Supplemental Table 1 Genomic coordinates of UCEs.

Because of the volume of information it contains, **Supplemental Table 1** is provided as a separate electronic file.

Supplemental Table 2: Distribution of human CNVs and INDELs among the alignable, nonrepetitive, intergenic, and genic portions of the human genome.

Data set			Percentage of genome					Intergenic/
Type	Source	Subset*	Alignable	Non-rpt	Intergenic	Intronic	Exonic	Genic
Genome			35	47	66	30	3	1.98
SDs	Eichler + Scherer	Complete	25	45	77	17	3	3.86
CNVs	Redon	Complete	30	46	72	24	3	2.73
		Recurrent	29	46	72	23	3	2.79
		Non-recurrent	34	47	71	25	2	2.60
	Wong	Complete	37	49	65	31	3	1.91
		Recurrent	34	49	63	31	4	1.82
	Simon-Sanchez	Complete	32	48	72	25	3	2.57
	de Smith	Complete	29	46	65	30	3	1.96
		Recurrent	28	45	66	28	3	2.07
		Non-recurrent	31	46	63	32	3	1.78
	Zogopoulos	Recurrent	30	49	68	26	3	2.31
	Korbel	Complete	27	42	65	31	4	1.88
	Pinto	Complete	28	46	78	19	2	3.69
	Wang	Complete	29	45	75	21	2	3.35
	Jakobsson	Recurrent	31	48	68	27	4	2.17
	Perry	Complete	25	44	73	18	3	3.45
		Recurrent	24	44	73	17	3	3.70
		Non-recurrent	28	44	72	23	3	2.79
	Kidd	Complete	25	42	75	20	2	3.43
INDELs	Mills	Complete	31	42	66	30	3	1.96
		Recurrent	23	29	69	28	3	2.28

The SD and CNV data sets are skewed with respect to the alignable portion of the human genome, defined as the 35% that is orthologous to the genome of the mouse. (Note that we have previously referred to this alignable portion as the "conserved" portion of the genome (DERTI *et al.* 2006).) This bias, however, cannot fully explain the strong depletion of UCEs among SDs and CNVs (**Supplemental Table 3**; also see DERTI *et al.* 2006). The percentages of the human genome, SDs, CNVs and INDELs that are repetitive are generally similar, therefore repetitive content is unlikely to underlie the depletion of UCEs among SDs, CNVs, and INDELs. This has been shown directly with regard to the depletion of UCEs from SDs (DERTI *et al.* 2006). The SD data set represents the union of those from Eichler and colleagues (SHE *et al.* 2004) and Scherer and colleagues (CHEUNG *et al.* 2003). *Recurrence denotes CNVs observed in more than one

individual in a study population, except in the case of Wong *et al.* (2007), in which recurrent CNVs were observed in > 3% of the study population. Non-rpt, non-repetitive.

Supplemental Table 3: Depletion of UCEs among CNVs cannot be fully explained by a general depletion of CNVs among the alignable regions of the human genome.

		Observed		Expected (bp)				Obs/
Data set	Subset*	N	bp	Mean	s.d.	Min	*P*	exp
Redon	Complete	78	20,205	22,866	2,381	14,662	0.1319	0.88
	Recurrent	40	10,845	14,441	2,000	8,169	0.0361	0.75
	Non-recurrent	38	9,360	8,399	1,548	3,785	0.7326	1.11
Wong	Complete	247	65,822	66,534	3,733	50,895	0.4244	0.99
	Recurrent	44	11,588	11,479	1,819	5,766	0.5239	1.01
Simon-Sanchez	Complete	1	208	4,664	1,157	1,455	0.0001	0.04
de Smith	Complete	12	2,649	5,957	1,286	2,577	0.0051	0.44
	Recurrent	4	1,042	3,720	1,033	839	0.0048	0.28
	Non-recurrent	8	1,607	2,227	782	203	0.2139	0.72
Zogopoulos	Recurrent	13	3,139	12,032	1,829	6,725	5.8×10^{-7}	0.26
Korbel	Complete	2	537	2,110	763	207	0.0196	0.25
Pinto	Complete	40	10,479	13,818	1,918	7,624	0.0409	0.76
Wang	Complete	5	1,254	3,554	1,012	697	0.0115	0.35
Jakobsson	Recurrent	17	3,922	10,768	1,714	6,210	3.3×10^{-5}	0.36
Perry	Complete	6	1,510	5,894	1,336	2,200	0.0005	0.26
	Recurrent	2	564	3,083	899	468	0.0025	0.18
	Non-recurrent	0	0	1,246	605	0	0.0197	0.00
Kidd	Complete	1	290	2,290	817	216	0.0072	0.13

To demonstrate that the depletion of UCEs among CNVs is not due to the distribution of CNVs relative to the alignable regions of the human genome, we conducted depletion analyses in which the 1,000 random sequences matched with the UCEs in number and length were drawn from only the portion of the genome that is alignable between the human and mouse genomes. Data pertain to the combined set of UCEs only. Depletion was observed in all cases where it had been observed previously (**Table 2** and **Supplemental Table 5**) except within the complete data set of CNVs from Redon *et al*. (2006). This exception may be due to the ambiguity and overestimation of the extents of the CNVs in Redon *et al.* (2006) (see KIDD *et al.* 2008; PERRY *et al.* 2008; AD unpublished) and, consistent with this, we find that UCEs are strongly depleted among the recurrent CNVs of Redon *et al*. (2006). Note that the overall weakened *P* values for these analyses are expected, given that the percentage of CNVs occurring within the alignable portion of the genome is slightly lower than that for the entire genome. These findings are consistent

with our earlier study (DERTI *et al.* 2006). *See note regarding recurrent CNVs and Wong *et al*. in legend to **Supplemental Table 2**.

Supplemental Table 4: Depletion of UCEs among CNVs stratified into deletions and duplications.

		Deletions							Duplications						
		Observed		Expected (bp)				Obs/	Observed		Expected (bp)				Obs/
Data set	UCE class	N	bp	Mean	s.d.	Min	*P*	exp	N	bp	Mean	s.d.	Min	*P*	exp
Simon-Sanchez	**Combined**	**1**	**208**	**1,953**	**744**	**0**	**0.0095**	**0.11**	**0**	**0**	**3,643**	**1,027**	**857**	**0.0002**	**0.00**
	Intergenic	0	0	1,181	573	0	0.0196	0.00	0	0	1,688	680	0	0.0065	0.00
	Intronic	0	0	269	289	0	0.1760	0.00	0	0	1,307	654	0	0.0228	0.00
	Exonic	1	208	246	265	0	0.4430	0.85	0	0	808	458	0	0.0388	0.00
de Smith	**Combined**	**10**	**2,206**	**6,136**	**1,321**	**2,476**	**0.0015**	**0.36**	**2**	**443**	**4,474**	**1,149**	**1,023**	**0.0002**	**0.10**
	Intergenic	2	286	2,752	870	666	0.0023	0.10	1	242	2,071	783	204	0.0097	0.12
	Intronic	0	0	2,167	847	0	0.0053	0.00	1	201	1,387	658	0	0.0357	0.14
	Exonic	8	1,920	1,275	596	0	0.8604	1.51	0	0	947	505	0	0.0304	0.00
Zogopoulos	**Combined**	**2**	**583**	**3,119**	**925**	**694**	**0.0031**	**0.19**	**11**	**2,556**	**13,386**	**1,934**	**8,425**	**$1.1x10^{-8}$**	**0.19**
	Intergenic	2	583	1,891	730	0	0.0366	0.31	8	1,800	6,090	1,233	2,254	0.0003	0.30
	Intronic	0	0	393	345	0	0.1273	0.00	2	504	4,439	1,140	1,499	0.0003	0.11
	Exonic	0	0	179	213	0	0.2003	0.00	1	252	2,943	867	684	0.0010	0.09
Korbel	**Combined**	**2**	**537**	**2,944**	**957**	**466**	**0.0059**	**0.18**							
	Intergenic	1	326	1,344	614	0	0.0487	0.24							
	Intronic	0	0	1,041	576	0	0.0354	0.00							
	Exonic	1	211	692	423	0	0.1277	0.30							
Pinto	**Combined**	**0**	**0**	**6,232**	**1,333**	**2,698**	**$1.5x10^{-6}$**	**0.00**	**40**	**10,479**	**12,514**	**1,916**	**7,219**	**0.1441**	**0.84**
	Intergenic	0	0	3,449	1,015	577	0.0003	0.00	29	7,397	6,455	1,346	2,424	0.7580	1.15
	Intronic	0	0	1,212	600	0	0.0217	0.00	7	1,991	3,209	1,004	553	0.1125	0.62
	Exonic	0	0	526	383	0	0.0848	0.00	4	1,091	1,605	635	0	0.2091	0.68
Jakobsson	**Combined**	**11**	**2,551**	**6,685**	**1,357**	**3,239**	**0.0012**	**0.38**	**6**	**1,371**	**6,431**	**1,346**	**2,563**	**$8.5x10^{-5}$**	**0.21**
	Intergenic	7	1,665	3,246	915	487	0.0420	0.51	1	218	2,964	905	509	0.0012	0.07
	Intronic	2	477	1,874	793	0	0.0391	0.25	5	1,153	2,168	865	203	0.1203	0.53
	Exonic	2	409	1,931	722	0	0.0175	0.21	0	0	1,186	568	0	0.0184	0.00
Kidd	**Combined**	**1**	**290**	**3,233**	**982**	**930**	**0.0014**	**0.09**							
	Intergenic	0	0	1,770	686	0	0.0049	0.00							
	Intronic	1	290	780	504	0	0.1655	0.37							
	Exonic	0	0	525	363	0	0.0740	0.00							

We determined the depletion of intergenic, intronic, and exonic UCEs among the deletions and duplications of those data sets that distinguished deletions from duplications (DE SMITH *et al.* 2007; JAKOBSSON *et al.* 2008; KIDD *et al.* 2008; KORBEL *et al.* 2007; PINTO *et al.* 2007; SIMON-SANCHEZ *et al.* 2007; ZOGOPOULOS *et al.* 2007). A CNV reported to be both a deletion and a duplication was considered in each subset. Significant depletion of the combined set of UCEs was observed among the duplications as well as the deletions of all CNV data sets examined, except for that of Pinto *et al*. (PINTO *et al.* 2007), for which significant depletion was found only among the deletion CNVs. With regard to differences among the three types of UCEs, the significance of depletion among duplications and deletions was generally driven by the nonexonic UCEs, although there were several instances where depletion of exonic UCEs among duplications was significant.

Supplemental Table 5: Recurrent CNVs and INDELs exhibit greater depletion of UCEs.

			Observed		Expected (bp)				
Data set	Subset*	UCE class	N	bp	Mean	s.d.	Min	*P*	Obs/exp
Redon	**Complete**	**Combined**	**78**	**20,205**	**27,123**	**2,651**	**18,998**	**0.0045**	**0.74**
		Intergenic	52	13,562	13,726	1,846	7,491	0.4646	0.99
		Intronic	9	2,221	7,420	1,483	1,981	0.0002	0.30
		Exonic	17	4,422	4,975	1,162	1,216	0.3171	0.89
	Recurrent	**Combined**	**40**	**10,845**	**18,456**	**2,279**	**11,869**	**0.0004**	**0.59**
		Intergenic	26	7,241	9,425	1,540	4,262	0.0781	0.77
		Intronic	5	1,161	4,737	1,205	1,458	0.0015	0.25
		Exonic	9	2,443	3,676	926	1,180	0.0915	0.66
	Non-recurrent	**Combined**	**38**	**9,360**	**8,671**	**1,551**	**4,190**	**0.6716**	**1.08**
		Intergenic	26	6,321	4,291	1,103	1,402	0.9671	1.47
		Intronic	4	1,060	2,662	901	555	0.0377	0.40
		Exonic	8	1,979	1,299	562	0	0.8869	1.52
Wong	**Complete**	**Combined**	**247**	**65,822**	**67,922**	**3,700**	**56,731**	**0.2852**	**0.97**
		Intergenic	132	34,916	31,322	2,585	23,286	0.9178	1.11
		Intronic	71	19,910	23,813	2,453	15,638	0.0558	0.84
		Exonic	44	10,996	13,435	1,593	8,552	0.0629	0.82
	Recurrent	**Combined**	**44**	**11,588**	**12,796**	**1,808**	**7,195**	**0.2520**	**0.91**
		Intergenic	21	5,435	5,736	1,227	1,935	0.4031	0.95
		Intronic	19	5,140	4,348	1,161	1,367	0.7524	1.18
		Exonic	4	1,013	3,232	872	786	0.0055	0.31
de Smith	**Complete**	**Combined**	**12**	**2,649**	**7,534**	**1,409**	**3,888**	**0.0003**	**0.35**
		Intergenic	3	528	3,403	918	1,041	0.0009	0.16
		Intronic	1	201	2,588	893	248	0.0038	0.08
		Exonic	8	1,920	1,537	599	0	0.7387	1.25
	Recurrent	**Combined**	**4**	**1,042**	**4,910**	**1,228**	**1,824**	**0.0008**	**0.21**
		Intergenic	0	0	2,224	762	332	0.0018	0.00
		Intronic	0	0	1,590	665	0	0.0084	0.00
		Exonic	4	1,042	1,039	532	0	0.5022	1.00
	Non-recurrent	**Combined**	**8**	**1,607**	**2,761**	**926**	**234**	**0.1063**	**0.58**
		Intergenic	3	528	1,188	575	0	0.1255	0.44
		Intronic	1	201	1,011	572	0	0.0784	0.20
		Exonic	4	878	496	368	0	0.8504	1.77
Perry	**Complete**	**Combined**	**6**	**1,510**	**8,606**	**1,555**	**3,421**	**2.5×10^{-6}**	**0.18**
		Intergenic	2	564	4,566	1,158	1,669	0.0003	0.12
		Intronic	0	0	1,766	740	0	0.0085	0.00
		Exonic	4	946	1,604	653	0	0.1568	0.59
	Recurrent	**Combined**	**0**	**0**	**6,701**	**1,404**	**2,748**	**9.1×10^{-7}**	**0.00**
		Intergenic	0	0	3,620	933	939	0.0001	0.00
		Intronic	0	0	1,308	622	0	0.0177	0.00
		Exonic	0	0	1,242	577	0	0.0157	0.00
	Non-recurrent	**Combined**	**6**	**1,510**	**1,869**	**756**	**0**	**0.3174**	**0.81**
		Intergenic	2	564	949	511	0	0.2256	0.59
		Intronic	0	0	508	403	0	0.1037	0.00
		Exonic	4	946	358	309	0	0.9715	2.64
Mills	**Complete**	**Combined**	**3**	**228**	**590**	**389**	**20**	**0.1760**	**0.39**
		Intergenic	2	227	273	244	0	0.4252	0.83
		Intronic	1	1	213	235	0	0.1835	0.00
		Exonic	0	0	135	187	0	0.2352	0.00
	Recurrent	**Combined**	**1**	**1**	**99**	**125**	**2**	**0.2165**	**0.01**
		Intergenic	0	0	46	80	0	0.2826	0.00
		Intronic	1	1	30	60	0	0.3144	0.03
		Exonic	0	0	19	62	0	0.3796	0.00

Analyses were carried out as described in **Table 2**. The data sets of Zogopoulos *et al*. (2007) and Jakobsson *et al*. (2008) only included recurrent CNVs and, therefore, are not included in this table. *See note regarding recurrent CNVs and the data set of Wong *et al*. in legend of **Supplemental Table 2**.

Supplemental Table 6: UCEs identified via alignments of additional species are significantly depleted among the union of second-generation human CNV data sets and among human SDs.

UCEs				Overlap with combined data sets of human CNVs							Overlap with human SDs						
				Observed		Expected (bp)					Observed		Expected (bp)				
Species	N	Mean (bp)	Class	N	bp	Mean	s.d.	Min	*P*	Obs/ exp	N	bp	Mean	s.d.	Min	*P*	Obs/ exp
HMR+	896	269	**Combined**	**9**	**2,337**	**13,169**	**1,914**	**7,077**	**$7.6x10^{-9}$**	**0.18**	**13**	**3,191**	**13,471**	**1,912**	**8,311**	**$3.8x10^{-8}$**	**0.24**
HDM+HC			Intergenic	3	890	6,717	1,350	2,775	$7.9x10^{-6}$	0.13	0	0	7,213	1,367	2,681	$6.6x10^{-8}$	0.00
			Intronic	1	290	3,268	993	667	0.0014	0.09	0	0	2,563	867	438	0.0016	0.00
			Exonic	5	1,157	2,518	807	602	0.0459	0.46	13	3,191	3,092	882	518	0.5447	1.03
HHrM	499	261	**Combined**	**6**	**1,345**	**7,053**	**1,365**	**3,706**	**$1.5x10^{-5}$**	**0.19**	**10**	**2,472**	**7,293**	**1,356**	**2,589**	**0.0002**	**0.34**
			Intergenic	2	453	3,263	909	982	0.0010	0.14	1	218	3,562	954	959	0.0002	0.06
			Intronic	0	0	1,722	688	0	0.0062	0.00	0	0	1,345	590	0	0.0113	0.00
			Exonic	4	892	1,792	670	202	0.0896	0.50	9	2,254	2,216	755	151	0.5201	1.02
HCowM	457	259	**Combined**	**4**	**981**	**6,543**	**1,263**	**2,737**	**$5.3x10^{-6}$**	**0.15**	**10**	**2,270**	**6,628**	**1,307**	**2,791**	**0.0004**	**0.34**
			Intergenic	1	290	3,455	967	944	0.0005	0.08	1	200	3,684	965	1,056	0.0002	0.05
			Intronic	0	0	1,473	649	0	0.0116	0.00	0	0	1,142	588	0	0.0261	0.00
			Exonic	3	691	1,374	580	0	0.1195	0.50	9	2,070	1,696	643	206	0.7196	1.22
Hop	684	263	**Combined**	**9**	**2,587**	**9,811**	**1,623**	**5,332**	**$4.3x10^{-6}$**	**0.26**	**8**	**2,125**	**10,158**	**1,626**	**5,296**	**$3.9x10^{-7}$**	**0.21**
			Intergenic	5	1,228	5,026	1,122	1,820	0.0004	0.24	1	289	5,505	1,135	2,070	$2.2x10^{-6}$	0.05
			Intronic	2	525	2,839	885	494	0.0045	0.18	0	0	2,229	833	395	0.0037	0.00
			Exonic	2	834	1,282	593	0	0.2250	0.65	7	1,836	1,590	657	0	0.6459	1.15
HPl	399	271	**Combined**	**5**	**1,219**	**5,902**	**1,263**	**1,998**	**0.0001**	**0.21**	**3**	**774**	**6,134**	**1,298**	**2,522**	**$1.8x10^{-5}$**	**0.13**
			Intergenic	4	1,012	3,291	957	673	0.0086	0.31	0	0	3,476	980	747	0.0002	0.00
			Intronic	1	207	1,581	663	200	0.0191	0.13	0	0	1,233	593	0	0.0188	0.00
			Exonic	0	0	791	473	0	0.0472	0.00	3	774	988	506	0	0.3362	0.78

See **Table 1** and **Materials and Methods** for full description of the union of second-generation CNV data sets. The outcomes of depletion analyses pertaining to human-mouse-rat (HMR) + human-dog (HDM) + human-chicken (HC), human-horse-mouse (HHrM), human-cow-mouse (HCowM), human-opossum (HOp), and human-platypus (HPl) UCEs are given along with the total number (N) and average length (Mean in bp) of the UCEs used in each analysis. Analyses were carried out as described in **Table 2.**

Supplemental Table 7: Assessment of the local depletion of nonexonic conserved elements (including UCEs) among the union of second-generation CNV data sets

Conserved elements		Overlap with CNVs							Overlap with SDs*	
		Observed		Expected (bp)						
Subset	% Identity	N	Bp	Mean	s.d.	Min	*P*	Obs/exp	*P*	Obs/exp
Intronic	100	1	290	695	454	0	0.186	0.42	0.124	0.00
	99	1	231	1,577	650	0	0.019	0.15	0.064	0.00
	98	7	1,599	2,824	857	448	0.076	0.57	0.159	0.49
	97	8	1,771	2,730	794	674	0.114	0.65	0.581	1.08
	96	13	2,939	4,612	1,041	1,675	0.054	0.64	0.313	0.82
	95	19	4,398	6,194	1,170	2,381	0.062	0.71	0.991	1.63
	94	17	4,047	4,246	992	1,615	0.421	0.95	1.000	3.16
	93	36	8,250	7,635	1,318	3,434	0.680	1.08	1.000	3.73
	92	22	4,678	6,963	1,238	2,284	0.032	0.67	1.000	3.80
	91	52	11,053	10,787	1,562	6,546	0.568	1.02	1.000	2.81
	90	42	9,226	10,867	1,573	5,981	0.148	0.85	1.000	3.76
Intergenic	100	3	890	2,253	814	221	0.047	0.40	0.048	0.00
	99	20	5,334	5,255	1,231	1,841	0.526	1.02	0.354	0.90
	98	35	8,493	8,328	1,464	3,859	0.545	1.02	0.489	1.00
	97	40	9,586	11,003	1,580	6,984	0.185	0.87	0.967	1.33
	96	43	9,513	13,752	1,859	8,883	0.011	0.69	1.000	1.50
	95	64	14,443	17,370	1,998	11,959	0.071	0.83	1.000	1.51
	94	72	16,199	20,133	2,104	13,720	0.031	0.80	1.000	1.76
	93	103	22,989	27,734	2,505	19,341	0.029	0.83	1.000	1.65
	92	129	29,278	39,145	2,982	28,943	4.68×10^{-4}	0.75	1.000	1.71
	91	158	34,819	42,072	2,926	34,089	0.007	0.83	1.000	1.86
	90	216	46,633	54,065	3,470	43,497	0.016	0.86	1.000	1.78

Because CNVs are recent as compared to the SDs, they may offer a glimpse of the temporal relationship between CNVs and SDs, i.e., how SDs came to be depleted of only the most highly conserved elements as these duplications were culled by natural selection from a presumed larger population of CNVs. In this series of studies, we assessed the depletion of conserved elements among the second-generation CNVs. Refer to **Table 1** for a full description of the union CNV data set, and to **Table 3** for the methodology and results of parallel analyses of the depletion of conserved elements among SDs; *P* values and obs/exp ratios from **Table 3** are duplicated here (*) in order to facilitate comparisons. The depletion of intergenic conserved elements among CNVs was assessed by comparison to sequences lying within 100 kb of the elements.

These studies show a strong depletion (i.e., low obs/exp) of intronic and intergenic UCEs, although these observations were not statistically significant, due likely to the minimal sizes of

these sets of elements. The results are comparable to those observed for the depletion of these elements among SDs. Significant but weaker depletion (i.e., higher obs/exp) is observed at lower conservation for both intronic and intergenic elements, possibly due to the greater sizes of these sets of elements. Based on the strength of depletion (obs/exp) alone, we note the following two similarities between depletion among CNVs and depletion among SDs (compare highlighted values of obs/exp): 1) for intronic elements, depletion is strongest for elements conserved at 99% or 100%, followed by depletion for elements conserved at 98% identity, and weaker or absent at all lower degrees of conservation; 2) for intergenic elements, the strongest depletion is observed with UCEs, followed by non-depletion of elements conserved at 99% and 98% identity. Given the striking differences between SDs and CNVs (age, length, and the inclusion of deletions in CNVs but not SDs), these similarities are noteworthy as they may be mutually reinforcing and therefore suggestive of a temporal connection between the depletion of these highly conserved elements among CNVs and the depletion among SDs (see below). Such an interpretation is consistent with our observation of a greater depletion of UCEs within recurrent CNVs as compared to rare CNVs (**Supplemental Tables 3** and **5**).

Starting at 97% identity, the patterns of depletion for SDs and CNVs diverge for both the intronic and intergenic elements. While the obs/exp ratio remains essentially above 1 for SDs, it fluctuates below 1 for CNVs, in some cases achieving statistical significance, albeit for weak effects. As dosage sensitive functions can be encoded in sequences that are not highly conserved, the depletion among CNVs, as a group, may reflect the dosage sensitive functions of at least a subset. That CNVs are depleted for imperfectly conserved sequences, while SDs are not, may be due to the differences between SDs and CNVs noted above. Alternatively, the different patterns of depletion may reflect differences in the nature of the dosage sensitive functions, a possibility further suggested by the discontinuity in depletion observed among CNVs for very highly and lesser conserved sequences. For example, enhancers providing essential functions are unlikely to be deleted and would therefore be depleted among CNVs, which include deletions, but could conceivably be duplicated without consequence and would therefore not necessarily be depleted among SDs, which consist only of duplications.

Finally, the differences in depletion patterns among SDs and CNVs may reflect natural selection. For example, genomes may have evolved to compensate for copy number changes of

lesser conserved sequences but still remained intolerant of changes in copy number of the most highly conserved elements, including UCEs. This proposed intolerance may reflect dosage-sensitive functions encoded by the elements or some dosage-sensitive aspect of the elements themselves, such as their participation in copy counting via comparison; the latter explanation would also offer a mechanistic basis for extreme sequence conservation.

Supplemental Table 8: Enrichment of TAATTA in intergenic and intronic UCEs relative to their flanks is greater than that of any other hexamer of 3A's and 3T's.

Motif	Occurrences in intergenic UCEs		Obs/Exp	Motif	Occurrences in intronic UCE		Obs/exp
	Observed	Expected			Observed	Expected	
TAATTA	268	76	3.52	TAATTA	161	68	2.35
ATTAAT	162	71	2.29	ATTAAT	125	67	1.87
AATTTA,TAAATT	315	155	2.03	AATTAT,ATAATT	203	160	1.27
AATTAT,ATAATT	298	158	1.88	TATTAA,TTAATA	173	137	1.26
TATTAA,TTAATA	231	138	1.68	AATTTA,TAAATT	203	166	1.22
ATTTAA,TTAAAT	324	196	1.65	ATATTA,TAATAT	132	120	1.10
ATTATA,TATAAT	169	103	1.64	ATTTAA,TTAAAT	205	204	1.01
ATATTA,TAATAT	175	118	1.49	ATTATA,TATAAT	114	124	0.92
AAATTT	131	94	1.40	AAATTT	95	106	0.90
TTATAA	66	58	1.14	AATATT	85	96	0.88
AATATT	110	101	1.09	TTATAA	53	69	0.77
TTTAAA	152	167	0.91	TTTAAA	106	176	0.60
ATATAT	54	83	0.65	ATATAT	40	88	0.46
TATATA	41	69	0.59	TATATA	27	68	0.40

The observed and expected numbers of occurrences of each hexamer permutation containing 3A's and 3T's were calculated for the intergenic and intronic UCEs, as was their ratio (obs/exp). The expected number of motif(s) is calculated by assuming a frequency of occurrence based on the density of motif(s) in the 1 kb 5' and 3' flanking regions. Reverse complements are considered together, with their occurrences being summed. TAATTA exhibited the greatest fold enrichment in both the intergenic and intronic UCEs. It should be noted that motif-x did not detect a number of the motifs listed here. This may reflect the fact that i) motif-x was run to search for over-represented motifs in the foreground sequences without any information regarding the content of the sequences flanking the UCEs (**Materials and Methods**), ii) the patterns were already included in more significant pentamers (e.g., adding an additional "T" onto TAAAT results in a substantial reduction of its overall significance) or, iii) these patterns did not meet the occurrence threshold imposed.

Supplemental Figure 1: The distributions of A+T content in the human genome and union of all ten CNV data sets showing a depletion of UCEs are not statistically different. Intergenic and intronic UCEs both have an average A+T content (~63%) that is higher than the genomic average (~59%). In order to demonstrate that depletion of UCEs among CNVs is not due to a paucity of CNVs near A+T rich genomic regions, we compared the distribution of A+T content among the union of all ten CNV data sets showing a depletion of UCEs (see **Table 2** for a list of these sets) to that of the human genome. We removed unsequenced nucleotides from both the human genome (hg17; blue bars) and the union of the ten CNV data sets (red bars), divided the remaining genome and CNVs into non-overlapping windows of 1 kb, and computed the A+T content in each window. We then randomly sampled 1,000 such non-overlapping windows from both A+T distributions and determined that the two distributions are not significantly different by a two-sample Mann-Whitney test ($P > 0.34$).

Supplemental Figure 2: Assessment of the depletion among human SDs of intergenic elements conserved at 90% identity or higher, relative to their flanking regions. For analyses of intergenic conserved elements, 1,000 sets of random sequences matched with the conserved elements in number and length were chosen from within 100 kb, 250 kb, 500 kb, 1.0 Mb, or 1.5 Mb lying 5' and 3' of the element conserved at the indicated % identity. (See **Table 3** for additional data regarding analyses in which random sequences were drawn from within 100 kb of the conserved element.) The SDs were taken from Scherer and colleagues (CHEUNG *et al.* 2003) and Eichler and colleagues (SHE *et al.* 2004). Significance of depletion is noted as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Supplemental Figure 3: Frequency of TAATTA increases sharply at the transition from

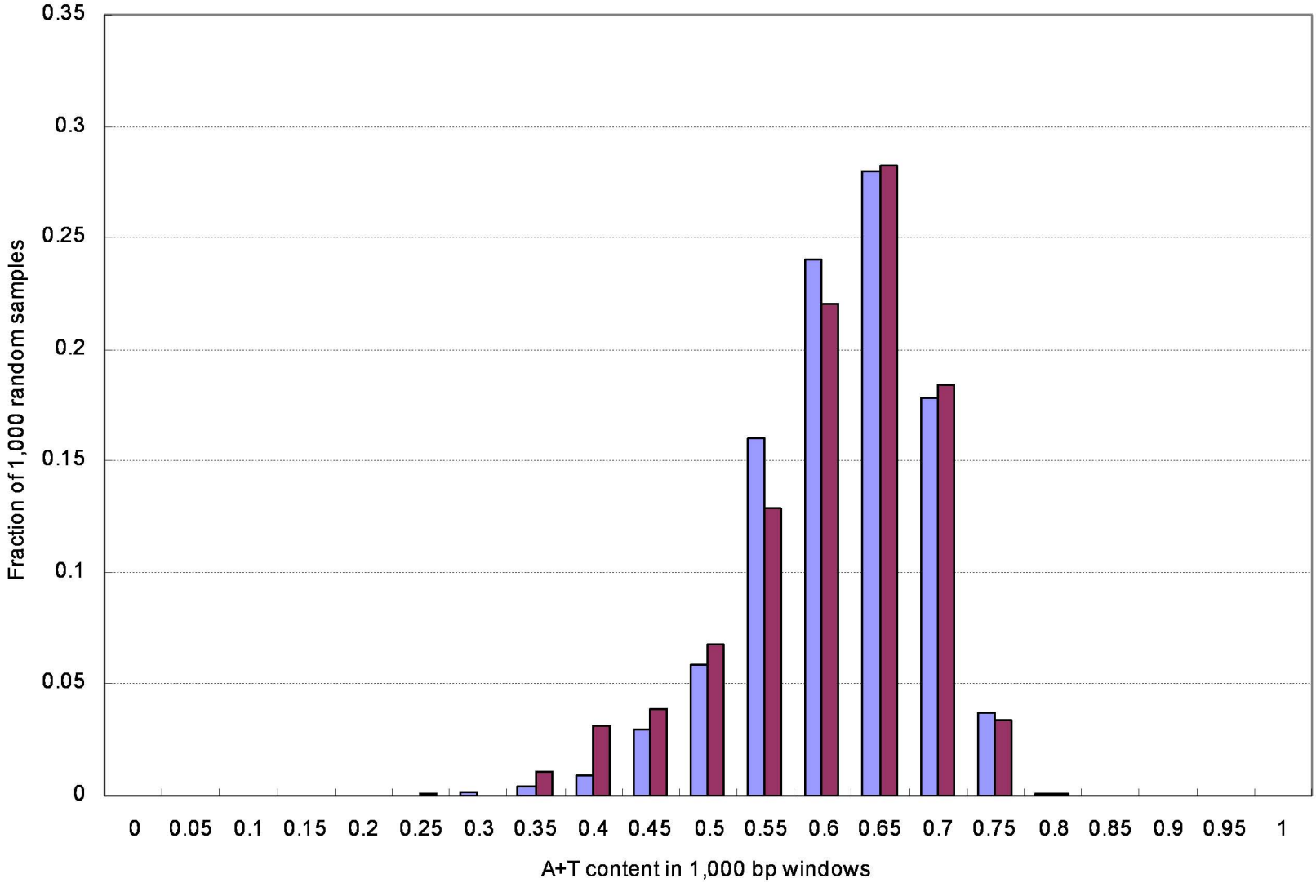
flanking sequences into the nonexonic UCEs.

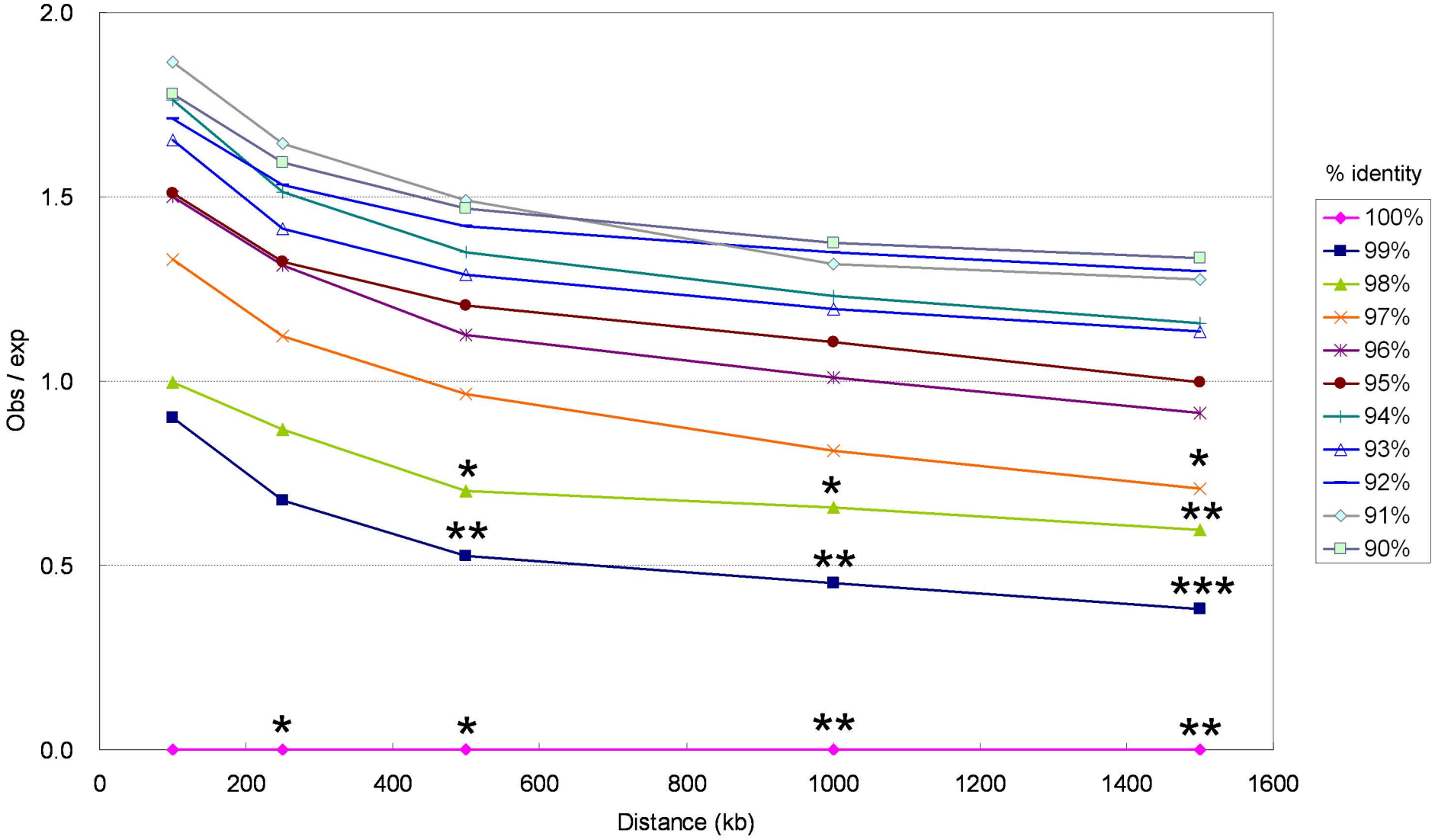
All 422 intergenic UCEs and their 5' and 3' flanking regions (4.5 times the length of the UCE) were divided into 100 segments, and corresponding segments across all 422 UCEs were then binned and assessed for the occurrence of TAATTA (y axis). UCEs correspond to bins 46-55, inclusive (black bar), and are clearly enriched for TAATTA as compared to flanking regions. A similar pattern is observed for the intronic UCEs, though the contrast in frequency of TAATTA within and outside the UCEs is less pronounced due to the relatively lower enrichment of the TAATTA motif among intronic UCEs.

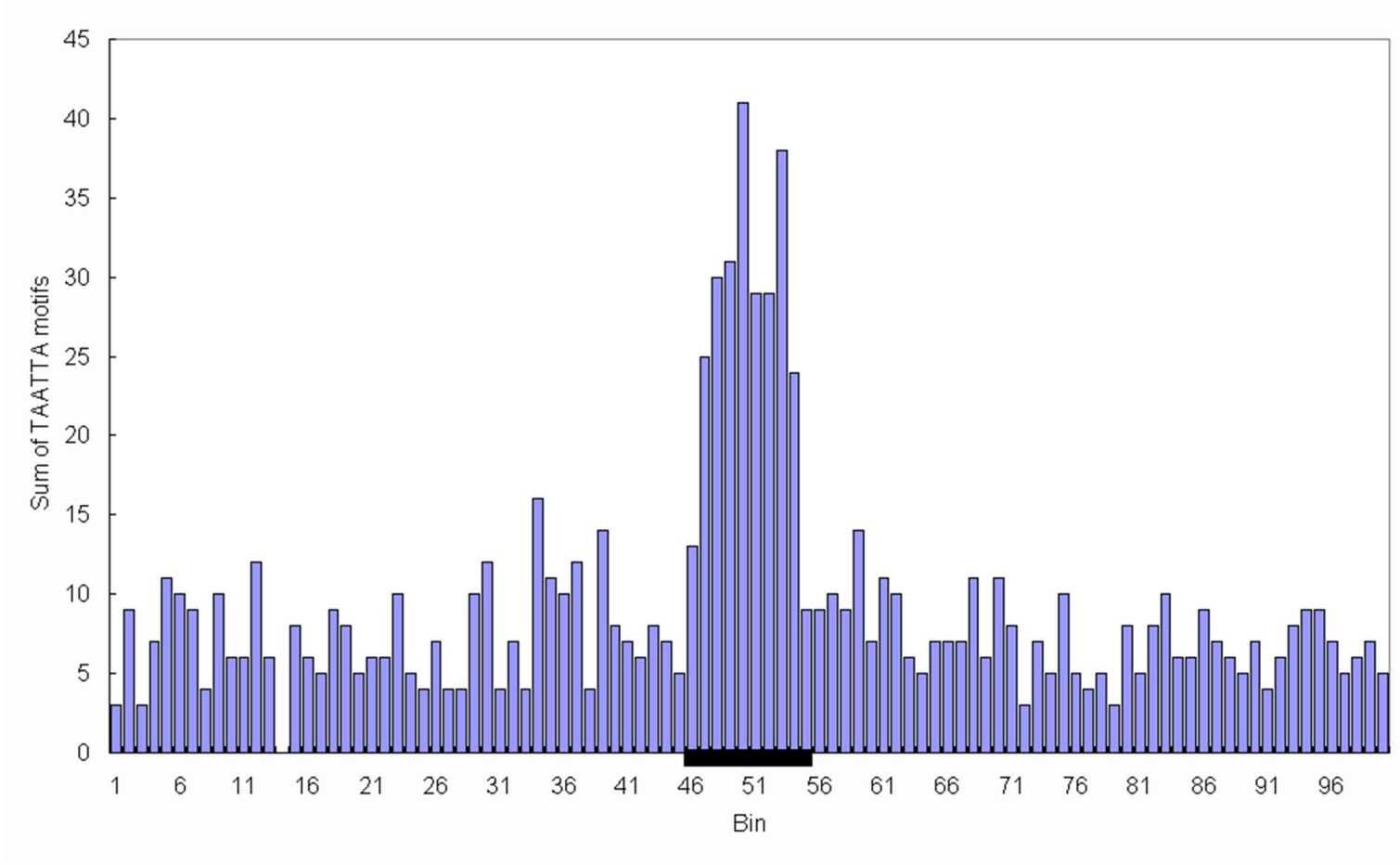
Supplemental Figure 4: A+T frequencies of intergenic conserved elements as the requirement for identity is lowered from 100% to 75%. Conserved elements are shown in red and their flanking sequences in blue.

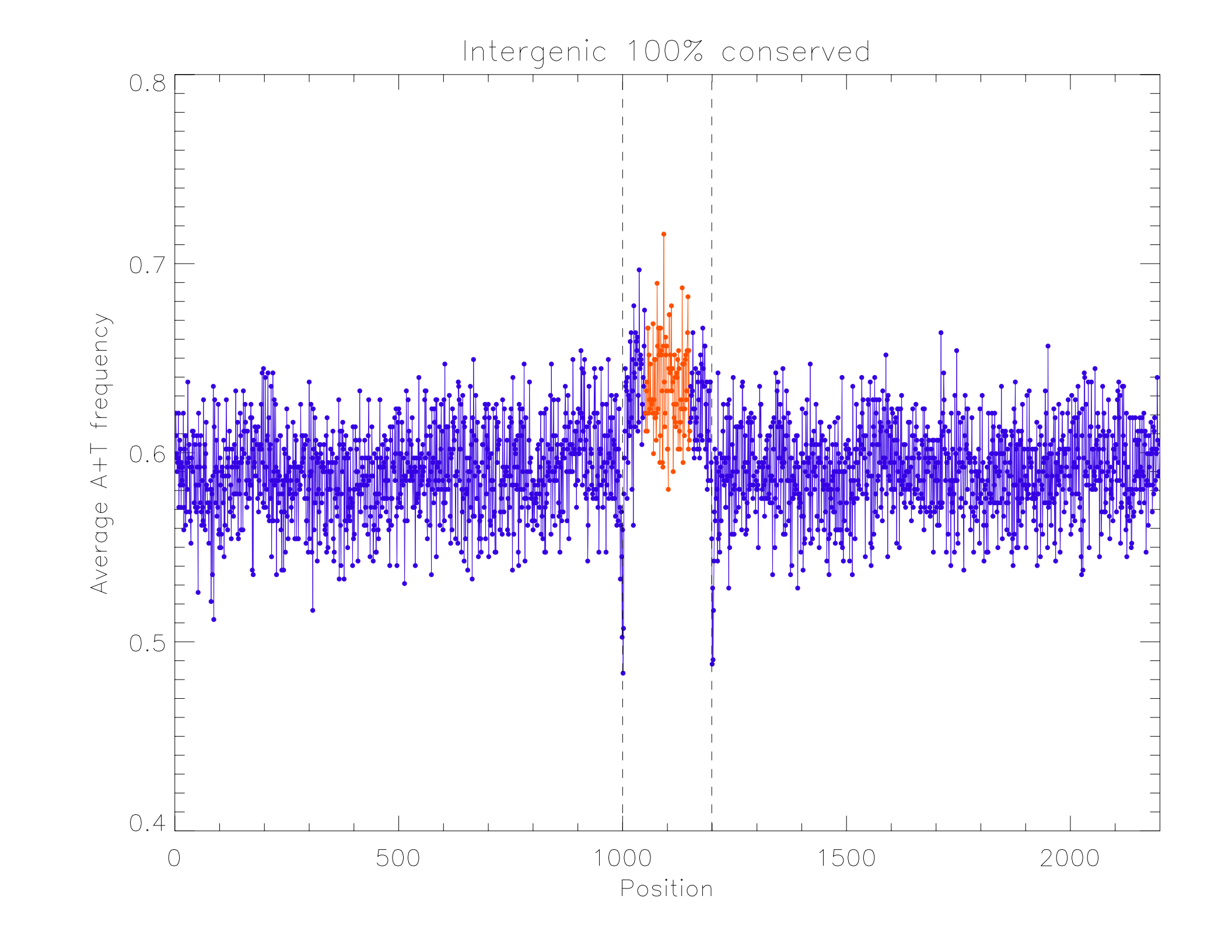
Supplemental Figure 5: A+T frequencies of intronic conserved elements as the requirement for identity is lowered from 100% to 75%. Conserved elements are shown in red and their flanking sequences in blue.

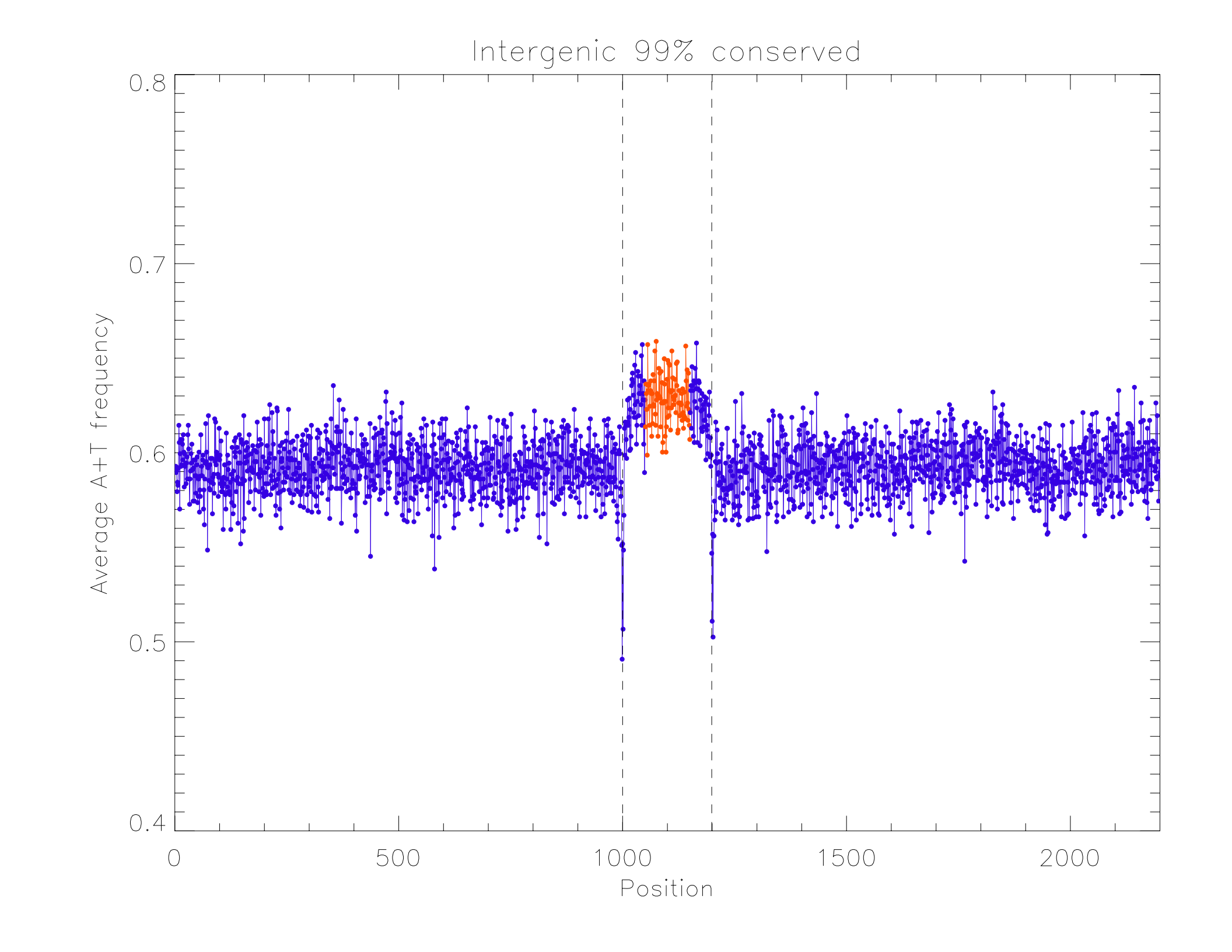
Supplemental Figure 6: A+T frequencies of exonic conserved elements as the requirement for identity is lowered from 100% to 75%. Conserved elements are shown in red and their flanking sequences in blue.

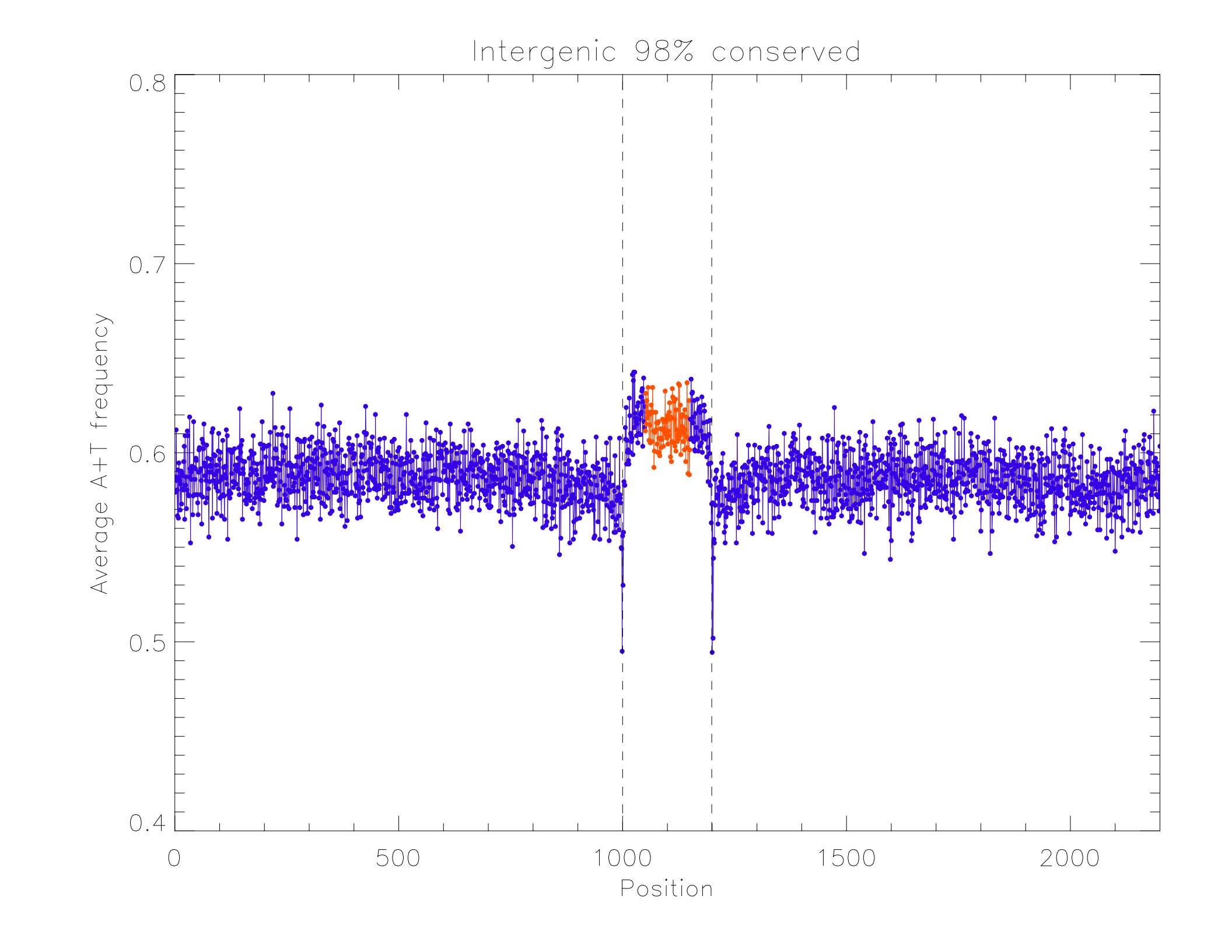

0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Fraction of 1,000 random samples
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1
A+T content in 1,000 bp windows

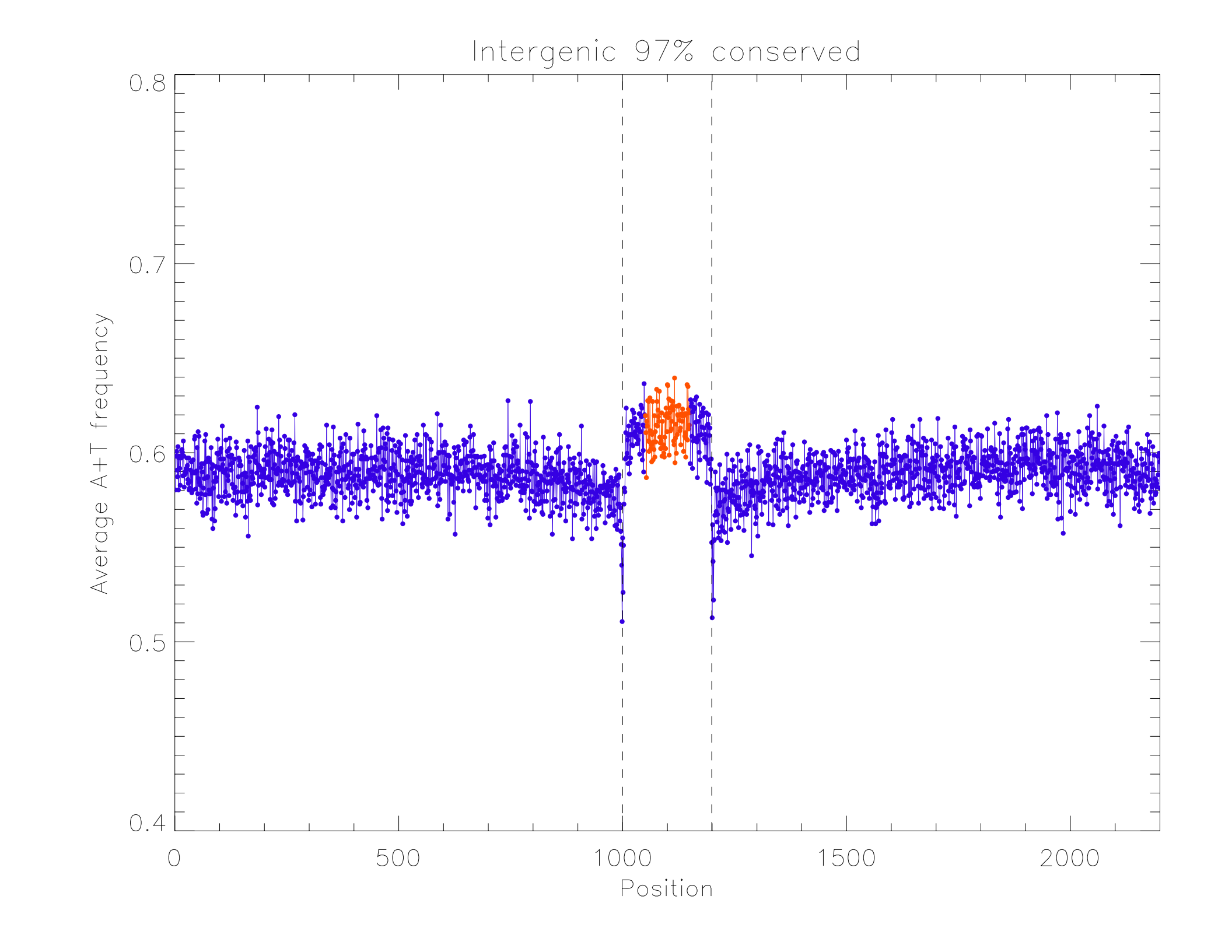

2.0
1.5
1.0
0.5
0.0
Obs / exp
0
200
400
600
800
1000
1200
1400
1600
Distance (kb)
% identity
100%
99%
98%
97%
96%
95%
94%
93%
92%
91%
90%
*
**
***

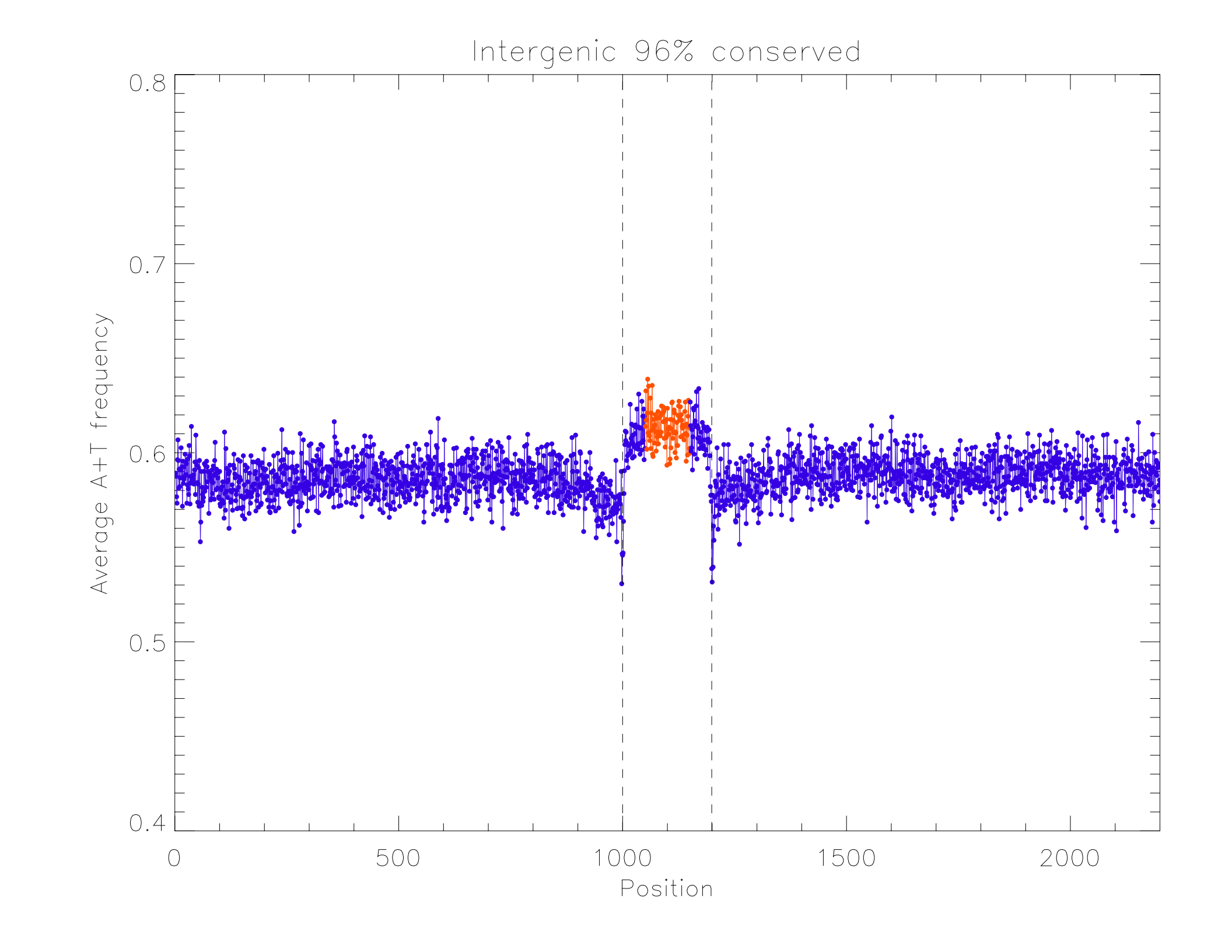

Sum of TAATTA motifs
0
5
10
15
20
25
30
35
40
45
1
6
11
16
21
26
31
36
41
46
51
56
61
66
71
76
81
86
91
96
Bin

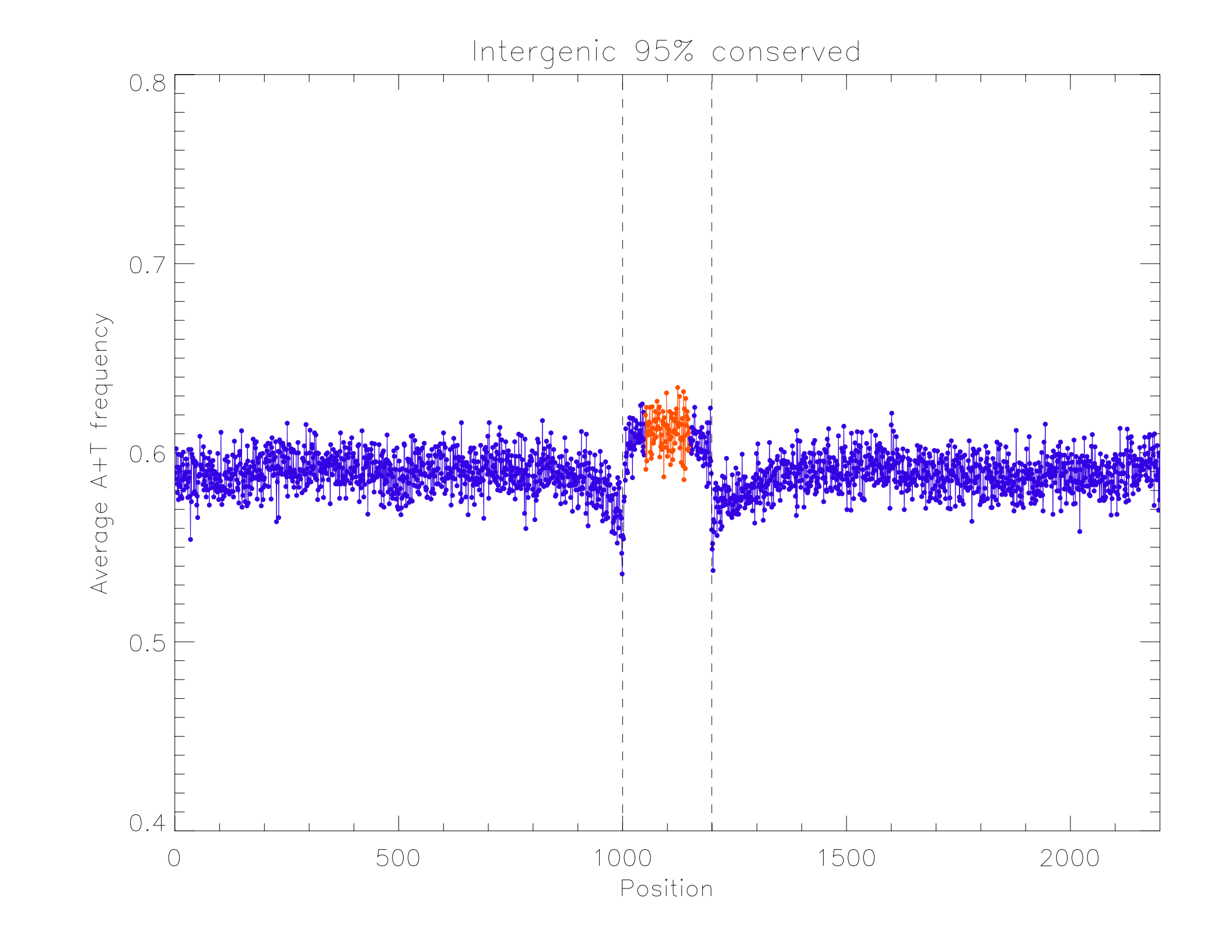

Intergenic 100% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

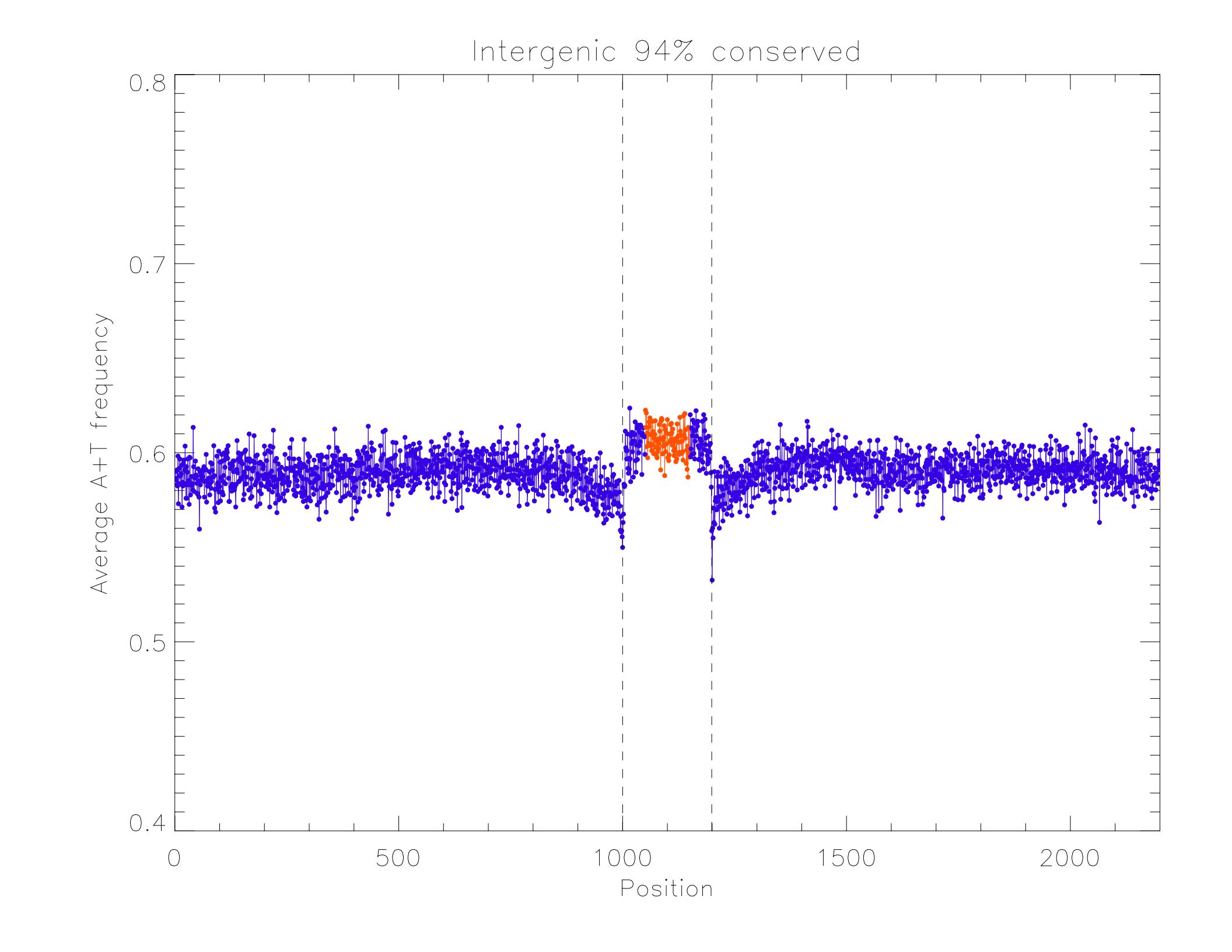

Intergenic 99% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

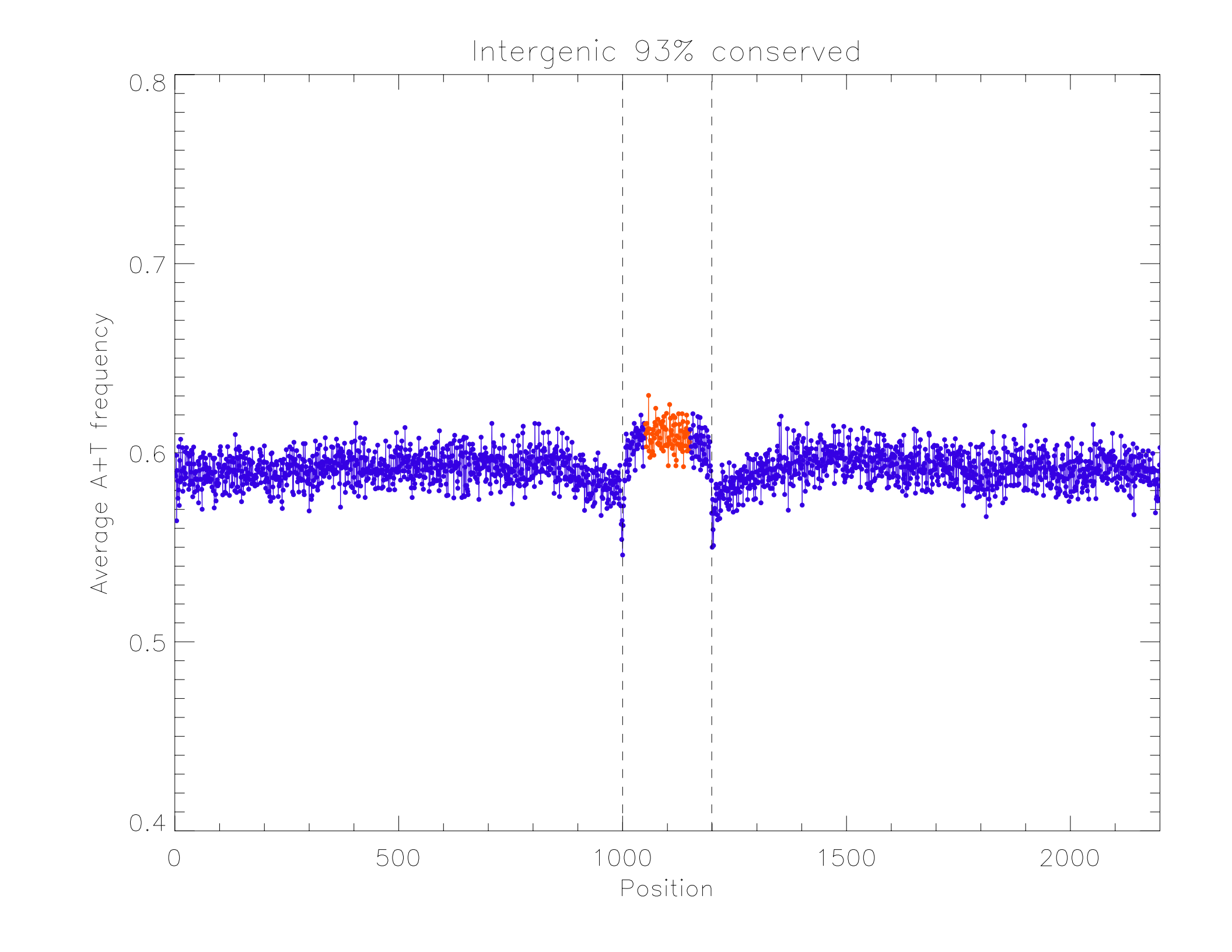

Intergenic 98% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

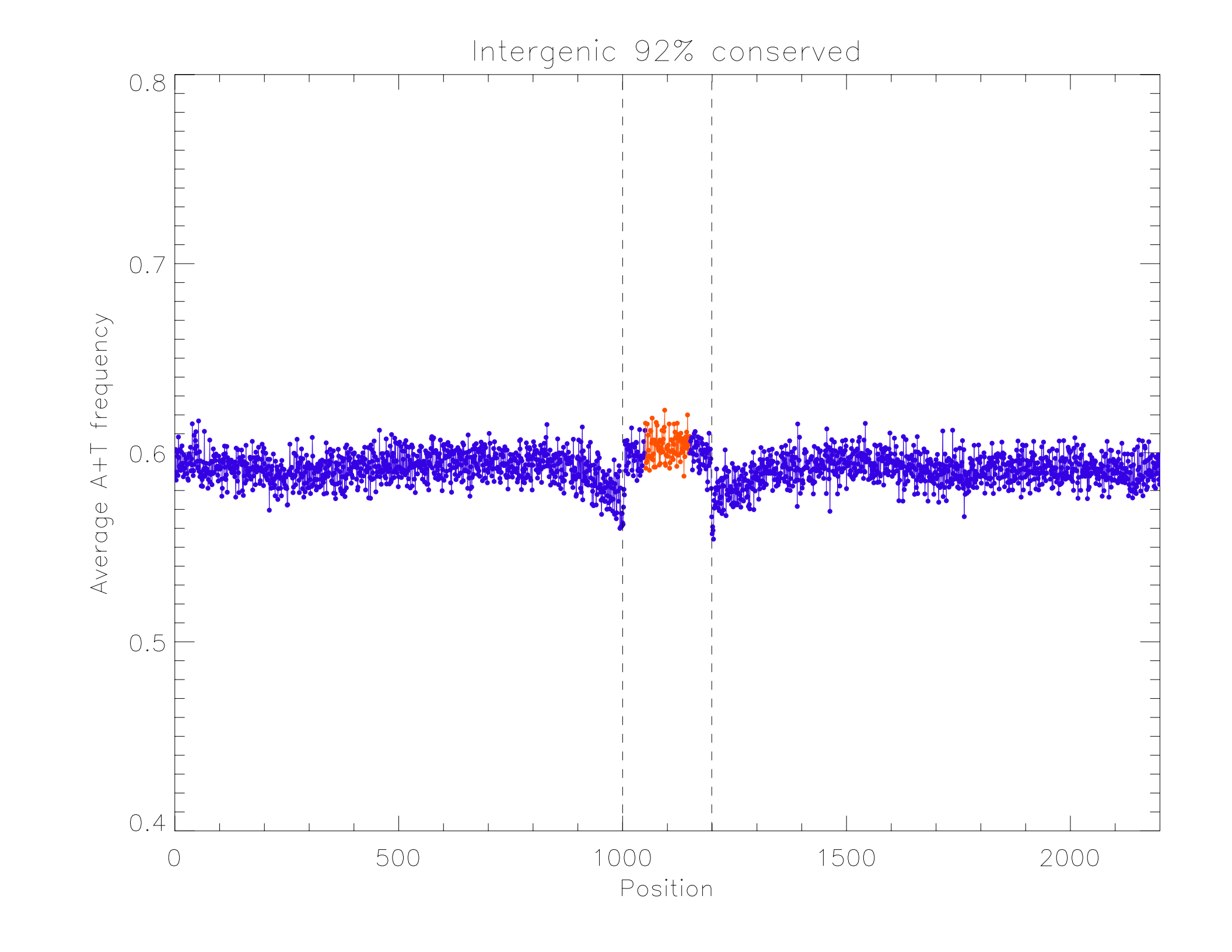

Intergenic 97% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

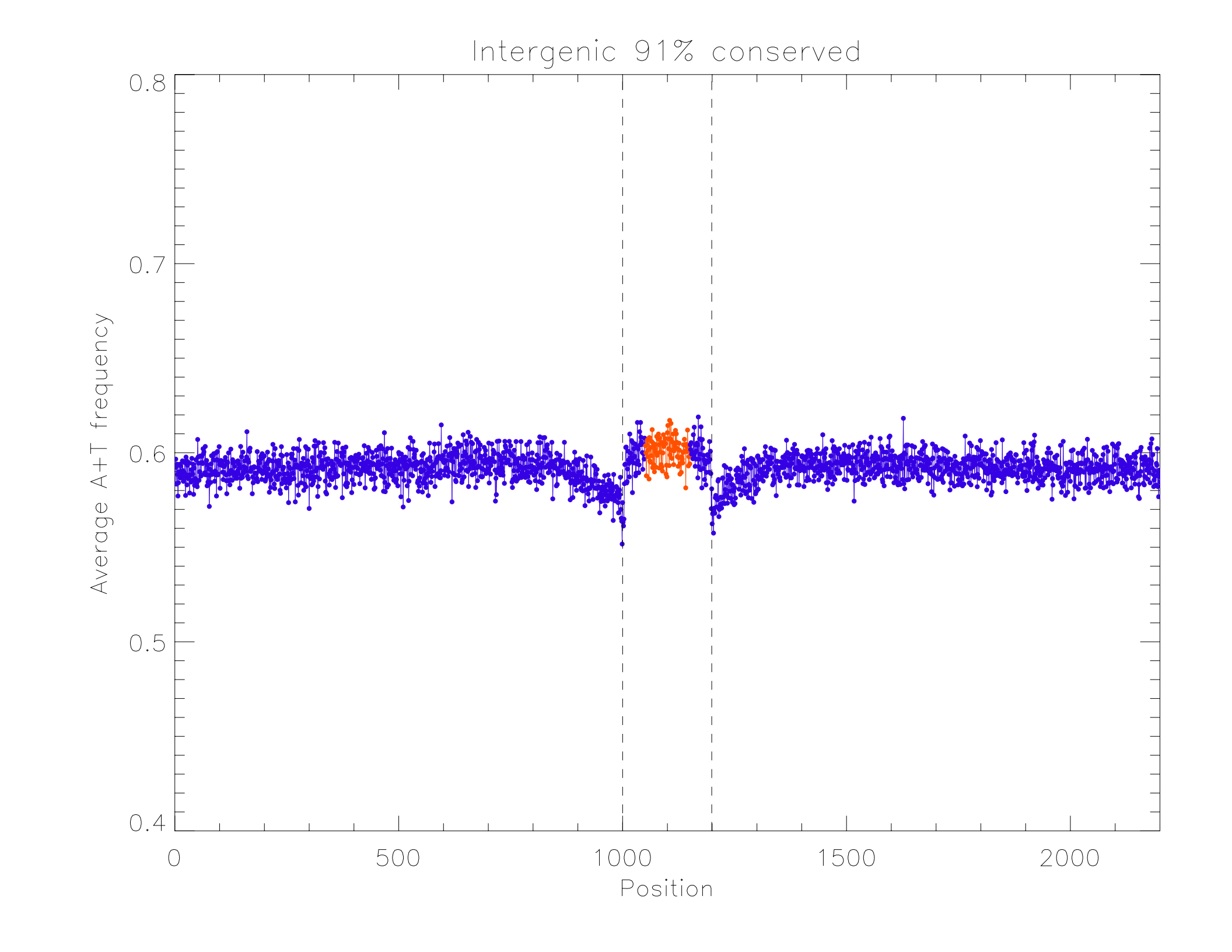

Intergenic 96% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

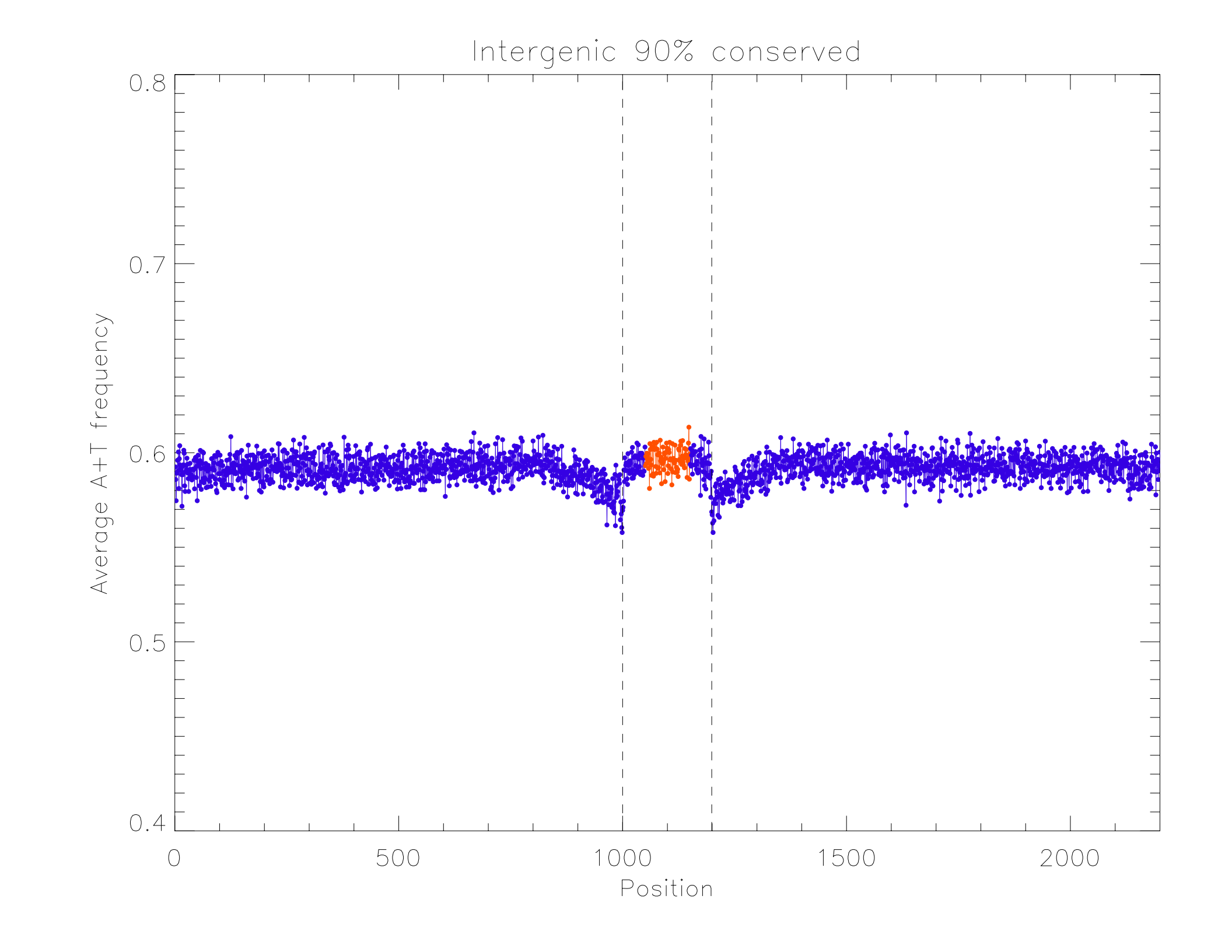

Intergenic 95% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

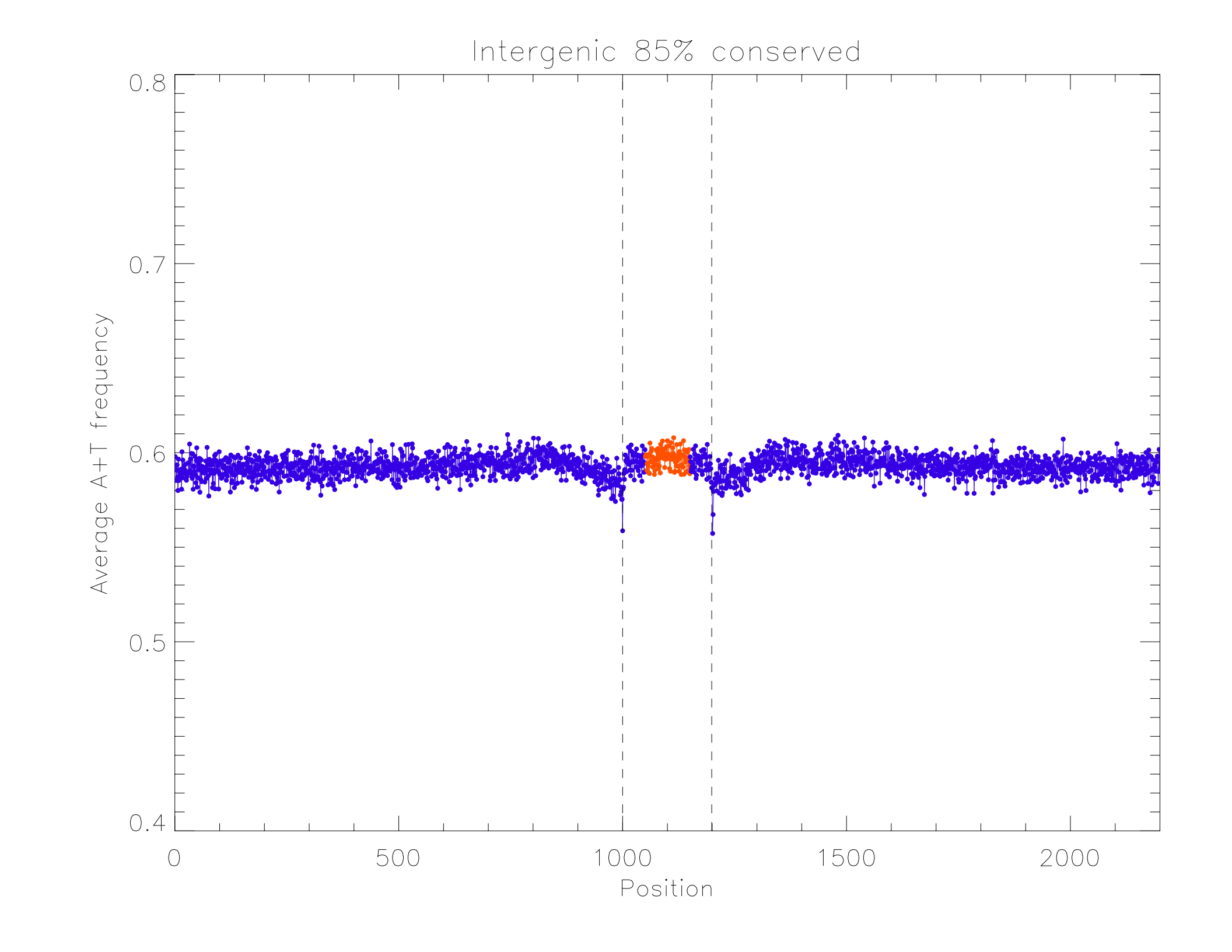

Intergenic 94% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

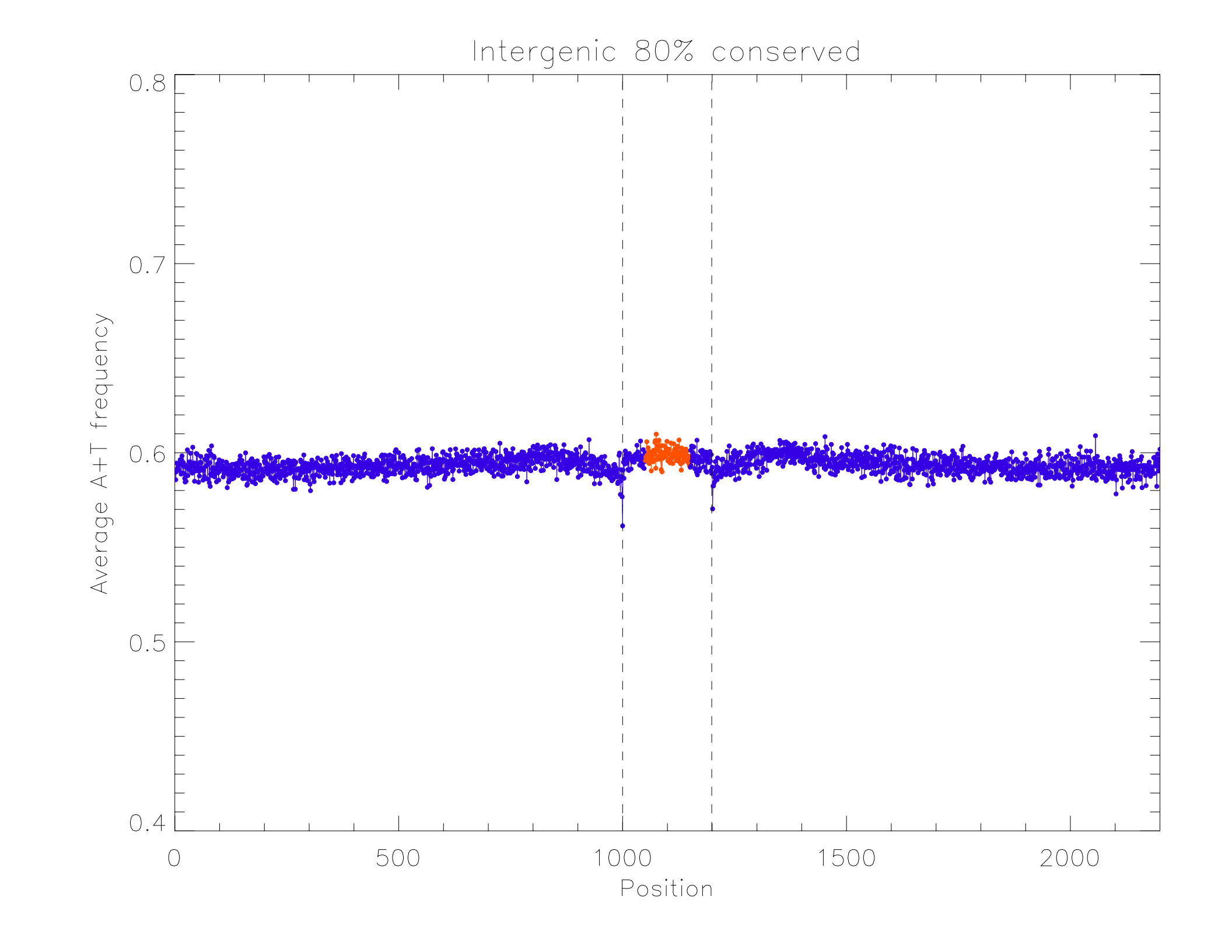

Intergenic 93% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

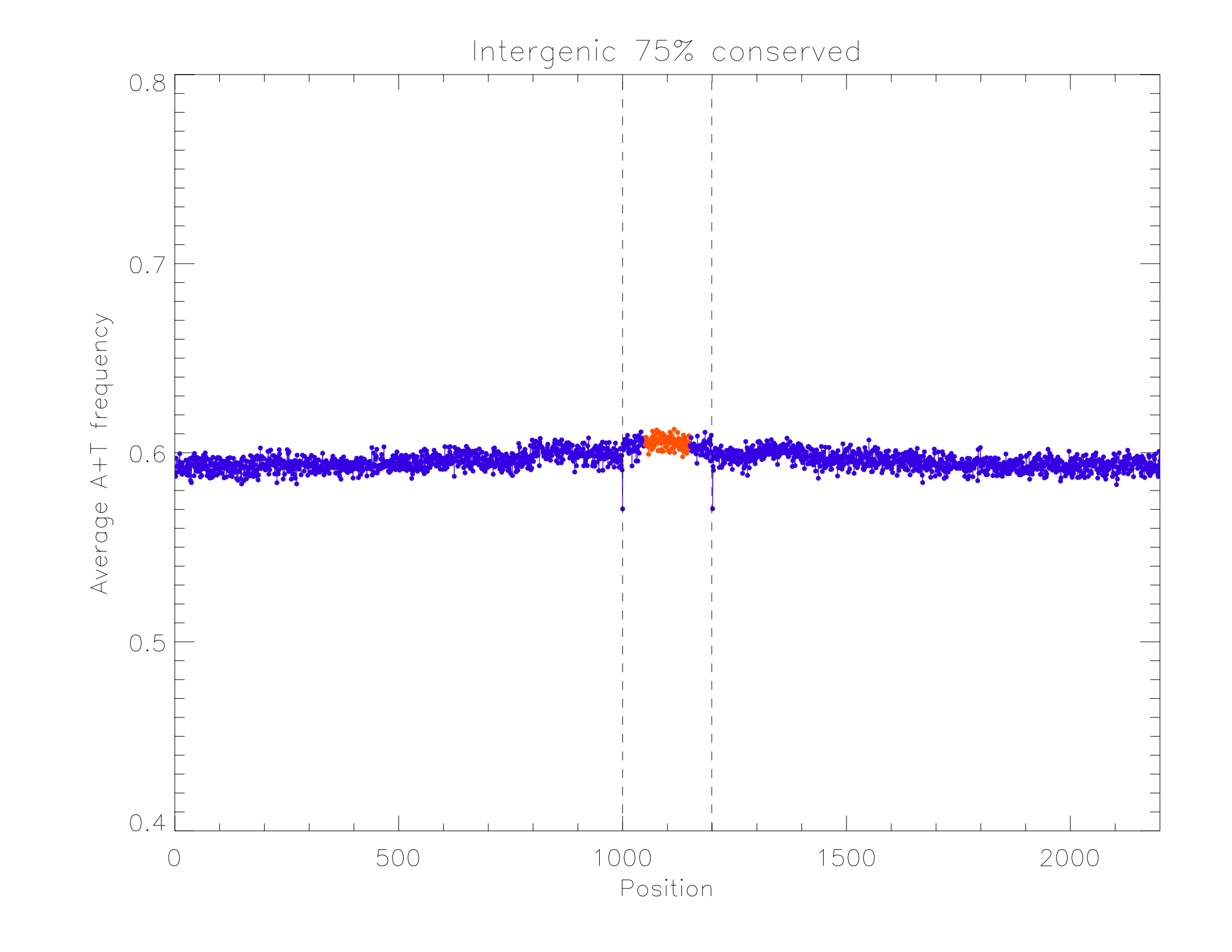

Intergenic 92% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

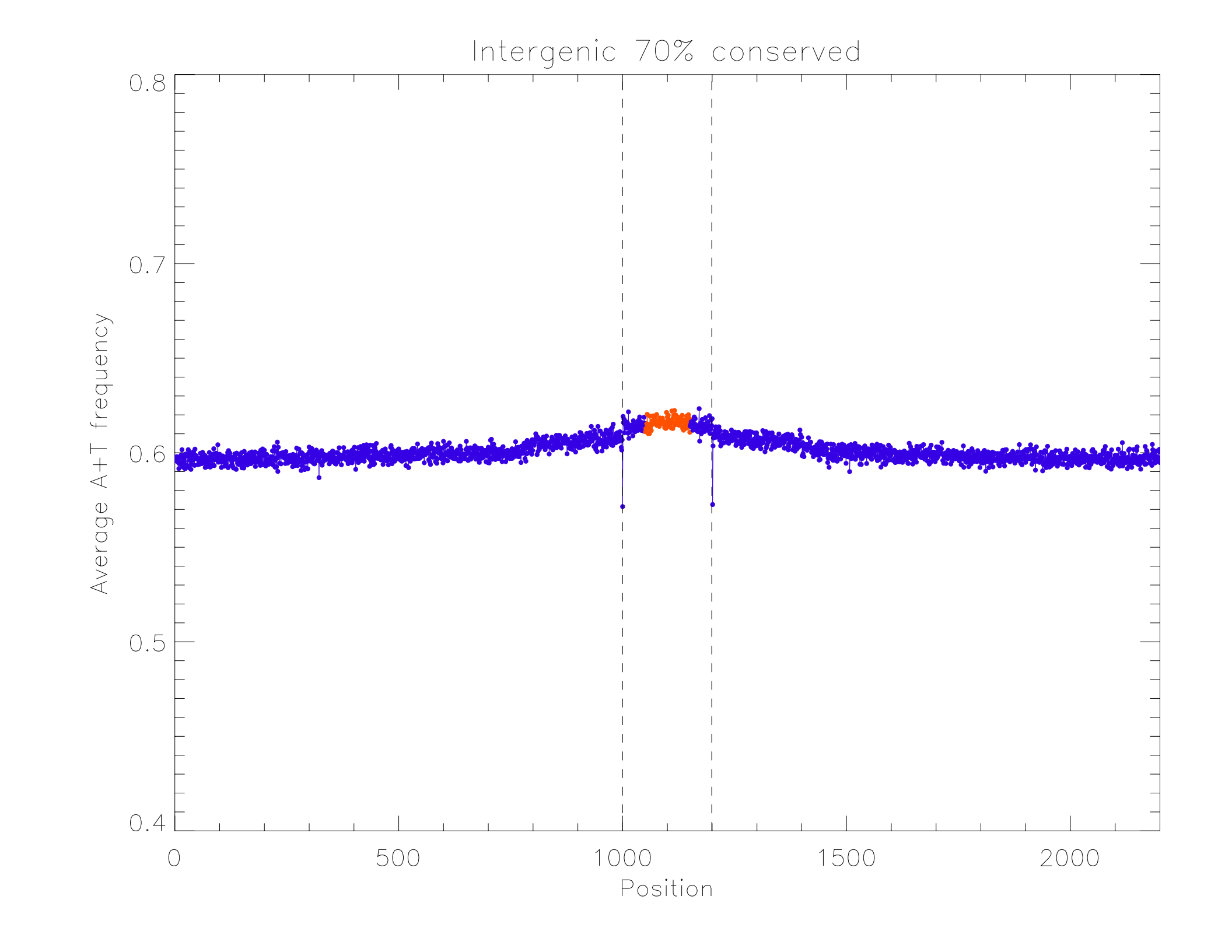

Intergenic 91% conserved
0.8
0.7
0.6
0.5
0.4
Average A+T frequency
0
500
1000
1500
2000
Position

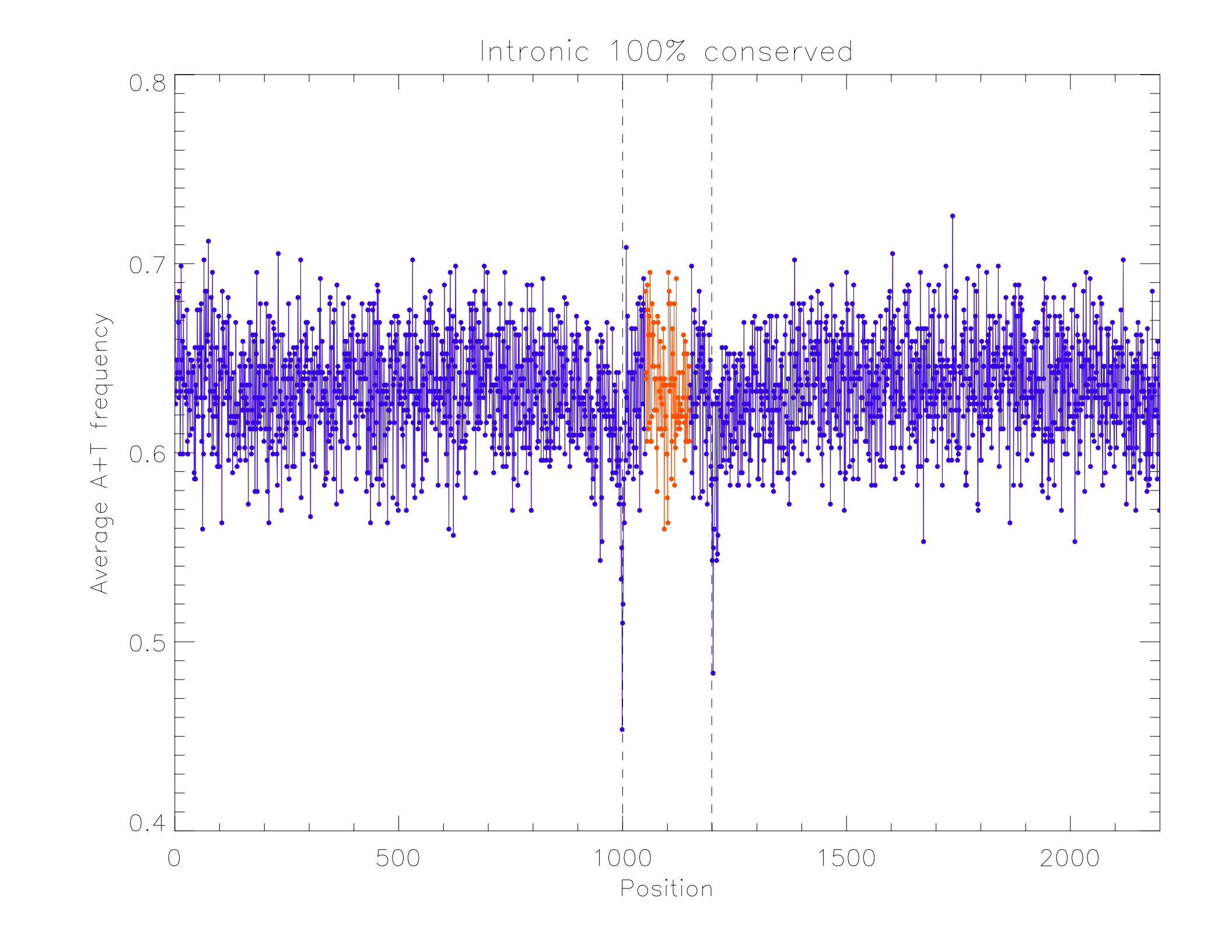

Intergenic 90% conserved
0.8
0.7
0.6
0.5
0.4
Average A+T frequency
0
500
1000
1500
2000
Position

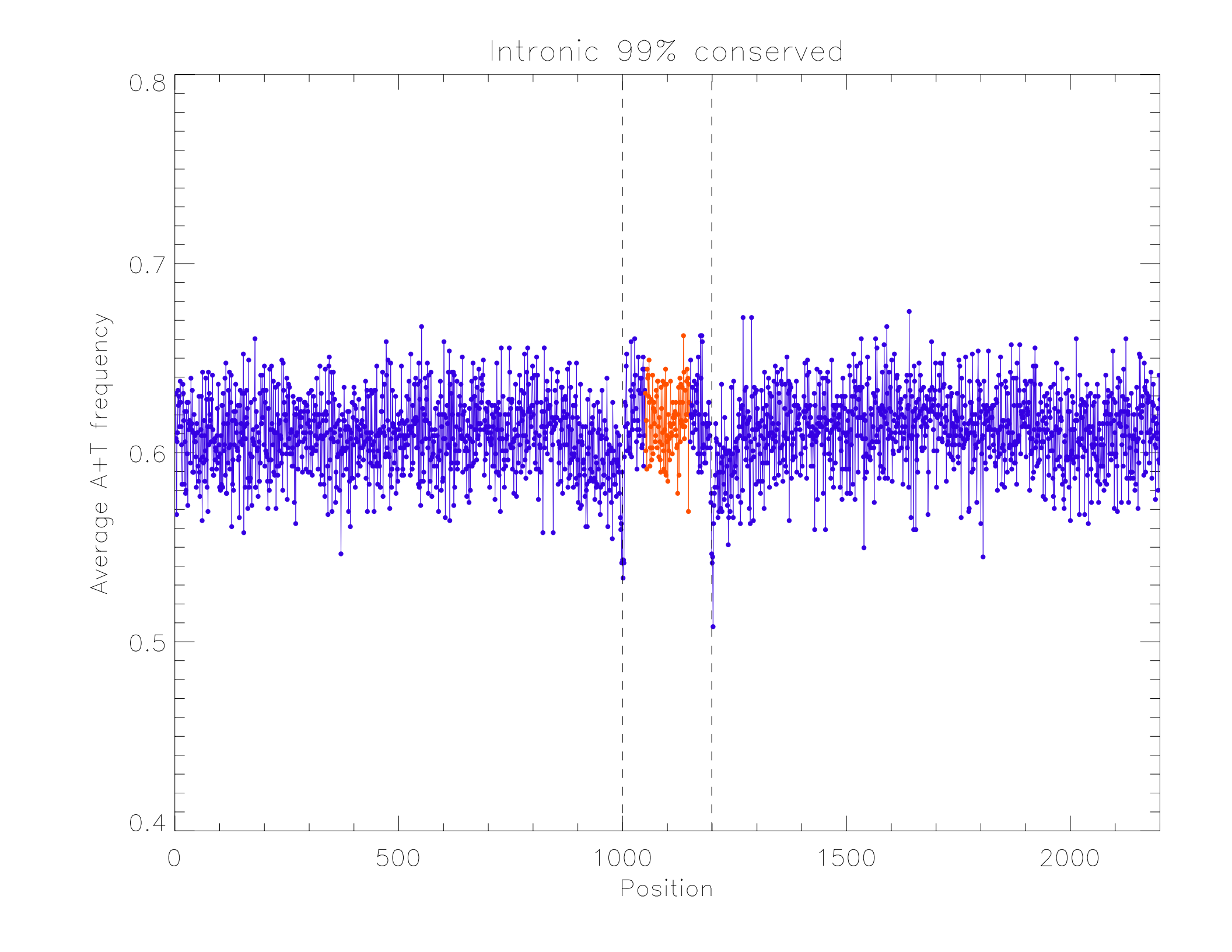

Intergenic 85% conserved
0.8
0.7
0.6
0.5
0.4
Average A+T frequency
0
500
1000
1500
2000
Position

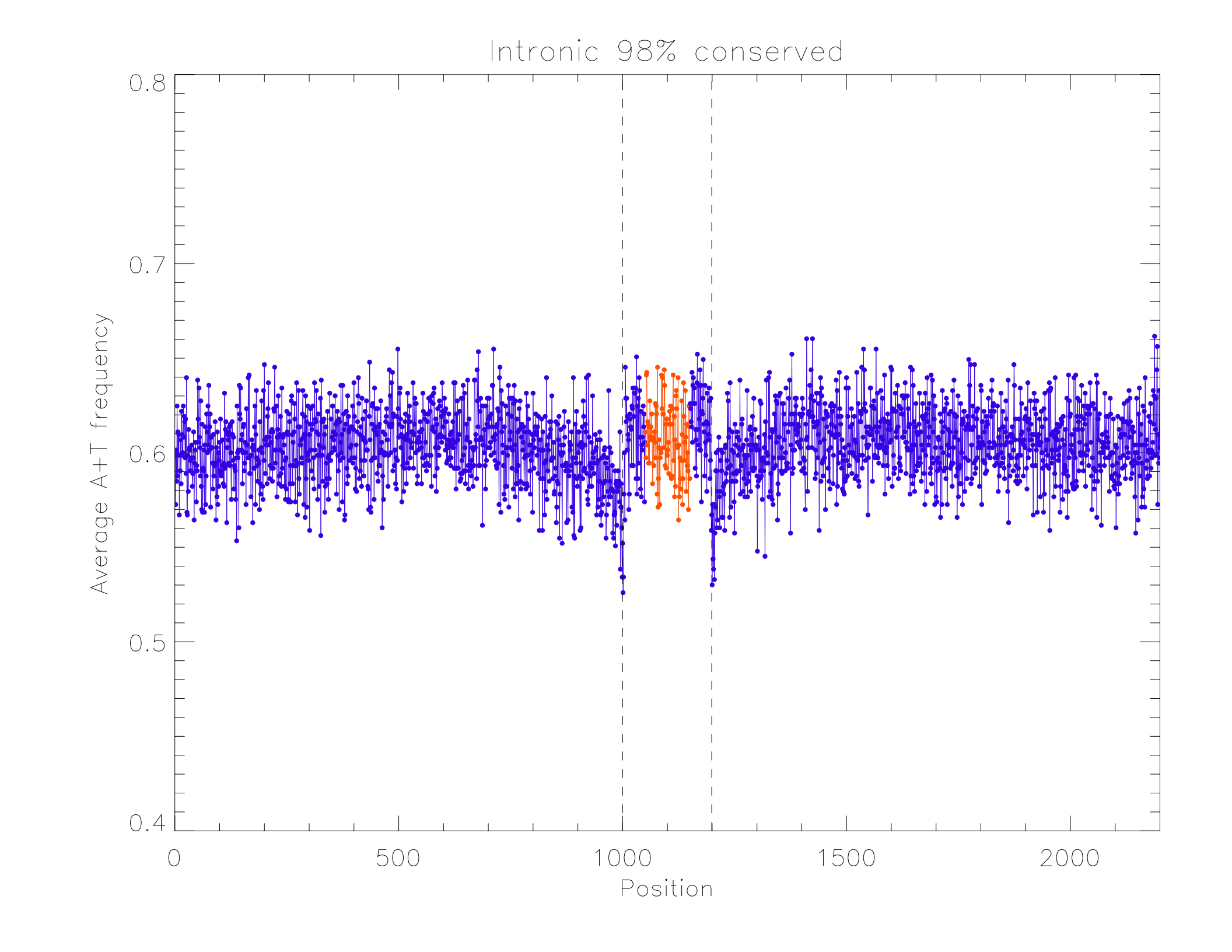

Intergenic 80% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

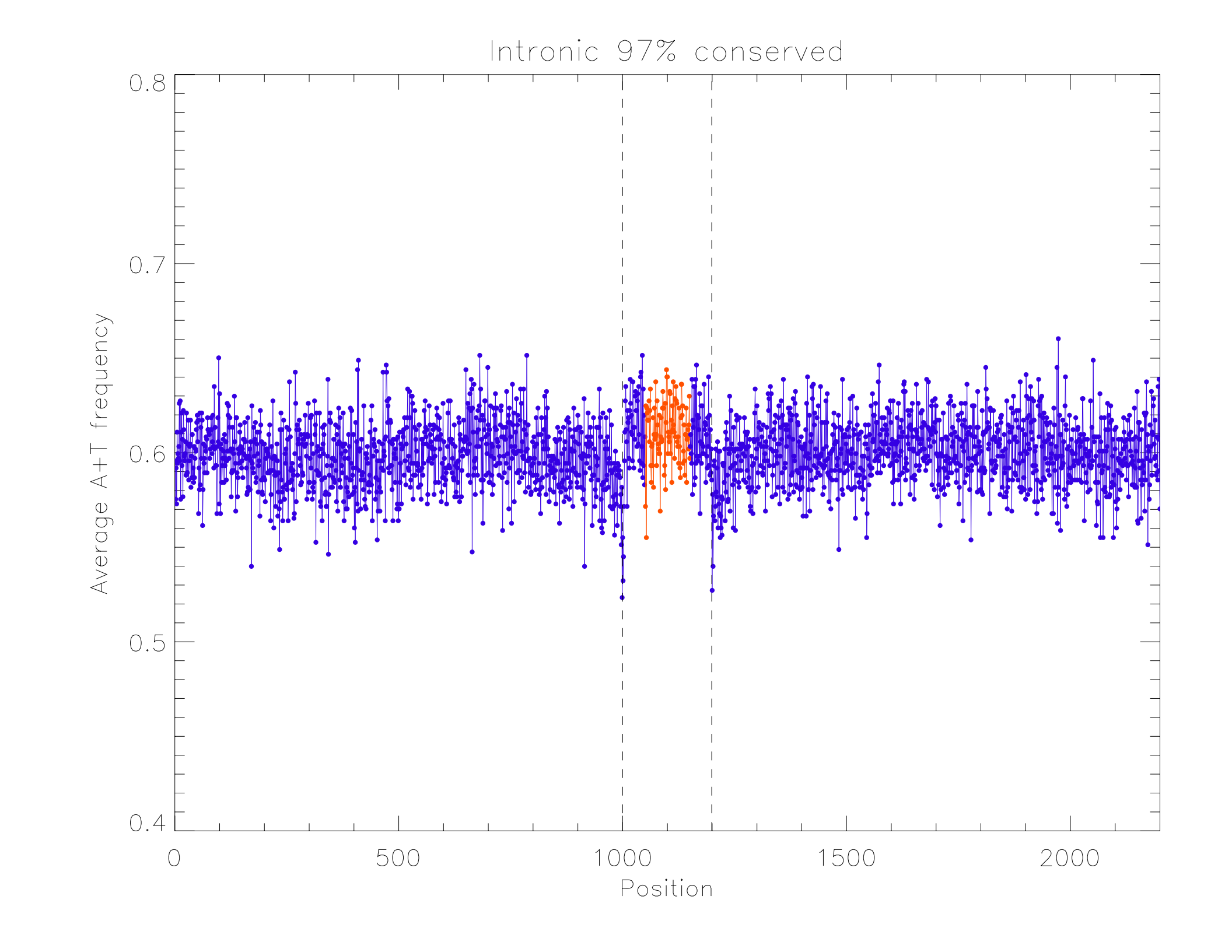

Intergenic 75% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

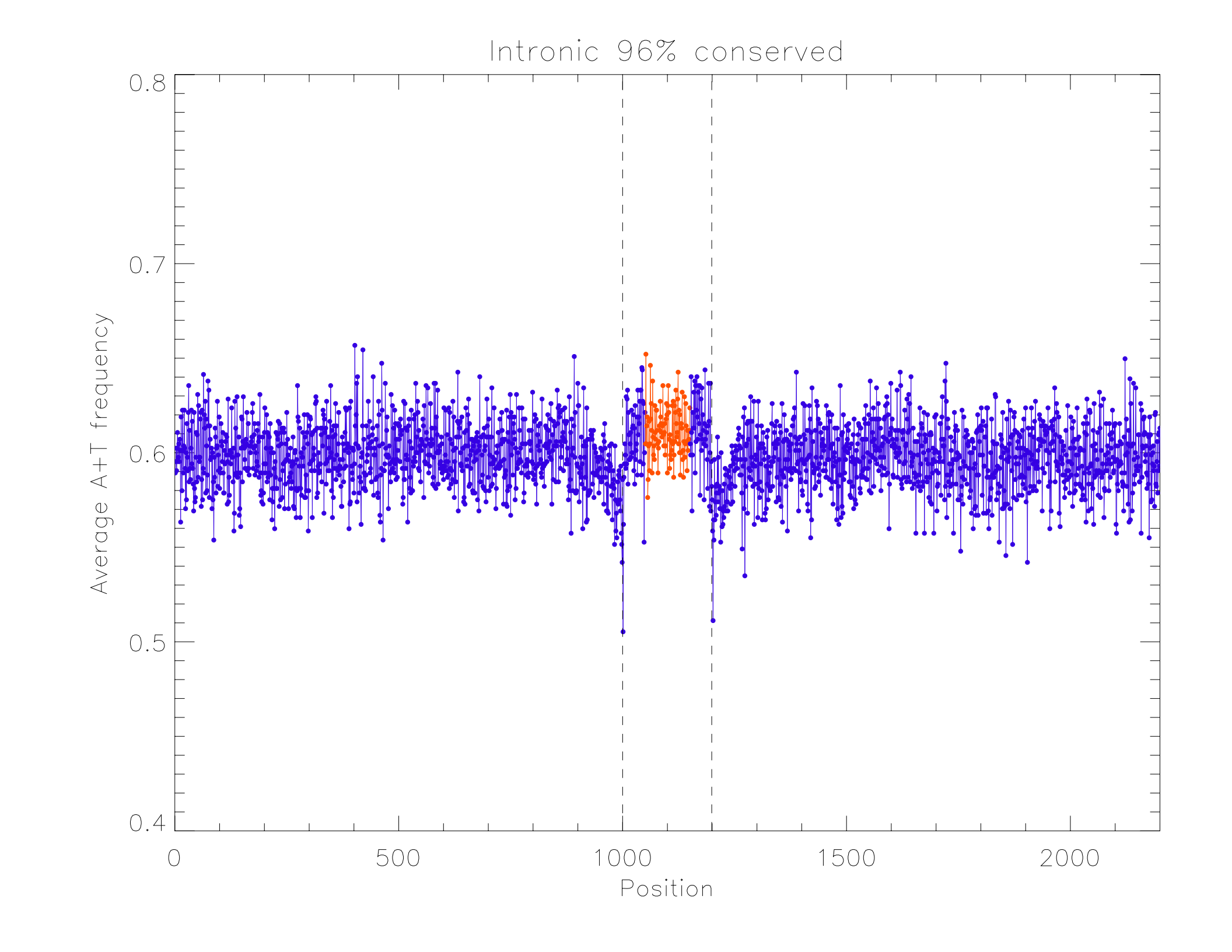

Intergenic 70% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

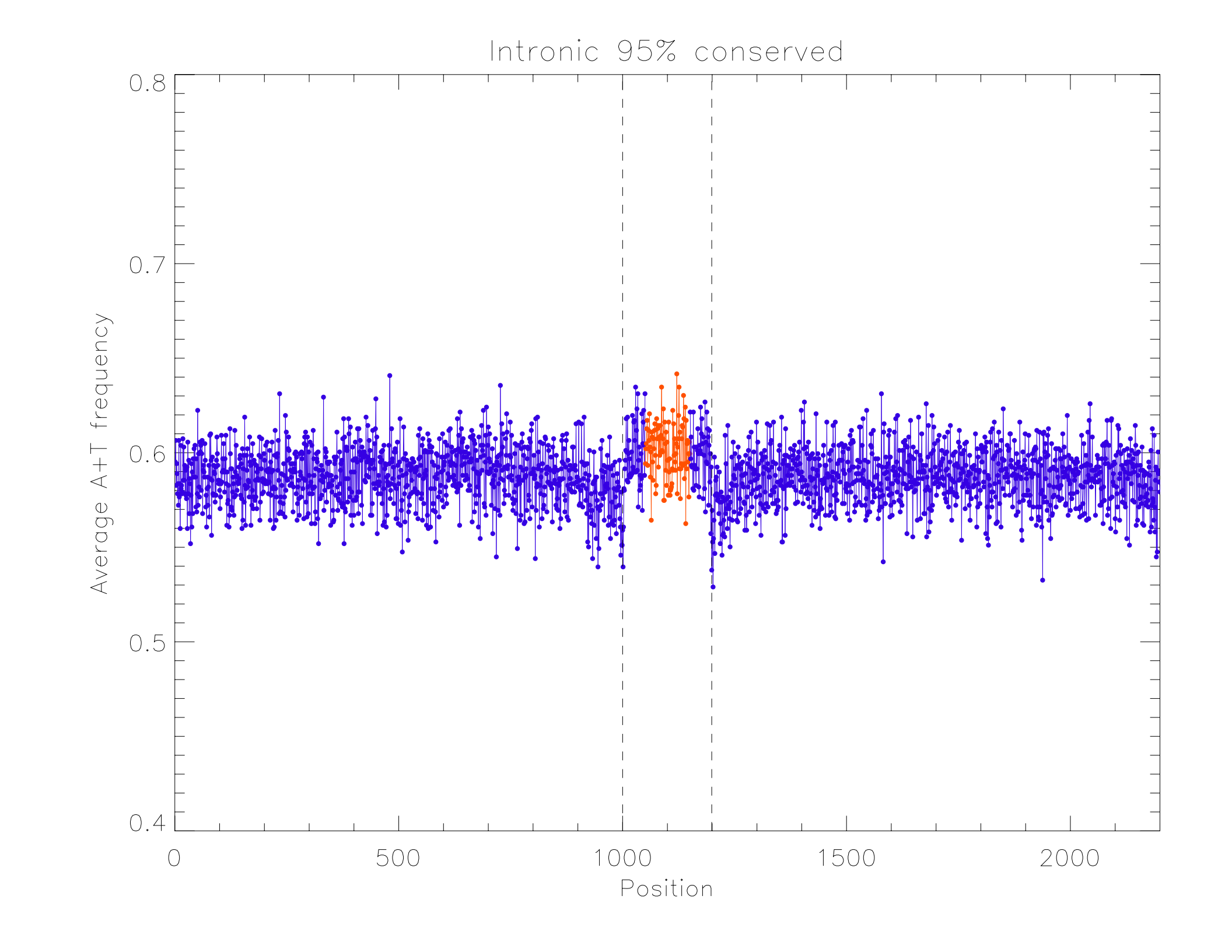

Intronic 100% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

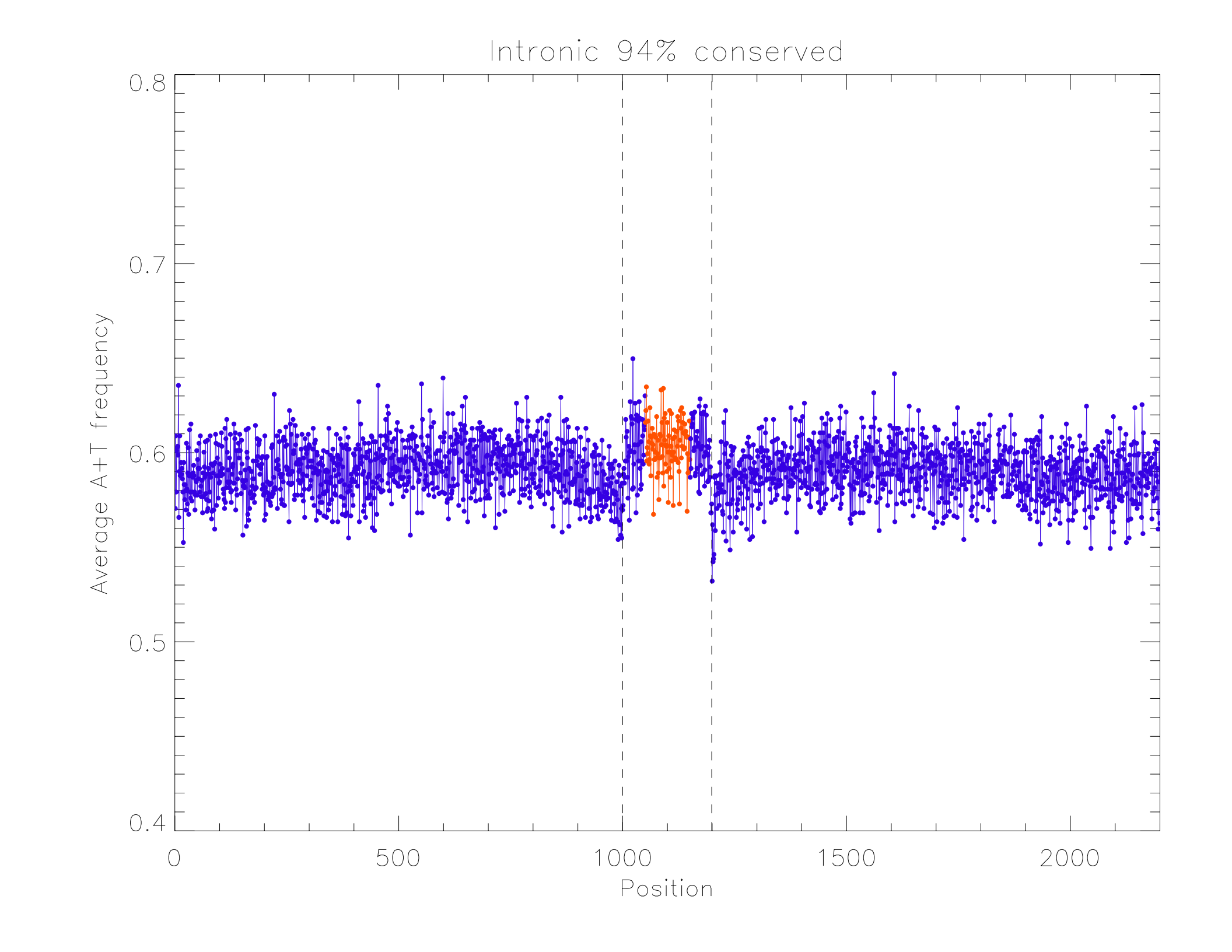

Intronic 99% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

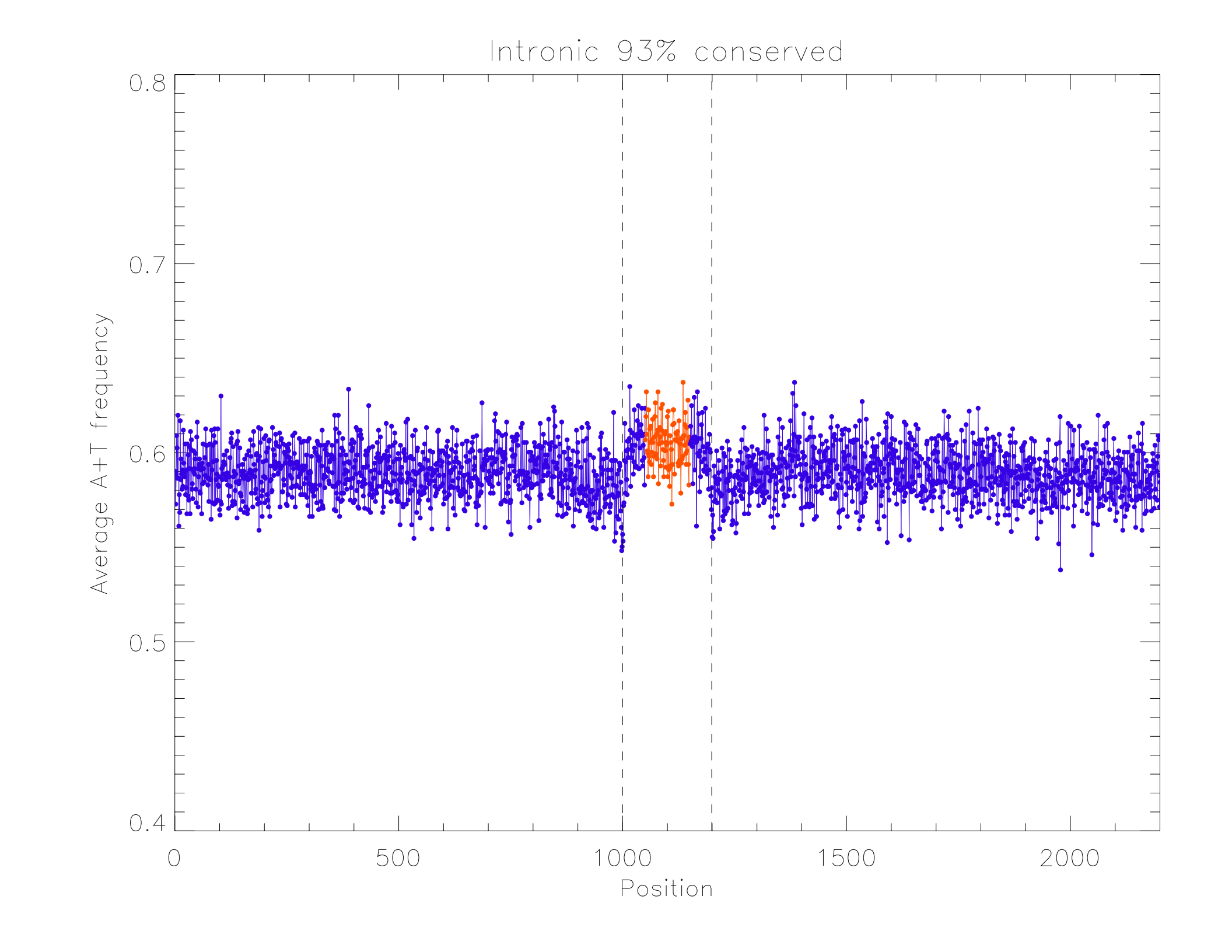

Intronic 98% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

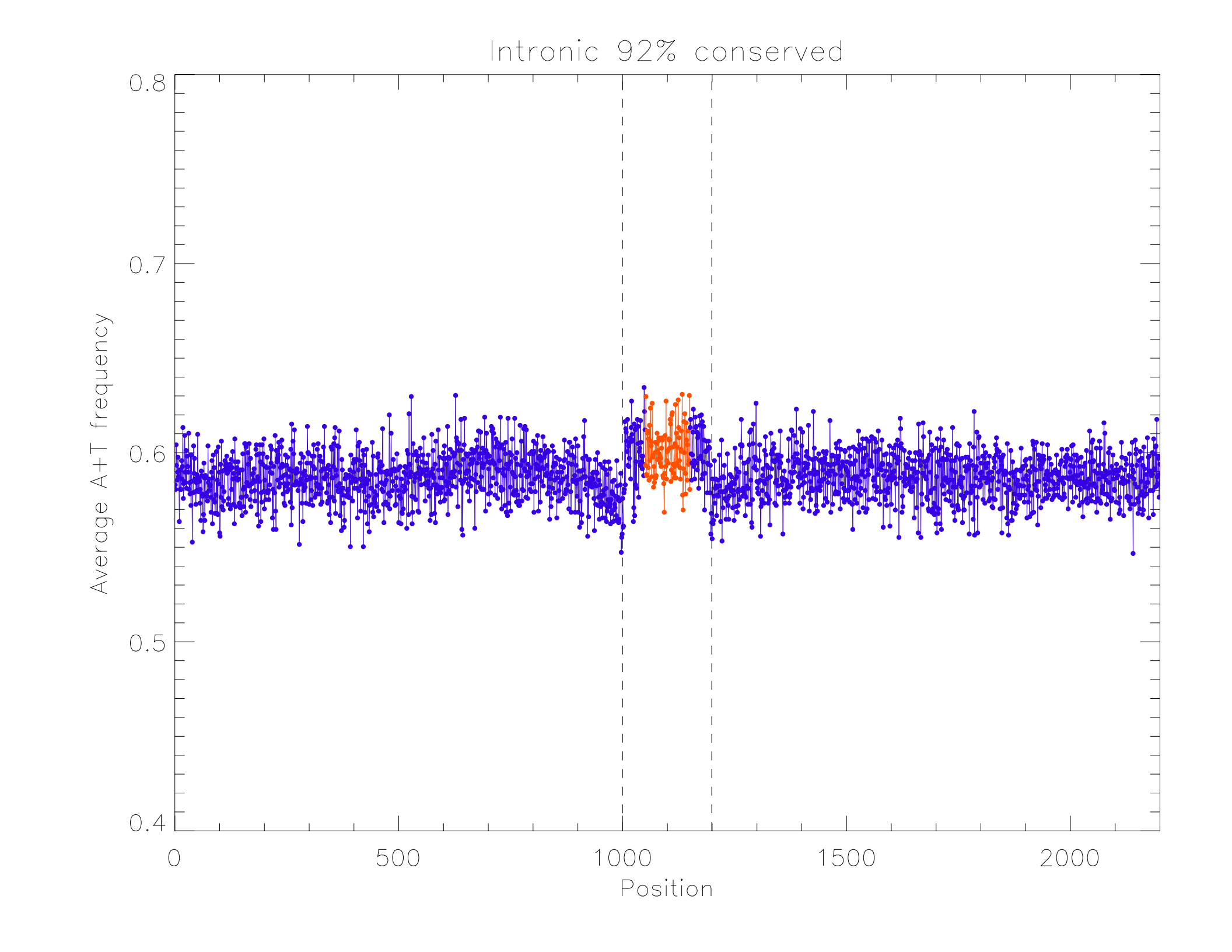

Intronic 97% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

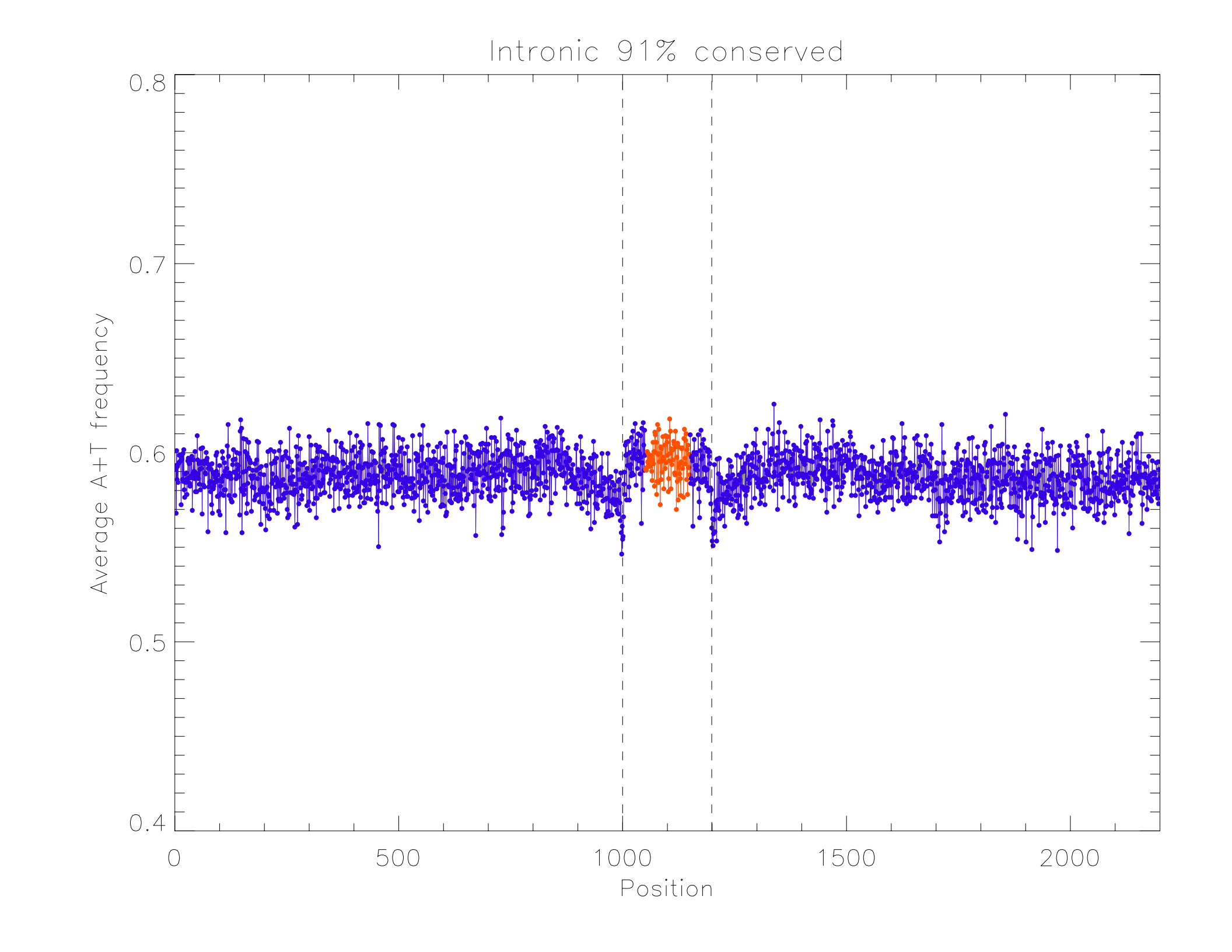

Intronic 96% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

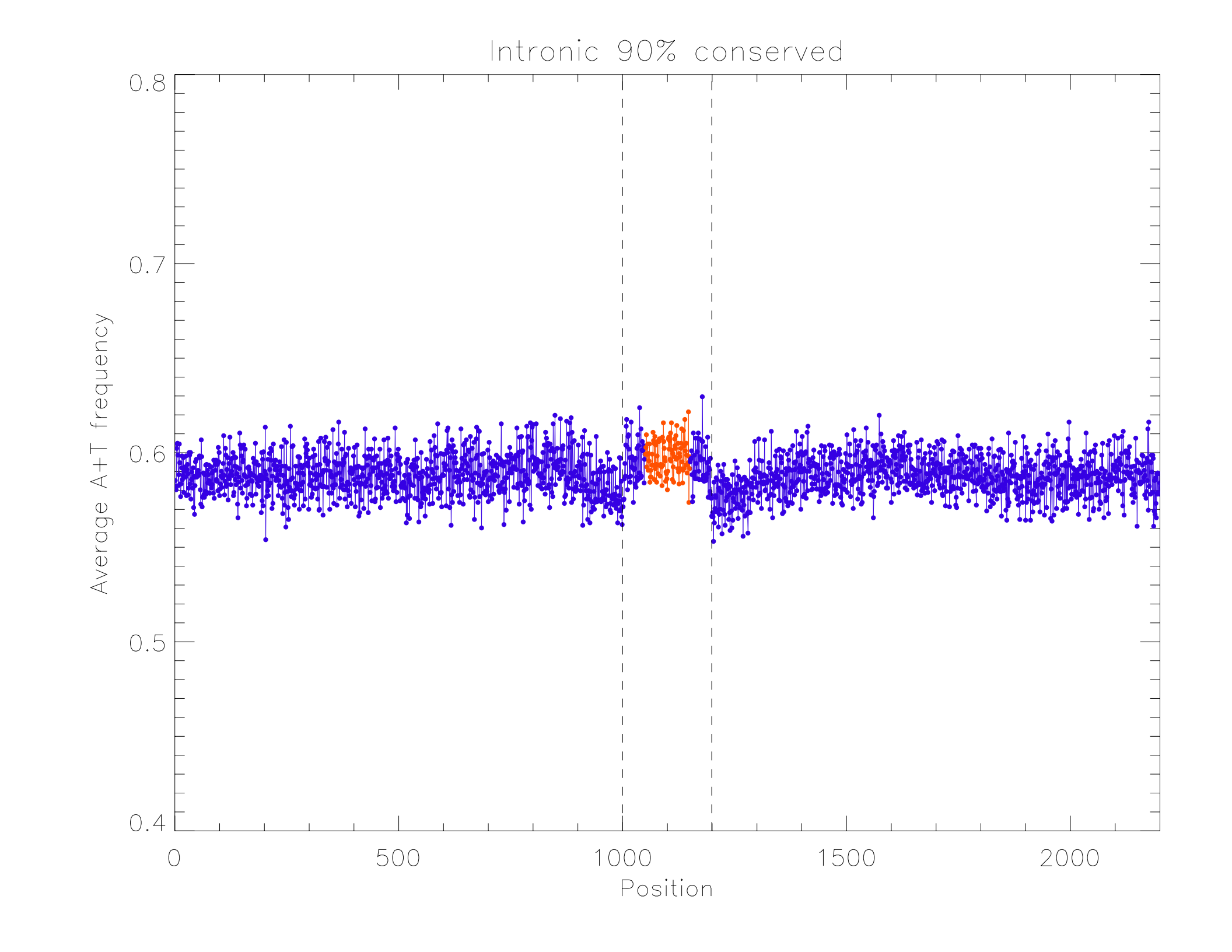

Intronic 95% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

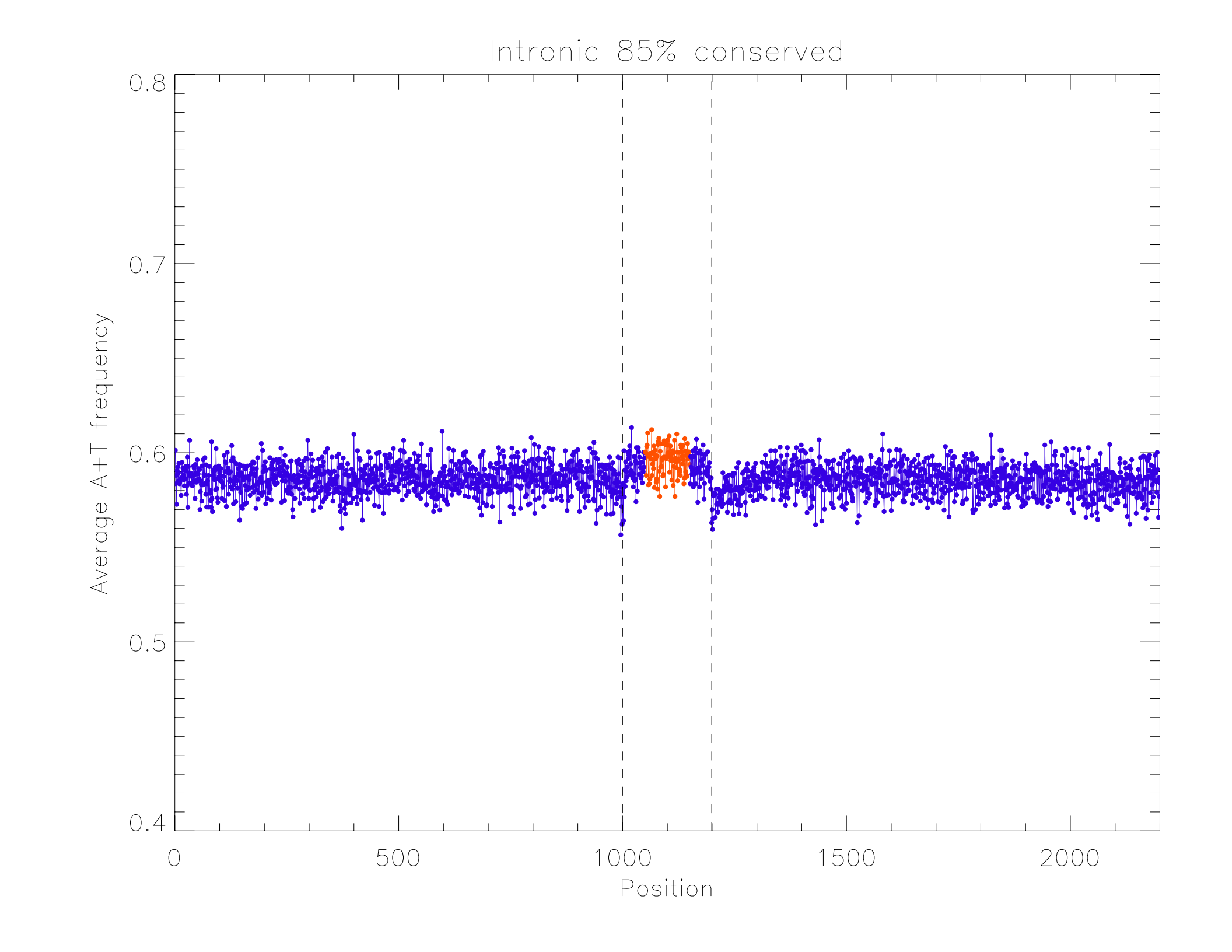

Intronic 94% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

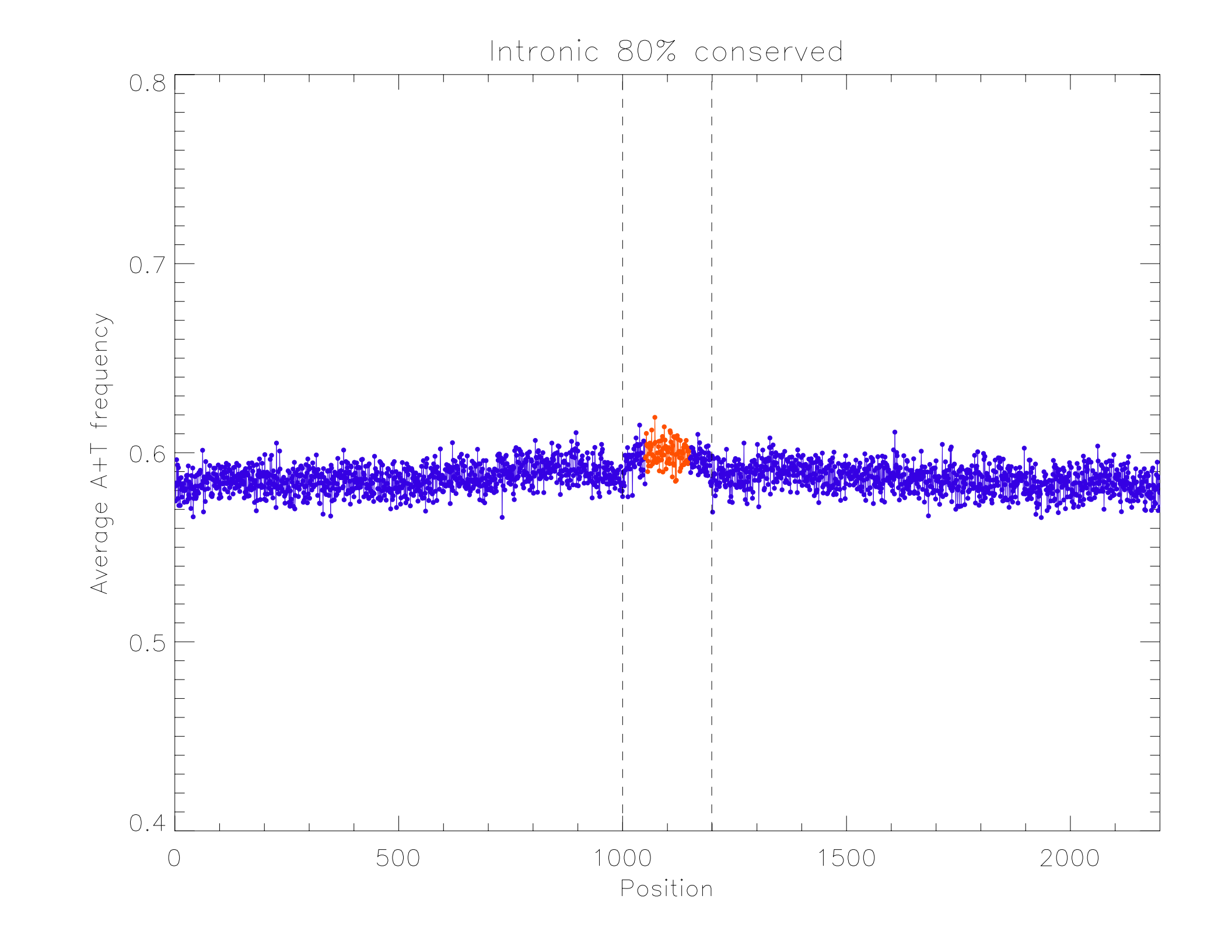

Intronic 93% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

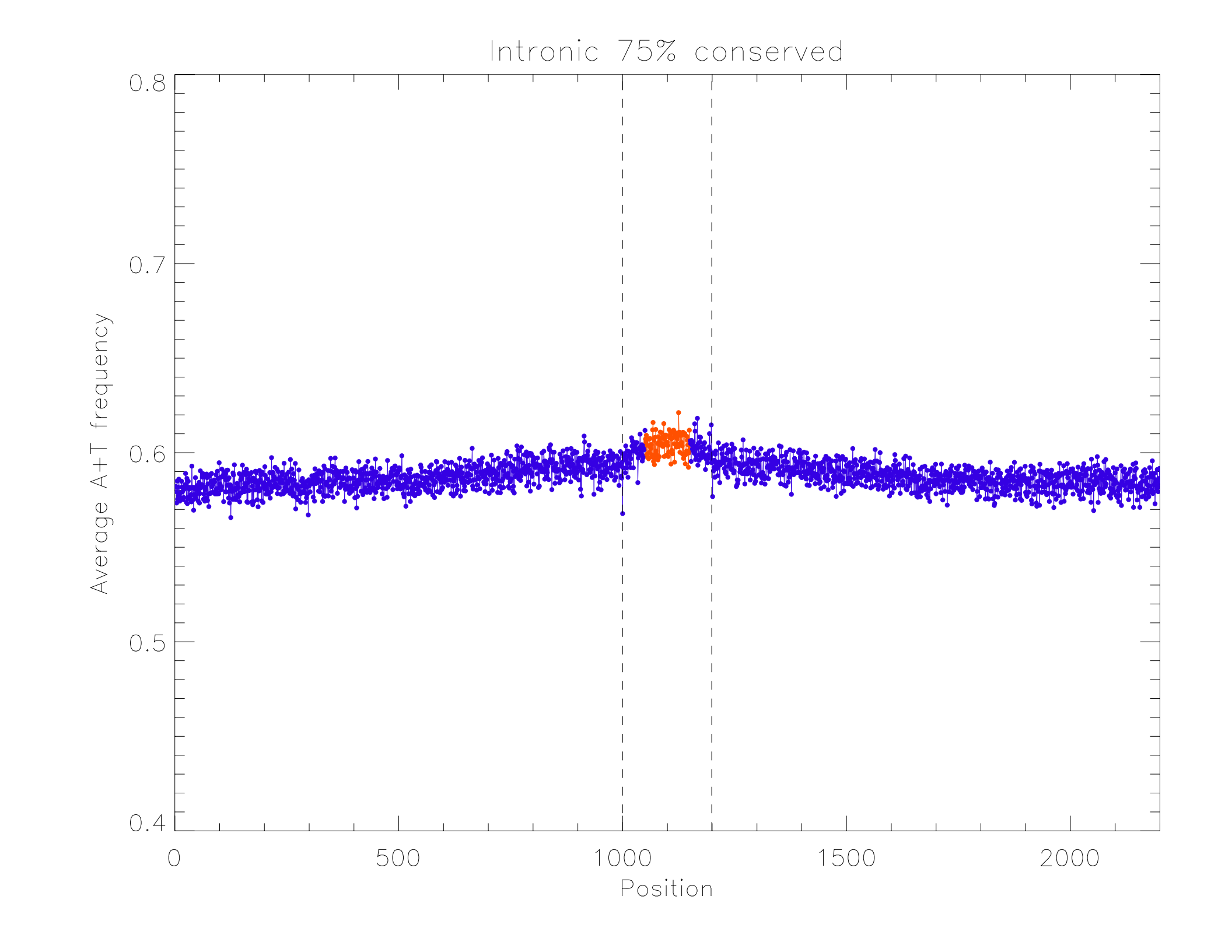

Intronic 92% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

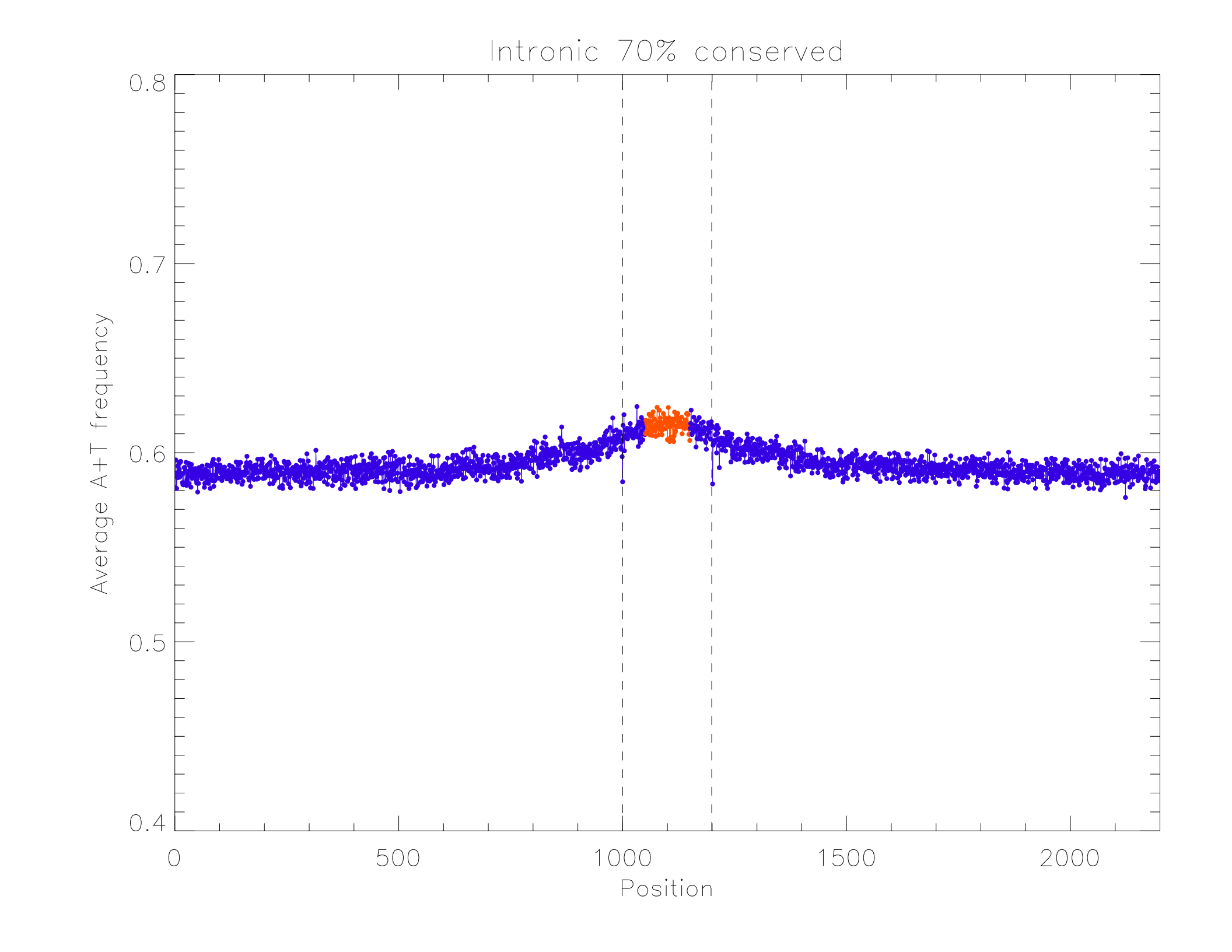

Intronic 91% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

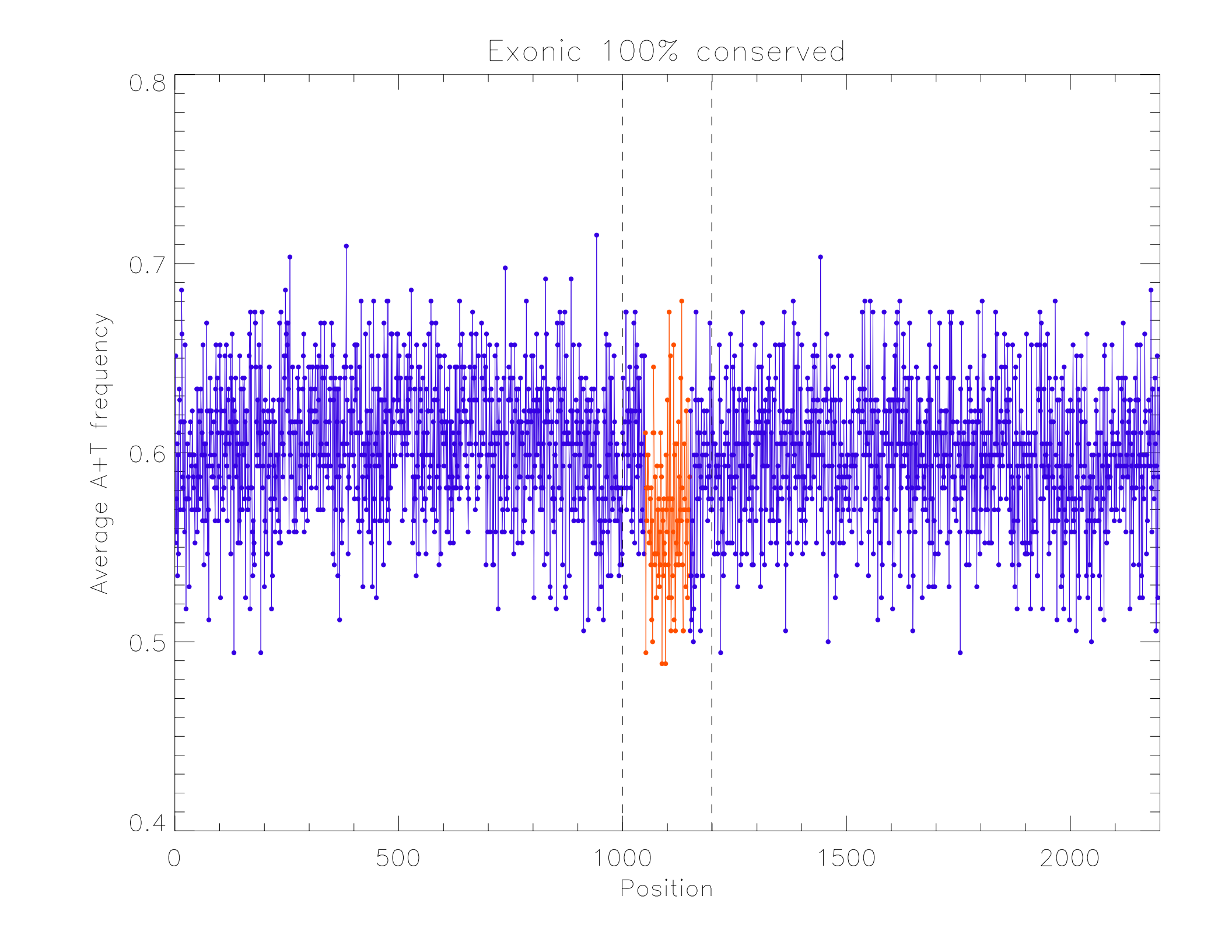

Intronic 90% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

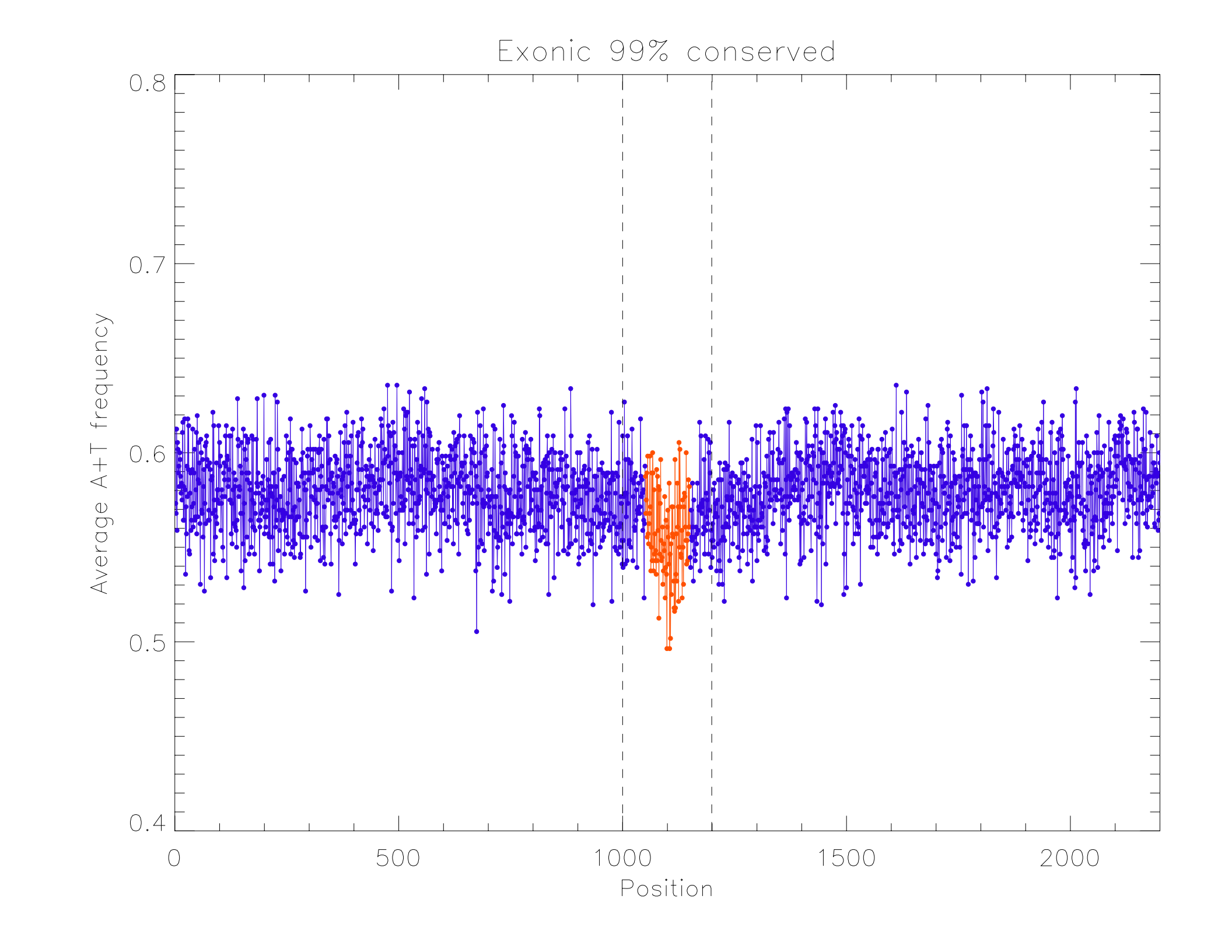

Intronic 85% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

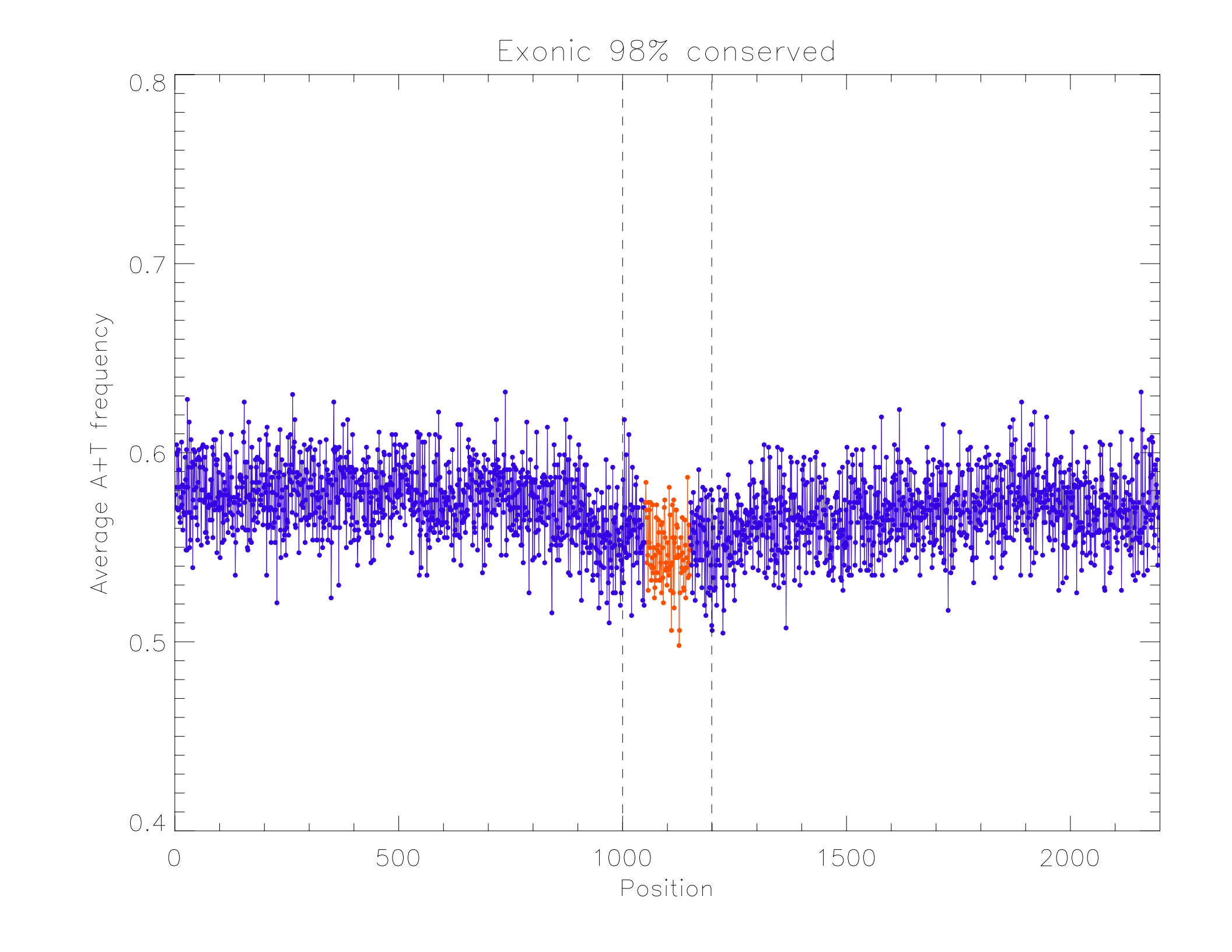

Intronic 80% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

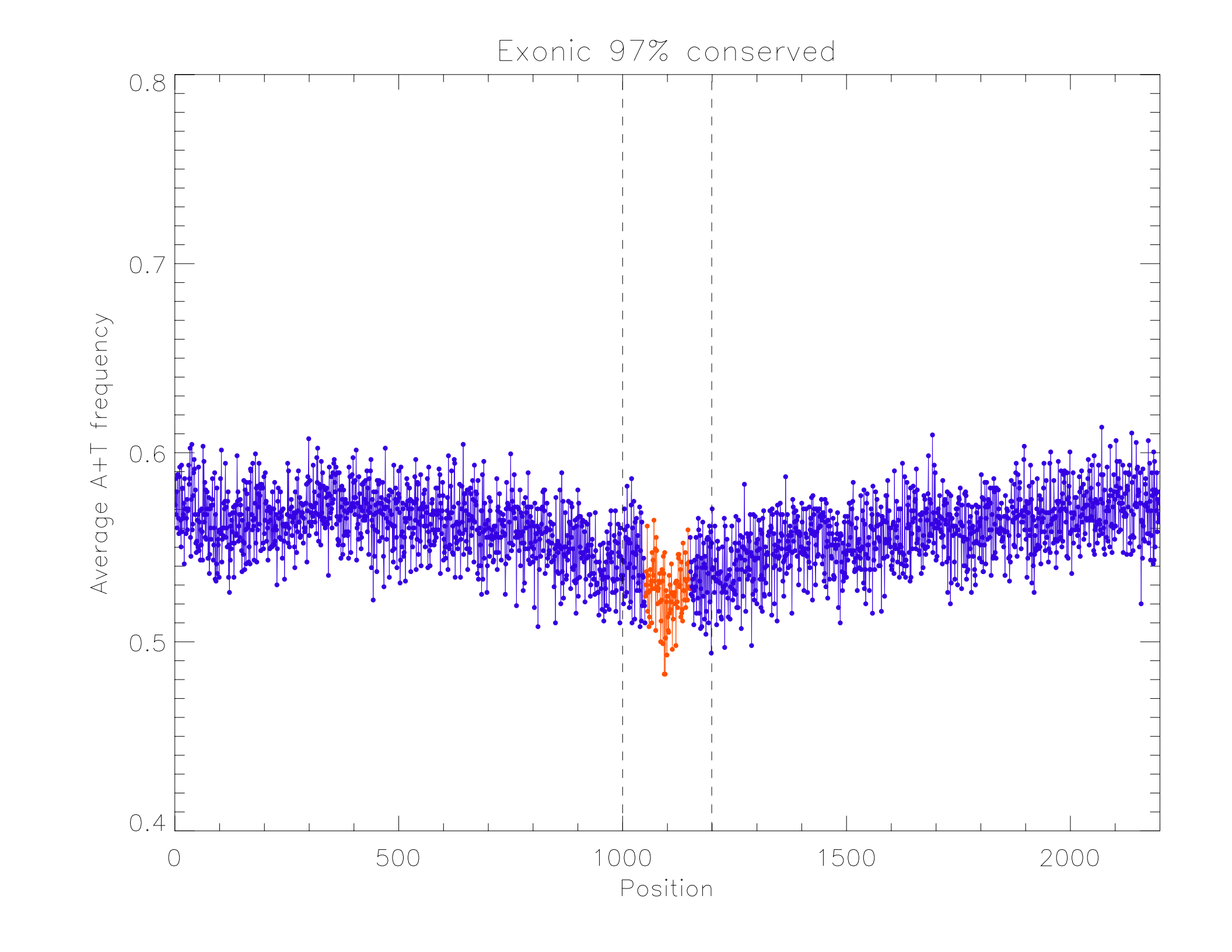

Intronic 75% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

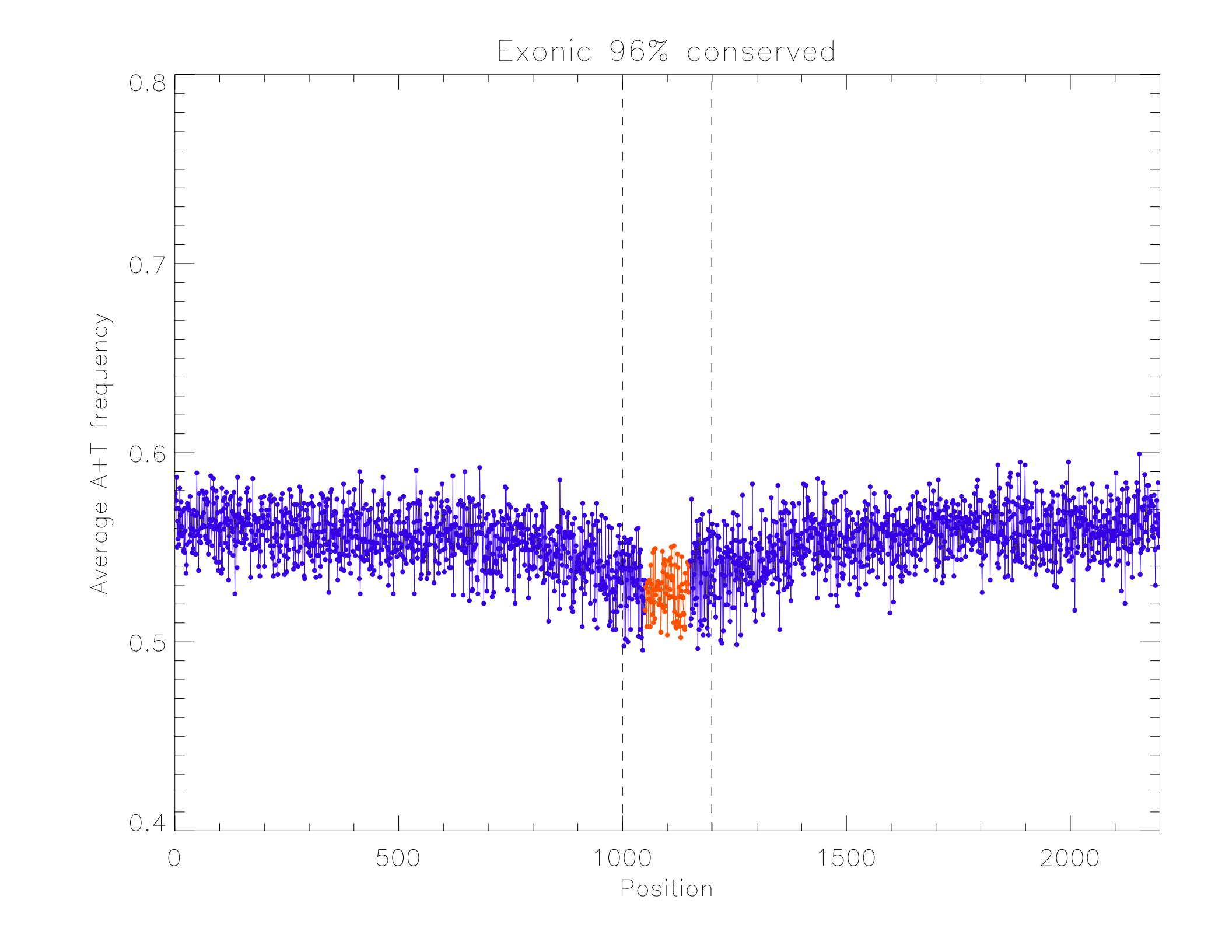

Intronic 70% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

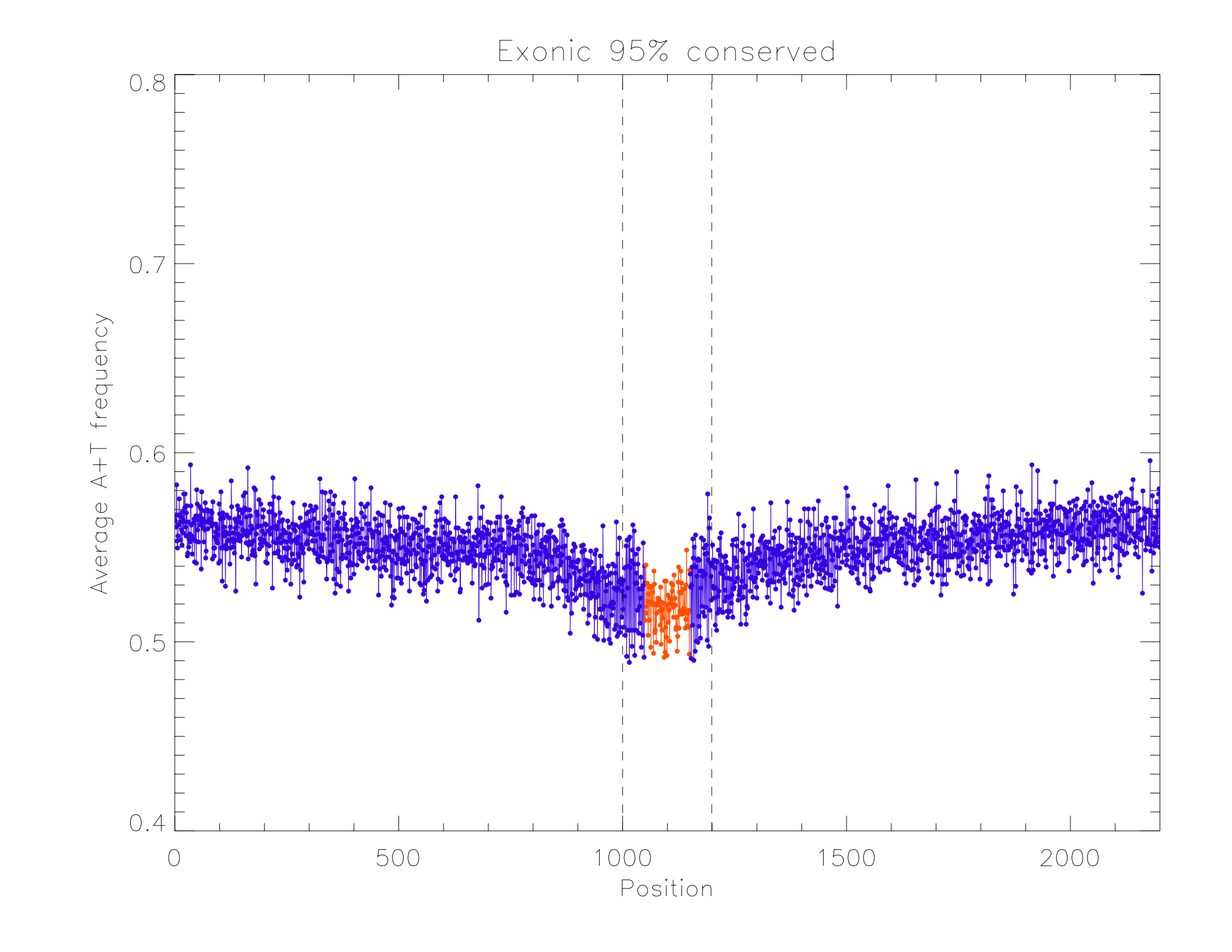

Exonic 100% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

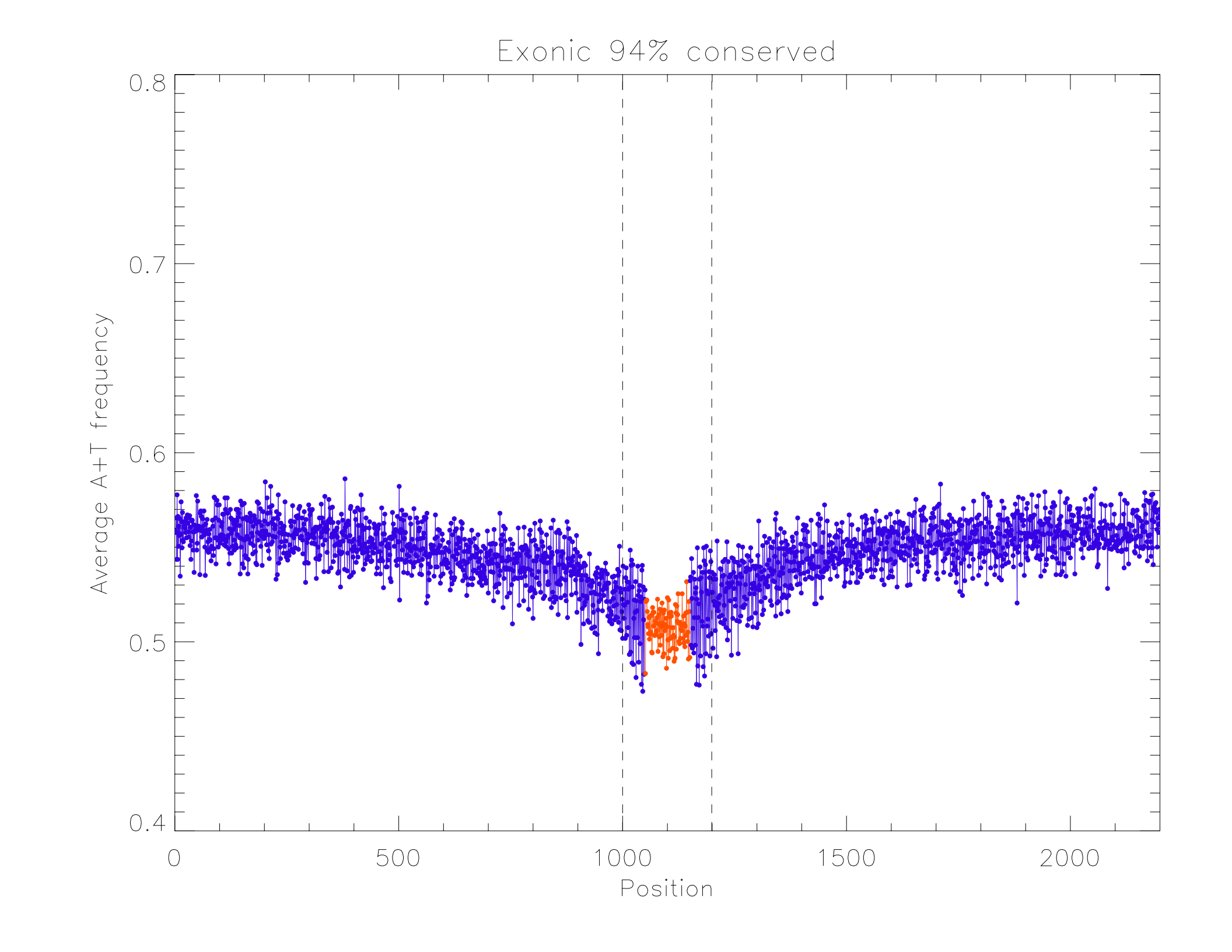

Exonic 99% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

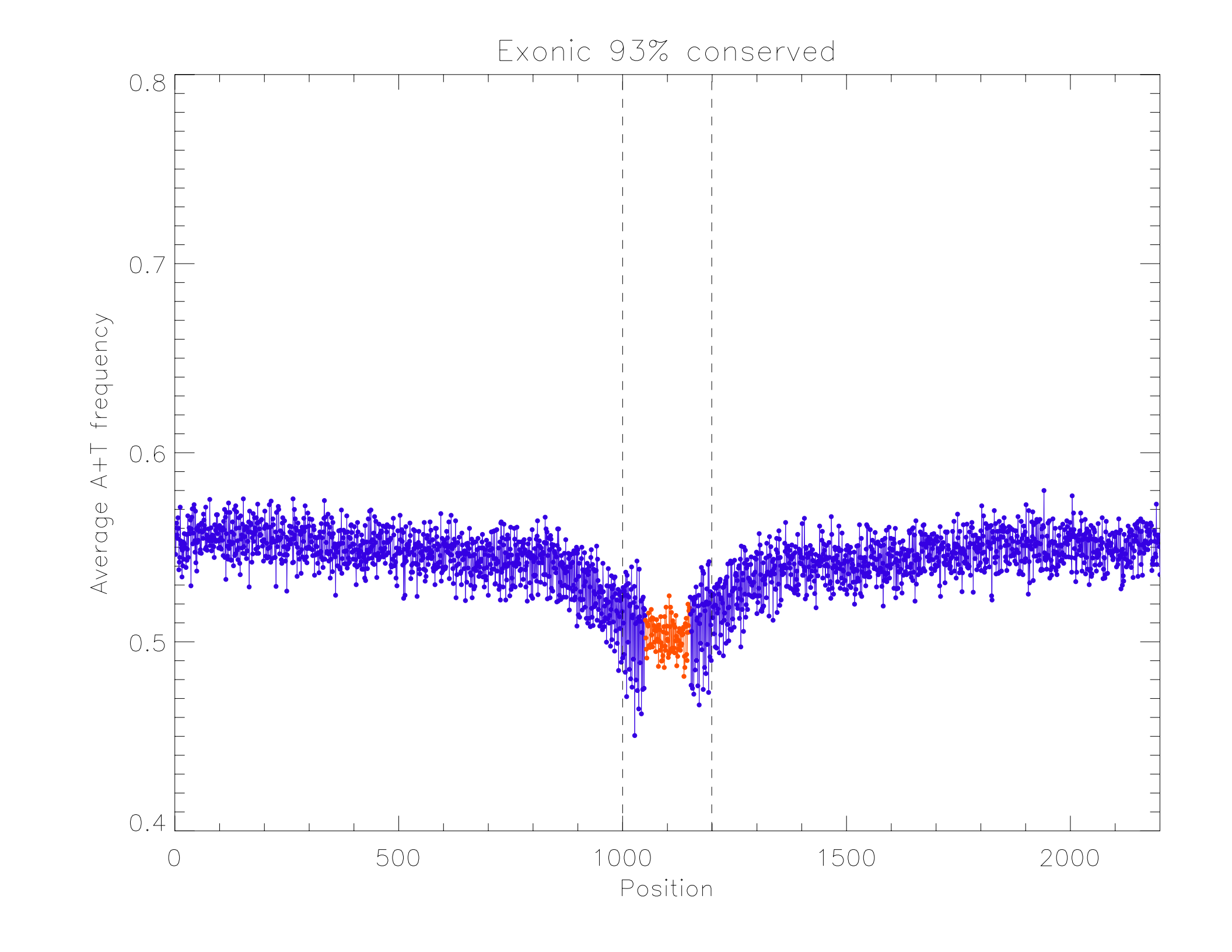

Exonic 98% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

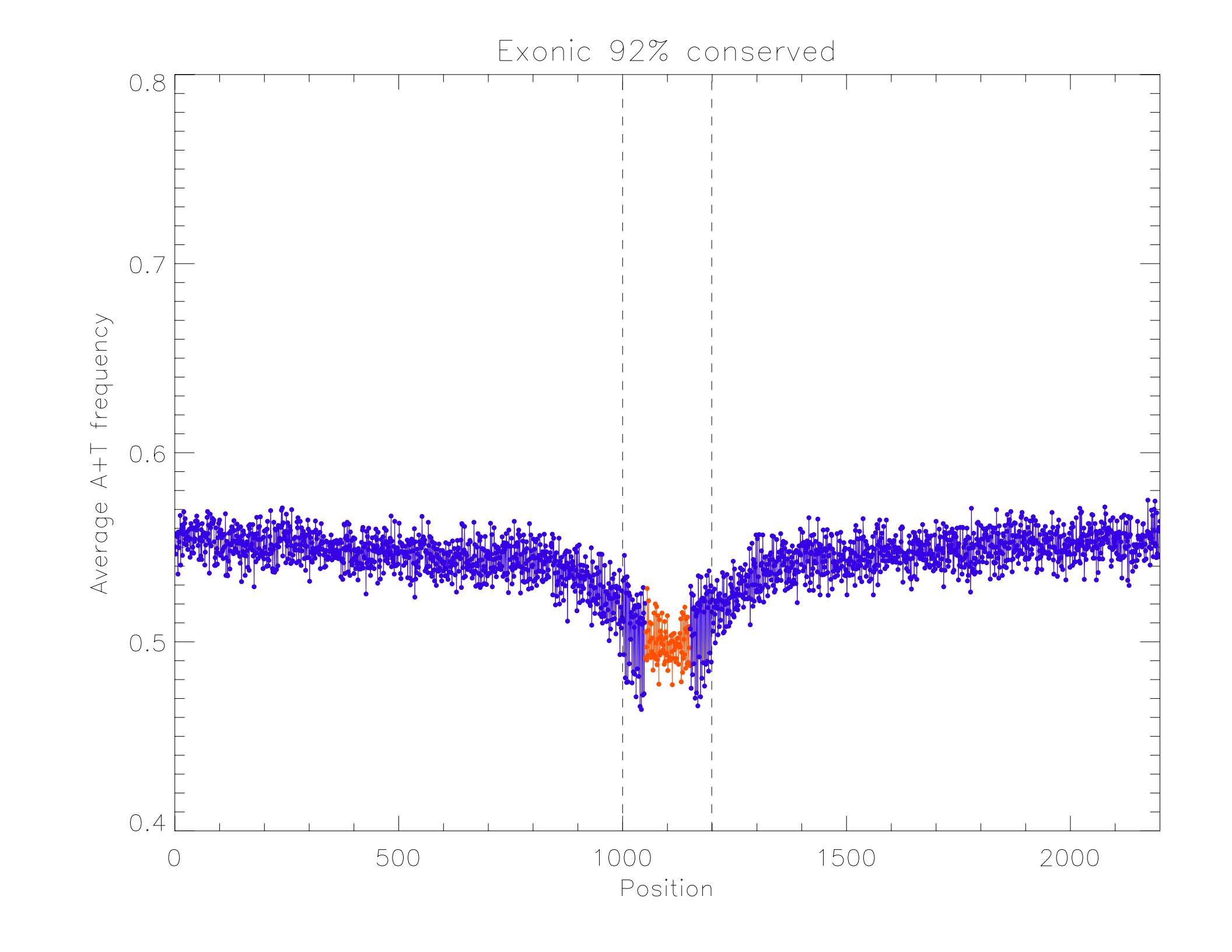

Exonic 97% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

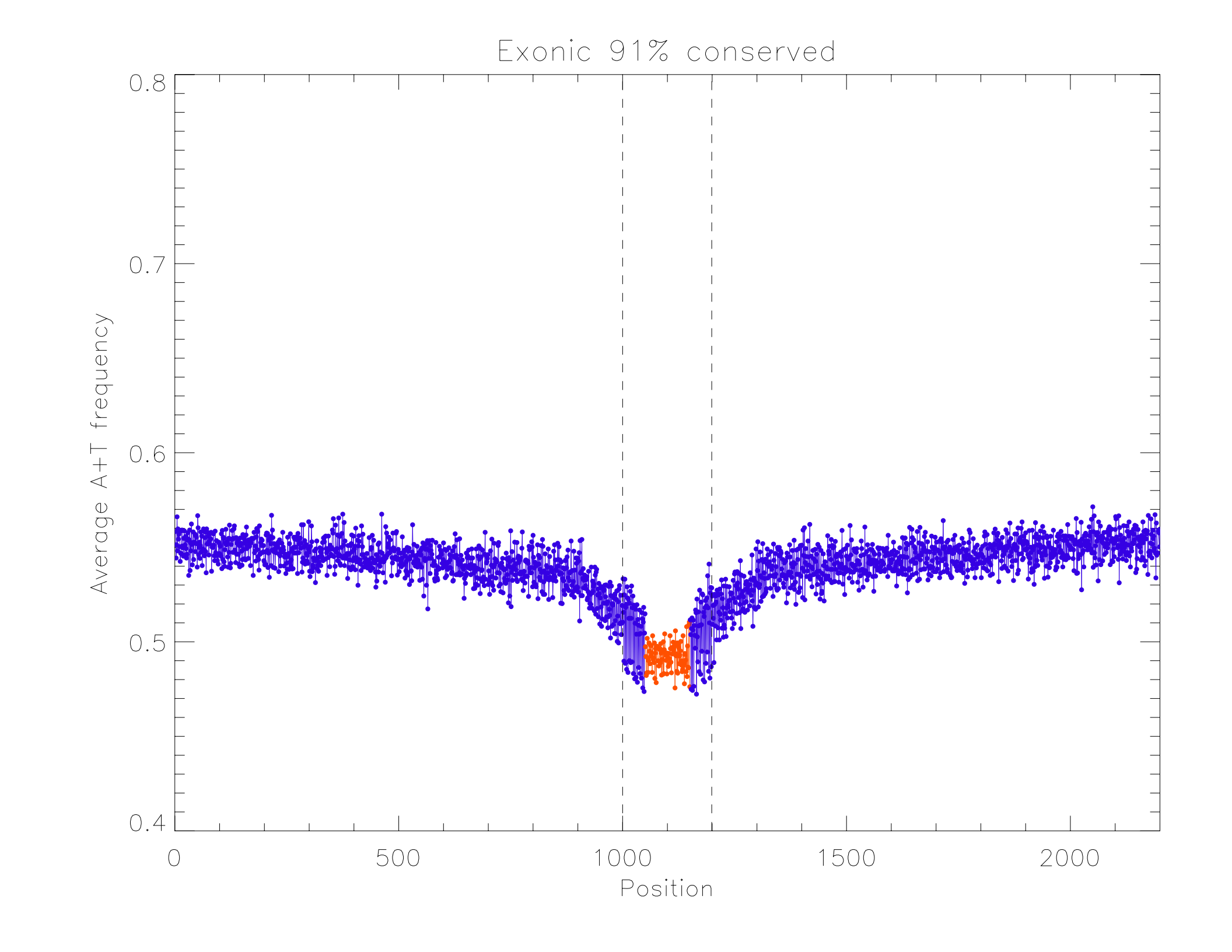

Exonic 96% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

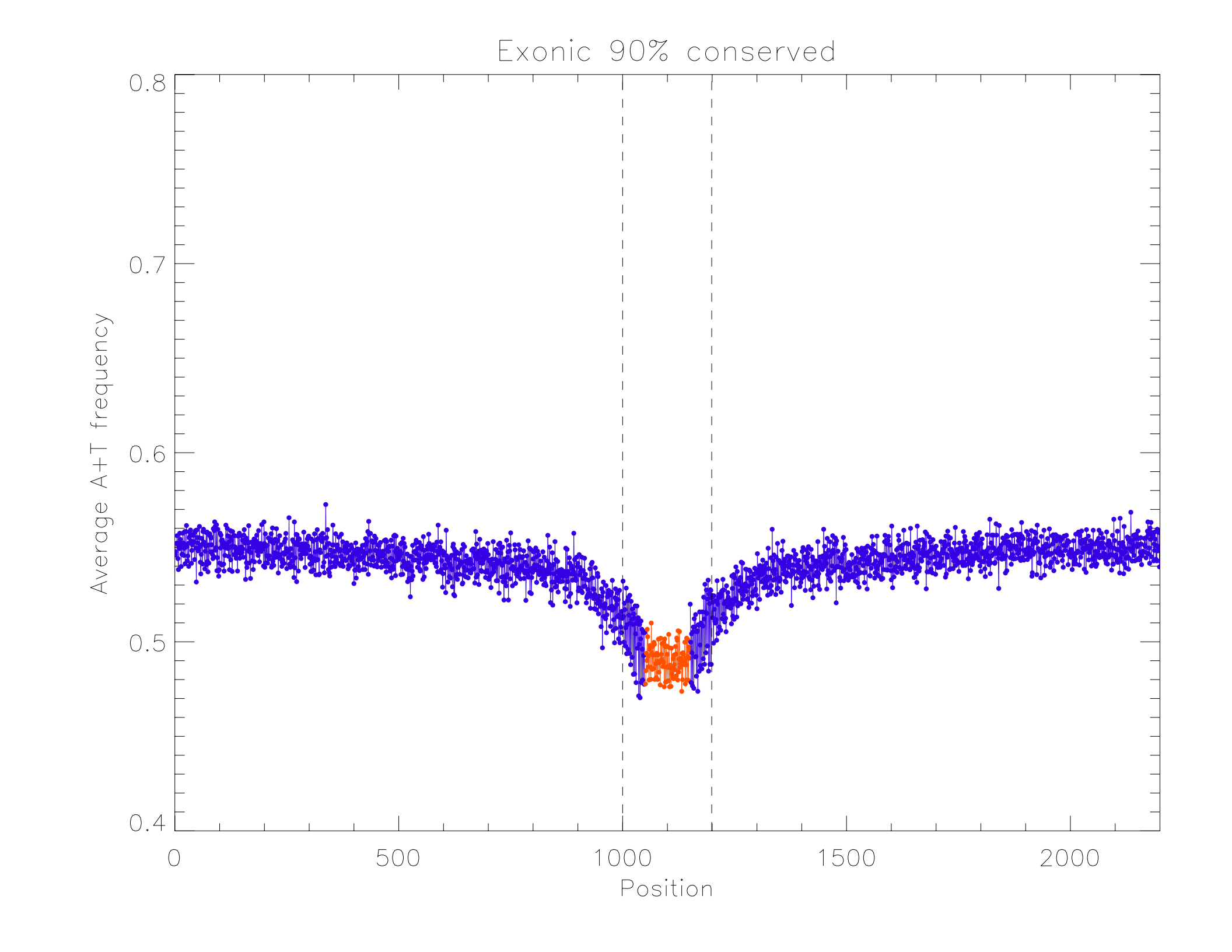

Exonic 95% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

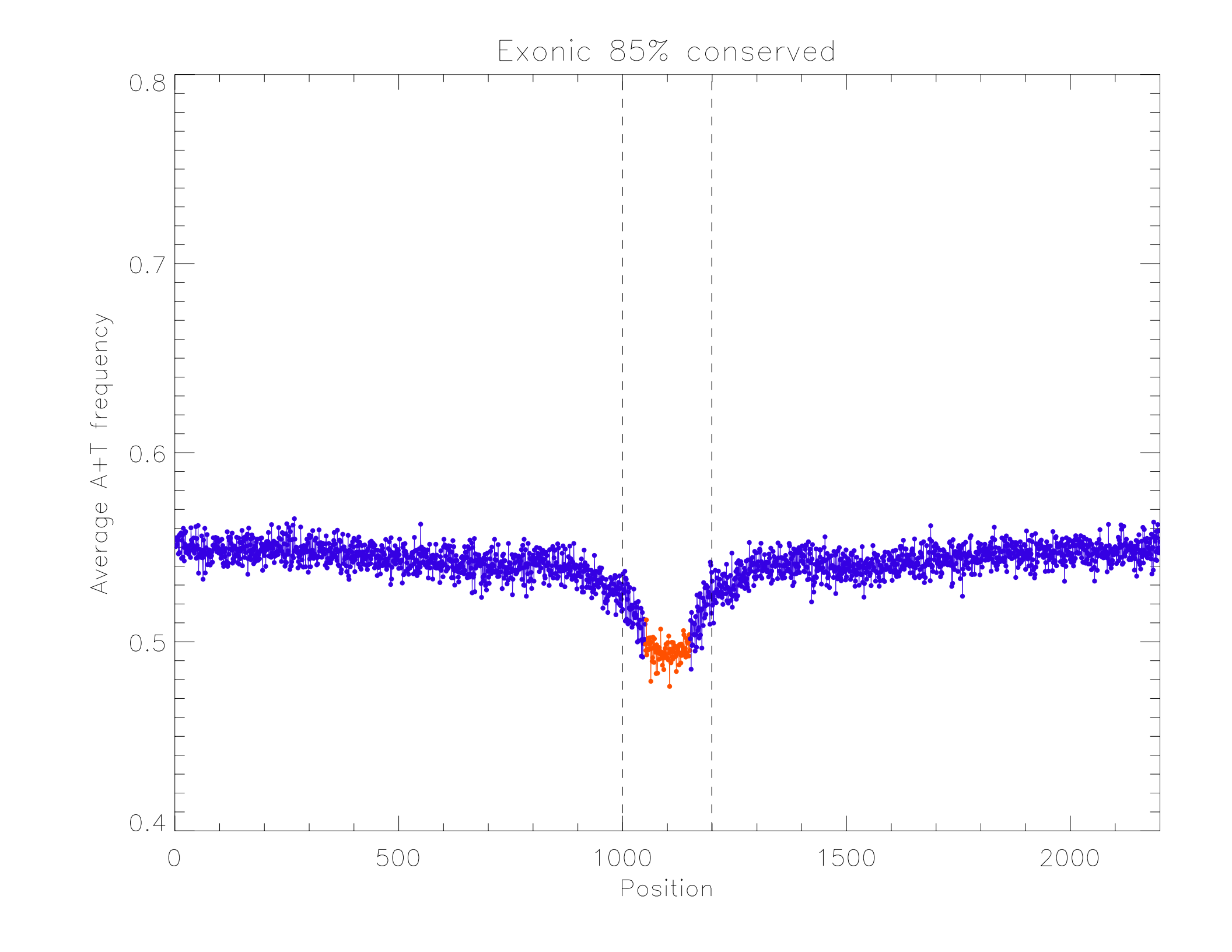

Exonic 94% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

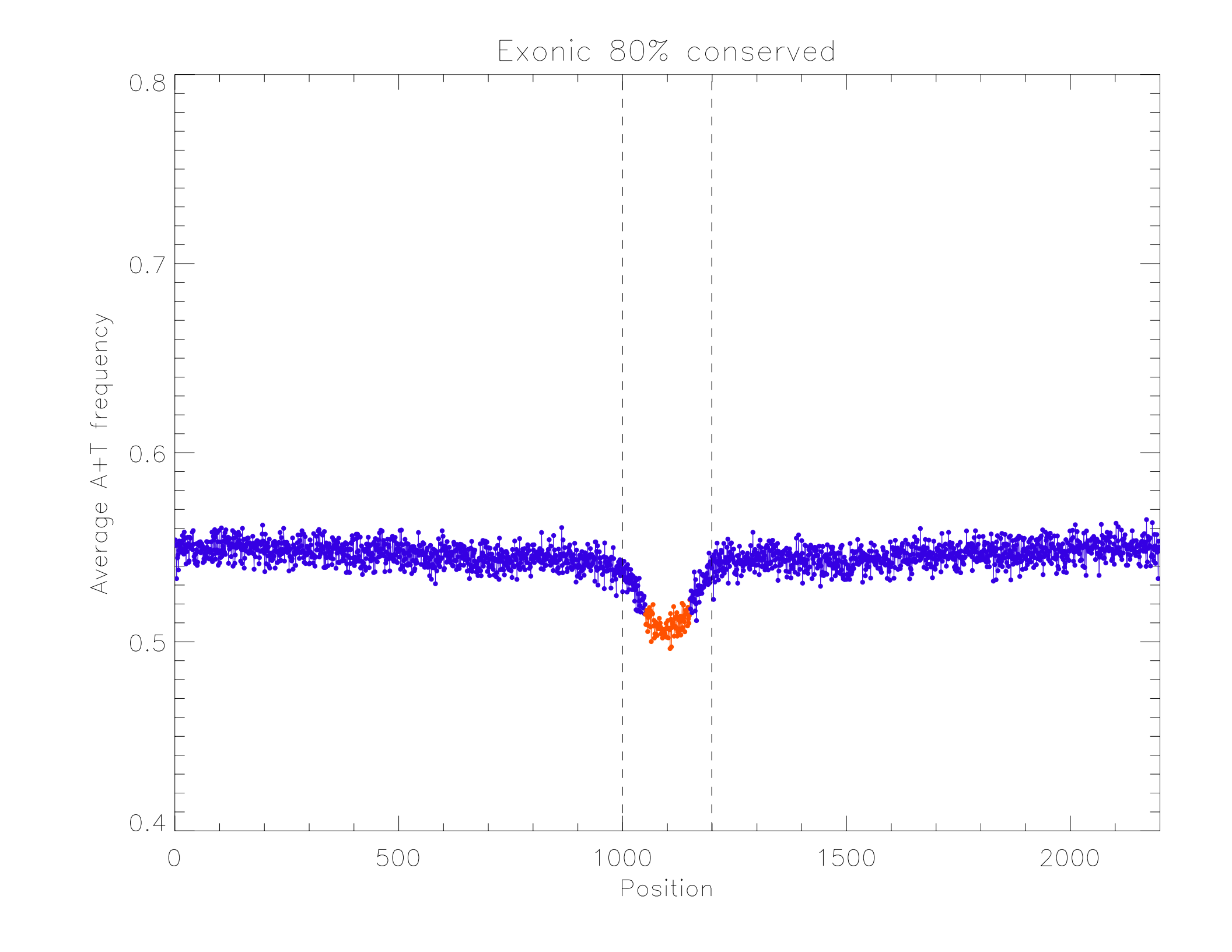

Exonic 93% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position

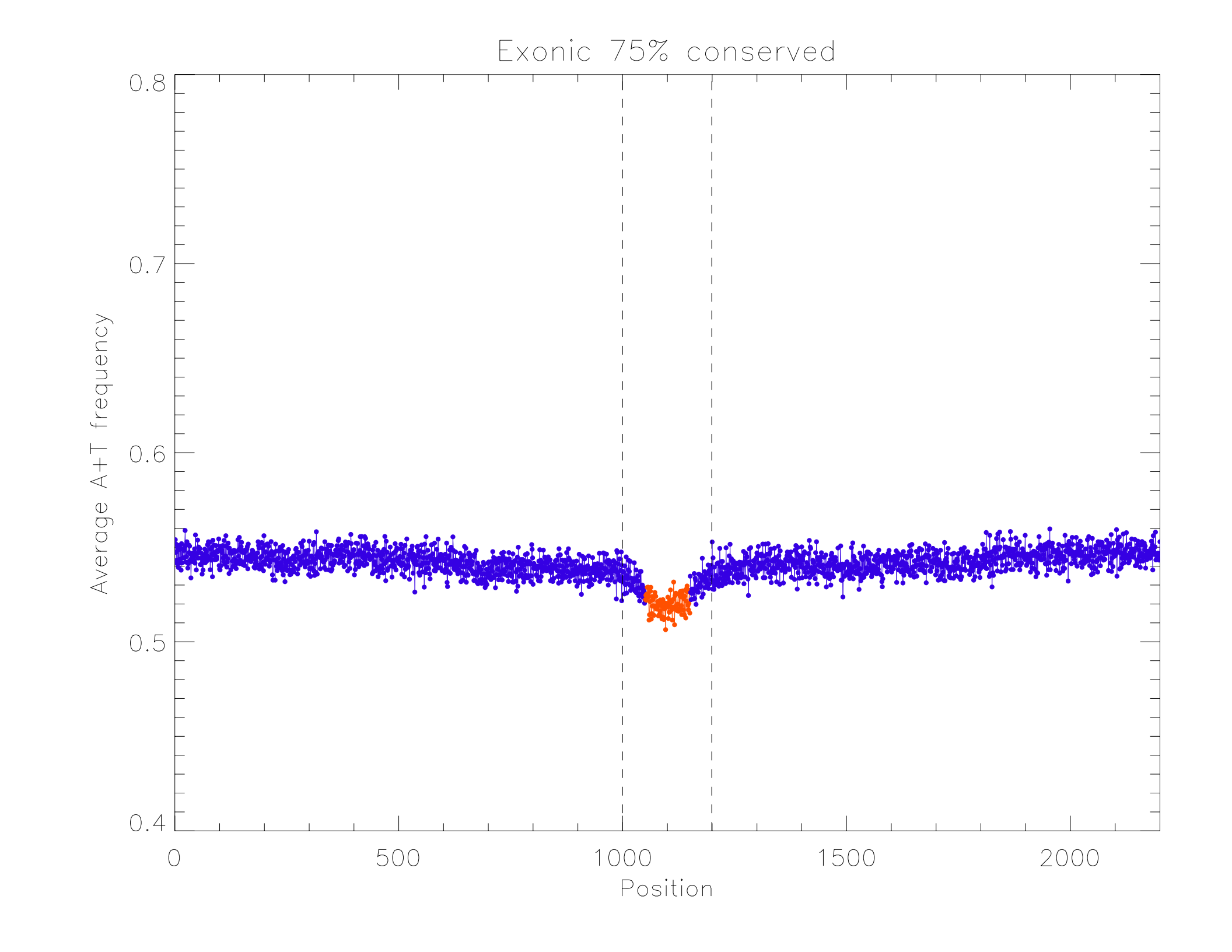

Exonic 92% conserved
0.8
0.7
0.6
0.5
0.4
Average A+T frequency
0
500
1000
1500
2000
Position

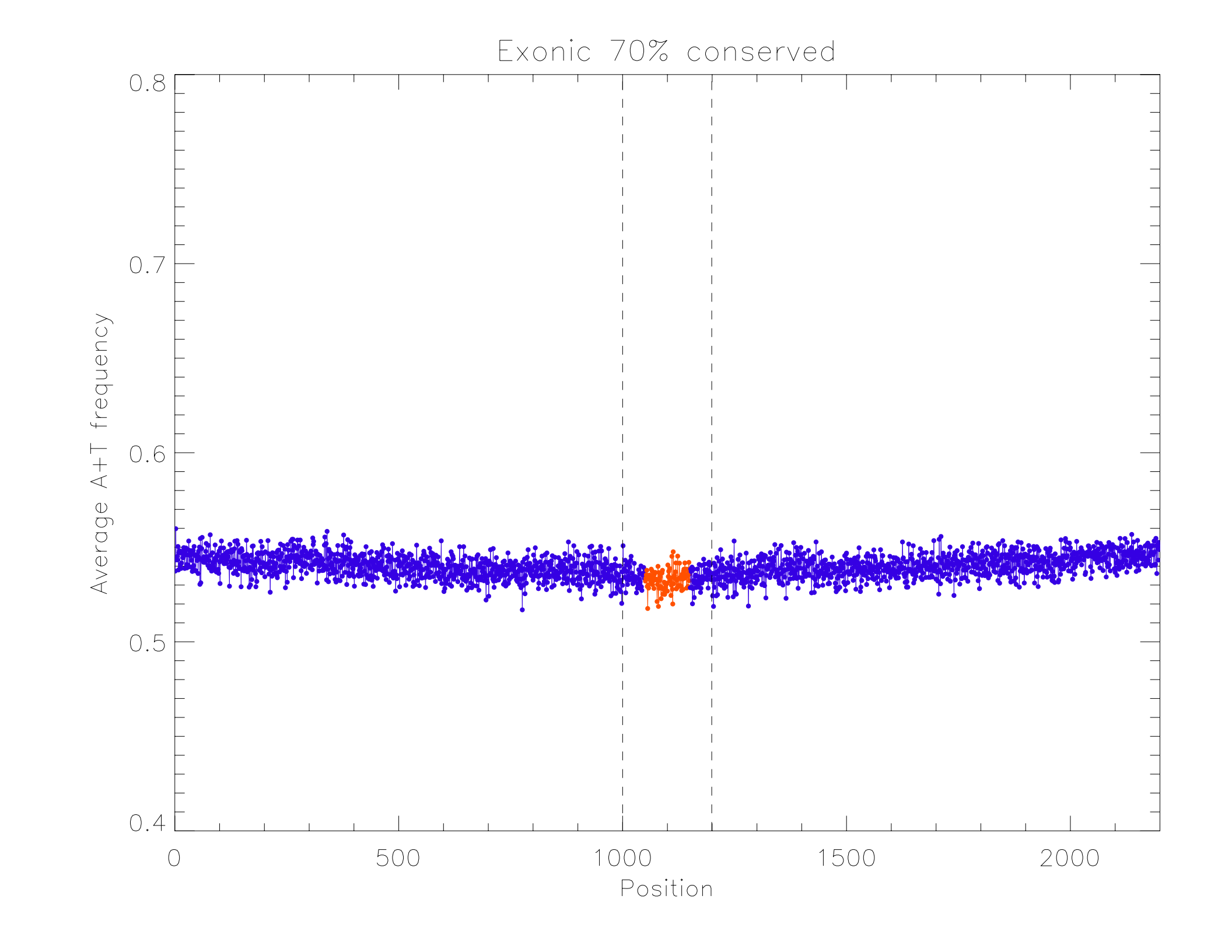

Exonic 91% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position


Exonic 90% conserved
0.8
0.7
0.6
0.5
0.4
Average A+T frequency
0
500
1000
1500
2000
Position


Exonic 85% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position


Exonic 80% conserved
Average A+T frequency
0.8
0.7
0.6
0.5
0.4
0
500
1000
1500
2000
Position


Exonic 75% conserved
0.8
0.7
0.6
0.5
0.4
Average A+T frequency
0
500
1000
1500
2000
Position


Exonic 70% conserved
0.8
0.7
0.6
0.5
0.4
Average A+T frequency
0
500
1000
1500
2000
Position