

# Genomics and epigenetics guided identification of tissue-specific genomic safe harbors

Dewan Shrestha<sup>1,2#</sup>, Aishee Bag<sup>3#</sup>, Ruiqiong Wu<sup>2</sup>, Yeting Zhang<sup>3</sup>, Xing Tang<sup>2</sup>, Qian Qi<sup>2</sup>, Jinchuan Xing<sup>3,4\*</sup>, Yong Cheng<sup>2,5\*</sup>

1. Department of Genetics, Genomics, and Informatics, College of Graduate Health Sciences, The University of Tennessee Health Science Center, Memphis, TN, USA
2. Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN, USA
3. Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
4. Human Genetic Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway, NJ, USA
5. Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

# These authors contributed equally to the project.

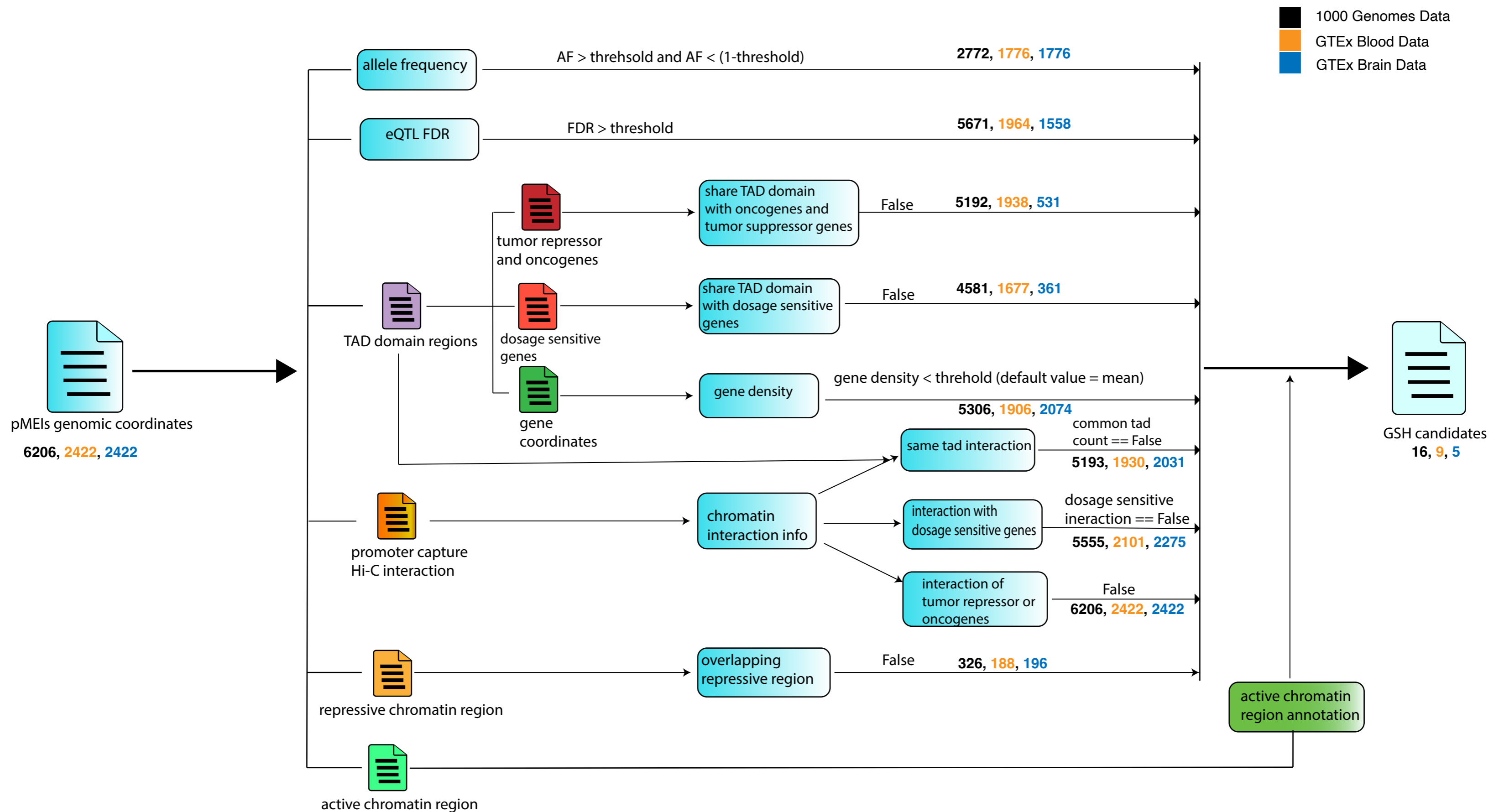
\* Corresponding authors: J.X. and Y.C.

Email addresses:

DS: dshrest2@uthsc.edu; AB: ab1603@rwjms.rutgers.edu; YZ: leafiezyt@gmail.com;  
RW: Ruiqiong.Wu@STJUDE.ORG; XT: tangx1986@gmail.com; QQ:  
Qian.Qi@STJUDE.ORG; JX: jinchuan.xing@rutgers.edu; YC: Yong.Cheng@STJUDE.ORG

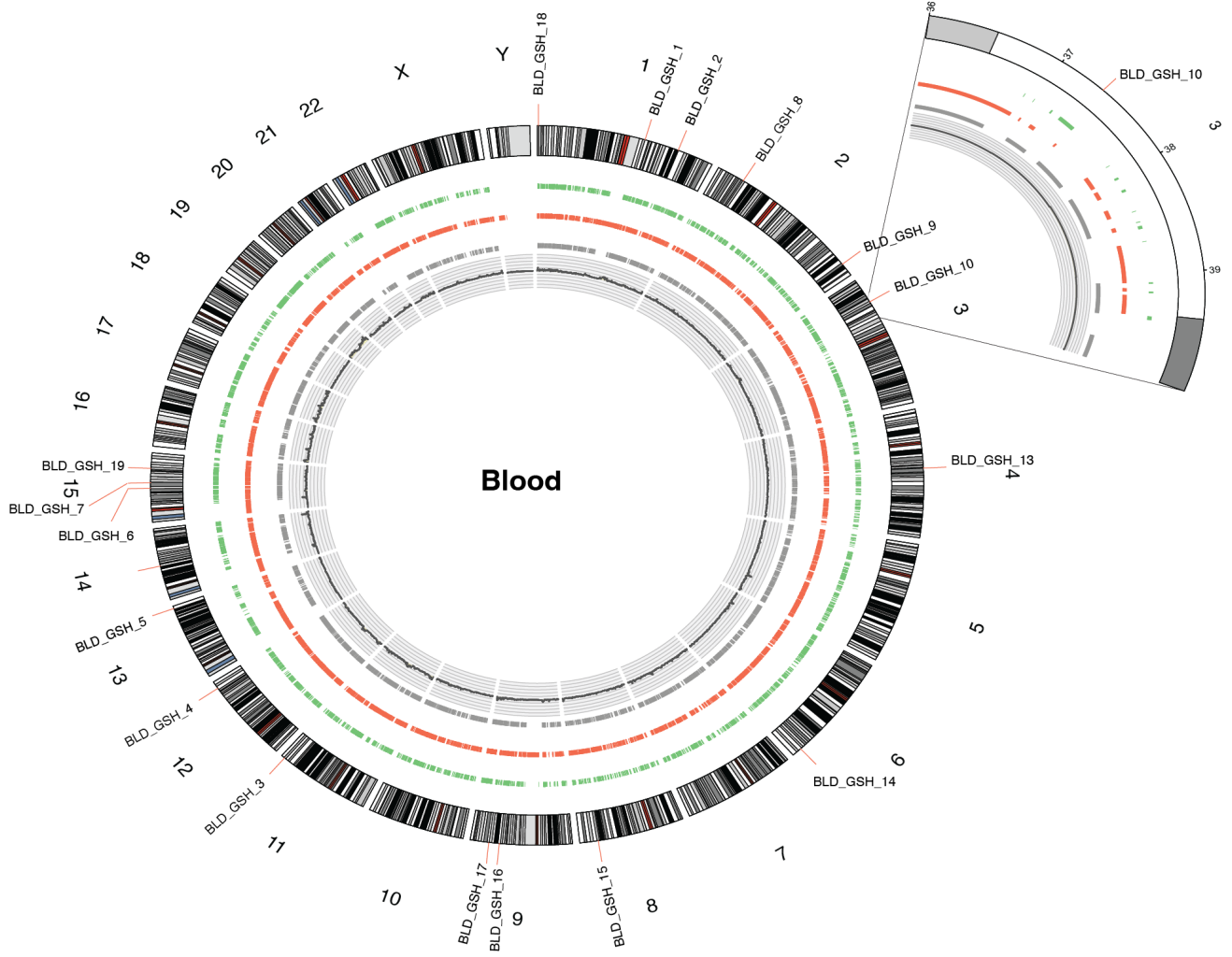
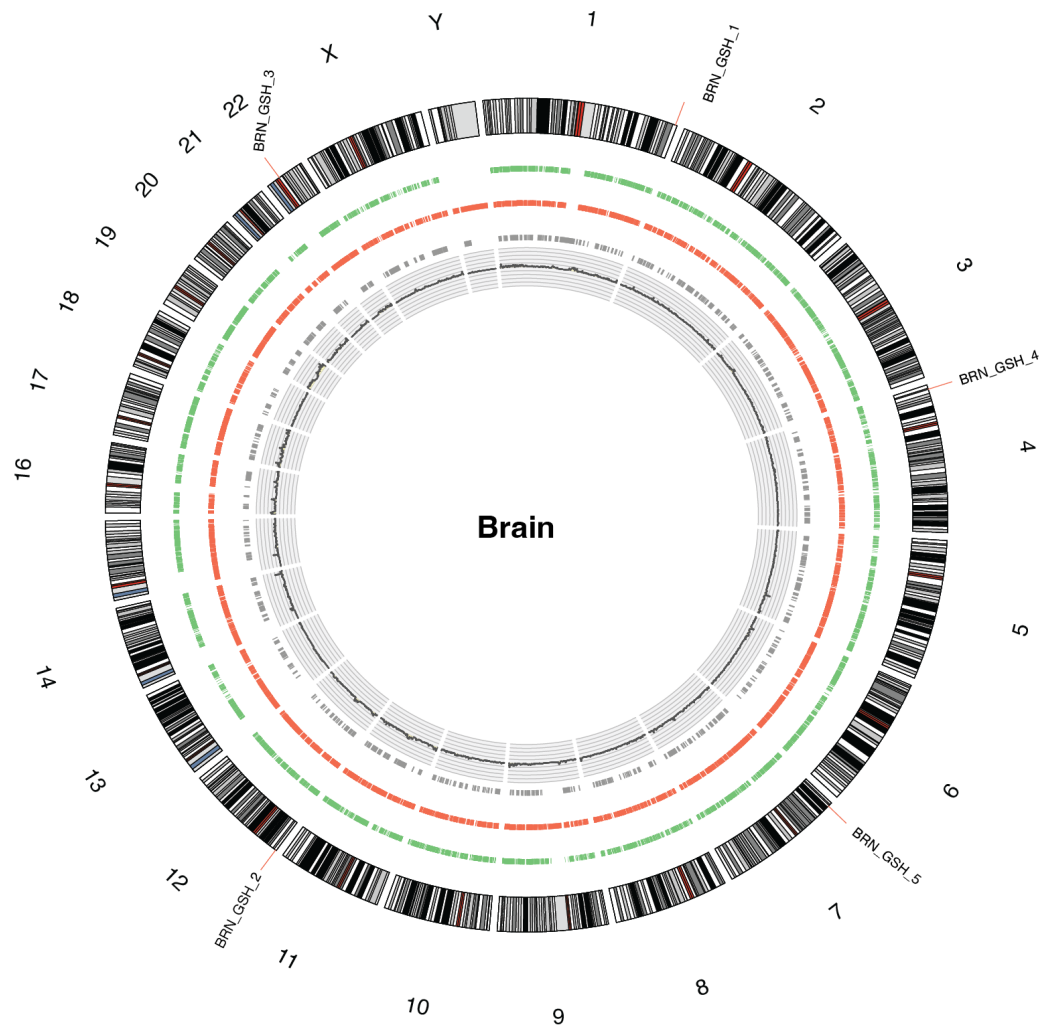
## **Additional file 1:**

Supplementary figures S1-9

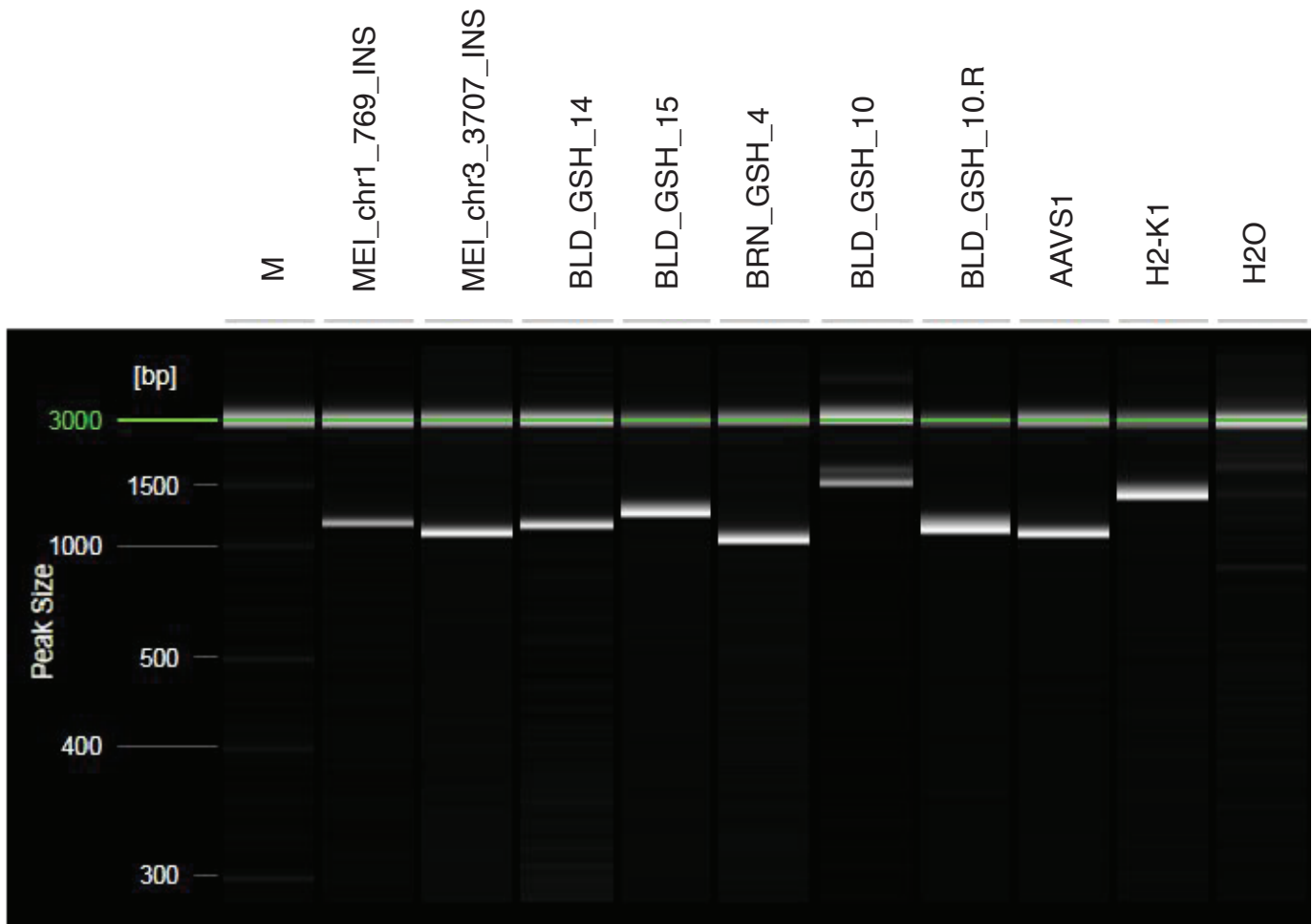


**Figure S1: Pipeline workflow for identification of GSH sites.**

The number in each of the filtration step is the number of pMEIs that passed the filtration criteria. Each color represents the specific dataset used. Black: 1000 Genomes Project; Orange: GTEx blood; Blue: GTEx brain. Because one pMEI can be filtered out at multiple steps, the numbers represented on each filtration steps are not exclusive.

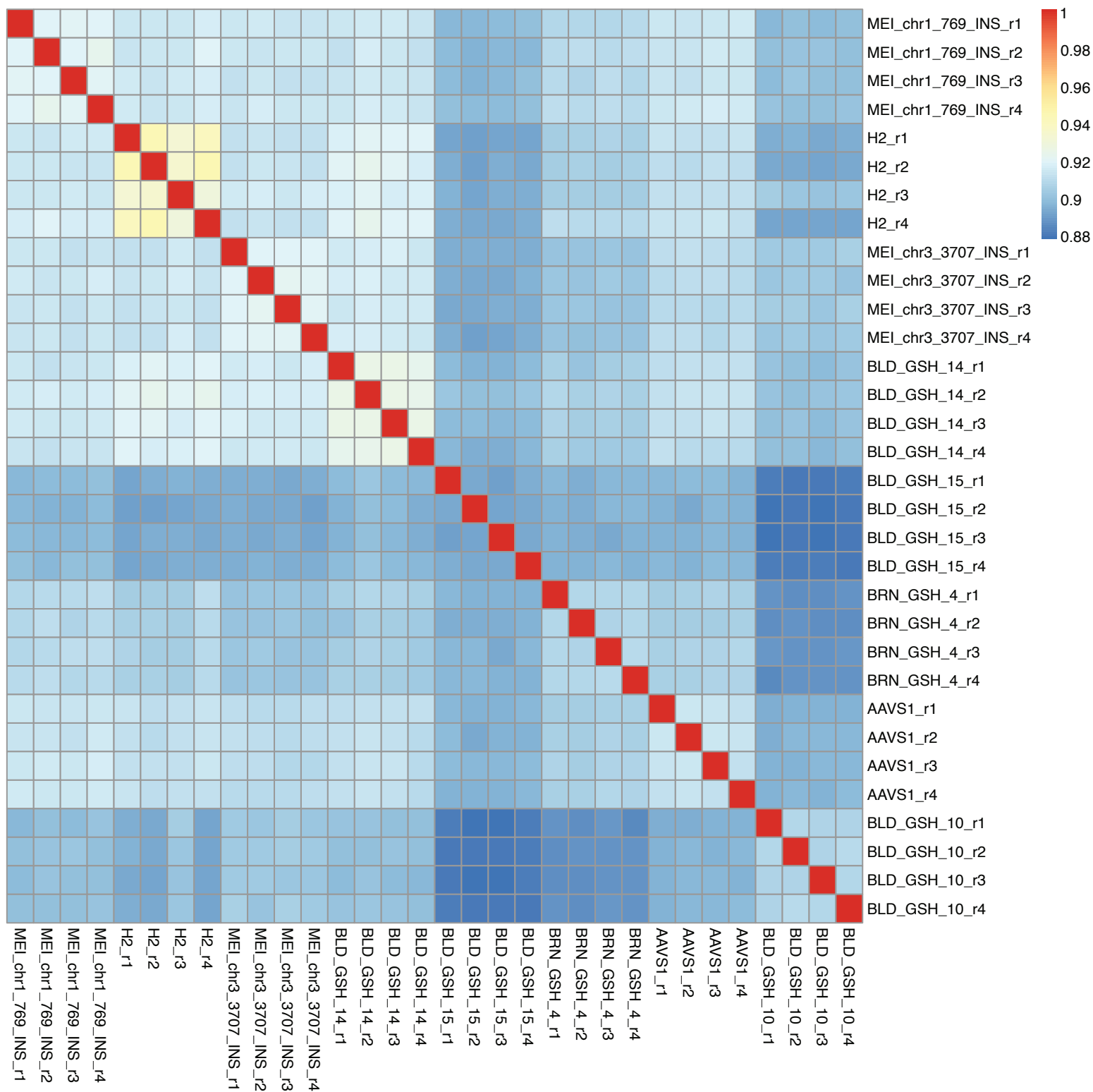
**a****b**

**Figure S2: Circos plot representation of GSH sites identified in blood (a) and brain (b) data.** The inner circle represents gene density. The grey bands outside of inner circle represents TADs. Repressive chromatin regions are represented by red bands and active regions by green bands. Validated BLD\_GSH\_10 locus on chromosome 3 in the blood data was enlarged as an example.

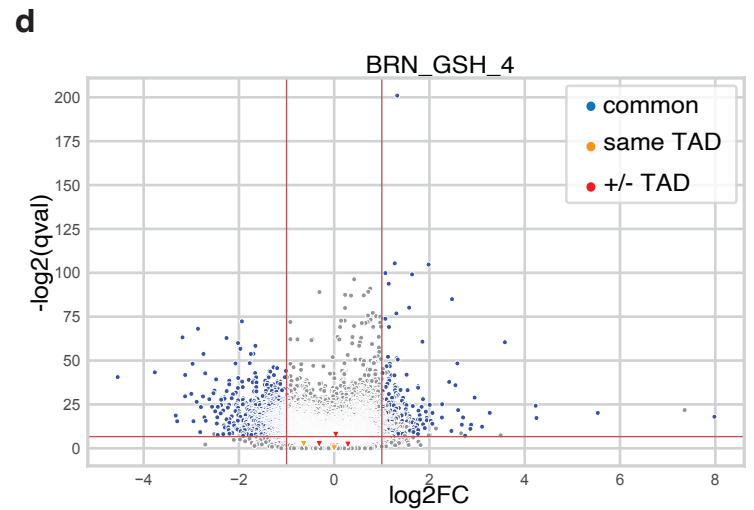
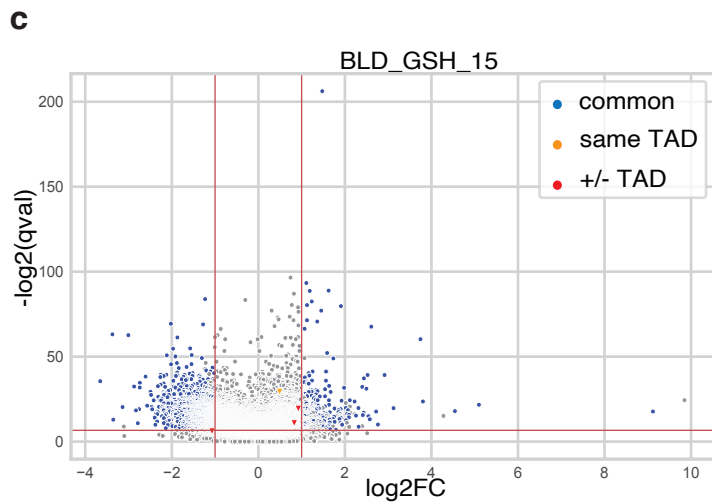
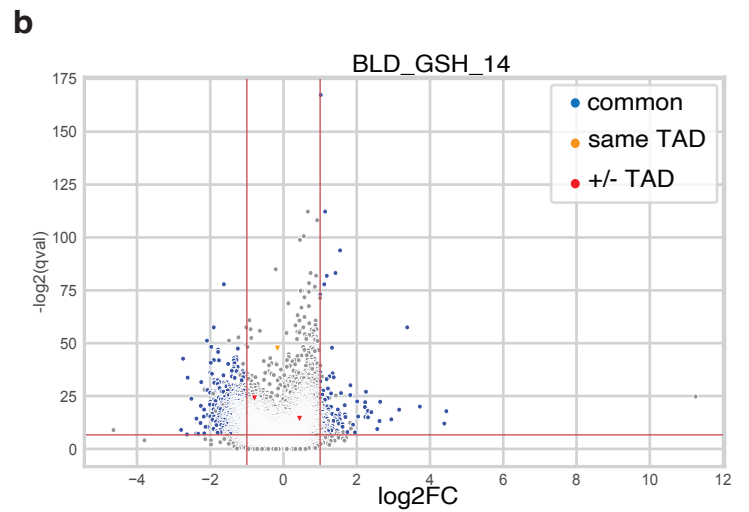
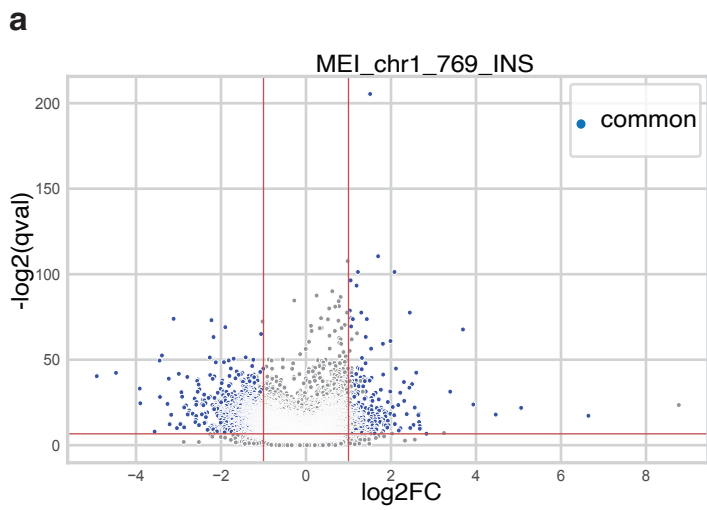


**Figure S3: PCR validation of GFP inserted cell line.**

The first lane represents the molecular marker, lane 2 and 3 represent GFP integrations in non-GSH MEI sites, lane 4-9 represent GFP integrations in GSH sites, lane 10 represents GFP positive HUDEP2 cells as a positive control and lane 11 is the negative control (water).

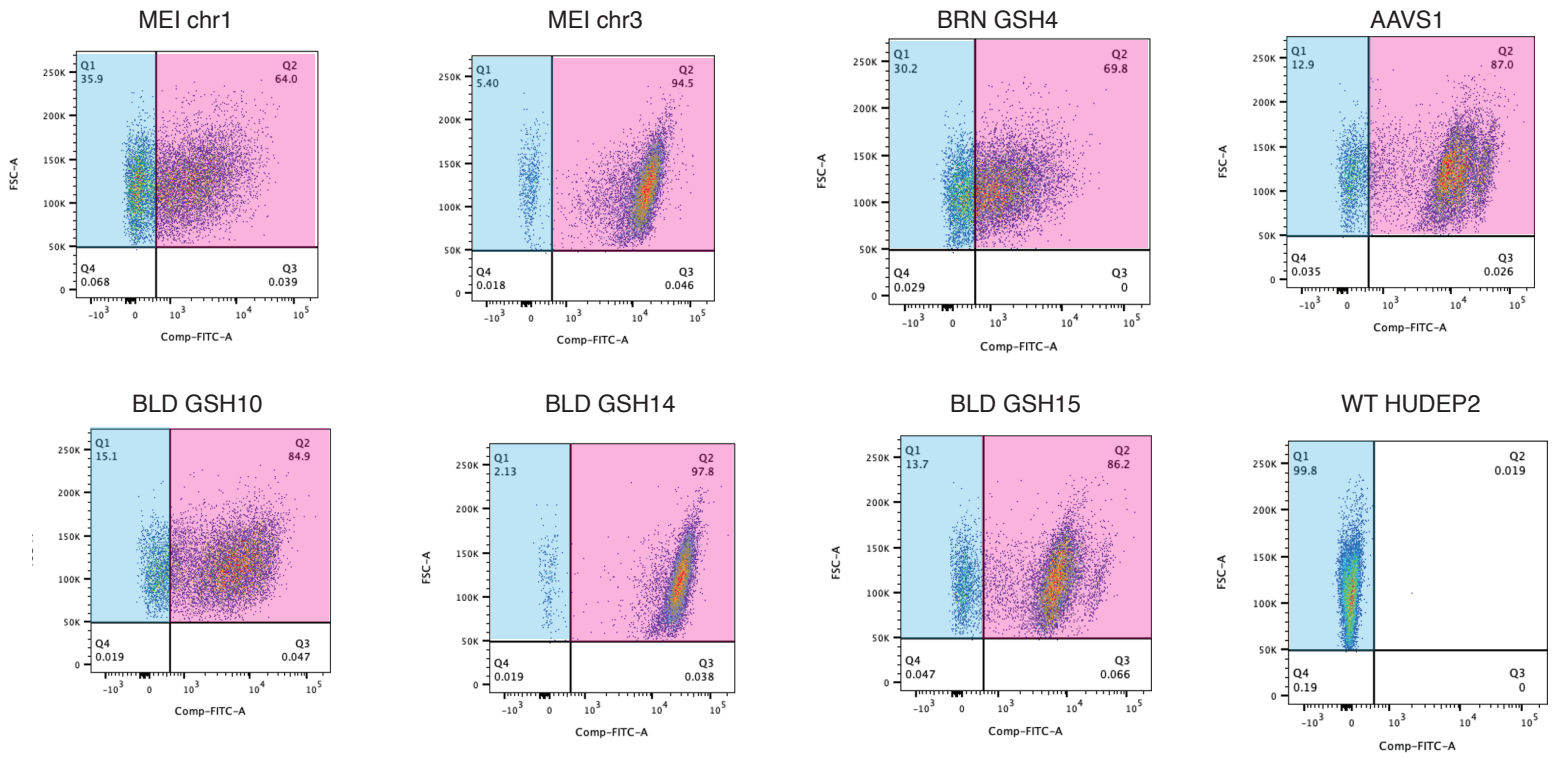


**Figure S4: Genome wide gene expression Correlation between GFP integrated cell lines and HUDEP2 cells.**  
 Genome-wide correlation plot between all GFP inserted HUDEP2 cell line and HUDEP2 WT cells.



**Figure S5: Volcano plot representation of Differentially Expressed Genes for four GSH integration sites.**  
 Common: DEGs share by more than two cell lines; Same TAD: genes within the same TAD of the GFP integration site; +/- TAD: genes in the TADs flanking to the GFP integration site.

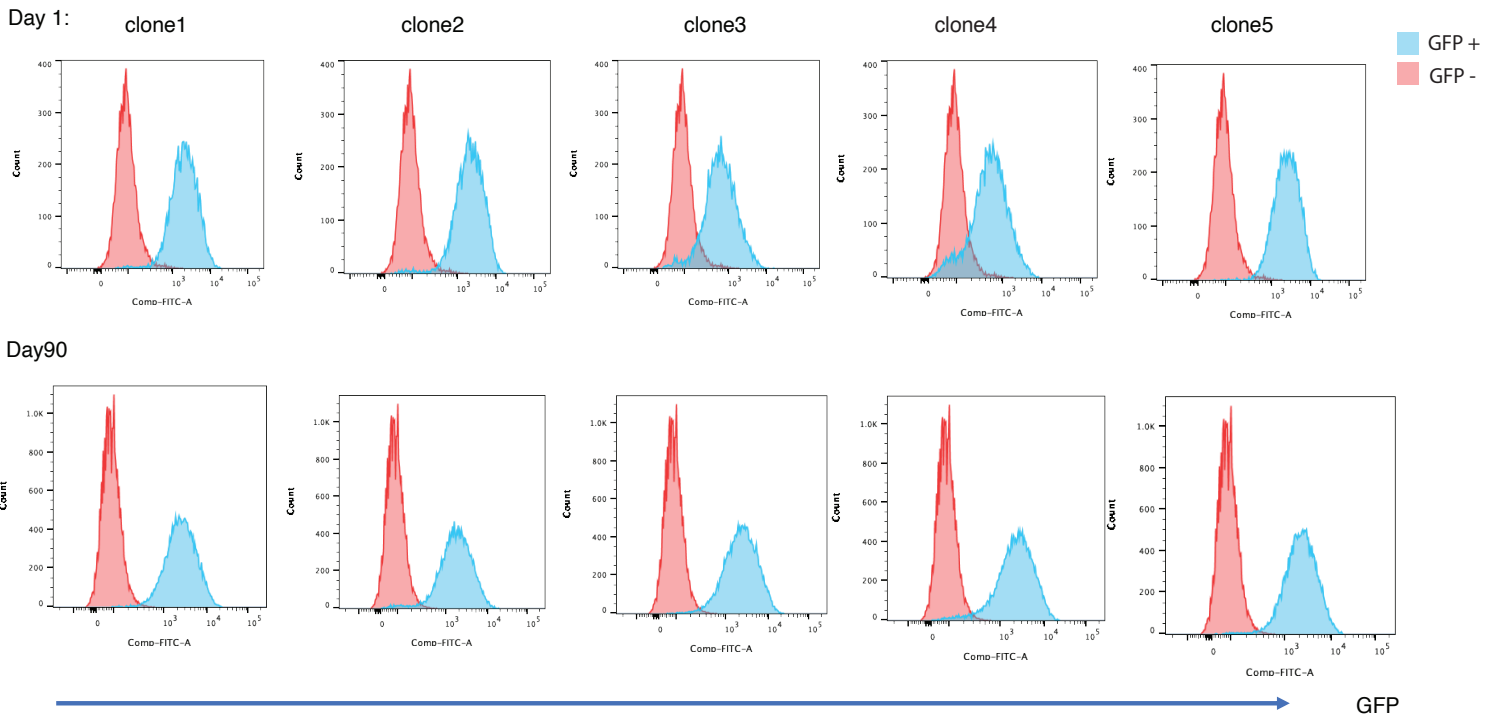
GFP -  
GFP +



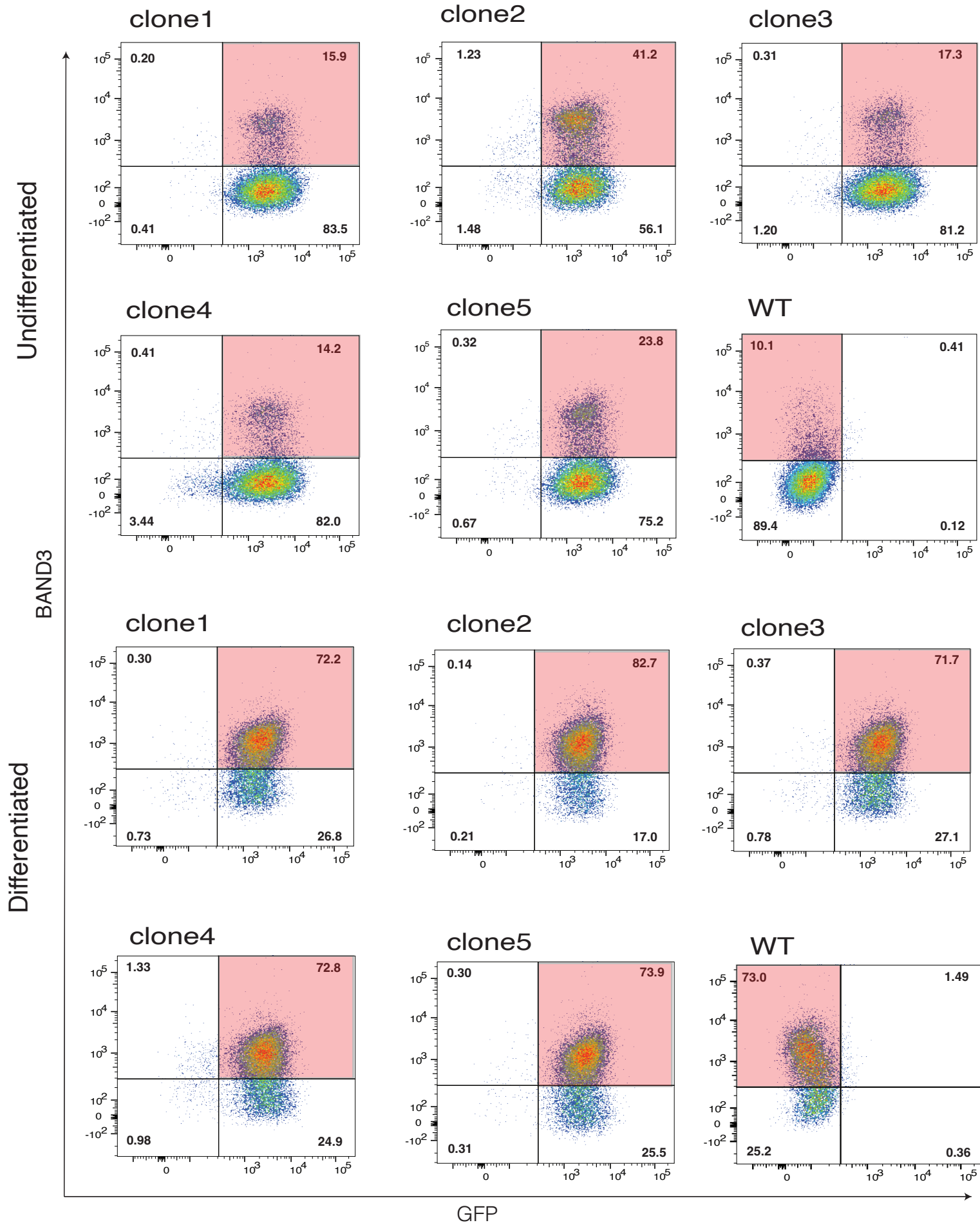
**Figure S6: GFP maintenance after one month based on different integration sites.**

The GFP expression compartment is highlighted in pink and GFP negative compartment is highlighted in blue.

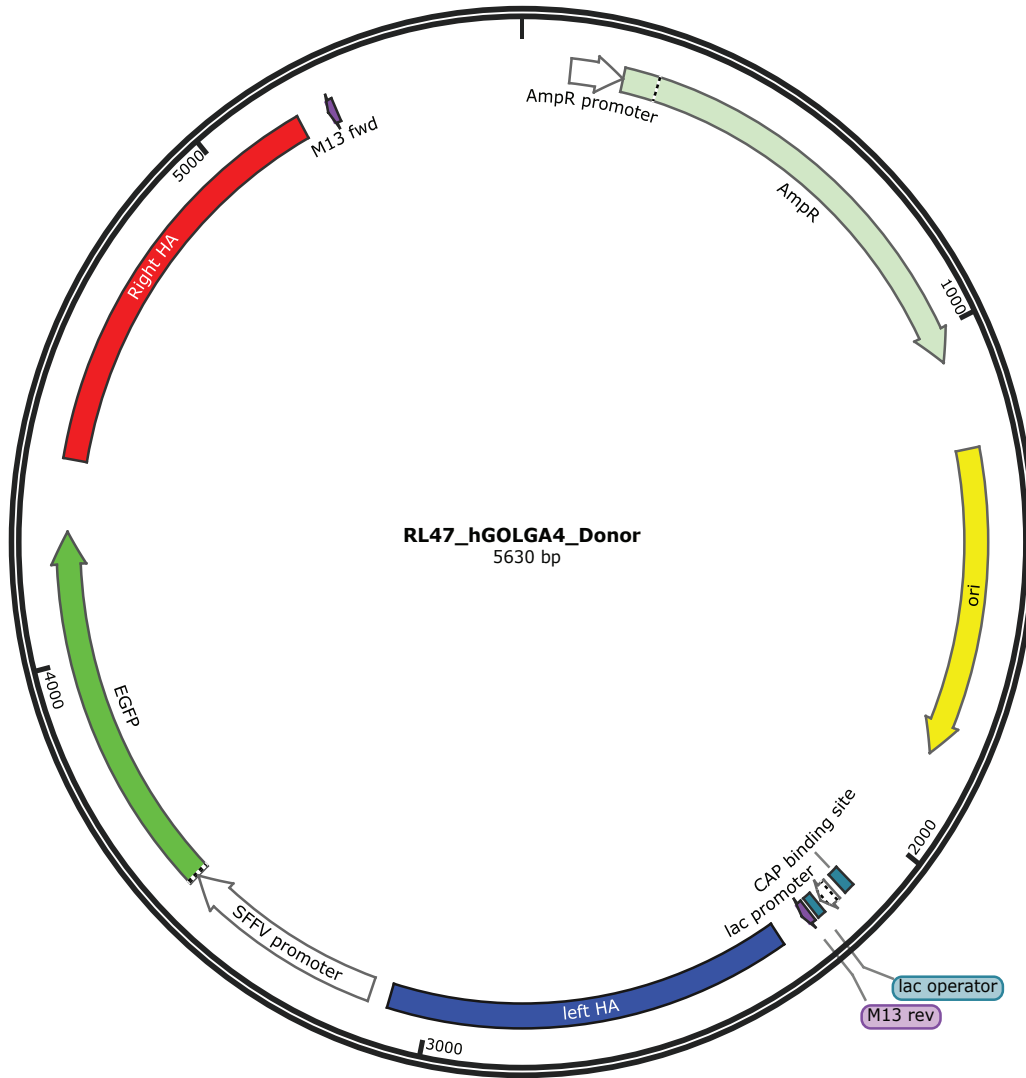




**Figure S7: GFP expression in Day 1 and Day 90 for 5 GFP clones of BLD\_GSH\_10.**



**Figure S8: Flow cytometry of GFP and Band3 staining of HUDEP2 GFP clones and HUDEP2 WT cells on day 0 and day 3 of differentiation. The GFP expression compartment is highlighted in red.**



**Figure S9: Plasmid vector used for the GFP insertion in HUDEP2 cells.**