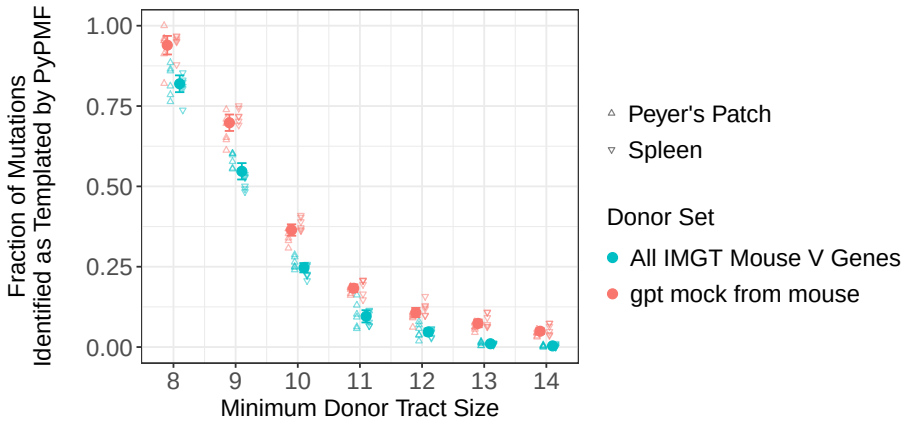
Supplement to "Lack of evidence for a substantial rate of templated mutagenesis in B cell diversification"*

Julia Fukuyama, Branden J Olson, and Frederick A Matsen IV

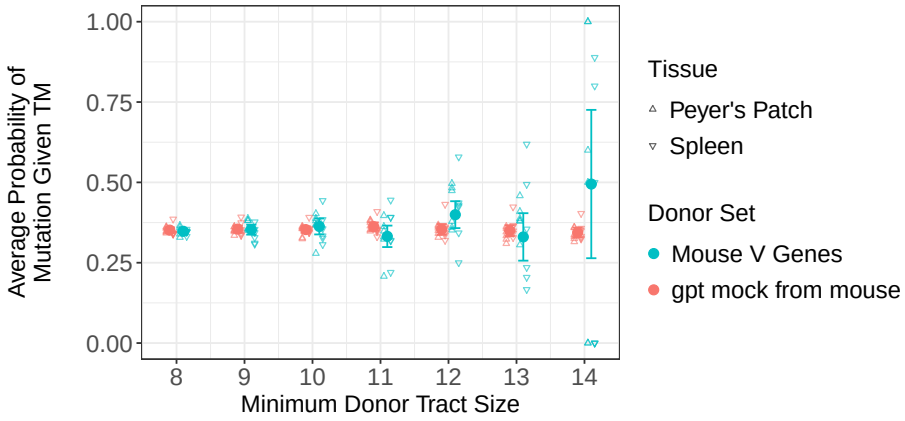
Donor set	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
gpt mock from Human	0.18	0.30	0.38	0.41	0.50	0.83
IMGT Human	0.23	0.31	0.36	0.40	0.46	0.74
gpt mock from mouse	0.18	0.39	0.49	0.48	0.58	0.79
IMGT Mouse	0.21	0.38	0.50	0.48	0.57	0.74

Supplemental Table I: Five-number summaries of the set of divergences between genes and root for four donor gene sets. The divergences for the *gpt* human mock set are similar to the divergences for the IMGT human set, and the divergences for the *gpt* mouse mock set are similar to the divergences for the IMGT mouse set.

*This research was supported by National Institutes of Health grants R01 GM113246, R01 AI120961, U19 AI117891, and R01 AI146028. The research of Frederick Matsen was supported in part by a Faculty Scholar grant from the Howard Hughes Medical Institute and the Simons Foundation.



Supplemental Figure 1: Hollow triangles represent the fraction of mutations explainable by templated mutagenesis in each sample, with upward-pointing triangles corresponding to samples from Peyer's patches and downward-pointing triangles corresponding to samples from the spleen. Reverse complements are included in each donor set. For each tract length, the filled circle and error bar represents the overall estimate of the probability of a mutation being explainable by templated mutagenesis plus or minus two standard errors. Points corresponding to samples from Peyer's patches and spleen are offset slightly to the left and right, respectively, to facilitate comparison and to avoid overplotting. This analysis was performed once on data from six individual mice, with two replicates per mouse corresponding to samples from Peyer's patches and spleen, yielding 12 total samples.



Supplemental Figure 2: Average probability of the observed mutations under a templated mutagenesis model, using *gpt* genes and their reverse complements (red) as well as the set of 129S1 V genes and their reverse complements (blue). Each point corresponds to one sample taken from either spleen or Peyer's patches. This analysis was performed once on data from six individual mice, with two replicates per mouse corresponding to samples from Peyer's patches and spleen, yielding 12 total samples.

Mice:

k	PyPMF rate	PyPMF FPR	UB (1)	UB (.99)	UB (.95)	UB (.9)
8	0.9	0.94	0	0	0	—
9	0.7	0.7	0.02	0.02	0.03	0.03
10	0.46	0.36	0.15	0.15	0.16	0.18
11	0.22	0.18	0.04	0.04	0.04	0.04
12	0.13	0.11	0.03	0.03	0.03	0.03
13	0.08	0.07	0.01	0.01	0.01	0.01
14	0.06	0.05	0.01	0.01	0.01	0.01

Humans:

k	PyPMF rate	PyPMF FPR	UB (1)	UB (.99)	UB (.95)	UB (.9)
8	0.88	0.91	0	0	0	—
9	0.6	0.65	0	0	0	0
10	0.34	0.29	0.07	0.07	0.08	0.08
11	0.22	0.14	0.09	0.09	0.1	0.1
12	0.16	0.08	0.09	0.09	0.1	0.1
13	0.15	0.05	0.1	0.1	0.11	0.11
14	0.14	0.03	0.1	0.11	0.11	0.12

Supplemental Table II: Upper bounds (UB) on the rate of templated mutagenesis in the VB1-8 (top) and the anti-Ebola sequences (bottom) computed for a range of tract lengths k and sensitivities when including reverse complements in the donor set. k denotes tract length, PyPolyMF rate is the naive PyPolyMF estimate of the rate of templated mutagenesis, PyPolyMF FPR is the PyPolyMF false positive rate, UB denotes upper bound, and the number in paretheses denotes the assumed sensitivity (true positive rate) of PyPolyMF.