GENERAL INFORMATIONS ON THE FEATURE –
e! v101 GRCg6a LNC-enriched –
(45,615 genes: 17,921 PCG & 25,082 LNC)

chr: chromosome

start: genomic coordinate of the start of the feature (gene/transcript)

end: genomic coordinate of the end of the feature

strand: strand of the feature (+ or -)

gnId: unique identifier of the gene corresponding to the feature

tpId: unique identifier of the transcript for the transcripts level annotation,
semi-column-separated unique identifiers of all the transcripts corresponding to the gene for the gene level annotation

source: source of the feature

version: version of the database used

	v1	v100	v103	v2018	v5	v94	Total
ALDB	2876	0	0	0	0	0	2876
ensembl	0	24319	0	0	0	2680	26999
INRA	0	0	0	10708	0	0	10708
NCBI	0	0	1293	0	0	0	1293
NONCODE	0	0	0	0	3702	0	3702
RefSeq	0	37	0	0	0	0	37
Total	2876	24356	1293	10708	3702	2680	45615

Figure 1. Number of models from each source with the version.

gnBiotype: biotype of the gene in a standardized way
lnc = long non-coding genes / antisense / lincRNA / sense_intronic
mir = miRNA
mis = misc_RNA. For INRA source, the “mis” genes correspond to the Transcript of Unknown Coding Potential (TUCP) from FEELnc’ “codpot” module.
mtr = Mt_rRNA
mtt = Mt_tRNA
pcg = protein coding gene
pse = pseudogene
rbz = ribozyme
rrn = rRNA
sca = scaRNA
sno = snoRNA
snr = snRNA
srn = sRNA

igg = IG_gene
pps = processed_pseudogene

	ALDB	ensembl	INRA	NCBI	NONCODE	RefSeq	Total
igg	0	99	0	0	0	0	99
lnc	2876	7223	10080	1201	3702	0	25082
mir	0	1312	0	7	0	0	1319
mis	0	135	348	3	0	0	486
mtr	0	0	0	0	0	2	2
mtt	0	0	0	0	0	22	22
pcg	0	17553	280	75	0	13	17921
pps	0	31	0	0	0	0	31
pse	0	283	0	2	0	0	285
rbz	0	2	0	0	0	0	2
rrn	0	54	0	1	0	0	55
sca	0	16	0	0	0	0	16
sno	0	218	0	0	0	0	218
snr	0	72	0	4	0	0	76
srn	0	1	0	0	0	0	1
Total	2876	26999	10708	1293	3702	37	45615

Figure 2. Number of each genes biotypes from each source.

tpBiotype: biotype of the transcript corresponding to the feature for the transcripts level annotation, semi-column-separated biotype of all the transcripts corresponding to the gene for the gene level annotation

gnName: name of the gene corresponding to the feature, after the following decision rule:
= hgncHsa1to1 > Chicken HGNC (hgncGga) > Mouse 1to1 HGNC (hgncMmu1to1) > Human 1to1 wikigene (wikigeneHsa1to1) > wikigeneGga > wikigeneMmu1to1 > Gene name as found in the field “gene_name” of the Ensembl GalGal5 GTF, v94 (gnNameGga) > Human 1tomany HGNC (hgncHsaXtoMany) > Mouse 1tomany HGNC (hgncMmuXtoMany) > Human 1tomany wikigene (wikigeneHsaXtoMany) > Mouse 1tomany wikigene (wikigeneMmuXtoMany) > gene ID (as in **gnId** column).
Note: all the 1-to-many and many-to-Many names are followed by “_XtoMany”

Table 1: number of genes with HGNC identifier, LOC, 1-to-Many or Many-to-Many and gene ID in the **gnName** column.

Type of name	**Number of genes**
HGNC identifier	15090
LOCxxx name	1097
containing "_XtoMany "	807
Gene ID (ENSGALG, INRAGALG, etc.)	28621

gnNameDesc: full name of the gene associated to the feature, originating from the same database as the **gnName** column.

gnNameGga: name associated to the gene in Ensembl Galgal6 (from BioMart)

hgncGga: HGNC symbol associated to the gene (from BioMart)

wikigeneGga: wikigene symbol associated to the gene (from BioMart)

hgncHsa1to1: HGNC symbol of the 1-to-1 human's orthologue of the gene (from BioMart)

wikigeneHsa1to1: wikigene symbol of the 1-to-1 human's orthologue of the gene (from BioMart)

hgncMmu1to1: HGNC symbol of the 1-to-1 mouse's orthologue of the gene (from BioMart)

wikigeneMmu1to1: wikigene symbol of the 1-to-1 mouse's orthologue of the gene (from BioMart)

gnNameDescGga: full name associated to the gene in Ensembl Galgal5 (from BioMart)

gnNameDescHsa1to1: full name of the 1-to-1 human's orthologue of the gene (from BioMart)

gnNameDescMmu1to1: full name of the 1-to-1 mouse's orthologue of the gene (from BioMart)

gnNameDescHsaXtoMany: full name of the X-to-many human's orthologue of the gene (from BioMart)

gnNameDescMmuXtoMany: full name of the X-to-many mouse's orthologue of the gene (from BioMart)

orthHsaType: type of orthology between the chicken genes and the human genes (from BioMart)

	igg	lnc	mir	mis	mtr	mtt	pcg	pps	pse	rbz	rrn	sca	sno	snr	srn	Total
many2many	18	0	2	3	0	0	256	0	0	0	0	0	14	6	0	299
one2many	0	0	12	13	0	0	1663	0	0	0	1	3	85	16	0	1793
one2one	1	0	124	28	0	0	12045	0	0	2	0	10	70	19	0	12299
Total	19	0	138	44	0	0	13964	0	0	2	1	13	169	41	0	14391

Figure 3: Number of genes with a 1-to-1, 1-to-many and many-to-many ortholog in human, as a function of the gene simple biotype (**gnBiotype**)

orthMmuType: type of orthology between the chicken genes and the mouse genes (from BioMart)

	igg	lnc	mir	mis	mtr	mtt	pcg	pps	pse	rbz	rrn	sca	sno	snr	srn	Total
many2many	0	0	4	0	0	0	275	0	0	0	0	0	15	6	0	300
one2many	1	0	11	11	0	0	1508	0	0	1	4	6	76	10	0	1628
one2one	0	0	103	32	0	0	12186	0	0	1	5	9	79	20	1	12436
Total	1	0	118	43	0	0	13969	0	0	2	9	15	170	36	1	14364

Figure 4: Number of genes with a 1-to-1, 1-to-many and many-to-many ortholog in mouse, as a function of the gene simple biotype (**gnBiotype**)

orthHsaGnId: Identifier of the human gene ID associated to the chicken gene (from BioMart)
Note: all the 1tomany and manyToMany ID are followed by "_XtoMany"

OrthMmuGnId: Identifier of the mouse gene ID associated to the chicken gene (from BioMart)
Note: all the 1tomany and manyToMany ID are followed by "_XtoMany"

hgncHsaXtoMany: HGNC symbol of the X-to-many human's orthologue of the gene (from BioMart)

hgncMmuXtoMany: HGNC symbol of the X-to-many mouses's orthologue of the gene (from BioMart)

wikigeneHsaXtoMany: Wikigene symbol of the X-to-many human's orthologue of the gene (from BioMart)

wikigeneMmuXtoMany: Wikigene symbol of the X-to-many mouse's orthologue of the gene (from BioMart)

exNb: number of exon(s) of the of the transcript(s) associated to the gene, in the format "**minimum** number of exon" (= number of exons of the transcript with the least exons) ; "**median** of the number of exon of the transcript(s)" ; "**maximum** number of exons" (= number of exons of the transcript with the most exons), semi-column-separated

exSz: Size of the exon(s) of the of the transcript(s) associated to the gene, in the format "**minimum** length of exon" (= smallest exon) ; "**median** size of exon of the transcript(s)" ; "**maximum** size of exon" (= longest exon), semi-column-separated

inSz: Size of the intron(s) of the of the transcript(s) associated to the gene, in the format "**minimum** length of intron" (= smallest intron); "**median** size of intron of the transcript(s)" ; "**maximum** size of intron" (= longest intron), semi-column-separated

tpSz: Size of the transcript(s) associated to the gene. Calculated as the sum of the length of all the exons associated to each transcript. Format: "**minimum** length of transcript" (= smallest transcript) ; "**median** size of transcript(s)" ; "**maximum** size of transcript" (= longest transcript), semi-column-separated

nbTp: number of transcript(s) associated to the gene.

gnSz: size of the gene. Calculated as the median of the size of all the transcripts associated to the gene.

LONG NON-CODING GENES ANNOTATION WITH RESPECT TO THE CLOSEST PROTEIN-CODING GENE

The column following concerns the FEELnc class annotation of the LNC (feelLnc) with respect to the nearest protein coding gene (feelLncPcg).

- feelLncPcgClassName, feelLncPcgClassType, feelLncPcgGnId, feelLncPcgGnName, feelLncPcgGnDist,

feelLncPcgClassName: Abbreviation of the FEELnc classification of the LNC with respect to the closest PCG

To transfer the FEELnc information from the transcript level to the gene level, an order of importance has to be decided.

The class names are composed of three parts:

- the first part (8 letters) is composed of the main class type. For the genic classes: lncgSSex, lncgSSin, lncgASex, lncgASin. For the intergenic classes: lincDivg, lincSSup, lincSSdw, lincConv.
- the second part (4 letters) concerns only the genic classes without subtype conflicts (see below), we add one of the three subtypes: Nested (Nest), Overlapping (Ovlp) or Containing (Cont)
- the third part (_n.n.n or _n.1.n) indicates that there are conflicts between annotation due to several PCGs related to the LNC locus.

Conflicts cases are of two types: the cases in which there are more than 1 annotation relative to one unique PCG (as indicated by the "n" in the middle and "1" at the end of "n.n.1"), and the case in which there more than 1 or more annotation relative to more than 1 PCG ("n.X.n" in the **feelLncPcgClassType** column). In these cases, we prioritized the annotation in the column « feelLncPcgClassName», which gives only 1 class per gene.
→ n.n.1 case: Genics have priority over intergenics (lncg > linc).
Among the genic, exonics have priority over intronics.
Among exonics and intronics, the subtypes nested / containing / overlapping have the same importance. They are kept if they do not produce conflicts and are removed if there are 2 or more subtypes.

→ n.X.n case: Same order of priority as previously (n.n.1 case).
Concerning the intergenics, there can be annotation conflicts between the several PCGs: the LNC can be classified as lincDivg with one PCG and lincConv with another one PCG (see figure 1 for an example). We prioritize the classes as following: Divg > SS > Conv. Same-strand have priority over Conv because it could suggest an error in the modelization: the LNC could be a 5'-part or 3'-part of the PCG. Between the same strand up and down (lincSSup and lincSSdw), we choose the closest. The third part of the class name of these genes is either "_n.n.n" or "_n.1.n".

lincConv	lincDivg	lincSSdw	lincSSup	lncgASex
3092	4760	5129	4148	2810
lncgASin	lncgSSex	lncgSSin	unclassified	
2788	2	68	2284	

Figure 5: Simplified class name present in the file for the LNC:PCG pairs

feelLncPcgClassType: gives information on three fields separated by a dot "." ($X_1.X_2.X_3$) about the classification done by FEELnc of the LNC transcript relatively to the closest PCG transcript (= LNC:PCG pair):

- X_1: number of transcripts of the LNC gene: "1" if 1 transcript, "n" if more than one transcript,
- X_2: number of feelnc class(es) associated to the LNC:PCG pair: "1" if 1 class, "n" if more than one class (the "unclassified" class does *not* count),
- X_3: number of PCG gene(s) concerned by this (these) annotation(s): "1" if 1 PCG gene, "n" if more than one PCG gene.

1.1.1: the LNC has 1 transcript, with 1 annotation associated to 1 PCG.
n.1.1: the LNC has several transcripts, all with the same annotation associated to the same PCG.
n.n.1: the LNC has several transcripts, with different annotations associated to the same PCG.
n.n.n: the LNC has several transcripts, with different annotations associated to different PCG.
n.1.n: the LNC has several transcripts, all with the same annotation but associated to different PCG.
unclassified: the LNC is either alone in a contig (beginning by AADN. or KQ), or no interactions were found within the 100 000 pb sliding window used by FEELnc.

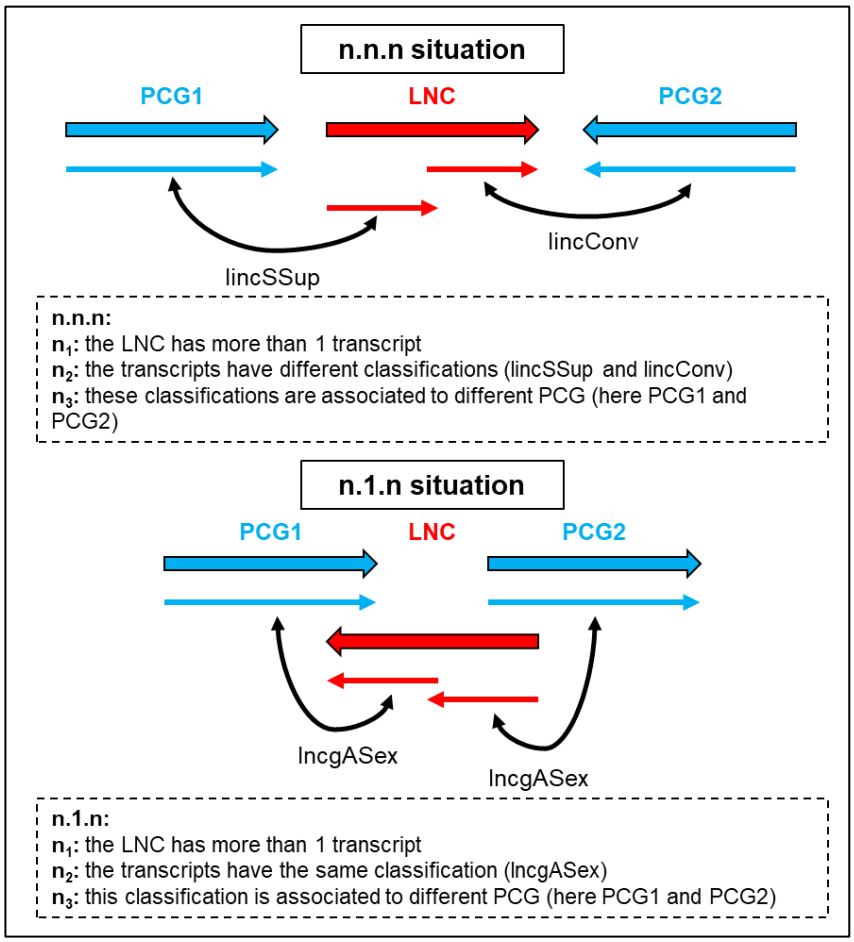

Figure 6. examples of configurations corresponding to a “n.n.n” type (top) or a “n.1.n” type (bottom)

	1.1.1	n.1.1	n.1.n	n.n.1	n.n.n	unclassified	Total
ALDB	2115	411	0	13	21	316	2876
ensembl	3960	1899	2	90	169	1103	7223
INRA	7835	1483	5	238	177	342	10080
NCBI	975	127	1	13	4	81	1201
NONCODE	2786	386	4	58	25	443	3702
Total	17671	4306	12	412	396	2285	25082

Figure 7. Repartition of the different classification types across databases for the LNC:PCG classification

feelLncPcgGnId: Unique identifier of the protein-coding gene relatively to which a LNC gene is classified by FEELnc

feelLncPcgGnName: Name of the protein-coding gene relatively to which a LNC gene is classified by FEELnc

feelLncPcgGnDist: Distance (in bp), as calculated by FEELnc, between the protein-coding gene relatively to which a LNC gene is classified by FEELnc and the LNC gene.

PROTEIN-CODING GENES ANNOTATION WITH RESPECT TO THE CLOSEST PROTEIN-CODING GENE

feelPcgPcgClassName: Abbreviation of the FEELnc classification of the PCG with respect to the closest PCG. Abbreviations are similar to those of LNC:PCG classification.

feelPcgPcgClassType: gives information on three fields separated by a dot "." (X1.X2.X3) about the classification done by FEELnc of the PCG transcript relatively to the closest PCG transcript, similarly to the LNC:PCG classification.

	1.1.1	n.1.1	n.1.n	n.n.1	n.n.n	unclassified	Total
ensembl	8970	3521	39	748	675	3600	17553
INRA	111	54	0	9	8	98	280
NCBI	57	3	0	0	1	14	75
RefSeq	13	0	0	0	0	0	13
Total	9151	3578	39	757	684	3712	17921

Figure 8. Repartition of the different classification types across databases for the PCG:PCG classification.

feelPcgPcgGnId: unique identifier of the coding gene relatively to which another protein-coding gene is classified by FEELnc.

feelPcgPcgGnName: name of the protein-coding gene relatively to which a PCG gene is classified by FEELnc

feelPcgPcgGnDist: distance (in bp) between the coding gene relatively to which a protein-coding gene is classified by FEELnc and the PCG

MICRO RNA AND SMALL RNA GENES ANNOTATION WITH RESPECT TO THE CLOSEST LONG NON-CODING GENE

feelMirLncClassName: Abbreviation of the FEELnc classification of the miRNA with respect to the closest LNC. Abbreviations are similar to those of LNC:PCG classification. From the point of view of procedure, we began by classifying the miRNA with respect to the LNC, obtaining one classification per miRNA (i.e., one LNC associated to each miRNA). We then reversed this classification in order to associate to each LNC every miRNA that was associated to the LNC.

feelMirLncGnId: Unique identifier of the miRNA relatively to which a LNC is classified by FEELnc

feelMirLncGnName: Name of the miRNA relatively to which a LNC gene is classified by FEELnc.

feelMirLncGnDist: Distance (in bp), as calculated by FEELnc, between the miRNA gene relatively to which a LNC gene is classified by FEELnc and the LNC gene

feelSmlLncClassName: Abbreviation of the FEELnc classification of the small RNA with respect to the closest LNC. Abbreviations are similar to those of LNC:PCG classification. From the point of view of procedure, we began by classifying the small RNA with respect to the LNC, obtaining one classification per small RNA (i.e., one LNC associated to each small RNA). We

then reversed this classification in order to associate to each LNC every small RNA that was associated to the LNC

feelSmlLncGnId: Unique identifier of the small RNA relatively to which a LNC is classified by FEELnc

feelSmlLncGnName: Name of the small RNA relatively to which a LNC gene is classified by FEELnc.

feelSmlLncGnDist: Distance (in bp), as calculated by FEELnc, between the small RNA gene relatively to which a LNC gene is classified by FEELnc and the LNC gene

DETAILS OF THE BY TRANSCRIPTS ANNOTATION OF THE LNC, PCG, MIRNA AND SMALL RNA WITH RESPECT TO THE CLOSEST RELEVANT GENE

In the following columns, there are as many fields (Class Name, gene ID, distance, etc.) as transcripts, and the fields are separated by semi-columns ";".

feelLncPcgClassNameByTp: abbreviation of the FEELnc classification for each transcript of the LNC with respect to the closest PCG. Abbreviations are similar to those of LNC:PCG classification at the gene level.

feelLncPcgGnIdByTp: unique identifier of the protein-coding gene relatively to which each transcript of the LNC is classified by FEELnc

feelLncPcgGnNameByTp: name of the protein-coding gene relatively to which each transcript of the LNC is classified by FEELnc

feelLncPcgGnDistByTp: distance (in bp) between each transcript of the LNC and the protein-coding gene relatively to which each transcript of the LNC is classified by FEELnc

feelPcgPcgClassNameByTp: abbreviation of the FEELnc classification for each transcript of the PCG with respect to the closest PCG. Abbreviations are similar to those of LNC:PCG classification at the gene level.

feelPcgPcgGnIdByTp: name of the protein-coding gene relatively to which each transcript of the PCG is classified by FEELnc

feelPcgPcgGnNameByTp: name of the protein-coding gene relatively to which each transcript of the PCG is classified by FEELnc

feelPcgPcgGnDistByTp: distance (in bp) between each transcript of the PCG and the protein-coding gene relatively to which each transcript of the LNC is classified by FEELnc

feelMirLncClassNameByTp: abbreviation of the FEELnc classification for each transcript of the miRNA with respect to the closest PCG. Abbreviations are similar to those of LNC:PCG classification at the gene level.

feelMirLncGnIdByTp: name of the protein-coding gene relatively to which each transcript of the miRNA is classified by FEELnc

feelMirLncGnNameByTp: name of the protein-coding gene relatively to which each transcript of the miRNA is classified by FEELnc

feelMirLncGnDistByTp: distance (in bp) between each transcript of the miRNA and the protein-coding gene relatively to which each transcript of the LNC is classified by FEELnc

feelSmlLncClassNameByTp: abbreviation of the FEELnc classification for each transcript of the small RNA with respect to the closest PCG. Abbreviations are similar to those of LNC:PCG classification at the gene level.

feelSmlLncGnIdByTp: name of the protein-coding gene relatively to which each transcript of the small RNA is classified by FEELnc

feelSmlLncGnNameByTp: name of the protein-coding gene relatively to which each transcript of the small RNA is classified by FEELnc

feelSmlLncGnDistByTp: distance (in bp) between each transcript of the small RNA and the protein-coding gene relatively to which each transcript of the LNC is classified by FEELnc

GO AND MGI TERMS

Go_gga_BP: Gene Ontology (GO) terms associated to biological process (BP) in hens

Go_gga_MF: Gene Ontology (GO) terms associated to molecular functions (MF) in hens

Go_gga_CC: Gene Ontology (GO) terms associated to cellular component (CC) in hens

Go_Hsa_BP: Gene Ontology (GO) terms associated to biological process (BP) in humans

Go_Hsa_MF: Gene Ontology (GO) terms associated to molecular functions (MF) in humans

Go_Hsa_CC: Gene Ontology (GO) terms associated to cellular component (CC) in humans

Go_Mmu_BP: Gene Ontology (GO) terms associated to biological process (BP) in mice

Go_Mmu_MF: Gene Ontology (GO) terms associated to molecular functions (MF) in mice

Go_Mmu_CC: Gene Ontology (GO) terms associated to cellular component (CC) in mice

MGIinputType: type of input from MGI

MGI_MPid: ID of the mammalian phenotype ontology from MGI

MGIterms: Terms associated to the mammalian phenotype ontology from MGI

EXPRESSION INFORMATION

top1TissuRosl: name of the most expressed tissue among 21 tissues

top1ExprRosl: expression in TPM of the most expressed tissue among 21 tissues

allTissuRosl: names of the 20 remaining tissues ranked in descending order of expression

allExprRosl: expression of the 20 remaining tissues ranked in descending order

top1TissuRprm: name of the most expressed tissue among 5 tissues

top1ExprRprm: expression in TPM of the most expressed tissue among 5 tissues

allTissuRprm: names of the 4 remaining tissues ranked in descending order of expression

allExprRprm: expression of the 4 remaining tissues ranked in descending