

Towards an accurate cancer diagnosis modelization: Comparison of Random Forest strategies

Ahmed DEBIT

10/05/2018



From Omics to clinics





From Omics to clinics

Omics

~20k genes (RNA-seq)



Clinics

Dozen of genes (biomarkers)



Signatures of biomarkers

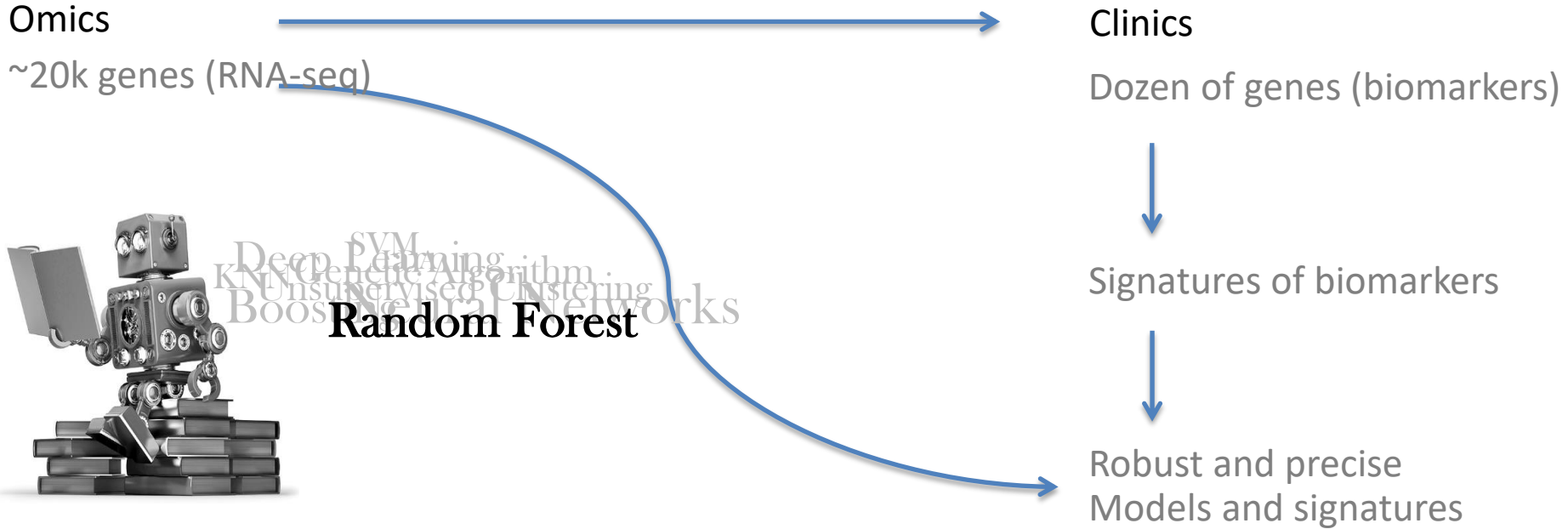


From Omics to clinics



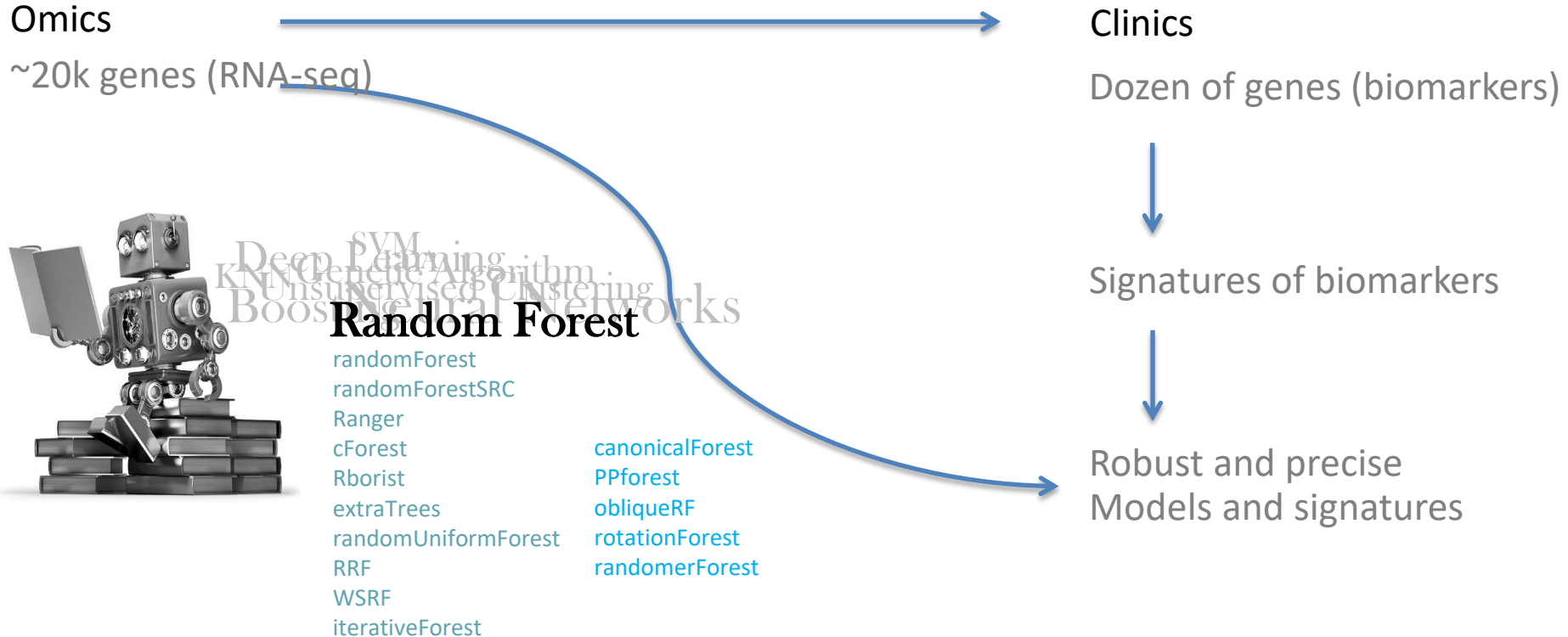


From Omics to clinics



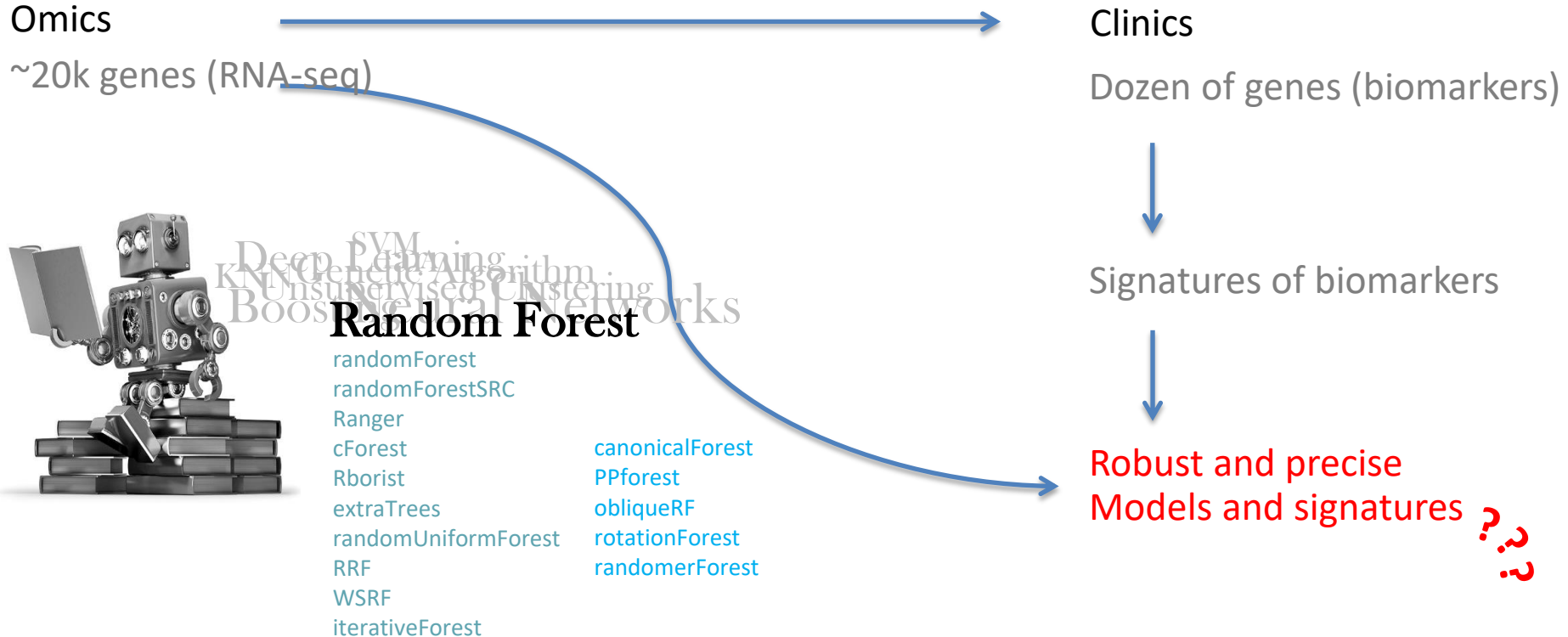


From Omics to clinics





From Omics to clinics





Comparison of RF methods

- Comparison of RF methods under same conditions:
 - **Same** training and the validation sets.
 - **Same** signatures.
 - **Same** computational nodes



Comparison of RF methods

- Comparison of RF methods under same conditions:
 - **Same** training and the validation sets.
 - **Same** signatures.
 - **Same** computational nodes
- For each signature, we'll focus on:
 - **50 resampling** to build the Training and the Validation set.
 - **25 modeling and validations.**
- Analysis of:
 - **Coefficient of Variation of 1,250 models & AUCs**
 - TCGA-BRCA (RPKM): 182 samples x 9560 genes
 - TCGA-LUSC (RPKM): 96 samples x 9262 genes

} 1,250 models & AUCs



Comparison of RF methods

- Comparison of RF methods under same conditions:
 - **Same** training and the validation sets.
 - **Same** signatures.
 - **Same** computational nodes
- For each signature, we'll focus on:
 - **50 resampling** to build the Training and the Validation set.
 - **25 modeling and validations.**
- Analysis of:
 - **Coefficient of Variation of 1,250 models & AUCs**
 - TCGA-BRCA (RPKM): 182 samples x 9560 genes
 - TCGA-LUSC (RPKM): 96 samples x 9262 genes

} 1,250 models & AUCs

Clinics

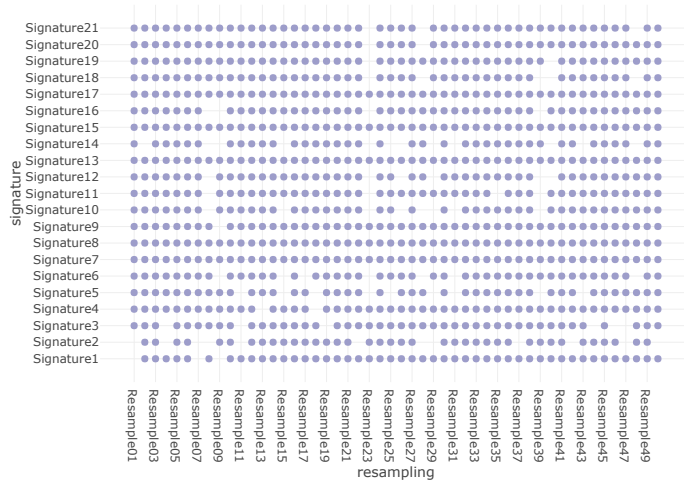


Hyper Stability : CV == 0

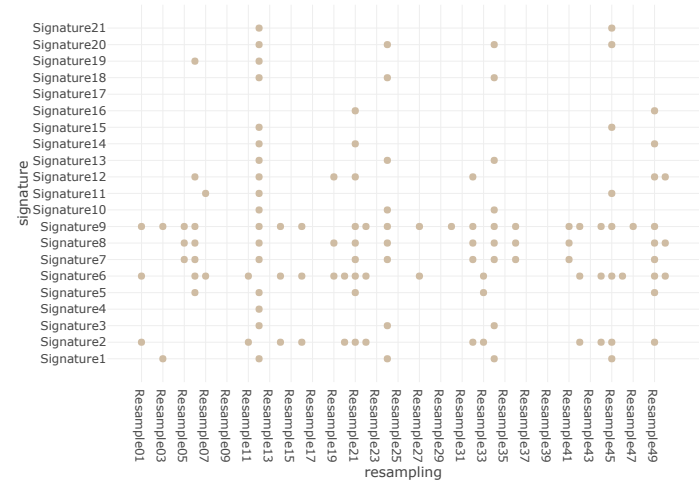


Hyper Stability discriminates RF methods

Best Methods

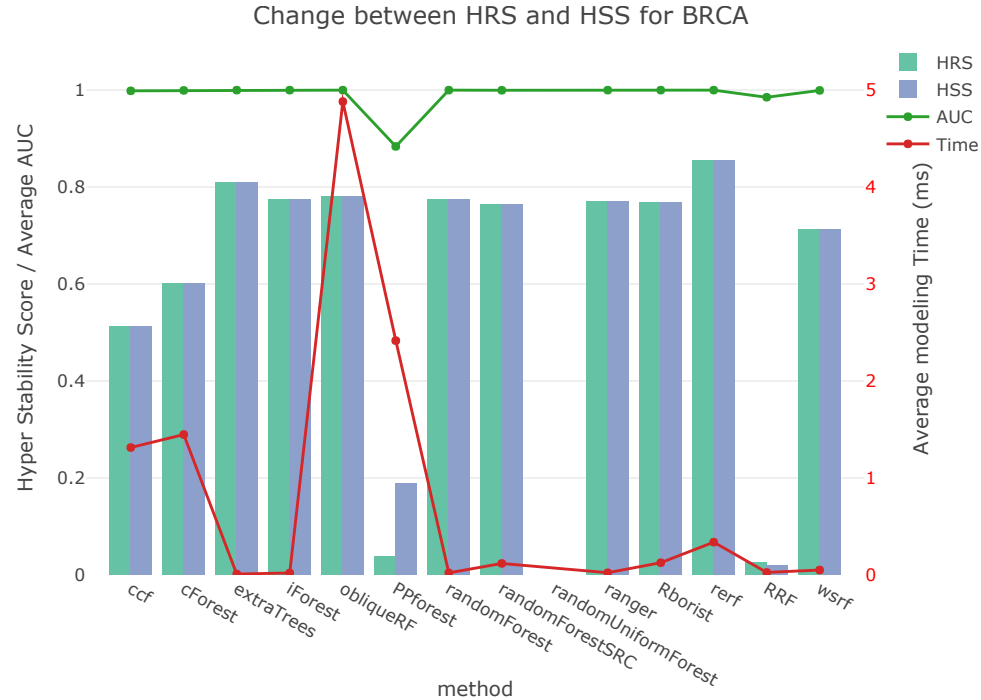


Worst Methods



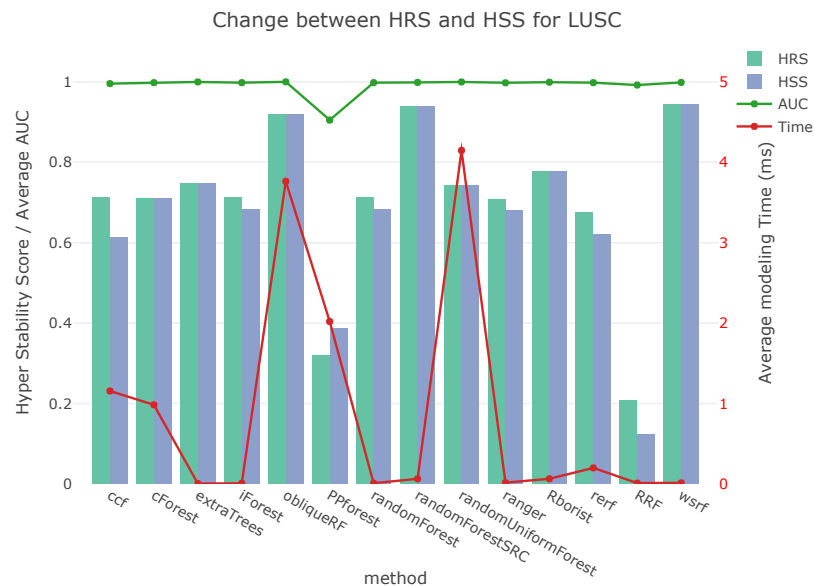
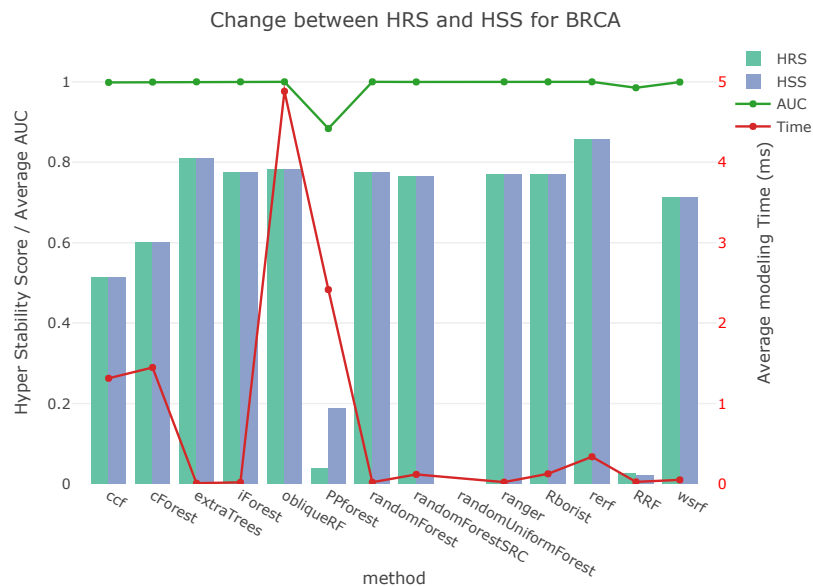


Hyper Stability Score helps finding the best method(s)





Hyper Stability Score is dataset dependent





Conclusions

- The AUC precision is dataset dependent
 - The Methods are dataset dependent.

- Trade-off:
 - AUC precision (hyper-stability)
 - Average AUC value
 - Modeling Time



Classification of classification methods

Towards robust signatures and predictions

