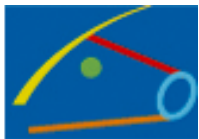


**Problèmes pratiques rencontrés dans la manipulations des données de la génomique**

**Exemple des bovins laitiers**

# Présentation du programme SAM

- **Programme SAM (2001-2010) dans 3 principales races bovines laitières : Holstein, Normande, Montbéliarde.**
- **Génotypages sur microsatellites puis à partir de l'été 2008 en routine sur puce BovineSNP50 d'Illumina.**
- **Les programmes de recherche CARTOFINE (2006-2008), CARTOFINE3 (2009-2010) et le consortium EUROGENOMICS (2010) ont apportés les typages des populations de référence des 3 races.**



# Présentation du programme SAM



**Populations de référence** : 16000 taureaux Holstein, 1500 Montbéliards, 1250 Normands

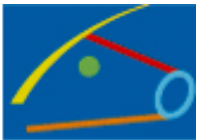
2520 femelles Holstein  
1121 Montbéliardes  
993 Normandes



2455 jeunes taureaux  
Holstein diffusés  
596 Montbéliards  
541 Normands

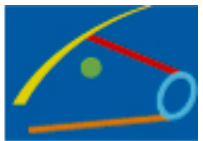


16048 candidats Holstein  
5596 Montbéliards  
2900 Normands

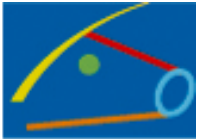


# Présentation du programme SAM

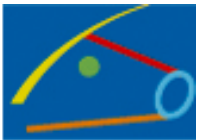
- **Cartographie de QTL et études de SG utilisent uniquement les populations de référence : 16000, 1500 et 1250 taureaux.**
- **Evaluations génomiques de routine utilisent tous les animaux typés : 37023, 8813 et 5684 animaux.**
  - Puce v1 de 54001 SNP jusqu'à oct. 2010
  - Puce v2 de 54609 SNP (=54001 – 1661 + 2269) à partir de nov. 2010



## 1 – Cartes génétiques ?



- **Cartes génétiques ?**
- **2 assemblages bovins disponibles : btau4.0 et UMD3.**
- **Étude sur 58349 SNP (dont ceux des puces v1 et v2)**



## ■ Cartes génétiques ?

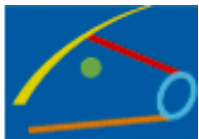
Btau4.0

UMD3

BTA	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	X	Y					
0	456	145	86	156	118	113	91	90	88	78	114	72	82	70	65	131	70	68	96	54	55	66	39	50	38	33	48	43	34	64	107						
1	105	3386	3		2	3				2	8	1												4			1				1						
2	72	2	2809	9	1						1		1			1																1					
3	65			2558				1		4						1																	1				
4	37		1		2564		1	1		1	13	1				2				1								1					1				
5	72	1				2251				1											1																
6	80						2559	2										1	1		1																
7	78																																				
8	54							2	2434		2			1															1		1		4				
9	64		1								2074			1		2					1					1	1										
10	58			1								2195																									
11	48	1	2				1																														
12	113																																				
13	61		2																																1		
14	114			1																																	
15	73								1																												
16	146																																				
17	29							1																													
18	37						1							1																							
19	47			1							1																										
20	21																																				
21	62			1																																	
22	21			27																																	
23	23											1																									
24	29		1				1																														
25	15				1			1																													
26	54																																				
27	14																																				
28	41																																				
29	49																																				
X	401		1						6																												
Y	1																																				14



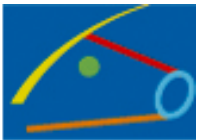
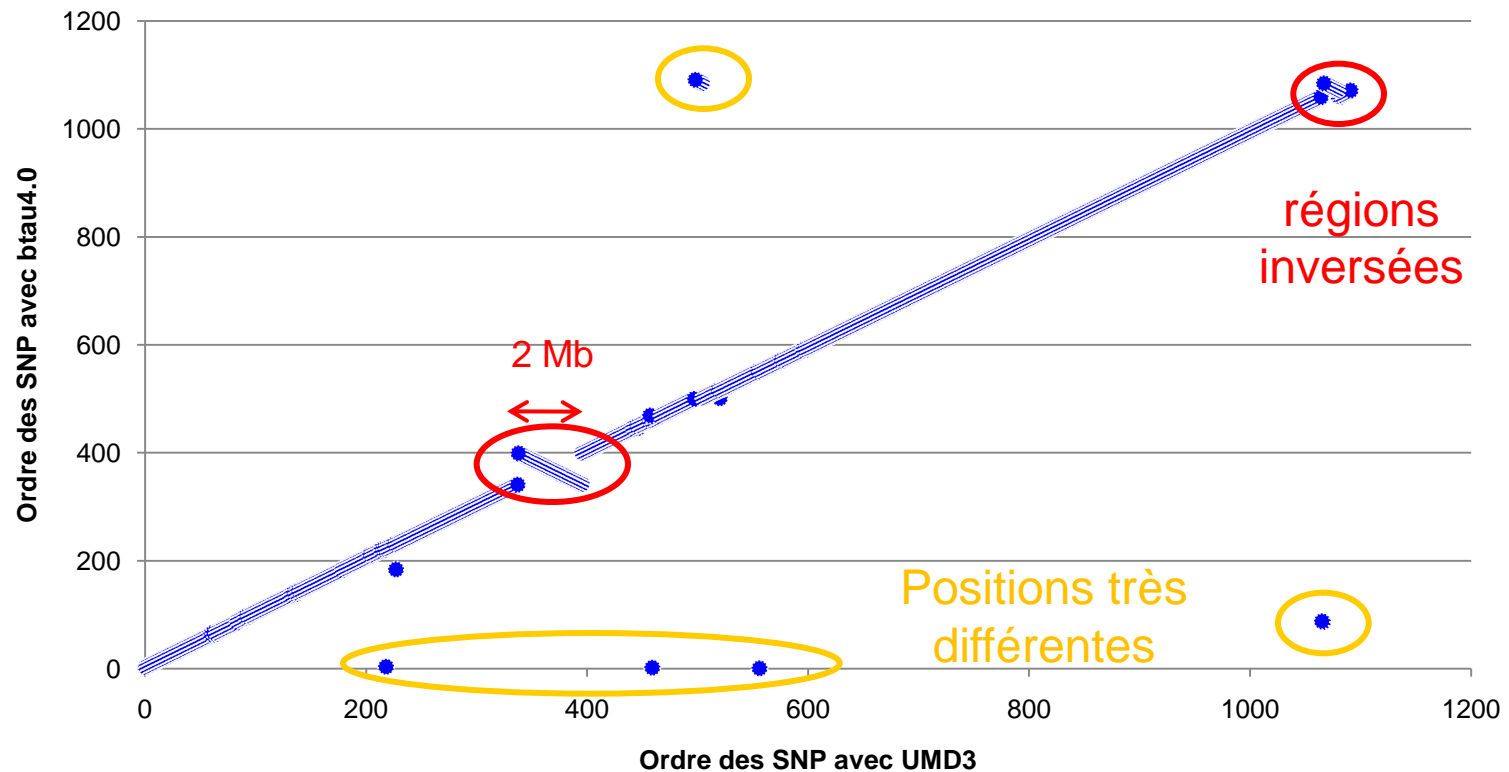
- **Cartes génétiques ?**
- **2540 SNP non positionnés sur btau4.0**
- **2820 SNP non positionnés sur UMD3**
  - dont 456 ni sur btau4.0, ni sur UMD3
- **53258 SNP positionnés sur les mêmes chromosomes**
- **167 SNP positionnés sur des chromosomes différents**
  - dont 27 sur BTA03 (btau4.0) vs BTA22 (UMD3)



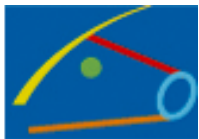


## ■ Cartes génétiques ?

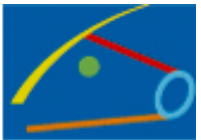
### Exemple de BTA26



- **Cartes génétiques?**
- **En pratique, utilisation de btau4.0 avec hypothèse 1 Mb = 1 cM**
- **Conséquences ?**
  - Phases, cartographies, imputations, etc.
- **Outil(s) pour construire cartes génétiques avec ce type de données ?**
  - Comment faire avec puce HD (~800000 SNP) ?



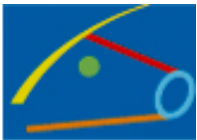
## 2 – Animaux retenus pour analyses



## ▪ Retirer animaux « douteux »

- 500 SNP très informatifs (MAF moyenne > 40%)
- Vérification compatibilité individu avec parent(s) typé(s)
  - Travail sur échantillon et non sur animal
- Si pb sur plus 5 SNP (1%), l'animal est retiré du système sauf si le pb concerne l'un des parents
  - Difficile de savoir si pb de filiation ou inversion d'échantillon

➔ **Etape indispensable avant autres travaux de préparation !**



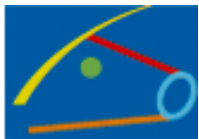
- Poids des « perf » (edc) ?

Nb filles	Holstein	Montbéliarde	Normande
<100	75%	83%	88%
<500	16%	9%	3%
<1000	2%	1%	1%
<10000	5%	5%	6%
<50000	1%	2%	2%
>50000	1%	0.1%	0.1%

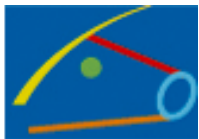
} ~90%

} ~10%

- Doit-on mettre un seuil minimal à l'edc ?
- Doit-on les borner ?
- Les poids ont-ils un impact sur problèmes de convergence ?



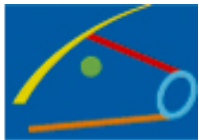
- **Problèmes de convergence pour analyses LD-LA ?**
  - avec ai-reml ou em-reml
  - parfois des problèmes de convergence si « petits » jeux de données ou caractère faiblement héritable
  - comment faire pour avoir moins de problèmes de convergence ?
    - Conserver des taureaux avec un seuil mini. sur edc (ex : >20)
    - Imposer une structure familiale comme pour régressions intra-pères ? Nb de fils minimum par père (ex: >5, >10, >15)
- ⇒ Cela semble donner de meilleurs résultats (constat lors du stage de Chris Hoze)
- Conséquences pratiques : hétérogénéité de situations par caractère nécessite la création de jeux de données spécifiques de groupe de caractères ayant les mêmes caractéristiques.



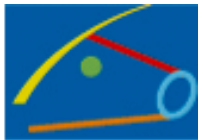
## ▪ **Autres règles classiquement utilisées**

(en dehors évaluations de routine)

- Pour analyses LD-LA, on conserve uniquement SNP avec  $MAF > 3\%$  => nécessite 2 passages de Verif\_typ pour déterminer les MAF des SNP => adapter les cartes et les fichiers de typages aux SNP conservés (en général entre 35000 et 40000 SNP).
- Verif\_typ : on conserve animaux reconstitués sur au moins 75% des SNP
- Dagphase (linkphase+beagle) pour construction de phases et imputations : dag créé avec familles de 5 fils ou plus.

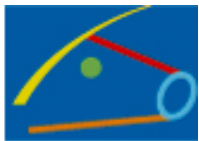
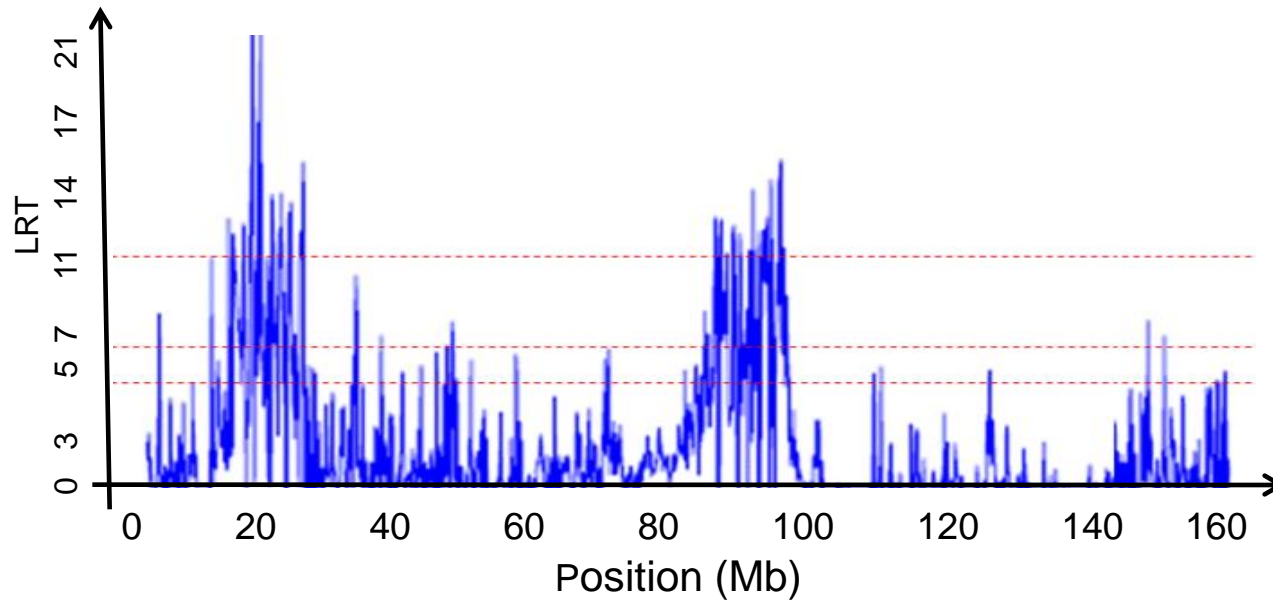


## 3 – Questions sur résultats des analyses LD-LA

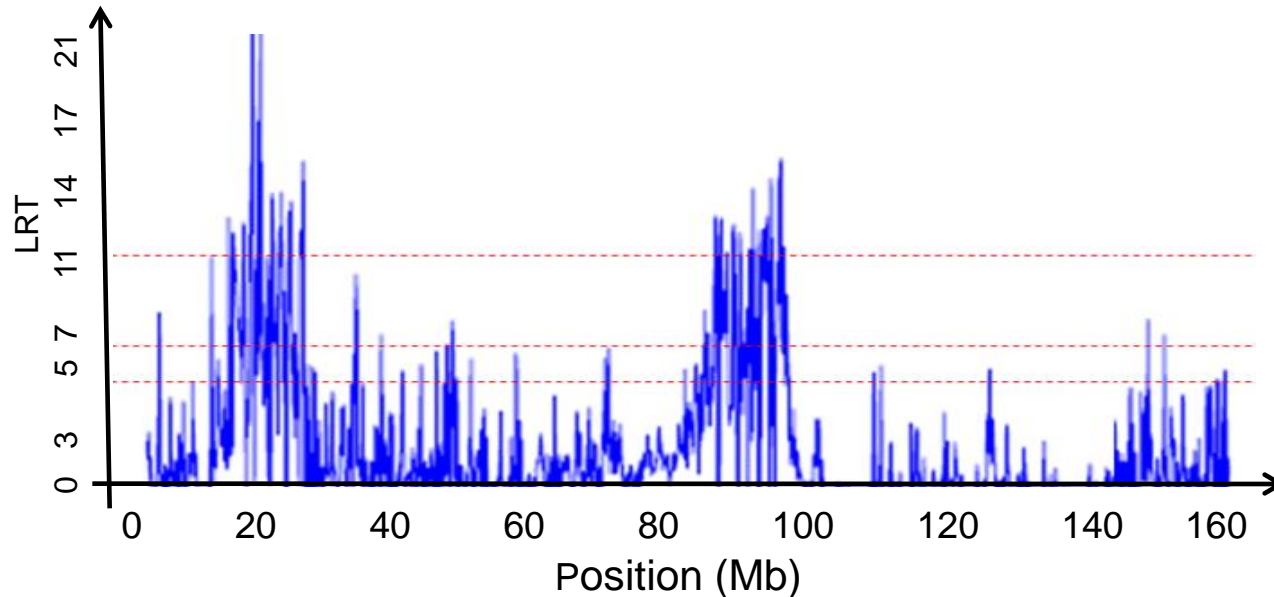




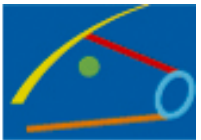
- Exemple : Race Montbéliarde, BTA1, FNAI



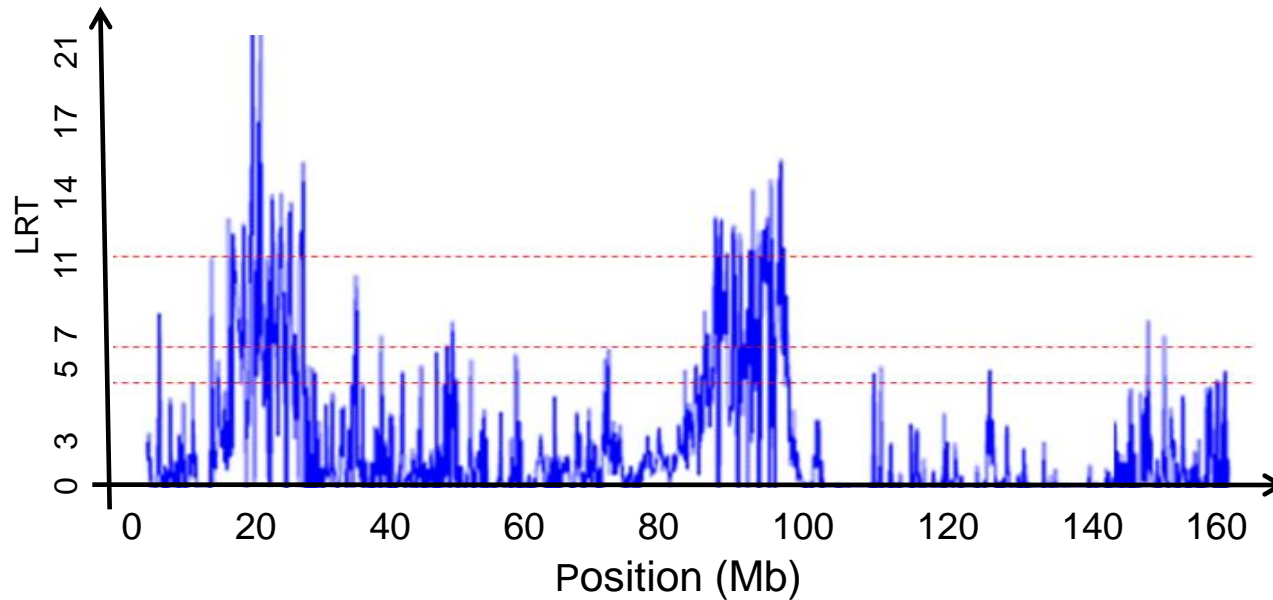
# Qu'est-ce qu'un QTL détecté ?



- **Plusieurs points ont des statistiques de tests proches dans une même région...**
- **En pratique, on retient un seul pic dans un intervalle de 4 Mb à la position du maximum de la statistique de test. Quelle serait la meilleure façon de faire ?**

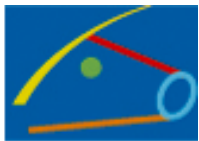


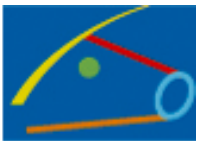
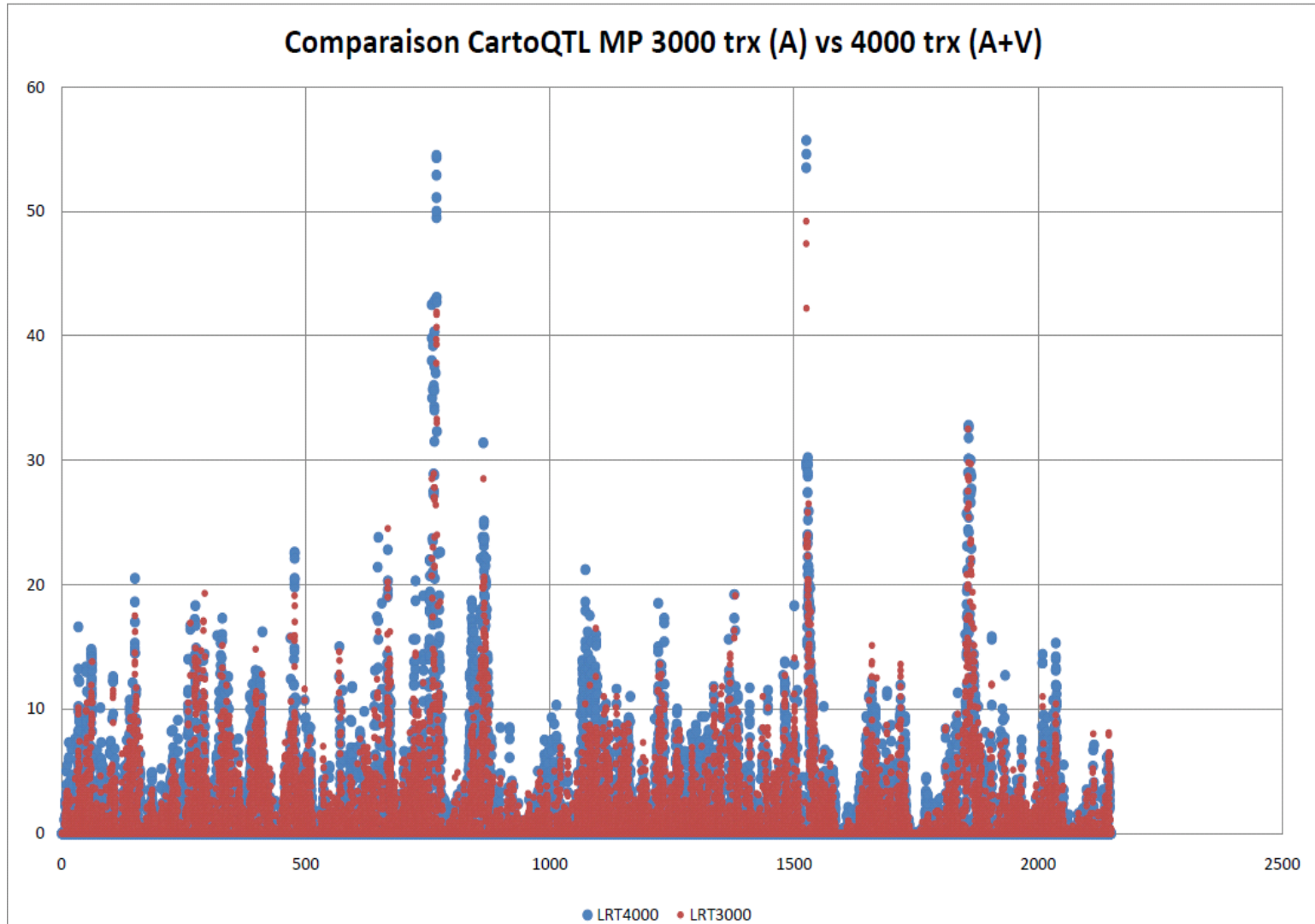
- Exemple : Race Montbéliarde, BTA1, FNAI



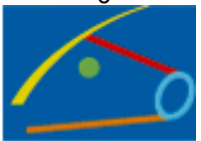
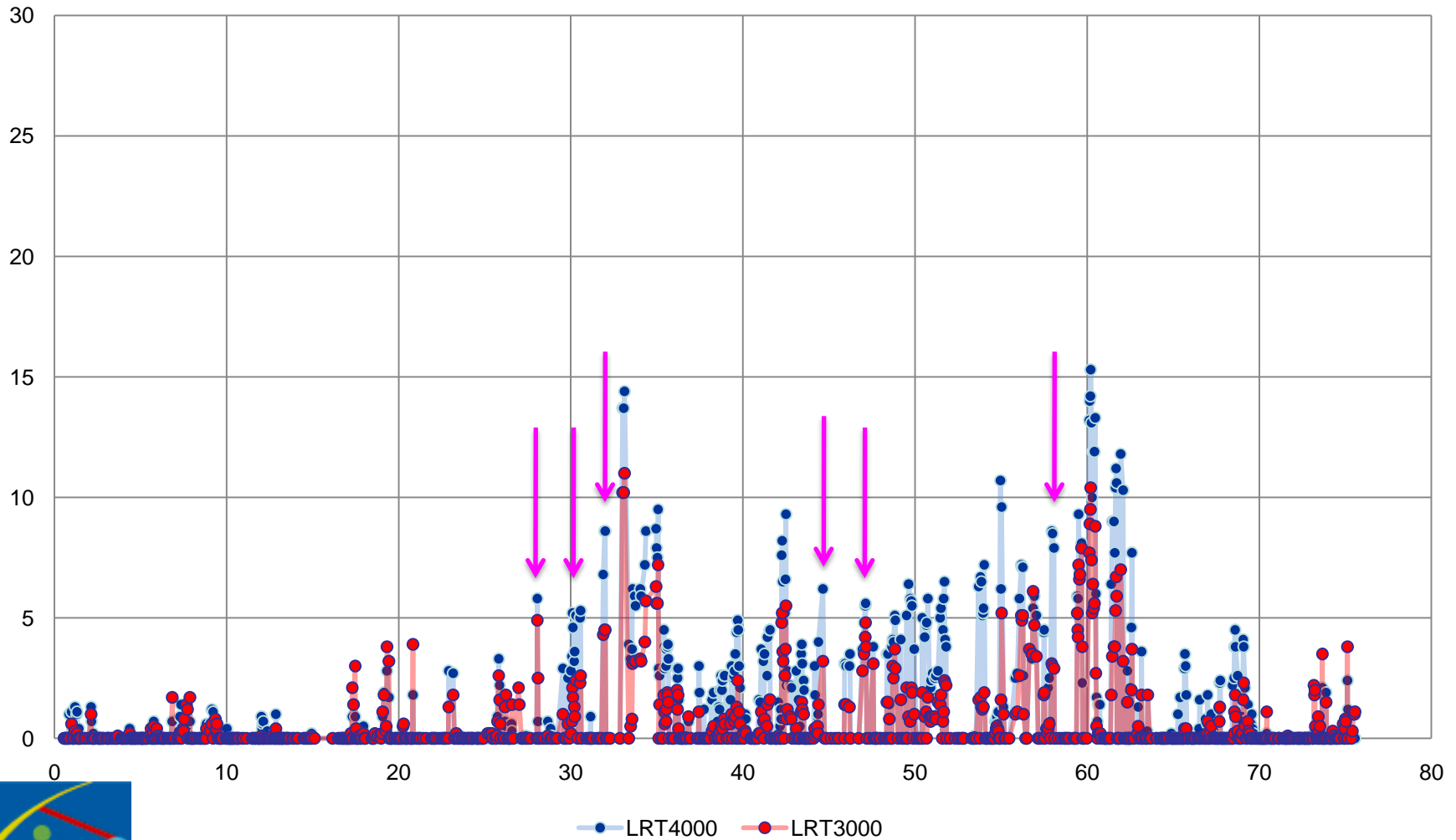
Exemple de FNAI en  
race Montbéliarde :

Nombre de QTL dont le LRT		
> 5	> 7	> 11
243	212	162





## MP - BTA20



- **Cartes génétiques ?**
- **Problèmes de convergence ?**
  - Impact des poids des « perf » ?
  - Doit-on conserver une structure familiale dans analyses LD-LA ? Si oui, avec quelles règles ?
- **Comment définir un « QTL » ?**
- **Comment calculer les seuils pour analyses LD-LA ?**  
*(Doit-on en calculer ?)*
- **Comment réaliser des analyses LD-LA multi-QTL ?**  
**Comment comparer uni-QTL vs multi-QTL ?**

