

Supplementary Information 4

Distribution of Phred/Phrap scores in the FANTOM2 sequences.

Phred/Phrap score	The number and percentage of bases in each sequence set					
	FANTOM1		FANTOM2new		FANTOM2	
0-9	594417	2.2%	519822	0.6%	1057167	0.9%
10-19	1507659	5.7%	1268584	1.4%	2667303	2.2%
20-29	3728312	14.1%	2805944	3.0%	6233971	5.2%
30-39	5587681	21.1%	6264479	6.7%	11858691	9.9%
40-49	5241080	19.8%	10582943	11.3%	15869033	13.2%
50-59	4253042	16.0%	13234397	14.2%	17499512	14.6%
60-69	2123940	8.0%	11833618	12.7%	13936480	11.6%
70-79	1623128	6.1%	14677583	15.7%	16383081	13.7%
80-89	829626	3.1%	11370425	12.2%	12370561	10.3%
>=90	1013762	3.8%	20870964	22.3%	22080926	18.4%
Total	26502647	100.0 %	93428759	100.0 %	119956725	100.0 %

The table has three distributions according to sequence sets. 'FANTOM1', 'FANTOM2-new' are distributions of Phred/Phrap scores in FANTOM1 21,076 sequence set and additional 39,694 sequences in FANTOM2 set respectively. 'FANTOM2' is a distribution in FANTOM2 60,770 sequence set.