Supplemental Data file

Supplemental data files (Excel):
The first set of 4 files (MACE-data) contains the gene entries with gene identifier, gene symbol, gene name, mean exp, mean mock, fold change log2, pvalue, -log10(pvalue), var-log2, chromosome number (with beginning and end), classification (into pseudogene (PG), non-annotated gene (NA), LINC RNA gene (LINC), MIR gene (MIR), SNO gene (SNO), MT gene (MT) and protein coding gene (PCG)). All data are given as in the Bioconductor output file with the exception of var-log2 which was simply calculated by ln(mean exp/mean mock)/ln2. The following files are provided:

1. Day 0 gene signature.xlsx
2. Day 3 gene signature.xlsx
3. Day 12 gene signature.xlsx
4. Day 28 gene signature.xlsx

The second set of 3 files (ATAQ-data) contain the gene entries with gene identifier, gene symbol, gene name, mean exp, mean mock, fold change log2, pvalue, var-log2, chromosome number (with beginning and end), classification (into pseudogene (PG), non-annotated gene (NA), LINC RNA gene (LINC), MIR gene (MIR), SNO gene (SNO) and protein coding gene (PCG)). All data are given as in the Bioconductor output file with the exception of var-log2 which was simply calculated by ln(mean exp/mean mock)/ln2. The following files are provided:

5. ATAC Day 0.xlsx
6. ATAC Day 3.xlsx
7. ATAC Day 28.xlsx
8. all-comparison_d3 vs d28_ATAQ

Methods

ATAC-Seq experiments
In particular, cells were grown in 6-well plate and treated with 1 µg/ml Doxycyclin for 48 hours. After that cells were harvested, and viability were checked in every samples. In all samples viability were more than 90%. Cells were then resuspended in cold PBS and cell numbers in each sample were counted. 50,000 cells were centrifuged at 500 RCF for 5 minutes at 4°C in a fixed angle centrifuge. After centrifugation, 900 µl of supernatant was aspirated with P1000 pipette and the remaining 100 µl of supernatant was carefully aspirated by pipetting with a P200 pipette tip to carefully avoid the cell pellet. Cells were resuspended in 50 µl cold ATAC-Seq Resuspension Buffer (RSB; 10 mM Tris-HCl pH 7.4, 10 mM NaCl, and 3 mM MgCl2 in water) containing 0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin, pipetting up and down 3 times. Cells were incubated on ice for 3 minutes and then 1 ml of cold ATAC-Seq Resuspension Buffer (RSB) containing 0.1% Tween-20 but no NP-40 and Digitonin was added. Tubes were inverted 3 times to mix and then centrifuged at 500 RCF for 10 minutes at 4°C in a fixed angle centrifuge to pellet down the nuclei. Supernatant was removed with two pipetting steps, as described before, and nuclei were resuspended in 50 µl of transposition mix (25 µl 2x TD buffer (20 mM Tris-HCl pH 7.4, 10 mM MgCl2, 20 % Dimethyl Formamide in water), 2.5 µl transposase 26 (100 nM final), 16.5 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20, and 5 µl water) by pipetting up and down six times. Transposition reactions were incubated at 37°C for 30 min in a thermomixer with shaking at 1000 rpm. The final ATAC-Seq experiment was performed at GenXPro (https://genxpro.net/) that used an Illumina HiSeq for the analysis, and the resulting data were analysed by bioinformatics tools.
**Differential gene expression profiling by MACE-Seq and Bioinformatic analyses**

The chimeric genes were either expressed constitutively, or induced for 48h with 1 \( \mu \text{g/ml} \) Doxycycline and total RNA were isolated from transfected cell lines. After testing the correct expression of both transgenes, differential gene expression (DGE) profiles were obtained by MACE (Massive Analysis of cDNA Ends) - Seq experiments following the manufacturer protocol (GenXPro, Frankfurt, Germany). Resulting data from 3 biological replicates of the cell line at the 4 different time points (day 0, day 3, day 12 and day 28) were compared with 3 biological replicates of mock-transfected cells. The MACE-libraries were prepared at GenXPro GmbH using the Massive Analysis of cDNA Ends (MACE) Library Preparation Kit (v2.0) from GenXPro GmbH. First, cDNA was generated using Oligo(dT) primers with distinct Oligo IDs per sample for subsequent pooling of up to 24 samples. After pooling, cDNA was fragmentated to an average size of 200 bp using the sonicator Biorupter Plus (Diagenode, Belgium). The distribution of cDNA fragment sizes was monitored using the automated microfluidic electrophoresis station LabChip GXII Touch HT platform (PerkinElmer, USA). The poly(A) containing cDNA fragments were purified using solid phase reversible immobilization (SPRI) beads (Agencourt AMPure XP, USA), end repaired and ligated to distinct 8-base pair UMI Adapters (also called TrueQuant adapters). Then, the library containing labelled and fragmentated cDNA was amplified by PCR, purified by SPRI beads (Agencourt AMPure XP, USA) and strand-specific sequenced using the HiSeq2500 (Illumina, USA).

Bioinformatic analysis was performed according to the analysis pipeline for MACE libraries by GenXPro GmbH. Unique Oligo IDs and UMIs on each transcript allowed initial demultiplexing and subsequent removal of PCR-duplicates. The remaining reads were trimmed for high-quality as well as adapter-free sequences and aligned to the human reference genome (Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.p13) using Bowtie 2. Resulting output data were implemented in the database program FileMaker for further analysis. All data received from the Bioconductor software from our RNA-Seq or ATAC-Seq experiments were incorporated into a FILEMAKER database program. All final data sets were exported from the FILEMAKER Database program as individual Excel documents for publication. Moreover, the FILEMAKER Database program was used to develop the GUDC, DAGT and DAGE/ST modules. In addition, we used the following server for further data analysis: ClustVis (https://biit.cs.ut.ee/clustvis/) for heatmap analyses; VolcaNoseR (https://huygens.science.uva.nl/VolcaNoseR/) for volcano plots; ShinyGO v0.61 (http://bioinformatics.sdstate.edu/go/) for pathway analyses.
Supplemental Figure legends:

Figure S1: Data dissection of the MACE- and ATAC-Seq experiment
The dissected data obtained from our MACE- and ATAC-Seq experiments are summarized. Upper panels display the identified gene signatures from days 0-28; the gene signatures were dissected into pseudogenes (PG), non-annotated genes (NA), LincRNA genes (LINC), microRNA genes (MIR), SNO genes (SNO), mitochondrial genes (MT) and protein coding genes (PCG), respectively. Lower panels display the dissected ATAC-Seq data. In the first panel, total number of entries, and the number of accessible (total up) and non-accessible chromatin (total down) is displayed. The last 4 columns represent the total number of the up- and down-regulated chromatin regions and their associated genes. These were calculated for up- and downregulated chromatin regions by a minimum of 2 reads, a p-value < 0.05 combined with log2 changes of >2 in case of up-regulated genes, while downregulated genes were identified by a minimum of 2 reads in the mock sample, a p-value < 0.05 combined with a log2 value of < -2.

Figure S2: Profiling mitochondrial gene transcription
Several mitochondrial genes were strongly transcribed in cells that express MLL-AF4 or both t(4;11) fusion proteins, even when the AF4-MLL transgene was shut down after day 3. Among those genes, ATP6, ATP8, CO1-CO3, CYB, ND4, as well both ribosomal RNAs RNR1 and RNR2 were highly expressed. It is not clear what kind of physiological consequences this causes in the cells, because respiration experiments did not show any significant difference between mock cells and cells expressing these t(4;11) fusion proteins.

Figure S3: GUDC module analysis
All 4 gene signatures were analyzed for the number of genes that were deregulated on each chromosome and displayed as percentages of all genes present on the same chromosome. Left: up-regulated genes in the t(4;11) cell culture model at all 4 timepoints (day 0 = blue; day 3 = green; day 12 = red; day 28 = orange). Right: down-regulated genes. For each timepoint the “mean of deregulated genes for all chromosomes” is given in percent. The graphs display a ‘fingerprint’ of the gene usage on each chromosome. The resulting graphical pattern can be used to deduce the genome-wide impact of the tested fusion proteins. It also demonstrates that target gene usage is not a random process.

Figure S4: Venn diagram and detailed pathway analysis of de novo genes
A. De novo genes signatures were analyzed along all 4 timepoints. All idiosyncratic and shared signature genes were analyzed by pathway analysis. Those where pathways could be identified are indicated by green numbers, while for all grey numbers, no pathway was identified. Pathways were found for days 3 and 28, as well intersections exhibiting 41, 95 and 189 genes, respectively. These pathways revealed G protein-coupled signaling, including Ca2+ signaling, as well as innate and humoral immune response, as well as B cell proliferation.
B. the Signature Tracing module revealed the same pathways, but could in addition attribute these pathways to the protein coding genes that arose at d3 and the combined signatures that derive from d3 and d28.

Figure S5: Venn diagram and detailed pathway analysis of shut-down genes
Shut-down genes signatures were analyzed along all 4 timepoints. All idiosyncratic and shared
signature genes were analyzed by pathway analysis. Those where pathways could be identified are indicated by red numbers, while for all grey numbers, no pathway was identified. Pathways were found for days 0, 3, 12 and 28. These signatures all show 3 major pathways, namely adhesion, migration and secretion. Idiosyncratic and intersection signatures - exhibiting 202 and 153 genes - identified again innate immune pathways and cell adhesion.

B. the Signature Tracing module revealed the same pathways (adhesion and migration), but could in addition attribute these pathways to the protein coding genes that arose at d0 and d3, while downregulated T-cell specific functions could be attributed to the d0-derived signature at day 12, while T-helper immune functions derived from the d3-signature at day28.

Figure S6:
MACE or ATAC-Seq experiments were performed and resulting data were analyzed by Bioconductor software to create output Excel files. Various other bioinformatic tools were used to analyze these data (volcano plots with huygens.science.uva.nl; heatmaps with biit.cs.ut.ee/clustvis; pathway analysis with bioinformatics.sdstate.edu). Circos plots display the genome-wide changes identified by RNA-Seq or ATAC Seq experiments. Finally, all Bioconductor output files were imported to the Filemaker database program, which allowed the creation of novel modules for refined analyses not provided by the Bioconductor package (GUDC, DAGT, DAGE/ST).
### Figure S1

#### MACE data dissection

| Genes Type          | PG | NA  | LINC | MIR | SNO | MT  | PCG |
|---------------------|----|-----|------|-----|-----|-----|-----|
| d0 upregulated genes (n=50) | 15 | 4   | 1    | 0   | 0   | 17  | 13  |
| d0 downregulated genes (n=42) | 2  | 1   | 0    | 0   | 0   | 1   | 38  |
| d3 upregulated genes (n=634) | 255| 52  | 13   | 5   | 2   | 14  | 92  |
| d3 downregulated genes (n=58) | 16 | 14  | 9    | 1   | 2   | 2   | 23  |
| d28 upregulated genes (n=554) | 287| 264 | 4    | 4   | 3   | 16  | 36  |
| d28 downregulated genes (n=53) | 5  | 7   | 0    | 1   | 0   | 2   | 38  |

#### ATAC Seq data dissection

| Timepoint | total entries | total up | total down | log2 > 1 | log2 > 2 | log2 < -2 | log2 < -1 |
|-----------|---------------|----------|------------|----------|----------|-----------|-----------|
| d0        | 46.127        | 23.092   | 23.035     | 536      | 243      | 281       | 576       |
| d3        | 48.129        | 19.627   | 28.502     | 479      | 189      | 432       | 846       |
| d28       | 53.103        | 24.357   | 28.746     | 607      | 314      | 180       | 606       |

| Timepoint | total entries | PG  | NA genes | LINC | MIR | SNO | PCG |
|-----------|---------------|-----|----------|------|-----|-----|-----|
| d0        | 46.127        | 3.718 | 7.151   | 921  | 417 | 105 | 10.786 |
| d3        | 48.129        | 3.834 | 7.248   | 773  | 368 | 72  | 10.736 |
| d28       | 53.103        | 5.676 | 9.572   | 1.018 | 675 | 117 | 11.450 |

| Timepoint | total entries | PG  | NA genes | LINC | MIR | SNO | PCG |
|-----------|---------------|-----|----------|------|-----|-----|-----|
| d0        | 46.127        | 6.223 | 9.739   | 894  | 1.047 | 179 | 10.672 |
Figure S4

S4A

De novo genes

S4B

de novo upregulated genes

Enrichment FDR  # genes functional category
0.0314613539739 4  Gas transport

Enrichment FDR  # genes functional category
0.00731310078209 120 G protein-coupled receptor signaling pathway
0.0176103431012 63 Response to bacterium
0.0253177592794036 59 Regulation of signaling receptor activity
0.003417368890311 54 Defense response to other organism
0.00177270027761 38 Humoral immune response

Enrichment FDR  # genes functional category
0.003806710676578 50 Regulation of signaling receptor activity
0.0130342132505113 37 Calcium ion homeostasis
0.003806710676578 32 Regulation of cytosolic calcium ion concentration
0.003782922352792 16 PLC-activating G protein-coupled receptor signaling
0.0017032132505113 15 Inorganic anion transmembrane transport
0.001899915725487 13 B cell proliferation
Figure S5

S5A
shut-down genes

S5B

| Enrichment FDR | # genes | functional category | 1249 |
|----------------|--------|---------------------|------|
| 0.04852694678230 | 117 | Cell adhesion | |
| 0.03287680402920 | 65 | Regulation of signaling receptor activity | |
| 0.04852694678230 | 28 | Positive regulation of ERK1 and ERK2 | |
| 0.04852694678230 | 19 | Lymphocyte migration | |

| Enrichment FDR | # genes | functional category | 328 |
|----------------|--------|---------------------|-----|
| 0.02943492190027 | 15 | Cell-cell adhesion via adhesion molecules | |
| 0.0412230902445 | 6 | Thymic T cell selection | |
| 0.0087029602924 | 5 | Negative thymic T cell selection | |
| 0.04057322098956 | 5 | Positive thymic T cell selection | |

| Enrichment FDR | # genes | functional category | 836 |
|----------------|--------|---------------------|-----|
| 0.0121445371806987 | 94 | Cell adhesion | |
| 0.0274817115489385 | 47 | Regulation of signaling receptor activity | |
| 0.0181771792229658 | 19 | Homogeneous cell adhesion via plasma membrane | |
| 0.0181771792229658 | 16 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 350 |
|----------------|--------|---------------------|-----|
| 0.002019634863984 | 105 | Tissue remodeling | |
| 0.02238789895939 | 4 | T-helper 1 type immune response | |

| Enrichment FDR | # genes | functional category | 555 |
|----------------|--------|---------------------|-----|
| 0.0274817115489385 | 47 | Regulation of signaling receptor activity | |
| 0.0490737879707205 | 16 | Lymphocyte migration | |
| 0.0490737879707205 | 110 | Cell adhesion via plasma membrane | |
| 0.0490737879707205 | 192 | Homogeneous cell adhesion via plasma membrane | |
| 0.0490737879707205 | 2973 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 655 |
|----------------|--------|---------------------|-----|
| 0.0490737879707205 | 192 | Homogeneous cell adhesion via plasma membrane | |
| 0.0490737879707205 | 2973 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 192 |
|----------------|--------|---------------------|-----|
| 0.0490737879707205 | 2973 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 2093 |
|----------------|--------|---------------------|-----|
| 0.0490737879707205 | 2973 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 2069 |
|----------------|--------|---------------------|-----|
| 0.0490737879707205 | 2973 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 2093 |
|----------------|--------|---------------------|-----|
| 0.0490737879707205 | 2973 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 2069 |
|----------------|--------|---------------------|-----|
| 0.0490737879707205 | 2973 | Acid secretion | |

| Enrichment FDR | # genes | functional category | 2093 |
|----------------|--------|---------------------|-----|
| 0.0490737879707205 | 2973 | Acid secretion | |

| Time | d0 | d3 | d12 | d28 |
|------|----|----|-----|-----|
| 1249 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
| 2973 |
| 2093 |
| 3957 |
| 2069 |
| 2093 |
| 3957 |
| 3171 |
| 492 |
Figure S6

[Diagram showing a workflow from "in vitro cells" to "bioinformatic analysis" with steps including transfection, stable integration selection, RNA-Seq, and ATAC-Seq.]

[Additional information on gene usage on different chromosomes (GUDC), differential analysis by gene type (DAGT), and differential analysis of de novo or shut-down gene expression (DAGE) and signature tracing (ST).]