

Additional File 2. Construction of motif dictionary by integration across seven sources. The table lists the number of motifs obtained from each source before and after filtering. The number of unique human genes associated with motifs is listed in parentheses. The initial set of 4378 motifs was filtered to remove redundant motifs and motifs with low information content, yielding the final set of 2935 motifs used in our analyses.

| Source | No. Motifs (pre-filter) | No. Motifs (post-filter) |
|-------------------------------------|-------------------------|--------------------------|
| hPDI ¹ | 436 (436) | 420 (420) |
| Jaspar ² | 633 (412) | 285 (228) |
| Jolma Taipale 2013 ³ | 843 (404) | 333 (245) |
| Kheradpour Kellis 2013 ⁴ | 292 (94) | 220 (84) |
| TRANSFAC ⁵ | 1799 (801) | 1449 (720) |
| UniPROBE ⁶ | 296 (277) | 192 (185) |
| Wang Weng 2013 ⁷ | 79 (110) | 36 (56) |
| Total | 4378 (1559) | 2935 (1422) |

¹Xie et al. 2010, *Bioinformatics* 26:287-289

²Mathelier 2014, *Nucleic Acids Res* 42:D142-147

³Jolma et al. 2013, *Cell* 152:327-339

⁴Kheradpour and Kellis 2014, *Nucleic Acids Res* 42:2976-2987

⁵Matys et al. 2006, *Nucleic Acids Res* 34:D108-110

⁶Robasky and Bulyk 2011, *Nucleic Acids Res* 39:D124-D128

⁷Wang et al. 2012, *Genome Res* 22:1798-1812