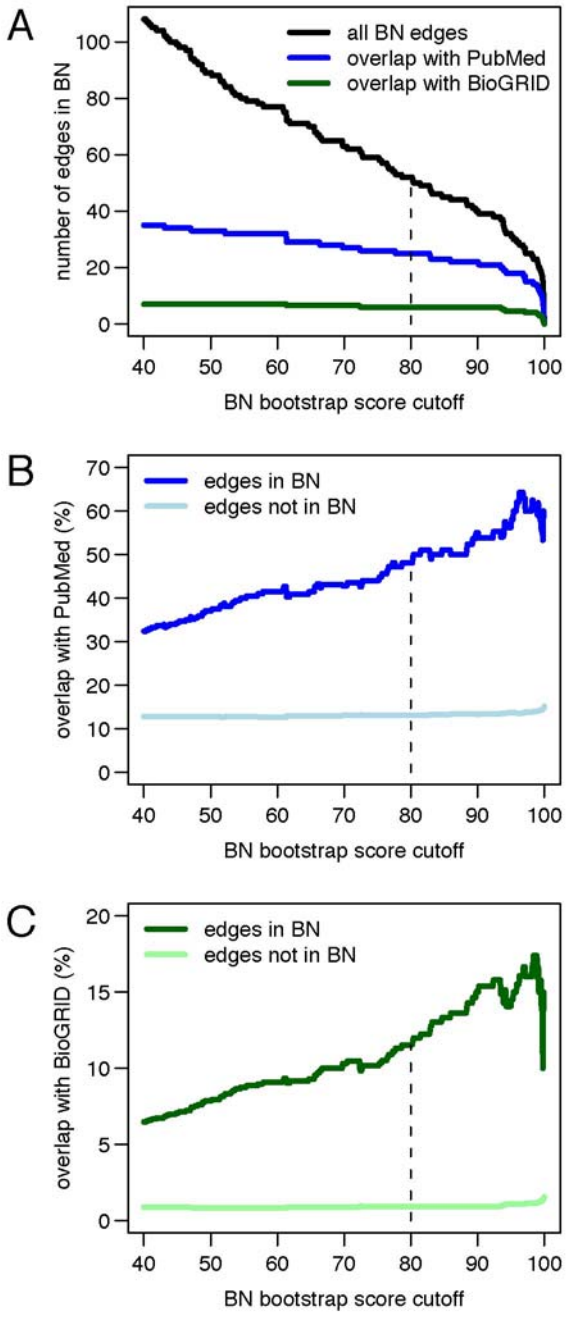
van Steensel *et al*: **Supplementary Information**

Suppl Figure S1: Overlap of BN with PubMed co-citations and BioGRID interactions.
Suppl Figure S2: Effect of binding data discretization on BNI performance.
Suppl Figure S3: Network distribution of genes expressed in specific embryonic tissues.
Suppl Figure S4: Analysis of the robustness of BNI.

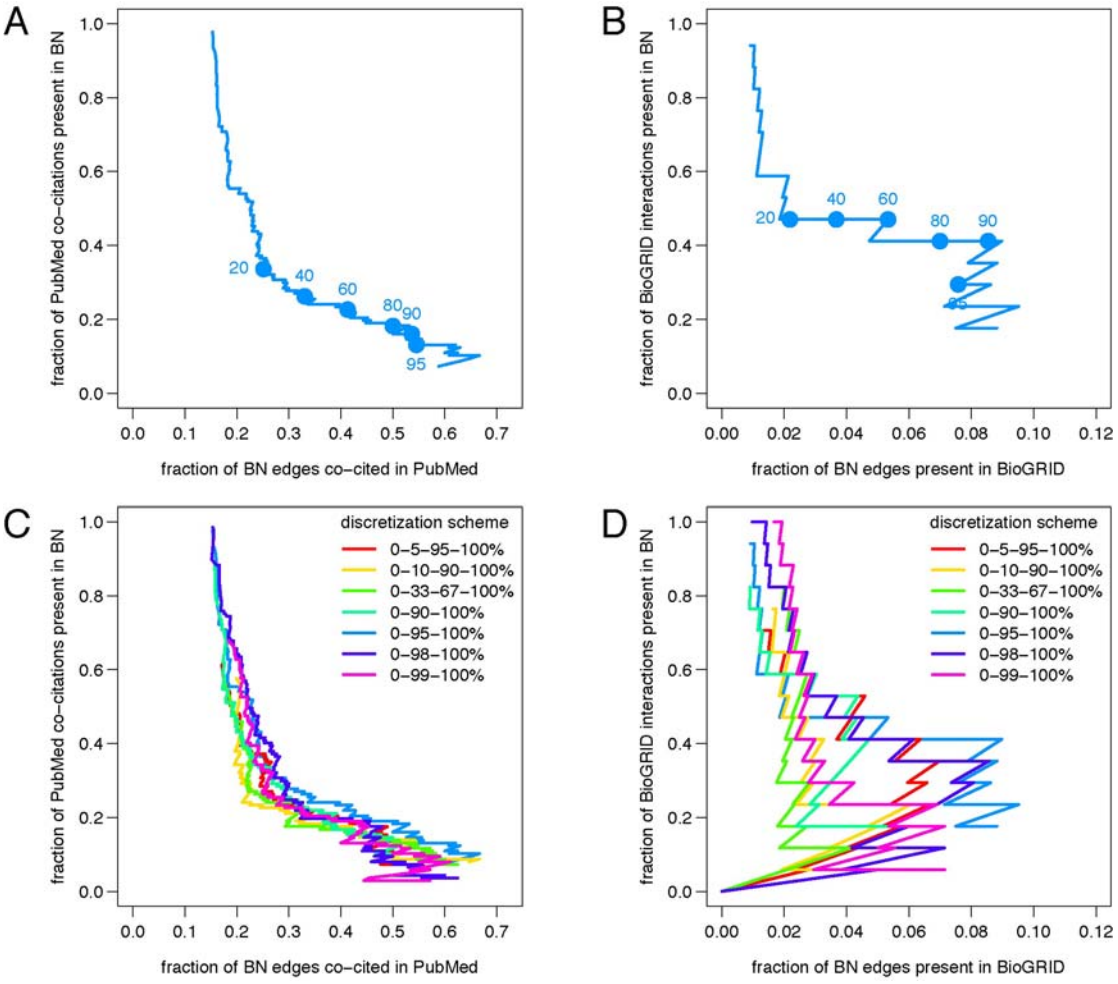
Suppl Dataset 1: Binding profiles of 43 chromatin components, with probe annotation. (Excel file)
Suppl Dataset 2: BNI bootstrap scores (based on 1000 bootstraps) for all pairs of chromatin components with scores > 0 (Tab-delimited file). This dataset was used to construct BN_{80}.
Suppl Dataset 3: Binding profiles of GAF, Jra, HP1 and Su(var)3-7 after RNAi knockdown of HP1, brm or white (Tab-delimited file).
Suppl Dataset 4: Parameter settings file for Banjo software (Text file).
Suppl Dataset 5: Search terms and results of PubMed co-citation analysis (Excel file)

Supplementary Figure S1. Overlap of BNI with databases of previously reported associations.

(A) Total number of edges in BN as a function of bootstrap score cutoff (black line), and the number of these edges that overlap with links between chromatin components as listed in PubMed (co-citations, blue) or BioGRID (physical or genetic interactions, green). **(B-C)** Dark colored lines: overlap with PubMed co-citations (B) or BioGRID interactions (C) as percentage of the total number of edges in the BN. Pale colored lines: same, but for all possible edges that are not part of the BN at the indicated cutoff. Vertical dotted line marks the 80% bootstrap score cutoff, which defines BN_{80}.



Supplementary Figure S2. Effect of binding data discretization scheme on BNI performance. (A-B) Performance curve of the BN obtained after binarization if the binding profiles using a 95th percentile threshold. Overlap is shown with PubMed co-citations (A) or interactions listed in BioGRID (B), plotted for all possible bootstrap score cutoffs, some of which are highlighted by solid circles. (C-D) The same overlap plots for various discretization schemes used to generate the input data for Banjo. The binarization scheme using the 95th percentile threshold yields a curve (light blue) that is located more to the upper right corner, indicating that it gives the most reliable performance of all schemes.



Supplementary Figure S3. Enrichment and depletion of tissue-specific genes among target genes of each chromatin component. Node colors depict enrichment (yellow) or depletion (blue) of genes that are expressed in the tissue as indicated above each graph, using terminology as in (Tomancak et al. 2007). Node sizes depict the statistical significance of the observed enrichment or depletion (two-sided binomial test), ranging from $P>10^{-3}$ (smallest nodes) to $P\leq10^{-8}$ (largest nodes). Only tissues with significant enrichment or depletion in at least one BN node are shown, one per page.

(see separate multi-page PDF document, available for download).

Supplementary Figure S4. Robustness of BN_{80}. Random combinations of 1-5 chromatin components (nodes) were removed from the original dataset (each time 50 combinations, or all 43 in the case of single deletions), and new networks were constructed by BNI, again using a bootstrap confidence threshold of 80%. The change in overlap of these reduced networks relative to the original 43-node network (Figure 2) was calculated as explained in the Methods section "Robustness of BNI". (A) Change in overlap of all edges. (B) Change in orientation of edges. The lower number of bootstraps used here reduces the accuracy of the bootstrap scores, and as a consequence causes a somewhat higher variability in the resulting networks; this background variability was estimated by repeating BNI 50 times on the original 43-component dataset (grey box, "0"). On average, for every one node difference, changes of connectivity are expected on ~2% of edges, and changes of orientation on ~1%. of them.