



INRAE



Méta-analyses d'association à partir des séquences complètes du génomme bovin : application aux caractères bouchers

Webinaire

16 juin 2022

Introduction

Méta-analyses d'association = combinaison des résultats de plusieurs analyses d'association (« GWAS ») réalisées sur plusieurs jeux de données indépendants → « *résultats enrichis* »

1. Principe et limites des Analyses d'Association
2. Principe et intérêt des Méta-Analyses d'Association
3. Quelques exemples et résultats de Méta-Analyses d'Association



Principe des Analyses d'Association « GWAS »

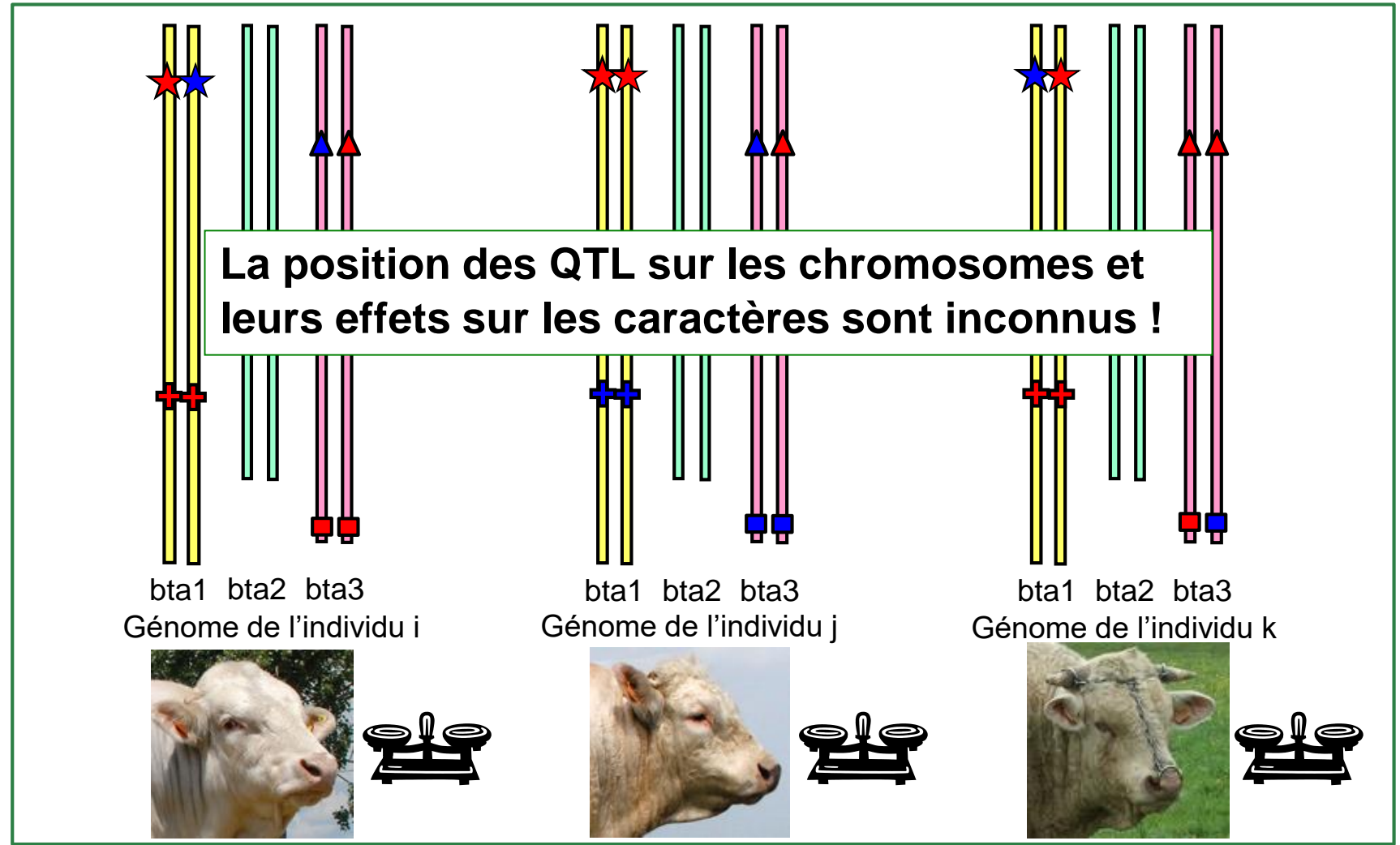
Variations dans la séquence des individus

Si effet sur performances : QTL

Position des QTL inconnue

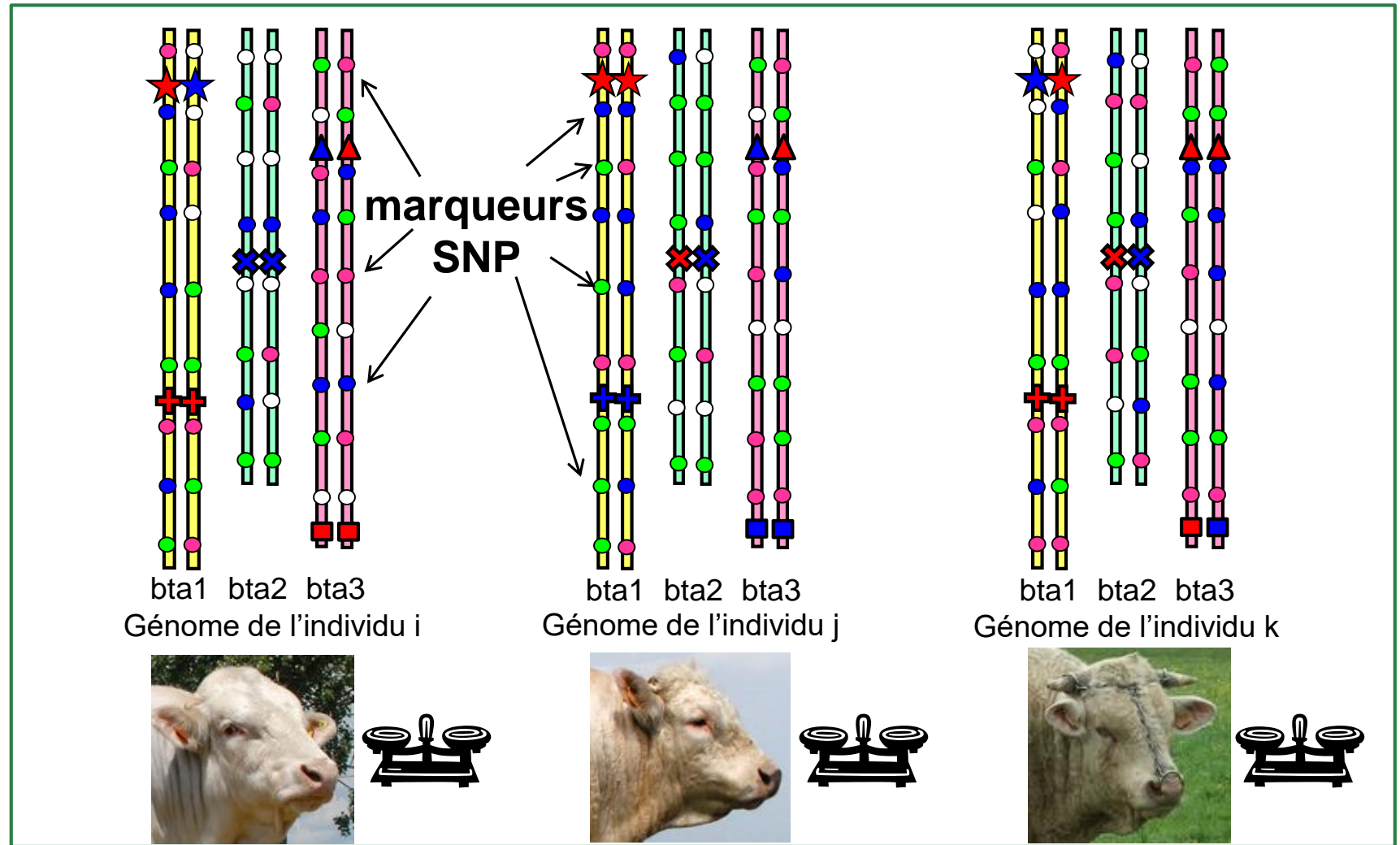
→ On veut les localiser pour :

- Connaitre le déterminisme génétique des caractères
- Mieux sélectionner les individus
- Améliorer l'efficacité de l'évaluation génomique



Principe des Analyses d'Association « GWAS »

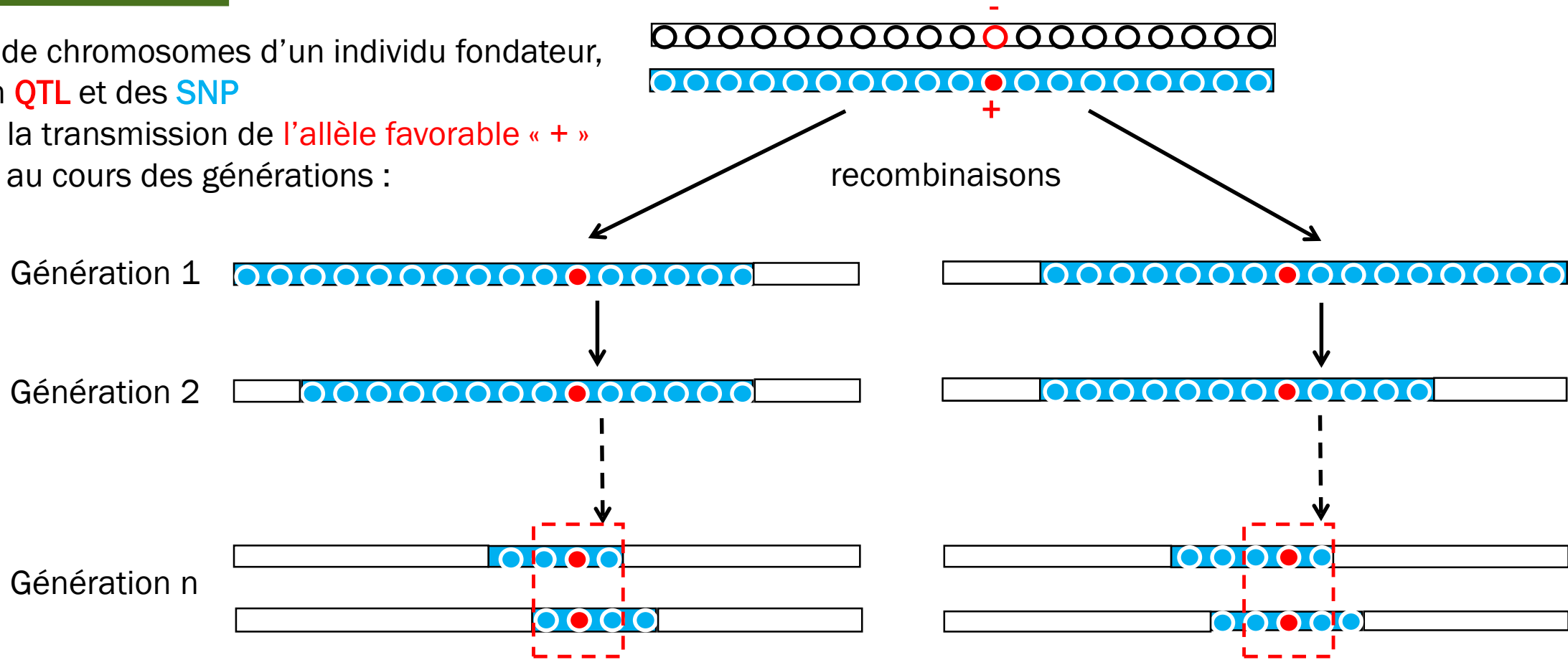
On génotype les individus pour un grand nombre de marqueurs SNP (2 allèles par marqueur) répartis sur tout le génome



Déséquilibre de liaison entre marqueurs et QTL ?

Une paire de chromosomes d'un individu fondateur,
portant un **QTL** et des **SNP**

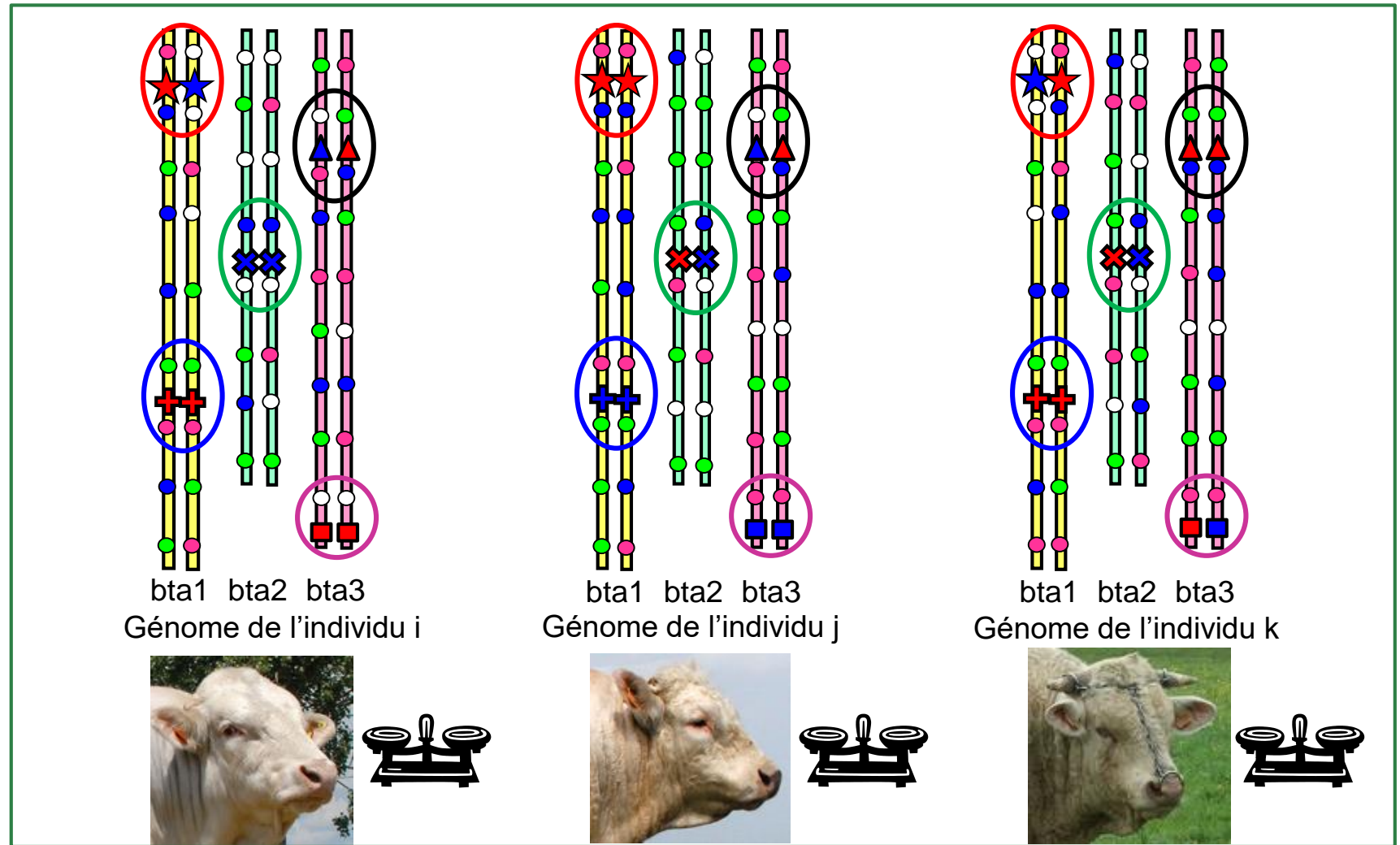
→ On suit la transmission de l'allèle favorable « + »
du **QTL** au cours des générations :



Fragment du chromosome ancestral conservé dans la descendance :
le QTL est en « **déséquilibre de liaison** » avec les SNP de ce segment chromosomique
= **association statistique** entre l'allèle au QTL et les allèles aux SNP du segment

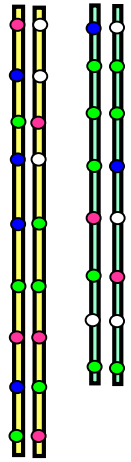
Principe des Analyses d'Association « GWAS »

Si on a suffisamment de SNP bien répartis sur tous les chromosomes, on est à peu près sûr d'avoir des SNP en déséquilibre de liaison avec les QTL ...



Principe des Analyses d'Association « GWAS »

- Pour chaque marqueur SNP j , on estime son effet α_j sur la performance avec le modèle suivant :



$$y_i = \mu + u_i + g_{ij}\alpha_j + e_{ij}$$

y_i : Performance de l'animal i

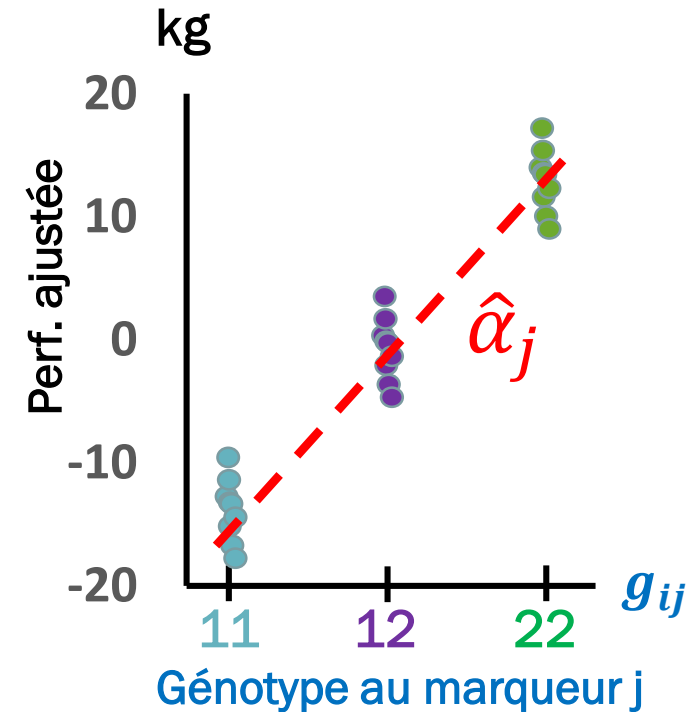
μ : moyenne

u_i : Valeur génétique additive de l'animal i non expliquée par le marqueur j

g_{ij} : Génotype l'animal i au marqueur j

α_j : Effet du marqueur j

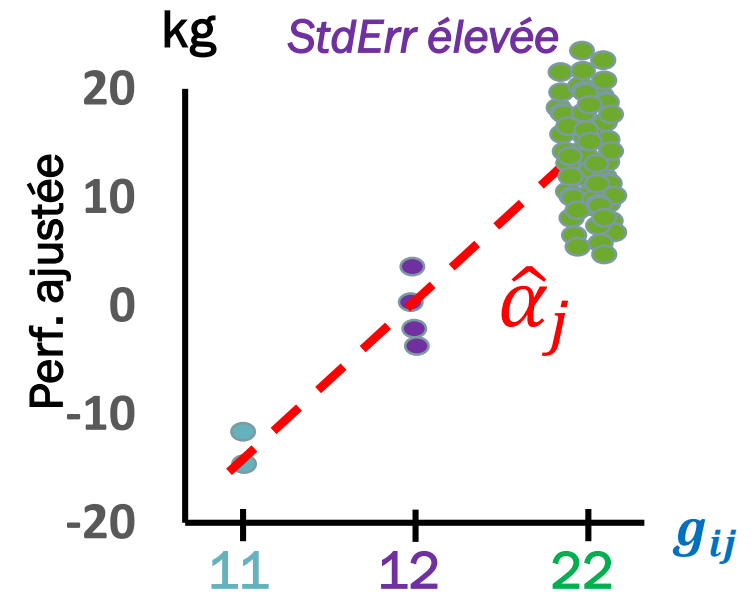
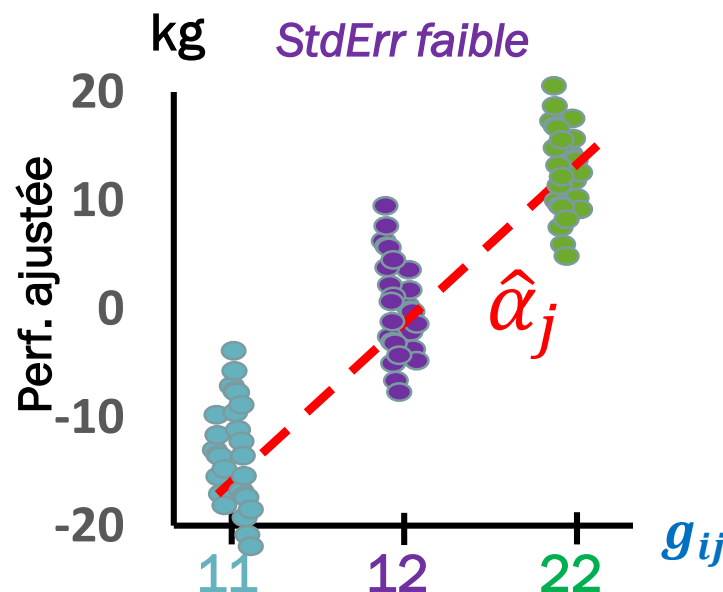
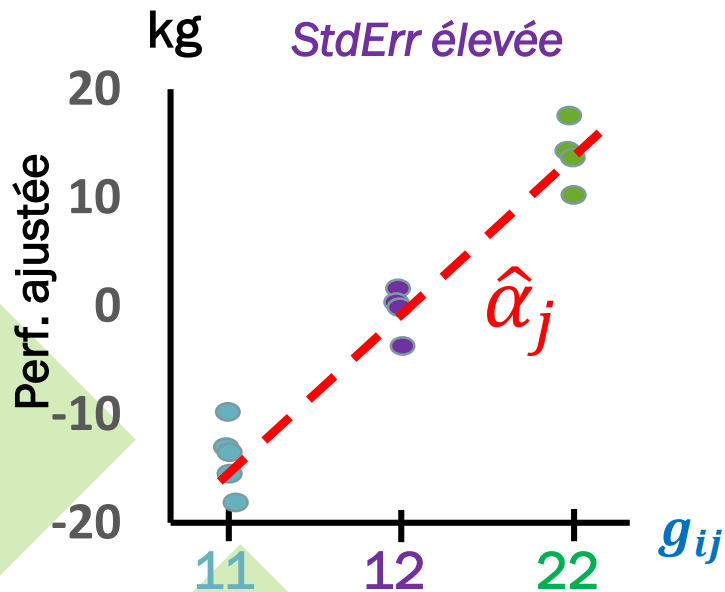
e_{ij} : Résidu non expliqué par le modèle



Effet $\hat{\alpha}_j$ d'un allèle 2 = +15kg

Principe des Analyses d'Association « GWAS »

- L'effet α_j d'un SNP est estimé avec une certaine erreur **StdErr** qui dépend :
 - De la taille du dispositif (nb individus génotypés avec performance)
 - De la fréquence de l'allèle le plus rare au SNP (MAF)



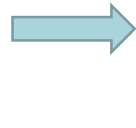
Ex : Caractère difficile ou couteux à mesurer

Allèle 1 du marqueur rare dans la population

Significativité de l'effet estimé pour le marqueur ?



Effet estimé $\hat{\alpha}_j$
StdErr_j

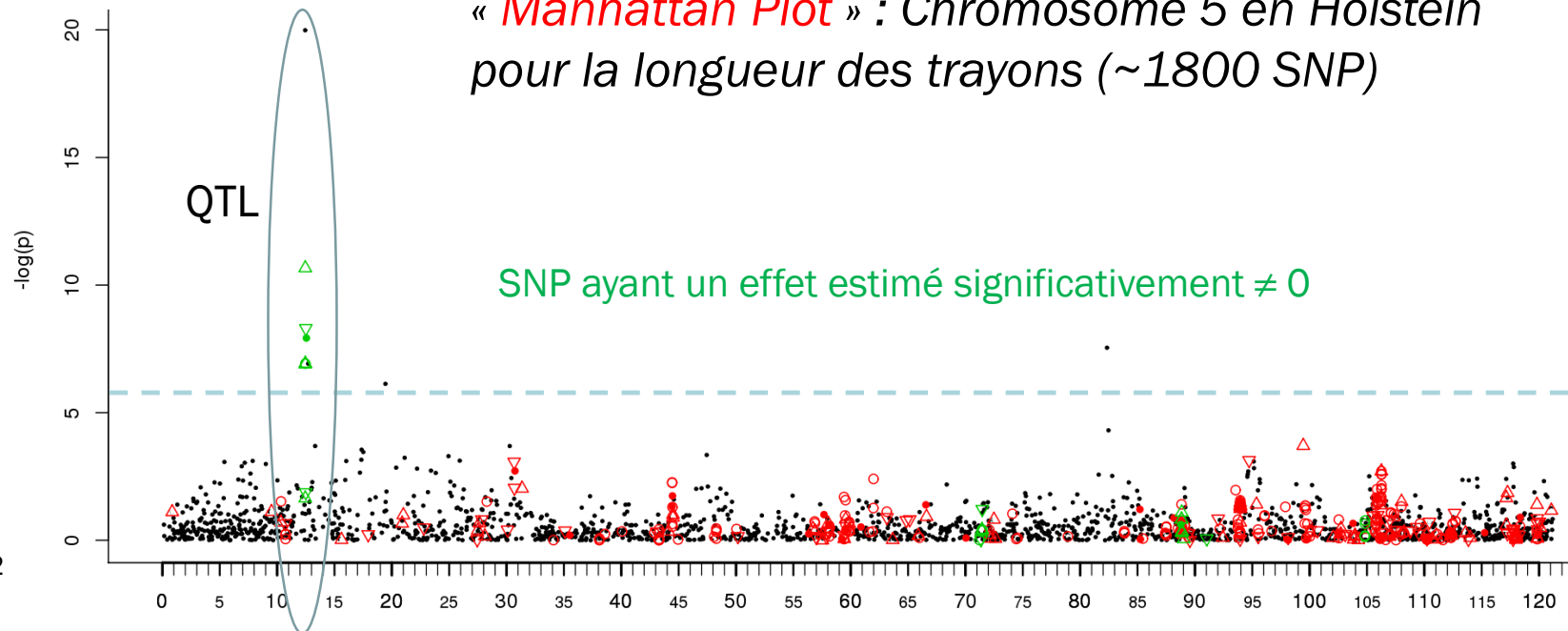
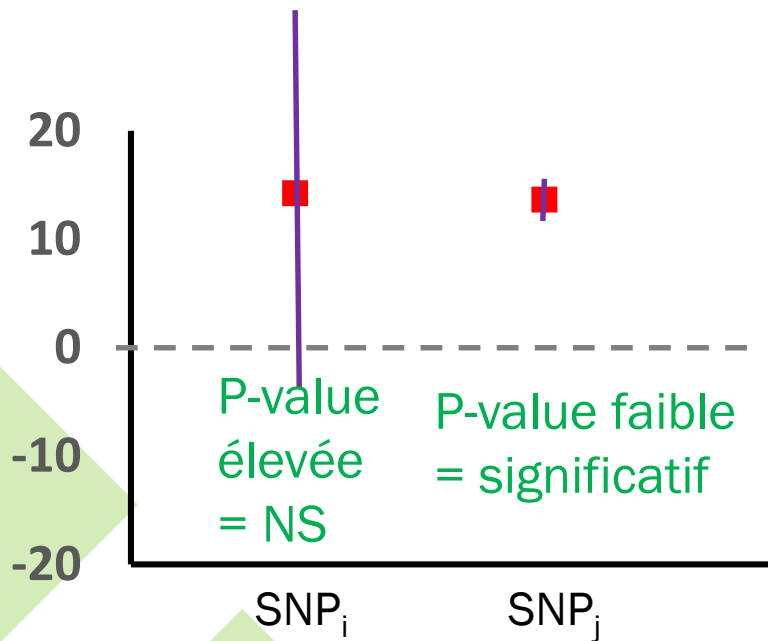


P-value = probabilité de se tromper en concluant que le SNP a un effet non nul sur les performances du caractère

Par commodité, on utilise $-\text{LOG}(\text{P-value})$

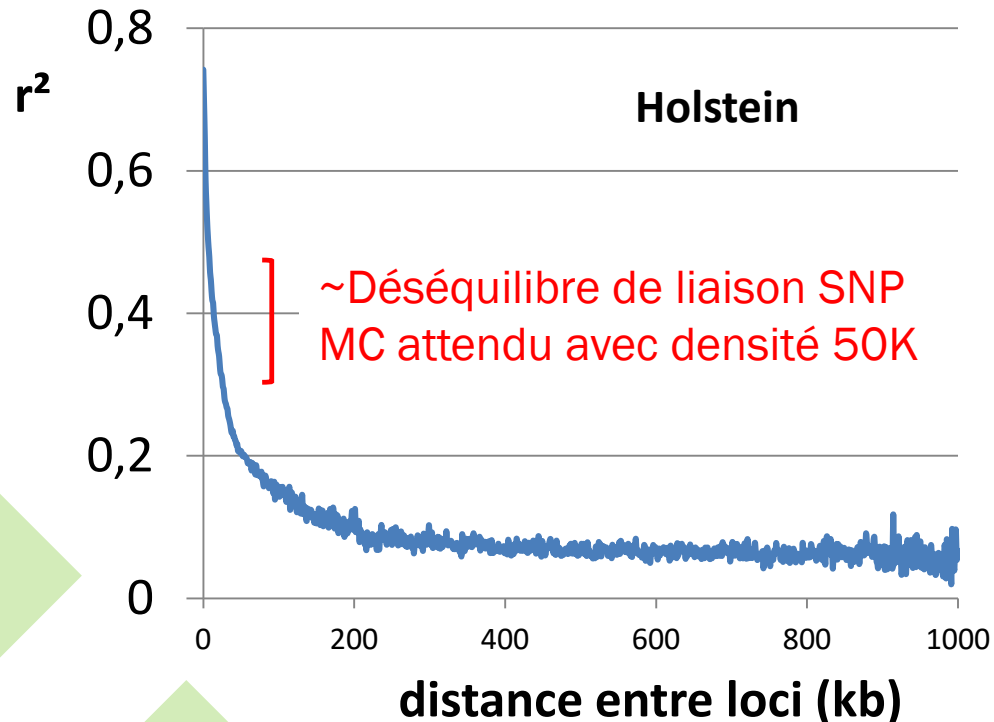
→ On cherche des marqueurs dont $-\text{LOG}(\text{P-value})$ très élevé

→ On calcule $\hat{\alpha}$ StdErr_j $-\text{LOG}(\text{P-value})$ pour chaque SNP



Densité de marqueurs ?

Peu de SNP → distance moyenne entre SNP et mutation causale (MC) élevée
 → Déséquilibre de liaison (r^2) SNP – mutation causale faible



D'après Hoze, 2013

Effet *apparent* d'un SNP = $\sqrt{r^2}$ * effet de la mutation causale

Faible densité de marqueurs :

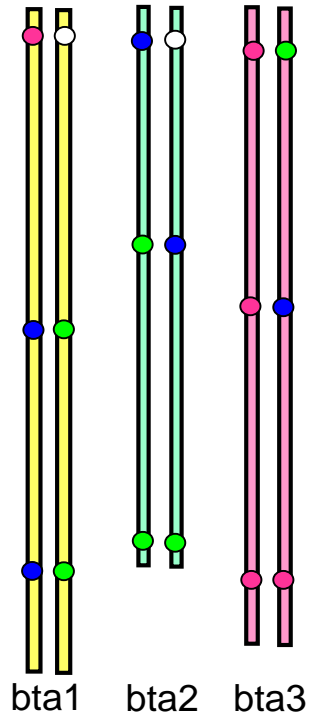
- moins de chance de détecter des SNP avec effet estimé significativement $\neq 0$
- faible précision de localisation de la mutation causale ...

Dans l'idéal, si on avait accès à la séquence de tous les individus du dispositif, on aurait le génotype à toutes les mutations causale ...

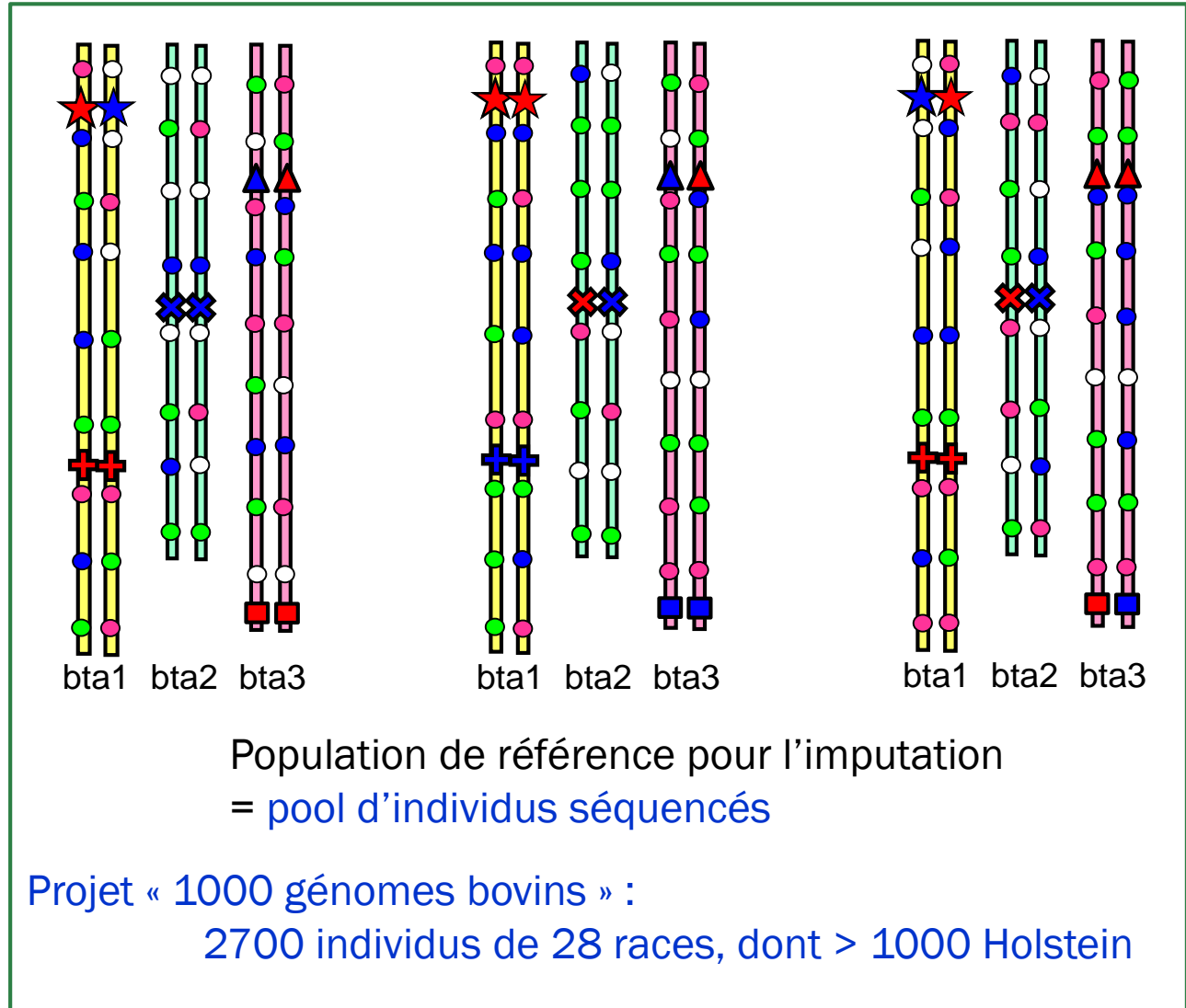
MAIS : la plupart des individus sont typés à basse ou moyenne densité ...

SOLUTION : l'imputation de génotypes !

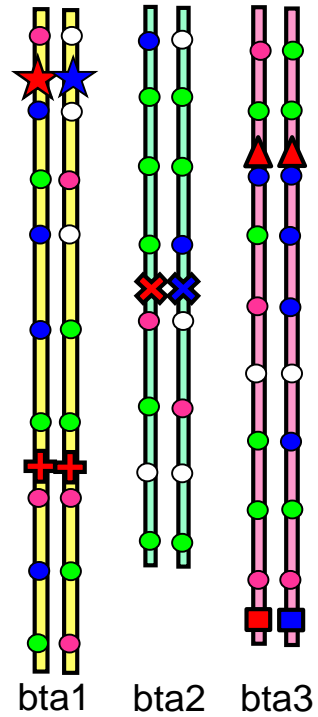
L'imputation de génotypes ?



individu génotypé en
basse densité

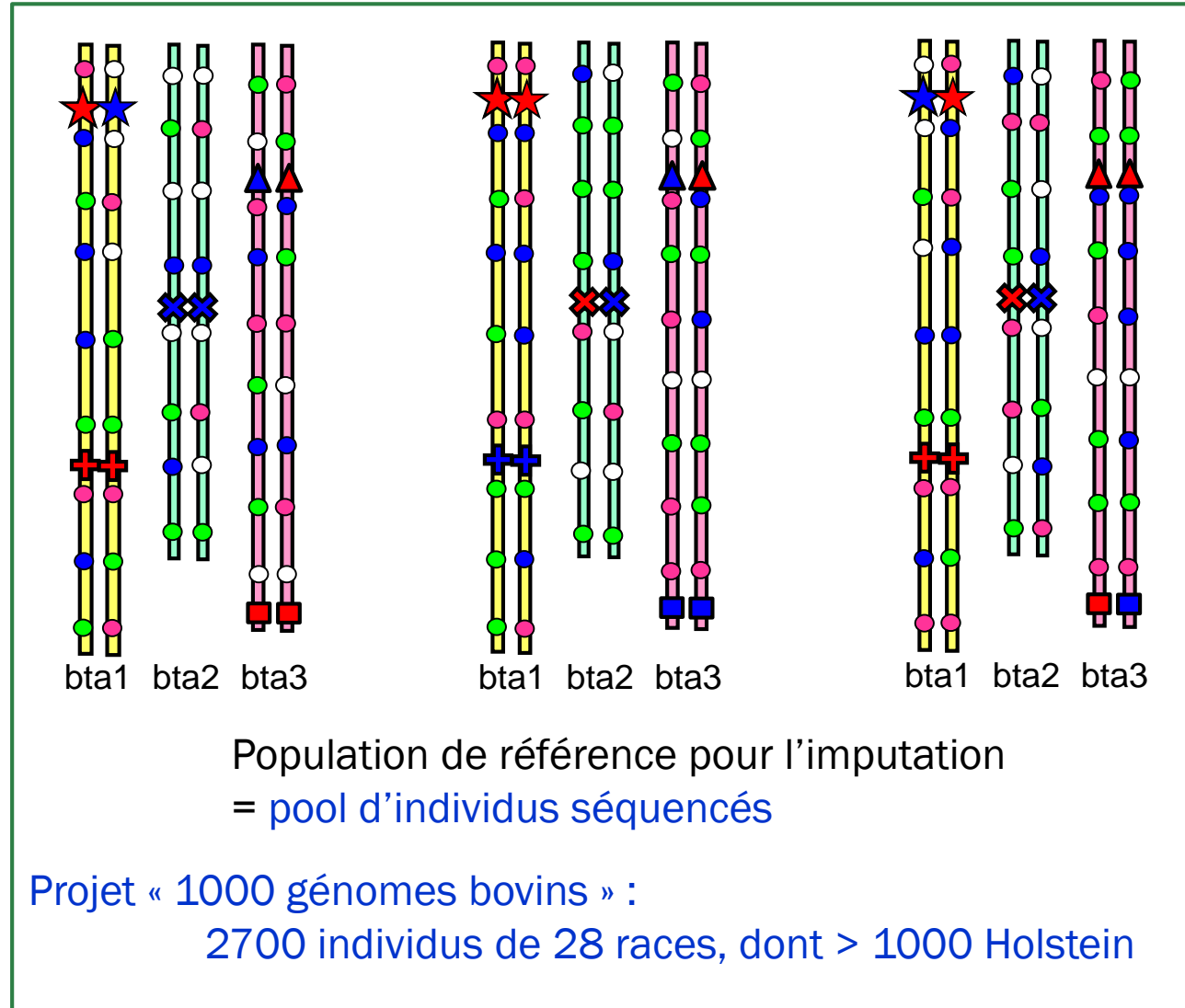


L'imputation de génotypes ?



individu génotypé en
basse densité

On reconstitue les génotypes manquants avec une (très) bonne précision selon MAF



Population de référence pour l'imputation
= pool d'individus séquencés

Projet « 1000 génomes bovins » :

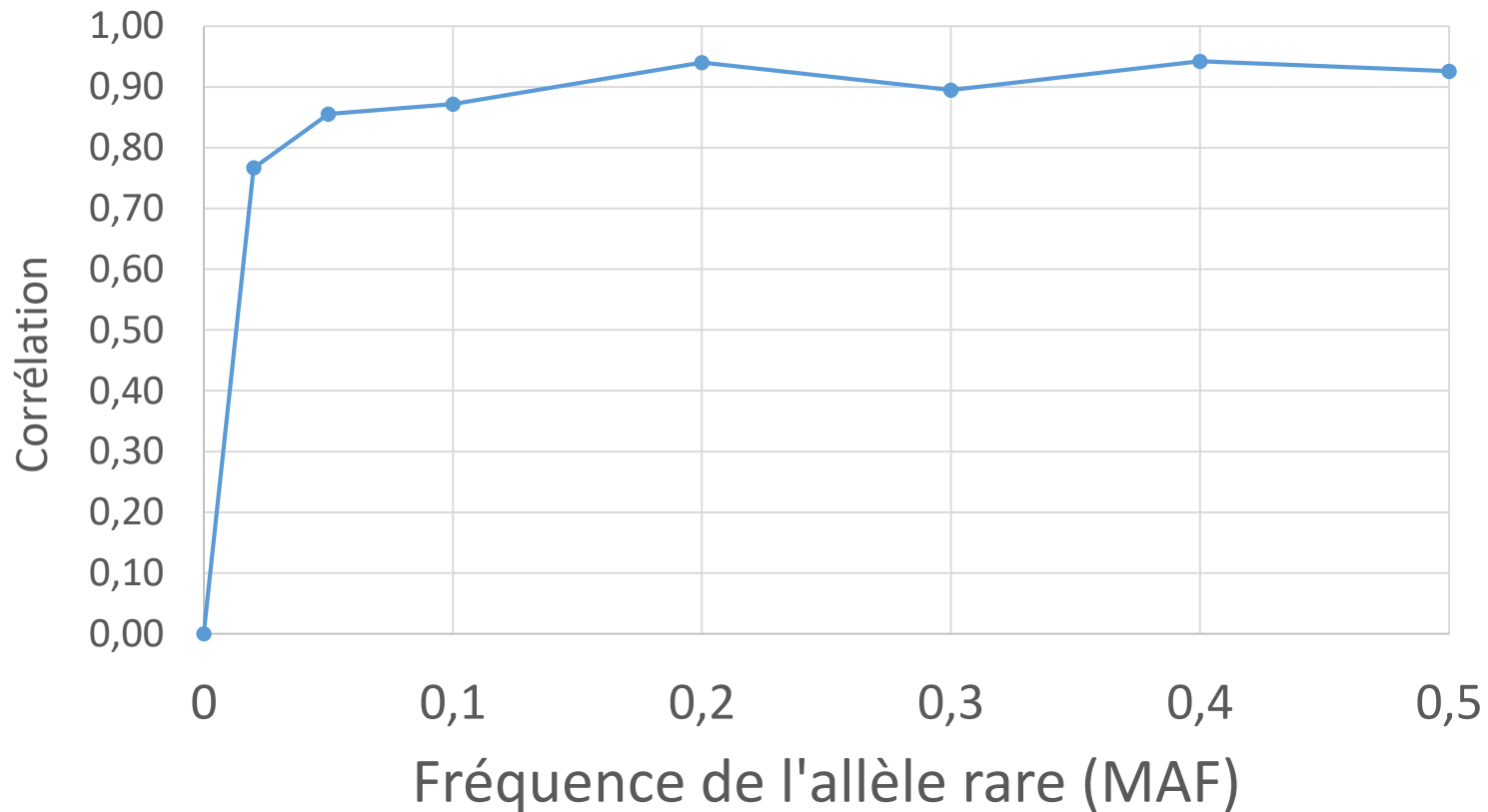
2700 individus de 28 races, dont > 1000 Holstein

Meilleurs résultats si 2 étapes :
(1) BD → HD (2) HD → SEQUENCE

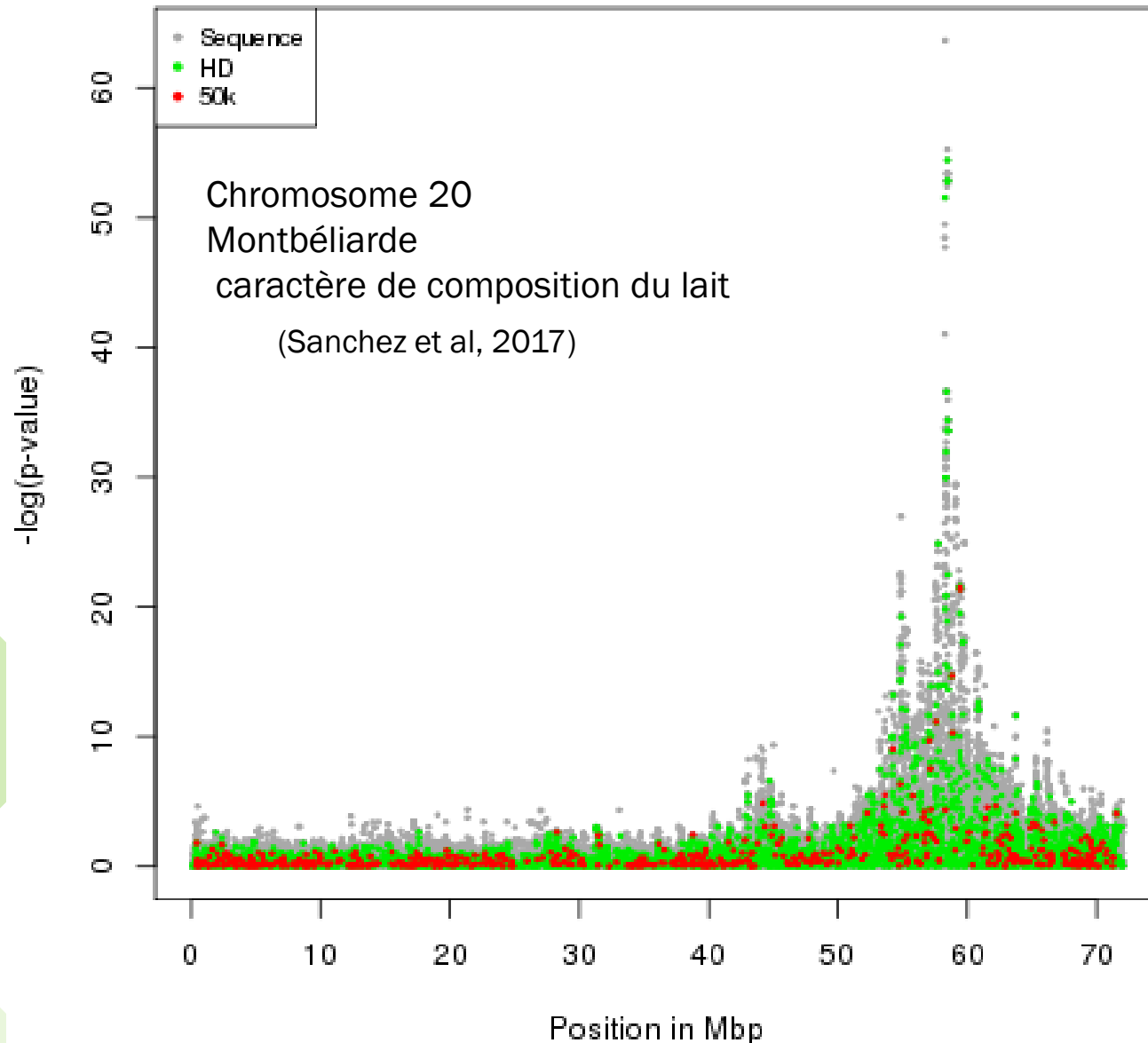
Précision de l'imputation de génotypes à la séquence

Précision de l'imputation

(correlation entre génotype vrai et génotype imputé)



Manhattan plots de GWAS 50K / 800K / Séquence



En augmentant la densité de marqueurs :

- Pics plus élevés = résultats plus significatifs *mutations causales ou SNP en plus fort DL avec mutation causales*
- *Intervalle de localisation des QTL plus fins*
- De nouveaux pics apparaissent = QTL supplémentaires

Malgré cela, en général, insuffisant pour identifier toutes les régions QTL et les mutations causales

- Nombre d'individus trop faible
- Allèle rare trop peu fréquent dans la population d'étude (fixation ?)
- Erreurs d'imputation → perte efficacité
- DL élevé à grande distance → intervalle de confiance très large

Une Solution : les Méta-Analyses d'Association ?



Solution pour dépasser (en partie) ces limites : augmenter la taille du dispositif et recommencer les GWAS

→ réunir les performances et les génotypes pour un même caractère de plusieurs populations d'études différentes :

Ex : Holstein Français, Holstein Allemand, Holstein Pays Bas

→ plus d'individus → **StdErr plus faible** → plus de chance de trouver des effets significatifs

Ex' : Holstein, Montbéliarde, Normande :

les différentes races ont une origine commune +/- lointaine → QTL communs

→ plus d'individus → **StdErr plus faible** → plus de chance de trouver des effets significatifs

Si les populations ont divergé il y a longtemps, nombreuses recombinaisons

→ haplotypes communs plus courts qu'intra race ? Associations SNP - mutations causales ≠ ?

→ pourrait permettre d'affiner la localisation des QTL ?

Une Solution : les Méta-Analyses d'Association ?



Plusieurs problèmes se posent:

- Regrouper plusieurs gros dispositifs → GWAS sur séquence trop longues, voire impossible à réaliser ?
- Les structures propriétaires des données peuvent ne pas vouloir partager leurs performances et leur génotypes (données coûteuses à recueillir, stratégiques pour sélection, ...)

→ Mettre **uniquement** en commun, pour chaque marqueur considéré dans les GWAS de chaque étude :

- Son **effet estimé**
- Son erreur d'estimation **StdErr** (la **P-Value** associée)

→ Combinaison pour déterminer la **P-value** de chaque marqueur **en combinant les données de tous les dispositifs**

AVANTAGES :

- Pas de partage de données élémentaires stratégiques
- Possibilité d'améliorer les résultats des GWAS intra-étude sans coût supplémentaire
- Création de relations de collaboration entre structures → terreau pour futures collaborations de recherche

Les Méta-Analyses d'Association

nature
genetics

LETTERS

<https://doi.org/10.1038/s41588-018-0056-5>

Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals

Aniek C. Bouwman¹, Hans D. Daetwyler^{2,3}, Amanda J. Chamberlain², Carla Hurtado Ponce^{2,4}, Mehdi Sargolzaei^{5,6}, Flavio S. Schenkel⁵, Goutam Sahana⁷, Armelle Govignon-Gion⁸, Simon Boitard⁹, Marlies Dolezal¹⁰, Hubert Pausch^{2,11,12}, Rasmus F. Brøndum⁷, Phil J. Bowman², Bo Thomsen⁹, Bernt Guldbbrandtsen⁷, Mogens S. Lund⁷, Bertrand Servin¹³, Dorian J. Garrick¹⁴, James Reecy¹⁴, Johanna Vilkki¹⁵, Alessandro Bagnato¹⁶, Min Wang^{2,3}, Jesse L. Hoff¹⁷, Robert D. Schnabel¹⁷, Jeremy F. Taylor¹⁷, Anna A. E. Vinkhuyzen^{18,19}, Frank Panitz⁹, Christian Bendixen⁹, Lars-Erik Holm⁹, Birgit Gredler²⁰, Chris Hozé^{8,21}, Mekki Boussaha⁸, Marie-Pierre Sanchez⁸, Dominique Rocha⁸, Aurelien Capitan^{8,21}, Thierry Tribout⁸, Anne Barbat⁸, Pascal Croiseau⁸, Cord Drögemüller²², Vidhya Jagannathan²², Christy Vander Jagt², John J. Crowley²³, Anna Bieber²⁴, Deirdre C. Purfield²⁵, Donagh P. Berry²⁵, Reiner Emmerling²⁶, Kay-Uwe Götz²⁶, Mirjam Frischknecht²⁰, Ingolf Russ²⁷, Johann Sölkner²⁸, Curtis P. Van Tassell²⁹, Ruedi Fries¹¹, Paul Stothard³⁰, Roel F. Veerkamp¹, Didier Boichard⁸, Mike E. Goddard^{2,4} and

NATURE GENETICS | www.nature.com/naturegenetics

1^{ère} grosse méta-analyse d'association dans le monde des animaux d'élevage (2018)

17 populations, 58 265 individus
25 millions de variants

→ 163 régions QTL

→ Collaboration avec des chercheurs d'équipes de très haut niveau ...

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

Les Méta-Analyses d'Association

van den Berg et al. *Genet Sel Evol* (2020) 52:37
<https://doi.org/10.1186/s12711-020-00556-4>




RESEARCH ARTICLE

Open Access



Meta-analysis for milk fat and protein percentage using imputed sequence variant genotypes in 94,321 cattle from eight cattle breeds

8 races, 94 321 individus,
Séquences imputées

Irene van den Berg^{1*} , Ruidong Xiang^{1,2}, Janez Jenko³, Hubert Pausch⁴, Mekki Boussaha⁵, Chris Schrooten⁶, Thierry Tribout⁵, Arne B. Gjuvsland³, Didier Boichard⁵, Øyvind Nordbø³, Marie-Pierre Sanchez⁵ and Mike E. Goddard^{1,2}

Les Méta-Analyses d'Association

LIMITES / DIFFICULTES :

Le caractère considéré doit être le même dans toutes les études dont on combine les résultats
caractères différents ↔ déterminismes génétiques (QTL) différents

- Unités de mesure différentes : ex kg vs livres → facile à corriger par standardisation des effets estimés et des StdErr
- Âges de mesure différents : ex Poids à âge type
- Conditions d'élevage différentes : ex alimentation ad libitum vs alimentation rationnée, bâtiment vs plein air,
- Définition du caractère : ex Poids à âge type vs GMQ ; mammites cliniques vs nb cellules somatiques ; nombre d'IA avant IA fécondante vs intervalle entre mise bas ; ...

→ Différents modèles existent pour Méta-Analyses d'Association : hypothèses +/- restrictives, +/- conservatives

Ex : utilisation de l'effet estimé (pour caractères identiques) vs utilisation du signe de l'effet (caractères proches)

Le projet européen BOVREG

Projet H2020 – 2019 - 2023
 Coordonné par FBN Christa KUHN &
 INRAE Dominique ROCHA

« Caractérisation des éléments de
 régulation fonctionnelle du
 génome en lien avec la diversité et
 la plasticité chez le bovin »







20 partenaires de 14 pays



Le projet européen BOVREG

WP4 / T4.1 : GWAS et méta-analyses à partir des séquences complètes pour des caractères d'efficacité, de résistance aux maladies et de fertilité

INRAE – GABI, équipe G2B : Didier BOICHARD, Pascal CROISEAU, Marie-Pierre SANCHEZ, Thierry TRIBOUT

<p>GWAS intra-population</p>						
<p>4 groupes de caractères</p>	<p>Résistance aux mammites</p>	<p>Production laitière & fertilité</p>	<p>Efficacité alimentaire</p>	<p>Production de viande</p>		
<p>Méta-analyses</p>						
	<p>8-13 populations < 120 000 anim.</p>	<p>7-12 populations < 125 000 anim.</p>	<p>3-9 populations < 13 000 anim.</p>	<p>3-10 populations < 25 000 anim.</p>		

Taureaux, vaches, jeunes bovins
Races pures laitières ou à viande, animaux croisés, lignées composites

Les Méta-Analyses (MA) d'Association du projet BOVREG



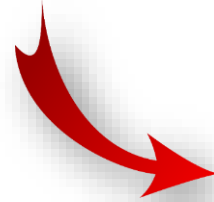
Imputations

2 étapes

50k → 777k (HD)
HD → Séquence

Animaux de la race génotypés HD
Animaux séquencés div. races (projet 1000G)

Fimpute / Minimac / Beagle



~ 20 millions de variants
dont les variants causaux



GWAS

GCTA

Mise en forme & envoi des
résultats aux partenaires
responsables des méta-analyses



Chr	SNP	bp	A1	A2	Freq	b	se	P
0	6:117:C:A	0	C	A	0.0550342	-0.163333	0.0616289	0.00804287
0	6:163:C:T	0	C	T	0.000620877	5.94381	20.5418	0.772312
0	6:203:G:A	0	G	A	0.00492704	0.0381459	2.45103	0.987583
0	6:209:C:T	0	C	T	0.00493752	0.475336	2.68661	0.859565
0	6:215:A:G	0	A	G	0.00539598	0.311183	2.47471	0.899934
0	6:345:G:A	0	G	A	0.0777834	-0.0400829	0.0789062	0.611467
0	6:346:A:G	0	A	G	0.286025	0.0226386	0.0400824	0.572209
0	6:482:AG:A	0	AG	A	0.00617638	5.80794	2.9337	0.0477338

Les MA des caractères bouchers du projet BOVREG

Un grand nombre et une grande diversité de populations

- ☒ 8 populations de 5 races pures françaises (NOR, MON, CHA, LIM, BLA)



- ☒ 4 populations de races pures suisses (BSW, OBR)



ETH zürich

- ☒ 2 populations croisées allemandes (HOL x CHA)



- ☒ 1 lignée composite canadienne (ANG, CHA, beef)



Les MA des caractères bouchers du projet BOVREG

Un grand nombre
et une grande
diversité de
caractères

- Croissance (6)
- Morphologie (6)
- Carcasse (21)

1	Croissance	Birth Weight	BW
2	Croissance	weight at month 15	W15
3	Croissance	weight at 18 months	W18
4	Croissance	weight at 24 months	W24
5	Croissance	average daily gain	ADG
6	Croissance	average daily gain during fattening	ADG
7	Morphologie	muscularity score	MS30
8	Morphologie	skeletal score	SS30
9	Morphologie	thickness of bones	TB30
10	Morphologie	Thighs	THIGHS
11	Morphologie	Wither	WITHER
12	Morphologie	Fat score	FS
13	Carcasse	carcass weight	CW
14	Carcasse	fat coverage	CF
15	Carcasse	meatiness	MT
16	Carcasse	Area of longissimus thoracis	ALT
17	Carcasse	Carcass conformation	CC
18	Carcasse	carcass fat score	FS
19	Carcasse	carcass yield	CY
20	Carcasse	Internal fat weight	IFW
21	Carcasse	length of the leg	LL
22	Carcasse	Rib Eye Area	REA
23	Carcasse	Weight at slaughter	WS
24	Carcasse	Maximum width of the thigh	WT
25	Carcasse	age at slaughter	AS
26	Carcasse	carcass grade	CG
27	Carcasse	average backfat thickness	ABT
28	Carcasse	hot carcass weight	CW
29	Carcasse	lean meat yield	LMY
30	Carcasse	fat content of 6th rib	FC6
31	Carcasse	fat content measured by ultrasound	FCU
32	Carcasse	muscular development	MD
33	Carcasse	skeletal development	SD

Regroupement des caractères dans 16 MA

MA	Type car.	Caractères	# car.	# pop.	# partenaires	# anim.
1	Croissance	W15/W18/ADG	3	7	4	18774
2	Croissance	BW	1	5	2	2720
3	Morphologie	MS30/THIGHS/CC	3	6	2	17418
4	Morphologie	MS30/WITHER/CC	3	6	2	17418
5	Morphologie	LL	1	5	2	3695
6	Morphologie	WT	1	5	2	3695
7	Morphologie	SS30/SD	2	4	2	12140
8	Carcasse	CW	1	7	4	19989
9	Carcasse	AS	1	6	2	12208
10	Carcasse	CY	1	5	2	3694
11	Carcasse	CG/LMY/MT/CC	4	10	5	25367
12	Carcasse	FS/ABT/FC6/FCU/CF	5	8	5	14622
13	Carcasse	WS	1	5	2	2636
14	Carcasse	ALT	1	5	2	3692
15	Carcasse	IFW	1	5	2	3686
16	Carcasse	REA	1	3	2	4453

- 1 à 5 caractères / MA
- 3 à 10 populations / MA
- 2 à 5 partenaires / MA
- 2600 à 20 000 animaux / MA

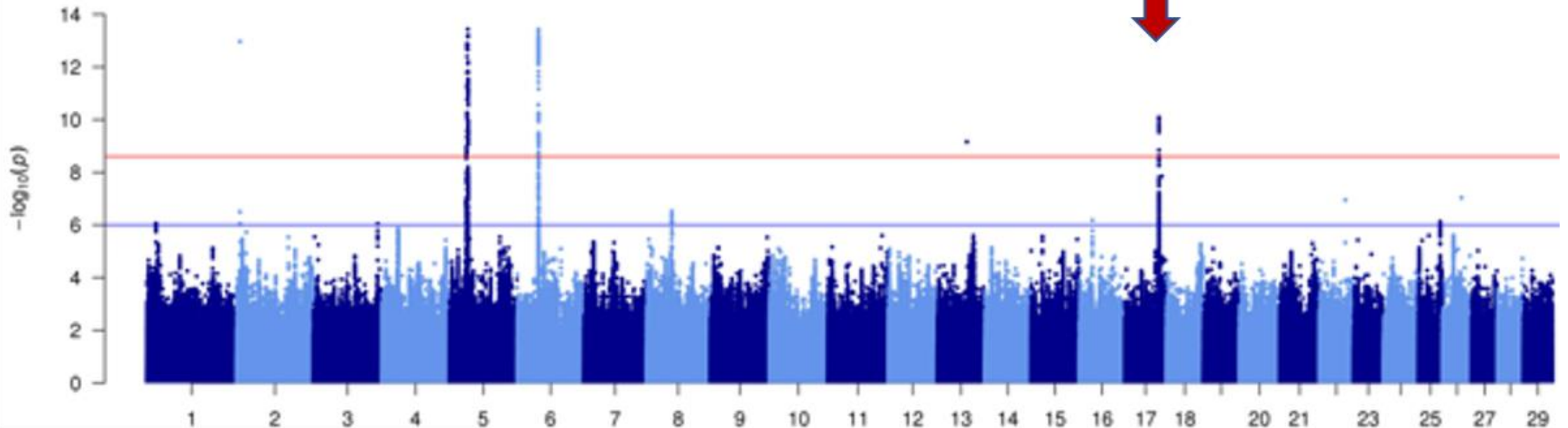
Les MA des caractères bouchers du projet BOVREG

Quelques exemples de résultats ...

QTL **non significatifs intra population** qui deviennent **significatifs en Méta-Analyse**

MA9 carcasse, **nouveau QTL sur chrom 17**

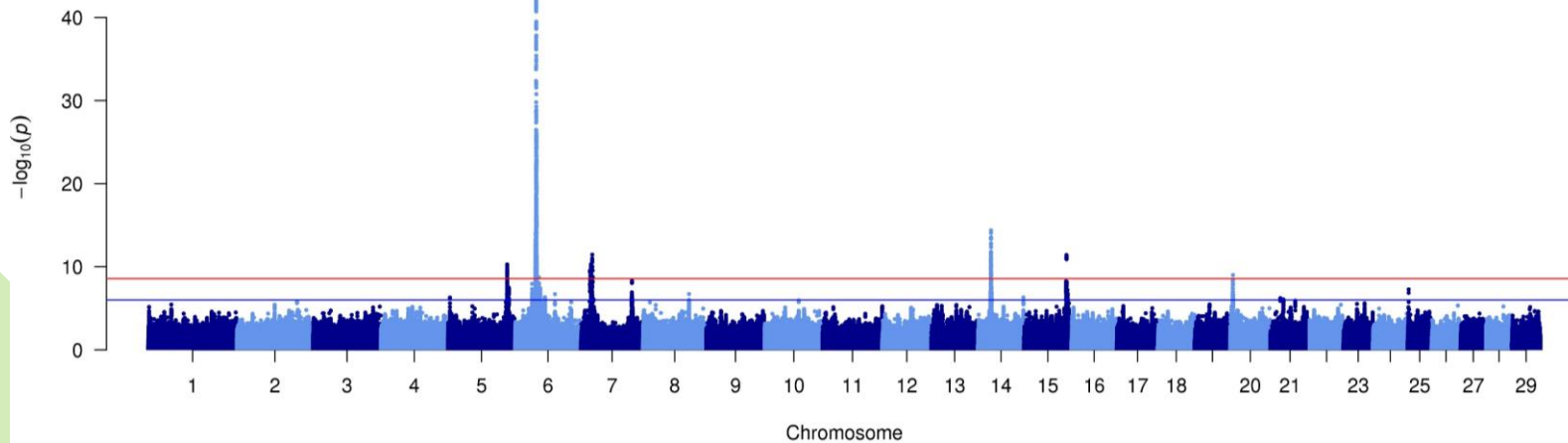
Méthode à effets fixés



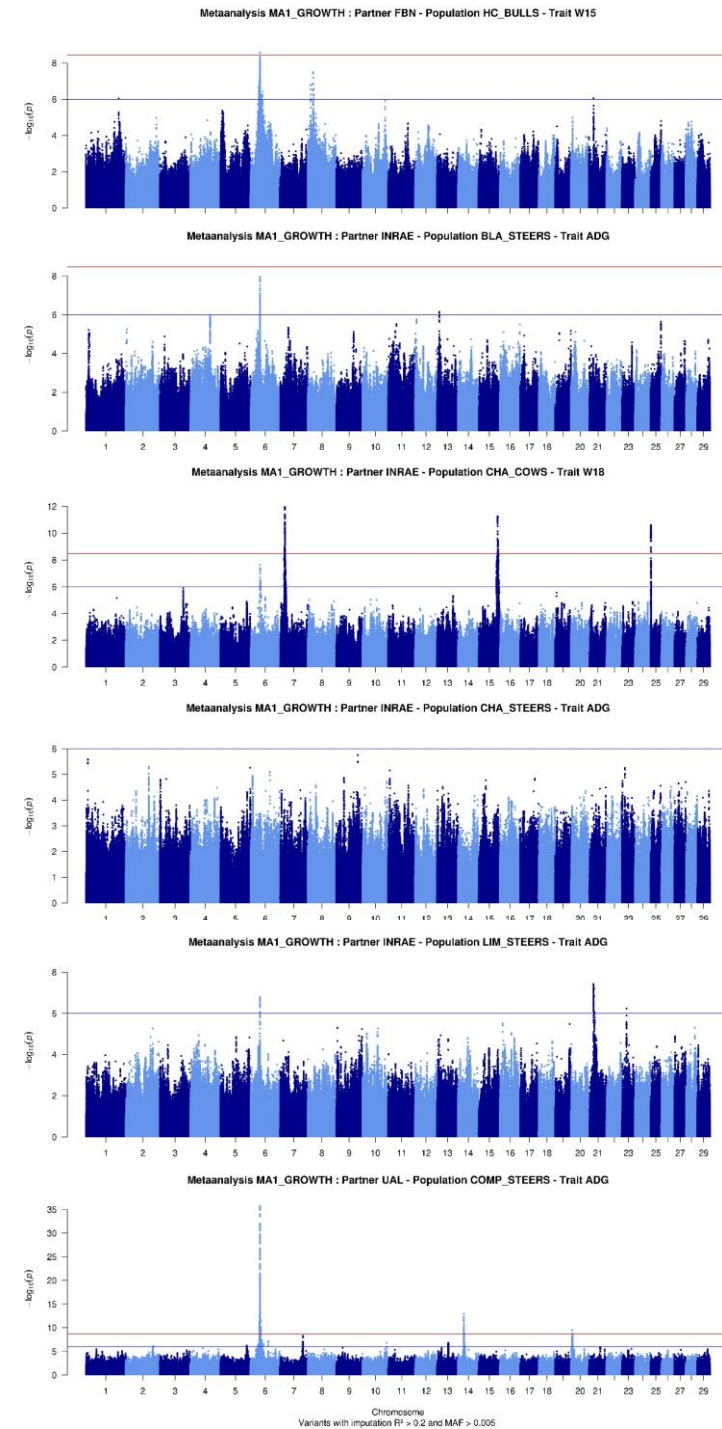
Les MA d'Association du projet BOVREG

- Dans certains cas, la MA semble mieux cibler les variants causaux
Ex: MA1 croissance

MA (effets fixés)



QTL identifié dans 10 des 16 MA, le pic de la MA est fin



Les MA d'Association du projet BOVREG

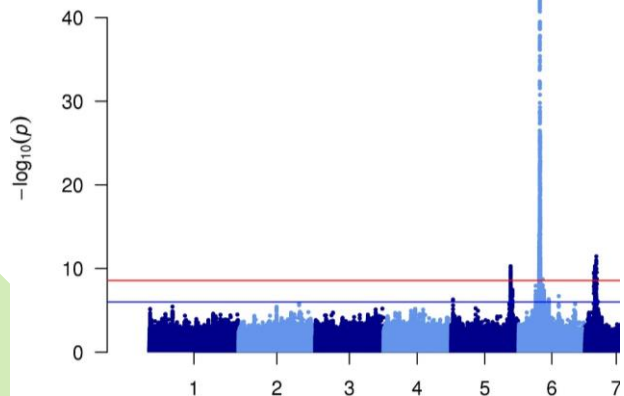
MA / GWAS intra-populations

- Dans certains cas, la MA semble mieux cibler les variants causaux

Ex: MA1 croissance

Dans tous les cas, le variant le plus significatif est dans ou a proximité du gène *LCORL* (\neq GWAS intra-populations)

MA (effets fixés)



MA	Type de caractères	-Log ₁₀ (p) max	IMPACT	NOM_GENE
MA1	Croissance	45.5	intergenic_region	<i>LCORL-SLIT2</i>
MA2	Croissance	22.0	intron_variant	<i>LCORL</i>
MA5	Morphologie	37.1	intron_variant	<i>LCORL</i>
MA6	Morphologie	21.1	intron_variant	<i>LCORL</i>
MA7	Morphologie	18.0	intron_variant	<i>LCORL</i>
MA8	Carcasse	31.4	intergenic_region	<i>LCORL-SLIT2</i>
MA11	Carcasse	11.9	frameshift_variant	<i>LCORL</i>
MA12	Carcasse	14.7	intron_variant	<i>LCORL</i>
MA13	Carcasse	29.2	intron_variant	<i>LCORL</i>
MA16	Carcasse	18.1	missense_variant	<i>LCORL</i>

QTL identifié dans 10 des 16 MA

LCORL => facteur de transcription

Associé à différents caractères (croissance, carcasse, stature, ingestion...) dans de nombreuses populations (e.g. Doyle et al., 2020)

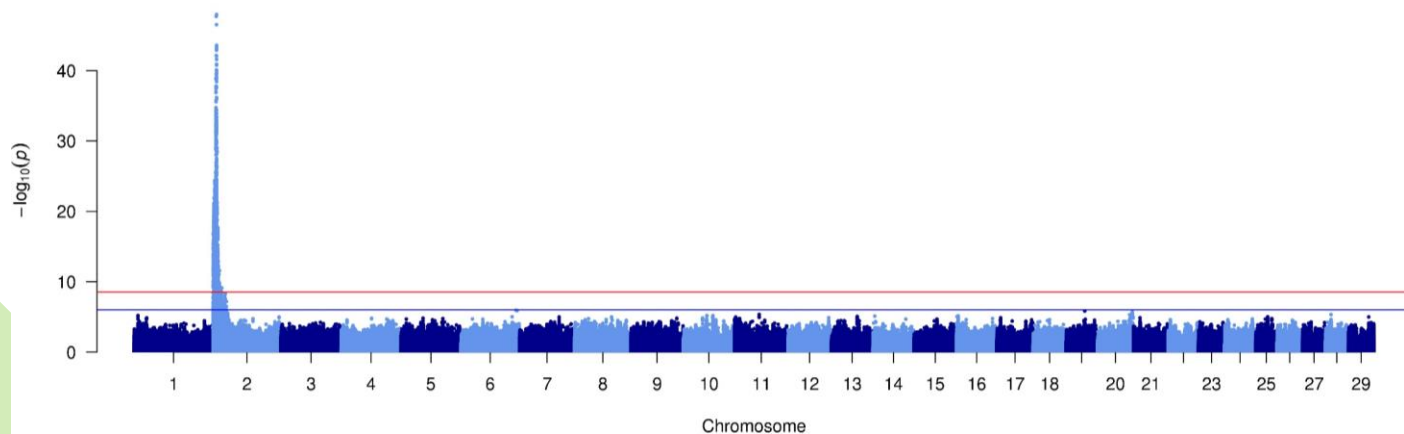
Si 1 seule mutation causale dans les différentes populations => le DL à moins grande distance entre races peut aider à mieux cibler la mutation causale



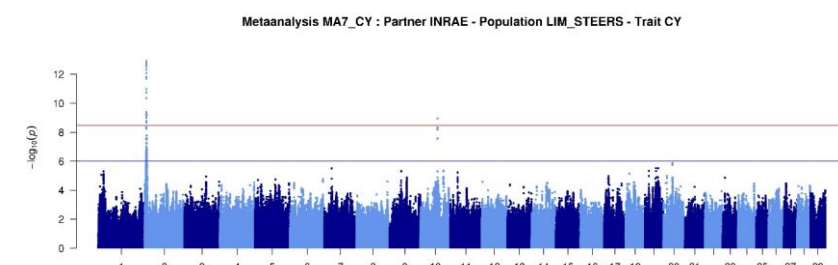
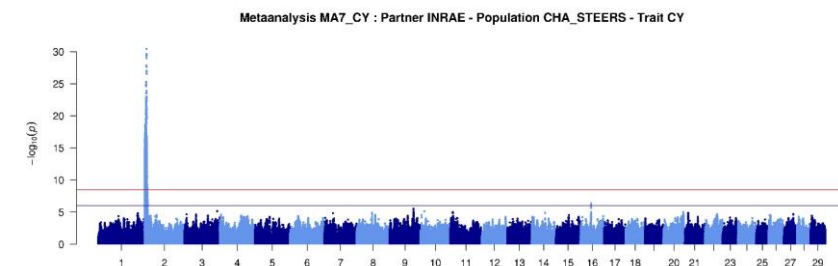
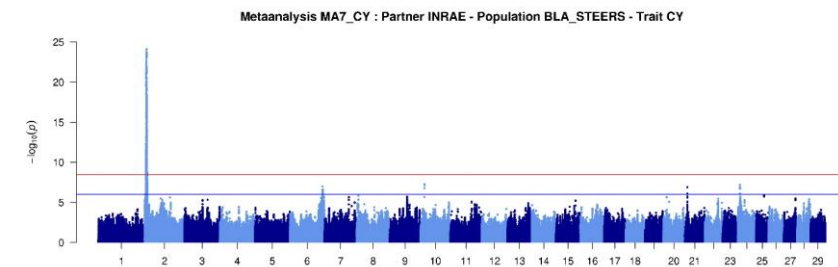
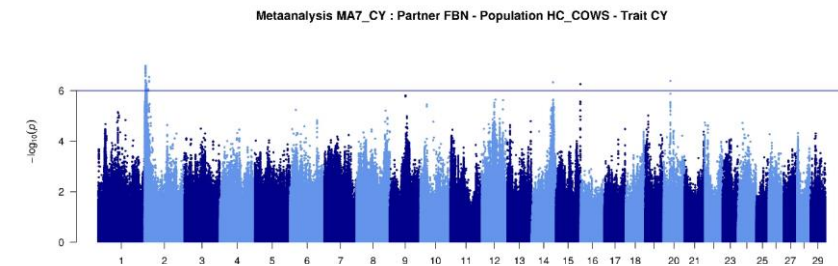
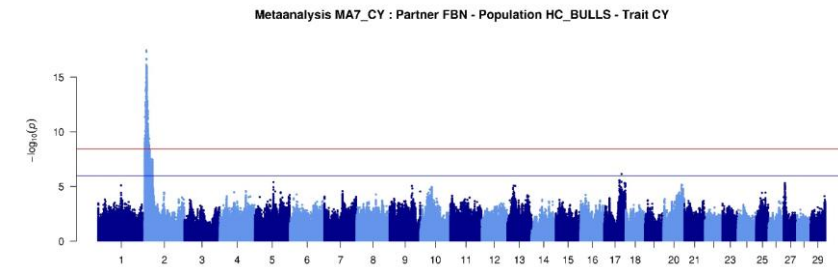
Les MA d'Association du projet BOVREG

- Dans d'autres cas, la MA semble « diluer » le signal
Ex: MA10 carcasse (carcass yield)

MA (effets fixés)



Le pic de la MA est plus large que les pics obtenus dans certaines GWAS intra-populations



Les MA d'Association du projet BOVREG

- Dans d'autres cas, la MA semble « diluer » le signal
Ex: MA10 carcasse (carcass yield)

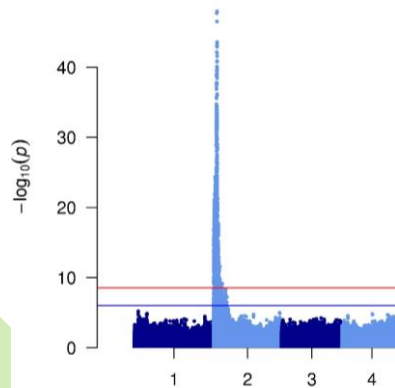
Chromosome 2
Gène *MSTN*
(myostatine)
Phénotype « culard »



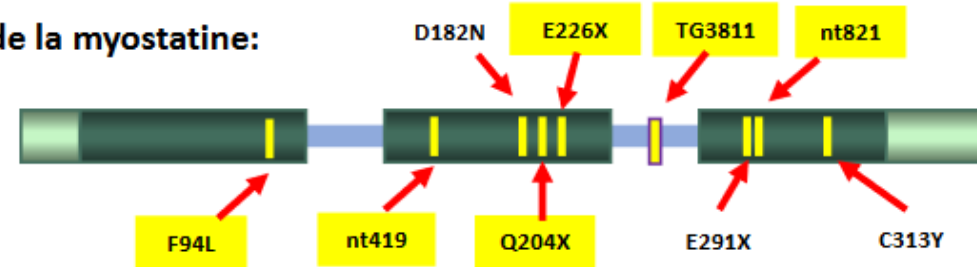
McPherron and Lee, 1997

Plusieurs mutations causales identifiées, différentes selon les races
Renand et al., 3R 2020

MA (effets fixés)



- 9 mutations connues dans le gène de la myostatine:



- Les plus fréquentes parmi les bovins génotypés sur les puces EuroGMD distribuées par Valogène.

Fréquences alléliques	n	+	F94L	Q204X	nt419	E226X	TG3811	nt821
Salers	1 099	99,6	0,1					0,2
Charolais	42 780	84,2	6,5	9,2				
Aubrac	775	8,1	87,0	0,2			2,0	2,6
Parthenais	1 350	0,4	0,2	0,3	4,3	7,6	0,3	86,9
Blonde d'Aquitaine	16 343	0,2	0,5	0,2	0,6		98,5	
Limousine	11 677	0,1	99,1	0,2				0,5



Si plusieurs mutations causales dans les différentes populations
=> dilution du signal GWAS

Conclusions

- L'UMT réalise des GWAS au niveau de la séquence depuis plusieurs années (premiers travaux en 2014 avec le projet BIG)
- Elle contribue à des MA ou réalise des MA à l'échelle internationale
- Ces travaux permettent de cibler précisément les variants causaux ou les variants en fort DL avec les variants causaux
- Ces variants sont régulièrement mis dans la partie recherche de la puce EuroGenomics et donc inclus dans les modèles d'évaluation pour améliorer la précision des valeurs génomiques des caractères d'intérêt et réduire le phénomène « d'érosion des effets des marqueurs » dans les évaluations génomiques



Merci

thierry.tribout@inrae.fr

