Supplementary Materials

Nucleotide weight matrices reveal ubiquitous mutational footprints of AID/APOBEC deaminases in human cancer genomes

Igor B. Rogozin, Abiel Roche-Lima, Artem G. Lada, Frida Belinky, Ivan A. Sidorenko, Galina V. Glazko, Vladimir N. Babenko, David N. Cooper, Youri I. Pavlov

Abbreviations: APO1 = APOBEC1, APO3A = APOBEC3A, APO3B = APOBEC3B, APO3C = APOBEC3C, APO3G = APOBEC3G, AID = Activation Induced Deaminase

Supplementary Table S1. Datasets of mutations induced by over-expression of AID/APOBEC enzymes in the yeast genome.

Enzyme	Number of mutations (sites)	Lada et al., 2011	Taylor et al., 2013	Lada et al., 2015	Lada et al., 2017
APO1	673	-	+	+	+
APO3A	723	-	+	-	-
APO3B	745	-	+	-	-
APO3C	67	-	+	-	-
APO3G	1147	+	+	+	+
AID	397	-	+	+	+

References:

Lada A.G., Krick C.F., Kozmin S.G., Mayorov V.I., Karpova T.S., Rogozin I.B., Pavlov Y.I. (2011) Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. Biochemistry (Mosc.), 76: 131-146.

Taylor BJ, Nik-Zainal S, Wu YL, Stebbings LA, Raine K, Campbell PJ, Rada C, Stratton MR, Neuberger MS. (2013) DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. Elife 2: e00534.

Lada A.G., Kliver S.F., Dhar A., Polev D., Masharsky A., Rogozin I.B., Pavlov Y.I. (2015) Disruption of transcriptional coactivator Sub1 leads to genome-wide re-distribution of clustered mutations induced by APOBEC in active yeast genes. PLoS Genet. 11: e1005217.

Lada A.G., Stepchenkova E.I., Zhuk A.S., Kliver S.F., Rogozin I.B., Polev D., Dhar A., Pavlov Y.I. (2017) Recombination is responsible for the increased recovery of drug-resistant mutants with hypermutated genomes in resting yeast diploids expressing APOBEC deaminases. Front. Genet. 8: 202.

Supplementary Table S2. Control study: correlation between AID/APOBEC mutable motifs and the context of somatic mutations in C:G sites in mitochondrial DNA.

Cancer tissue type	#Mutations	Test	APO1	APO3A	APO3B	APO3C	APO3G	AID
Bladder	39	Ratio	.912	.913	.908	.958	.897	.937
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Blood	66	Ratio	.887	.917	.907	.902	.912	.901
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Brain	4	Ratio	.901	1.018	1.028	.923	.985	1.007
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test		0.089	0.031			0.342
Breast	283	Ratio	.918	.932	.925	.946	.932	.912
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Cervix	22	Ratio	.935	.901	.927	.946	.866	.875
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Colon	81	Ratio	.892	.904	.898	.913	.886	.906
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Kidney	7	Ratio	.917	.864	.911	.971	.931	.901
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Liver	49	Ratio	.818	.849	.828	.880	.833	.881
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Lung	55	Ratio	.930	.954	.929	.976	.987	.904
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Ovary	34	Ratio	.889	.883	.858	.923	.827	.857
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Prostate	79	Ratio	.927	.916	.930	.965	.938	.937
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						

Skin	16	Ratio	.867	.836	.871	.904	.878	.901
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Stomach	23	Ratio	.824	.867	.873	.939	.947	.938
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
Uterus	40	Ratio	.966	.979	.972	.946	.971	.867
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						

NSE indicates absence of significant excess of mutations in AID/APOBEC mutable motifs (NSE, no significant excess) suggesting that there is no association between mutagenesis and motifs. The correlation was measured using Student *t*-test and Monte Carlo (MC) tests. The asterisk denotes that the corresponding $P < 0.0005$ (critical value = 3.291) is a conservative estimate of the critical overall value of the *t*-test taking into account the Bonferroni correction for multiple testing ($16 \times 6 = 96$). "Ratio" is the mean weight of mutated sites divided by the mean weight of non-mutated sites. Results of the MC test are shown for cases where the Ratio > 1 (shown in green).

Supplementary Table S3. Control study: fractions of random matrices with a significant correlation between AID/APOBEC pseudo-mutable motifs (shuffled sites of mutations) and the context of somatic mutations [IGOR: at?] C:G sites.

Cancer tissue type	#Mutations	APO1	APO3A	APO3B	APO3C	APO3G	AID
Bladder	38750	0.	0.	0.	0.	0.	0.
Blood	10633	0.	0.	0.	0.	0.	0.
Brain	30926	0.	0.	0.	0.04	0.	0.03
Breast	48035	0.	0.	0.	0.	0.	0.
Cervix	41454	0.	0.	0.	0.	0.	0.
Colon	175109	0.	0.03	**0.12**	**0.21**	0.	**0.35**
Kidney	32382	0.	0.	0.	0.	0.	0.
Liver	74161	0.	0.	0.	0.01	0.	0.
Lung	180284	0.	0.	0.	0.	0.	0.
Ovary	22340	0.	0.	0.	0.	0.	0.
Pancreas	35165	0.	0.	0.	0.03	0.	0.
Prostate	16703	0.	0.	0.	**0.30**	0.	0.
Rectum	30018	0.	0.03	**0.18**	0.	0.	**0.17**
Skin	244248	**0.81**	**0.33**	0.	**0.96**	0.	0.
Stomach	115652	0.	0.	**0.15**	**0.15**	0.	**0.19**
Uterus	55999	0.	0.	0.	0.09	0.	0.

The significance of the excess of somatic mutations observed in a pseudo-mutable context was measured using the Student *t*-test and Monte Carlo (MC) tests (see Materials and Methods for details). The threshold value was corrected using the conservative Bonferroni correction for multiple testing (16 tissues x 6 mutable contexts = 96, $P < 0.0005$, critical t-test value = 3.291). The expected false discovery rate should be around 0.05, large fraction values (>0.10) are shown in green.

Supplementary Table S4. Control study: fraction of random matrices with a significant correlation between AID/APOBEC pseudo-mutable motifs (randomly sampled sites from the yeast genome) and the context of somatic mutations at C:G sites.

Cancer tissue type	#Mutations	APO1	APO3A	APO3B	APO3C	APO3G	AID
All somatic mutations							
Bladder	38750	**0.65**	**0.46**	**0.56**	**0.57**	**0.64**	**0.50**
Blood	10633	**0.19**	0.04	0.	**0.16**	0.01	0.
Brain	30926	0.	0.02	0.	**0.17**	0.01	0.01
Breast	48035	0.	0.07	0.	**0.35**	**0.10**	**0.21**
Cervix	41454	**0.70**	**0.56**	**0.64**	**0.58**	**0.73**	**0.61**
Colon	175109	0.	0.04	0.02	**0.25**	0.	0.
Kidney	32382	0.	0.02	0.01	**0.18**	0.01	0.02
Liver	74161	0.	0.04	0.01	**0.16**	0.01	0.
Lung	180284	0.07	0.02	0.04	**0.29**	0.01	0.09
Ovary	22340	0.	0.02	0.01	**0.15**	0.01	0.
Pancreas	35165	0.01	0.04	0.02	**0.20**	0.01	0.01
Prostate	16703	0.01	0.03	0.02	**0.23**	0.01	0.
Rectum	30018	0.01	0.04	0.02	**0.19**	0.02	0.02
Skin	244248	**0.79**	**0.72**	**0.80**	**0.63**	**0.84**	**0.70**
Stomach	115652	0.	0.04	0.01	**0.24**	0.01	0.03
Uterus	55999	0.01	0.04	0.02	**0.22**	0.02	0.02
Somatic mutations with a C:G-rich DNA context (≥50% C+G in the 10 nucleotide region around sites of somatic mutations)							
Bladder	38750	**0.42**	**0.28**	**0.30**	**0.47**	**0.32**	**0.41**
Cervix	41454	**0.46**	**0.29**	**0.37**	**0.40**	**0.39**	**0.48**
Skin	244248	**0.74**	**0.66**	**0.71**	**0.57**	**0.75**	**0.71**

The significance of the excess of somatic mutations observed in a pseudo-mutable context was measured using the Student *t*-test and Monte Carlo (MC) tests (see Materials and Methods for details). The threshold value was corrected using the conservative Bonferroni correction for multiple testing (16 tissues x 6 mutable contexts = 96, P < 0.0005, critical t-test value = 3.291). The expected false discovery rate should be around 0.05, large fraction values (>0.10) are shown in green.

Supplementary Table S5. Nucleotide composition of the DNA context of somatic mutations (±3 nucleotides).

	A	T	G	C
Bladder	**.212**	**.328**	**.212**	**.248**
Blood	.230	.224	.294	.252
Brain	.224	.213	.311	.252
Breast	.221	.300	.234	.245
Cervix	**.206**	**.350**	**.200**	**.244**
Colon	.196	.239	.328	.237
Kidney	.231	.255	.251	.263
Liver	.215	.236	.268	.281
Lung	.222	.254	.246	.278
Ovary	.234	.244	.262	.259
Pancreas	.214	.225	.309	.252
Prostate	.207	.215	.296	.281
Rectum	.207	.278	.294	.221
Skin	**.177**	**.356**	**.175**	**.293**
Stomach	.198	.219	.342	.242
Uterus	.209	.269	.306	.217

Sites with mutations in G nucleotides were converted to the complementary strand

Supplementary Table S6. Correlation between AID/APOBEC mutable motifs and the context of C:G > T:A somatic mutations.

Cancer tissue type	#Muta-tions	Test	APO1	APO3A	APO3B	APO3C	APO3G	AID
Bladder	23127	Ratio	1.487*	1.463*	1.461*	1.179*	1.101*	1.051*
		t-test	112.845	134.038	120.517	71.132	41.798	21.415
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.657		.627	.138	.921	.500
Blood	6331	Ratio	.946	1.026*	.978	1.021*	1.009	1.065*
		t-test	NSE	4.330	NSE	4.356	NSE	12.015
		MC test		<0.001		<0.001		<0.001
		Fraction		.190		.195		
Brain	23354	Ratio	.855	.975	.888	.992	.960	1.063*
		t-test	NSE	NSE	NSE	NSE	NSE	21.831
		MC test						<0.001
		Fraction						
Breast	10327	Ratio	1.233*	1.261*	1.231*	1.098*	1.039*	1.050*
		t-test	34.183	48.309	37.605	25.377	9.677	12.992
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.476			.123	.866	
Cervix	23034	Ratio	1.573*	1.542*	1.538*1	1.207*	1.113*	1.060*
		t-test	138.455	168.841	50.422	82.775	49.688	25.686
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.737		.707	.165	.165	.339
Colon	132513	Ratio	.838	1.003	.893	.989	.916	1.080*
		t-test	NSE	NSE	NSE	NSE	NSE	65.985
		MC test						<0.001
		Fraction						
Kidney	16690	Ratio	.972	1.024*	.995	1.020*	.989	1.039*
		t-test	NSE	6.486	NSE	6.899	NSE	11.814
		MC test		<0.001		<0.001		<0.001
		Fraction		.230		.215		
Liver	42221	Ratio	.915	.968	.946	.997	.998	1.014*
		t-test	NSE	NSE	NSE	NSE	NSE	6.400
		MC test						<0.001
		Fraction						
Lung	58116	Ratio	1.164*	1.179*	1.179*	1.077*	1.078*	1.014*
		t-test	59.858	82.536	72.640	48.475	42.941	8.518
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.393			.091	.815	
Ovary	10973	Ratio	.960	1.031*	.990	1.012*	.992	1.041*
		t-test	NSE	6.851	NSE	3.355	NSE	9.813
		MC test		<0.001		<0.001		<0.001
		Fraction		.199		.205		
Pancreas	25544	Ratio	.921	1.039*	.966	1.010*	.977	1.058*
		t-test	NSE	13.290	NSE	4.034	NSE	21.396
		MC test		<0.001		<0.001		<0.001
		Fraction		.178		.291		

Prostate	9506	Ratio *t*-test MC test Fraction	.866 NSE	1.001 NSE	.910 NSE	1.013* 3.411 <0.001 .230	.962 NSE	1.082* 18.442 <0.001
Rectum	20315	Ratio *t*-test MC test Fraction	.956 NSE	1.086* 24.646 <0.001 .239	1.009 NSE	1.028* 10.270 <0.001	.965 NSE	1.076* 25.245 <0.001
Skin	230979	Ratio *t*-test MC test Fraction	1.438* 331.007 <0.001 .617	1.333*3 00.669 <0.001	1.402*3 69.514 <0.001 .527	1.189*2 23.837 <0.001 .281	1.148* 178.865 <0.001 .980	1.038* 46.377 <0.001 .667
Stomach	92882	Ratio *t*-test MC test Fraction	.799 NSE	.970 NSE	.859 NSE	.970 NSE	.901 NSE	1.070* 48.316 <0.001
Uterus	41875	Ratio *t*-test MC test Fraction	.931 NSE	1.089* 35.553 <0.001 .292	.980 NSE	1.046* 23.344 <0.001	.929 NSE	1.107* 52.688 <0.001

NSE means that No Significant Excess of mutable motifs was found. The significance of the observed excess was measured using the Student *t*-test and Monte Carlo (MC) tests. The asterisk denotes that the corresponding $P < 0.0005$ (critical value = 3.291) is a conservative estimate of the critical overall value of the *t*-test taking into account the Bonferroni correction for multiple testing ($16 \times 6 = 96$). “Ratio” is the mean weight of mutated sites divided by the mean weight of non-mutated sites. The predicted fraction of mutations induced by AID/APOBEC proteins (“Fraction”) is shown when the significant excess of somatic mutations in a mutable motif comparisons was detected; all cases where there was a significant difference between observed and expected distributions ($P > 0.05$) were discarded.

Supplementary Table S7. Correlation between AID/APOBEC mutable motifs and the context of C:G > G:C somatic mutations.

Cancer tissue type	#Muta-tions	Test	APO1	APO3A	APO3B	APO3C	APO3G	AID
Bladder	11020	Ratio	1.698*	1.575*	1.614*1	1.178*	1.162*	1.015*
		t-test	146.887	149.530	53.222	58.002	55.175	4.832
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction		.795	.819	.758	.266	
Blood	2095	Ratio	.911	.939	.922	.936	.903	1.006
		t-test	NSE	NSE	NSE	NSE	NSE	NSE
		MC test						
		Fraction						
Brain	3306	Ratio	1.160*	1.101*	1.126*	1.036*	1.013	1.033*
		t-test	14.417	11.126	11.777	5.912	NSE	4.534
		MC test	<0.001	<0.001	<0.001	<0.001		<0.001
		Fraction	.394	.363	<0.001	.147		.550
Breast	22746	Ratio	1.553*	1.451*	1.487*	1.138*	1.111*	1.019*
		t-test	143.860	141.757	141.726	61.792	49.665	8.114
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction		.718	.733	.109	.193	.234
Cervix	12993	Ratio	1.802*	1.667*	1.705*	1.214*	1.174*	1.033*
		t-test	224.947	228.869	237.999	76.699	71.539	11.911
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction		.872	.915	.648	.308	
Colon	6336	Ratio	1.080*	1.032*	1.039*	.988	.949	1.029*
		t-test	9.495	4.781	5.040	NSE	NSE	5.476
		MC test	<0.001	<0.001	<0.001			<0.001
		Fraction		.321				
Kidney	5867	Ratio	1.129*	1.070*	1.114*	1.027*	1.043*	1.017
		t-test	15.831	10.701	15.046	6.015	7.853	NSE
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	
		Fraction	.380	.334		.153		
Liver	8314	Ratio	1.142*	1.091*	1.126*	1.029*	1.043*	1.012
		t-test	20.167	16.430	19.542	7.435	8.894	NSE
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	
		Fraction	.368	.329		.129		
Lung	32118	Ratio	1.289*	1.250*	1.273*	1.081*	1.076*	1.032*
		t-test	78.368	86.746	82.654	41.739	33.698	14.806
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.522			.093	.856	.359
Ovary	5609	Ratio	1.153*	1.105*	1.131*	1.031*	1.031*	1.022*
		t-test	17.649	15.263	16.282	6.634	5.451	3.928
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.395			.082	.780	
Pancreas	3426	Ratio	1.232*	1.158*	1.195*	1.049*	1.032*	1.024*
		t-test	20.503	17.298	18.502	7.899	4.606	3.340
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	0.001
		Fraction	.448	.405		.103	.716	

Prostate	1387	Ratio	1.130*	1.088*	1.106*	1.012	1.042*	.994
		t-test	7.583	6.648	6.730	NSE	3.587	NSE
		MC test	<0.001	<0.001	<0.001		<0.001	
		Fraction	.396	.327			.839	
Rectum	1975	Ratio	1.120*	1.086*	1.103*	1.032*	.973	1.055*
		t-test	8.026	7.207	7.400	3.960	NSE	6.052
		MC test	<0.001	<0.001	<0.001	<0.001		<0.001
		Fraction	.401	.368	.570	.094		
Skin	5027	Ratio	1.141*	1.099*	1.135*	1.032*	1.075*	.999
		t-test	15.681	14.045	16.676	6.394	11.708	NSE
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	
		Fraction	.364			.115		
Stomach	4960	Ratio	1.120*	1.061*	1.075*	.993	.954	1.018
		t-test	12.627	8.030	8.504	NSE	NSE	NSE
		MC test	<0.001	<0.001	<0.001			
		Fraction	.390	.349	.558			
Uterus	2221	Ratio	1.210*	1.158*	1.163*	1.035*	1.011	1.023
		t-test	14.995	13.702	12.451	4.621	NSE	NSE
		MC test	<0.001	<0.001	<0.001	<0.001		
		Fraction	.465	.442		.857		

NSE means that No Significant Excess of mutable motifs was found. The significance of the observed excess was measured using the Student *t*-test and Monte Carlo (MC) tests. The asterisk denotes that the corresponding $P < 0.0005$ (critical value = 3.291) is a conservative estimate of the critical overall value of the *t*-test taking into account the Bonferroni correction for multiple testing (16 x 6 = 96). “Ratio” is the mean weight of mutated sites divided by the mean weight of non-mutated sites. The predicted fraction of mutations induced by AID/APOBEC proteins (“Fraction”) is shown when a significant excess of somatic mutations in a mutable motif comparisons was detected; all cases where there was a significant difference between observed and expected distributions ($P > 0.05$) were discarded.

Supplementary Table S8. Correlation between AID/APOBEC mutable motifs and the context of C:G > A:T somatic mutations.

Cancer tissue type	#Muta-tions	Test	APO1	APO3A	APO3B	APO3C	APO3G	AID
Bladder	4603	Ratio	1.295*	1.259*	1.291*	1.088*	1.045*	1.020*
		t-test	29.792	32.106	32.703	15.633	7.384	3.348
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.522		.490	.116	.892	.444
Blood	2207	Ratio	.883	.938	.926	.990	.937	1.068*
		t-test	NSE	NSE	NSE	NSE	NSE	7.287
		MC test						<0.001
		Fraction						
Brain	4266	Ratio	.995	.995	1.015	1.011	1.004	1.042*
		t-test	NSE	NSE	NSE	NSE	NSE	6.606
		MC test						<0.001
		Fraction						
Breast	14962	Ratio	1.146*	1.124*	1.152*	1.046*	1.015*	1.019*
		t-test	27.304	28.109	31.022	15.073	4.222	5.583
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.411	.361		.122	.831	
Cervix	5427	Ratio	1.358*	1.314*	1.327*	1.138*	1.058*	1.058*
		t-test	40.229	43.802	41.438	26.786	10.963	11.523
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.594		.560	.180	.911	.460
Colon	36260	Ratio	1.149*	1.126*	1.177*	1.098*	1.093*	1.021*
		t-test	44.656	48.604	62.061	48.784	42.269	9.439
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.432	.366		.238	.872	.635
Kidney	9825	Ratio	1.071*	1.063*	1.099*	1.036*	1.041*	1.007
		t-test	11.351	12.512	17.191	9.867	9.186	NSE
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	
		Fraction	.326	.281		.117		
Liver	23626	Ratio	1.000	1.021*	1.039*	1.009*	.995	1.015*
		t-test	NSE	6.654	10.378	3.538	NSE	5.391
		MC test		<0.001	<0.001	<0.001		<0.001
		Fraction		.238		.098		
Lung	90050	Ratio	.991	1.005*	1.039*	.997	1.066*	.973
		t-test	NSE	3.376	21.854	NSE	42.629	NSE
		MC test		.081	<0.001		<0.001	
		Fraction		.199				
Ovary	5758	Ratio	.991	.971	.993	.987	.973	1.017
		t-test	NSE	NSE	NSE	NSE	NSE	2.940
		MC test						
		Fraction						
Pancreas	6195	Ratio	1.054*	1.038*	1.065*	1.026*	.990	1.015
		t-test	6.687	5.866	8.702	5.289	NSE	NSE
		MC test	<0.001	<0.001	<0.001	<0.001		
		Fraction	.337			.113		

Prostate	5810	Ratio	1.018*	1.075*	1.070*	1.052*	1.362*	.941
		t-test	2.715	14.553	12.186	12.186	53.228	NSE
		MC test	.018	<0.001	<0.001	<0.001	<0.001	
		Fraction	.143	.483		.680		
Rectum	7728	Ratio	1.290*	1.241*	1.272*	1.128*	1.022*	1.074*
		t-test	39.488	42.170	42.426	29.901	5.138	16.987
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.577	.509	.542	.294	.043	.708
Skin	8242	Ratio	1.145*	1.142*	1.189*	1.103*	1.252*	.969
		t-test	22.752	28.022	35.087	25.986	48.166	NSE
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	
		Fraction	.322	.239	.253	.114		
Stomach	17810	Ratio	.995	.995	1.056*	1.041*	1.116*	.972
		t-test	NSE	NSE	14.678	14.868	34.117	NSE
		MC test			<0.001	<0.001	<0.001	
		Fraction				.176		
Uterus	11903	Ratio	1.160*	1.136*	1.169*	1.106*	1.085*	1.059*
		t-test	28.320	30.494	34.360	31.664	22.785	16.397
		MC test	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
		Fraction	.473	.408	.416		.910	.687

NSE means that No Significant Excess of mutable motifs was found. The significance of an excess was measured using the Student *t*-test and Monte Carlo (MC) tests. The asterisk denotes that the corresponding $P < 0.0005$ (critical value = 3.291) is a conservative estimate of the critical overall value of the *t*-test taking into account the Bonferroni correction for multiple testing (16 x 6 = 96). “Ratio” is the mean weight of mutated sites divided by the mean weight of non-mutated sites. The predicted fraction of mutations induced by AID/APOBEC proteins (“Fraction”) is shown when a significant excess of somatic mutations in mutable motif comparisons was detected; all cases where there was a significant difference between observed and expected distributions ($P > 0.05$) were discarded.

Supplementary Figure S1. AID/APOBEC weight matrices W(b,j).

Position	-3 (3)	-2 (4)	-1 (5)	C Mutation	+1 (7)	+2 (8)	+3 (9)
APOBEC1							
A	-.857	-.314	-2.466	-	.052	.154	.119
T	1.013	.554	1.298	-	.444	.307	.168
G	-1.316	-.820	-4.361	-	-.623	-.381	-.353
C	-.779	.074	-2.571	-	-.196	-.260	-.012
Smax = 5.148							
Smin = -9.409							
APOBEC3A							
A	-.160	-.050	-2.615	-	.476	.243	-.039
T	.396	.516	1.268	-	-.090	.256	.219
G	-.559	-1.222	-2.569	-	.098	-.258	-.258
C	.090	.046	-1.917	-	-1.041	-.454	.040
Smax = 4.025							
Smin = -7.082							
APOBEC3B							
A	-.032	.261	-2.799	-	.400	.261	-.026
T	.330	.176	1.252	-	.039	.352	.218
G	-.401	-.153	-3.554	-	-.229	-.335	-.273
C	-.030	-.456	-1.294	-	-.429	-.655	.039
Smax = 3.807							
Smin = -7.098							
APOBEC3C							
A	.239	-.499	-.332	-	.399	.362	.239
T	.286	1.020	.664	-	.053	.104	.244
G	-.473	-1.390	-.697	-	-.003	.114	-.211
C	-.293	-1.392	-.213	-	-.832	-1.169	-.476
Smax = 3.914							
Smin = -7.233							
APOBEC3G							
A	.290	-.933	-3.004	-	.695	.249	.160
T	.488	.043	-1.132	-	-.193	.188	.225
G	-.479	-2.500	-3.982	-	-.254	-.132	-.349
C	-1.071	.954	1.349	-	-1.093	-.481	-.151
Smax = 4.798							
Smin =-10.496							
AID							
A	.404	.774	.391	-	.141	.027	.072
T	.012	.218	-.311	-	.068	.095	.031
G	-.321	-2.734	.342	-	.119	.007	-.107
C	-.294	-.944	-.917	-	-.401	-.109	.026
Smax = 2.347							
Smin = -5.001							

Supplementary Figure S2. The overall distribution of fraction of somatic C:G > T:A mutations associated with AID/APOBEC deamination (APOBEC1, APOBEC3A, ABOPEC3B and AID deaminases).