

# **The heterochromatin landscape in migrating cells and the importance of H3K27me3 for migration-associated transcriptional changes**

Tamar Segal, Mali Salmon-Divon and Gabi Gerlitz

## **Supplementary Methods**

**ChIP-qPCR.** The described ChIP-seq protocol was scaled down by a factor of 4, thus for each point lysate from  $0.25 \times 10^7$  was taken. In addition to the above antibodies an additional IP point was carried out with 3 $\mu$ l of control rabbit IgG (Millipore, PP64B, CA, USA). Isolated DNA subjected to qPCR using Fast SYBR® Green Master Mix (Applied Biosystems, 4385612, Vilnius, Lithuania), according to the manufacturer's protocol. Each experiment was repeated three times in duplicates. Standard errors (SE) and Wilcoxon signed-rank test were calculated. A list of the primers used appears in Sup. Table 1.

**Reverse transcription of mRNA and qPCR.** Validation of the RNA-seq results of selected mRNAs was performed by reverse transcription followed by qPCR. In short, mature RNA was reverse-transcribed by GoScript™ Reverse Transcription System using polyT primers according to the manufacturer's protocols (Promega, A5000, Madison, USA). cDNA was subjected to qPCR using Fast SYBR® Green Master Mix (Applied Biosystems, 4385612, Vilnius, Lithuania), according to the manufacturer's protocol. The relative expression was normalized to RPS29 and each experiment was repeated at least three times in duplicates. Standard errors (SE) and T.test were calculated. A list of the primers used appears in Sup. Table 1.

**Heatmap of ChIP signal across genes.** A matrix containing the normalized ChIP-seq score across  $\pm 3000$  bp around the TSS of Refseq genes was prepared using bwtool [1]. To keep intervals in 5'→3' direction, regions around the TSS of genes transcribed from the negative

strand of the reference genome were reversed. An in-house script was written to calculate the average score across every 20 bp, leading to a signal matrix of 300 signal points around the TSS per gene. K mean clustering algorithm was applied to divide the genes into clusters based on signal intensity around their TSS. The "elbow" method [2] was used in order to determine the optimal number of clusters. Briefly, a plot of the "within groups sum of squared errors" by the "number of clusters extracted" was generated. The estimated optimal number of clusters is the number under which the elbow or bend in the plot is found. Heatmaps presenting the distribution of TSS signal within each cluster were created by Complex Heatmap [3] and EnrichedHeatmap [4] R packages. For correlation calculation, the maximum signal value across 300 signal points was detected for each gene, and Pearson correlation coefficient was computed between these modification signal value and the gene expression level. For H4K20me1, correlation calculation was done separately for signal upstream the TSS and downstream the TSS.

#### Primers used in PCR analysis

| <b>Gene name</b>         | <b>Forward primer (5'→3')</b> | <b>Reverse primer (5'→3')</b> |
|--------------------------|-------------------------------|-------------------------------|
| <b><i>E-cadherin</i></b> | ACTTCTGCTAGACTCCTGCC          | ATCATGGGGCACGAGTCAG           |
| <b><i>Gapdh</i></b>      | TTGTGCAGTGCCAGGTGAAAA         | CCATCACGTCCTCCATCATCC         |
| <b><i>Line</i></b>       | GCCTAAGCCACAGCAGCAG           | ACACAGGGTGCTGCCTCAG           |
| <b><i>Ctgf</i></b>       | CCTGCGACCCACACAAGG            | GCTGCTTTGGAAGGACTCAC          |
| <b><i>Cyr61</i></b>      | AGACCCGGATCTGTGAAGTG          | GGGCCGGTATTTCTTGACAC          |
| <b><i>Fos</i></b>        | AGAGCGGGAATGGTGAAGAC          | CTGCAGCCATCTTATTCCGT          |
| <b><i>Id1</i></b>        | CTGAACGGCGAGATCAGTG           | TTTTCCTCTTGCCCTCCTGAA         |
| <b><i>Mgarp</i></b>      | TGGAACATCTGGCTCCAATA          | CAACTCCGCCTTTGTTTGT           |

|                  |                       |                        |
|------------------|-----------------------|------------------------|
| <i>Nppb</i>      | GCACAAGATAGACCGGATCG  | CTTCAAAGGTGGTCCCAGAG   |
| <i>Plk2</i>      | CAGATGCCCCTGAACAATTT  | GGAGGTAGAGCCGAGGTCTT   |
| <i>Ddit4</i>     | ACCAGTTCGCTCACCTTC    | AAACGATCCCAGAGGCTAGG   |
| <i>Hk2</i>       | GGCCGTGGTAAATGACACAG  | CCACCAGTTCCACATTACGC   |
| <i>Kdm3a</i>     | AAGGACCAGCAGAAGGACTC  | GCCAATCGCTTCACTGTCAT   |
| <i>Tnfsf9</i>    | ACTACACAACAGGGCTCTCC  | GTCTTCTTCGTACCTCAGACCT |
| <i>Bmt2</i>      | CACTGGGCAAAAACCTTGTA  | GTCTTAGTGGCGAGGACAGC   |
| <i>Narf</i>      | GAAGGGCAAACCTCCCATACC | CCGGAGTAGATCCCTTCCAT   |
| <i>Slc2a1</i>    | TTGTTGTAGAGCGAGCTGGA  | AAGGCCACAAAGCCAAAGAT   |
| <i>Rps29</i>     | TCGTTGGGCGTCTGAAGGCAA | CGGAAGCACTGGCGGCACAT   |
| <i>Serpinb9b</i> | CCAGAGTTGTTGGCAGATGAC | ACTGGCCTTTTCTCATCCTTGT |
| <i>Selenbp1</i>  | CTAACCCACAGAAGCCCCG   | GCTGGATCATCTGAGGGCC    |
| <i>Runx2</i>     | TGAGATTTGTGGGCCGGAG   | AGCTTCTGTCTGTGCCTTCT   |
| <i>Bmt2</i>      | CACTGGGCAAAAACCTTGTA  | GTCTTAGTGGCGAGGACAGC   |
| <i>Tnfsf9</i>    | ACTACACAACAGGGCTCTCC  | GTCTTCTTCGTACCTCAGACCT |
| <i>Pgbd5</i>     | TACAAGGTCCAGCCCTTCCT  | TTTCCGCTTTTTCCTCTTCC   |
| <i>Hsd17b7</i>   | GGACGTTGCTCCTACCCATA  | CGCGTATTTGGTCAGAGGAT   |
| <i>Frs3</i>      | GTGTTCGAGGGCAGAGGAC   | CAGAGACCGTGTAGCCAGC    |
| <i>Adamts1</i>   | GACCCGAGAGCCAGAACAC   | CTACGAGCATGGTTTCCACA   |

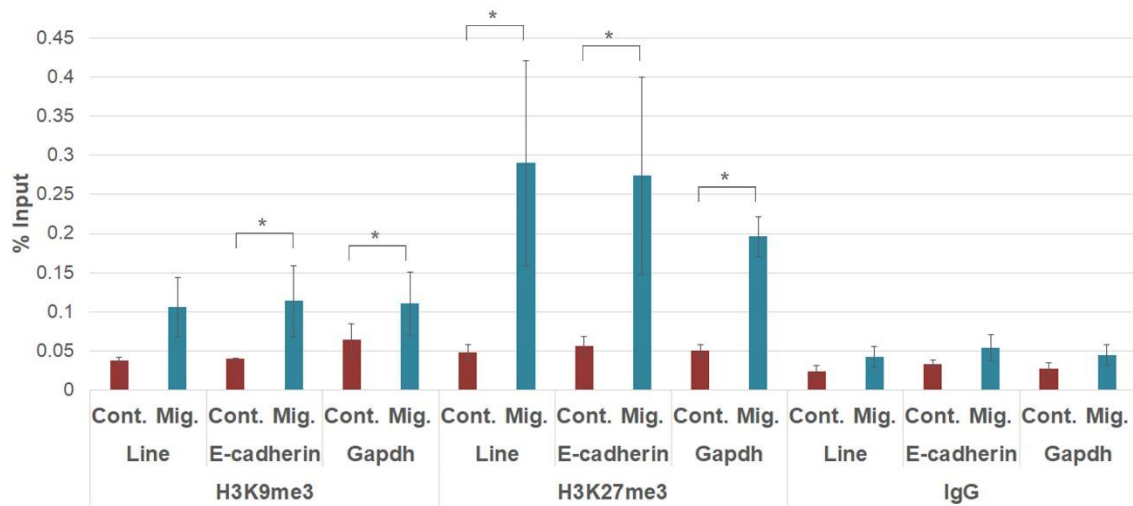
## References

- [1] A. Pohl, M. Beato, bwtool: a tool for bigWig files, *Bioinformatics*. 30 (2014) 1618–1619.  
doi:10.1093/bioinformatics/btu056.

- [2] D.J. Ketchen, C.L. Shook, The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique, *Strateg. Manag. J.* 17 (1996) 441–458.  
doi:10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.
- [3] Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, *Bioinforma. Oxf. Engl.* 32 (2016) 2847–2849.  
doi:10.1093/bioinformatics/btw313.
- [4] Z. Gu, R. Eils, M. Schlesner, N. Ishaque, EnrichedHeatmap: an R/Bioconductor package for comprehensive visualization of genomic signal associations, *BMC Genomics.* 19 (2018) 234.  
doi:10.1186/s12864-018-4625-x.

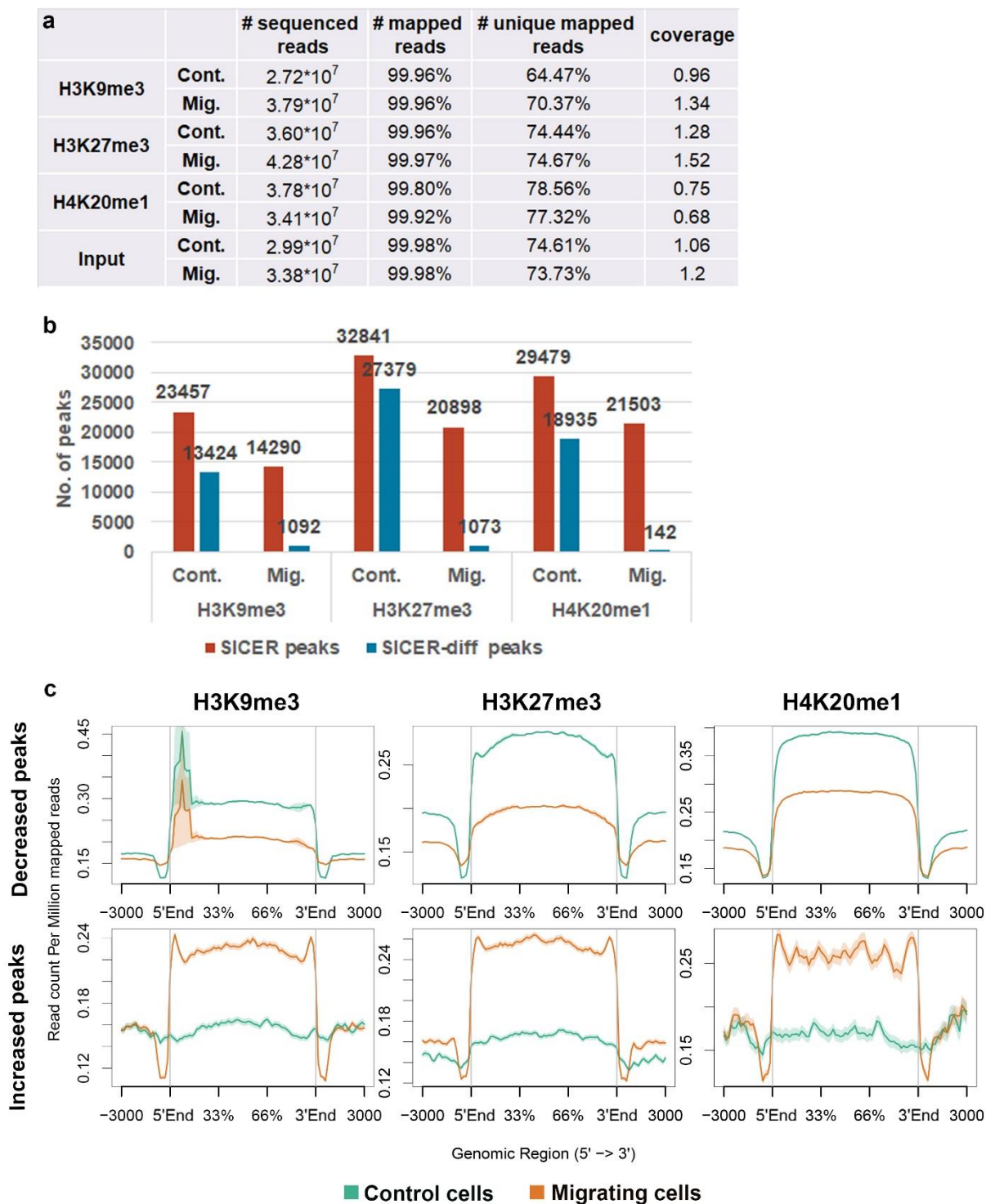
## Supplementary Figures

Figure S1



**Figure S1: Increased levels of heterochromatin markers in the promoters of E-cadherin, Gapdh and Line upon induction of migration.** ChIP-qPCR with antibodies against H3K9me3, H3K27me3 and control IgG. The immunoprecipitated DNA was analyzed with primers corresponding to the promoters of E-cadherin, Gapdh and Line in control cells (Cont.) and migrating cells (Mig.). Mean percentage of input  $\pm$  SE, (n= 3) is represented in the graph. \*P value<0.05 (Wilcoxon rank sum test).

Figure S2



**Figure S2: Enrichment of H3K9me3, H3K27me3 and H4K20me1 reads and peaks following ChIP-seq in control and migrating cells.** (a) ChIP-seq statistics observed in control cells (Cont.) and migrating cells (Mig.). The percentage of reads mapped to the mouse genome (mm10) was calculated out of the total number of sequenced reads. Similarly, the percentage of reads mapped uniquely to the genome was calculated out of the total mapped reads. Coverage was calculated by dividing the number of sequenced nucleotides by the genome length. (b) The number of peaks and differential peaks (FDR<0.05) detected by SICER or SICER-diff, respectively, for the three modifications in control and migrating cells. (c) Distribution of H3K9me3, H3K27me3 and H4K20me1 signals in control cells (green line) and migrating cells (orange line) across differential peak regions. "Decreased peaks" (top) and "Increased peaks" (bottom) refer to regions where the signal was lower or higher, respectively, in migrating cells compared to control cells. In each graph, the region between 5'End and 3'End represents the analyzed peak body. Since peaks are of variable sizes, the coverage vectors were normalized to be equal length prior to average coverage calculation. The units for the abscissa are bp.

Figure S3

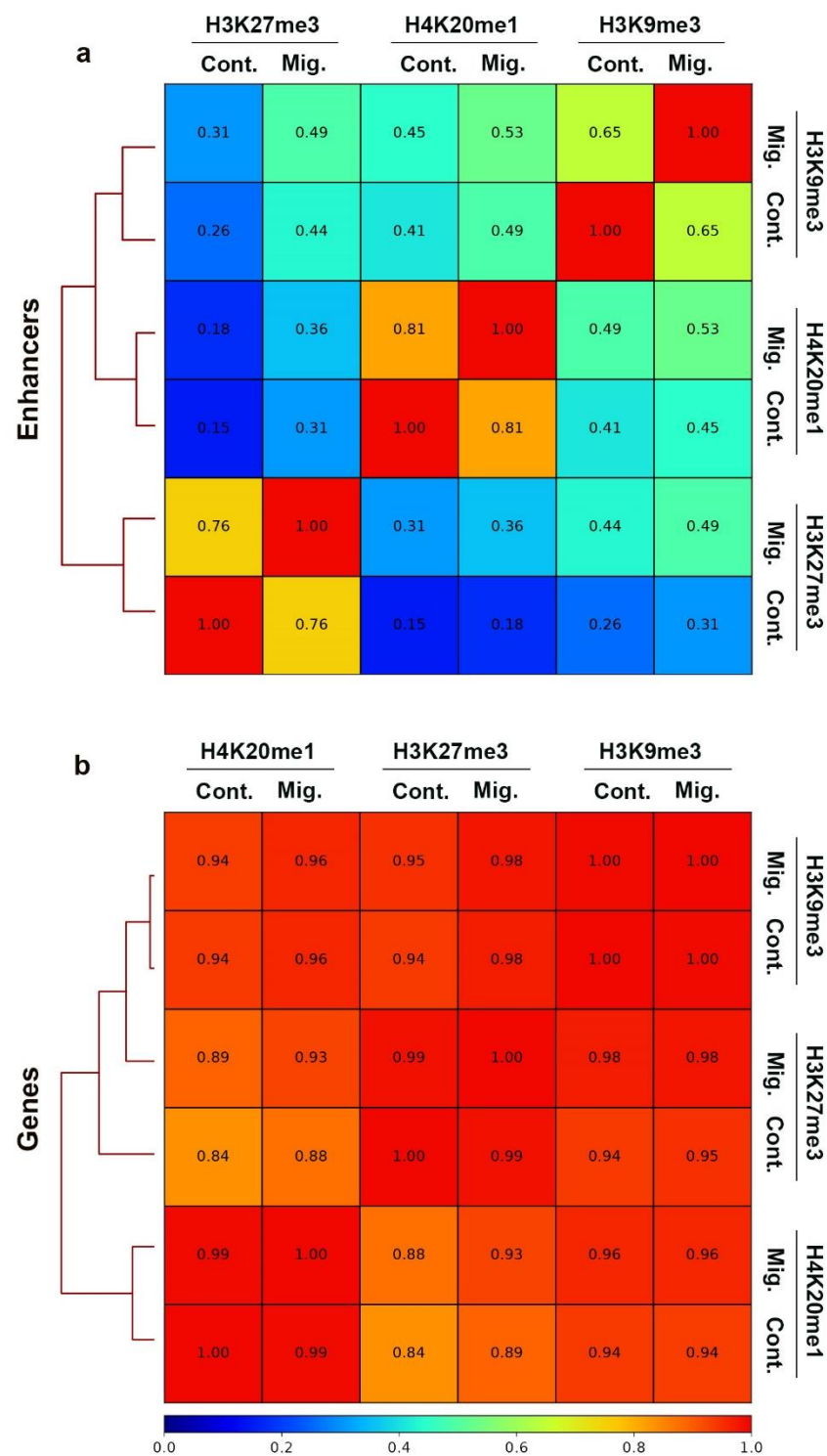


Figure S3: Correlation between H3K9me3, H3K27me2 and H4K20me1 ChIP-seq signal across specific genomic elements. Spearman correlation coefficients of the ChIP-seq signals



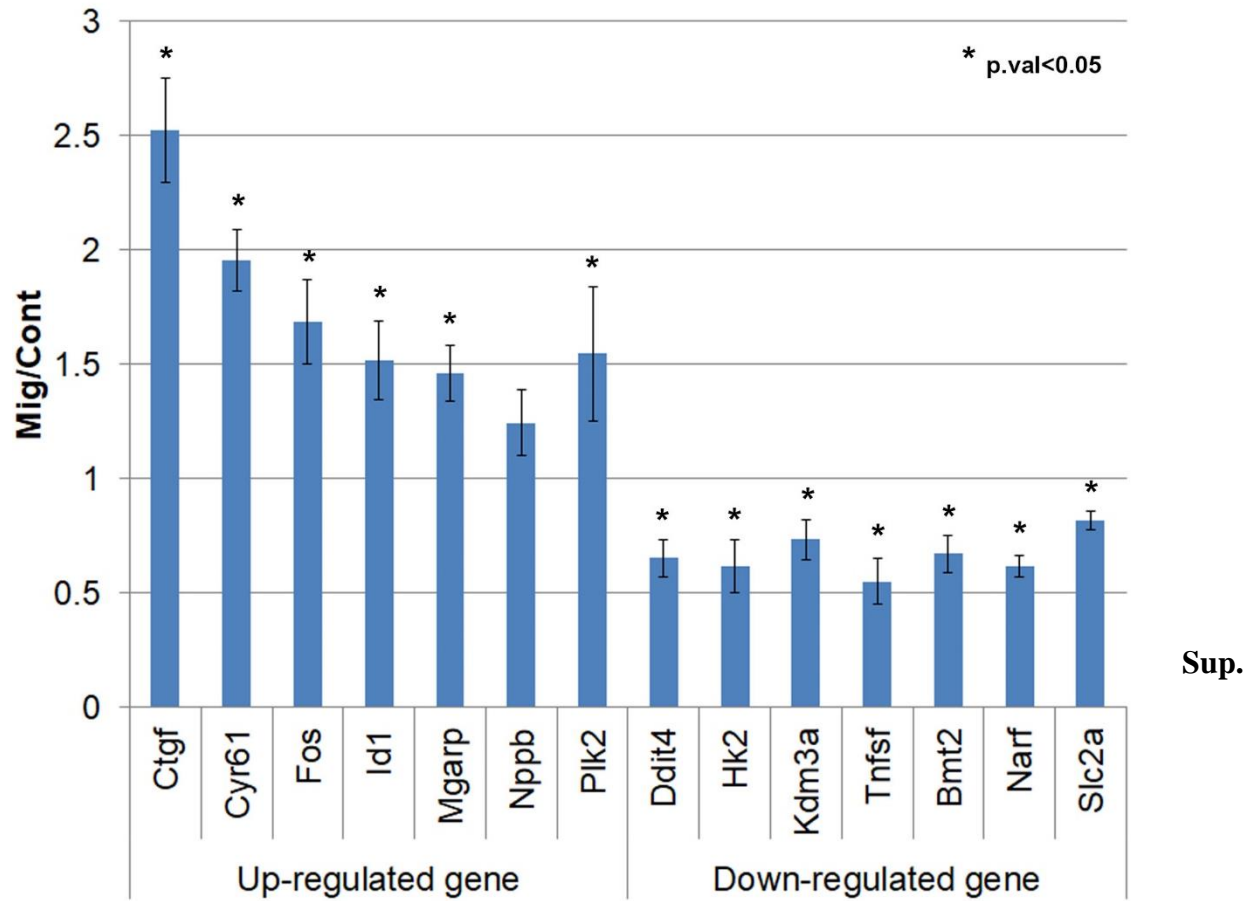
were calculated from reads coverage of consecutively equally sized 10kb bins for enhancers (a) and RefSeq genes (b).

**Figure S4**

|                          |               | #<br>sequenced<br>reads | # mapped<br>reads | # unique<br>mapped<br>reads |
|--------------------------|---------------|-------------------------|-------------------|-----------------------------|
| <b>Cont.</b>             | <b>Rep. 1</b> | $3.80 \cdot 10^7$       | 95.12%            | 94.06%                      |
|                          | <b>Rep. 2</b> | $5.50 \cdot 10^7$       | 98.40%            | 92.91%                      |
|                          | <b>Rep. 3</b> | $3.75 \cdot 10^7$       | 98.18%            | 92.99%                      |
|                          | <b>Rep. 4</b> | $2.66 \cdot 10^7$       | 94.86%            | 93.09%                      |
|                          | <b>Rep. 5</b> | $2.35 \cdot 10^7$       | 94.70%            | 93.10%                      |
| <b>Mig.</b>              | <b>Rep. 1</b> | $3.08 \cdot 10^7$       | 97.66%            | 94.00%                      |
|                          | <b>Rep. 2</b> | $4.93 \cdot 10^7$       | 97.82%            | 92.84%                      |
|                          | <b>Rep. 3</b> | $3.25 \cdot 10^7$       | 98.26%            | 92.72%                      |
|                          | <b>Rep. 4</b> | $2.81 \cdot 10^7$       | 95.43%            | 92.80%                      |
|                          | <b>Rep. 5</b> | $3.01 \cdot 10^7$       | 95.18%            | 92.81%                      |
| <b>Mig. +<br/>GSK343</b> | <b>Rep 1</b>  | $2.32 \cdot 10^7$       | 95.13%            | 92.57%                      |
|                          | <b>Rep 2</b>  | $2.57 \cdot 10^7$       | 95.40%            | 93.16%                      |
|                          | <b>Rep 3</b>  | $2.65 \cdot 10^7$       | 94.56%            | 92.57%                      |
| <b>Cont. +<br/>DRB</b>   | <b>Rep 1</b>  | $2.43 \cdot 10^7$       | 98.53%            | 92.70%                      |
|                          | <b>Rep 2</b>  | $2.41 \cdot 10^7$       | 98.46%            | 92.07%                      |
|                          | <b>Rep 3</b>  | $2.21 \cdot 10^7$       | 98.45%            | 92.66%                      |
| <b>Mig. +<br/>DRB</b>    | <b>Rep 1</b>  | $2.21 \cdot 10^7$       | 95.71%            | 92.02%                      |
|                          | <b>Rep 2</b>  | $2.29 \cdot 10^7$       | 98.57%            | 91.23%                      |
|                          | <b>Rep 3</b>  | $2.47 \cdot 10^7$       | 98.41%            | 91.64%                      |

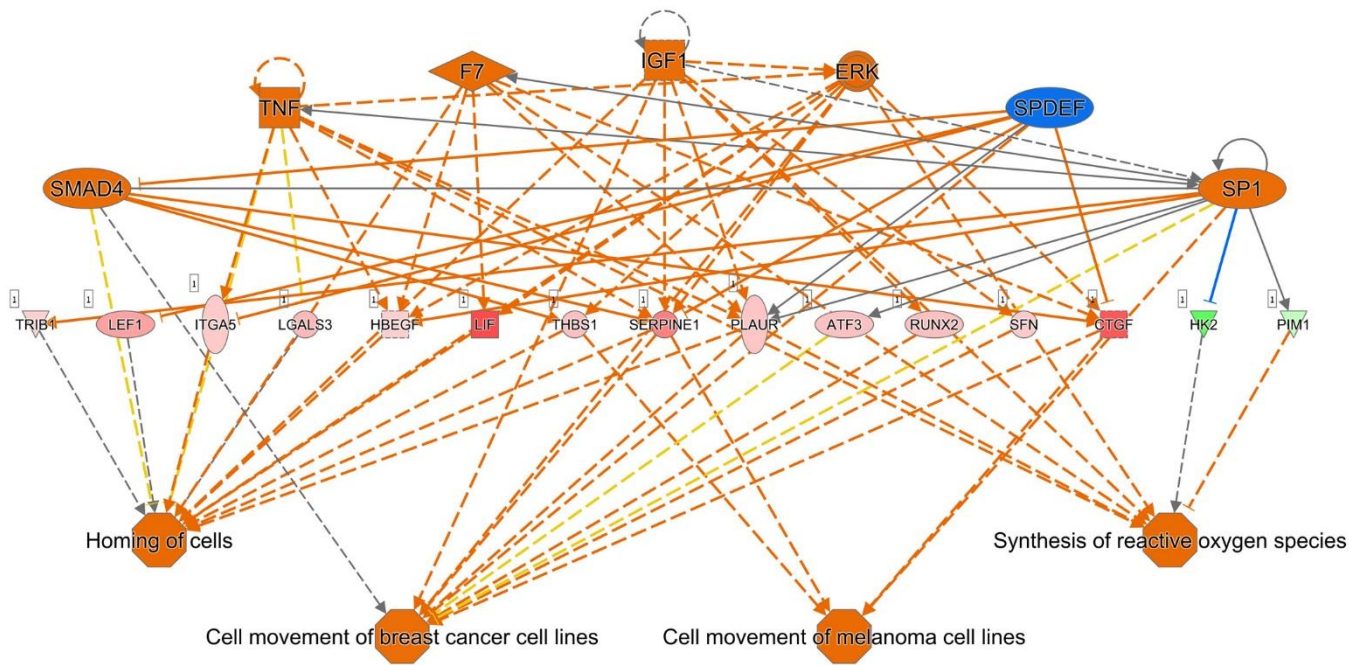
**Figure S4: Transcriptome profiling of control and migrating cells using RNA-seq.** RNA-seq statistics observed in control cells (Cont.), migrating cells (Mig.), GSK343-treated cells and DRB-treated cells. The percentage of mapped reads was calculated out of the number of sequenced reads. The percentage of unique mapped reads was calculated out of the total mapped reads.

**Figure S5**



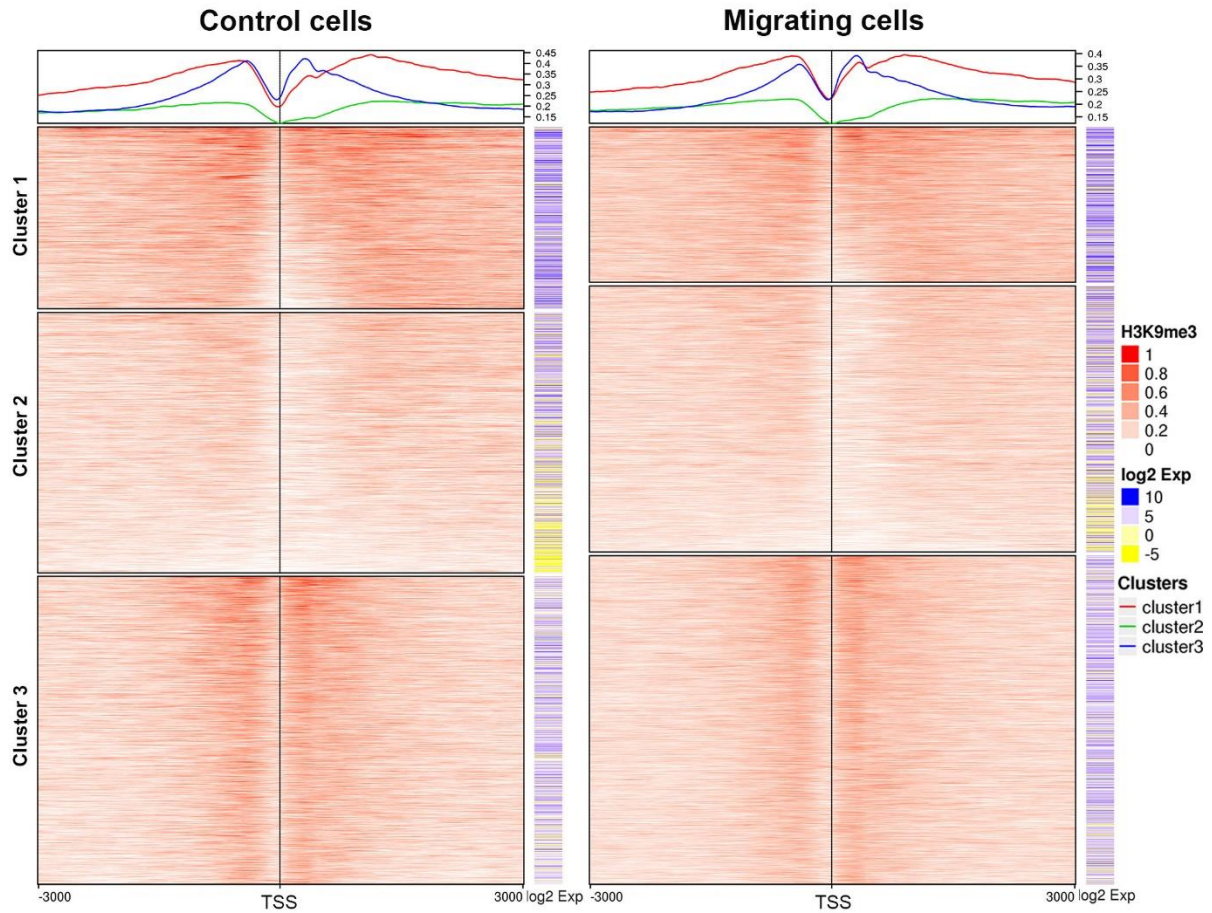
**Figure S5: Validation of RNA-seq based transcriptome profiling in control and migrating cells by RT-qPCR.** Relative RNA levels of 7 migration up-regulated genes and 7 migration down regulated genes in the RNA-seq analysis were determined by RT-qPCR. The bars represent the average ratio between the RNA levels in migrating cells and the RNA levels in control cells in 3 repetitions  $\pm$  SE. Statistical significance was calculated by student's T-test, \*  $p < 0.05$ .

**Figure S6**



**Figure S6: Predicted regulatory network effects identified from differential gene expression profile upon induction of migration.** Upstream regulators predicted to be activated (orange) or inhibited (blue) appear in the upper layer. The prediction is based on the migration-induced differentially-expressed genes that appear in the middle layer (red represents upregulation and green represents downregulation). The shape of the nodes reflects the functional class of each gene product: transcription regulator (vertical ellipse), cytokine/growth factor (square), and complex/group/other (circle). In the lower layer, the expected phenotypic consequences of changes in gene expression are shown by considering the Ingenuity Knowledge Base (absolute z-score > 2 and P-value < 0.05). The orange octagonal symbols represent activation predicted relationships., Solid and dashed lines between genes represent known direct and indirect gene interactions, respectively. Orange (predicted to be activated) lines represent relationships with causal consistency, yellow for inconsistency, and gray for non-predicted effect.

**Figure S7**

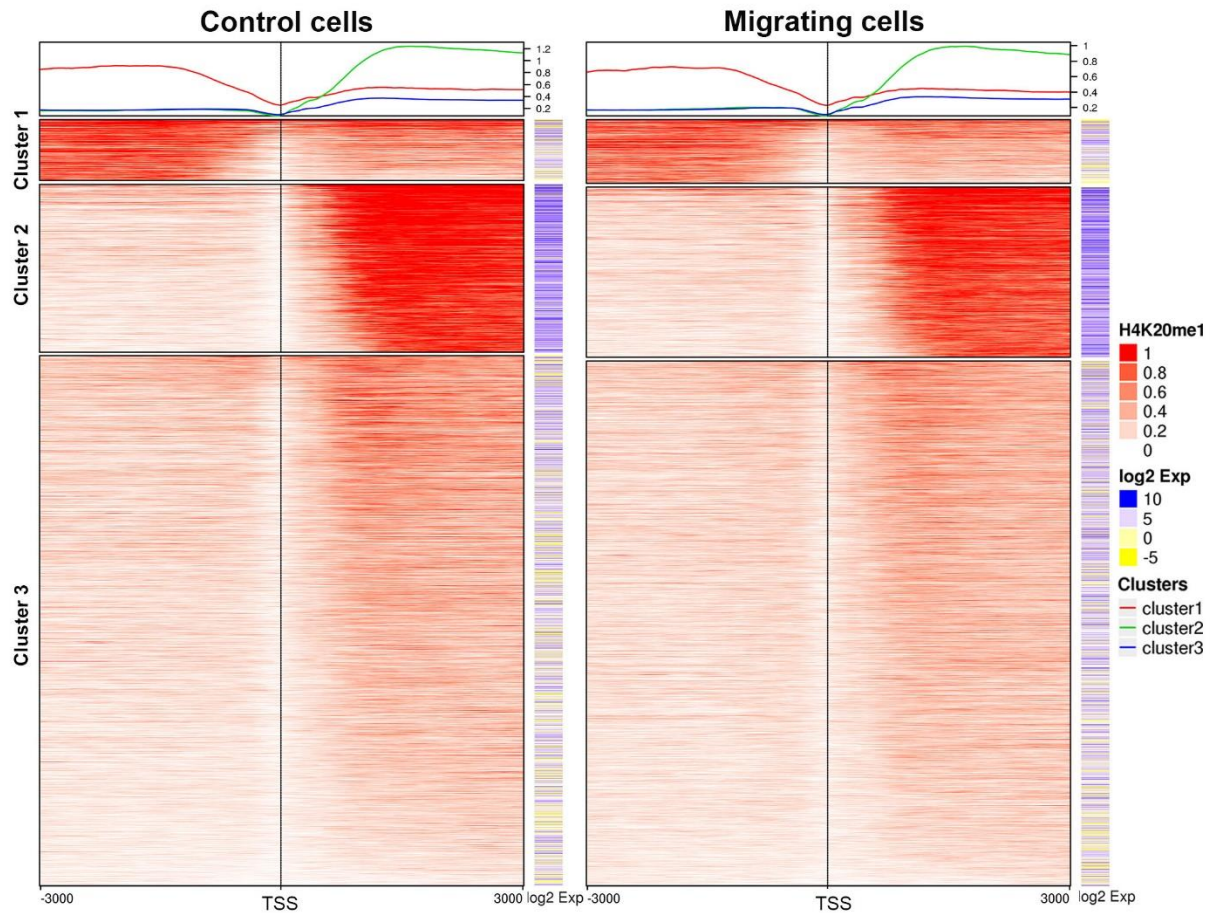


**Figure S7: Association between H3K9me3 signal around the TSS and gene expression.**

ChIP-seq score of H3K9me3 around the TSS of H3K9me3-enriched genes  $\pm 3000$ bp was fetched and the average score across every 20 nucleotides was calculated and assembled into a matrix for creating the heatmap. The analysis was done for both control (left) and migrating cells (right). Each row in the heatmap represents a gene ( $n = 13,226$  in control cells,  $n = 10,554$  in migrating cells). Genes were clustered according to their signal intensity using KNN clustering method ( $k=3$ ). The expression level of every gene is presented next to its corresponding ChIP-seq signal line. The average ChIP-seq score distribution across genes belong to each cluster is presented above the heatmap.



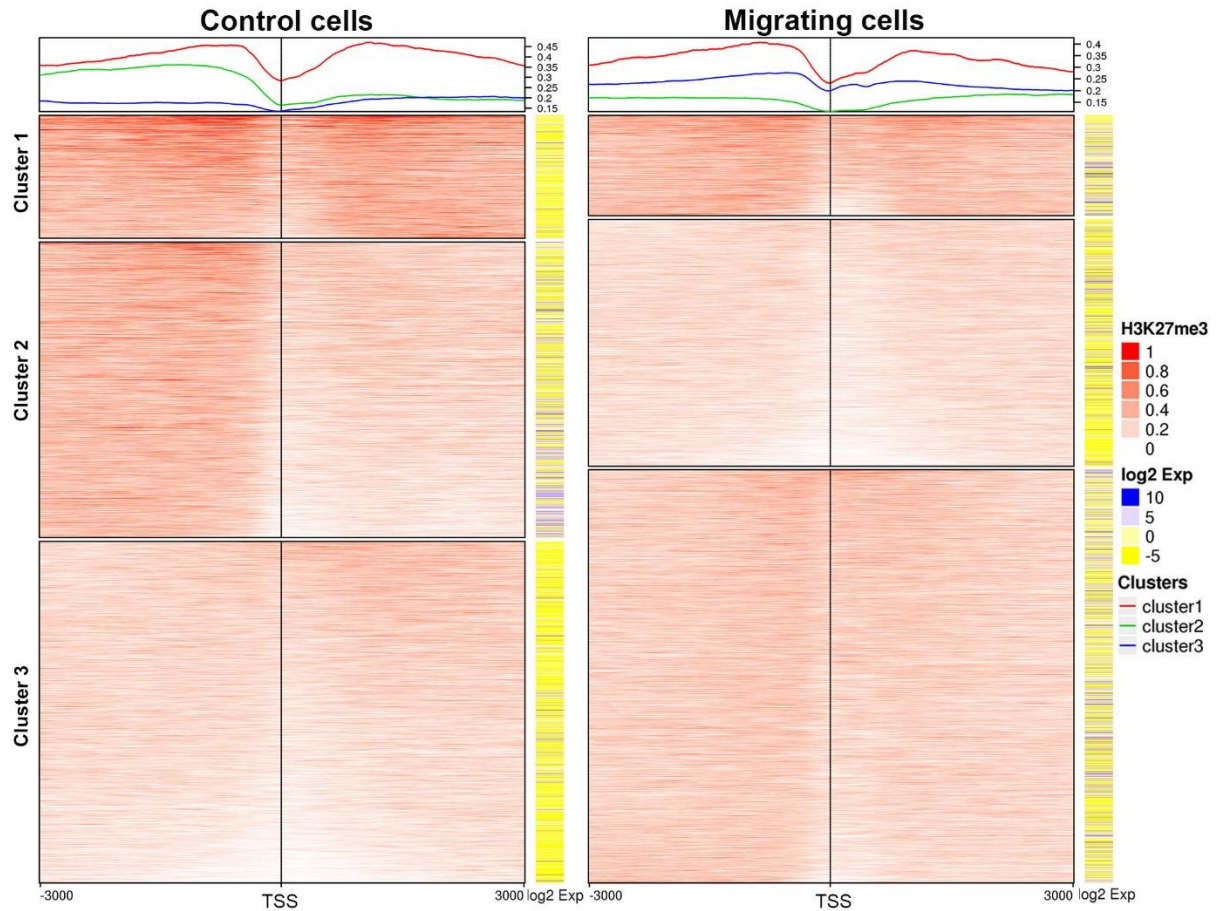
**Figure S8**



**Figure S8: Association between H4K20me1 signal around the TSS and gene expression.**

ChIP-seq score of H4K20me1 around the TSS of H4K20me1-enriched genes  $\pm 3000$ bp was fetched and the average score across every 20 nucleotides was calculated and assembled into a matrix for creating the heatmap. The analysis was done for both control (left) and migrating cells (right). Each row in the heatmap represents a gene ( $n = 18,056$  in control cells,  $n = 15,865$  in migrating cells). Genes were clustered according to their signal intensity using KNN clustering method ( $k=3$ ). The expression level of every gene is presented next to its corresponding ChIP-seq signal line. The average ChIP-seq score distribution across genes belong to each cluster is presented above the heatmap.

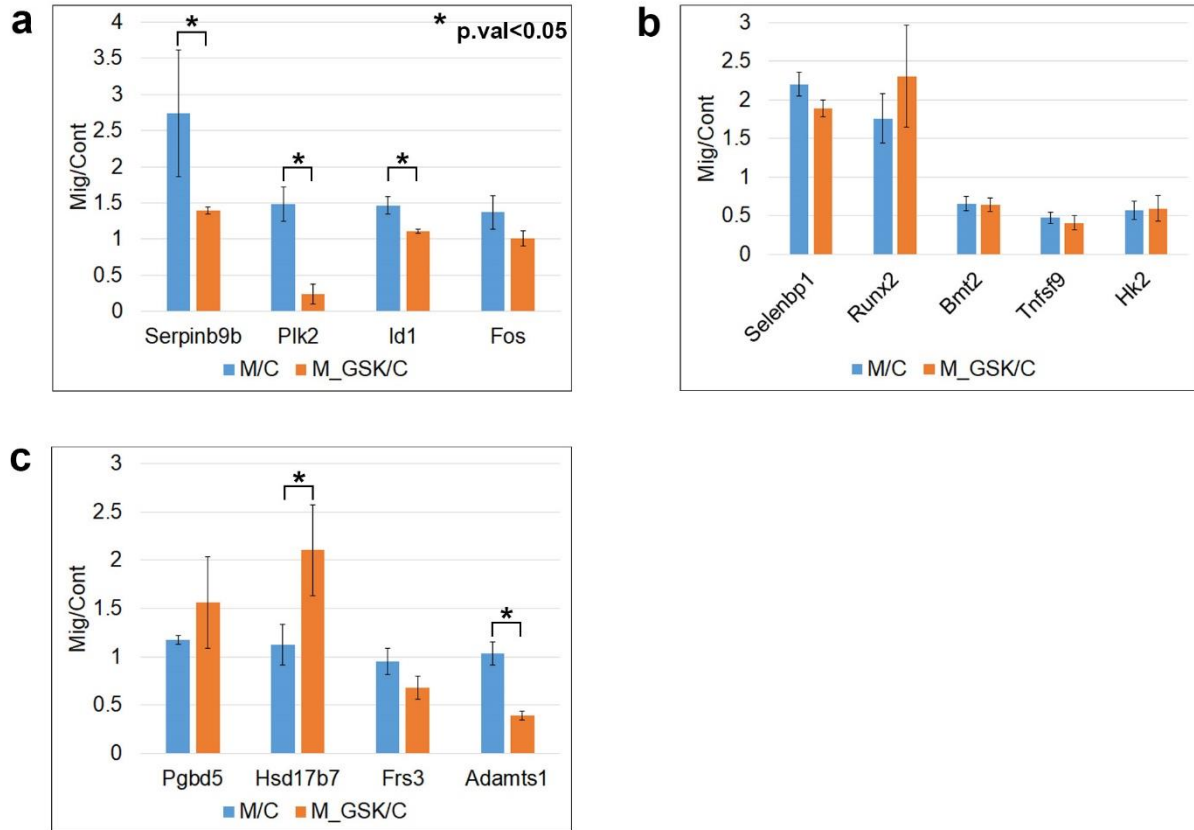
**Figure S9**



**Figure S9: Association between H3K27me3 signal around the TSS and gene expression.**

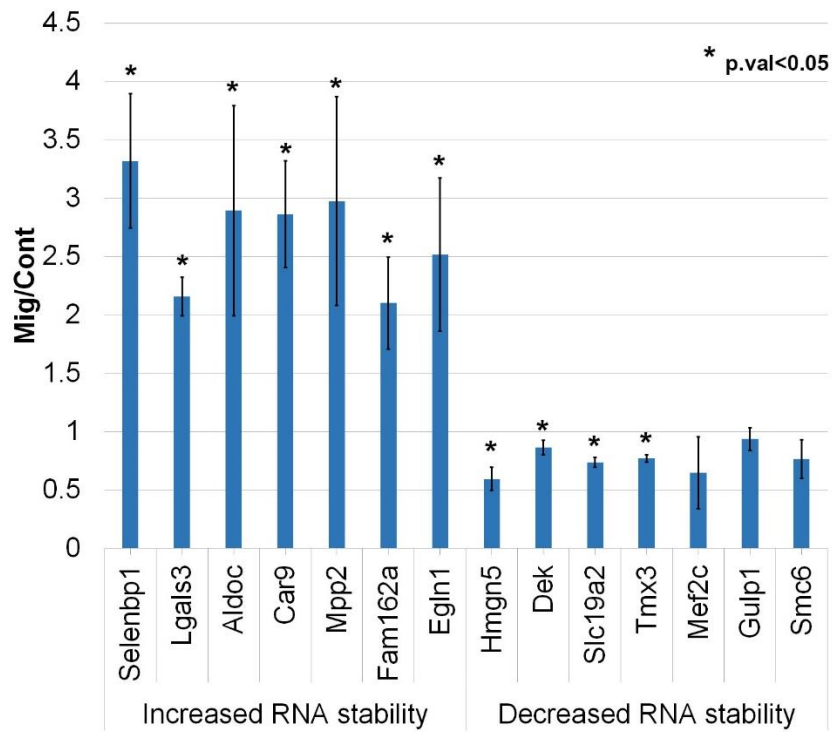
ChIP-seq score of H3K27me3 around the TSS of H3K27me3-enriched genes  $\pm 3000$ bp was fetched and the average score across every 20 nucleotides was calculated and assembled into a matrix for creating the heatmap. The analysis was done for both control (left) and migrating cells (right). Each row in the heatmap represents a gene ( $n = 9,713$  in control cells,  $n = 7,633$  in migrating cells). Genes were clustered according to their signal intensity using KNN clustering method ( $k=3$ ). The expression level of every gene is presented next to its corresponding ChIP-seq signal line. The average ChIP-seq score distribution across genes belong to each cluster is presented above the heatmap.

**Figure S10**



**Figure S10: Validation of RNA-seq based transcriptome profiling in migrating cells treated with GSK343 by RT-qPCR.** Relative RNA levels of K27me-dependent genes (a), K27me-independent genes (b) and K27me-buffered genes (c) in the RNA-seq analysis were determined by RT-qPCR. The blue bars and the orange bars represent the average ratio  $\pm$  SE between the RNA levels in migrating cells and those in control cells or between the RNA levels in migrating cells treated with GSK343 and those in control cells, respectively.  $n=3$ , statistical significance was calculated by student's T-test, \*  $p<0.05$ .

**Figure S11**



**Figure S11: Validation of RNA-seq based RNA stability profiling in control and migrating cells by RT-qPCR.** 7 genes with migration-increased RNA stability levels and 7 genes with migration-reduced RNA stability levels in the RNA-seq analysis were detected by RT-qPCR in control and migrating cells treated with DRB. The bars represent the average ratio between the RNA levels in migrating cells and the RNA levels in control cells in 3 repetitions  $\pm$  SE. Statistical significance was calculated by student's T-test, \*  $p < 0.05$ .



## **Legends to Supplementary Tables**

**Sup. Table 1: Altered gene expression upon induction of migration.** Differentially expressed genes between migrating cells and control cells with fold change  $>1.3$  and  $FDR < 0.05$ .

**Sup. Table 2: The dependence of migration-induced transcriptional changes on H3K27 methylation.** Differentially expressed genes between migrating cells and control that did not change between migrating cells treated with GSK343 and control cells were termed "H3K27me-dependent genes". Differentially expressed genes between migrating cells and control cells and also between migrating cells treated with GSK343 and control cells were termed "H3K27me-independent genes". Genes that were not changed between migrating cells and control cells but differentially expressed between migrating cells treated with GSK343 and control cells were termed "H3K27me-buffered genes". Genes with fold change  $>1.3$  and a  $FDR < 0.05$  are listed.

**Sup. Table 3: Migration induced changes in RNA stability.** Differentially expressed genes between DRB-treated migrating cells and DRB-treated control cells (fold change  $>1.3$  and  $FDR < 0.05$ ).

## **Supplementary Scripts**

**Script 1: Script to generate heatmaps presented in supplementary data.**