

Tim Hulsen<sup>1\*</sup>, Peter Groenen<sup>2</sup>, Jacob de Vlieg<sup>1,2</sup>, Martijn A. Huynen<sup>1</sup>  
<sup>1</sup>CMBI / NCMLS, Radboud University Nijmegen, The Netherlands

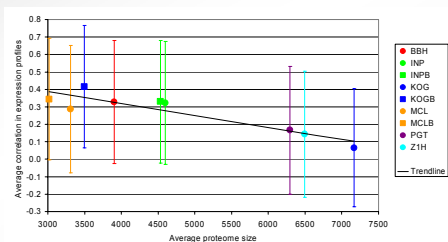
\*T.Hulsen@cmbi.ru.nl  
<sup>2</sup>NV Organon, Oss, The Netherlands

**1** The **transfer of functional annotation** from model organism proteins to human proteins is one of the main applications of comparative genomics. To do this various methods are used to analyze cross-species orthologous relationships according to the operational definition of **orthology**. Often the definition of orthology is incorrectly interpreted as prediction of **cross-species functionally equivalent proteins**, while in fact it is only an indication for **evolutionary related proteins**, which do not necessarily perform the same function. Nevertheless, the quality of the orthology prediction is crucial for the transfer of functional annotation (and other related information). Therefore, ortholog identification methods should be used that return protein pairs with the **highest possible reliability** that allow for the exploration of functional similarity. To measure the similarity in function of proteins from different species we use **functional genomics data** like expression data and protein interaction data. We tested several of the most popular **ortholog identification methods** and performed an ortholog analysis on the proteins involved in the **trypsin inhibition pathway**.

**2** Studied ortholog identification methods  
**BBH**: Best Bidirectional Hit    **INP**: InParanoid  
**KOG**: KOG                            **MCL**: OrthoMCL  
**PGT**: PhyloGenetic Tree        **Z1H**: Z-value > 100

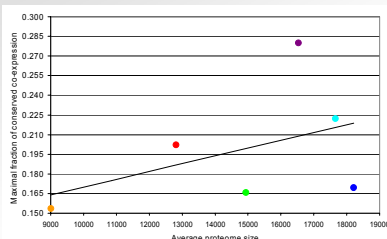
**3** Direct conservation of functional parameters: **1-1**

Correlation of expression profiles



Other benchmarks:  
 equal **SwissProt** name, equal **InterPro** acc.nr.

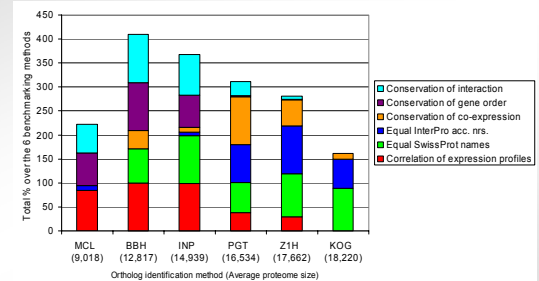
**4** Pairwise conservation of functional parameters: **2-2**



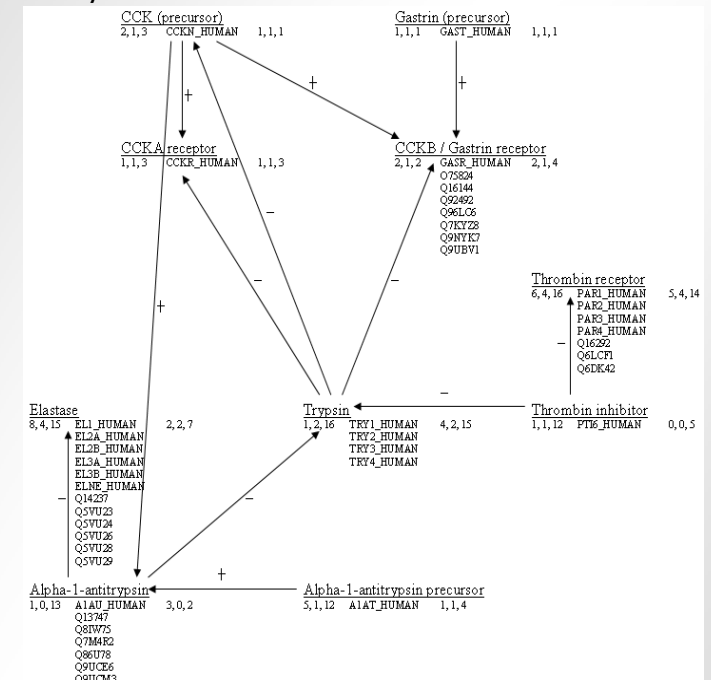
Conservation of co-expression

Other benchmarks:  
 conservation of **gene order**, cons. of **interaction**

**5** Overall benchmarking scores



**6** As a practical example in which orthologies can be used, we studied the **trypsin inhibition pathway**, which is known to differ between human, mouse and rat (**high CCK levels cause pancreatic cancer** in rat and mouse, but not in human). For each **human** protein in the pathway, the number of orthologous proteins in **mouse** and **rat** were defined, in three ways: by **SwissProt** annotation, by the **BBH** method, and by the **PGT** method.



Although the differences in **CCK levels** between human, mouse and rat can not be easily explained by these orthologies, the **BBH** method seems to be the better method for solving a problem like this.

**7** Conclusion: **Best Bidirectional Hit** and **InParanoid** methods score quite good in our benchmarking. Methods predicting **few orthologous relationships**, like BBH, are especially useful when applying them to complete **pathways**.

(Submitted to Bioinformatics)