Supplemental information

Development of an *in vitro* genotoxicity assay

to detect retroviral vector-induced

lymphoid insertional mutants

Antonella L. Bastone, Violetta Dziadek, Philipp John-Neek, Friederike Mansel, Jenni Fleischauer, Eric Agyeman-Duah, Dirk Schaudien, Oliver Dittrich-Breiholz, Adrian Schwarzer, Axel Schambach, and Michael Rothe
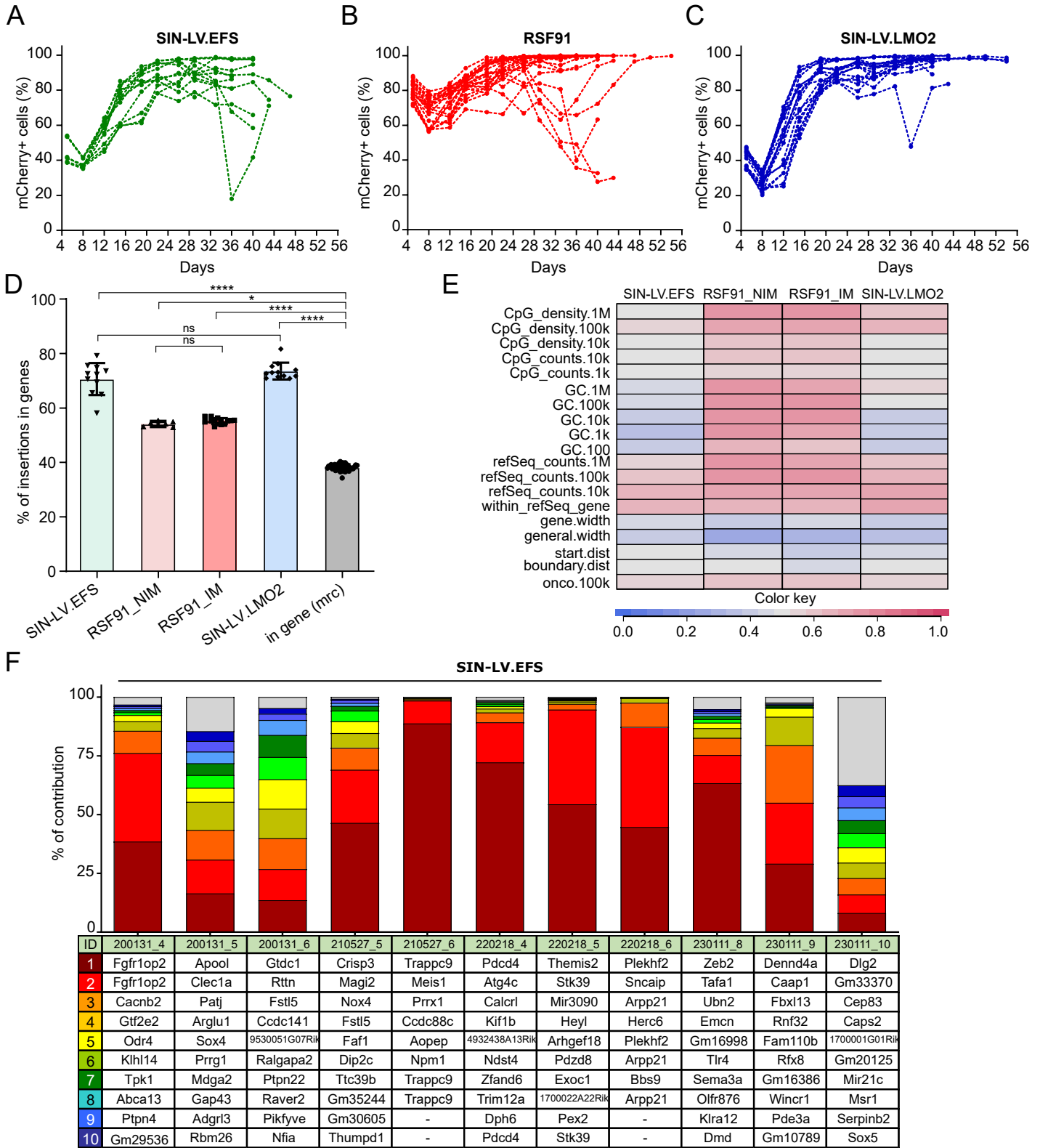
Figure S1: ISA of transduced samples. (A), **(B)**, and **(C)**: percentage of transduced cells over time determined by mCherry expression by flow cytometry. Each line represents an individual replicate. **(D)** Percentage of insertions within genes for each vector construct, measured in early time point samples (before day 20 of culture). The percentages follow the expected insertional preferences for the different vector types. Matched random control (mrc) sites generated in silico by the pipeline were used for comparison. Statistical analysis was performed with Kruskal-Wallis with Dunn′s post-hoc test and Bonferroni-Holm for multiple testing. *p<0.05; ****p<0.0001; ns = not significant. Indicated is the mean ± SD. **(E)** Genomic features associated with the insertions of transduced samples. The latest time point available for each sample were clustered together according to the vector used and development/absence of the DN2 block. As described in Berry et al. 2017, the colored tiles represent the direction and magnitude of deviation of the integration site dataset from the mrc distribution for each genomic feature, quantified using the receiver operating characteristic (ROC) curve area method. In the color key, ROC areas between 0.0-0.5 suggest the genomic feature was disfavored in the insertion site dataset compared to mrc while ROC values between 0.5 and 1.0 indicate enrichment in the dataset. **(F)** Top 10 insertion sites of SIN-LV.EFS-transduced samples, depicted in color. All remaining integrations were grouped in the gray bar.
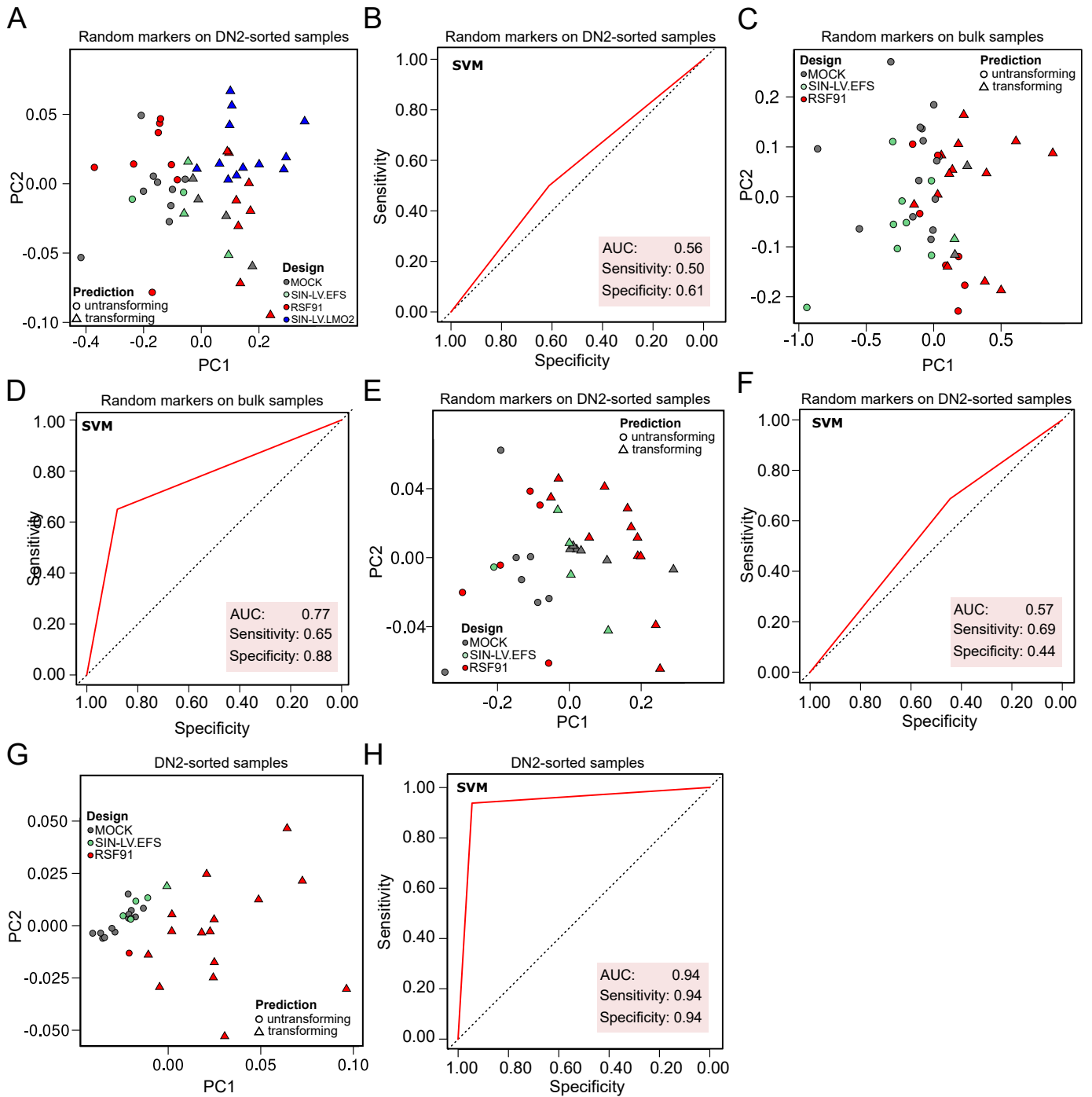
Figure S2: Separation of classes using random markers. PCA **(A)** and ROC analysis **(B)** of all DN2-sorted samples using randomly selected markers. **(C)** and **(D)** illustrate the PCA **(C)** and ROC analysis in **(D)** of bulk culture samples using random markers. In **(E)** and **(F)**, the same analysis with randomly selected markers on DN2 cells as shown in **(A)** and **(B)** was repeated excluding the SIN-LV.LMO2 group to deduct the influence of the class imbalance within the DN2-sorted samples (more transforming than untransforming). **(G)** and **(H)** illustrate the PCA **(G)** and ROC analysis **(H)** using the DN2 features on DN2-sorted samples excluding SIN-LV.LMO2, to validate the approach used in **(E)** and **(F)**. In this case, the exclusion of SIN-LV.LMO2 does not change the model performance and overall classification of DN2-sorted samples.
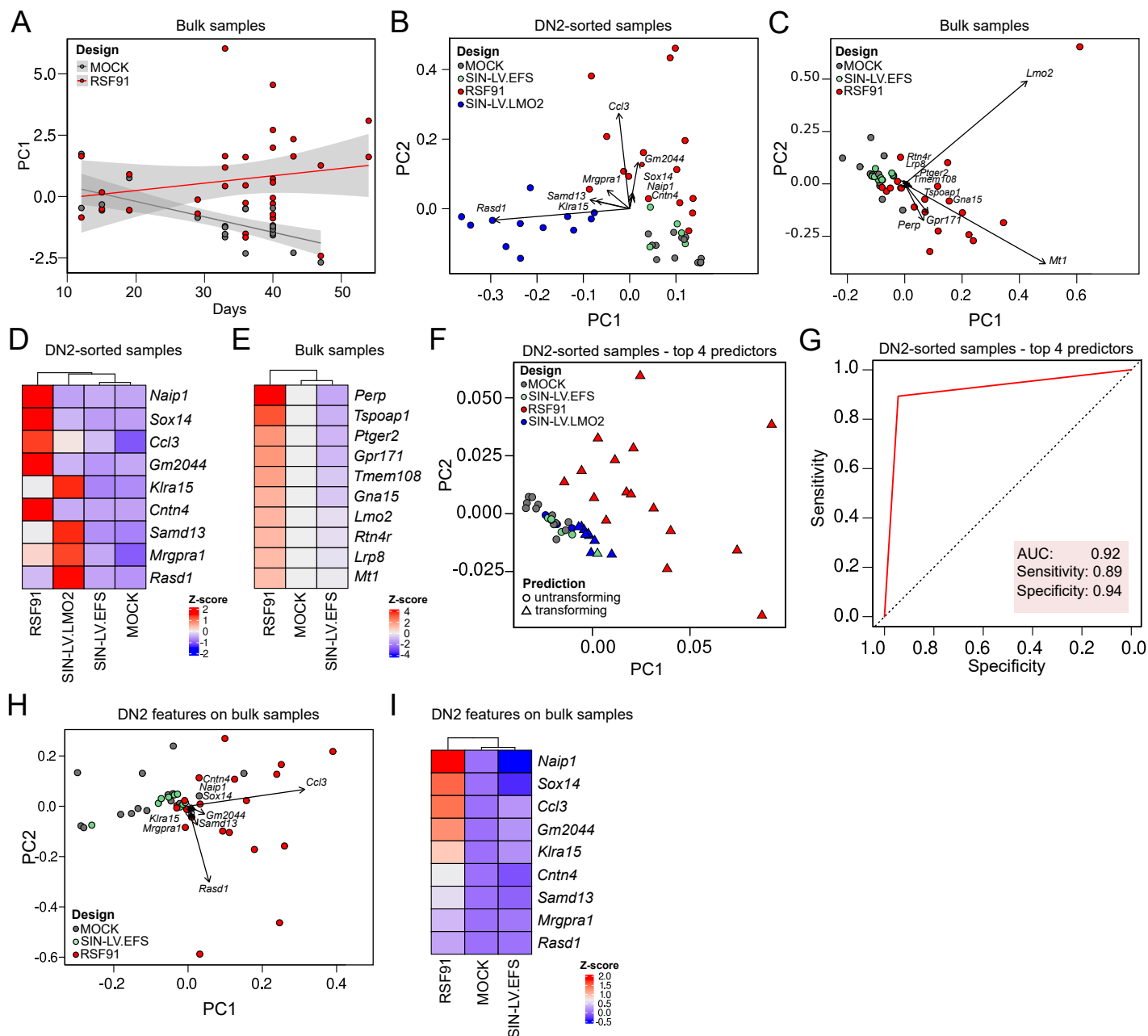
Figure S3: Mutagenic vectors induced specific gene expression changes in lymphoid samples. **(A)** PC1 of bulk samples taken at different time points using the bulk feature set. Starting from a relative close origin, mock and RSF91-transduced samples diverge over time following a linear tendency, based on their expression of the selected bulk features. The filled lines indicate the running median over time, while the dotted lines correspond to the first and third quantiles. **(B)** and **(C)** show the contribution of each of the selected genes to PC1 and PC2 when classifying the DN2-sorted samples with DN2 predictors **(B)** and the bulk cultures with bulk predictors **(C)**. The length of the arrows indicates relative intensity of the gene´s influence and the arrow orientation correlates with the contribution to each of the principle components as well as the correlation between each gene. **(D)** and **(E)**: heatmaps showing the expression of each of the selected predictors accross the different groups. **(D)**: expression of DN2 predictors on DN2-sorted samples. **(E)**: expression of bulk predictors on bulk culture samples. The row z-score scale indicates the degree of deviation from the mean expression for each gene. **(F)** SVM prediction of DN2-sorted samples using the top 4 most frequently found DN2 predictors (*Naip1*, *Sox14*, *Ccl3*, *Gm2044*). **(G)** ROC curve for the analysis in **(F)**. **(H)** contribution of the DN2 features to PC1 and PC2 when classifying the bulk cultures. **(I)** Expression of the DN2 predictors in the different vector groups on bulk cultures.
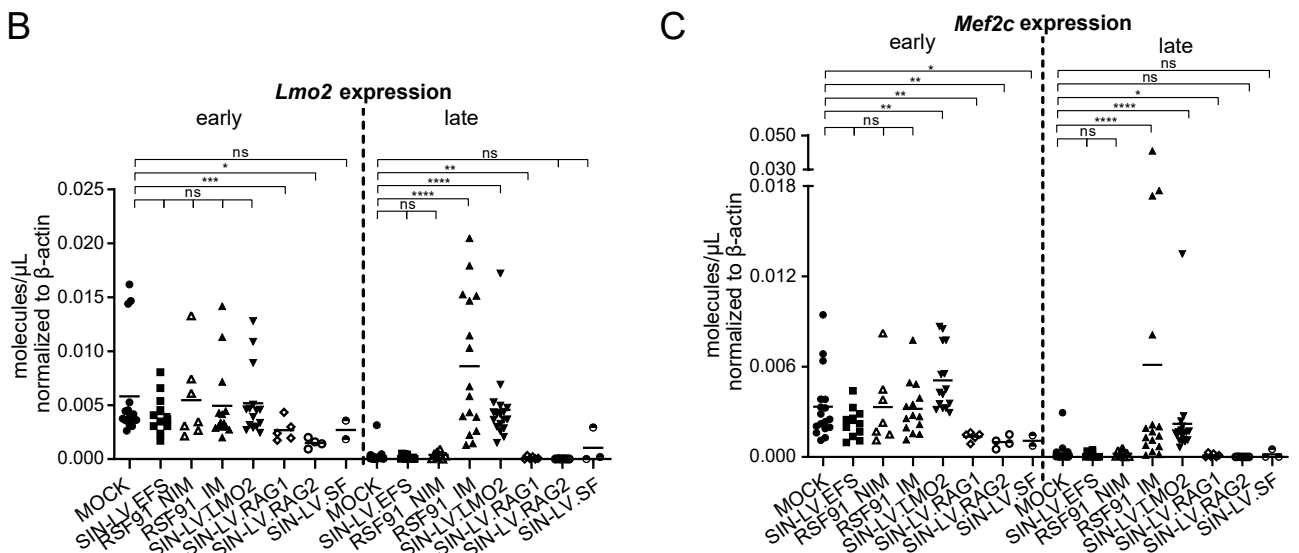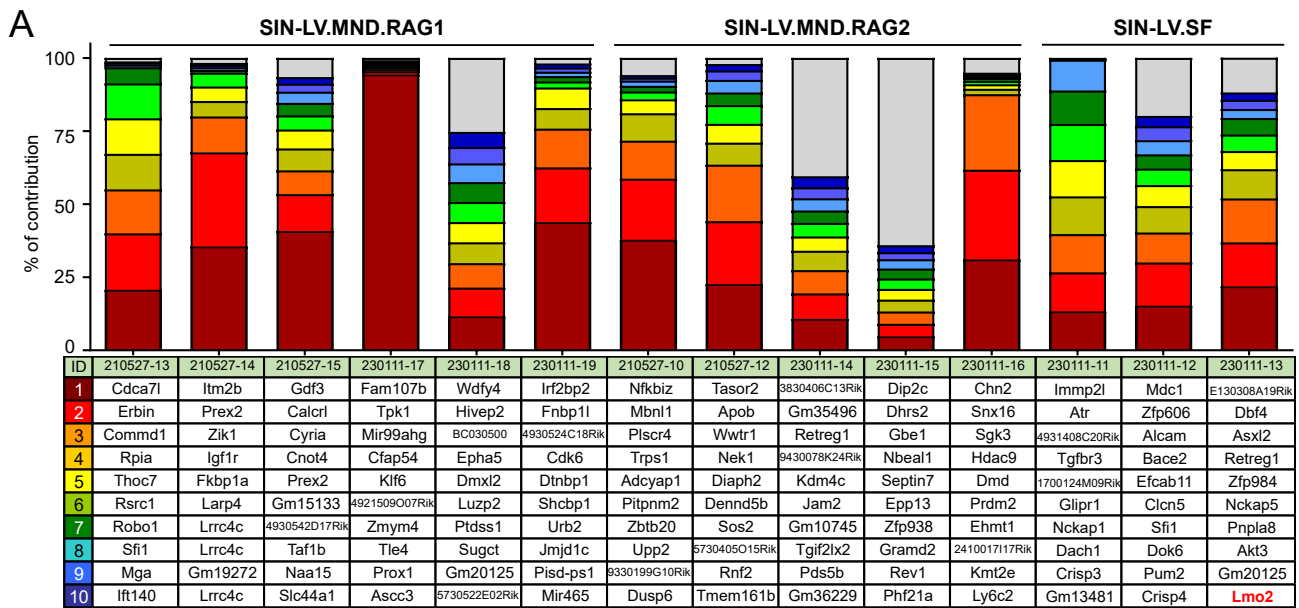
Figure S4: Phenotype and ISA of additional vectors in the lymphoid assay. (A) Schematic representation of the additional SIN-LV tested, with the internal SFFV promoter driving the expression of the fluorescent protein eGFP, or the internal MND promoter (viral origin) driving the expression of either RAG1 or RAG2. **(B)** Comparison of the DN2 percentages at the end of culture for all vector types tested, measured by flow cytometry. **(C)** Cell proliferation at the end of culture expressed as expansion factor per day for all vector types tested. The total cell number on the last day of culture of each sample was divided by the number of seeded cells in the previous time point multiplied by the timespan between counts. **(D)** Cohens d values reflecting the effect size between the mean expansion factors shown in **(C)**. **(E)** VCN by ddPCR of all samples that yielded a successful INSPIIRED reaction, using the latest time point available. Three samples were excluded due to very low VCN values not matching the percentage of mCherry+ cells detected by flow cytometry (190614-8, 210527-5, 210527-6). For the samples highlighted in red, less than 10 ng were used for VCN determination, which could have affected the quality of the result. Further details on samples included can be found in methods and in Supplementary Table S1. **(F)** Number of unique insertion sites obtained after alignment to the mouse genome. **(G)**, **(H)**, and **(I)**: sequencing statistics calculated with the INSPIIRED pipeline for the overall clonality assessment of the samples. A low UC50 **(G)**, low Shannon **(H)** and high Gini **(I)** indices correlate with an oligoclonal insertion site pattern. Early time points refer to samples taken between days 12-15, and late time points, to samples taken after day 30. In all cases, statistical analysis was performed with Kruskal-Wallis with Dunn´s post-hoc test and Bonferroni-Holm correction for multiple testing to account for the difference in variance between the groups. Each data point of the scattered plots represent an individual sample. **p<0.01; ****p<0.0001; ns = not significant. eGFP: enhanced green fluorescent protein.

4

Figure S5: ISA and expression of relevant T cell proto-oncogenes in samples transduced with the additional SIN-LV vectors. (A) Top 10 insertion sites (depicted in colors) of SIN-LV.MND.RAG1/RAG2- and SIN-LV.SF-transduced samples. The remaining insertions of each sample were grouped together in the grey bar. **(B)** and **(C)**: *Lmo2* and *Mef2c* expression determined by ddPCR. For late time point measurements of SIN-LV.MND.RAG2, both transcription factors were downregulated compared to mock (*Lmo2*: SIN-LV.MND.RAG2$_{median}$ = 1.12×10^{-5}, mock$_{median}$ = 8.27×10^{-5}; *Mef2c*: SIN-LV.MND.RAG2$_{median}$ = 0.00, mock$_{median}$ = 3.74×10^{-5}). Early time points refer to samples taken between days 12-15, and late time points, to samples taken after day 30. Statistical analysis was performed with Kruskal-Wallis with Dunn´s post-hoc test and Bonferroni-Holm correction for multiple testing to account for the difference in variance between the groups. Each data point of the scattered plots represents an individual sample. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; ns = not significant.
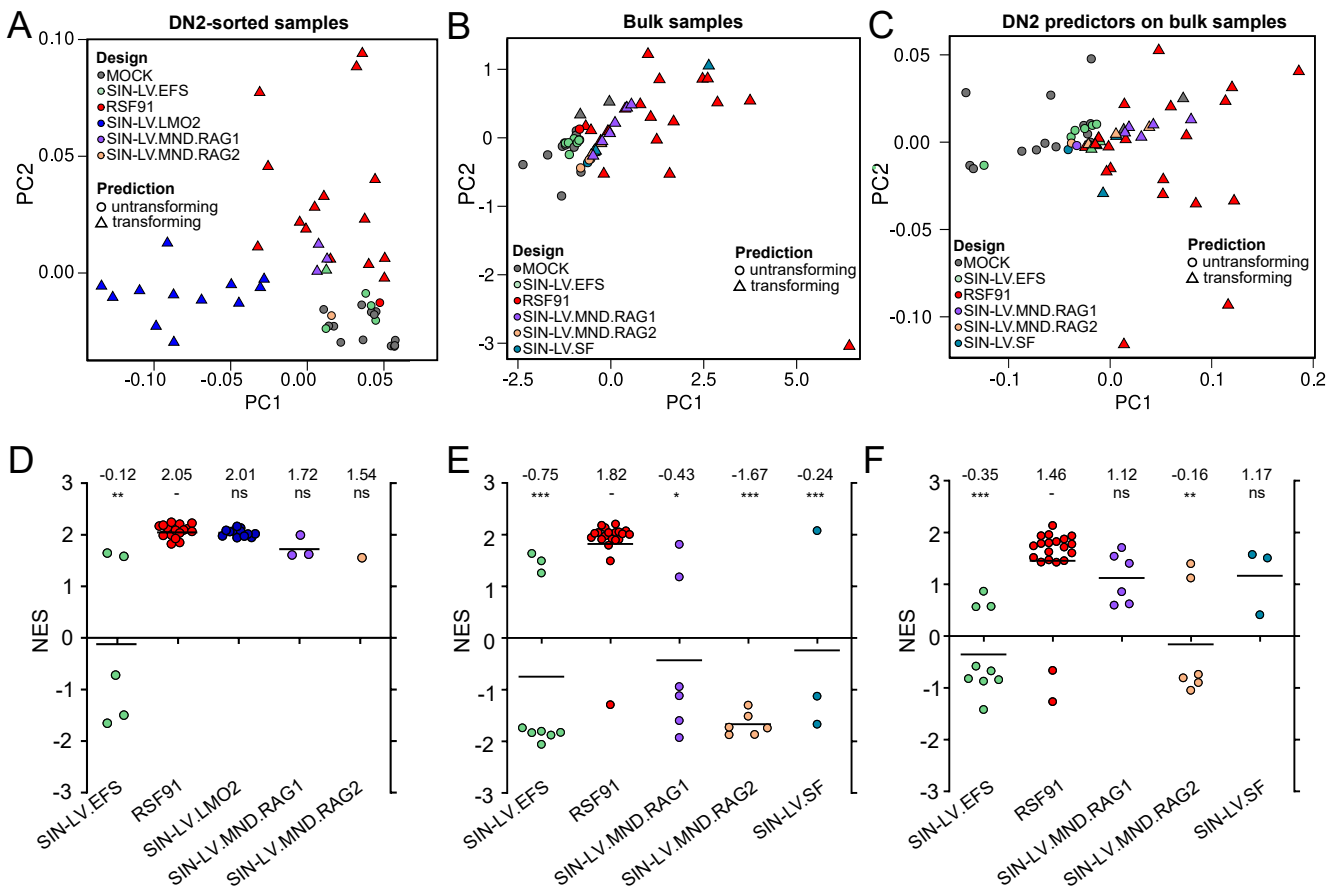
Figure S6: Genotoxicity prediction of additional SIN-LV vectors using SAGA-XL. (A), **(B)**, and **(C)** illustrate the PCA analysis of DN2-sorted samples with DN2 features **(A)**, bulk samples with bulk features **(B)**, and bulk samples with the genotoxicity predictors found in the DN2-sorted population **(C)**. **(D)**, **(E)**, and **(F)** display the NES values obtained with SAGA-XL-GSEA using the DN2 features on DN2-sorted samples **(D)**, the bulk features on bulk samples **(E)**, and the DN2 features on bulk samples **(F)**. The mean NES values for each vector type are depicted above the graphs. Statistical comparison vs RSF91 is indicated below the correspondent mean NES and was performed with Kruskal-Wallis with Dunn´s post hoc test and Bonferroni-Holm correction. *p<0.05; **p<0.01; ***p<0.001; ns = not significant.
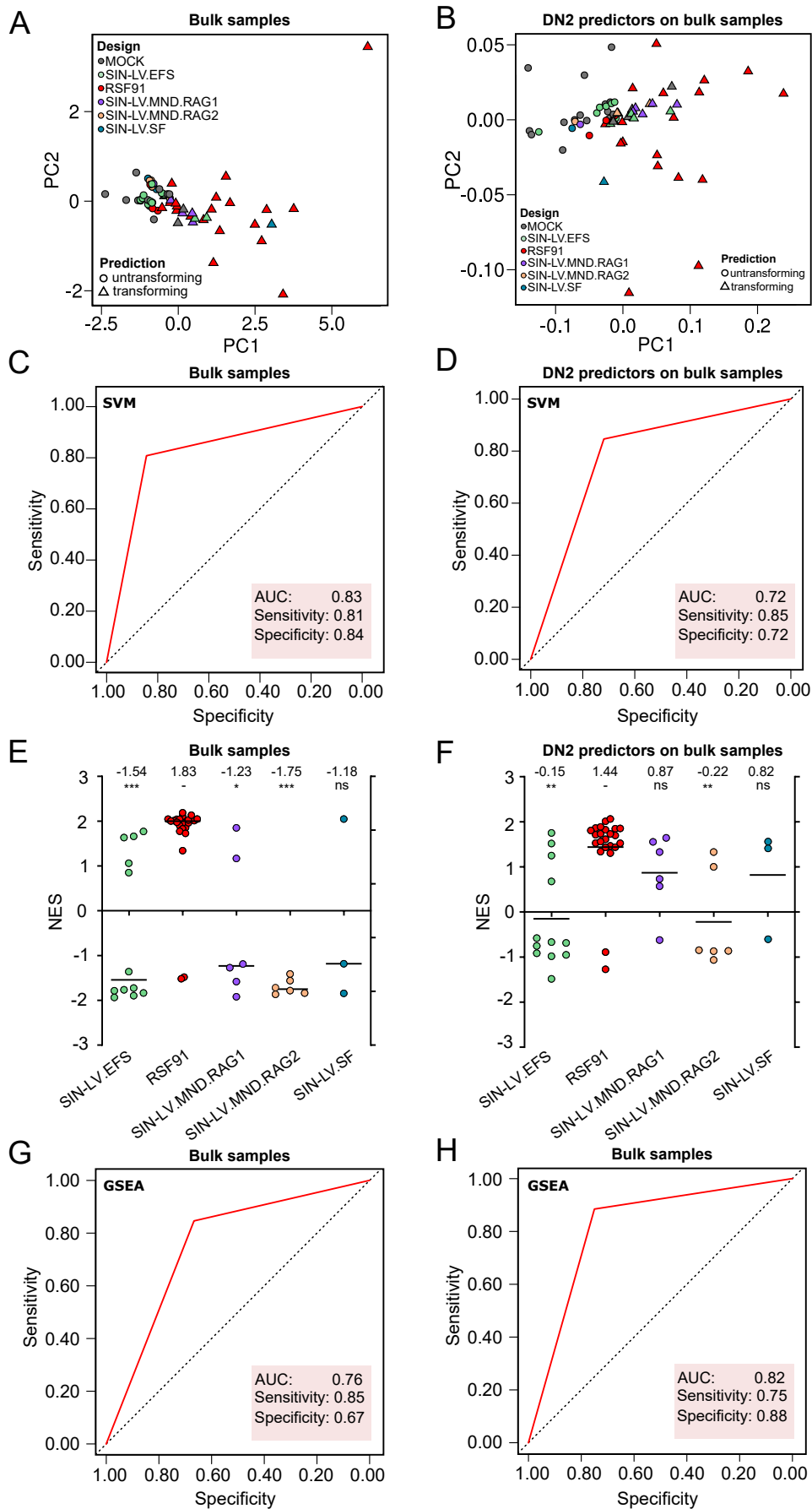
Figure S7: Validation of SAGA-XL. All available bulk culture samples, including some mock and RSF91-transduced samples unseen by the pipeline, were analysed with SAGA-XL using the bulk predictors **(A)** or the DN2 features **(B)**. **(C)** and **(D)**: ROC curves for the prediction models from **(A)** and **(B)**, respectively. **(E)** and **(F)**: SAGA-XL-GSEA results using the bulk predictors **(E)** or the features found on the DN2 population **(F)**. The mean NES values for each vector type are depicted above the graphs. Statistical comparison vs RSF91 samples is indicated below the correspondent mean NES and was performed with Kruskal-Wallis with Dunn´s post hoc test and Bonferroni-Holm correction. **(G)** and **(H)**: ROC curves for the SAGA-XL-GSEA approach.
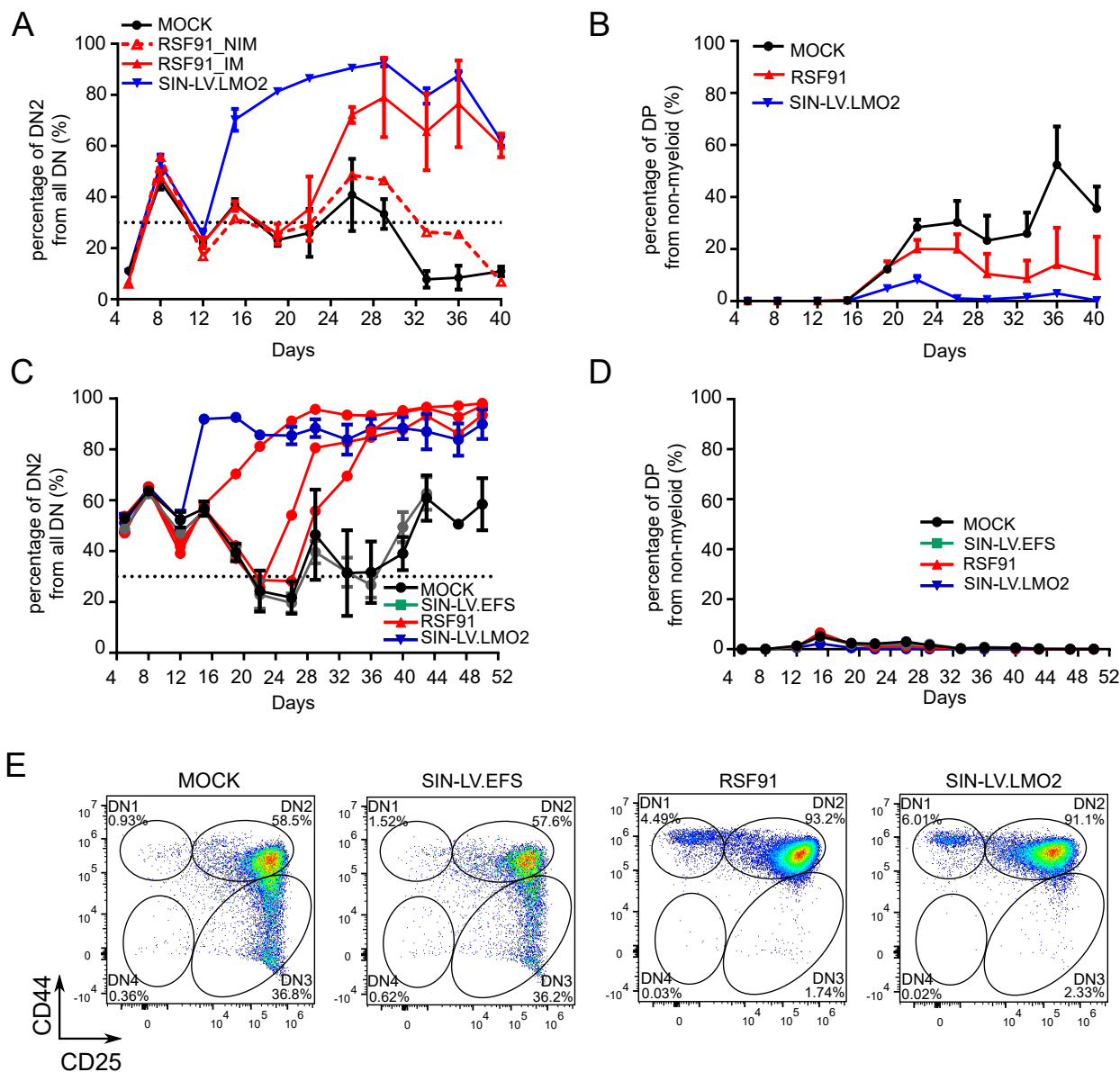
Figure S8: Interassay variability based on differentiation of mock samples. (A) and **(B)** show an example of a successful experiment in terms of T cell differentiation (n = 4). The DN2 population of mock cultures, followed by flow cytometry analysis, gradually decreased during long-term culture **(A)**, while DP T cells were generated in increasing amounts **(B)**. **(C)** and **(D)** depict experiment ID 200703, which failed to fulfill the inclusion criteria (n=3). **(C)** Kinetics of the DN2 population followed by flow cytometry. The DN2 population of mock cultures and SIN-LV.EFS-transduced samples increased over time at a similar pace than RSF91-transduced cultures. **(D)** Percentage of DP cells over time measured by flow cytometry analysis for the same samples depicted in **(C)**. No DP cells were generated for any of the samples. **(E)** Flow cytometry analysis of day 43 from the experiment depicted in **(C)** and **(D)** for mock and samples transduced with the different vectors. The high amount of DN2 cells by the end of culture for mock samples prevented a proper assessment of the DN2-block in the transduced samples. In all cases, the DN stages 1-4 were analyzed based on the CD44 and CD25 markers and the DP cells, based on CD4 and CD8 markers. Cells expressing the myeloid markers CD11b and Gr-1 were first excluded by gating. Data are presented as mean ± SD.
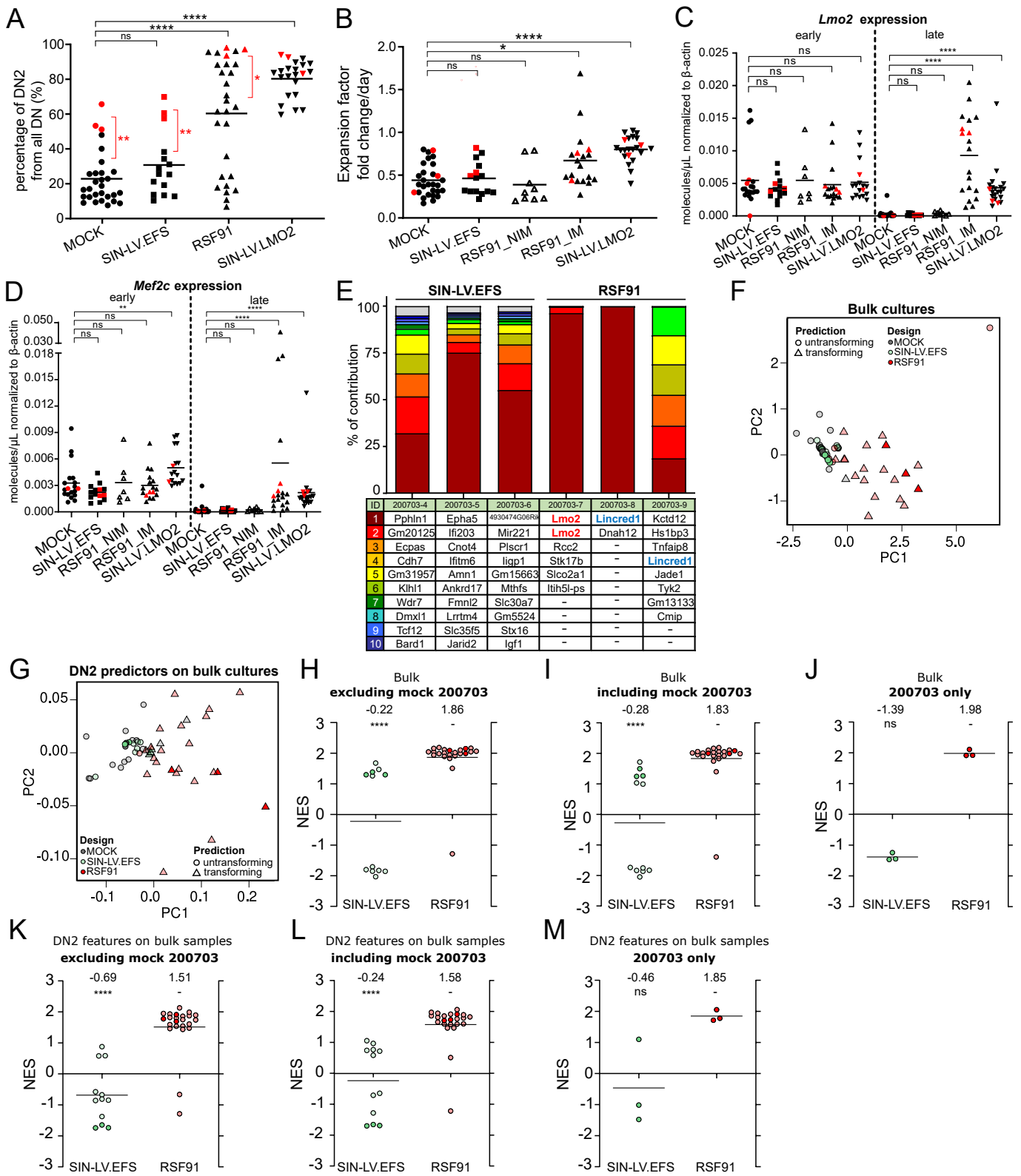
Figure S9: Influence of impaired mock differentiation on the assay readout. In **(A)**, **(B)**, **(C)** and **(D),** the samples from experiment ID 200703 with impaired mock differentiation presented in Supplementary Figure S8 are depicted in red and were pooled with previously analysed samples that passed the inclusion criteria (shown in main Figures 2-5) to see their influence on the overall data distribution within and between the groups. The significance level for the comparison within groups is shown in red and was performed with Mann-Whitney-U test. When not specified, the differences were not significant. For statistical comparison between groups, Kruskal-Wallis using Dunn's post-hoc and Bonferroni-Holm correction for multiple testing were performed. **(A)** Percentage of DN2 cells by the end of culture. **(B)** Expansion factor calculated as total cell number on last day of culture of each sample divided by number of seeded cells in previous time point multiplied by timespan between counts. **(C)** *Lmo2* and **(D)** *Mef2c* expression measured by ddPCR in early (before day 20) and the latest time point available. **(E)** ISA of experiment ID 200703. The top 10 insertions are depicted in colors and the remaining ones were grouped together in a grey bar. **(F)** and **(G)** show the SAGA-XL-SVM prediction for experiment ID 200703 (n_{mock}=2, $n_{SIN-LV.EFS}$=3, n_{RSF91}=3) depicted in bright colors, overlayed to the previously analysed samples (included in Figures 2-5) in dimmer colors, using the bulk predictors in bulk samples **(F)** or the DN2 features in bulk samples **(G)**. **(H)-(M)** show SAGA-XL-GSEA results using the bulk predictors on bulk samples **(H-J)** or DN2 features on bulk samples **(K-M)**. In all cases, the samples from experiment ID 200703 are depicted in bright colors and all previously analysed samples, in dimmer colors. To test the influence of the samples from experiment ID 200703 on the overall results, three strategies were evaluated: **(H)** and **(K)** excluded the mock from experiment ID 200703 from the calculations; **(I)** and **(L)** pooled the mock from ID 200703 with all previously measured mock; **(J)** and **(M)** only analysed samples from the 200703 batch. The mean NES values for each vector type are depicted above the graphs. Statistical comparison vs RSF91 in **(H)-(M)** is indicated below the correspondent mean NES and was performed with non-parametric Mann-Whitney-U test. Each data point of the scattered plots represent an individual sample. *p<0.05; **p<0.01; ****p<0.0001; ns = not significant.

Table S1: Sample information