



**HAL**  
open science

## Gene orthology detection for long noncoding RNA (lncRNA)

Fabien Degalez, Coralie Alain, Laetitia Lagoutte, Sandrine Lagarrigue

► **To cite this version:**

Fabien Degalez, Coralie Alain, Laetitia Lagoutte, Sandrine Lagarrigue. Gene orthology detection for long noncoding RNA (lncRNA). 39. Congress of the international society for animal genetics (ISAG), Jul 2023, Cape Town, South Africa. , pp.61, 2023, Abstracts. hal-04217585

**HAL Id: hal-04217585**

**<https://hal.inrae.fr/hal-04217585>**

Submitted on 15 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# GENE ORTHOLOGY DETECTION FOR LONG NON CODING RNA (lncRNA)

F. Degalez, L. Lagoutte, F. Lecerf, C. Allain, S. Lagarrigue

fabien.degalez@inrae.fr

PEGASE, INRAE, Institut Agro, 35590 Saint Gilles, France

## CONTEXT

→ Long non-coding RNAs (lncRNAs), with at least 200 nt & low potential coding capabilities are a major component of regulatory elements in genomes. More than 15,000 loci have been modeled in human.

→ However, the role of most of them remains to be clarified.

- Exploring lncRNA conservation between species is an approach to strengthen the annotation of lncRNAs by inferring function in one species from another one more studied such as human or mouse, as has been done previously for protein coding genes (PCGs).

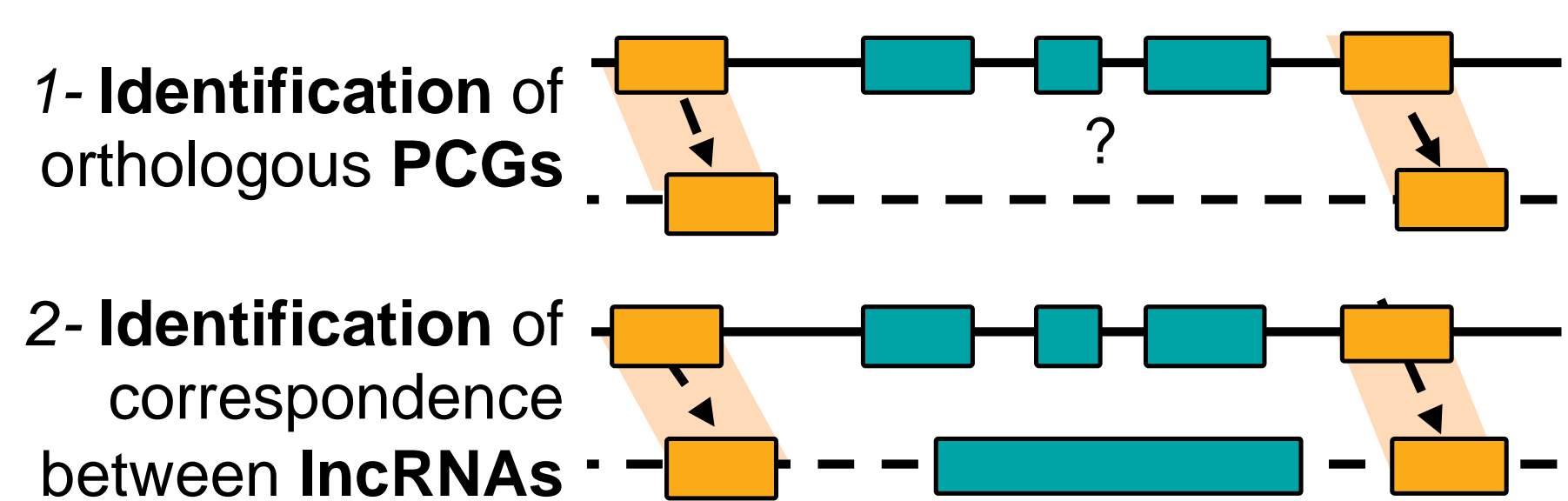
- However, unlike PCGs, lncRNA sequences are not well conserved across species [1]. Therefore, no lncRNA orthologs are reported in reference databases such as Ensembl BioMart, regardless of species.

- In this context, we have developed a workflow combining 3 approaches (Method 1, 2 and 3) that can be used for any species of interest and have applied it on 11 species covering a large phylogenetic scale from mammals to chicken.

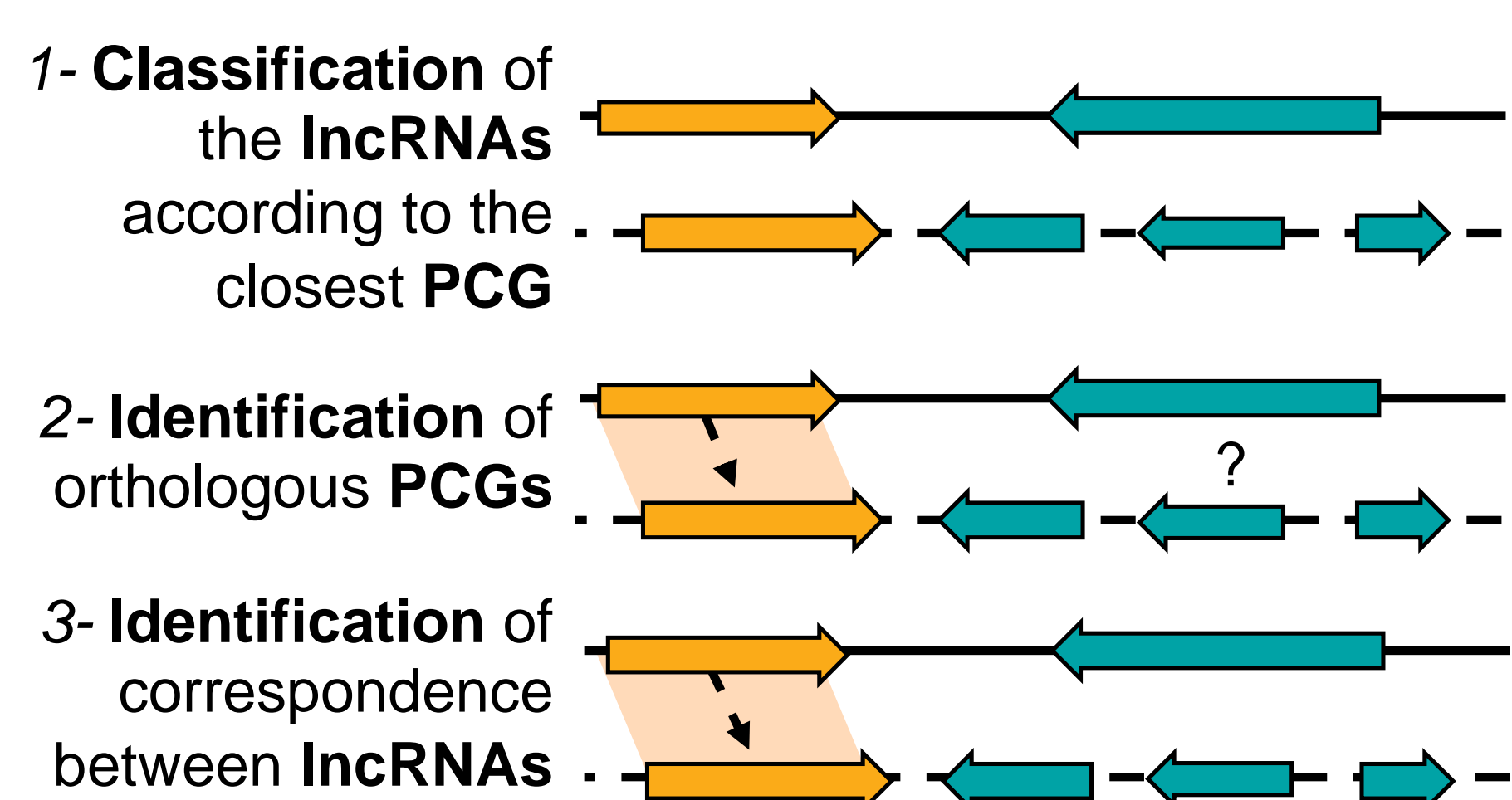
## METHOD

Gitlab – f.degalez/lncrna\_orthologfinder

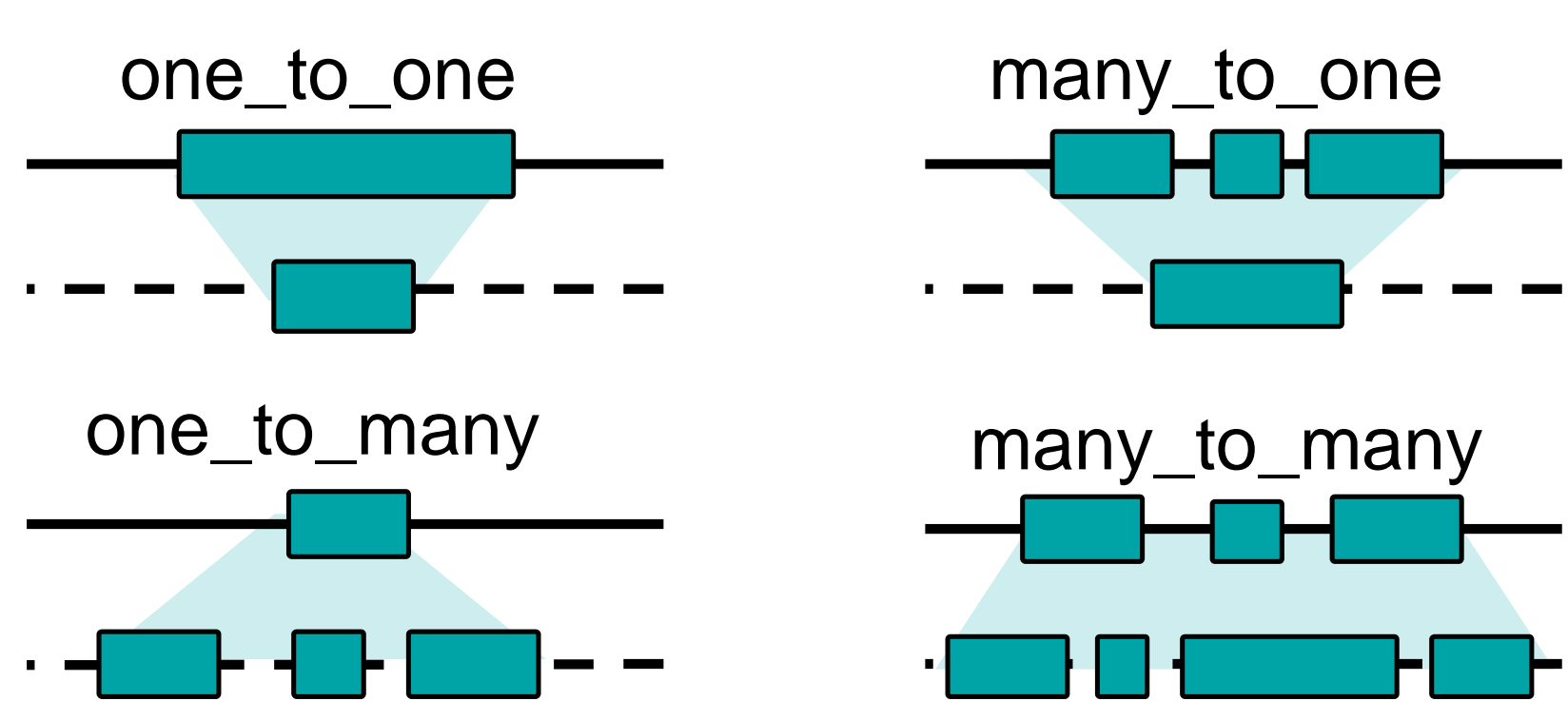
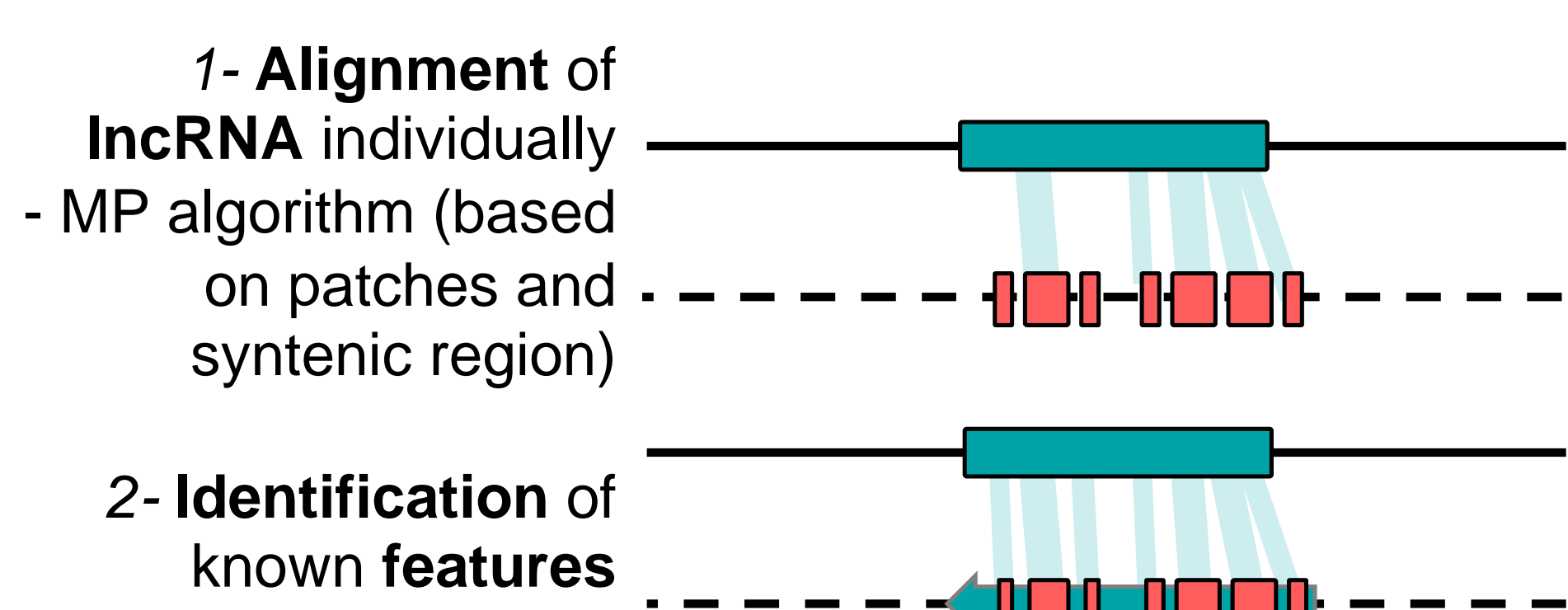
### Method 1 – Synteny 2 PCGs



### Method 2 – Synteny 1 PCG & FEELnc config.



### Method 3 – Alignment



## RESULTS - NUMBERS

1 Table 1: Orthologous lncRNAs between HSA & 10 species (mammals, birds & fish).

\* The reference Ensembl chicken gene atlas is enriched in lncRNA loci [2]

Total	PCG	lncRNA	Target	Source : Human							
				Total : 61552 - PCG : 19994 - lncRNA : 17734							
				method 1		method 2		method 3			
				1to1	many	Tot	Match	U			
55414	21884	9949	Mouse	457	876	1333	1214	3190	4404	8660	10867
31121	20500	6479	Dog	112	230	342	665	1543	2208	-	2372
30371	20912	7241	Horse	144	515	659	746	1775	2521	7684	9216
27607	21861	1480	Cow	65	99	164	260	480	740	8947	9342
27271	21343	2705	Goat	73	171	244	319	821	1140	8824	9378
31908	21280	6790	Pig	150	343	493	735	1785	2520	8885	10205
45289	17859	24835	* Chicken	244	858	1102	1368	4955	6323	4816	9142
17970	16226	1034	Turkey	18	37	55	137	275	412	-	437
22150	16619	4757	Zebrafish	60	213	273	548	1221	1769	4835	6099
32520	25432	2222	Zebrafish	14	23	37	156	355	511	-	529

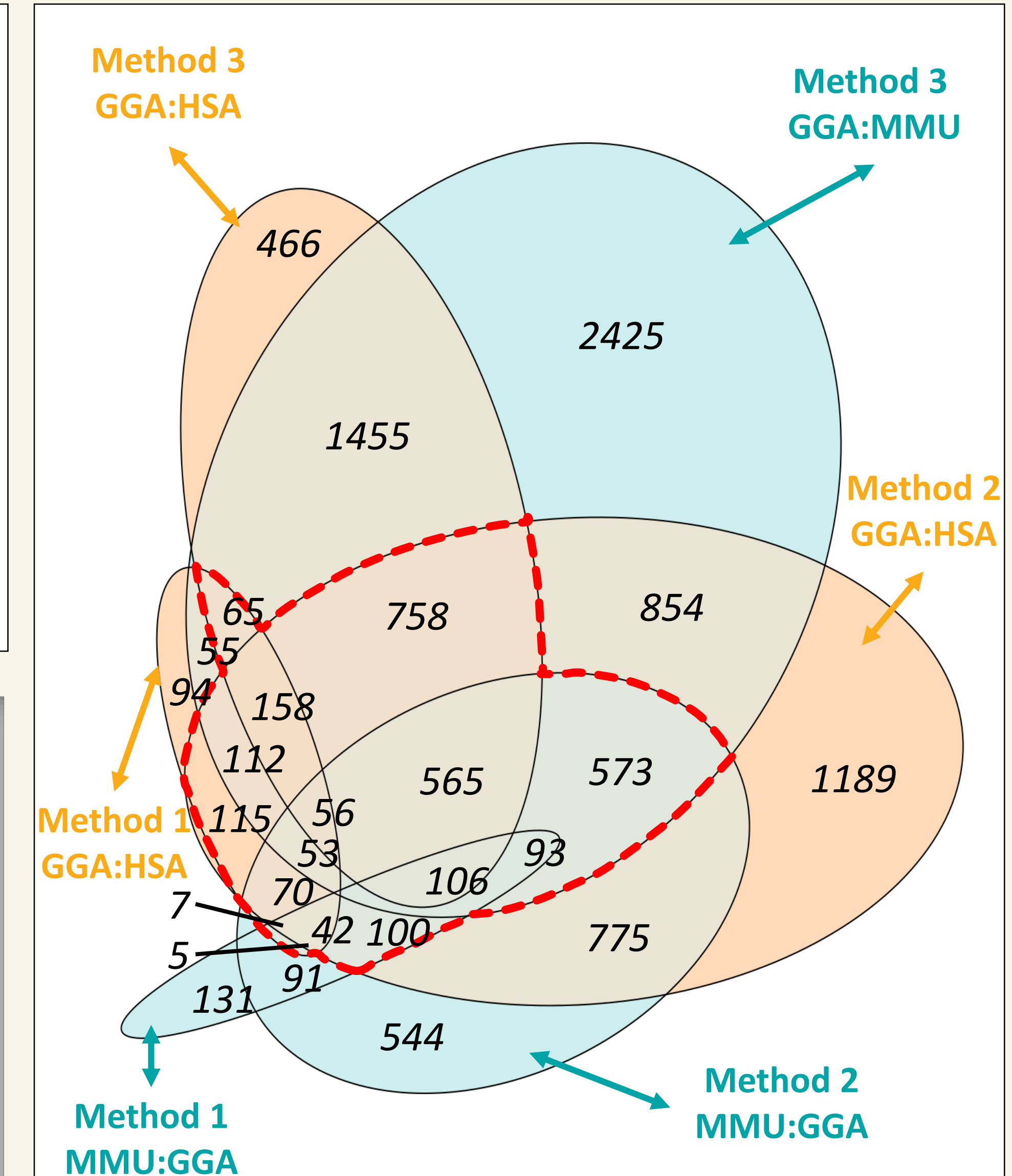
2 Fig. 1: One of the 23 lncRNAs in HSA identified as orthologous in 9 species by at least one method.

Orthologous transcripts are either antisense (1) or divergent (2) with respect to the closest PCG, suggesting partial knowledge of TSS position in some species. lncRNA capture combined with long-read sequencing is in progress.

HSA	INTS6	1	INTS6-AS1	Ref.
MMU	INTS6	-	14-62997428:63061164	Meth3
CFA	INTS6	2	ENSFCAG845021720	Meth2
ECA	ENSECAG32008	2	INTS6	Meth2&3
BTA	INTS6	-	12-20902200:20992516	Meth3
CHI	12-65613622:65700765	-	INTS6	Meth3
SSC	11-16322402:16376747	-	INTS6	Meth3
GGA	INTS6	2	INRA2889-2289	Meth2&3
GUT	INTS6	-	1-60247507:60264009	Meth3

3 Fig. 2: Subset of 3609 orthologous HSA:GGA lncRNAs considered more reliable by combining several methods.

Out of the 9142 HSA:GGA orthologous lncRNAs detected by at least 1 method (Table 1), 3609 (39%) lncRNAs were selected (red section) as more reliable because found orthologous by at least 2 methods (2765 cases) or by the method 1 or method 2 secured by 2 methods for GGA:MMU (1873 cases)



The results presented here consider annotations and/or genome assemblies that may have been updated, such as the enriched annotation for the chicken genome (see poster P63). New analyses are currently in progress to include the major modifications.

## RESULTS – FUNCTIONAL ANNOTATION

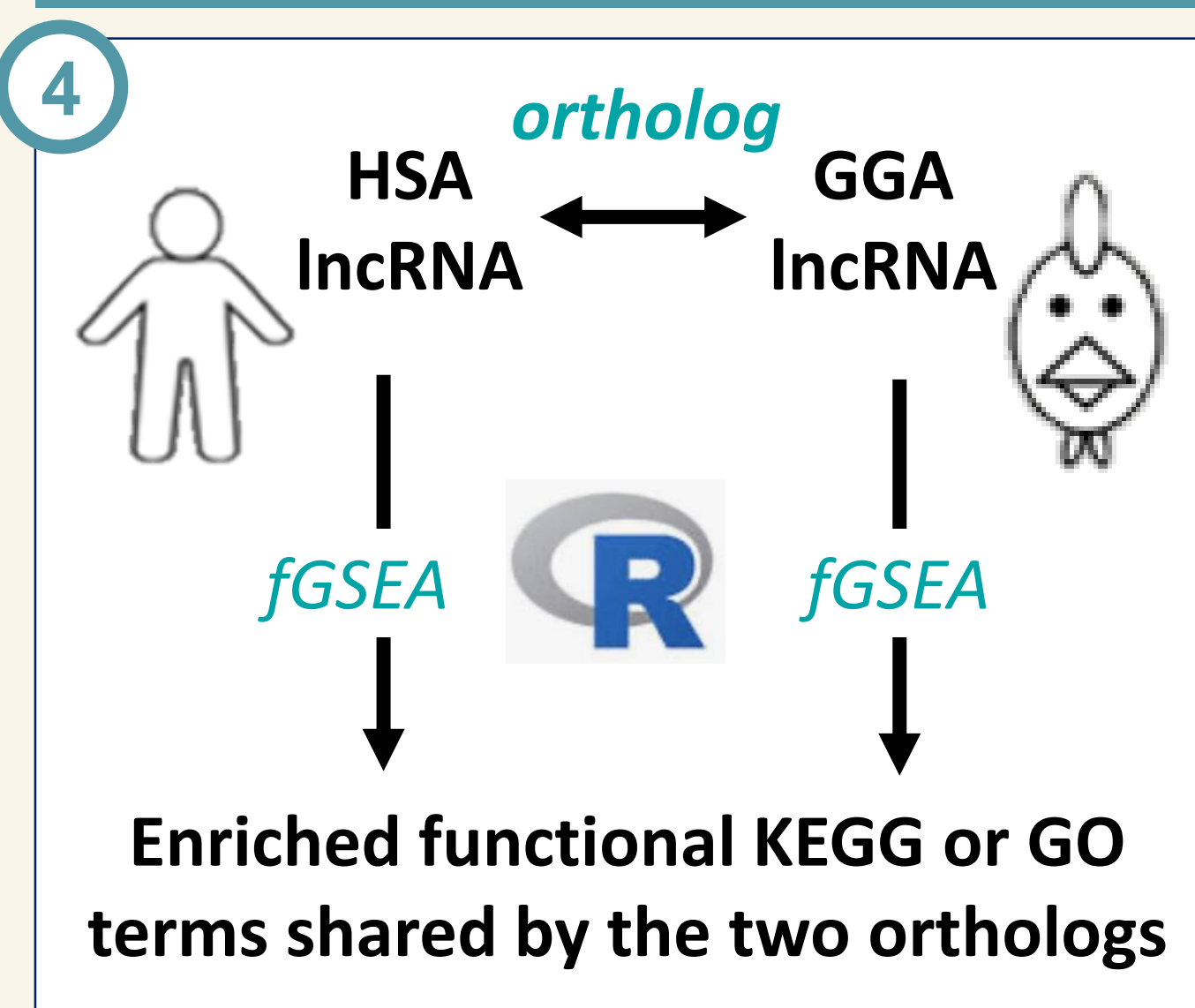


Fig. 4: Several tissue-specific orthologous lncRNAs had co-expressed genes associated to the same enriched terms related to the tissue function.

Here, the case of liver-specific lncRNA orthologs in GGA and HSA associated to "Complement and coagulation cascades" (10 / 82); FDR: 1.29 e-14.

