

## SUPPLEMENTARY TABLES

**Supplementary Table 1. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – protein interactors dataset (version 1).**

<b>Interactors dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.736	0.743
Chi <sup>2</sup>	0.730	0.738
Decision Stump	0.731	0.738
Asymmetric Optimal Prediction	0.724	0.742
Log Odds Ratio	0.725	0.713
Auto-Filter	0.759	<b>0.801</b>
Baseline (no filter method)	0.717	

**Supplementary Table 2. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – GO terms dataset (version 1).**

<b>GO terms dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.753	0.744
Chi <sup>2</sup>	0.740	0.725
Decision Stump	0.719	0.734
Asymmetric Optimal Prediction	0.725	0.750
Log Odds Ratio	0.721	0.708
Auto-Filter	0.779	<b>0.818</b>
Baseline (no filter method)	0.767	

**Supplementary Table 3. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – physiology phenotypes dataset (version 1).**

<b>Phenotypes dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.759	0.722
Chi <sup>2</sup>	0.711	0.729
Decision Stump	0.714	<b>0.761</b>
Asymmetric Optimal Prediction	0.755	0.741
Log Odds Ratio	0.724	0.706
Auto-Filter	0.719	0.728
Baseline (no filter method)	0.741	

**Supplementary Table 4. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – GenAge/GenDR dataset (version 1).**

<b>GenAge/GenDR dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.709	0.727
Chi <sup>2</sup>	0.740	0.751
Decision Stump	<b>0.757</b>	0.742
Asymmetric Optimal Prediction	0.739	0.744
Log Odds Ratio	0.708	0.721
DDFMS	0.702	0.725
Baseline	0.683	

**Supplementary Table 5. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – protein interactors dataset (version 2).**

<b>Interactors dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.695	<b>0.752</b>
Chi <sup>2</sup>	0.661	0.722
Decision Stump	0.689	0.704
Asymmetric Optimal Prediction	0.651	0.737
Log Odds Ratio	0.656	0.662
Auto-Filter	0.656	0.688
Baseline (no filter method)	0.747	

**Supplementary Table 6. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – GO terms dataset (version 2).**

<b>GO terms dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.736	0.754
Chi <sup>2</sup>	0.692	0.73
Decision Stump	0.726	<b>0.772</b>
Asymmetric Optimal Prediction	0.69	0.768
Log Odds Ratio	0.667	0.713
Auto-Filter	0.669	0.708
Baseline (no filter method)	0.765	

**Supplementary Table 7. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – physiology phenotypes dataset (version 2).**

<b>Phenotypes dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.694	<b>0.718</b>
Chi <sup>2</sup>	0.66	0.716
Decision Stump	0.695	0.701
Asymmetric Optimal Prediction	0.654	0.704
Log Odds Ratio	0.644	0.716
Auto-Filter	0.644	0.688
Baseline (no filter method)	0.715	

**Supplementary Table 8. Median AUC values obtained by Random Forest using the single filter vs filter ensemble methods in a pre-processing phase – GenAge/GenDR dataset (version 2).**

<b>GenAge/GenDR dataset</b>	<b>Single filter</b>	<b>Filter ensemble</b>
Information Gain	0.653	0.69
Chi <sup>2</sup>	0.638	0.685
Decision Stump	0.678	0.681
Asymmetric Optimal Prediction	0.64	<b>0.72</b>
Log Odds Ratio	0.617	0.69
Auto-Filter	0.628	0.692
Baseline	0.701	