INRAO

Méthodes d'analyse intégrative spatiotemporelle de données omiques. Transcriptome et développement embryonnaire bovin.

V Duranthon (BREED, PHASE) & D Laloë (GABI, GA)

> Introduction

Ere des données massives

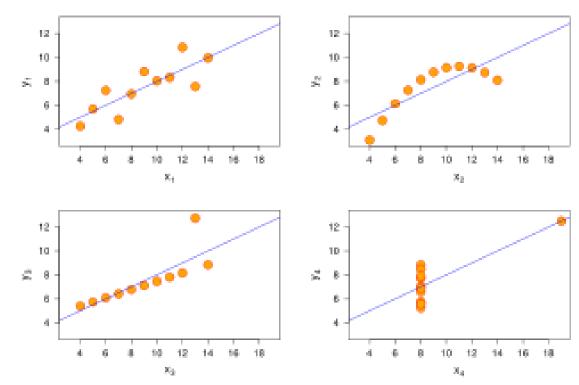
- Volume important
- p >> n
- Contrôle / organisation des données (Design, connaissance des données,...)
- Qu'est-ce qu'on cherche exactement ?



> Introduction

Vieux pots et bonne soupe

- Analyse exploratoire des données (Tukey, Anscombe)
 - Les données avant le modèle
 - Pas (trop) d'hypothèses sur les distributions
 - Accent sur la visualisation



Introduction

Vieux pots et bonne soupe

Analyses factorielles (ACP, AC, ACM,...) (back to Pearson...)

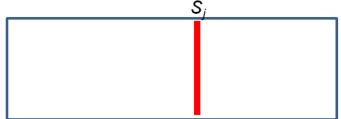
- Tables de données -> points dans des espaces vectoriels (espace des données, des variables)
- Réduction de dimensions, synthèse, formalisation, visualisation via un mécanisme unique (schéma de dualité, SVD)

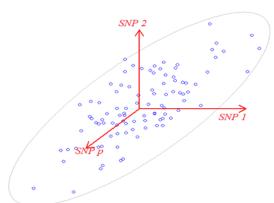


La ième observation définie par *p* variables

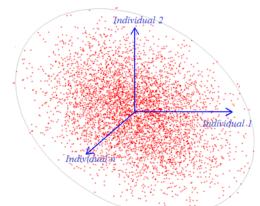


La ième variable definie par *n* observations



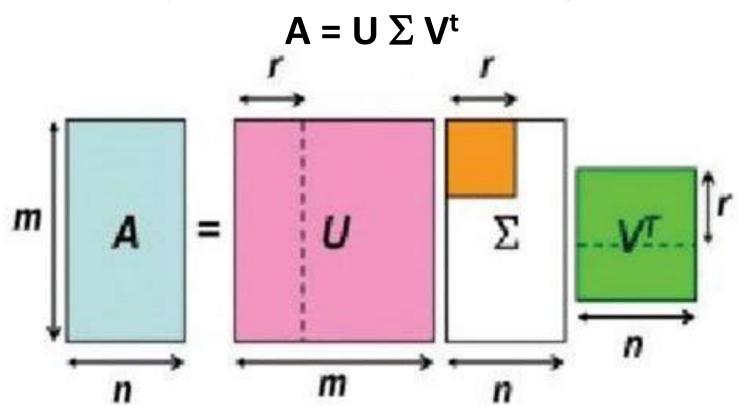


Un nuage de *n* points (observations) dans un hyperspace à *p* dimensions (variables)



Un nuage de *p* points (variables) dans un hyperspace à *n* dimensions (observations)

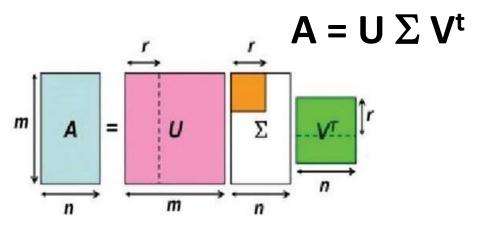
Décomposition en valeurs singulières



 Σ : matrice diagonale des valeurs singulières σ = λ ½

Théorème de Eckart-Young -> optimisation / approximation de rang r

Décomposition en valeurs singulières Schéma de dualité



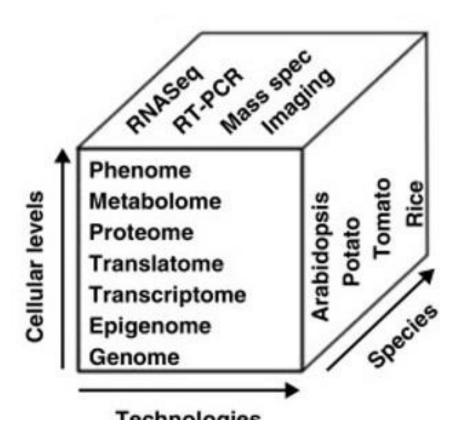
ACP: tableau de variables quantitatives

A C: tableau de contingence (table d'abondance)

ACM : tableau de variables qualitatives



Organisation hiérarchique de la complexité biologique (ou de l'information ?)





Available online at www.sciencedirect.com

ScienceDirect



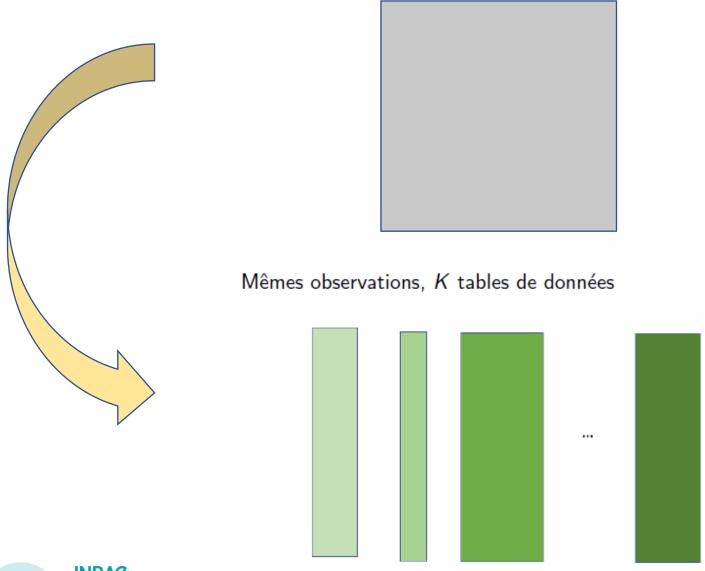


More effort — more results: recent advances in integrative 'omics' data analysis

Dhivyaa Rajasundaram^{1,2,3} and Joachim Selbig^{1,2}



D'une à plusieurs tables



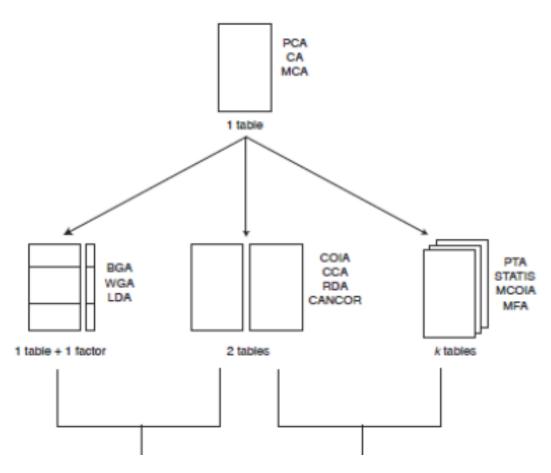
INRAe

K blocs de données

D'une à plusieurs tables

Thioulouse, 2011. Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. Ann.

Appl. Stat. 5 (4) 2300 - 2325



INRA

> Intégration de données

Analyses multitables

Ensemble d'individus décrits par plusieurs groupes de variables

- Relations entre tableaux (interstructure) Coefficient de corrélation
- Recherche d'un compromis Droite de régression
- Intrastructure (Ecart au compromis) Ecart à la droite de régression





BoSeX-Dim



Effets du sexe chromosomique et de l'environnement sur le développement précoce des bovins

Gènes à expression sexuellement dimorphique et développement précoce chez les bovins : impact de l'environnement V Duranthon

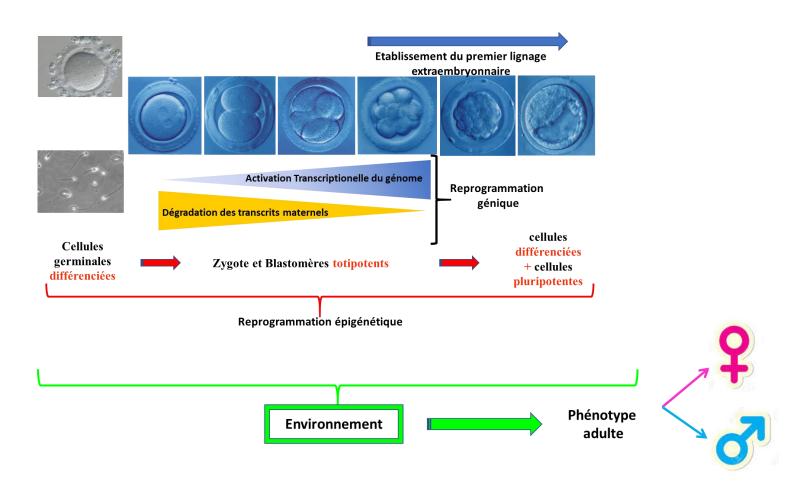
Partenaires :1. INRAE BREED équipes EPEE, DGP, MECP2, PEPPS GABI équipe GiBBS

2. Allice





Contexte biologique



INRAe

Contexte biologique

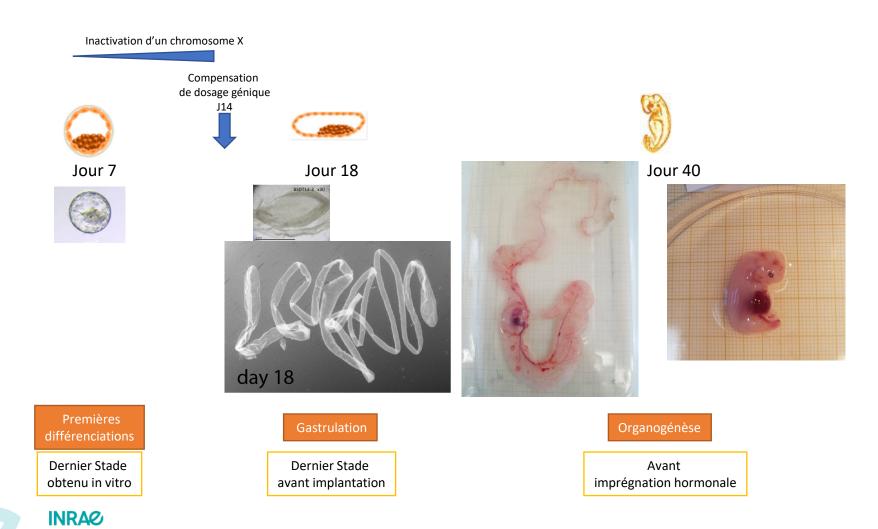
- L'environnement dans lequel se déroulent les premiers stades du développement embryonnaire influence le phénotype de l'individu à naître (concept de programmation péri-conceptionnelle).
- La manifestation phénotypique de ces effets de l'environnement diffère souvent selon le sexe de l'individu.
- Il existe donc des différences entre individus mâles et femelles avant toute imprégnation hormonale, et en particulier pendant la période préimplantatoire. Elles sont dues aux chromosomes sexuels:
 - présence du Y
 - nombre de X

L'effet du nombre de X peut différer selon les espèces pour deux raisons:

- le stade à partir duquel l'inactivation d'un X est effective dans les embryons femelles
- le nombre de gènes qui échappent à l'inactivation

INRAO

Trois stades clefs du développement



Plan expérimental pour les analyses de transcriptome



Total:160 transcriptomes

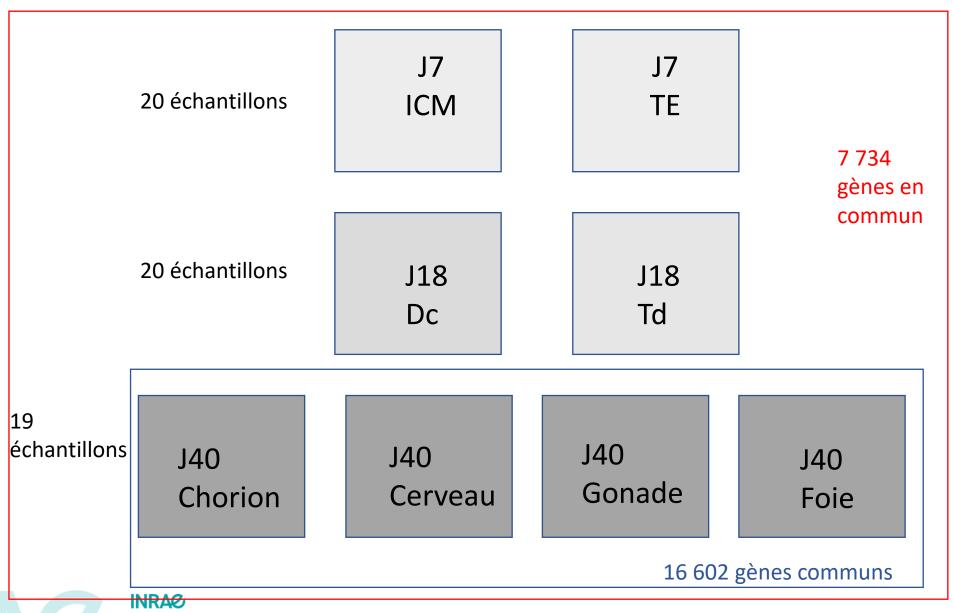
INRAe

> Intégration de données ?

- Blocs différents de données :
 - Transcriptome, méthylome, métabolome, morphologie, génome
- Répétition d'un même bloc de données:
 - Transcriptome:
 - 3 temps : D7, D18, D40
 - 8 tissus (ICM, TE;TD, Dc; Chorion, Gonade, Cerveau, Foie



TOTAL 156 échantillons

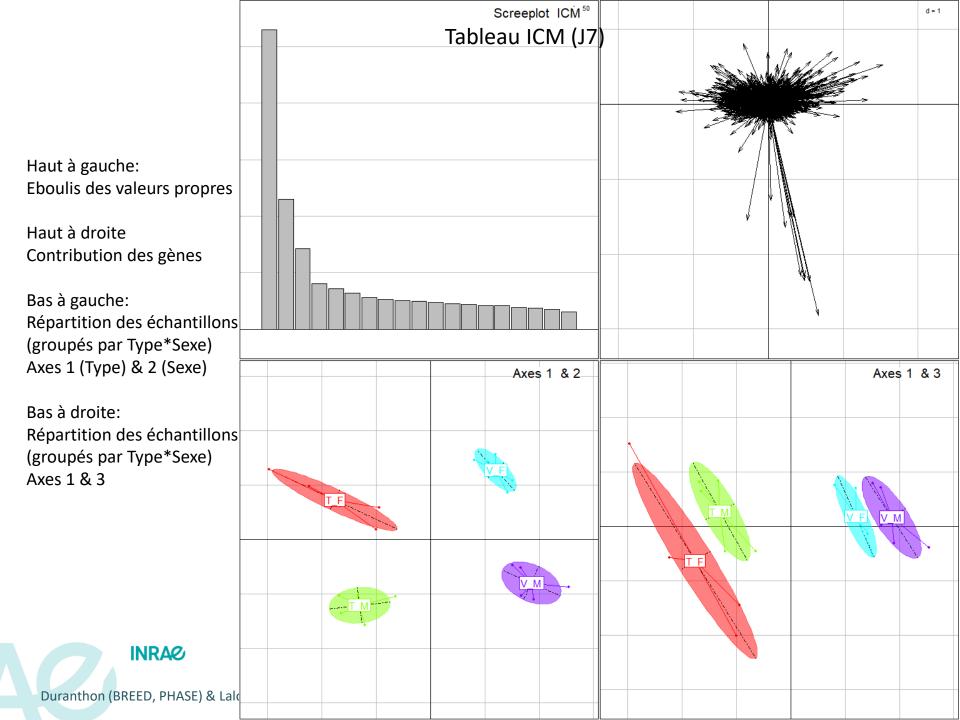


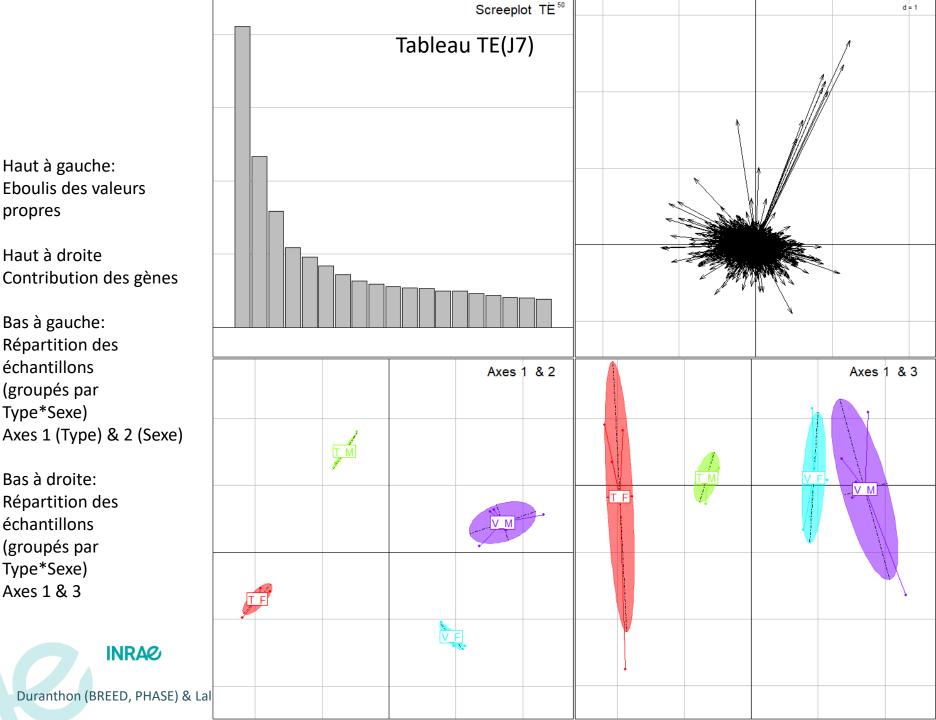
Prétraitement

- Transformation log (fonction RLOG de DESEQ2)
- Tableaux bicentrés ligne*colonne
 - Elimination effets taille
 - + robuste



- ACP tableau par tableau





Haut à gauche:

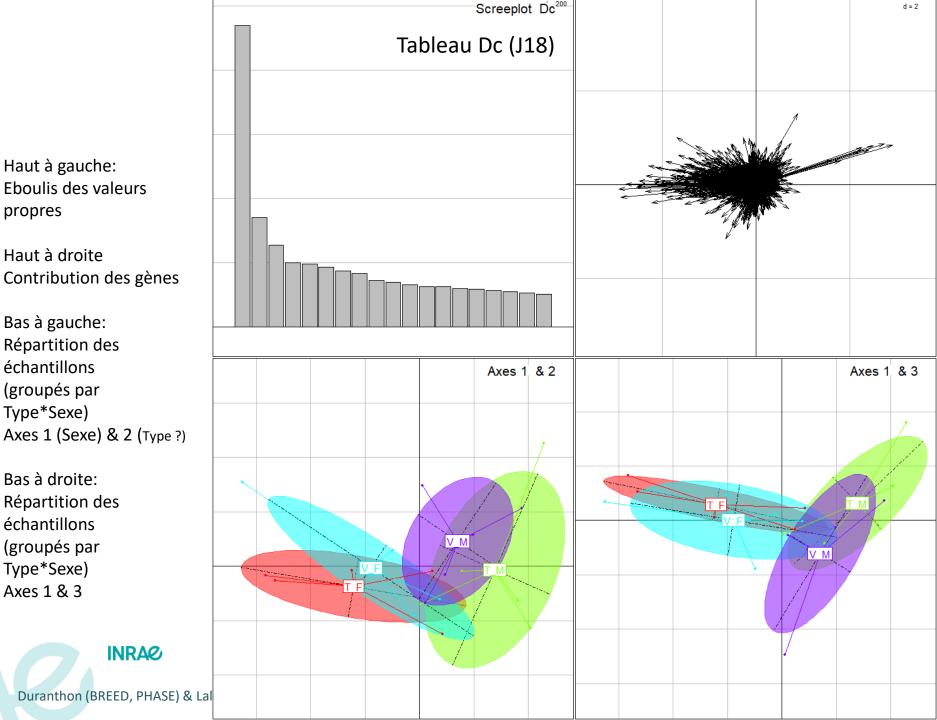
Haut à droite

Bas à gauche: Répartition des échantillons

(groupés par Type*Sexe)

Bas à droite: Répartition des échantillons (groupés par Type*Sexe) Axes 1 & 3

propres



Haut à gauche:

Haut à droite

Bas à gauche: Répartition des échantillons

(groupés par Type*Sexe)

Bas à droite: Répartition des échantillons (groupés par Type*Sexe) Axes 1 & 3

propres

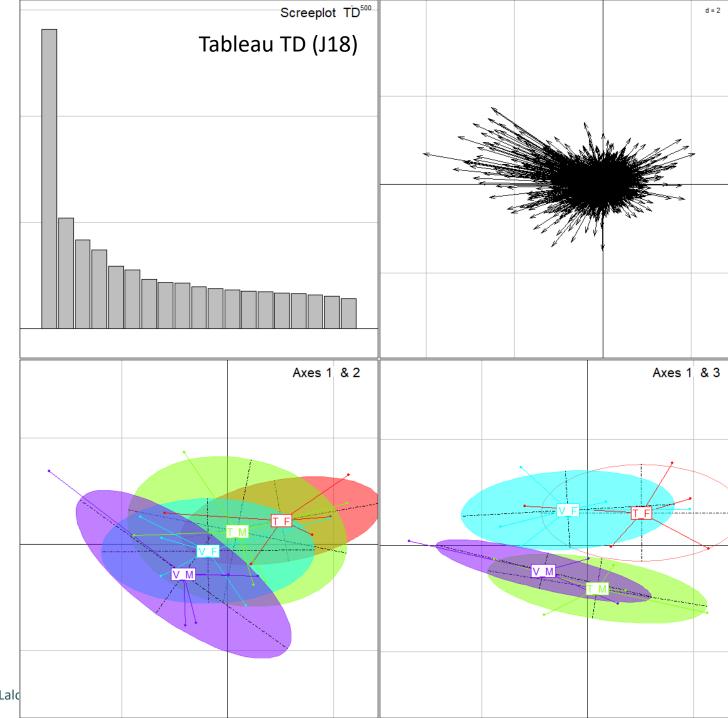
Haut à droite Contribution des gènes

Bas à gauche: Répartition des échantillons (groupés par Type*Sexe) Axes 1 (Type ?) & 2 (Sexe)

Bas à droite: Répartition des échantillons (groupés par Type*Sexe) Axes 1 & 3

INRAe

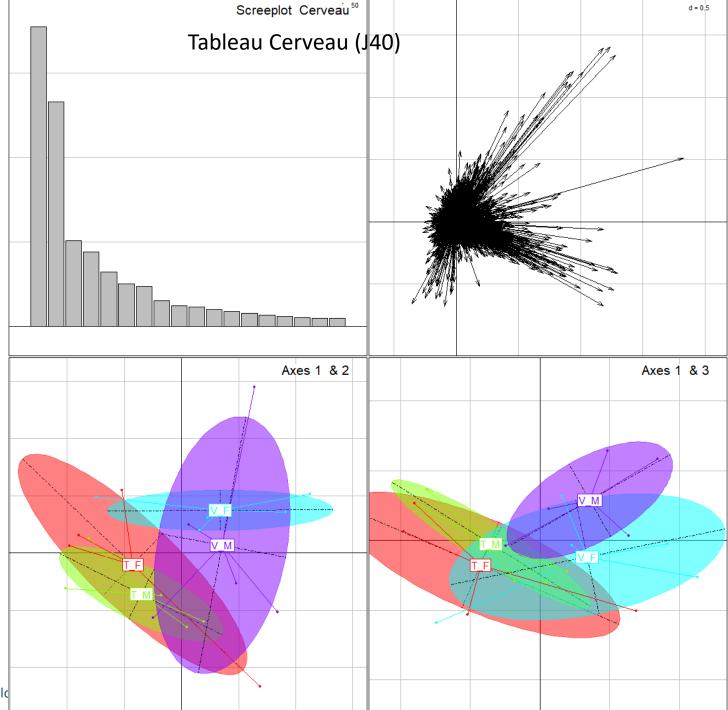
Duranthon (BREED, PHASE) & Lalc



Haut à droite Contribution des gènes

Bas à gauche: Répartition des échantillons (groupés par Type*Sexe) Axes 1 (Type ?) & 2 (Sexe)

Bas à droite: Répartition des échantillons (groupés par Type*Sexe) Axes 1 & 3



INRAe

Duranthon (BREED, PHASE) & Lalc

Haut à droite Contribution des gènes

Répartition des échantillons (groupés par Type*Sexe) Axes 1 (Sexe ?) & 2 (Type)

Répartition des échantillons (groupés par Type*Sexe) Axes 1 & 3

Bas à gauche: Axes 1 & 2 Axes 1 & 3 Bas à droite: V_M INRAe Duranthon (BREED, PHASE) & Lalc

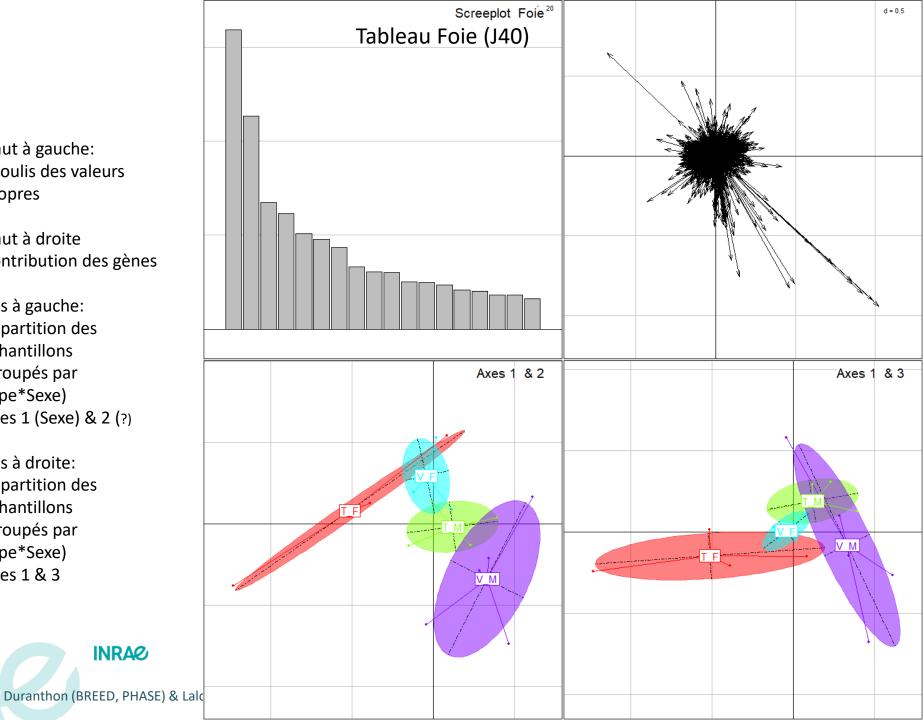
Screeplot Chorion 50

Tableau Chorion (J40)

Haut à droite Contribution des gènes

Bas à gauche: Répartition des échantillons (groupés par Type*Sexe) Axes 1 (Sexe) & 2 (?)

Bas à droite: Répartition des échantillons (groupés par Type*Sexe) Axes 1 & 3



Haut à droite Contribution des gènes

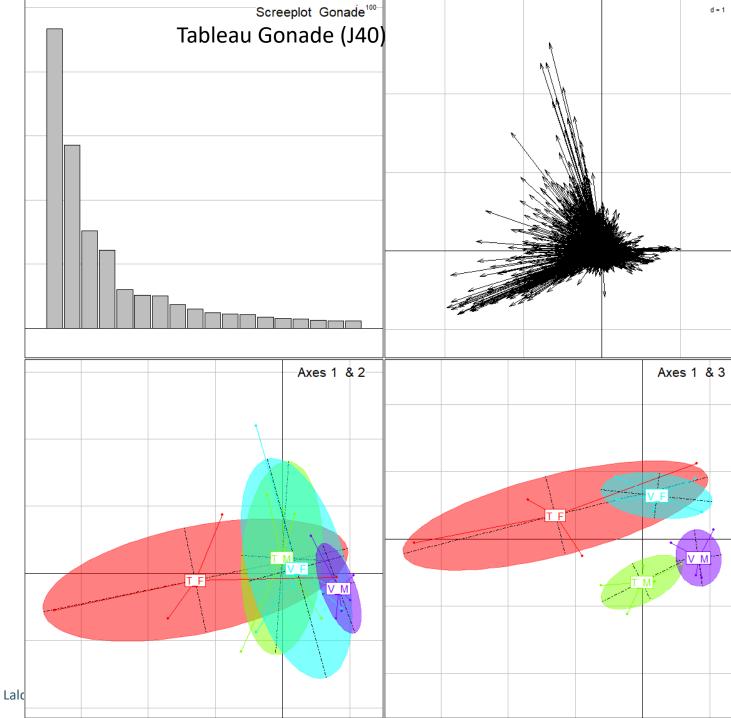
Bas à gauche: Répartition des échantillons (groupés par Type*Sexe) Axes 1 (?) & 2 (?)

Bas à droite:

Répartition des échantillons (groupés par Type*Sexe) Axes 1 & 3 (Type*Sexe)



Duranthon (BREED, PHASE) & Lalc



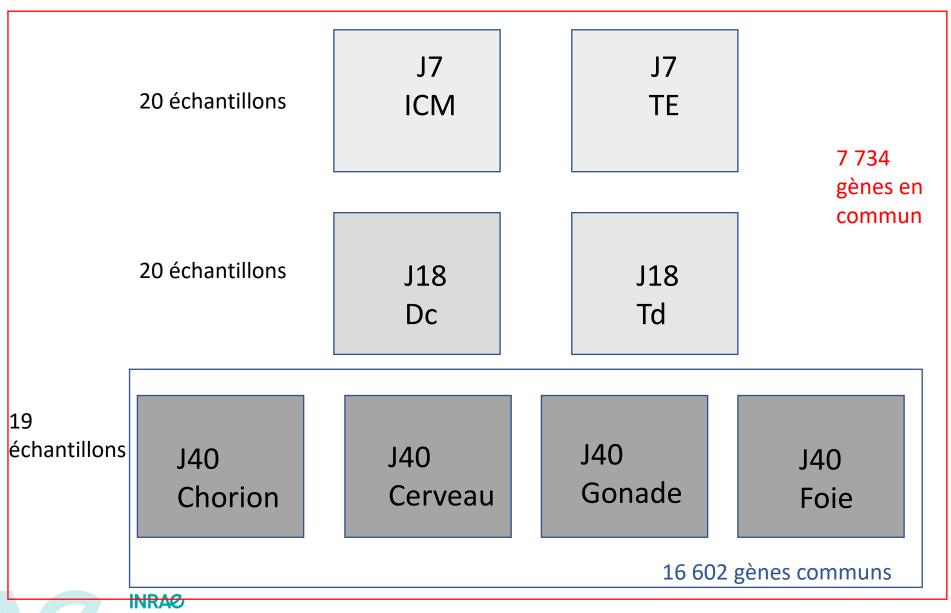
- ICM (J7): l'axe 1 oppose le type, et l'axe 2 le sexe;
- TE (J7): idem;
- DC (J18); Axe 1, oppose le sexe;
- TD (J18): on retrouve les oppositions de J7 avec un axe 1 distinguant le type et un axe 3 le sexe;
- Cerveau (J40): Axe 1 : type
- Chorion (J40): ?
- Foie (J40) : Les axes ½ opposent les sexes
- Gonade (J40): Axe 3 : sexe



INRAO

Analyse transcriptome J40 Analyse triadique partielle (ACP multi-tissus)

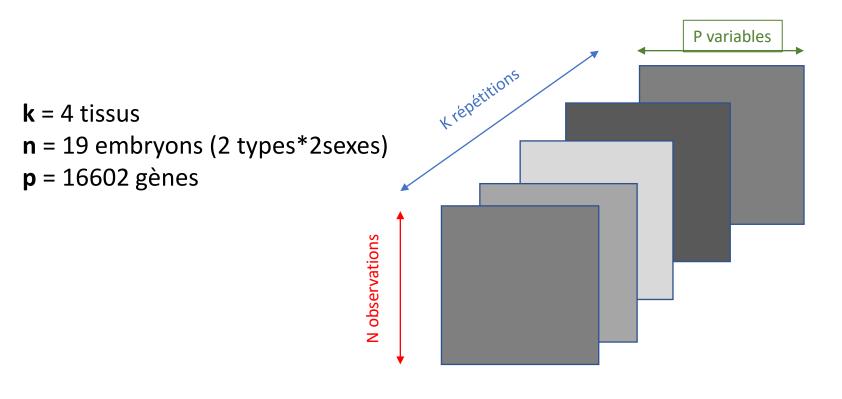
TOTAL 156 échantillons



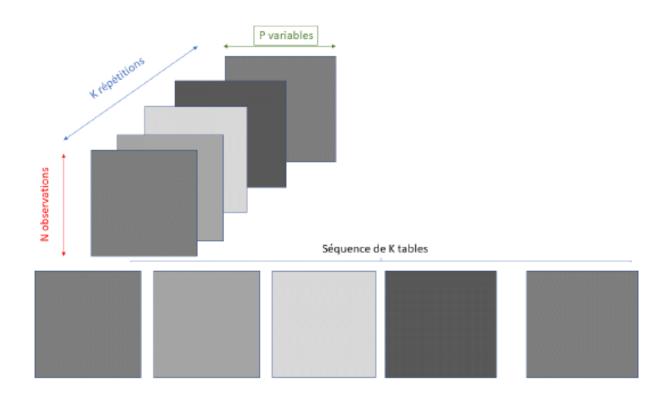
> Intégration de données ? Prise en compte de la structure

Structure en cube:

Mêmes observations; même variables; Répétition (espace, temps)

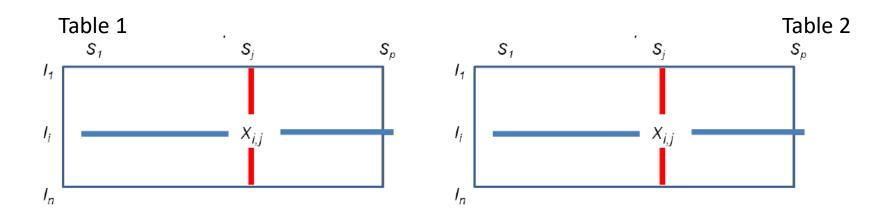


D'une structure en cube à une séquence de K tables L'Analyse Triadique Partielle





Analyse Triadique Partielle

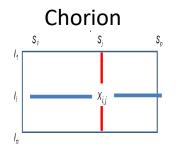


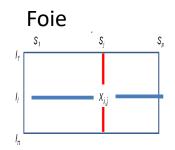
1. Relation entre tableaux :

Comment quantifier la ressemblance entre deux tables ?

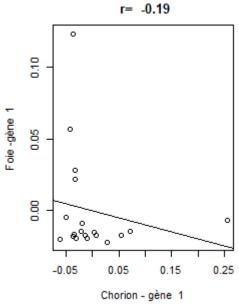
Analyse Triadique Partielle

1. Comment quantifier la ressemblance entre deux tables ?





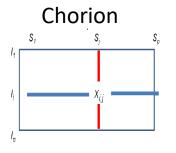
1 gène : Corrélation entre expression dans le chorion et expression dans le foie r=-0,19

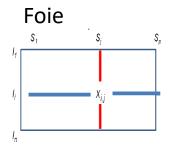


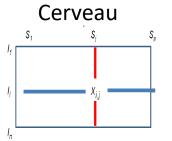
Gène 1 - Chorion vs Foie

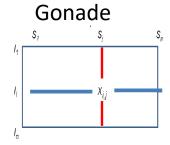
Analyse Triadique Partielle

1. Comment quantifier la ressemblance entre deux tables ?



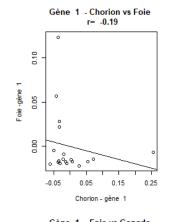


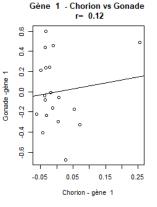


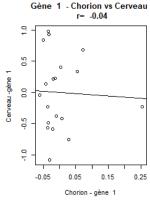


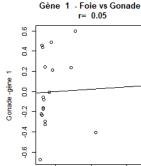
Gène 1

	chorio n	foie	gonade	cerveau
chorion	1	-0,19	0,12	-0,04
foie	-0,19	1	0,05	0,07
gonade	0,12	0,05	1	0,22
cerveau	-0,04	0,07	0,22	1





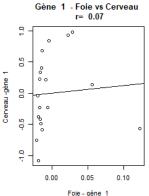


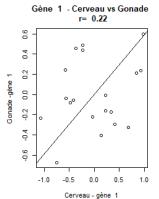


0.10

0.05

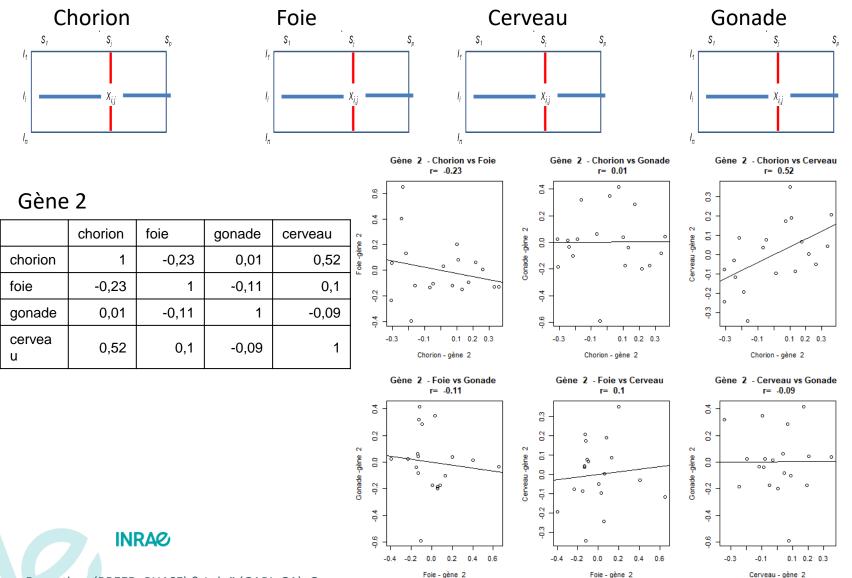
Foie - gène 1





INRAO

1. Comment quantifier la ressemblance entre deux tables ?



1. Comment quantifier la ressemblance entre deux tables ?

Gène 1

	chorion	foie	gonad e	cerveau
chorion	1	-0,19	0,12	-0,04
foie	-0,19	1	0,05	0,07
gonade	0,12	0,05	1	0,22
cerveau	-0,04	0,07	0,22	1

Gène 2

	chorion	foie	gona de	cerveau
chorion	1	-0,23	0,01	0,52
foie	-0,23	1	-0,11	0,1
gonade	0,01	-0,11	1	-0,09
cerveau	0,52	0,1	-0,09	1

Corrélations moyennes entre tables (RV)

	cerveau	chorion	foie	gonade
cerveau	1.00	0.09	0.12	0.14
chorion	0.09	1.00	0.12	0.06
foie	0.12	0.12	1.00	0.06
gonade	0.14	0.06	0.06	1.00

Gène 16602

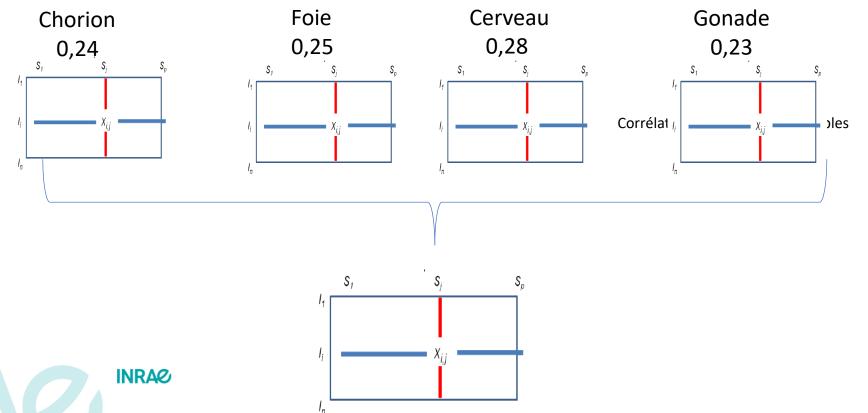
	chorion	foie	gonade	cerveau
chorion	1	0,52	0,18	0,03
foie	0,52	1	0,01	0,17
gonade	0,18	0,01	1	0,3
cerveau	0,03	0,17	0,3	1

INRAO

Etablissement d'un compromis: table moyenne

	cerveau	chorion	foie	gonade
cerveau	1.00	0.09	0.12	0.14
chorion	0.09	1.00	0.12	0.06
foie	0.12	0.12	1.00	0.06
gonade	0.14	0.06	0.06	1.00

Moyenne pondérée : plus les tables se ressemblent, plus leur poids dans la moyenne est importante



2. Établissement d'un compromis : moyenne pondérée des tables

- Pondérée : plus les tables se ressemblent, plus elle pèsent dans le calcul de la moyenne
- Coordonnées des embryons : comment les individus s'organisent (par exemple en fonction du type et du sexe
- Coordonnée / Contribution des gènes: Quel poids ont les gènes dans la construction de cette organisation

3 .Variations autour du compromis : Quel est le rôle de chaque tissu dans cette construction

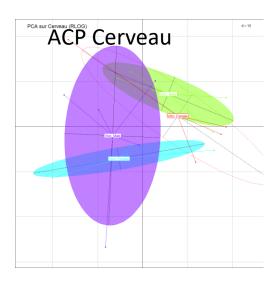
 Coordonnées « partielles » des embryons :
 Coordonnée / Contribution partielle des gènes

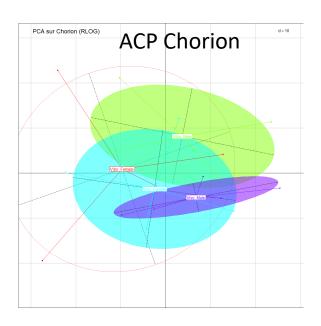
Pour chaque axe et chaque gène:

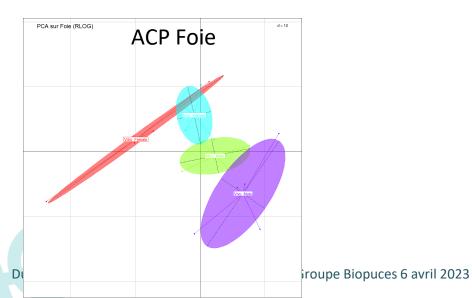
- Une coordonnée moyenne (compromis);
- 4 coordonnées partielles (foie, gonade, cerveau, chorion)
 - Si les 4 coordonnées sont +/- les mêmes : gène « cohérent » quel que soit le tissu
 - Si les coordonnées sont différentes, l'expression du gène est différent selon les tissus

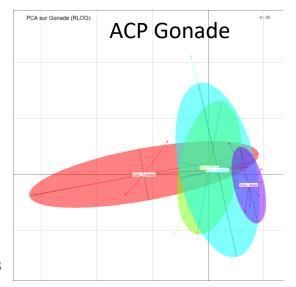
INRAO

Des ACP par tissu...

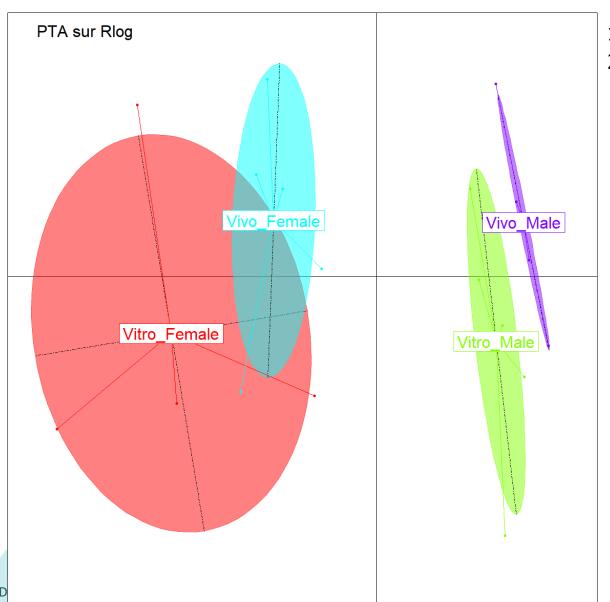








... A l'analyse du compromis entre les 4 tissus



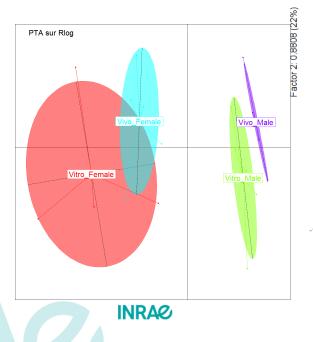
1^{er} axe différenciant les sexes2^{ème} axe différenciant les types

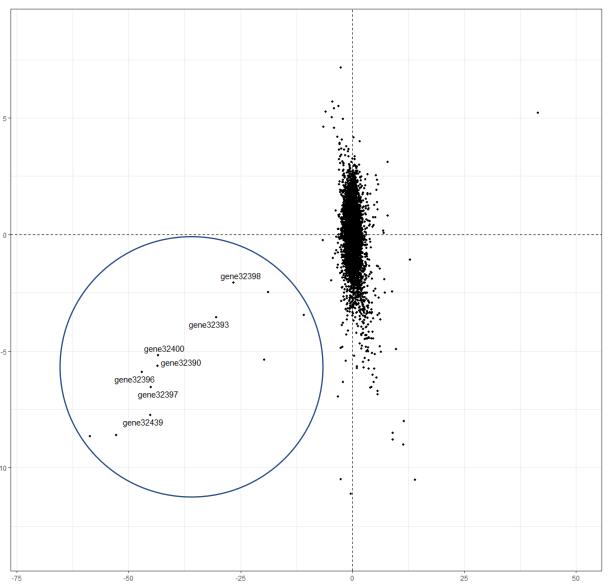
Pour chaque axe et chaque gène:

- Une coordonnée moyenne (compromis);
- 4 coordonnées partielles (foie, gonade, cerveau, chorion)
 - Si les 4 coordonnées sont +/- les mêmes : gène « cohérent » quel que soit le tissu
 - Si les coordonnées sont différentes, l'expression du gène est différent selon les tissus

ACP sur les coordonnées partielles des gènes sur le premier axe

Le 1^{er} axe exhibe des gènes « cohérents » et influents pour la différenciation entre sexes





Factor 1: 1.9019 (47.5%)

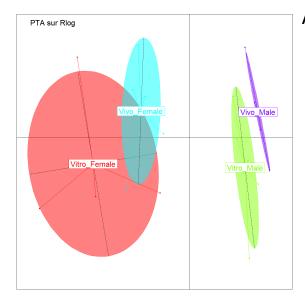
Duranthon (BREED, PHASE) & Laloë (GABI

ACP sur les coordonnées des gènes sur le premier axe

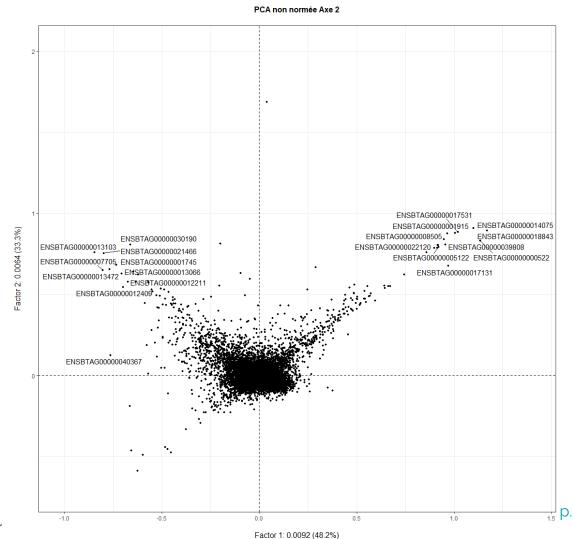
С	cerveau	chorion	foie	gonade	moy	ampli
gene32392	2,10	1,80	1,60	2,31	1,95	0,71
gene32401	1,84	1,66	1,50	2,01	1,75	0,51
gene32396	1,54	1,55	1,33	1,99	1,60	0,66
gene32439	1,43	1,44	1,41	1,85	1,53	0,44
gene32397	1,42	1,54	1,34	1,82	1,53	0,48
gene32390	1,62	1,14	1,13	1,95	1,46	0,82
gene32400	1,55	1,28	1,14	1,88	1,46	0,74
ENSBTAG00 000054758	-1,60	-0,81	-1,05	-2,13	-1,40	1,32
gene32393	0,91	1,09	0,90	1,32	1,06	0,41



- Ce sont des gènes localisés sur le Y ou sur le X*: pas très étonnant qu'on les retrouve ici.
- Les résultats valident l'approche: preuve de concept
- Regardons ce que cela donne pour l'effet condition...



ACP sur les coordonnées des gènes sur le deuxième axe



INRAO





С	cerveau	chorion	foie	gonade	moy	ampli	Gene name
ENSBTAG00000010991	1,292	0,067	0,041	-1,129	0,068	2,421	TTR
ENSBTAG00000018843	-0,039	-0,173	-0,019	-1,466	-0,424	1,447	SERPINA1
ENSBTAG00000014075	0,029	-0,196	-0,032	-1,424	-0,406	1,453	ITIH2
ENSBTAG00000000522	-0,045	-0,236	0,006	-1,398	-0,418	1,404	AHSG
ENSBTAG00000017531	0,018	-0,060	-0,003	-1,352	-0,349	1,370	FETUB
ENSBTAG00000001915	0,002	0,010	-0,031	-1,337	-0,339	1,347	АРОН
ENSBTAG00000017294	0,024	0,013	0,015	-1,303	-0,313	1,327	ORM1
ENSBTAG00000008505	0,029	-0,039	-0,081	-1,270	-0,340	1,299	АРОВ
ENSBTAG00000039808	-0,009	-0,024	-0,016	-1,252	-0,325	1,244	SERPINA6
ENSBTAG00000001638	-0,008	0,047	-0,014	-1,222	-0,299	1,269	FGA
ENSBTAG00000001271	-0,002	0,024	-0,030	-1,223	-0,308	1,246	PLG
ENSBTAG00000005122	-0,007	0,008	-0,052	-1,219	-0,317	1,227	KNG1
ENSBTAG00000009212	0,000	-0,020	0,008	-1,206	-0,304	1,214	APOA2
ENSBTAG00000017131	-0,042	-0,274	-0,017	-1,172	-0,377	1,155	AFP
ENSBTAG00000022120	0,023	-0,052	-0,033	-1,194	-0,314	1,216	FGB
ENSBTAG00000006745	-0,009	0,064	-0,034	-1,152	-0,283	1,216	FGG
ENSBTAG00000013103	1,113	0,227	0,030	0,142	0,378	1,083	COL1A1
ENSBTAG00000021466	1,110	0,136	-0,017	0,112	0,335	1,127	COL3A1
ENSBTAG00000007705	1,049	0,085	-0,059	0,184	0,315	1,108	COLEC12
ENSBTAG00000030190	1,015	0,273	-0,075	-0,033	0,295	1,090	COL6A3
ENSBTAG00000001745	1,011	0,116	0,025	0,111	0,315	0,986	LUM

En conclusion

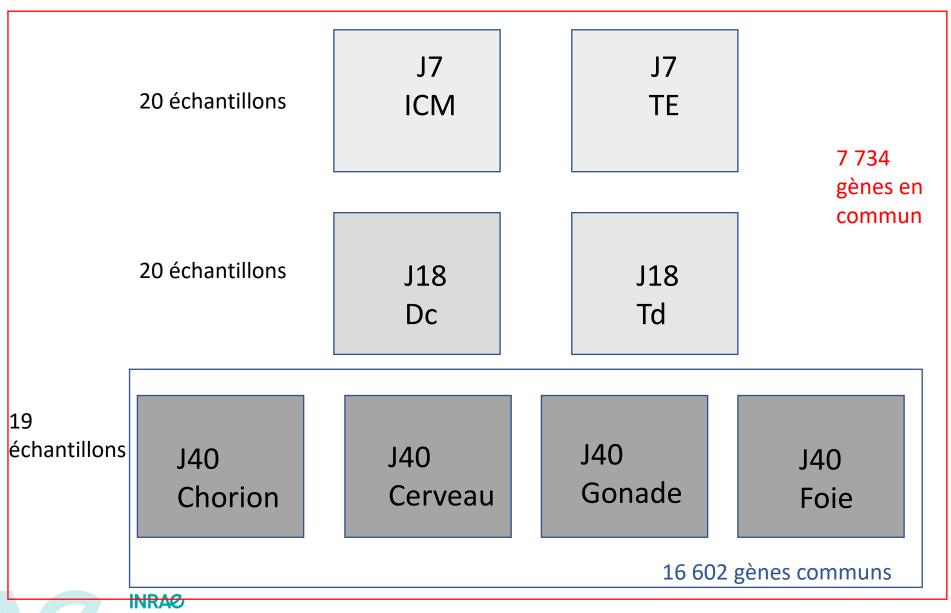
- Augmente la structuration des observations (1^{er} axe= sexe; 2nd axe = vivo/vitro)
 par rapport à l'analyse tissu / tissu
- Permet d'établir une typologie de l'action des gènes selon la différenciation
- 1^{er} axe : gènes cohérents (même action sur les quatre tissus) : « gene »
- 2^{ème} axe : gènes bcp moins cohérents, différentiels sur un seul tissu:
 - Surexpression vitro pour la gonade
 - Surexpression vivo pour le cerveau,
 - Combinaison des deux (TTR)

INRAO

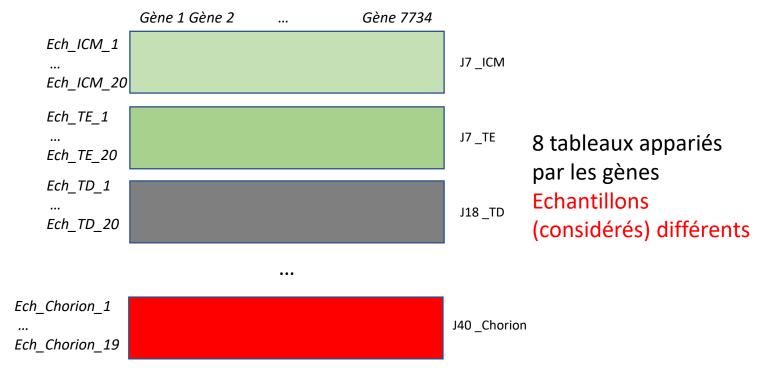
Analyse Intégrative sur les 8 transcriptomes BoSexDim



TOTAL 156 échantillons



Principe de l' Analyse Factorielle Multiple sur les 8 transcriptomes BoSexDim

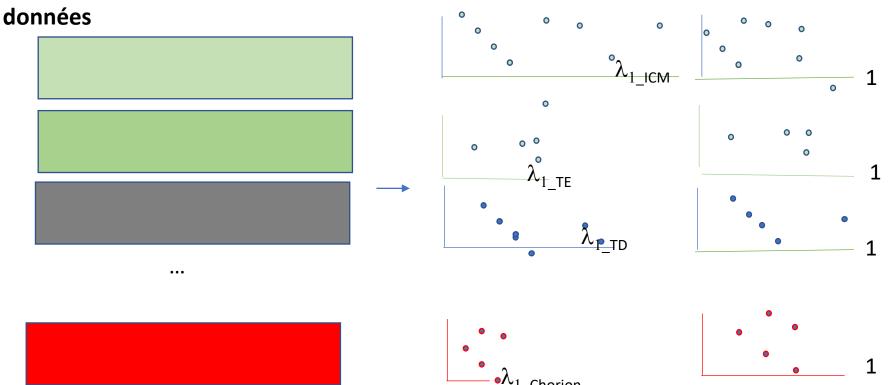


=156 échantillons

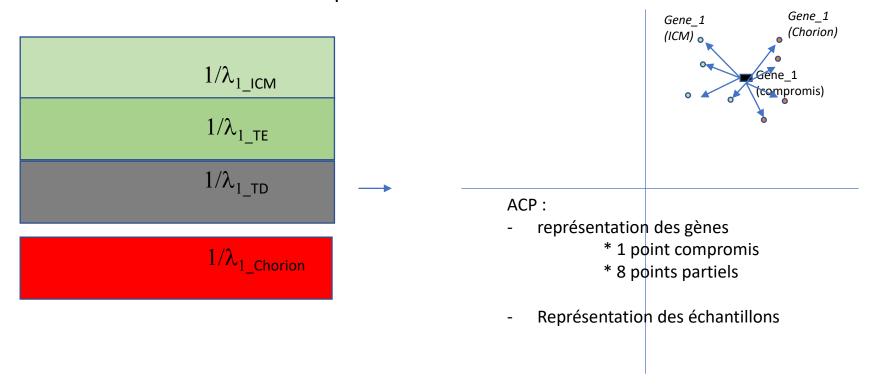
=8 blocs de données

Etape de pondération par $1/\lambda_1$:

Standardisation par l'échelle du premier axe de l'ACP de chaque bloc de



ACP sur la combinaison des blocs pondérés





Marche pas On ne prend pas en compte la structuration des échantillons en sexe*type

-> Trouver autre chose...

Ce ne sont pas les mêmes échantillons, mais il y a une structure en sexe*type (cf plus haut)

-> Analyse sur les moyennes des échantillons par sexe*type



Analyse triadique partielle 8 transcriptomes

On retrouve une structure en cube, avec

n=4

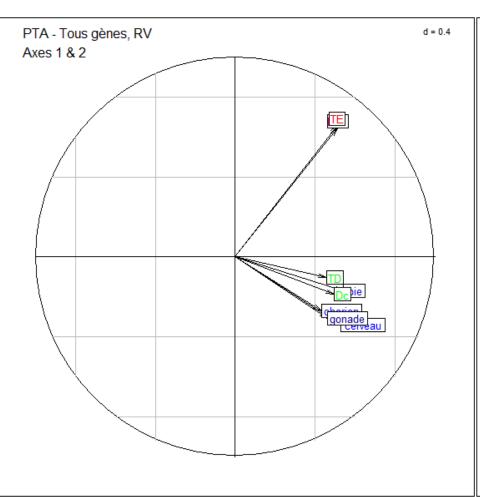
K = 8

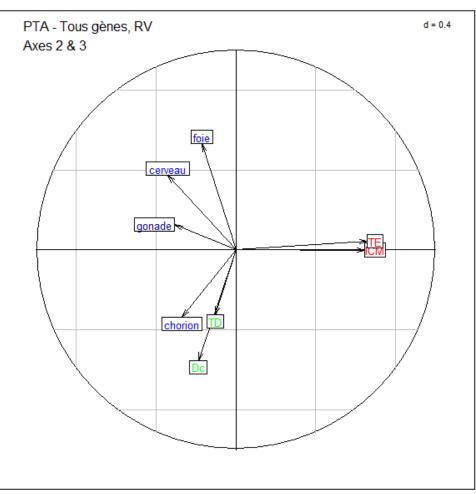
P=26205

Relations entre tissus RV

	cerveau	chorion	foie	gonade	ICM	TE	Dc	TD
cerveau	1.00	0.14	0.26	0.17	0.08	0.08	0.14	0.15
chorion	0.14	1.00	0.11	0.15	0.09	0.06	0.17	0.09
foie	0.26	0.11	1.00	0.17	0.12	0.14	0.11	0.13
gonade	0.17	0.15	0.17	1.00	0.08	0.07	0.14	0.09
ICM	0.08	0.09	0.12	0.08	1.00	0.40	0.11	0.09
TE	0.08	0.06	0.14	0.07	0.40	1.00	0.10	0.12
Dc	0.14	0.17	0.11	0.14	0.11	0.10	1.00	0.18
TD	0.15	0.09	0.13	0.09	0.09	0.12	0.18	1.00

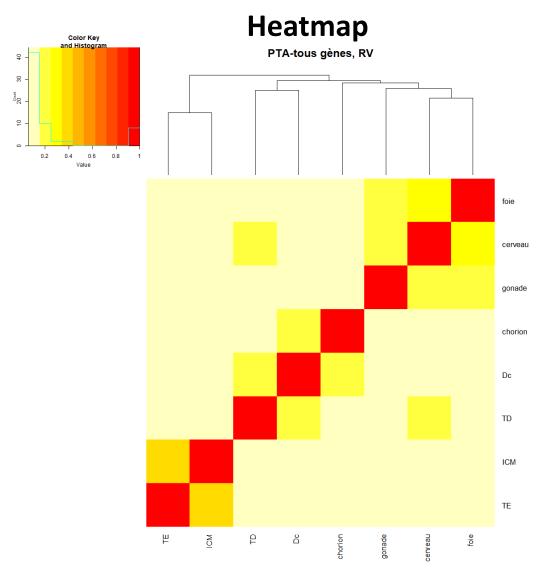
Analyse Triadique partielle 8 transcriptomes Relations entre tissus Cercle des corrélations



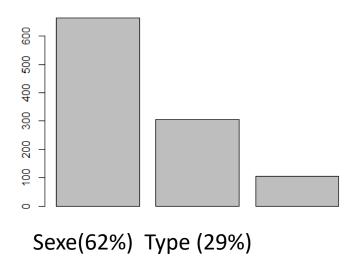


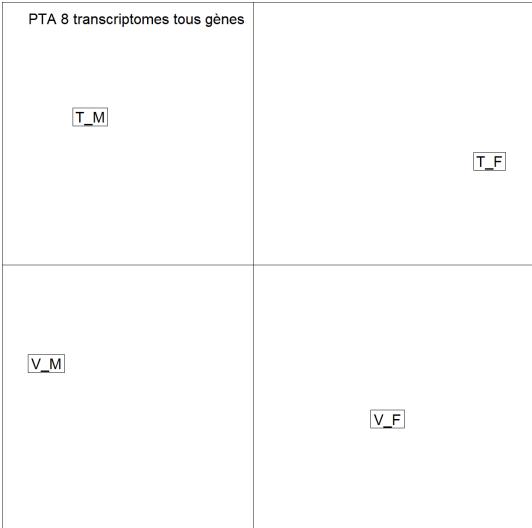
INRA

Analyse Triadique partielle 8 transcriptomes Relations entre tissus

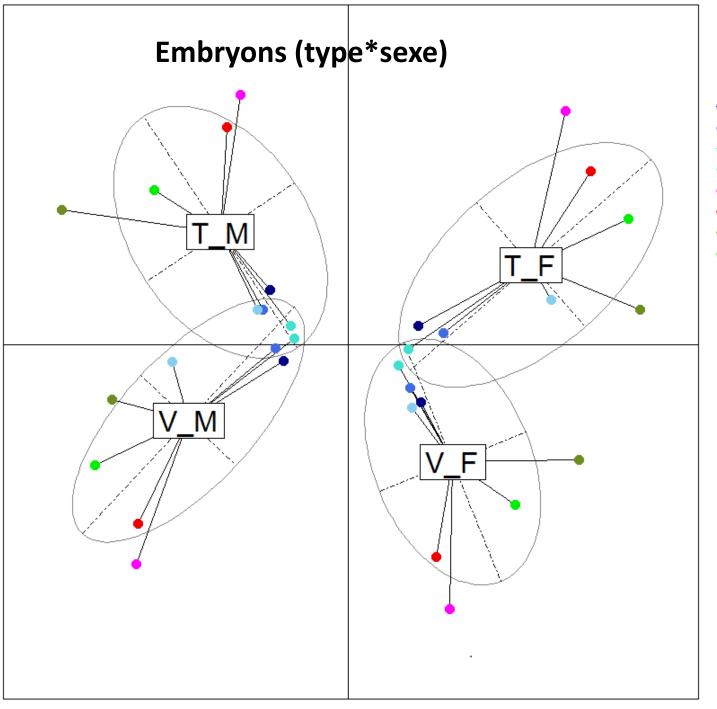


Embryons (type*sexe)



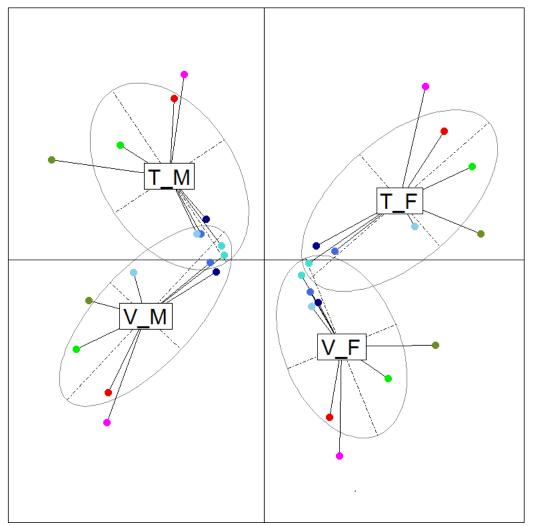


INRAe



- cerveau
- chorion foie
- gonade ICM TE

Embryons (type*sexe)



J7: Type/Sexe

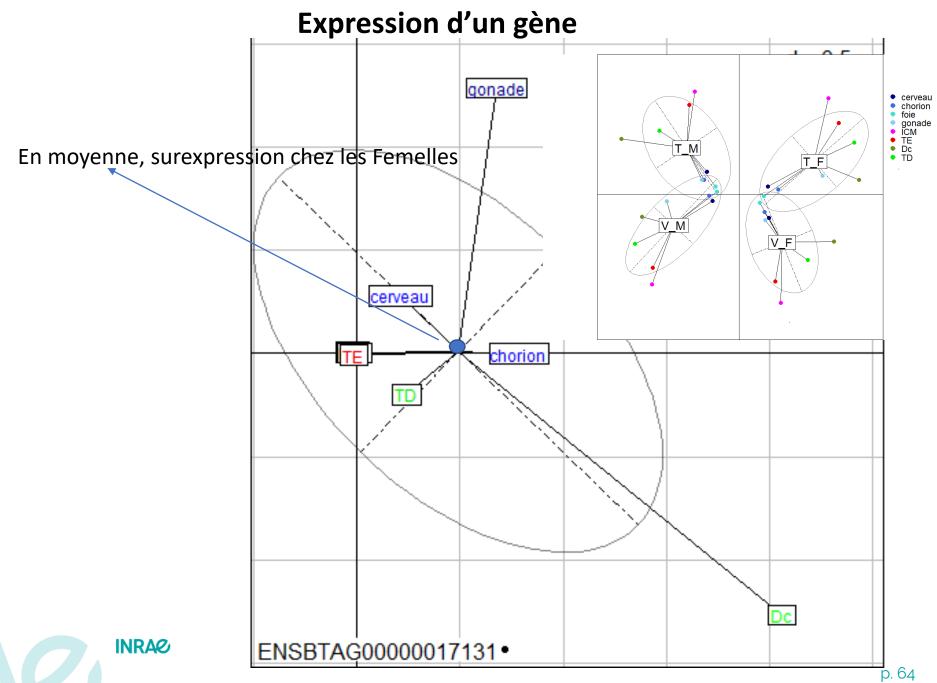
J18: Sexe / Type

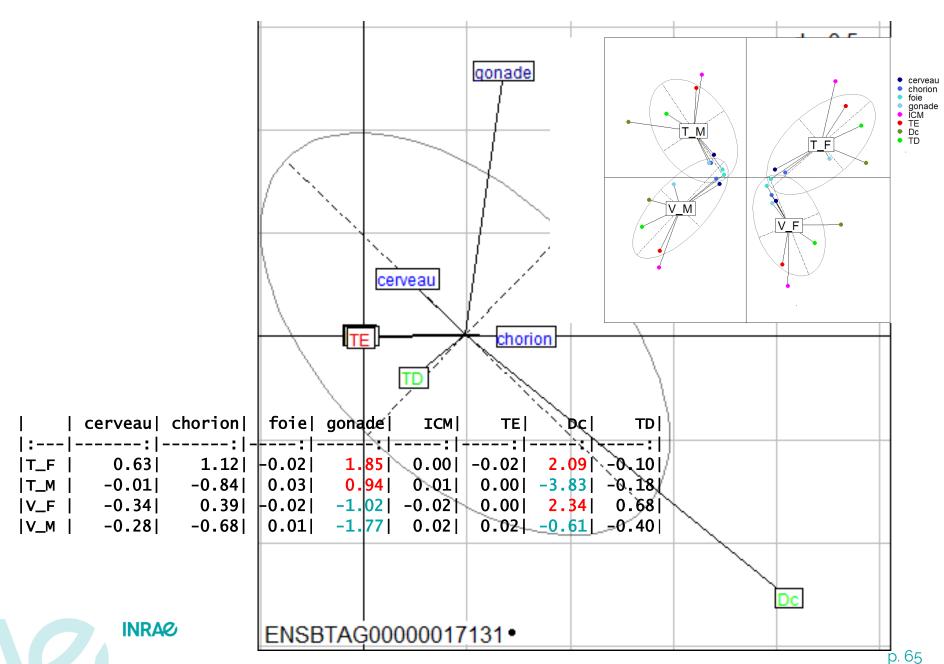
J40: Sexe

cerveau chorion foie

gonade ICM TE

Dc TD





Intégration des transcriptomes via l'analyse triadique partielle

- Structuration commune aux 8 transcriptomes
- Spécificité des tissus
- « Comportement » des gènes selon les tissus
- Structuration des transcriptomes
- -> aller plus loin dans la clusterisation des gènes...



Remerciements

Ludivine Laffont
Sylvie Ruffini
Nathalie Peynot
Nathalie Daniel
Vincent Brochard

Alice Jouneau Isabelle Hue Eugénie Canon Christophe Richard

Hélène Jammes Aurélie Chaulot-Talmon Valérie Gélin Marjolaine Andre Geneviève Jolivet Catherine Archilla
Anne Aubert Frambourg
Luc Jouneau
Béatrice Mandon-Pépin
Eric Pailhoux

Aurélie Bonnet
Daniel Le Bourhis
Olivier Desnoes
L. Le Berre
Pascal Salvetti
Laurent Schibler

Florence Jaffrézic Andrea Rau

