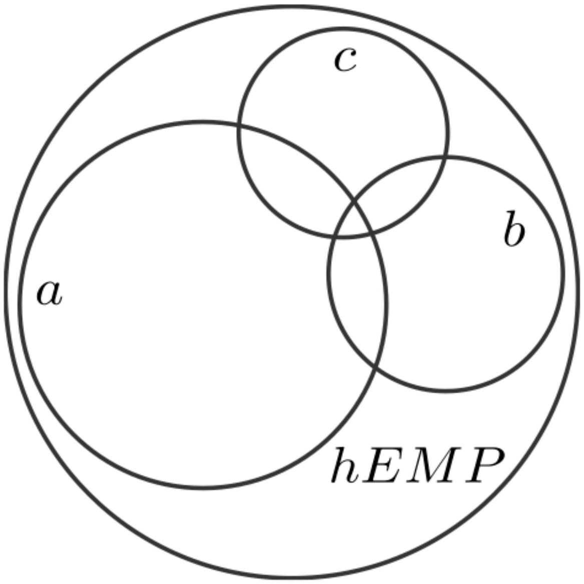
Supplementary Figure 4

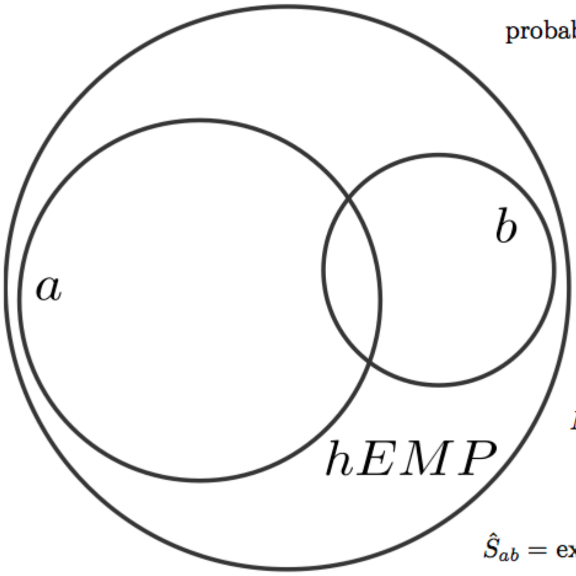
The model we use for calculating expectations assumes that the sets of lentiviral tags in the lineages are all randomly and independently drawn from the set of tags in the hEMP population before differentiation. This random draw models chance fate decisions, cell sampling during sorting, genomic DNA sampling during isolation and library preparation, and sampling during sequencing. In the graphical example shown below, the size of the circle indicates the size of the set of lentiviral tags. Lineage *a* has more tags than lineage *b*, which has more tags than lineage *c*. All three sets are contained by the total set of tags in the hEMP population.



We illustrate below how the expected numbers of shared lentiviral tags between lineages *a* and *b* (^S_{ab}) would be calculated under this model using the Chao2 estimates *N* of the number of tags in the original tagged hEMP population:

probability that tag in $hEMP$ is sampled and sequenced in $a = \frac{S_a}{N}$

probability that tag in $hEMP$ is sampled and sequenced in $b = \frac{S_b}{N}$



$$\hat{S}_{ab} = N \times \frac{S_a}{N} \times \frac{S_b}{N}$$

$N =$ Chao2 estimate of # of tags in $hEMP$ population

$S_a =$ # of tags sequenced in population a

$S_b =$ # of tags sequenced in population b

$\hat{S}_{ab} =$ expectation for # of IS expected to be detected in both a and b, assuming complete multipotency