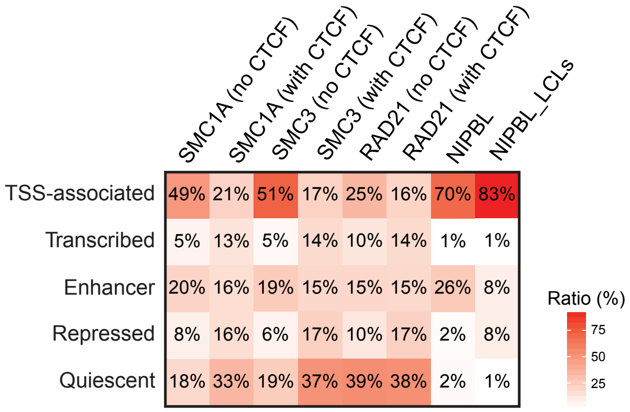

Figure S1. Comparison of the genome-wide occupancy between NIPBL and cohesin subunits ChIP-Seq datasets. Heatmap showing genome-wide proportions of simplified chromatin states in regions occupied by SMC1A, SMC3 and RAD21 with or without CTCF in GM12878 cells and by NIPBL in GM12878 cells and LCLs. Different subunits of the cohesin complex similarly occupy the genome. NIPBL genome-wide occupancy pattern is similar in GM12878 and LCLs. The color scale indicates the ratio of overlap.

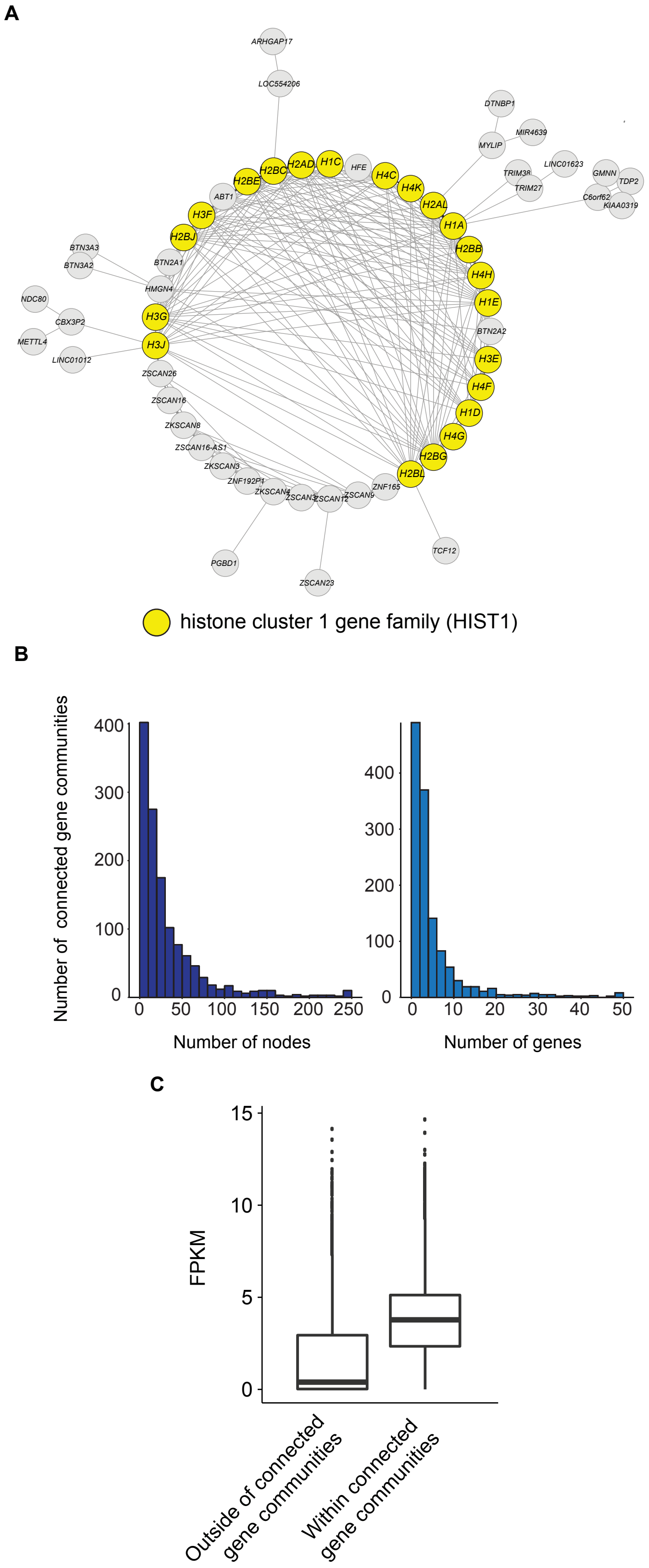

Figure S2. Characterization of connected gene communities in normal lymphoblastoid cells.
(**A**) Graphical representation of the connected gene community containing genes from the histone cluster 1 family (HIST1) as represented by yellow nodes. 21 of the 58 genes are connected. (**B**) Composition of connected gene communities. The left panel shows the number of connected gene communities in function of their number of nodes. The right panel shows the number of connected gene communities in function of their number of genes. (**C**) Boxplot representation of the expression level of genes found outside or within connected gene communities. Genes in connected gene communities are expressed at a higher level than genes found outside of these network

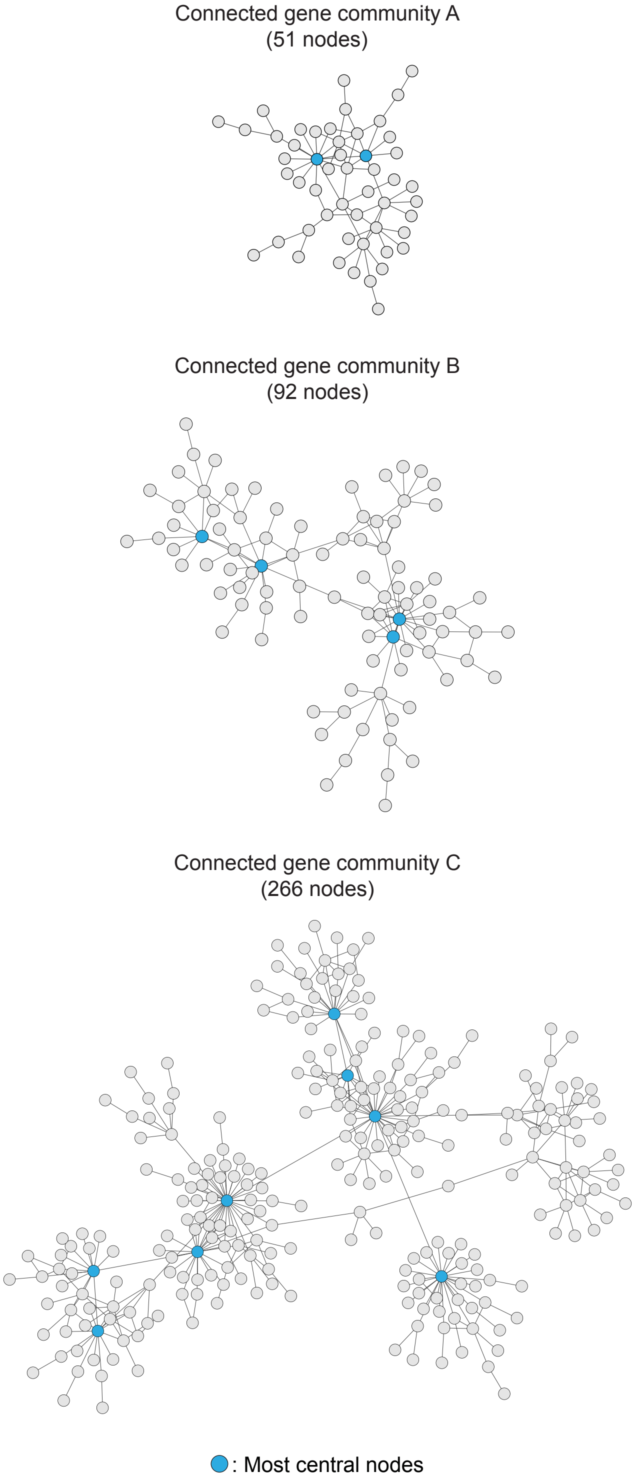

Figure S3. Identification of the central nodes within connected gene communities.
Graphical representation of three connected gene communities with different sizes. The nodes with the highest centrality scores in each community are represented in blue.

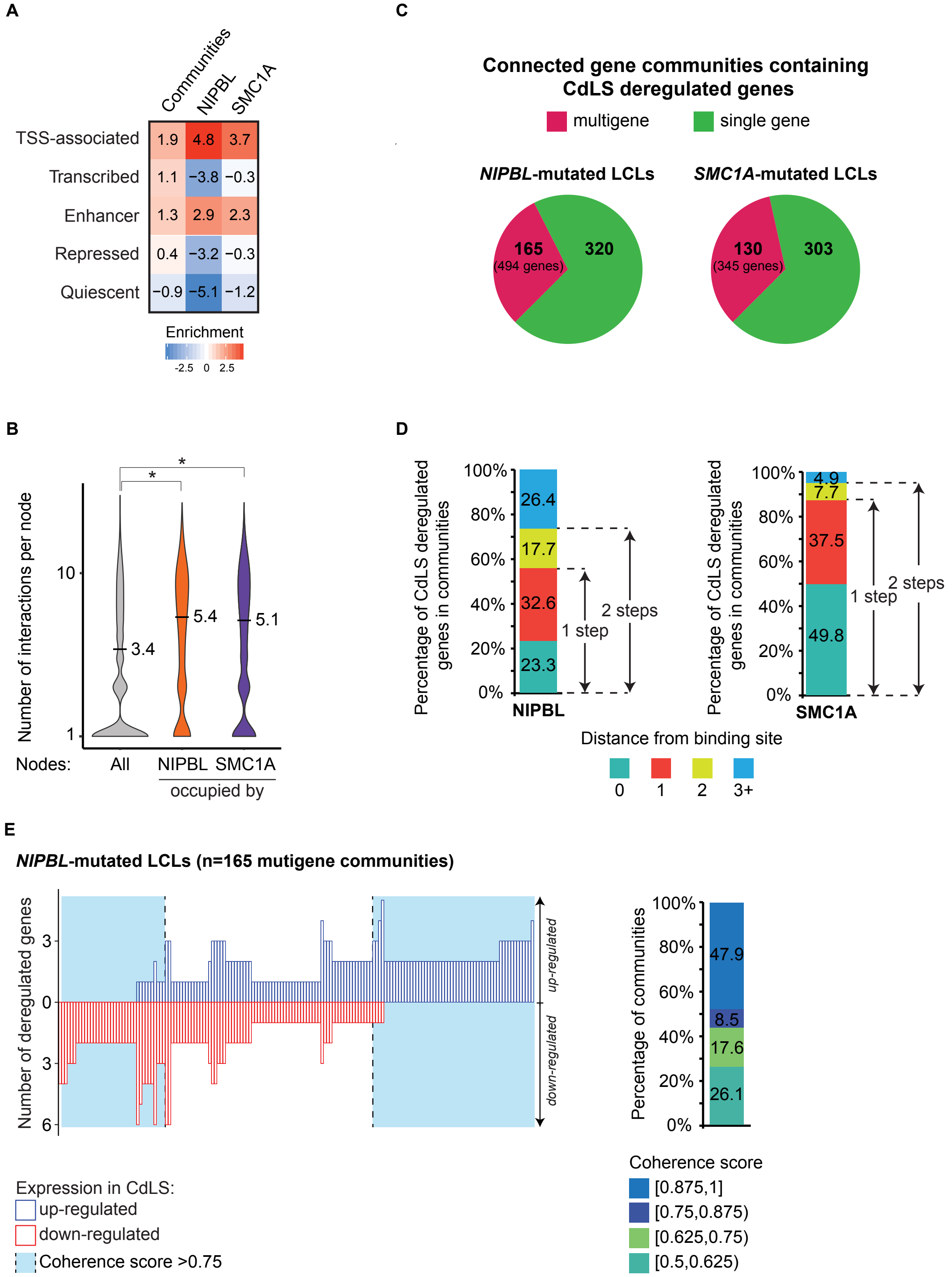
A
Communities
NIPBL
SMC1A
TSS-associated
1.9
4.8
3.7
Transcribed
1.1
−3.8
−0.3
Enhancer
1.3
2.9
2.3
Repressed
0.4
−3.2
−0.3
Quiescent
−0.9
−5.1
−1.2
Enrichment
-2.5
0
2.5
B
Number of interactions per node
10
1
*
*
3.4
5.4
5.1
Nodes:
All
NIPBL
SMC1A
occupied by
C
Connected gene communities containing CdLS deregulated genes
multigene
single gene
NIPBL-mutated LCLs
SMC1A-mutated LCLs
165
(494 genes)
320
130
(345 genes)
303
D
Percentage of CdLS deregulated genes in communities
100%
80%
60%
40%
20%
0%
26.4
17.7
32.6
23.3
1 step
2 steps
NIPBL
4.9
7.7
37.5
49.8
1 step
2 steps
SMC1A
Distance from binding site
0
1
2
3+
E
NIPBL-mutated LCLs (n=165 mutigene communities)
Number of deregulated genes
3
0
3
6
up-regulated
down-regulated
Percentage of communities
100%
80%
60%
40%
20%
0%
47.9
8.5
17.6
26.1
Expression in CdLS:
up-regulated
down-regulated
Coherence score >0.75
Coherence score
[0.875,1]
[0.75,0.875)
[0.625,0.75)
[0.5,0.625)

Figure S4. Promoter Capture Hi-C-defined connected gene communities identify similar properties for NIPBL- and SMC1A-occupied nodes. (**A**) Heatmap showing the enrichment of simplified chromatin states within connected gene communities compared to all NIPBL- and SMC1A-occupied regions. The enrichment fold was calculated relative to the genome. NIPBL and SMC1A are enriched at TSS-associated and enhancer regions, similar to connected gene communities. The color scale indicates the enrichment versus the genome. (**B**) Violin plots representing the connectivity of NIPBL- and SMC1A-occupied nodes. The number of interactions for all nodes is displayed in grey while those for NIPBL- and SMC1A-occupied nodes are displayed in orange and purple respectively. On average, nodes occupied by NIPBL and SMC1A are involved in 5.4 ($p<2.2e-16$, Wilcoxon rank sum test) and 5.1 ($p<2.2e-16$, Wilcoxon rank sum test) interactions compared to 3.4 for all nodes within connected gene communities. (**C**) Distribution of CdLS-deregulated genes within connected gene communities. Genes deregulated in NIPBL-mutated LCLs are found in 485 connected gene communities (165 multigene). Genes deregulated in SMC1A-mutated cells are found in 433 connected gene communities (130 multigene). (**D**) Gene deregulated in CdLS are connected to NIPBL- and SMC1A-occupied nodes. Graphical representation of the proportion of CdLS deregulated genes in function of the distance from a NIPBL- or SMC1A-occupied node. A distance of 0 corresponds to the gene locus deregulated in CdLS being directly occupied by NIPBL or SMC1A +/- 1kb from the TSS. A distance of 1, 2 or 3 corresponds to the number of steps from the occupied node. (**E**) Coordinated gene expression changes in connected gene communities are associated with mutations in NIPBL. Left panel - Quantification of the number of up- and down-regulated genes in CdLS within individual multigene connected gene communities. Communities with at least 75% of coherency are highlighted in light blue. Right panel - Distribution of coherence score of connected gene communities. A coherence score of 1 represents that all CdLS-modulated genes are deregulated in the same direction.