Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process.

Dorota Sikora*, Lynda Rocheleau*, Earl G. Brown and Martin Pelchat[1]

Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, K1H 8M5, Canada.

* These authors contributed equally to this work.

[1]Corresponding author: Phone: (613) 562-5800 ext. 8846

Fax: (613) 562-5452

E-mail: mpelchat@uottawa.ca

Table S1: Relative numbers of each nucleotide present at the 5’ ends of viral mRNAs obtained from infected A549 cells. The nucleotide positions (-15 to 7) are numbered according to the host/virus junction (vertical line), defined as the phosphodiester bond located upstream of position G2 on IAV mRNA.

IAV		Nucleotide position																				
mRNA		-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	2	3	4	5	6	7
PB2																						
	A	13625	88000	121718	452994	849610	1064018	582189	593672	476726	569818	475635	449721	536516	849081	630008	2823	410	1171	3064484	3057845	3062793
	C	31	6285	114269	172368	484324	444930	680283	719954	782091	845340	955219	663750	922677	813331	936907	486	3063577	1761	89	2606	1286
	G	6250	16153	98065	301677	481185	752497	1053829	1008333	1120834	1042348	1190272	1432314	1137117	916080	1394606	3057250	97	3056213	241	4106	262
	T	1164	14596	48118	90112	259795	548051	510035	509274	503658	485467	385543	516035	468579	486419	103396	4358	833	5772	103	360	576
PB1																						
	A	165207	146591	607107	1027360	1786258	1757484	1134943	864487	1103671	891809	1155491	1039597	1206296	1277402	1940919	7560	1774	17880	5391700	5394367	5394220
	C	33444	153739	128837	496753	840891	1152777	1770044	2015569	1786427	1844806	1423703	1571602	1575591	1931770	1404741	1045	5394090	982	3498	971	1213
	G	29628	109982	194509	339320	821202	1061459	1220738	1370349	1241922	1321580	1255485	1544127	1449690	1110034	1707750	5379740	137	5376205	1159	1443	1338
	T	56707	78436	175436	336158	663216	1328104	1219490	1106466	1246215	1325181	1555979	1241890	1165644	1078015	343871	8936	1280	2214	924	500	510
PA																						
	A	37874	36700	212453	334702	263567	244529	188997	207956	161945	209885	174014	136773	97958	155868	664562	2310	155	1315278	1313766	1314659	1314898
	C	9246	26636	52173	135176	247906	279455	259500	305543	251793	339049	493392	375727	362402	825356	279306	78	1315629	173	753	620	595
	G	18523	23901	82771	187897	290840	301828	377876	252860	429741	429568	252916	483855	712104	115831	185252	1312991	140	322	1313	522	304
	T	4471	36461	47225	137309	218298	348219	353697	420875	366280	292029	372777	316884	142551	217962	186909	650	105	256	197	228	232
HA																						
	A	76444	120163	408742	642800	849507	742733	723245	623808	489089	447842	500611	390464	281935	531228	1326146	2376	35	3144900	3131381	3138859	3148836
	C	25230	40925	122013	332257	404559	536907	610673	560664	691058	908884	989766	704012	855042	1952610	531280	1000	3148785	717	8063	3311	247
	G	10306	47774	214784	307979	602259	730763	876371	919072	1080656	1044693	824533	1317298	1569825	236940	843267	3142984	48	2127	8045	5297	451
	T	21739	98686	87732	333010	590979	873991	689830	822002	715090	660565	790468	734914	439920	428817	448921	3254	746	1870	2125	2147	80
NP																						
	A	29141	69450	230915	397380	462177	459641	350636	291722	238885	204047	232351	198854	131284	228247	775808	2089	13	1561322	1559873	1559386	1561445
	C	1973	12312	42693	91595	180742	212927	239975	284267	269355	381011	526799	362032	318849	1005255	333080	250	1561153	40	418	408	30
	G	9933	35216	96615	243754	319842	448054	549287	514273	714554	730841	501420	785028	982398	149216	306335	1558582	20	287	1164	1836	159
	T	4806	17045	43720	111384	199714	268044	258029	320857	215675	169674	264360	207517	123673	176478	146436	738	473	10	204	29	25
NA																						
	A	102183	234851	669689	1428329	1465091	1281649	926296	731923	719661	718993	550529	485741	401042	537850	2473796	10859	2169	5079960	5081092	5078171	5077466
	C	23945	71441	264386	576378	789442	1110058	822914	1202350	1270683	1482489	1616477	1488921	1313111	3625400	1154702	1297	5080064	861	1404	1513	796
	G	15764	160896	351844	599334	1060309	1481549	1808305	1808390	1997272	2007527	1917314	2168715	2981418	447570	1007831	5072086	908	3028	1810	4035	6460
	T	17997	45812	232260	423496	863571	866770	1204095	1059610	865184	731577	923637	932059	383687	473968	448467	554	1655	947	490	1077	74
M																						
	A	46457	135414	330537	573580	711801	817277	557786	596429	431646	383429	451959	377331	305844	604263	1092470	5297	407	2552985	2551650	2549502	2553296
	C	8541	39148	76143	184699	365840	417924	495561	439977	521266	784798	656864	522239	668320	1335260	642747	330	2552952	74	864	2042	30
	G	34514	53036	161376	335263	494734	559965	756270	730552	941316	928060	841476	1211203	1335013	317732	570904	2547803	62	768	1010	1813	485
	T	8377	13377	113707	226518	384789	570970	567316	630976	540495	398596	568903	434172	237959	294170	247728	419	428	22	325	492	38
NS																						
	A	91277	266786	855007	1733888	1956312	1740772	1417973	1149297	964743	798235	686189	735760	475282	804253	3020035	16556	444	6081400	6071218	6071710	6080066
	C	26032	92774	184880	465828	808641	1013811	1050855	1384678	1186863	1560167	2353993	1430482	1257807	4212273	1209861	1771	6081195	96	2046	3945	1329
	G	66408	117285	349566	871814	1107874	1678265	1874249	1627548	2489189	2752653	1800246	2863522	3851041	477949	1244681	6053797	312	1448	8214	7211	804
	T	35635	67871	248222	523080	970313	1221952	1336697	1557191	1158851	823275	1174711	1039352	492934	587410	608455	10908	1081	88	1554	166	833

Table S2: Relative numbers of each nucleotide present at the 5’ ends of viral mRNAs obtained from infected M-1 cells. The nucleotide positions (-15 to 7) are numbered according to the host/virus junction (vertical line), defined as the phosphodiester bond located upstream of position G2 on IAV mRNA.

IAV		Nucleotide position																				
mRNA		-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	2	3	4	5	6	7
PB2																						
	A	26887	61677	196513	450431	956437	1197022	663242	714769	631678	530174	491465	498172	358634	874361	594305	5132	684	1973	3280298	3268244	3278325
	C	3419	25551	49389	146045	381175	595407	737431	646736	779621	705343	990685	774104	905145	1016421	1013293	537	3278884	2535	78	6407	1556
	G	7225	27052	135056	280438	528010	733488	1121682	1183931	1226992	1431459	1257505	1612425	1604358	969377	1565830	3268858	56	3268709	297	5776	215
	T	4815	14215	40551	119777	270864	550812	601295	580377	494454	499242	496518	382666	407916	420476	107322	6223	1126	7533	77	323	654
PB1																						
	A	91812	289641	547734	1166896	1758050	1875985	1325485	1156329	1175589	1157927	1289375	1274160	1391811	1443586	2180196	8548	2933	24915	5863802	5869578	5870552
	C	12283	60407	236358	506078	1057071	1420388	1784297	1893767	1936176	1846372	1746836	1455389	1805050	2077721	1472346	7654	5868540	818	1716	1165	1271
	G	19503	65877	196276	336328	760796	1097174	1248907	1493665	1560158	1473824	1396263	1683915	1627194	1326311	1843137	5849522	150	5846052	6623	2025	1288
	T	14899	37119	117255	299936	651196	1306567	1423333	1243339	1137212	1340041	1401600	1435384	1024991	1004896	377489	7444	1545	1383	1027	400	57
PA																						
	A	55009	161124	575522	1154497	1219427	1530309	1142641	986105	737003	642988	658298	644562	595493	785219	2543381	11899	885	5158005	5139917	5154845	5155115
	C	9132	62716	184331	482379	727791	710653	989141	948974	895095	1211189	1722068	1412392	1411290	3158964	939482	442	5154471	897	8037	2683	4599
	G	33165	54974	298205	482409	823142	1142961	1377890	1411280	1993072	2023641	1258572	1900913	2661703	456120	992932	5145548	1162	1395	12743	1932	1031
	T	17075	39272	102406	403097	852445	1058207	1017211	1206198	1049224	946188	1314933	1132747	466556	742005	685784	3690	5061	1282	882	2119	834
HA																						
	A	33875	123703	331546	778178	1021900	1040344	834647	683714	618861	607170	608137	480650	367615	621744	1450127	5802	104	3703714	3702078	3693904	3710344
	C	10326	36678	123359	241407	480136	617043	554428	799749	755817	926755	1243625	740544	947103	2332163	747052	260	3710455	300	2108	4436	338
	G	18548	41150	161255	402040	544787	873317	1044729	1035625	1390034	1438781	1044403	1795634	2044492	366060	1102185	3703928	218	5797	6290	9122	484
	T	8982	42732	123294	266595	561270	778190	902503	833294	643894	553062	715243	676485	343048	387315	411846	1220	433	1399	734	3748	44
NP																						
	A	82600	271967	901214	1992417	2219692	2236844	1933219	1557609	1151479	926438	989711	873314	635782	1028919	3566548	12781	191	7471091	7453930	7450428	7472923
	C	13327	45532	178012	457736	787395	1071757	1059651	1270418	1264337	1544536	2603361	1736492	1312507	4901221	1751012	3171	7470085	418	5935	6034	369
	G	28679	124321	437792	1001787	1585957	2257133	2624166	2544999	3663295	3877324	2427068	3941999	5014811	768171	1546109	7451279	425	2912	12505	17963	1326
	T	13369	46354	133155	460669	973062	1223305	1204828	1489379	859777	780111	1256790	876362	491781	773488	611060	7498	4028	308	2359	304	111
NA																						
	A	93215	386616	850134	1327961	1475682	1418418	853640	867286	871324	826585	424112	745303	237436	819864	2666002	12989	2969	5151257	5152373	5151279	5149313
	C	14990	12189	141414	502129	568318	709667	1109775	1245615	1066377	1436329	2193223	1482569	1324634	3371239	1025288	1765	5151571	662	2347	1745	1138
	G	76218	180090	422041	731171	1315101	1665571	1834757	1854978	2122224	2152102	1611051	2154850	3165148	420181	1117372	5142284	1596	4866	2604	4476	7256
	T	14896	90298	87547	467557	931450	1063821	1104224	952705	907655	657776	899829	772652	428295	546477	349225	849	1751	1102	563	387	180
M																						
	A	75023	309137	736942	1323563	2073523	1963996	1540785	1361678	1091634	1159171	1126548	1003161	807798	1389086	2570946	10578	2648	6457717	6454399	6440672	6458021
	C	25431	73080	141282	410006	755785	1005963	1050708	1062112	1248786	1554673	1705103	1241223	1429357	3346467	1686896	1324	6456566	521	2709	13261	190
	G	40664	96516	303828	595488	1030319	1494477	1819644	2005295	2453207	2509627	2210520	2923268	3627442	968086	1425324	6447838	249	2600	2744	5253	2222
	T	33102	46869	170747	448866	709189	1317190	1453913	1475042	1226288	994129	1283162	1274187	587237	753953	777779	1205	1482	107	1093	1759	512
NS																						
	A	75043	160620	874337	1920229	2016424	2009849	1717722	1260879	908683	917045	827430	753051	520922	925347	3184554	14417	663	6682654	6668840	6670511	6681355
	C	20878	49886	127398	514021	848253	1090443	1108214	1272446	1331711	1544100	2361831	1465776	1211785	4357620	1408238	2090	6684579	199	5117	5465	1593
	G	56181	104394	316619	863332	1317658	1992987	2306619	2290016	3112427	3282623	2297862	3585894	4482207	716044	1543650	6659033	284	3601	11072	10270	2759
	T	16805	59006	122785	402173	1015952	1163693	1167551	1504986	1051583	782460	1123871	872440	469121	686978	550069	10971	985	57	1482	265	804

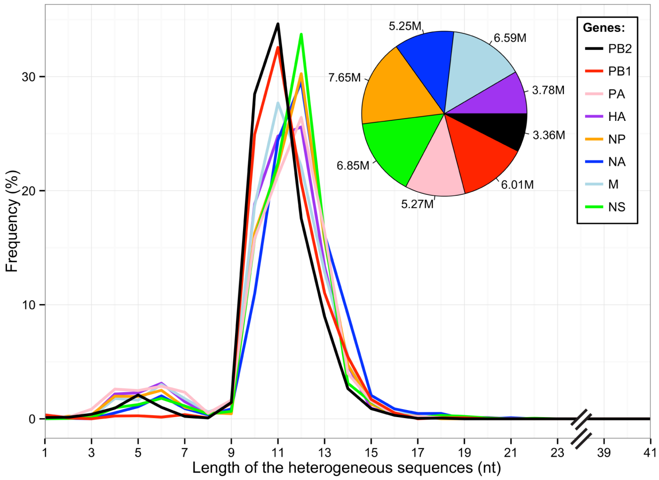
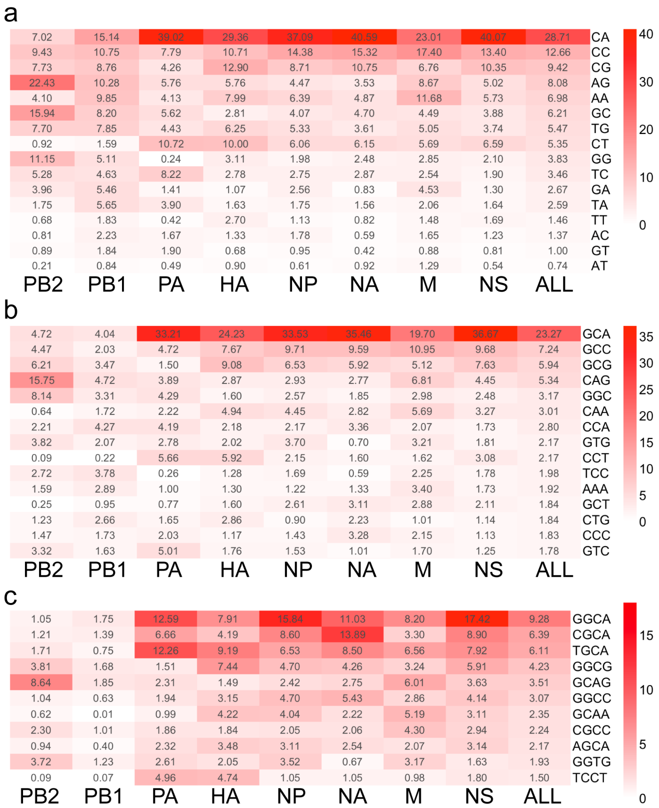


Figure S1: Length distribution of the heterogeneous sequences obtained from infected M-1 cells. All sequences representing each of the eight IAV transcripts isolated from infected M-1 cells and located between the ligated RACE primer and position G2 on IAV mRNA are included. Inset: proportion of IAV sequences corresponding to each transcript obtained following high-throughput sequencing.



a

PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
7.02	15.14	39.02	29.36	37.09	40.59	23.01	40.07	28.71	CA
9.43	10.75	7.79	10.71	14.38	15.32	17.40	13.40	12.66	CC
7.73	8.76	4.26	12.90	8.71	10.75	6.76	10.35	9.42	CG
22.43	10.28	5.76	5.76	4.47	3.53	8.67	5.02	8.08	AG
4.10	9.85	4.13	7.99	6.39	4.87	11.68	5.73	6.98	AA
15.94	8.20	5.62	2.81	4.07	4.70	4.49	3.88	6.21	GC
7.70	7.85	4.43	6.25	5.33	3.61	5.05	3.74	5.47	TG
0.92	1.59	10.72	10.00	6.06	6.15	5.69	6.59	5.35	CT
11.15	5.11	0.24	3.11	1.98	2.48	2.85	2.10	3.83	GG
5.28	4.63	8.22	2.78	2.75	2.87	2.54	1.90	3.46	TC
3.96	5.46	1.41	1.07	2.56	0.83	4.53	1.30	2.67	GA
1.75	5.65	3.90	1.63	1.75	1.56	2.06	1.64	2.59	TA
0.68	1.83	0.42	2.70	1.13	0.82	1.48	1.69	1.46	TT
0.81	2.23	1.67	1.33	1.78	0.59	1.65	1.23	1.37	AC
0.89	1.84	1.90	0.68	0.95	0.42	0.88	0.81	1.00	GT
0.21	0.84	0.49	0.90	0.61	0.92	1.29	0.54	0.74	AT

b

PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
4.72	4.04	33.21	24.23	33.53	35.46	19.70	36.67	23.27	GCA
4.47	2.03	4.72	7.67	9.71	9.59	10.95	9.68	7.24	GCC
6.21	3.47	1.50	9.08	6.53	5.92	5.12	7.63	5.94	GCG
15.75	4.72	3.89	2.87	2.93	2.77	6.81	4.45	5.34	CAG
8.14	3.31	4.29	1.60	2.57	1.85	2.98	2.48	3.17	GGC
0.64	1.72	2.22	4.94	4.45	2.82	5.69	3.27	3.01	CAA
2.21	4.27	4.19	2.18	2.17	3.36	2.07	1.73	2.80	CCA
3.82	2.07	2.78	2.02	3.70	0.70	3.21	1.81	2.17	GTG
0.09	0.22	5.66	5.92	2.15	1.60	1.62	3.08	2.17	CCT
2.72	3.78	0.26	1.28	1.69	0.59	2.25	1.78	1.98	TCC
1.59	2.89	1.00	1.30	1.22	1.33	3.40	1.73	1.92	AAA
0.25	0.95	0.77	1.60	2.61	3.11	2.88	2.11	1.84	GCT
1.23	2.66	1.65	2.86	0.90	2.23	1.01	1.14	1.84	CTG
1.47	1.73	2.03	1.17	1.43	3.28	2.15	1.13	1.83	CCC
3.32	1.63	5.01	1.76	1.53	1.01	1.70	1.25	1.78	GTC

c

PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
1.05	1.75	12.59	7.91	15.84	11.03	8.20	17.42	9.28	GGCA
1.21	1.39	6.66	4.19	8.60	13.89	3.30	8.90	6.39	CGCA
1.71	0.75	12.26	9.19	6.53	8.50	6.56	7.92	6.11	TGCA
3.81	1.68	1.51	7.44	4.70	4.26	3.24	5.91	4.23	GGCG
8.64	1.85	2.31	1.49	2.42	2.75	6.01	3.63	3.51	GCAG
1.04	0.63	1.94	3.15	4.70	5.43	2.86	4.14	3.07	GGCC
0.62	0.01	0.99	4.22	4.04	2.22	5.19	3.11	2.35	GCAA
2.30	1.01	1.86	1.84	2.05	2.06	4.30	2.94	2.24	CGCC
0.94	0.40	2.32	3.48	3.11	2.54	2.07	3.14	2.17	AGCA
3.72	1.23	2.61	2.05	3.52	0.67	3.17	1.63	1.93	GGTG
0.09	0.07	4.96	4.74	1.05	1.05	0.98	1.80	1.50	TCCT

Figure S2: Analysis of the nucleotide motifs present at the 3' ends of the heterogeneous sequences of viral mRNAs in human cells. A heatmap is shown to compare the frequencies (%) of all possible dinucleotide (**a**), and trinucleotide (**b**) and tetranucleotide (**c**) motifs found at the 3' ends of the heterogeneous sequences used by each IAV mRNA. For (**b**) and (**c**), only the motifs which were found ~5% or more of the time by at least one of the transcripts are presented. All motifs are represented in the 5' to 3' orientation.

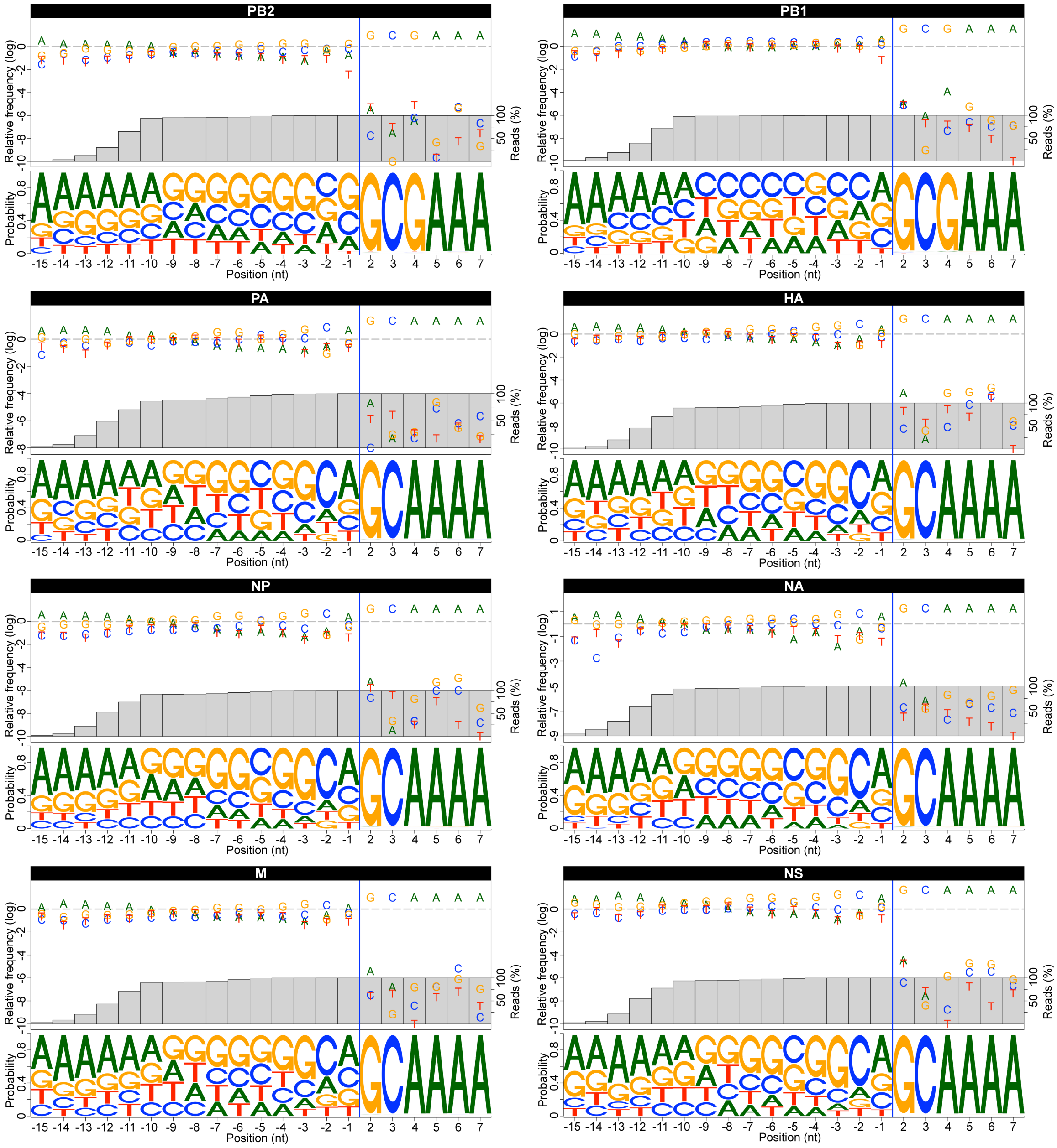
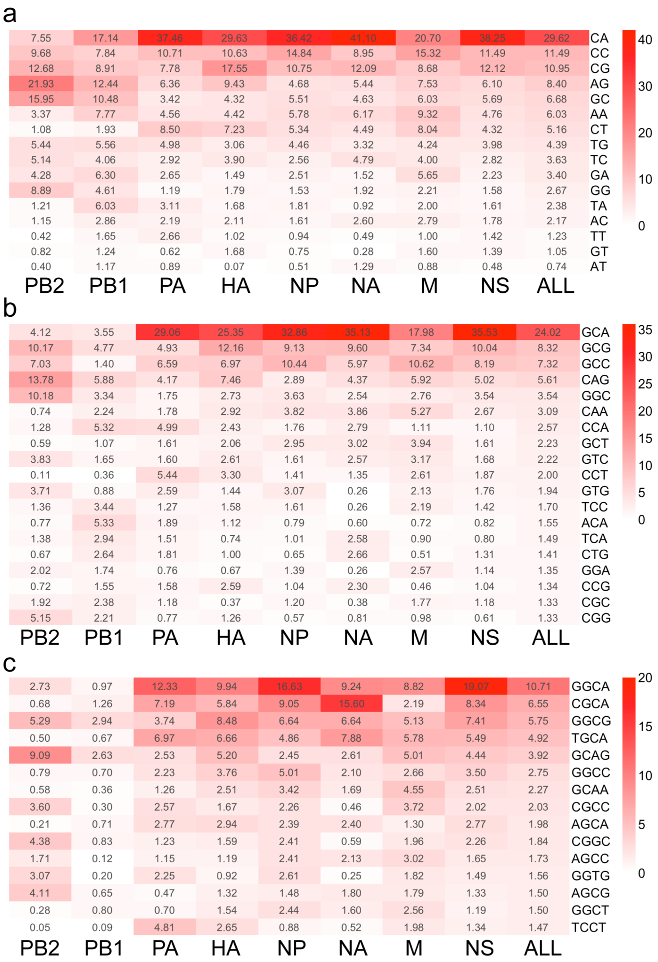


Figure S3: Sequence variation present in the heterogeneous sequences located at the 5'-end of viral mRNAs obtained from infected M-1 cells. For each transcript, the top panel shows the relative nucleotide frequency at each position (log scaled), and the bottom panel shows the Logo representation of the nucleotide variation. The nucleotides are numbered according to the host/virus junction (blue line), defined as the phosphodiester bond located upstream of position G2 on IAV mRNA. The grey bars represent the percentage of the population of reads included in the calculation. Only sequence reads that appeared at least twice were used in the analysis. For all IAV mRNA, only nucleotide frequencies observed from 15 nt upstream of the G2 and up to 5 nt downstream of G2 are represented. All motifs are represented in the 5' to 3' orientation.



a

PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
7.55	17.14	37.46	29.63	36.42	41.10	20.70	38.25	29.62	CA
9.68	7.84	10.71	10.63	14.84	8.95	15.32	11.49	11.49	CC
12.68	8.91	7.78	17.55	10.75	12.09	8.68	12.12	10.95	CG
21.93	12.44	6.36	9.43	4.68	5.44	7.53	6.10	8.40	AG
15.95	10.48	3.42	4.32	5.51	4.63	6.03	5.69	6.68	GC
3.37	7.77	4.56	4.42	5.78	6.17	9.32	4.76	6.03	AA
1.08	1.93	8.50	7.23	5.34	4.49	8.04	4.32	5.16	CT
5.44	5.56	4.98	3.06	4.46	3.32	4.24	3.98	4.39	TG
5.14	4.06	2.92	3.90	2.56	4.79	4.00	2.82	3.63	TC
4.28	6.30	2.65	1.49	2.51	1.52	5.65	2.23	3.40	GA
8.89	4.61	1.19	1.79	1.53	1.92	2.21	1.58	2.67	GG
1.21	6.03	3.11	1.68	1.81	0.92	2.00	1.61	2.38	TA
1.15	2.86	2.19	2.11	1.61	2.60	2.79	1.78	2.17	AC
0.42	1.65	2.66	1.02	0.94	0.49	1.00	1.42	1.23	TT
0.82	1.24	0.62	1.68	0.75	0.28	1.60	1.39	1.05	GT
0.40	1.17	0.89	0.07	0.51	1.29	0.88	0.48	0.74	AT

b

PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
4.12	3.55	29.06	25.35	32.86	35.13	17.98	35.53	24.02	GCA
10.17	4.77	4.93	12.16	9.13	9.60	7.34	10.04	8.32	GCG
7.03	1.40	6.59	6.97	10.44	5.97	10.62	8.19	7.32	GCC
13.78	5.88	4.17	7.46	2.89	4.37	5.92	5.02	5.61	CAG
10.18	3.34	1.75	2.73	3.63	2.54	2.76	3.54	3.54	GGC
0.74	2.24	1.78	2.92	3.82	3.86	5.27	2.67	3.09	CAA
1.28	5.32	4.99	2.43	1.76	2.79	1.11	1.10	2.57	CCA
0.59	1.07	1.61	2.06	2.95	3.02	3.94	1.61	2.23	GCT
3.83	1.65	1.60	2.61	1.61	2.57	3.17	1.68	2.22	GTC
0.11	0.36	5.44	3.30	1.41	1.35	2.61	1.87	2.00	CCT
3.71	0.88	2.59	1.44	3.07	0.26	2.13	1.76	1.94	GTG
1.36	3.44	1.27	1.58	1.61	0.26	2.19	1.42	1.70	TCC
0.77	5.33	1.89	1.12	0.79	0.60	0.72	0.82	1.55	ACA
1.38	2.94	1.51	0.74	1.01	2.58	0.90	0.80	1.49	TCA
0.67	2.64	1.81	1.00	0.65	2.66	0.51	1.31	1.41	CTG
2.02	1.74	0.76	0.67	1.39	0.26	2.57	1.14	1.35	GGA
0.72	1.55	1.58	2.59	1.04	2.30	0.46	1.04	1.34	CCG
1.92	2.38	1.18	0.37	1.20	0.38	1.77	1.18	1.33	CGC
5.15	2.21	0.77	1.26	0.57	0.81	0.98	0.61	1.33	CGG

c

PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
2.73	0.97	12.33	9.94	16.63	9.24	8.82	19.07	10.71	GGCA
0.68	1.26	7.19	5.84	9.05	15.60	2.19	8.34	6.55	CGCA
5.29	2.94	3.74	8.48	6.64	6.64	5.13	7.41	5.75	GGCG
0.50	0.67	6.97	6.66	4.86	7.88	5.78	5.49	4.92	TGCA
9.09	2.63	2.53	5.20	2.45	2.61	5.01	4.44	3.92	GCAG
0.79	0.70	2.23	3.76	5.01	2.10	2.66	3.50	2.75	GGCC
0.58	0.36	1.26	2.51	3.42	1.69	4.55	2.51	2.27	GCAA
3.60	0.30	2.57	1.67	2.26	0.46	3.72	2.02	2.03	CGCC
0.21	0.71	2.77	2.94	2.39	2.40	1.30	2.77	1.98	AGCA
4.38	0.83	1.23	1.59	2.41	0.59	1.96	2.26	1.84	CGGC
1.71	0.12	1.15	1.19	2.41	2.13	3.02	1.65	1.73	AGCC
3.07	0.20	2.25	0.92	2.61	0.25	1.82	1.49	1.56	GGTG
4.11	0.65	0.47	1.32	1.48	1.80	1.79	1.33	1.50	AGCG
0.28	0.80	0.70	1.54	2.44	1.60	2.56	1.19	1.50	GGCT
0.05	0.09	4.81	2.65	0.88	0.52	1.98	1.34	1.47	TCCT

Figure S4: Analysis of the nucleotide motifs present at the 3’ ends of the heterogeneous sequences of IAV mRNAs in mouse cells. A heatmap is shown to compare the frequencies (%) of all possible dinucleotide (**a**), and trinucleotide (**b**) and tetranucleotide (**c**) motifs found at the 3’ ends of the heterogeneous sequences used by each IAV mRNA. For (**b**) and (**c**), only the motifs which were found ~5% or more of the time by at least one of the transcripts are presented. All motifs are represented in the 5’ to 3’ orientation.

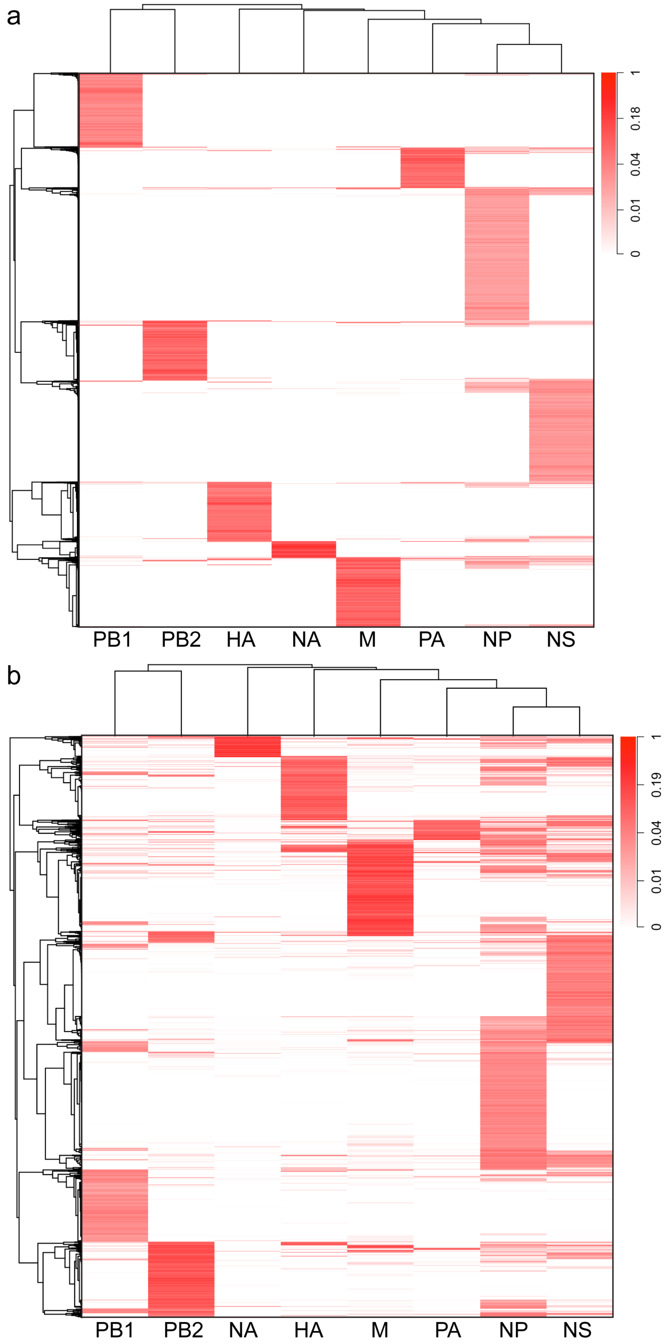


Figure S5: Mouse sequences/genes used by the eight viral transcripts. (**a**) Heatmap representation showing the distribution of the heterogeneous sequences (vertical axis) associated with each IAV transcript. (**b**) Heatmap representation showing the distribution of the host genes (vertical axis) potentially used by each IAV transcript. The heterogeneous sequences/host genes are hierarchically clustered. Data in each column have been normalized to the maximum value in that column. To simplify the representation, only sequences/genes that appears at least 1000 times were used.

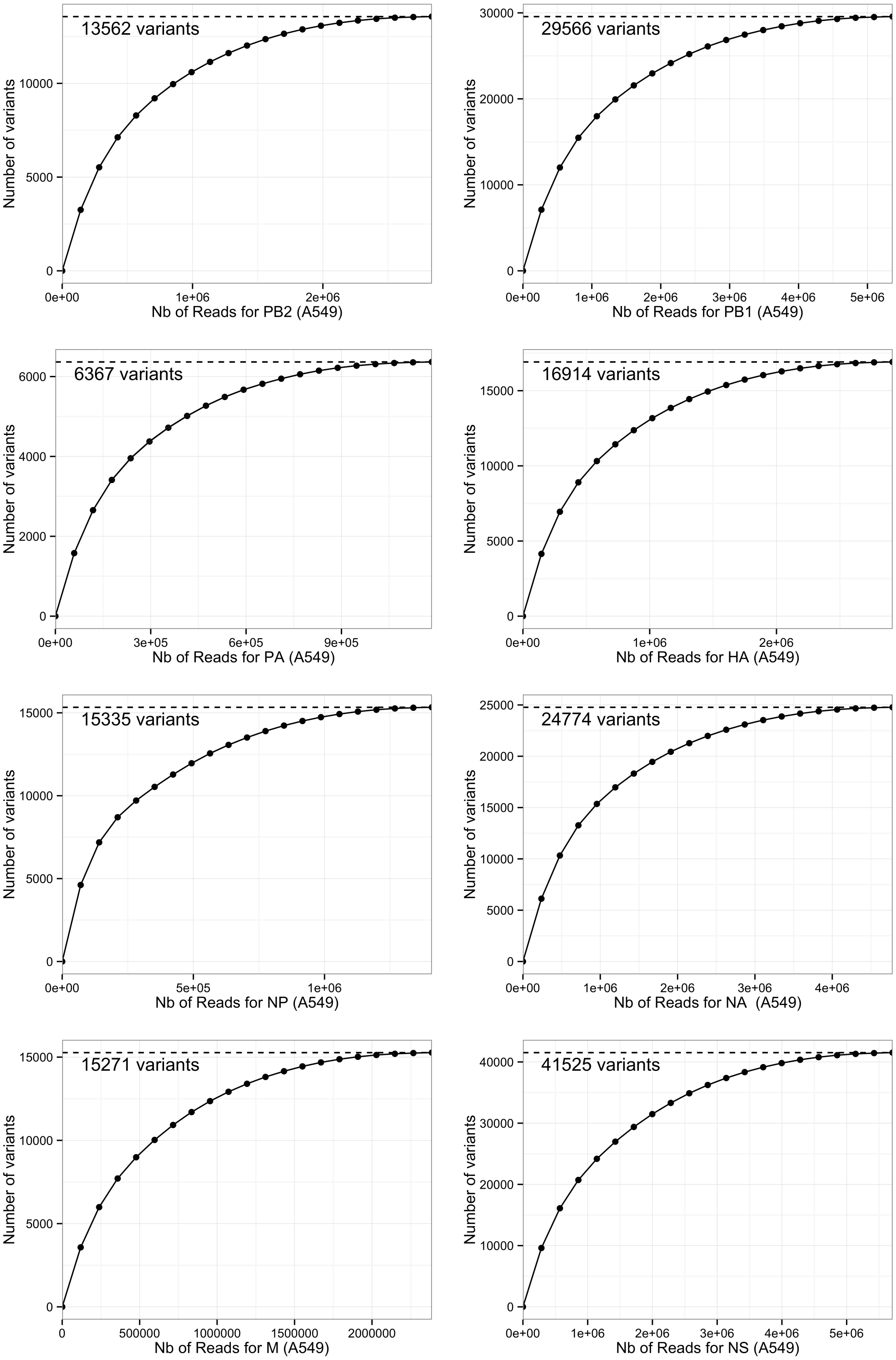


Figure S6: Rarefaction curves of the human heterogeneous sequences associated with each viral transcript. The number of unique heterogeneous sequence variants is plotted as a function of the number of reads obtained from IAV-infected human (A549) cells. Only sequence reads that appeared at least twice were used in the analysis.

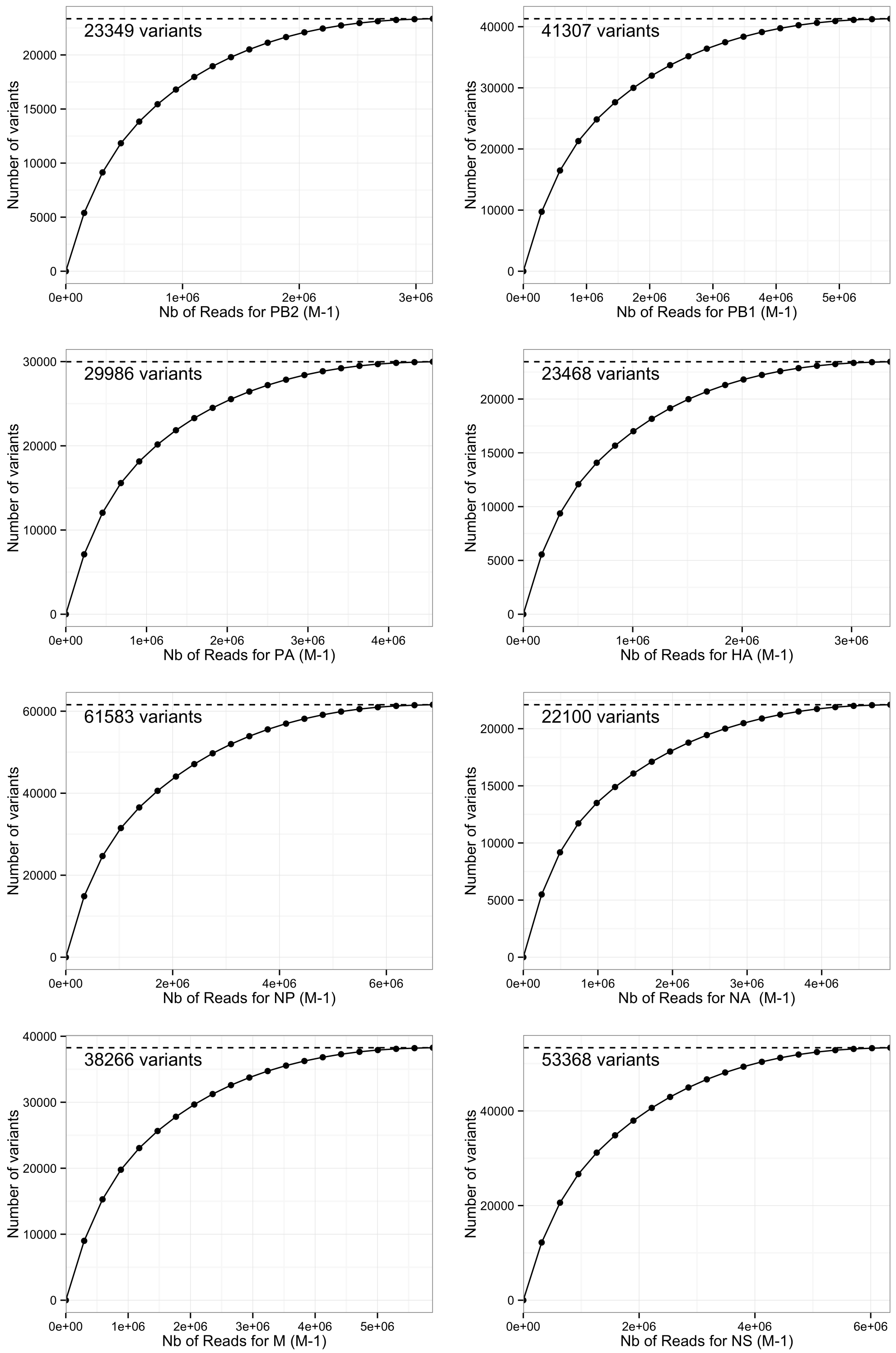
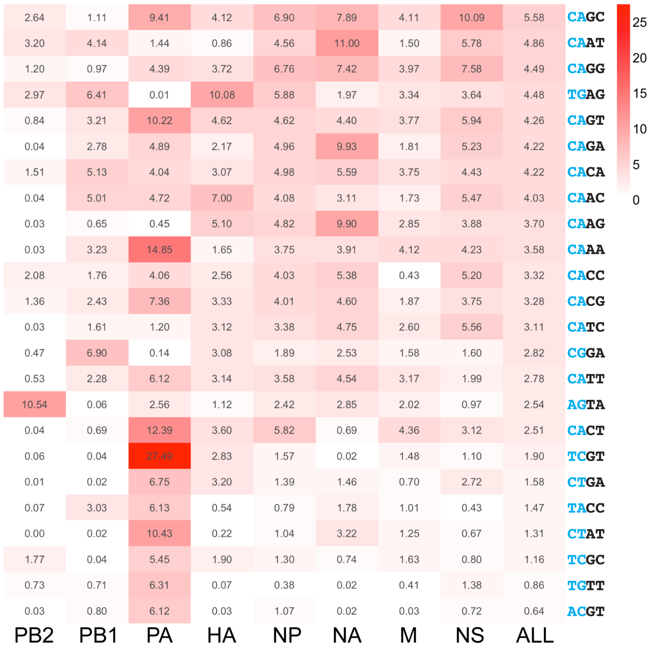


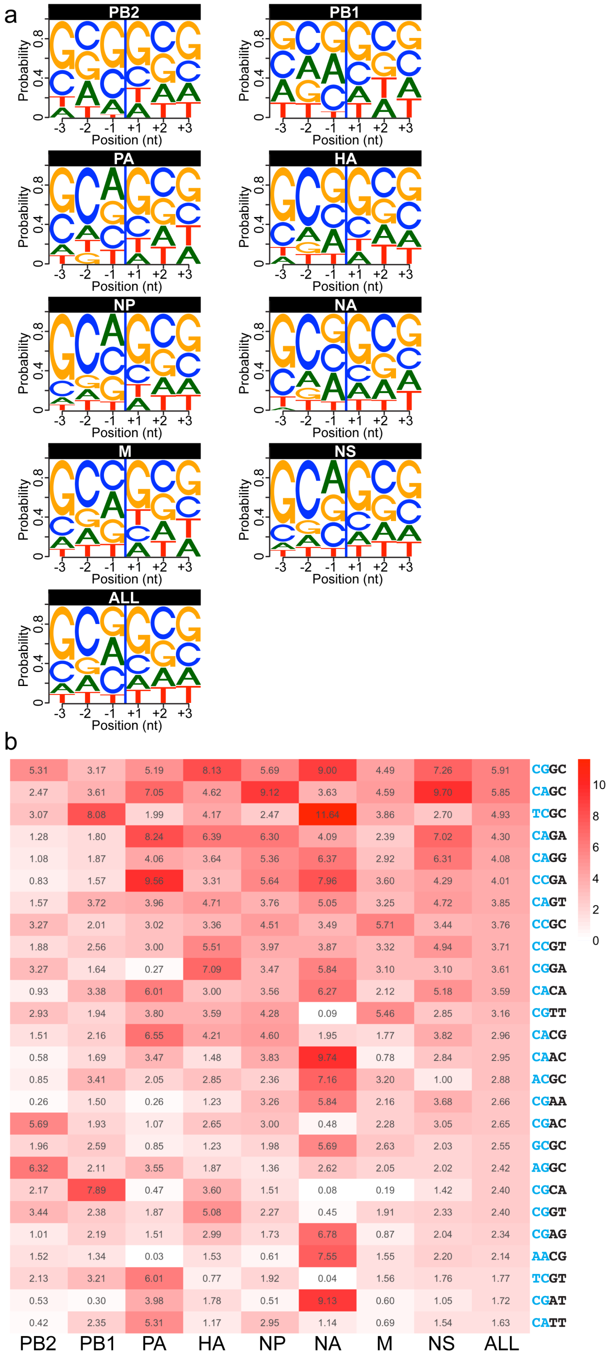
Figure S7: Rarefaction curves of the mouse heterogeneous sequences associated with each viral transcript. The number of unique heterogeneous sequence variants is plotted as a function of the number of reads obtained from IAV-infected mouse (M-1) cells. Only sequence reads that appeared at least twice were used in the analysis.

Figure S8: Enrichment of Gene Ontology (GO) term corresponding to the genes targeted by influenza A/HongKong/1/1968 (H3N2) virus cap-snatching in mouse cells. Data are presented as level 3 GO categorization for biological process. Data in each column have been normalized to the maximum value in that column. To account for bias due to the gene set we used, mRNA expression profile of mock infected cells (i.e. "MOCK) was used for the GO terms enrichment analysis.



PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
2.64	1.11	9.41	4.12	6.90	7.89	4.11	10.09	5.58	CAGC
3.20	4.14	1.44	0.86	4.56	11.00	1.50	5.78	4.86	CAAT
1.20	0.97	4.39	3.72	6.76	7.42	3.97	7.58	4.49	CAGG
2.97	6.41	0.01	10.08	5.88	1.97	3.34	3.64	4.48	TGAG
0.84	3.21	10.22	4.62	4.62	4.40	3.77	5.94	4.26	CAGT
0.04	2.78	4.89	2.17	4.96	9.93	1.81	5.23	4.22	CAGA
1.51	5.13	4.04	3.07	4.98	5.59	3.75	4.43	4.22	CACA
0.04	5.01	4.72	7.00	4.08	3.11	1.73	5.47	4.03	CAAC
0.03	0.65	0.45	5.10	4.82	9.90	2.85	3.88	3.70	CAAG
0.03	3.23	14.85	1.65	3.75	3.91	4.12	4.23	3.58	CAAA
2.08	1.76	4.06	2.56	4.03	5.38	0.43	5.20	3.32	CACC
1.36	2.43	7.36	3.33	4.01	4.60	1.87	3.75	3.28	CACG
0.03	1.61	1.20	3.12	3.38	4.75	2.60	5.56	3.11	CATC
0.47	6.90	0.14	3.08	1.89	2.53	1.58	1.60	2.82	CGGA
0.53	2.28	6.12	3.14	3.58	4.54	3.17	1.99	2.78	CATT
10.54	0.06	2.56	1.12	2.42	2.85	2.02	0.97	2.54	AGTA
0.04	0.69	12.39	3.60	5.82	0.69	4.36	3.12	2.51	CACT
0.06	0.04	27.49	2.83	1.57	0.02	1.48	1.10	1.90	TCGT
0.01	0.02	6.75	3.20	1.39	1.46	0.70	2.72	1.58	CTGA
0.07	3.03	6.13	0.54	0.79	1.78	1.01	0.43	1.47	TACC
0.00	0.02	10.43	0.22	1.04	3.22	1.25	0.67	1.31	CTAT
1.77	0.04	5.45	1.90	1.30	0.74	1.63	0.80	1.16	TCGC
0.73	0.71	6.31	0.07	0.38	0.02	0.41	1.38	0.86	TGTT
0.03	0.80	6.12	0.03	1.07	0.02	0.03	0.72	0.64	ACGT

Figure S9: Analysis of nucleotide sequences enriched at the cleavage site of human pre-mRNA. Heatmap representation of the frequencies (%) of tetranucleotide motifs composed of the last two nucleotides of the heterogeneous sequences (blue) and the next two nucleotides of the host pre-mRNA (black). The "ALL" dataset represents the sum of all the others dataset. All motifs are represented in the 5' to 3' orientation.



PB2	PB1	PA	HA	NP	NA	M	NS	ALL	
5.31	3.17	5.19	8.13	5.69	9.00	4.49	7.26	5.91	CGGC
2.47	3.61	7.05	4.62	9.12	3.63	4.59	9.70	5.85	CAGC
3.07	8.08	1.99	4.17	2.47	11.64	3.86	2.70	4.93	TCGC
1.28	1.80	8.24	6.39	6.30	4.09	2.39	7.02	4.30	CAGA
1.08	1.87	4.06	3.64	5.36	6.37	2.92	6.31	4.08	CAGG
0.83	1.57	9.56	3.31	5.64	7.96	3.60	4.29	4.01	CCGA
1.57	3.72	3.96	4.71	3.76	5.05	3.25	4.72	3.85	CAGT
3.27	2.01	3.02	3.36	4.51	3.49	5.71	3.44	3.76	CCGC
1.88	2.56	3.00	5.51	3.97	3.87	3.32	4.94	3.71	CCGT
3.27	1.64	0.27	7.09	3.47	5.84	3.10	3.10	3.61	CGGA
0.93	3.38	6.01	3.00	3.56	6.27	2.12	5.18	3.59	CACA
2.93	1.94	3.80	3.59	4.28	0.09	5.46	2.85	3.16	CGTT
1.51	2.16	6.55	4.21	4.60	1.95	1.77	3.82	2.96	CACG
0.58	1.69	3.47	1.48	3.83	9.74	0.78	2.84	2.95	CAAC
0.85	3.41	2.05	2.85	2.36	7.16	3.20	1.00	2.88	ACGC
0.26	1.50	0.26	1.23	3.26	5.84	2.16	3.68	2.66	CGAA
5.69	1.93	1.07	2.65	3.00	0.48	2.28	3.05	2.65	CGAC
1.96	2.59	0.85	1.23	1.98	5.69	2.63	2.03	2.55	GCGC
6.32	2.11	3.55	1.87	1.36	2.62	2.05	2.02	2.42	AGGC
2.17	7.89	0.47	3.60	1.51	0.08	0.19	1.42	2.40	CGCA
3.44	2.38	1.87	5.08	2.27	0.45	1.91	2.33	2.40	CGGT
1.01	2.19	1.51	2.99	1.73	6.78	0.87	2.04	2.34	CGAG
1.52	1.34	0.03	1.53	0.61	7.55	1.55	2.20	2.14	AACG
2.13	3.21	6.01	0.77	1.92	0.04	1.56	1.76	1.77	TCGT
0.53	0.30	3.98	1.78	0.51	9.13	0.60	1.05	1.72	CGAT
0.42	2.35	5.31	1.17	2.95	1.14	0.69	1.54	1.63	CATT

Figure S10: Analysis of nucleotide sequences enriched at the cleavage site in targeted mouse pre-mRNAs. (**a**) Logo representation of the last three nucleotides of the heterogeneous sequence and the next three nucleotides found in the host pre-mRNA downstream of the heterogeneous sequences. The nucleotides are numbered according to the heterogeneous sequences/pre-mRNA junction (blue line). (**b**) Heatmap representation of the frequencies (%) of tetranucleotide motifs composed of the last two nucleotides of the heterogeneous sequences (blue) and the next two nucleotides of the host pre-mRNA (black). The "ALL" dataset represents the sum of all the individual datasets. All motifs are represented in the 5' to 3' orientation.