Electronic Supplementary Material (ESI) for Analyst.

Supplementary Information

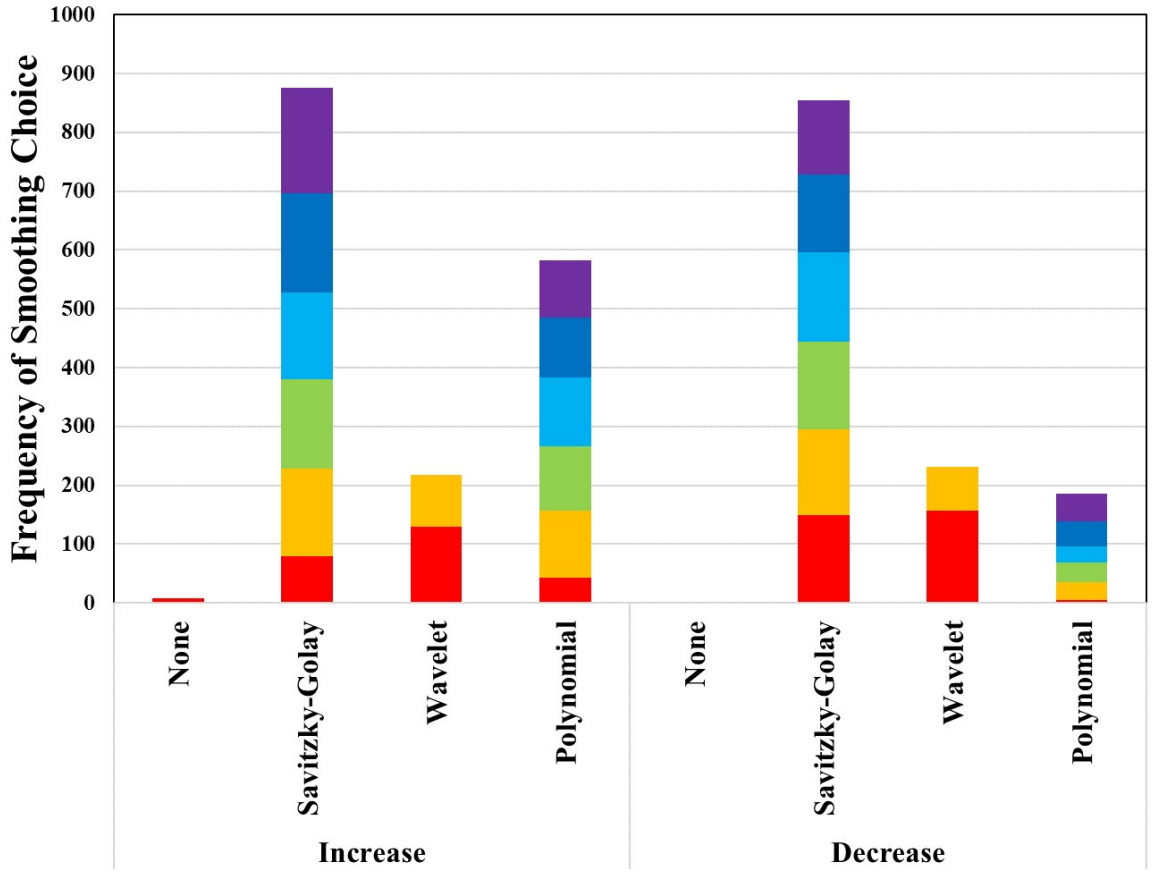

Figure S1: Exploring the impact of smoothing on classification performance using the overall metric. The frequency of each smoothing permutation is shown with regards to an **increase** in performance compared to the raw unprocessed data, and also the **decrease** in performance. If a smoothing procedure has user defined parameters, these are described within the individual bar charts (1→ 6 from red→purple).

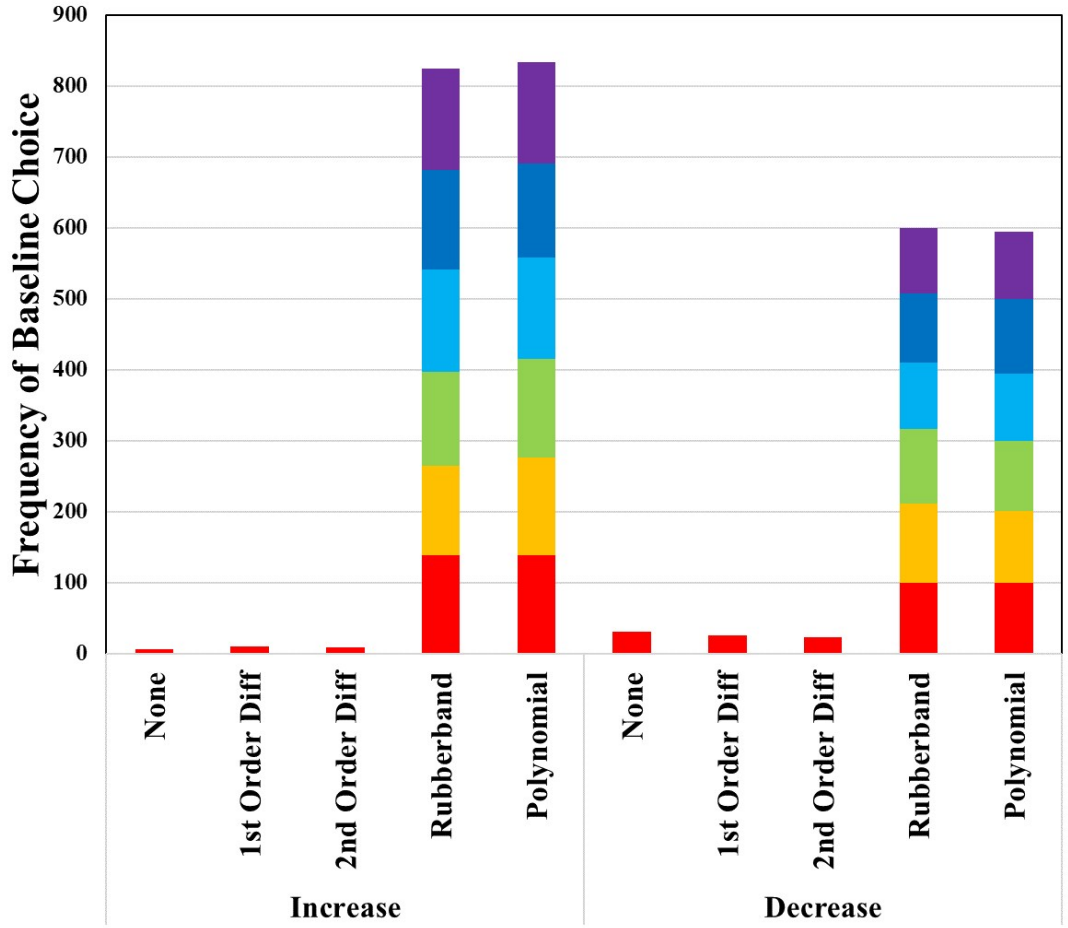

Figure S2: Exploring the impact of baseline correction on classification performance using the overall metric. The frequency of each smoothing permutation is shown with regards to an **increase** in performance compared to the raw unprocessed data, and also the **decrease** in performance. If a smoothing procedure has user defined parameters, these are described within the individual bar charts (1→ 6 from red→purple).

Table S1: Diagnostic performance metrics for the top twelve pre-processing permutations applied in alternative orders in the full separate re-analysis of the data. The values from the original BSNC ordering differ from those previously reported, due to the separate analysis of the data. Each order combination is denoted by (B) binning, (S) smoothing, (N) normalisation and (C) baseline correction. Shaded colouring is ranked separately for each metric, highlighting the highest and lowest values with green colour representing the highest value and red representing the lowest value.

	BSNC			BCSN			BCNS			BSCN			BNCS			BNSC		
Permutation	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**
100220	5.319	0.930	0.892	5.304	0.929	0.890	5.284	0.928	0.885	5.283	0.927	0.886	5.293	0.928	0.889	5.304	0.927	0.896
100020	5.262	0.925	0.884	5.262	0.925	0.884	5.262	0.925	0.884	5.262	0.925	0.884	5.262	0.925	0.884	5.262	0.925	0.884
416220	5.215	0.923	0.868	5.227	0.924	0.872	5.260	0.927	0.878	5.213	0.922	0.869	5.167	0.918	0.860	5.118	0.913	0.854
416320	5.192	0.923	0.857	5.056	0.911	0.832	5.236	0.925	0.873	5.165	0.915	0.871	5.092	0.929	0.806	5.090	0.938	0.787
100210	5.172	0.924	0.847	5.241	0.921	0.885	5.221	0.920	0.879	5.218	0.916	0.890	5.249	0.923	0.883	5.266	0.923	0.891
416210	5.122	0.924	0.824	5.136	0.917	0.849	5.212	0.923	0.868	5.137	0.915	0.857	5.104	0.919	0.832	5.083	0.917	0.828
416310	5.101	0.918	0.830	5.125	0.917	0.847	5.132	0.917	0.848	5.164	0.918	0.861	5.021	0.936	0.763	5.047	0.932	0.779
100010	5.084	0.918	0.825	5.084	0.918	0.825	5.084	0.918	0.825	5.084	0.918	0.825	5.084	0.918	0.825	5.084	0.918	0.825
424010	5.044	0.914	0.817	5.038	0.915	0.815	4.946	0.906	0.799	5.066	0.912	0.831	5.029	0.913	0.813	5.071	0.914	0.831
412136	5.038	0.917	0.807	4.950	0.915	0.777	4.919	0.915	0.765	4.956	0.912	0.787	5.032	0.920	0.798	5.051	0.923	0.797
424210	5.023	0.913	0.811	4.981	0.906	0.814	5.025	0.908	0.826	5.104	0.914	0.844	5.032	0.911	0.821	5.113	0.917	0.839
412132	5.009	0.913	0.805	5.021	0.921	0.792	5.051	0.921	0.802	5.081	0.924	0.807	5.069	0.922	0.807	5.056	0.922	0.803

Table S2: Percentage change in diagnostic performance metrics for the top twelve pre-processing permutations applied in alternative orders in the full separate re-analysis of the data. Here you can see that Each order combination is denoted by (B) binning, (S) smoothing, (N) normalisation and (C) baseline correction. Shaded colouring is ranked separately for each metric, highlighting the highest and lowest values.

	BCSN			BCNS			BSCN			BNCS			BNSC		
Permutation	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**
100220	-0.281	-0.153	-0.210	-0.666	-0.300	-0.778	-0.684	-0.344	-0.751	-0.487	-0.290	-0.418	-0.290	-0.396	0.439
100020	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
416220	0.233	0.117	0.505	0.870	0.397	1.258	-0.042	-0.042	0.208	-0.925	-0.474	-0.857	-1.852	-1.076	-1.581
416320	-2.606	-1.297	-2.940	0.858	0.194	1.867	-0.515	-0.919	1.610	-1.917	0.663	-5.968	-1.957	1.583	-8.201
100210	1.327	-0.276	4.437	0.934	-0.415	3.817	0.876	-0.837	5.010	1.481	-0.096	4.285	1.811	-0.113	5.237
416210	0.271	-0.745	3.022	1.763	-0.183	5.311	0.296	-1.037	3.971	-0.344	-0.612	0.933	-0.758	-0.795	0.426
416310	0.469	-0.212	2.036	0.616	-0.124	2.149	1.243	-0.100	3.717	-1.562	1.953	-8.060	-1.054	1.510	-6.094
100010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
424010	-0.113	0.047	-0.283	-1.940	-0.930	-2.238	0.428	-0.182	1.760	-0.299	-0.106	-0.450	0.542	-0.046	1.688
412136	-1.740	-0.284	-3.698	-2.360	-0.290	-5.214	-1.633	-0.611	-2.491	-0.116	0.285	-1.111	0.256	0.658	-1.175
424210	-0.848	-0.777	0.289	0.040	-0.523	1.843	1.613	0.168	3.958	0.167	-0.218	1.211	1.785	0.485	3.447
412132	0.250	0.826	-1.617	0.852	0.905	-0.352	1.443	1.199	0.338	1.204	1.004	0.322	0.949	0.968	-0.208

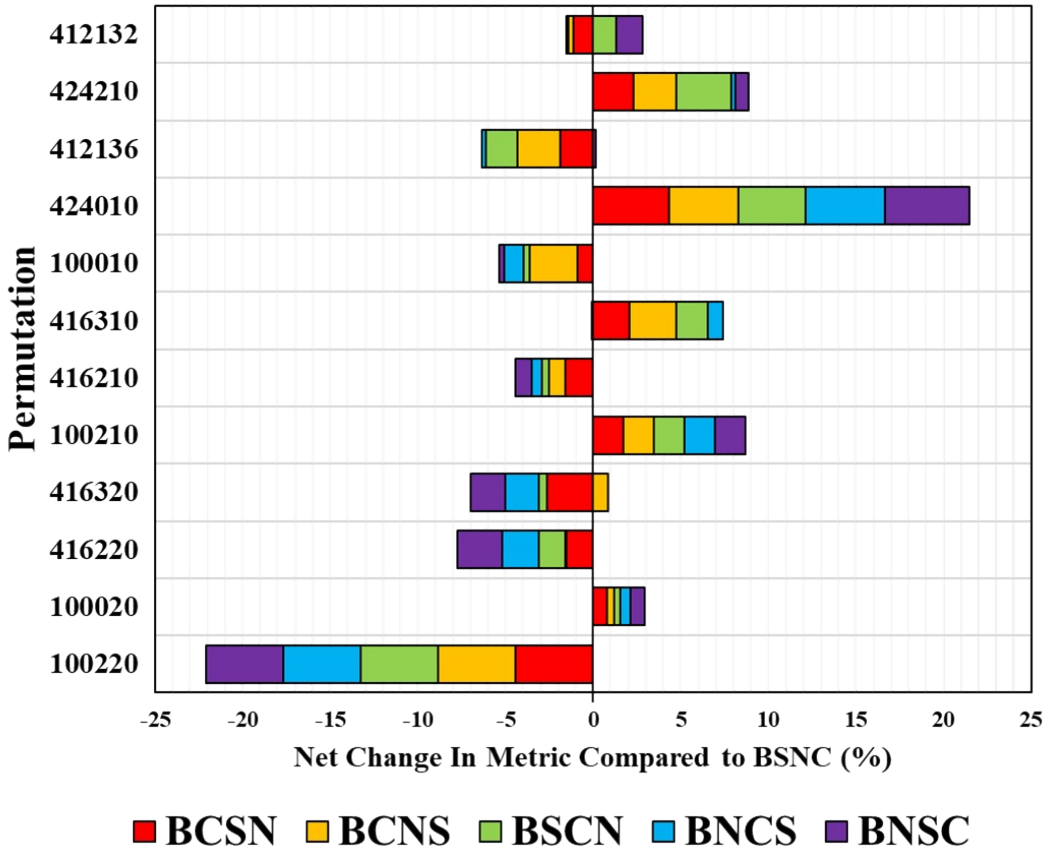

Figure S3: Percentage change in diagnostic performance metrics for the top twelve pre-processing permutations, comparing the previously reported metrics from the overall performance analysis and the newly generate re-order study. By comparing the two separate datasets, alterations in performance based on re-training the algorithm can be seen.. Here you can see that Each order combination is denoted by (B) binning, (S) smoothing, (N) normalisation and (C) baseline correction. Shaded colouring is ranked separately for each metric, highlighting the highest and lowest values.

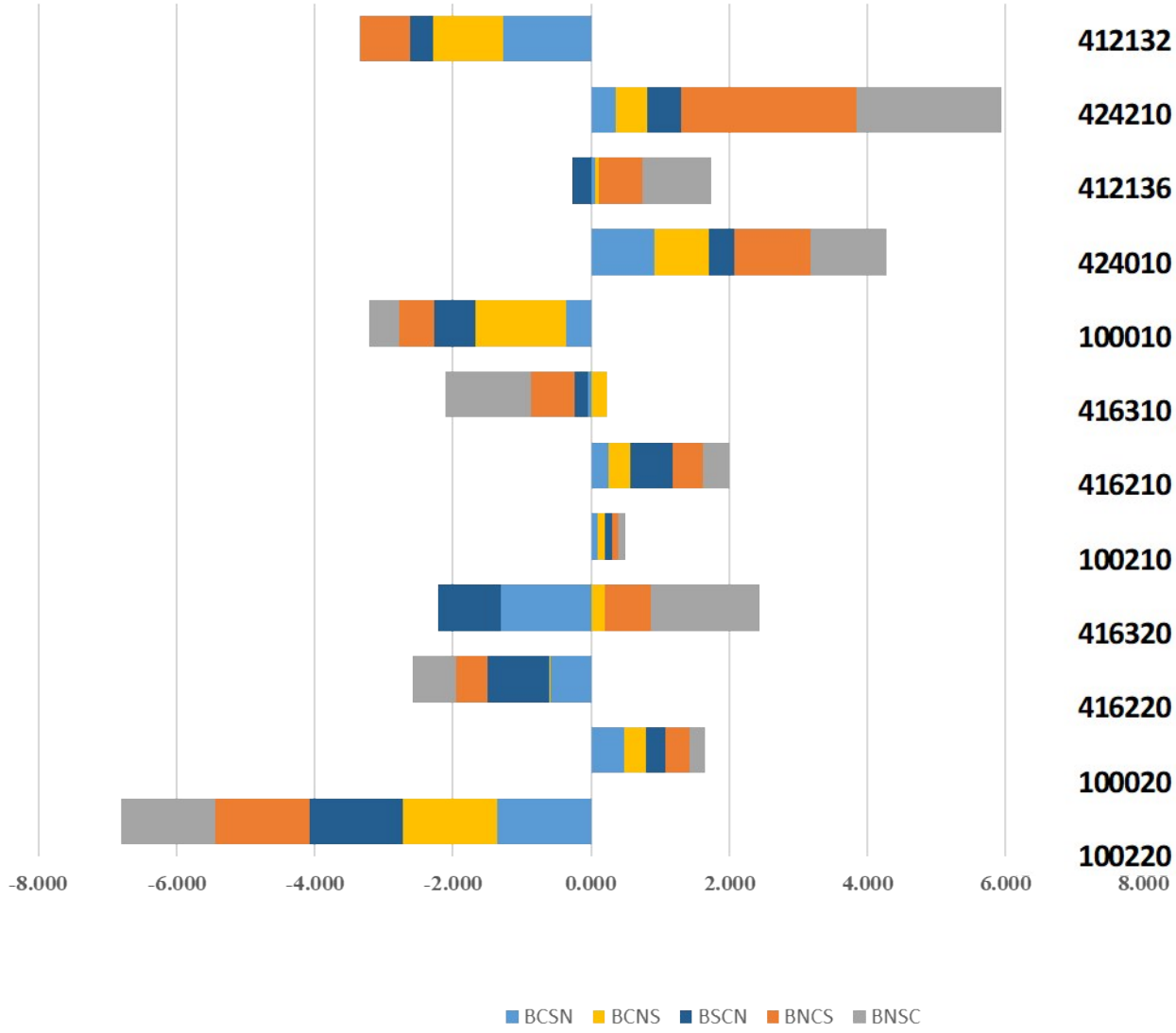

Figure S4. The comparative change (%) in sensitivity by altering the order of the top twelve pre-processing permutations. Each order combination is compared to the previously stated Binning (B), Smoothing (S), Normalisation (N) and Baseline Correction (C) order.

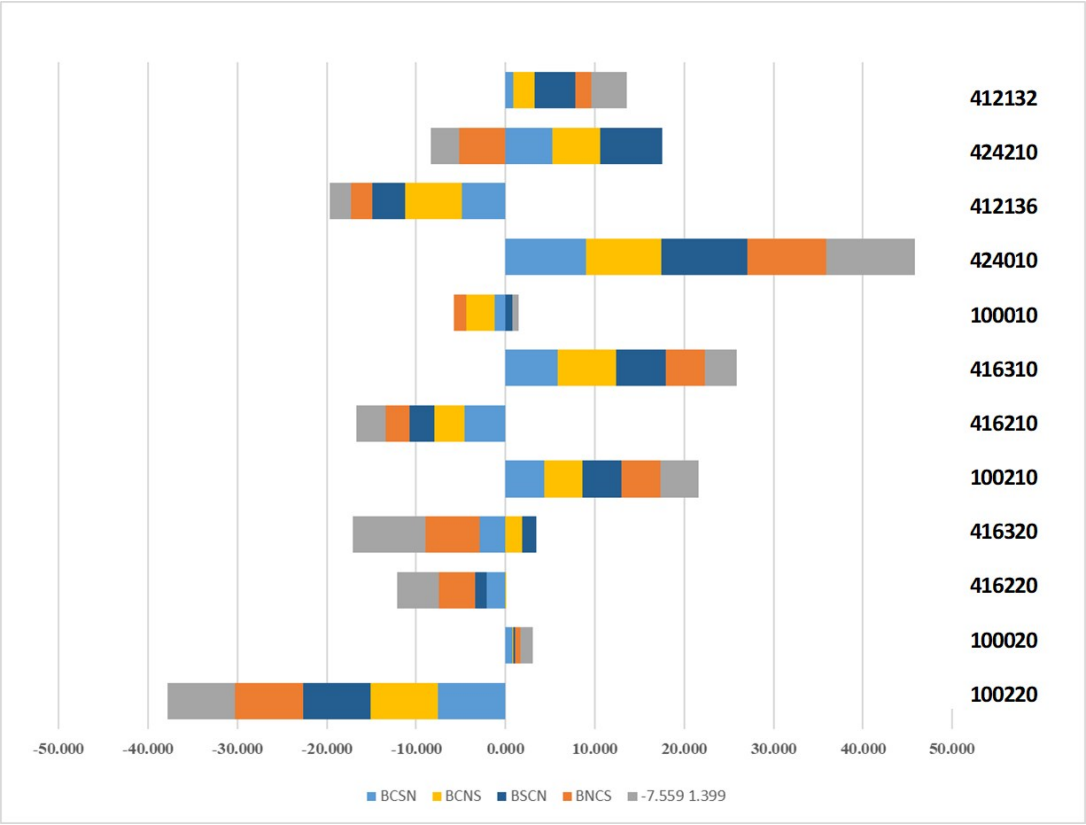

Figure S5. The comparative change (%) in specificity by altering the order of the top twelve pre-processing permutations. Each order combination is compared to the previously stated Binning (B), Smoothing (S), Normalisation (N) and Baseline Correction (C) order.

Table S3. The comparative change (%) in diagnostic performance compared to a standard random forest (RF) binary classifier by altering the order of the top twelve pre-processing permutations. (SVM – Support vector machine)

	Random Forest Fed SVM			Genetic Algorithm Fed SVM		
Permutation	**Overall**	**Sens**	**Spec**	**Overall**	**Sens**	**Spec**
100220	-0.1	0.6	-2.6	-0.4	0.4	-2.7
100020	0.5	1.4	-3.2	0.8	0.4	-0.3
416220	1.1	0.7	0.0	0.0	2.5	-6.9
416320	0.5	1.5	-3.2	-3.5	-0.5	-8.2
100210	0.7	2.6	-5.2	-0.5	1.2	-4.9
416310	0.9	3.2	-6.5	-0.6	2.9	-9.2
416210	3.1	0.3	6.1	-4.0	0.6	-11.8
100010	2.8	2.0	1.3	4.1	2.7	2.8
424210	1.2	2.7	-4.4	-5.0	-0.7	-11.0
424010	0.5	1.4	-2.8	-1.6	1.2	-7.5
412132	6.2	2.6	8.1	7.4	4.0	7.4
412136	2.4	0.7	3.5	7.2	3.6	8.1