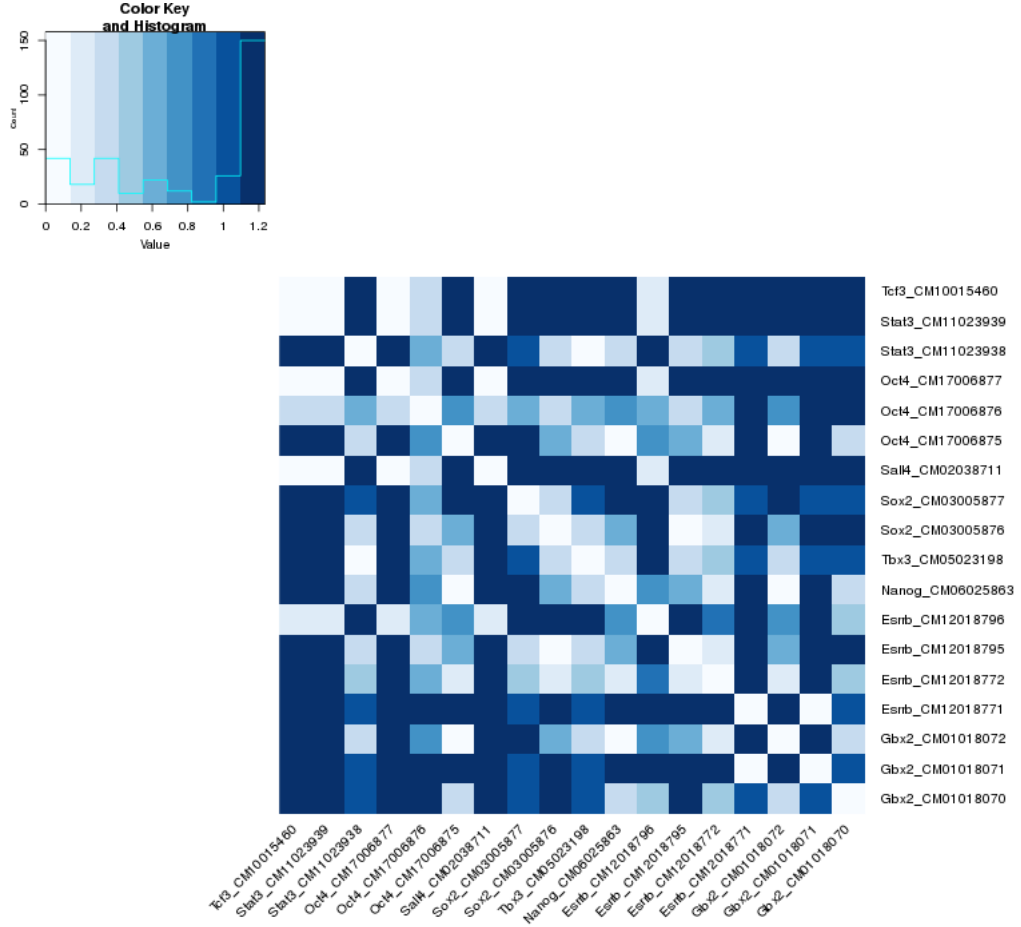


Figure 2: Spearman's  $\rho$  correlation matrix between all the CRMs (for TFs involved in the original model). Original matrix  $M$  is such as  $M_{i,j} = \delta_{TF_j \in CRM_i}$  with  $\delta$  being the Kronecker symbol (associated function: *heatmap\_c2*).

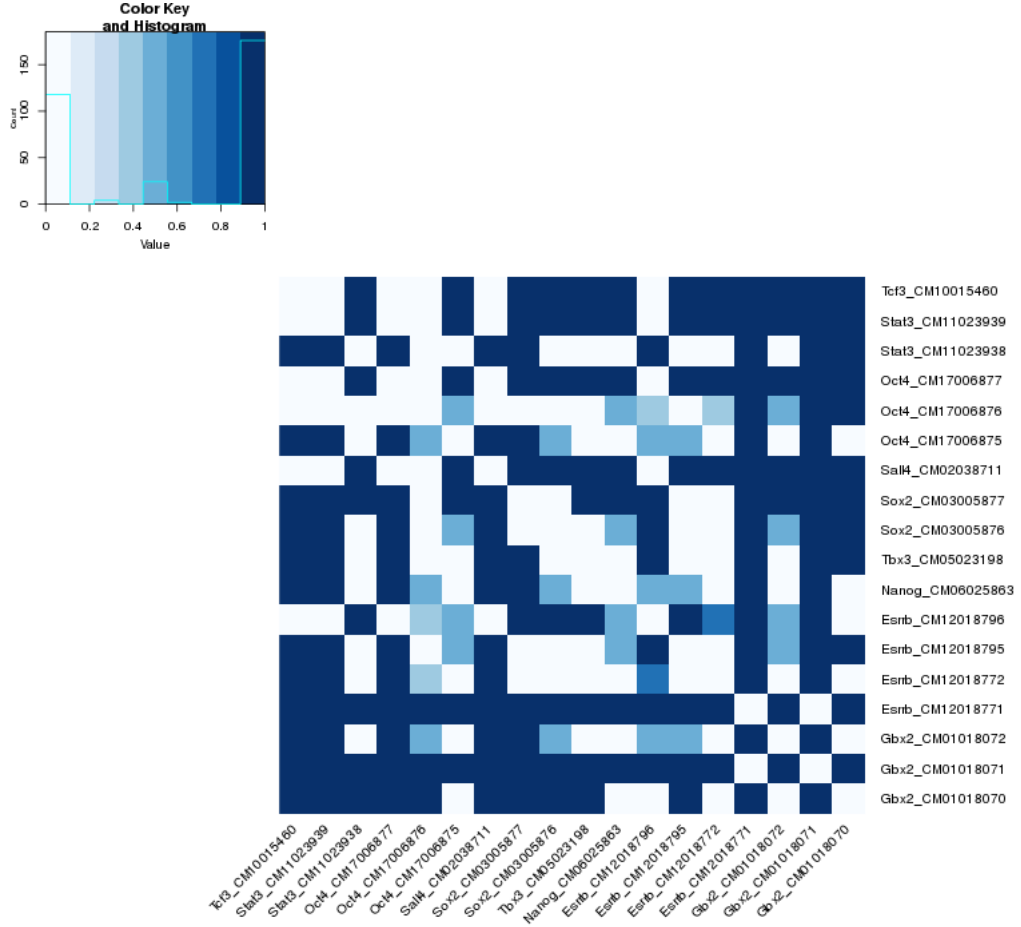


Figure 3: Overlap matrix between all the CRMs (for TFs involved in the original model). Original matrix  $M$  is such as  $M_{i,j} = 1 - \frac{|CRM_i \cap CRM_j|}{\min(|CRM_i|, |CRM_j|)}$  (associated function: *overlapMatrix*).

## References

- [1] S-J Dunn, Graziano Martello, Boyan Yordanov, Stephen Emmott, and AG Smith. Defining an essential transcription factor program for naive pluripotency. *Science*, 344(6188):1156–1160, 2014.
- [2] Donna Karolchik, Robert Baertsch, Mark Diekhans, Terrence S Furey, Angie Hinrichs, YT Lu, Krishna M Roskin, Matthias Schwartz, Charles W Sugnet, Daryl J Thomas, et al. The ucsc genome browser database. *Nucleic acids research*, 31(1):51–54, 2003.
- [3] Alexei A Sharov, Dawood B Dudekula, and Minoru SH Ko. Cisview: a browser and database of cis-regulatory modules predicted in the mouse genome. *DNA research*, 13(3):123–134, 2006.

