

INRAE



**➤ H2020 BovReg
Meta-analyses GWAS à partir des séquences
complètes du génome bovin**

Marie-Pierre SANCHEZ – INRAE, UMR GABI, équipe G2B



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 815668

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

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



20 partenaires de 14 pays

- Allemagne (2)
- Belgique (2)
- Danemark (1)
- Espagne (1)
- France (2)
- Finlande (1)
- Italie (1)
- Pays-Bas (2)
- Pologne (1)
- Portugal (1)
- Royaume-Uni (3)
- Suisse (1)

- Australie (1)
- Canada (1)



GABI (G2B et GiBBS) dans 8 WP (moléculaire/bioinformatique/quantitatif)

WP4 – Didier BOICHARD, Pascal CROISEAU, Marie-Pierre SANCHEZ, Thierry TRIBOUT
Analyse intégrée génotypes / phénotypes

WP7 – Didier BOICHARD, Pascal CROISEAU, Thierry TRIBOUT Prise en compte de l'information biologique dans les modèles de prédiction génomique

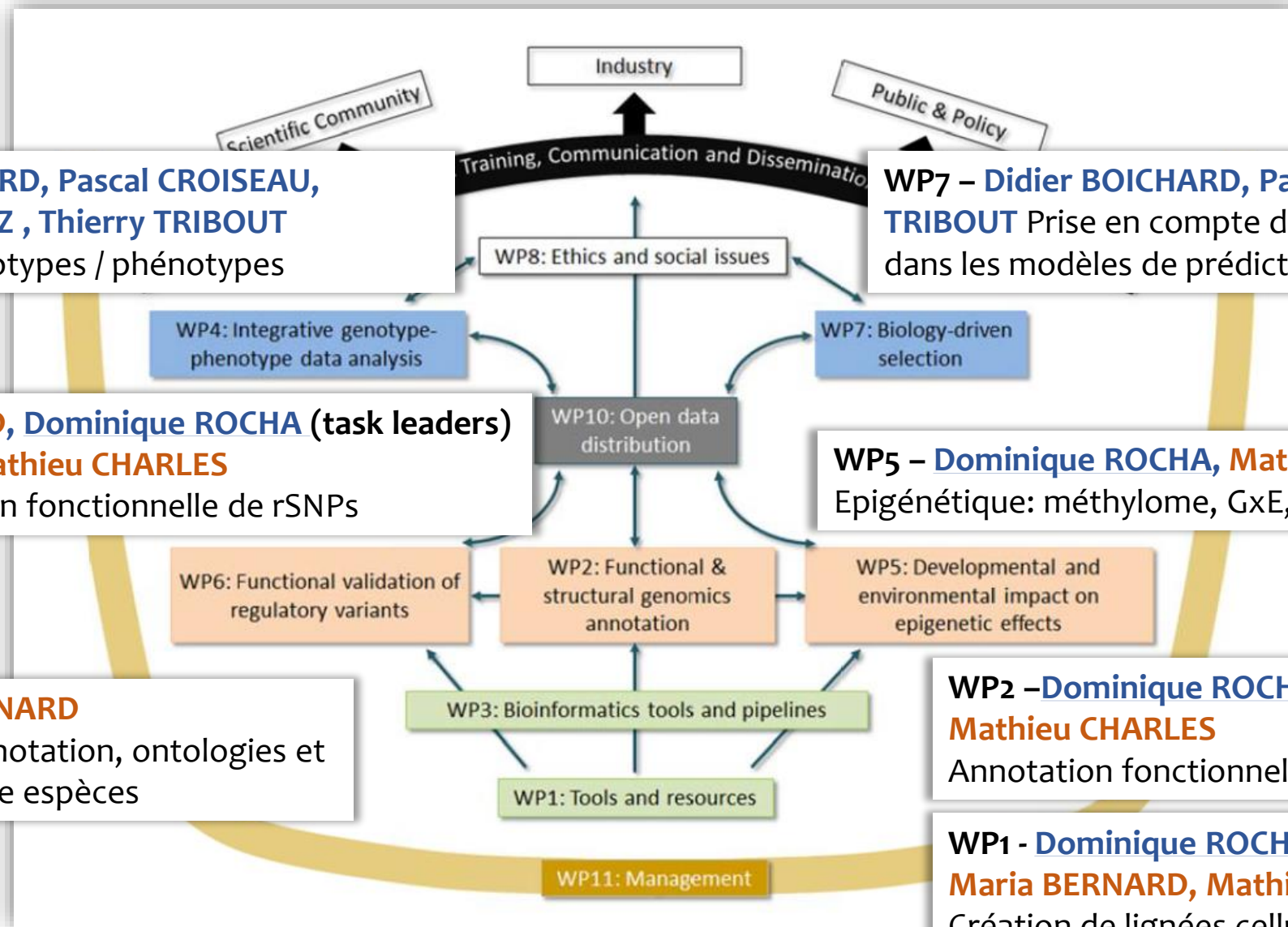
WP6 – Maria BERNARD, Dominique ROCHA (task leaders) Maureen FEMENIA, Mathieu CHARLES
Annotation et validation fonctionnelle de rSNPs

WP5 – Dominique ROCHA, Mathieu CHARLES
Epigénétique: méthylome, GxE, effet transgénérationnel

WP3 – Maria BERNARD
Pipelines pour annotation, ontologies et comparaison entre espèces

WP2 – Dominique ROCHA, Maureen FEMENIA, Mathieu CHARLES
Annotation fonctionnelle et structurale

WP1 - Dominique ROCHA (WP leader) Maria BERNARD, Mathieu CHARLES
Création de lignées cellulaires bovines
Protocoles ChIP-seq automatisés



INRAE

BovReg

SAPS - 21 janvier 2022



p. 4

WP4 – Analyse intégrée des génotypes et phénotypes

WP leader: Hubert PAUSCH (ETH, Suisse) - 11 partenaires sur les 20 impliqués dans BovReg

Objectifs: mettre en évidence les variants génétiques qui expliquent les variations dans l'expression des gènes et les caractères clés de BovReg en utilisant les données de séquence, du transcriptome, du métabolome & les phénotypes

T4.1 – Hubert PAUSCH (ETH, Suisse)

GWAS et méta-analyses à partir des séquences complètes pour des caractères d'efficience, de résistance aux maladies et de fertilité

T4.2 – Carole CHARLIER (GIGA, Belgique)

Impact des sites d'insertion des éléments génétiques mobiles sur les phénotypes

T4.3 – Christa KUHN (FBN, Allemagne)

Analyses eQTLs et mQTLs

T4.4 – Emily CLARK (UEDIN, Royaume-Uni)

Intégration des données -omiques dans les GWAS pour hiérarchiser les variants candidats



WP4 – T4.1

INRAE - GABI, équipe G2B

Didier BOICHARD

Pascal CROISEAU

Marie-Pierre SANCHEZ

Thierry TRIBOUT



KO meeting
Sept. 2019

WP4 – T4.1

INRAE - GABI, équipe G2B

Didier BOICHARD

Pascal CROISEAU

Marie-Pierre SANCHEZ

Thierry TRIBOUT

INRAE  **Université de Limoges - GAMAA**








Romain PHILIPPE



T4.1 – GWAS intra-populations & méta-analyses

Task leader: Hubert PAUSCH (ETH, Suisse) - 9 partenaires

GWAS = « Genome-wide association analyses » pour identifier les régions du génome (QTL) associées aux phénotypes

| | | | | |
|--|--|---|--|--|
| GWAS intra-population |  |  |  | |
| 4 groupes de caractères | Résistance aux mammites | Production laitière & fertilité | Efficacité alimentaire | Production de viande |
| Méta-analyses |  |  |  |  |
| | 8-13 populations < 120 000 anim. | 7-12 populations < 125 000 anim. | 3-9 populations < 13 000 anim. | 3-10 populations < 25 000 anim. |
| Taureaux, vaches, jeunes bovins Races pures laitières ou à viande, animaux croisés, lignées composites | | | | |

T4.1 – GWAS intra-populations

Harmonisation de la stratégie d'analyse entre les différents partenaires



Imputations

2 étapes

50k → 777k (HD)

Animaux de la race génotypés HD

HD → Séquence

Animaux séquencés div. races (projet 1000G)

Fimpute / Minimac / Beagle

~ 20 millions de variants
dont les variants causaux

Modèle mixte testant l'effet de chaque variant indépendamment et incluant un effet polygénique estimé avec une matrice de parenté construite à partir des génotypes 50k (prise en compte de la structure de la population)



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 |

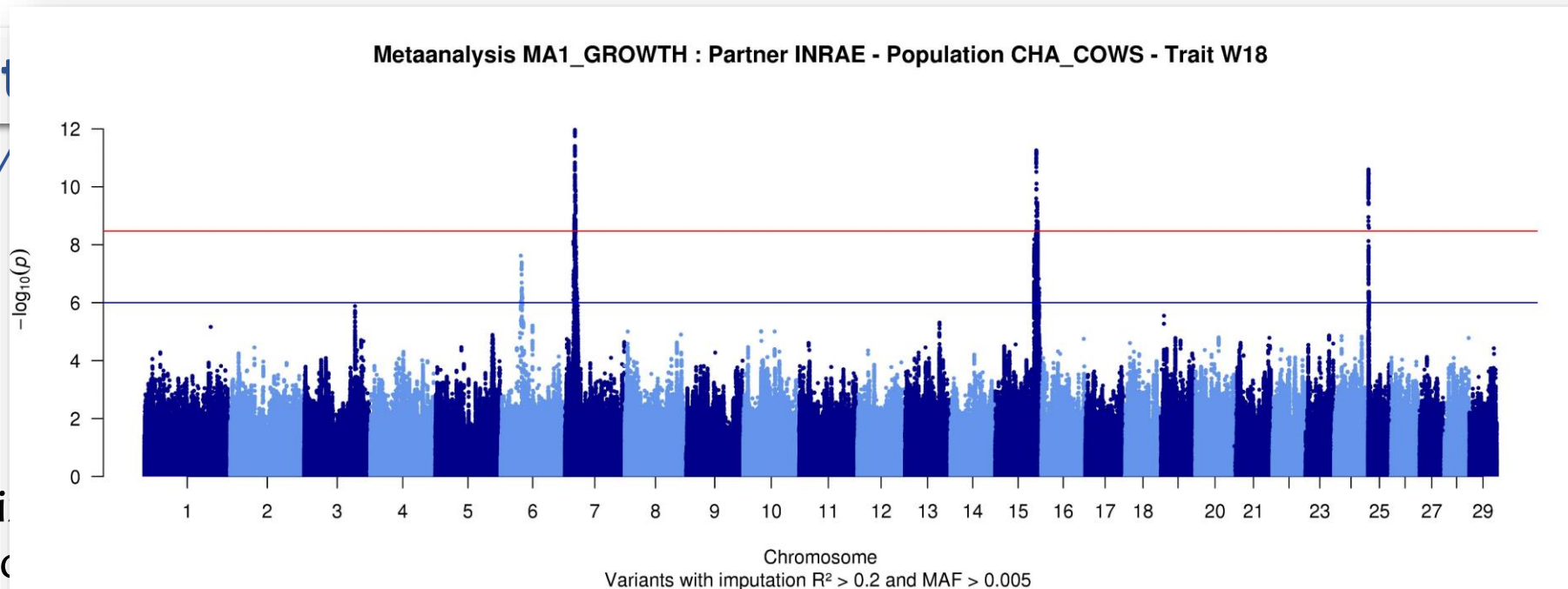
T4.1 – GWAS intra-populations

Harmonisation de la stratégie d'analyse entre les différents partenaires



Impute

Fimpute /



et 1000G)

Modèle mi
un effet po

des génotypes 50k (prise en compte de la structure de la population)

IAS

GCTA

| 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 |

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



T4.1 – GWAS intra-population: contributions INRAE (GABI / GAMAA)




116
populations
x
caractères

38 000
animaux



| Race | Taureaux <i>Prod. laitière, mammite, fertilité + morphologie</i> | Vaches <i>Croissance, morphologie</i> | Jeunes bovins (JB) <i>Carcasse + qualité de viande</i> |
|--------------------|---|--|---|
| Montbéliarde | 3100 taureaux 8+2 caractères | | 4200 JB 3 caractères |
| Normande | 2700 taureaux 8+4 caractères | | 2700 JB 3 caractères |
| Holstein | 10000 taureaux 8 caractères | | |
| Charolaise | | 8000 vaches 7 caractères | 4300 JB (3 caractères) 1000 JB (14+10 caractères) |
| Limousine | | | 1000 JB 14+10 caractères |
| Blonde d'Aquitaine | | | 1000 JB 14+8 caractères |

T4.1 – Contribution de GABI (G2B) aux MA « laitières »

| | | |
|----------------------------------|--|--|
| GWAS intra-population |  | |
| 4 groupes de caractères | Résistance aux mammites 2 caractères | Production laitière & fertilité 8 caractères |
| Méta-analyses |  |  |
| | 8-13 populations < 120 000 anim. | 7-12 populations < 125 000 anim. |
| Contribution G2B | 3 populations 15 800 anim. | 3 populations 15 800 anim. |

T4.1 – INRAE responsable des MA « viande »

GWAS
intra-population

4 groupes de
caractères

Méta-analyses

Contribution INRAE



Production de
viande

INRAE

3-10 populations
< 25 000 anim.

MA

Production de viande **INRAE** **GABI – G2B**

Qualité de viande **INRAE** **Université de Limoges** **GAMAA**



T4.1 – INRAE responsable des MA « viande »

GWAS
intra-population



4 groupes de
caractères

**Production de
viande**

Méta-analyses

INRAE

3-10 populations
< 25 000 anim.

Contribution INRAE

MA

Production de viande **INRAE GABI – G2B**



T4.1 – MA « viande » – populations analysées

Un grand nombre et une grande diversité de populations

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



INRAE  Université de Limoges

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



ETH zürich

- ☒ 2 populations croisées allemandes (HOL x CHA)



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



UNIVERSITY OF ALBERTA



T4.1 – MA « viande » – caractères analysés

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 | Carcass | CW | 1 | 7 | 4 | 19989 |
| 9 | Carcass | AS | 1 | 6 | 2 | 12208 |
| 10 | Carcass | CY | 1 | 5 | 2 | 3694 |
| 11 | Carcass | CG/LMY/MT/CC | 4 | 10 | 5 | 25367 |
| 12 | Carcass | FS/ABT/FC6/FCU/CF | 5 | 8 | 5 | 14622 |
| 13 | Carcass | WS | 1 | 5 | 2 | 2636 |
| 14 | Carcass | ALT | 1 | 5 | 2 | 3692 |
| 15 | Carcass | IFW | 1 | 5 | 2 | 3686 |
| 16 | Carcass | 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

T4.1 – MA « viande » – méthodes

2 méthodes testées

Logiciel METAL (Willer et al., 2010)

z-score

Pour chaque variant, les probabilités (p-value) associées à ses effets estimés dans les différentes études sont combinés et pondérés par la taille de l'échantillon

Effets fixés

Pour chaque variant, ses effets estimés dans les différentes études sont combinés et pondérés par leur variance d'erreur (indirectement par la taille de l'échantillon)

➔ La méthode à effets fixés généralement plus puissante mais nécessite que les caractères des différentes études soient identiques et mesurés dans la même unité
=> standardisation des effets par l'écart-type génétique du caractère, propre à chaque population
Effet d'un variant significatif si $-\log_{10}(p\text{-value}) \geq 8,7$ => seuil 5% après correction Bonferroni (~25M variants)

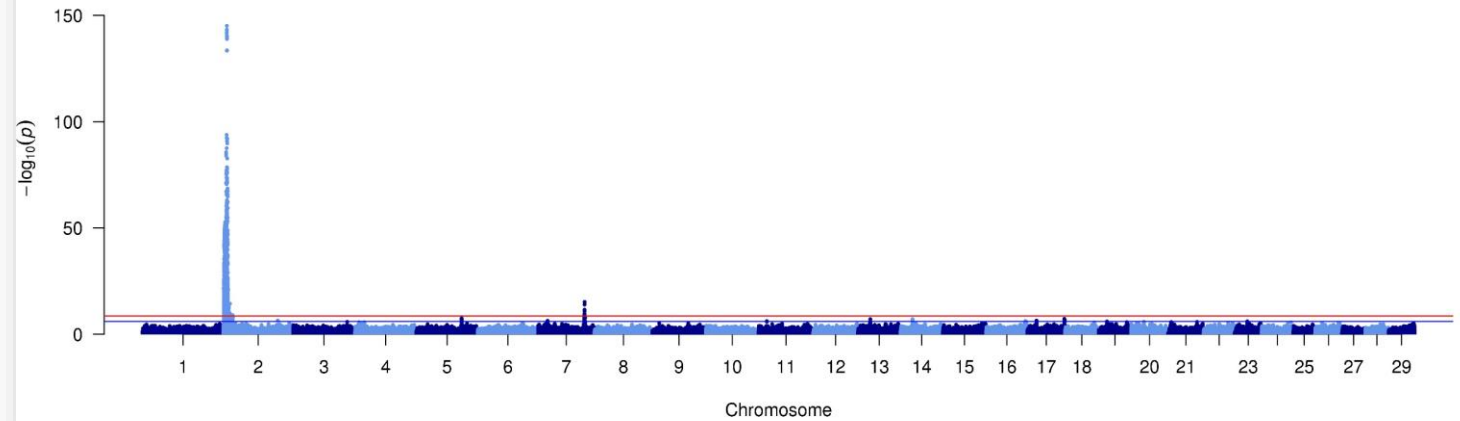
T4.1 – MA « viande » – résultats

Résultats significatifs et donc QTL détectés pour 15 des 16 MA sur 11 autosomes bovins

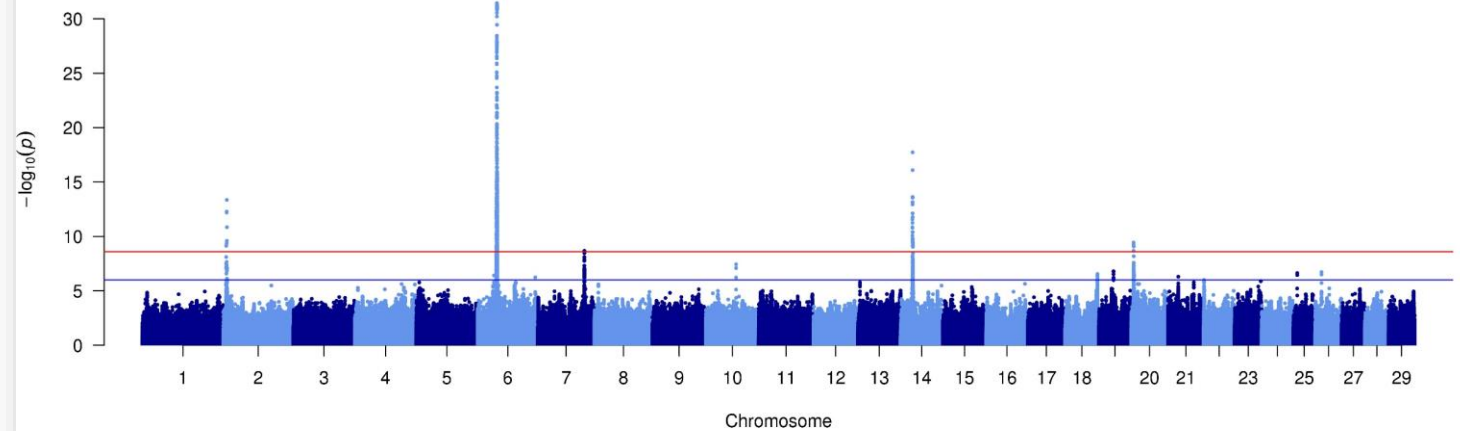
Les QTL les plus significatifs sont localisés

Sur le chromosome 2

MA2 morphologie (méthode à effets fixés)



MA8 carcasse (méthode à effets fixés)



Sur le chromosome 6

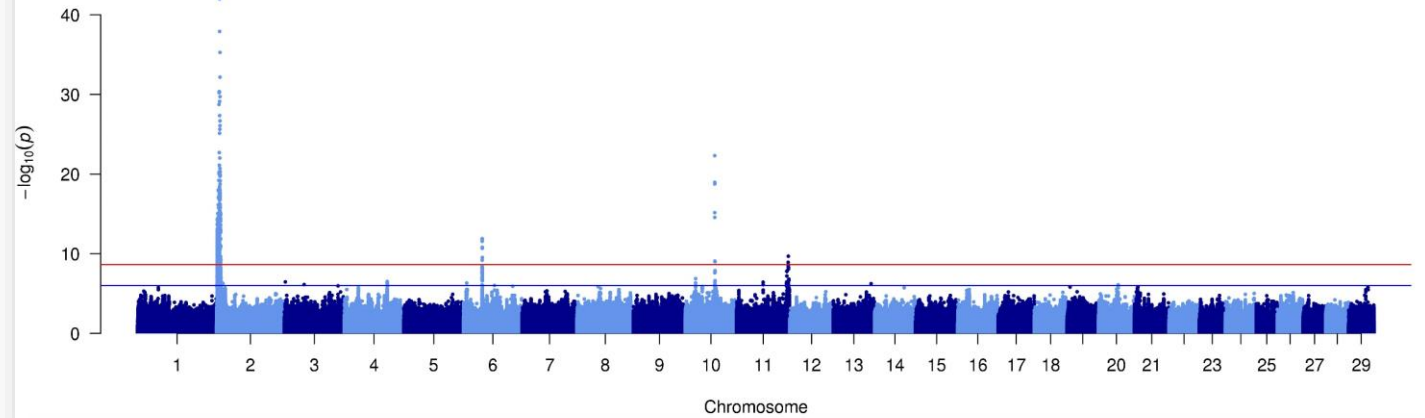
T4.1 – MA « viande » – résultats

Méthode à effets fixés / Méthode du z-score:

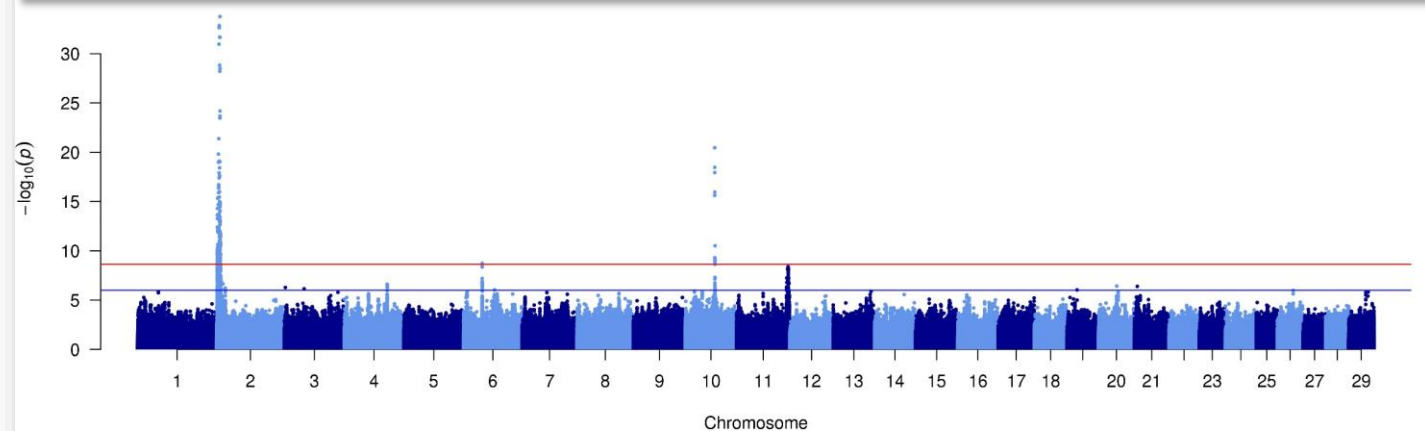
- Les QTL sont généralement plus significatifs
- Plus de QTL détectés

Ex: MA11 carcasse

Méthode à effets fixés



Méthode z-score



T4.1 – MA « viande » – résultats

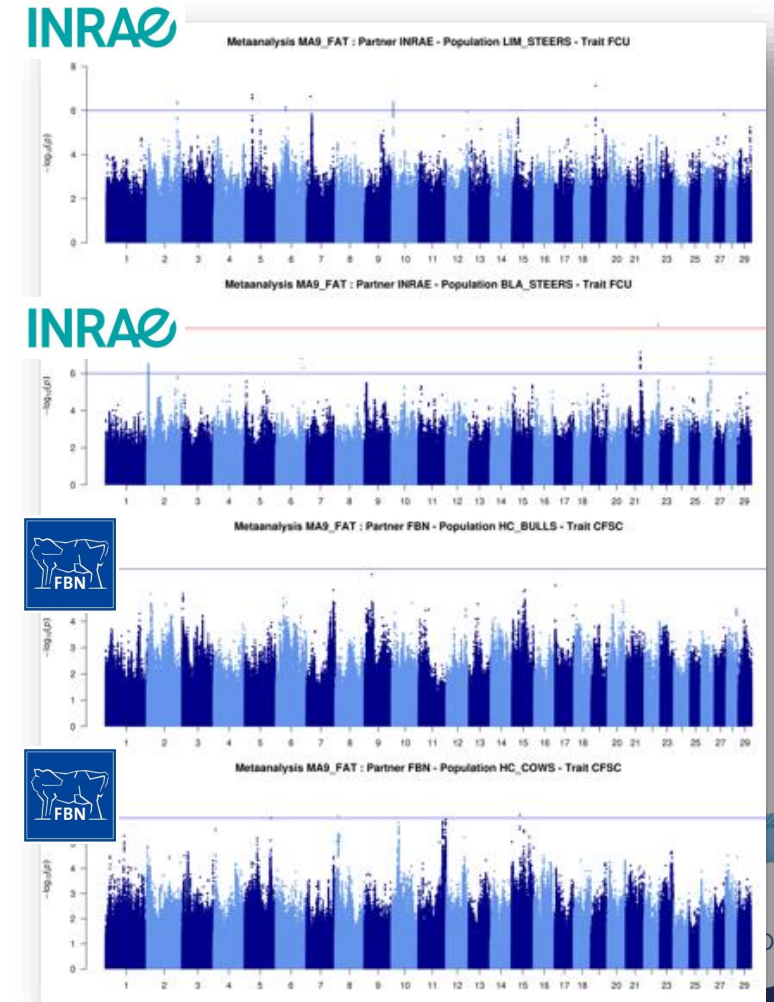
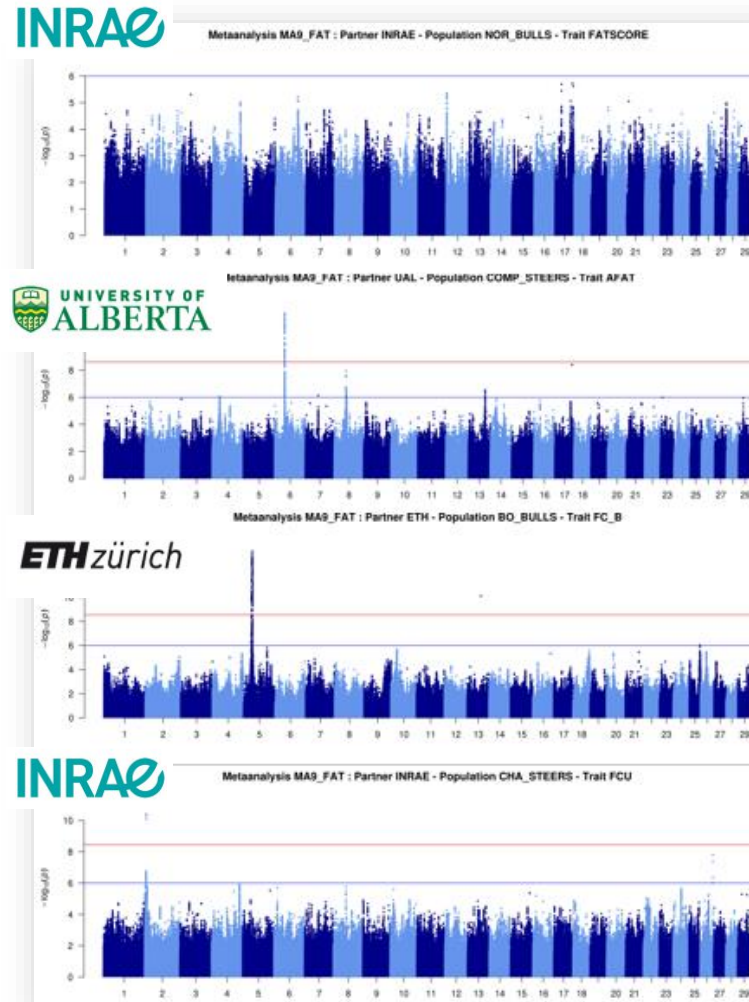
MA / GWAS intra-populations

- confirment les QTL des GWAS intra-populations avec des effets plus significatifs la plupart du temps
- mettent en évidence des QTL non détectés dans les GWAS intra-populations

Ex: MA9 carcasse

GWAS intra-populations

=> 3 QTL sur chrom. 2, 5 et 6



INRAE

BovReg

SAPS - 21 janvier 2022



T4.1 – MA « viande » – résultats

MA / GWAS intra-populations

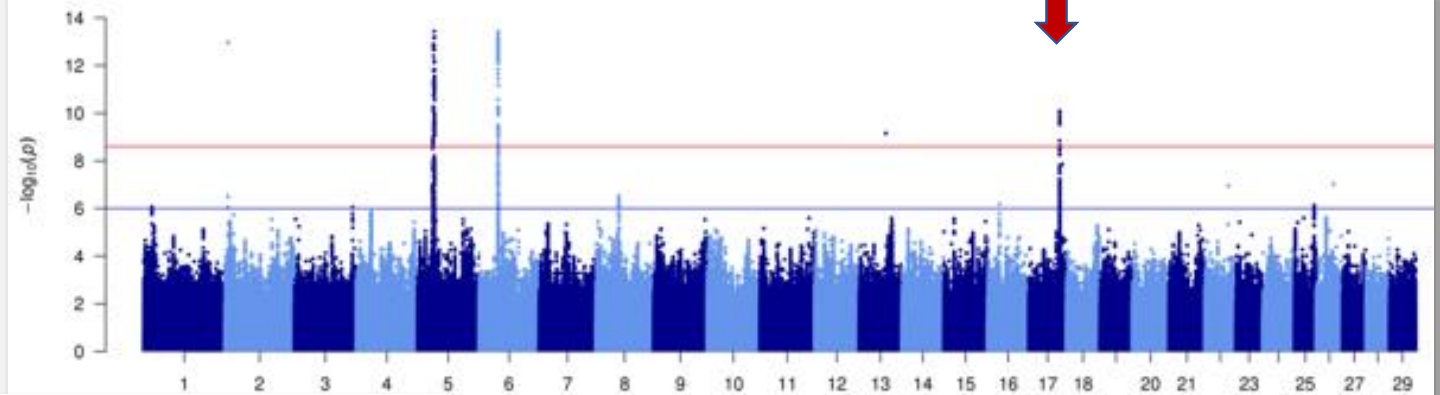
- confirment les QTL des GWAS intra-populations avec des effets plus significatifs la plupart du temps
- mettent en évidence des QTL non détectés dans les GWAS intra-populations

Ex: MA9 carcasse

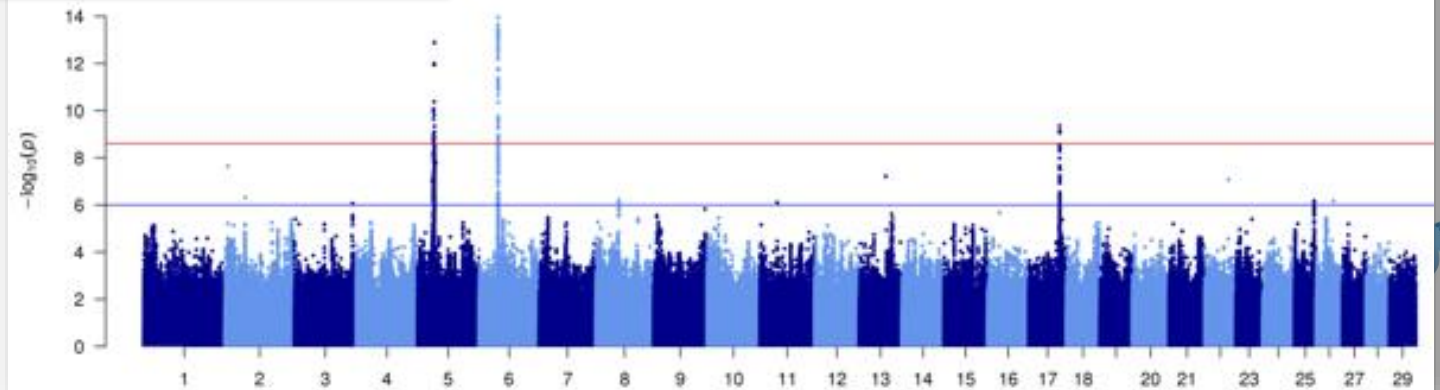
Méta-analyses

=> 3 QTL sur chrom. 2, 5 et 6
+ 1 nouveau QTL sur chrom. 17

Méthode à effets fixés



Méthode z-score



T4.1 – MA « viande » – résultats

MA / GWAS intra-populations – Comparaison des annotations des variants TOP1

| Annotation fonctionnelle | GWAS intra-populations | MA effets fixés |
|-----------------------------------|------------------------|-----------------|
| downstream_gene_variant | 1.9 | 13.9 |
| frameshift_variant | 1.9 | 2.8 |
| intergenic_region | 37.7 | 27.8 |
| intron_variant | 39.6 | 36.1 |
| missense_variant | 1.9 | 2.8 |
| stop_gained | 11.3 | 11.1 |
| synonymous_variant | 1.9 | 0.0 |
| upstream_gene_variant | 3.8 | 5.6 |
| | | |
| % variants TOP1 dans gènes | 62.3 | 72.2 |

Les MA permettent-elles de cibler plus facilement les variants causaux?
=> Situations variées selon les QTL

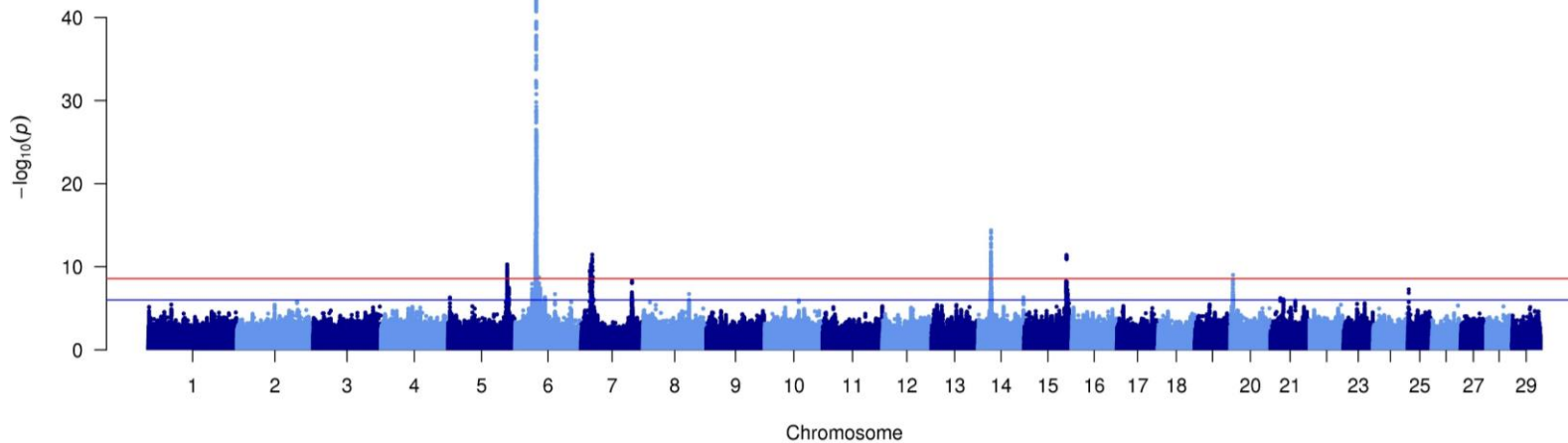
T4.1 – MA « viande » – résultats

MA / GWAS intra-populations

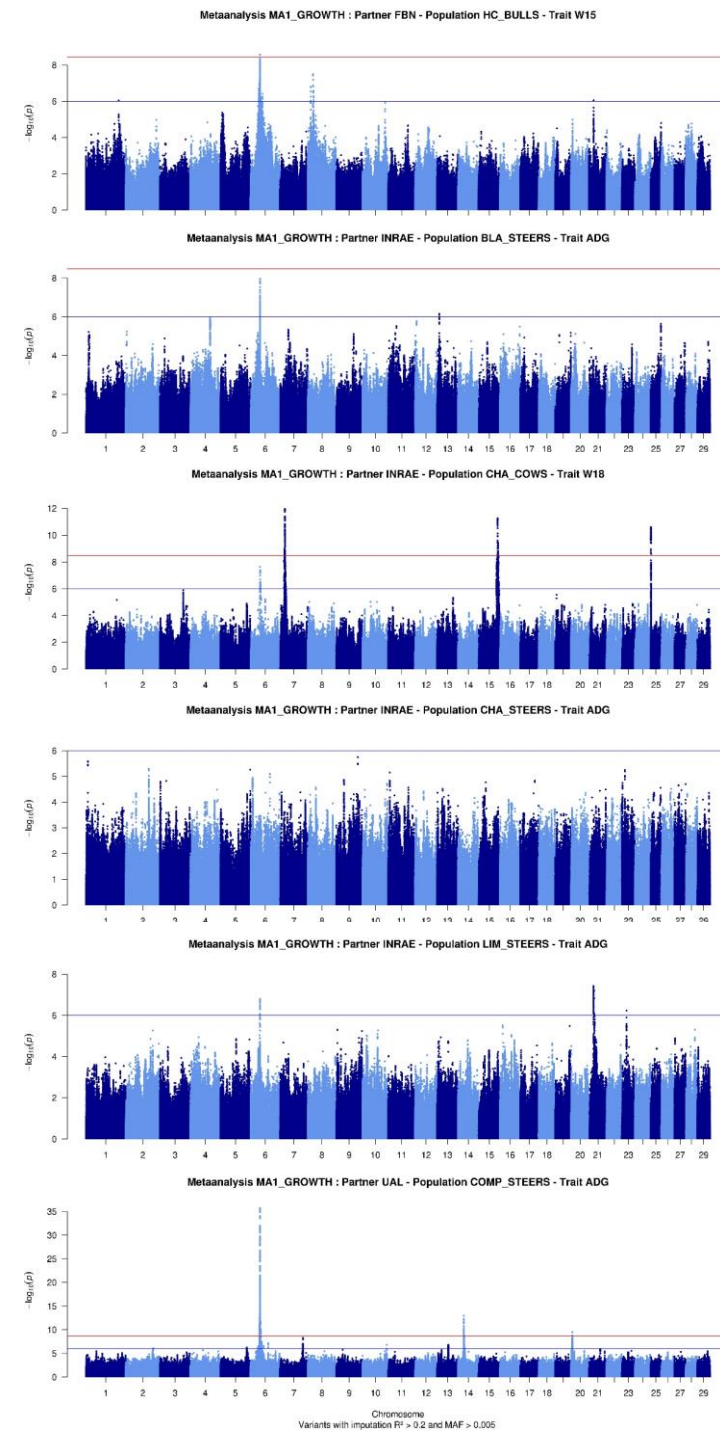
- Dans certains cas, la MA semble mieux cibler les variants causaux

Ex: MA1 croissance

MA (effets fixés)



QTL identifié dans 10 des 16 MA



INRAE

BovReg

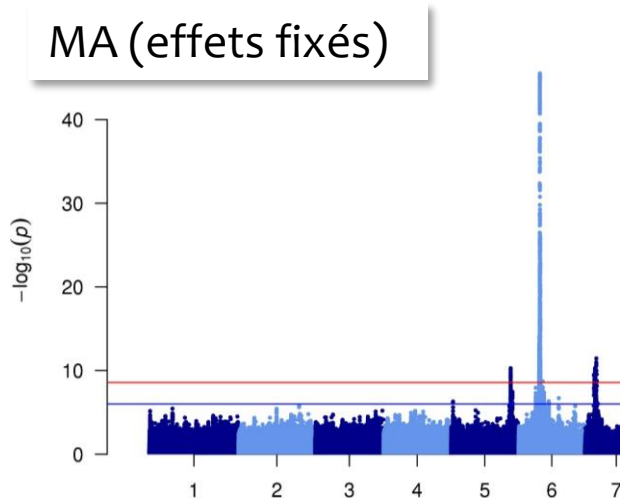
SAPS - 21 janvier 2022

T4.1 – MA « viande » – résultats

MA / GWAS intra-populations

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

Ex: MA1 croissance



QTL identifié dans 10 des 16 MA

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

| 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> |

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 (MA) peut aider à mieux cibler la mutation causale



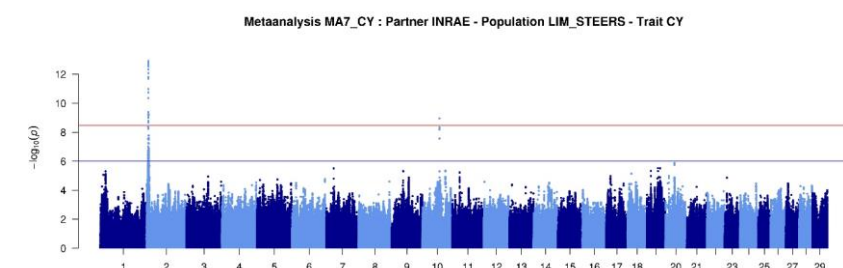
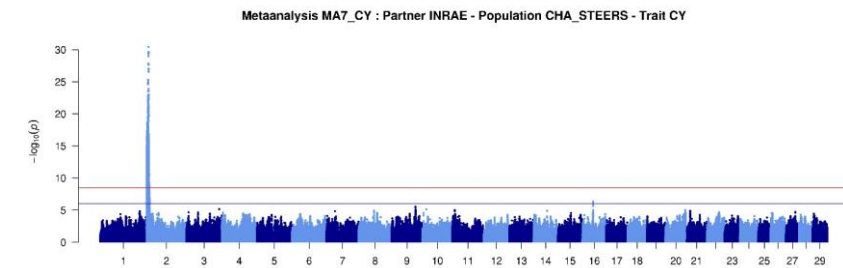
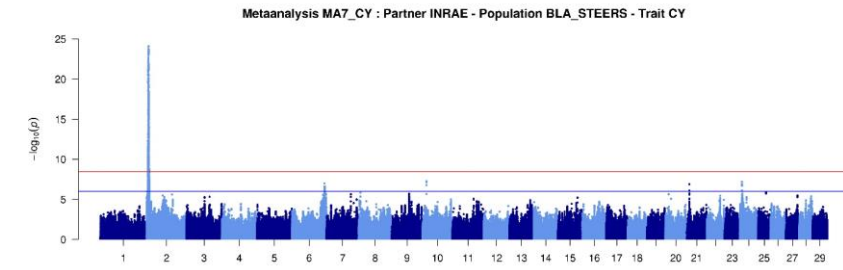
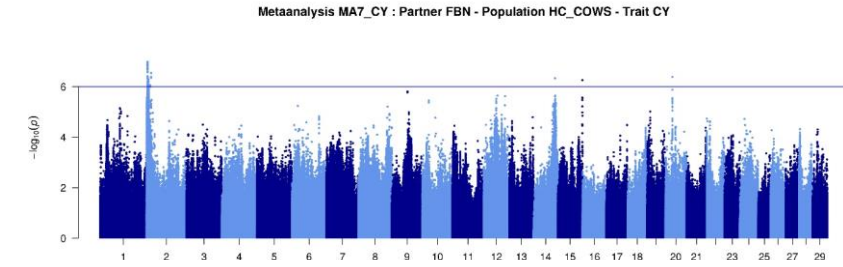
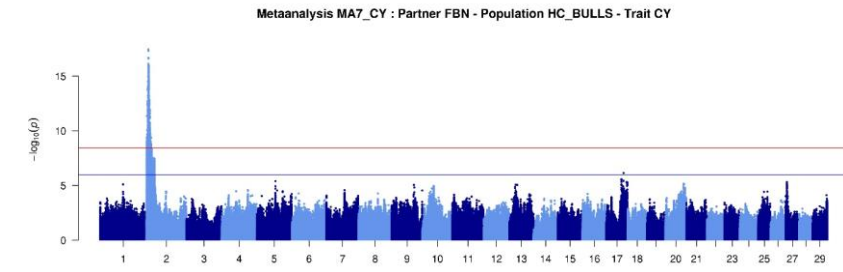
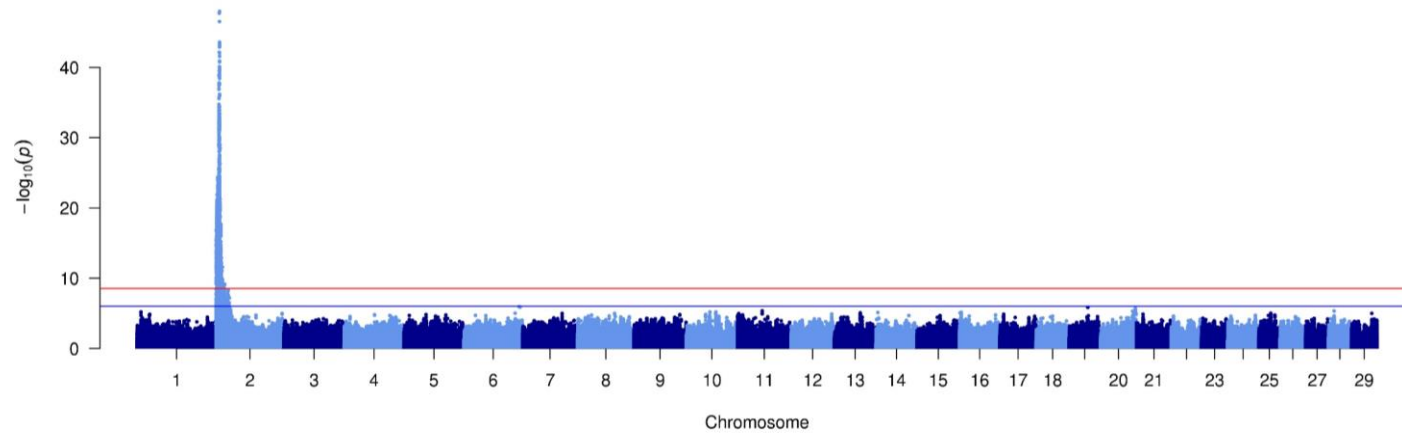
T4.1 – MA « viande » – résultats

MA / GWAS intra-populations

- Dans d'autres cas, la MA semble « diluer » le signal

Ex: MA10 carcasse (carcass yield)

MA (effets fixés)



INRAE

BovReg

SAPS - 21 janvier 2022

T4.1 – MA « viande » – résultats

MA / GWAS intra-populations

- Dans d'autres cas, la MA semble « diluer » le signal

Ex: MA10 carcasse (carcass yield)

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

Chromosome 2

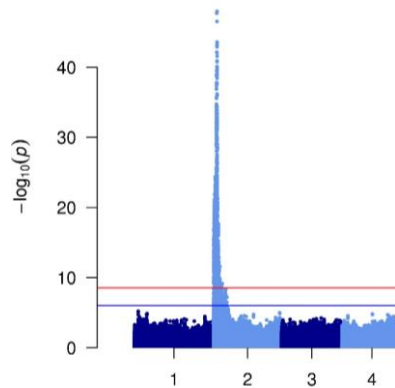
Gène **MSTN**
(myostatine)

Phénotype « culard »

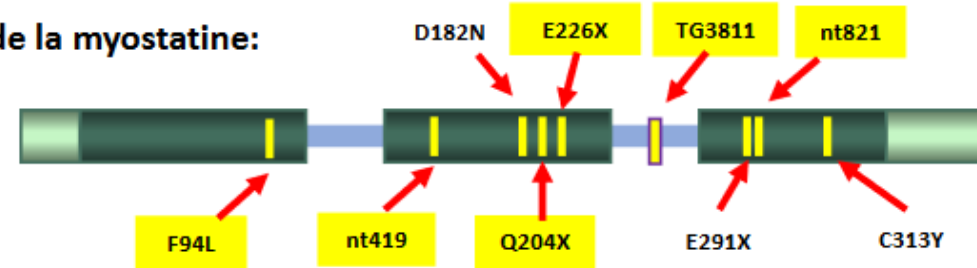


McPherron and Lee, 1997

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



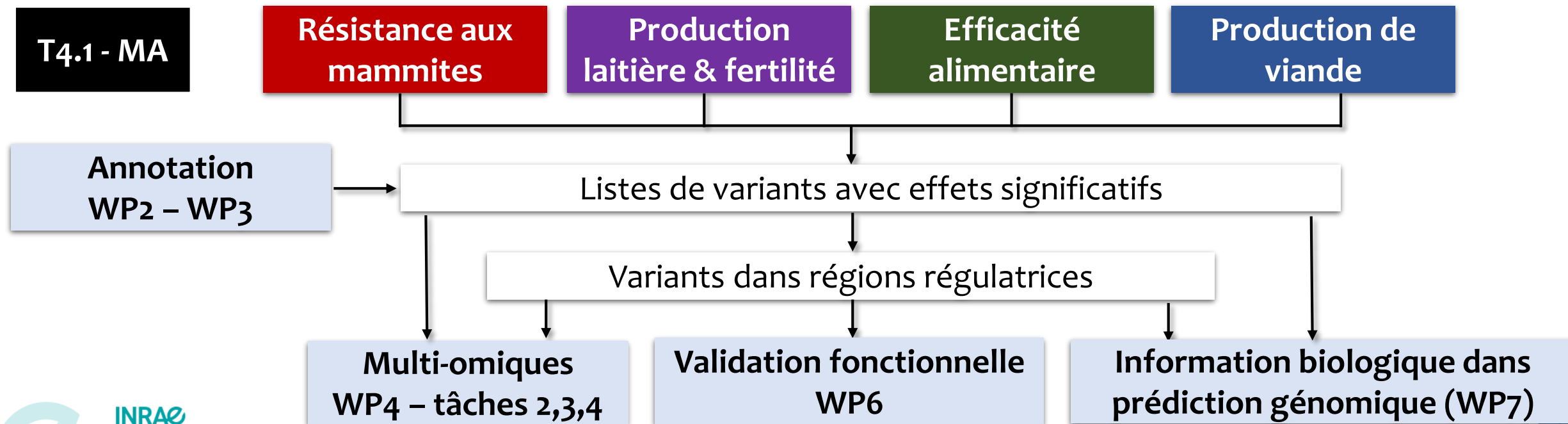
T4.1 – MA « viande » – discussion / conclusions

Les **méta-analyses**, réalisées à partir de résultats de GWAS (sans les données élémentaires)

=> QTL avec des effets plus significatifs (nouveaux QTL)

=> Semblent permettre de mieux cibler les variants causaux dans certains cas (même mutation)

Livrables de la tâche 4.1 et articulation avec les autres tâches / WP



Article soumis au congrès mondial de génétique animale (WCGALP 2022)

Sequence-based GWAS meta-analyses for beef production traits

M.P. Sanchez^{1*}, T. Tribout¹, N. Kadri², P.K. Chitneedi³, S. Maak³, C. Hozé^{1,4},
M. Boussaha¹, P. Croiseau¹, R. Philippe⁵, M. Spengeler⁶, C. Kühn³,
Y. Wang⁷, H. Pausch², D. Boichard¹

¹Université Paris Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; ²ETH, 8092 Zürich, Switzerland; ³Research Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany; ⁴Allice, 75595 Paris, France; ⁵INRAE, USC1061 GAMAA, Université de Limoges, 87060 Limoges, France; ⁶QualitasAG, 6300 Zug, Switzerland; ⁷University of Alberta, Edmonton, Canada. *marie-pierre.sanchez@inrae.fr

Merci!



INRAE

BovReg

SAPS - 21 janvier 2022



p. 28