



**GeT**  
Génome et  
Transcriptome

## CartoSeq :

Identification en masse des variants génétiques  
influençant les caractères de production chez les trois  
principales races laitières françaises

Diane Esquerré

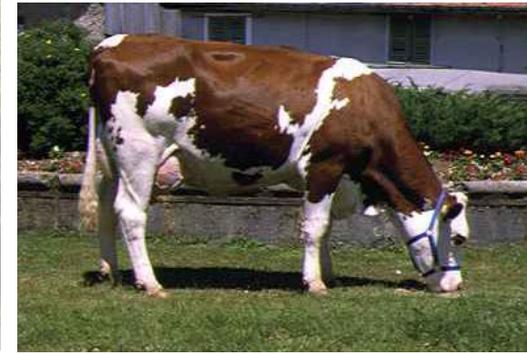
GeT-Plage

30 septembre 2013



# Projet Cartoseq

3 races étudiées: Holstein, Normande, Montbéliarde



Co-financement ANR / Apis-Gene 2011-2013 : 750 000 € + 1 année de prolongation

4 partenaires :



- UMR INRA GABI, G2B : D Boichard, C Grohs, B Weiss, D Rocha



le réseau de la génétique animale

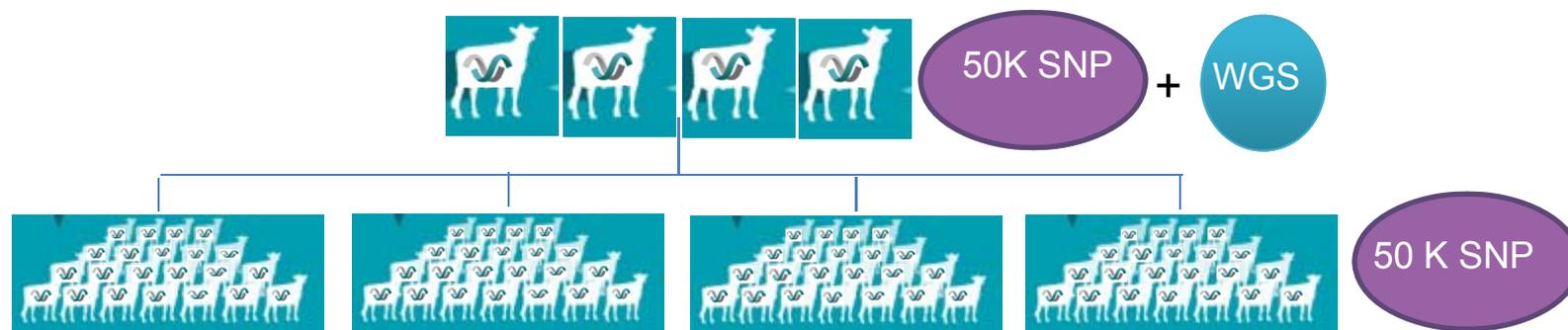


- Plateforme Génomique : D Esquerré, O Bouchez, J Barbieri



- SIGENAE : C Klopp, A Djari, P Dehais, S Rodriguez

# Contexte du projet



- ❖ Insémination artificielle & sélection intensive des taureaux reproducteurs  
⇒ Diversité génétique réduite dans les races bovines laitières (le nombre d'ancêtres efficaces est de l'ordre de 50 individus)
- ❖ Des milliers de taureaux d'IA et leur descendants ont été phénotypés et génotypés sur puces 50K et 800K SNP (projet ANR/Apis-Gène CartoFine et Sélection Génomique)



Cartoseq: Séquençage des ancêtres clés du cheptel bovin français.



# Séquençage 2011-2013

Races	Nb Animaux	Nb lignes HiSeq 2x100pb	Nb Séquences générées	Quantité de séq générées Gb
Holstein	27	58	15 753 596 633	1 575
Normande	20	30	9 359 217 250	936
Montbéliarde	20	28	8 712 184 265	871
<b>Total</b>	<b>67</b>	<b>116</b>	<b>~ 34 milliard</b>	<b>~ 3 400 Gb</b>



15 Flow cells HiSeq 2000  
 16X de couverture en moyenne par animal  
 5,7 Tb de données sauvegardées sur NG6





# Séquençage 2011-2013

TruSeq DNA sample  
Prep Kit (Illumina)

TruSeq sequencing  
kit V2 (Illumina)

Mars 2011

22Gb/lane

85 à 95 % seq alignées

TruSeq sequencing  
kit V3 (Illumina)

Nov 2012

35Gb/lane

85 à 95 % seq alignées

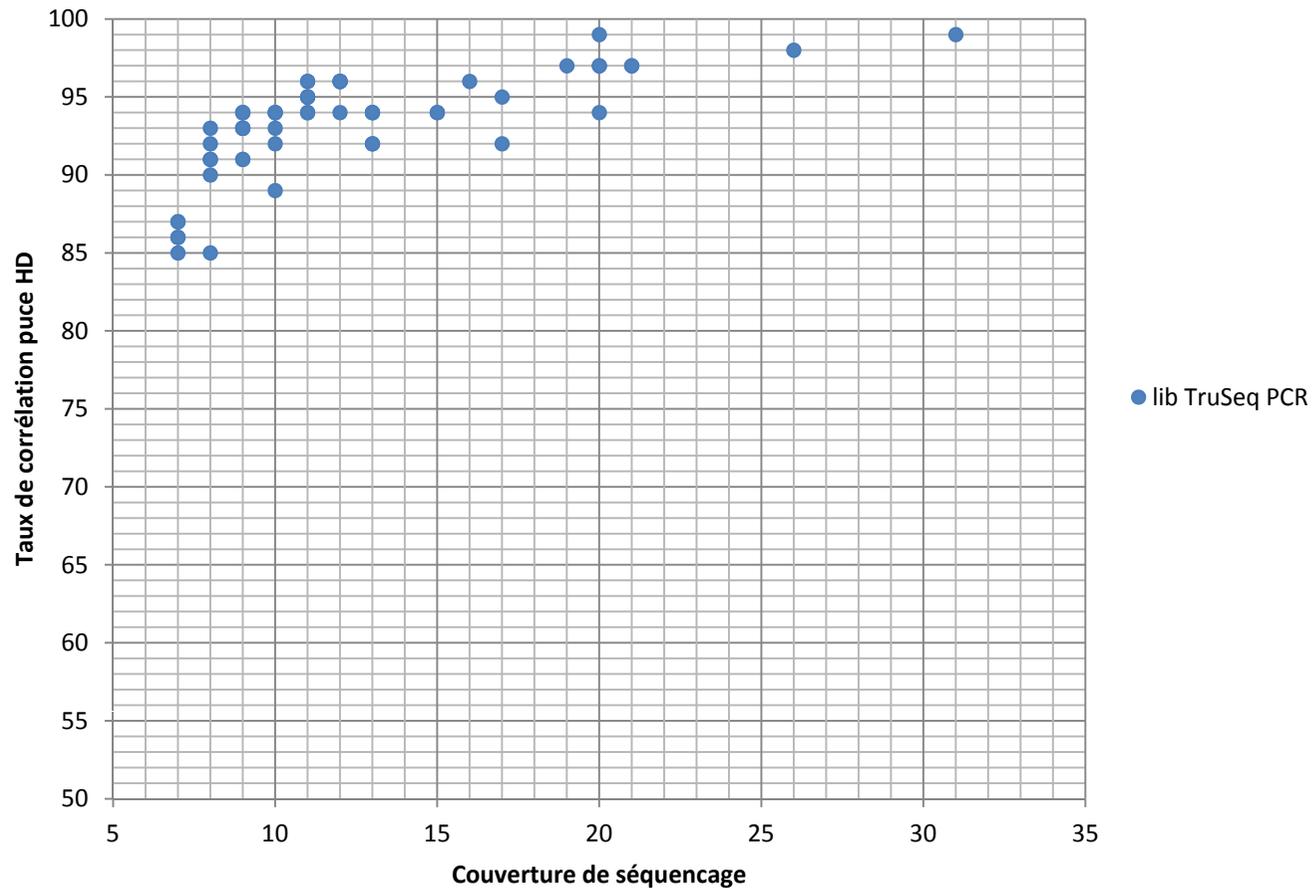
Nextflex PCR free DNA sample  
Prep kit (Bioscientific)

Janv 2013

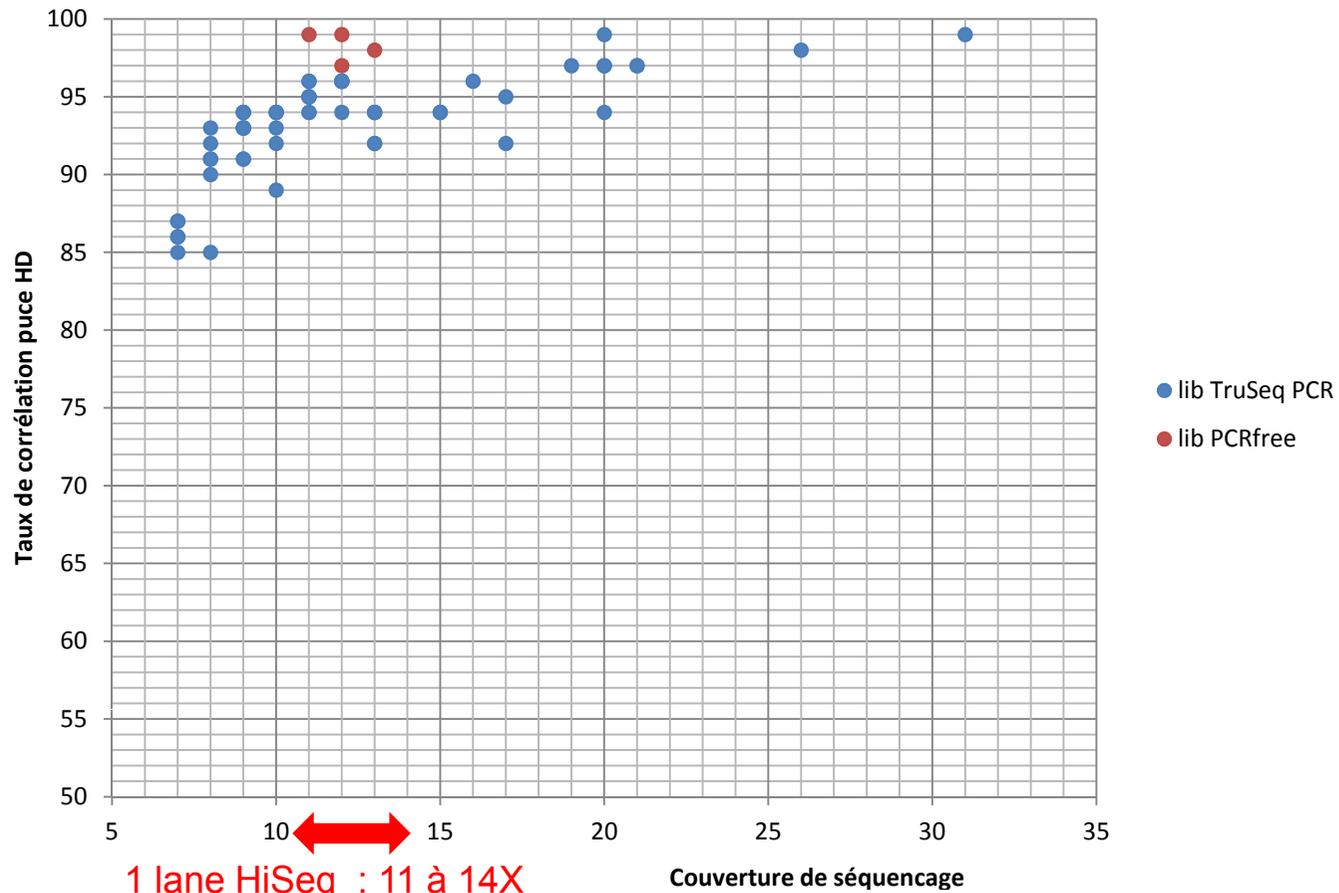
40 Gb/lane

>97% seq alignées

# Corrélation séquençage WGS/puces HD

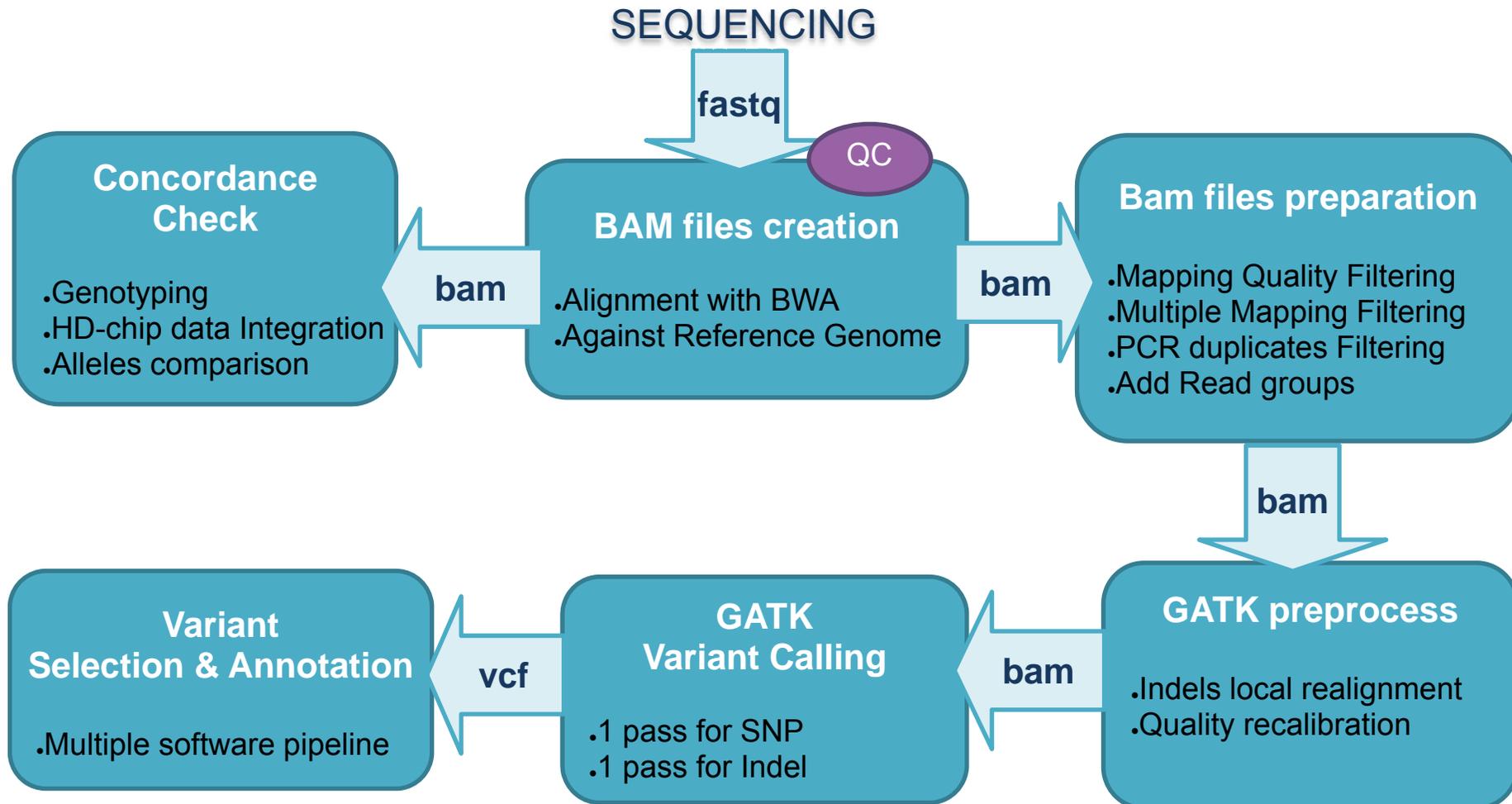


# Corrélation séquençage WGS/puces HD



1 lane HiSeq : 11 à 14X

97 – 99% de corrélation en « PCR free »



<http://www.1000bullgenomes.com/>



1000 bull genomes project

1000 bull genomes project consortium

Comité de pilotage

- Ruedi Fries (Technische Universität München, Germany)
- Mogens Lund/Bernt Guldbandsent (Aarhus University, Denmark)
- Didier Boichard (INRA, France)
- Paul Stothard (University of Alberta, Canada)
- Roel Veerkamp (Wageningen UR, Netherlands)
- Ben Hayes/Mike Goddard (DFL)
- Curt Van Tassell (United States Department of Agriculture)

**The Project**

The 1000 bull genomes project aims to provide, for the bovine research community, a large database for imputation of genetic variants for genomic prediction and genome wide association studies in all cattle breeds. The project aims to develop a resource to allow project partners to impute full genome sequence in bulls and cows that have been genotyped with SNP arrays. This could be for example for the purposes of genomic prediction, genome wide association, and discovery of causal mutations.

**Project Resources**

To join the 1000 bull genomes project you are required to contribute a minimum of 25 animals sequenced at 10.5X coverage after quality control (or 250x equivalent), and be approved by the project steering committee.

Séquençage des ancêtres clés & mise en place d'une base de données recensant tous les variants

→ 28 millions SNP et Indels identifiés

→ pour permettre l'imputation de toute la séquence génomique sur les animaux génotypés sur puces SNP en prenant en compte les informations de filiation.

→ GWAS, prédiction génomique et identification des mutations causales

Ex1: Détection de mutations génétiques récessives par recherche de déficit en homozygotes

Ex2: Détection de mutations impliquées dans les syndromes sporadiques dominants

## 2/ Détection de mutations génétiques récessives

⇒ Races laitières

→ Problème de baisse de fertilité :



CVM affected calf  
(A. Gentile)

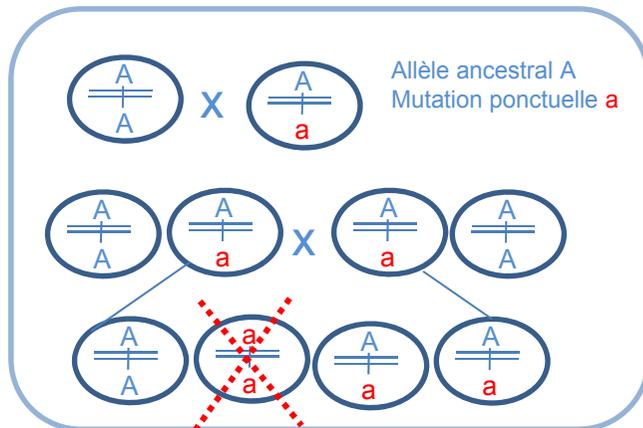
- 50% expliquée par la sélection pour la production de lait
- Autres facteurs ...

→ Population génétique réduite → Consanguinité → responsable de l'émergence régulière de mutations récessives (ex: BLAD, CVM)

⇒ Apparition et expression de mutations génétiques récessives létales durant la phase embryonnaire

Recherche d'haplotypes présentant un déficit en homozygotes dans les données de génotypage (SG: 48878 Holstein, 16833 Montbéliarde, 1466 Normande)

9 régions montrent une perte significative en homozygotes et une baisse de fertilité associée



Haplo	BTA	Interval (Mb)
BY	21	20.2-22.3
HH1	5	61.4-66.2
HH5	3	45.8-47.6
HH6	3	49.4-52.6
HH3	8	94.0-96.5
HH4	1	1.9-3.3
MH1	19	27.6-29.4
MH2	29	27.9-29.1
NH5	7	3.6-4.6

Analyse  
données WGS



# Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2

Sébastien Fritz<sup>1,2</sup>, Aurelien Capitan<sup>1,2</sup>, Anis Djari<sup>3</sup>, Sabrina C. Rodriguez<sup>2,3</sup>, Anne Barbat<sup>2</sup>, Aurélie Baur<sup>1,2</sup>, Cécile Grohs<sup>2</sup>, Bernard Weiss<sup>2</sup>, Mekki Boussaha<sup>2</sup>, Diane Esquerré<sup>4</sup>, Christophe Klopp<sup>3</sup>, Dominique Rocha<sup>2</sup>, Didier Boichard<sup>2\*</sup>

1 à 3 porteurs



15 à 45 000 polymorphismes



Filtres:

- Concordant avec le statut QTL
- Non homozygote
- Allele absent dans les autres races et sur la référence

500 à 1,500 candidats positionnels après filtres



Filtres:

- Dans les séquences codantes
- Conservé dans les autres espèces
- Predit comme très délétère par Sift et Polyphen



1 à 8 mutations dans le codant (dont 0 à 2 mutations délétères)

Haplo	BTA	Interval (Mb)	Candidate mutation	Gene	Amino acid changes
BY	21	20.2-22.3	g.21184870_21188198del	FANCI	p.V877Lfs27X
HH1	5	61.4-66.2	g.63150400C>T	APAF1	p.Q579X
HH5	3	45.8-47.6	g.43412427C>A	SLC35A3	p.V180F
HH6	3	49.4-52.6			
HH3	8	94.0-96.5	g.95410507T>C	SMC2	p.F1135S
HH4	1	1.9-3.3	g.1277227A>C	GART	p.N290T
MH1	19	27.6-29.4	g.27956790C>T	SHBG	p.Q52X
MH2	29	27.9-29.1	g.28879810C>T	SLC37A2	p.R12X
NH5	7	3.6-4.6	No candidate		

7 mutations candidates

### 3/ Détection de mutations impliquées dans les syndromes sporadiques dominants

Syndrome du « veau bouledogue », forme létale de chondrodysplasie

Prévalence de ~1% dans les descendance d'un seul taureau d'élite mondial séquencé (Holstein)

Effet dominant, mosaïcisme germlinal suspecté



Analyse données NGS : 2 cas et 239 controles:	Bulldog calf #1	Bulldog calf #2	Shared by both calves
Average sequence coverage in non repetitive regions (sd)	9.0 (3.4)	12.0 (4.8)	
Number of putative sequence variants detected	5,690,540	6,350,410	3,826,591
Heterozygous polym. absent from the whole 1000 bull genome database (and affecting the 1 <sup>ary</sup> structure of a protein)	7,693 (23)	11,020 (45)	446 (2)
Mutation in functional candidate	g.32475732G>A on BTA5 (i.e. COL2A1 p.G960R)		

Mutation identique chez l'homme dans COL2A1, responsable de nombreux désordres squelettiques .

# Merci !

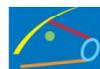


G2B:

- Didier Boichard
- Dominique Rocha
- Cécile Grohs
- Bernard Weiss
- Mekki Boussaha
- François Guillaume
- Irene van den Berg



- Johanna Barbieri
- Olivier Bouchez
- Emeline Lhuillier
- Gérald Salin
- Frédéric Escudié
- Claire Kuchly



**UNCEIA**

le réseau de la génétique animale

- Sébastien Fritz
- Aurélien Capitan



- Christophe Klopp

- Anis Djari
- Sabrina Rodriguez



## Imputation to the Genome Sequence

