Integration of GWAS, CNV and selection signature reveals candidate genes for fat regulation in chickens

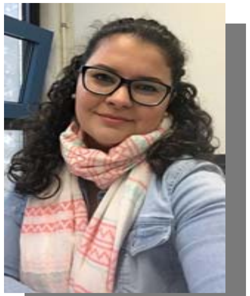
Dr. Luiz Lehmann Coutinho

University of São Paulo (USP)
Luiz de Queiroz College of Agriculture (ESALQ)

Auckland - NZ
February 2018

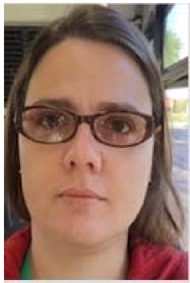
Integration of GWAS, CNV and selection signature reveals candidate genes for abdominal fat regulation in chickens

Gabriel C. M. Moreira
PhD candidate
University of São Paulo
🇧🇷



Thaís F. Godoy
PhD candidate
University of São Paulo
🇧🇷

Clarissa Boschiero
Post-Doc
Noble Research Institute
🇧🇷 🇺🇸



Aline S. M. Cesar
Post-Doc
University of São Paulo
🇧🇷

James M. Reecy
Professor / Researcher
Iowa State University
🇺🇸

Mônica C. Ledur
Researcher
EMBRAPA Suínos e Aves
🇧🇷

Dorian J. Garrick
Chief Scientist
Massey University
🇳🇿

Vinícius H. da Silva
PhD Student
Wageningen University & Research
🇧🇷 🇳🇱

R.P.M.A. Crooijmans
Assistant professor
Wageningen University & Research
🇳🇱

M.A.M. Groenen
Professor / Researcher
Wageningen University & Research
🇳🇱

Luiz L. Coutinho
Professor / Researcher
University of São Paulo
🇧🇷

Methods - *populations*

- **EMBRAPA F2 Chicken Resource Population (F2-TC)**

White Plymouth Rock
White Cornish breeds

White Leghorn

Pure broiler-line (TT)

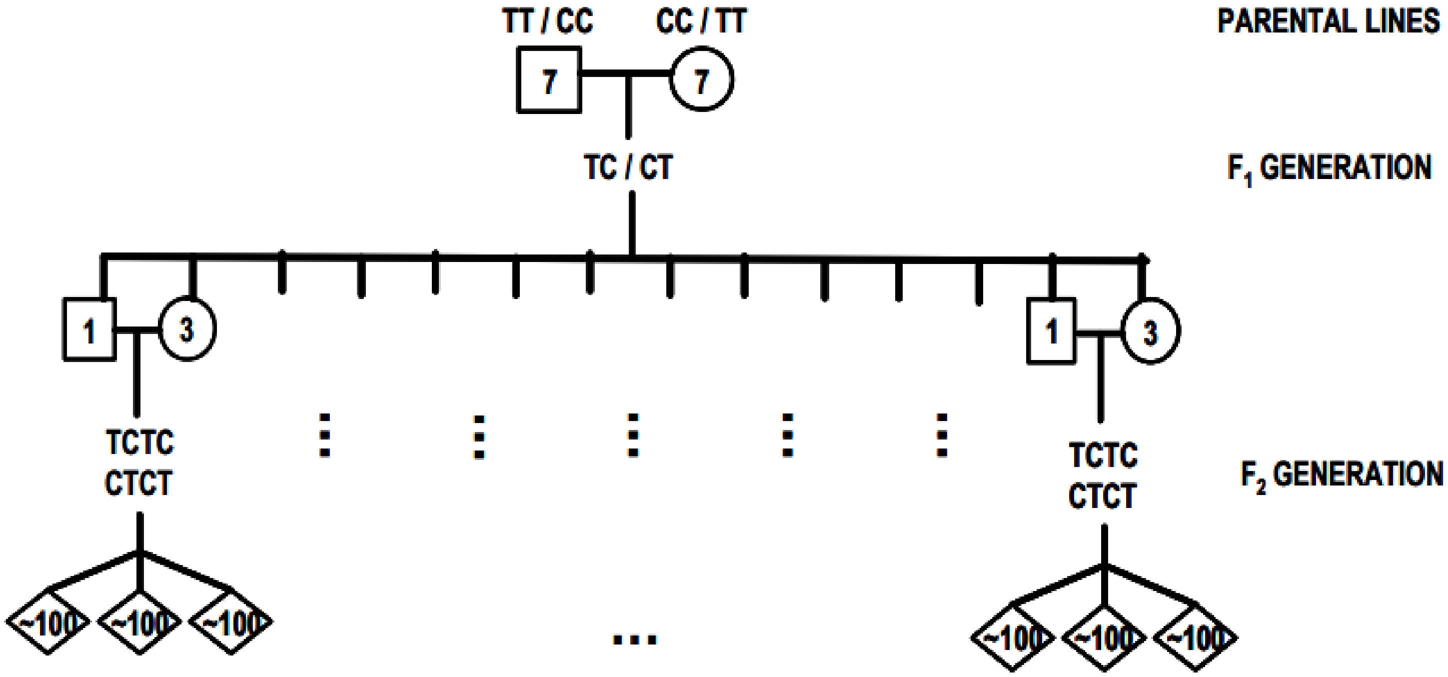
X

Pure layer-line (CC)

Line	weight (g)	carcass (g)	breast (g)	fat (g)
TT	2,395	1,776	486	57.2
CC	514	330	73	0.8

Methods - *population*

- **EMBRAPA F2 Chicken Resource Population (F2-TC)**



Rosário, M.F. do; Ledur, M.C.; Moura, A.S.A.M.T.; Coutinho, L.L.; Garcia, A.A.F. *Genotypic characterization of microsatellite markers in broiler and layer selected chicken lines and their reciprocal F1s.* **Scientia Agricola, Piracicaba, v. 2, n. 66, p. 150–158, 2009.

➢ GWAS: Identification of genomic windows associated with carcass fat content traits

Gabriel C. M. Moreira
PhD candidate
University of São Paulo

Methods

497 chickens from F2 generation

High-density SNP chip
600K Affymetrix

DishQC ≧ 0.82
Sample call rate ≧ 90%
SNP call rate ≧ 98%
MAF ≧ 2%

480 samples
371,785 SNPs

Traits measured:**
Carcass fat content (grams) - CFCg
Carcass fat content (dry matter basis) -CFCDM

GWAS – GenSel software
BayesC ⇢ ***To estimate the variance componentes - priors***
BayesB ⇢ **π=0.9988**
- **41,000 MCMC interaction – 1,000 discarded**
- **947 1-Mb non-overlapping windows**

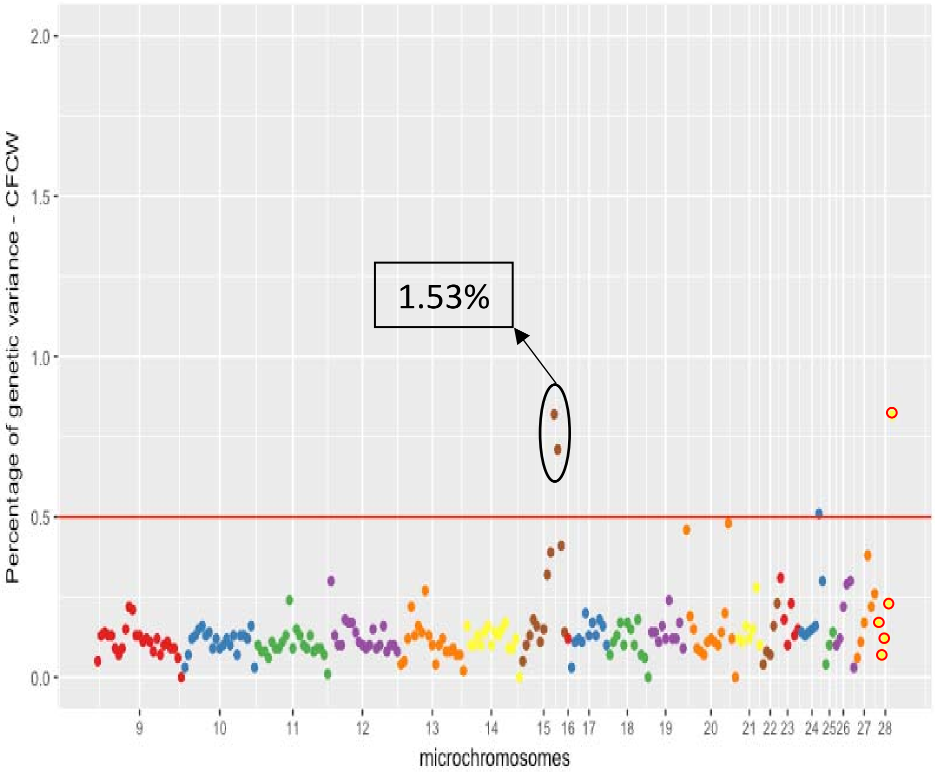
Effects:
Sex and hatch (fixed)
BW42 (covariate) – for CFCg

** Nones, K., et al, 2012. Quantitative trait loci associated with chemical composition of the chicken carcass, Animal Genetics, 43:570–576.

Results – *genomic windows CFCW*

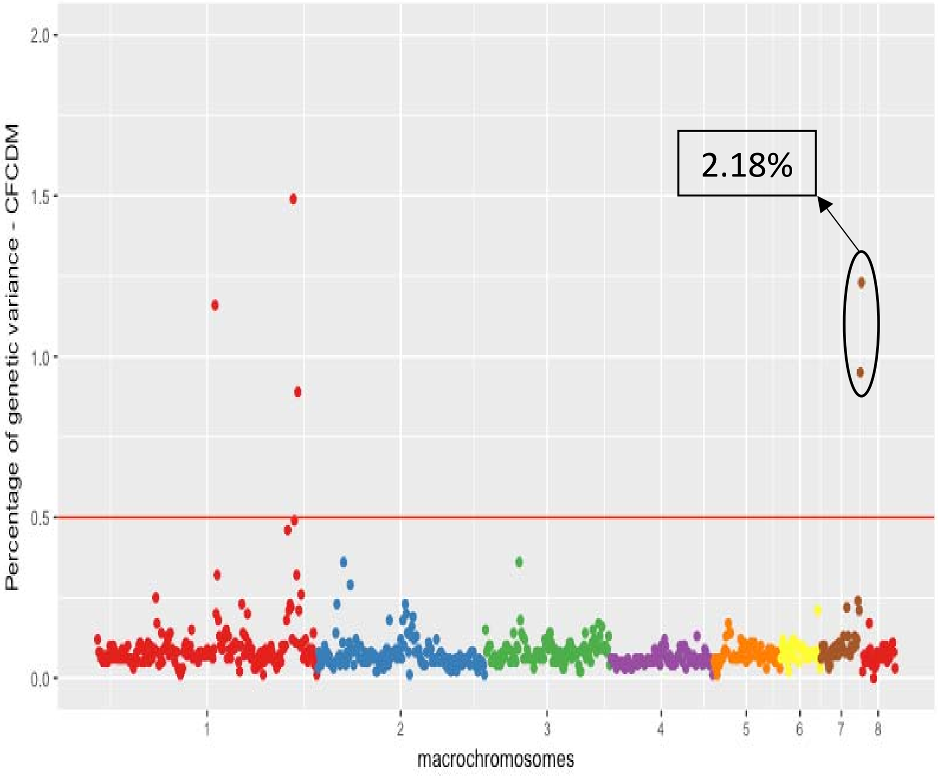
3.27%

1.31%

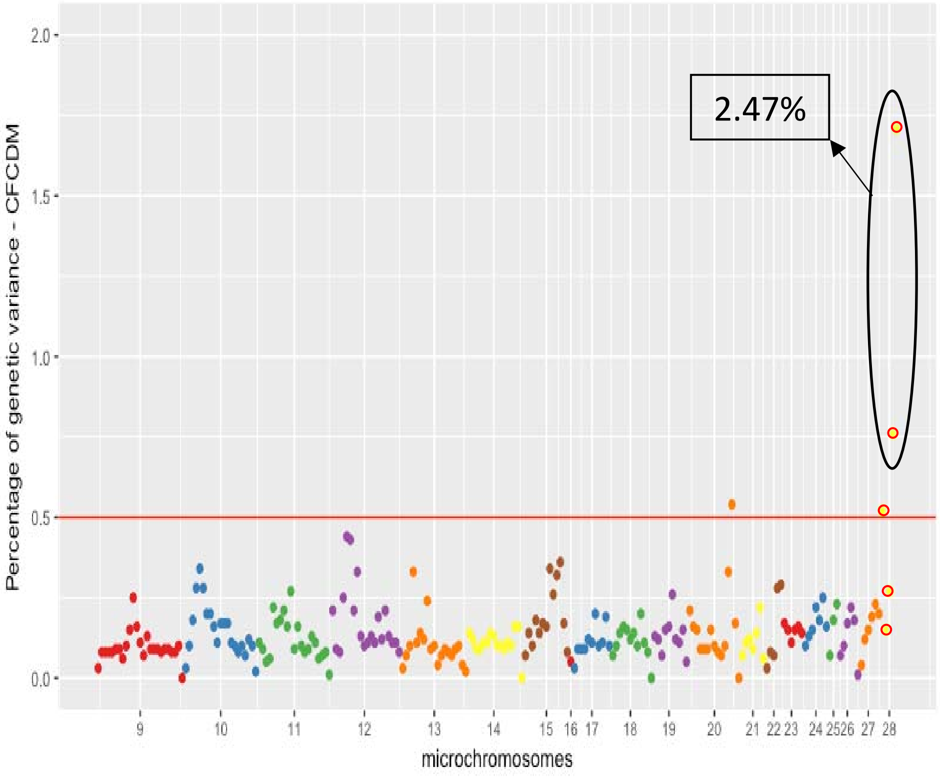


1.53%

Results – *genomic windows CFCDM*

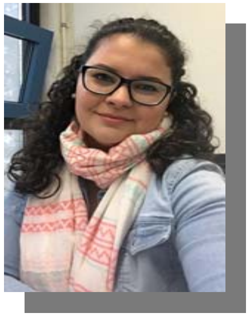


2.18%



2.47%

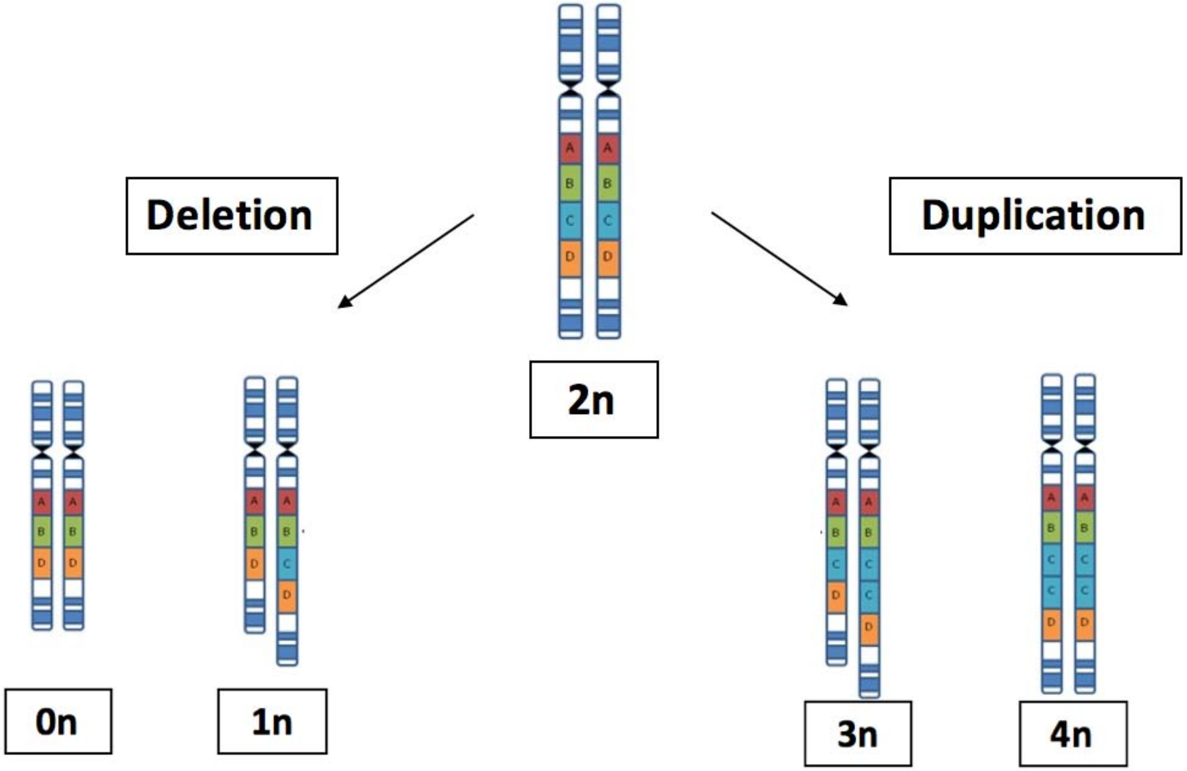
➢ Identification of inherited copy number variations (CNVs) located within the genomic windows associated (GGA7)



Thaís F. Godoy
PhD candidate
University of São Paulo

Copy Number Variation

- A segment of DNA (≥ 1kb) that is present at a variable copy number in comparison with a reference genome. (Feuk et al., 2006)



Deletion

Duplication

2n

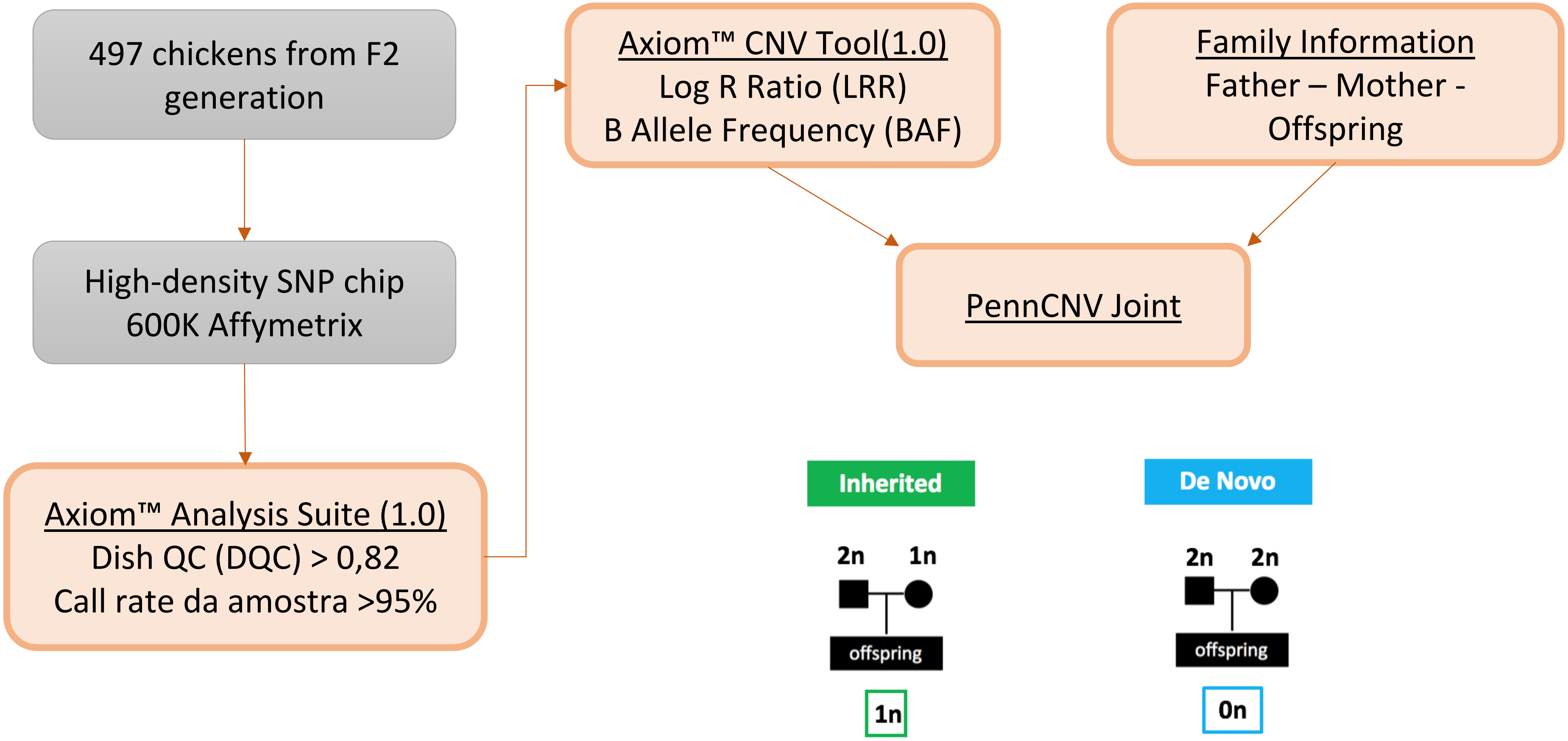
0n

1n

3n

4n

Methods



497 chickens from F2 generation

High-density SNP chip
600K Affymetrix

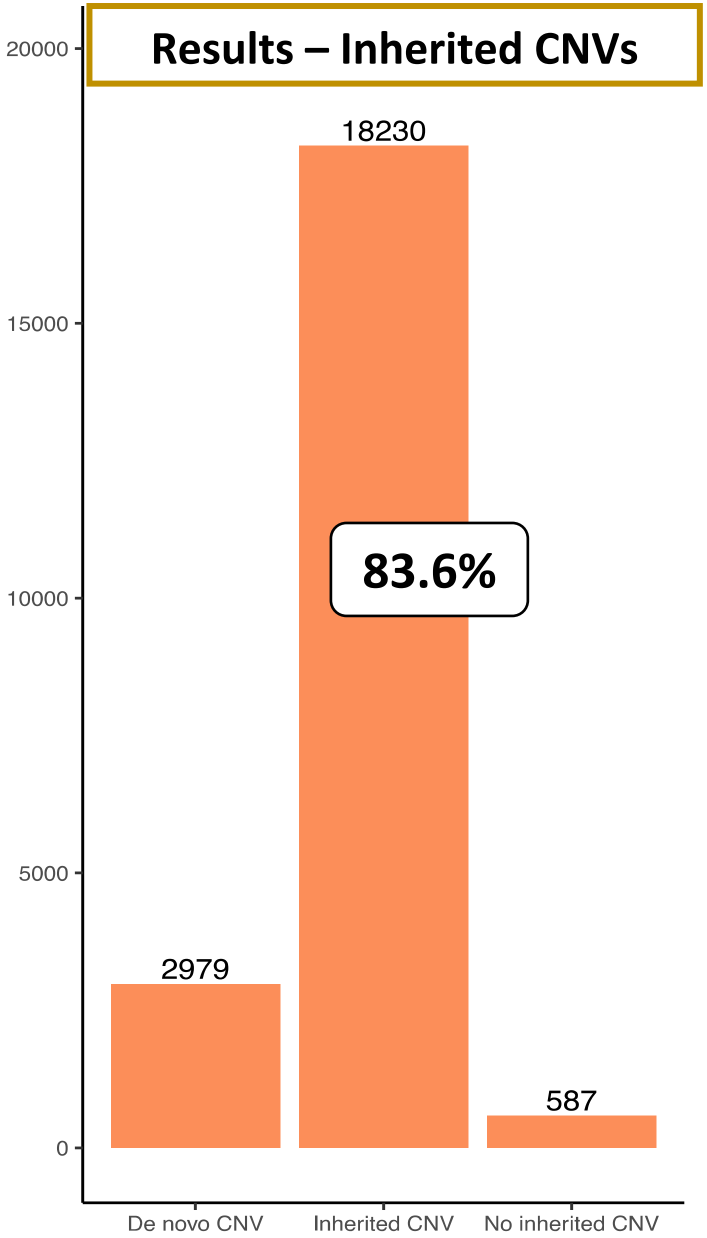
Axiom™ Analysis Suite (1.0)
Dish QC (DQC) > 0,82
Call rate da amostra >95%

Axiom™ CNV Tool(1.0)
Log R Ratio (LRR)
B Allele Frequency (BAF)

Family Information
Father – Mother - Offspring

PennCNV Joint

Results – Inherited CNVs



83.6%

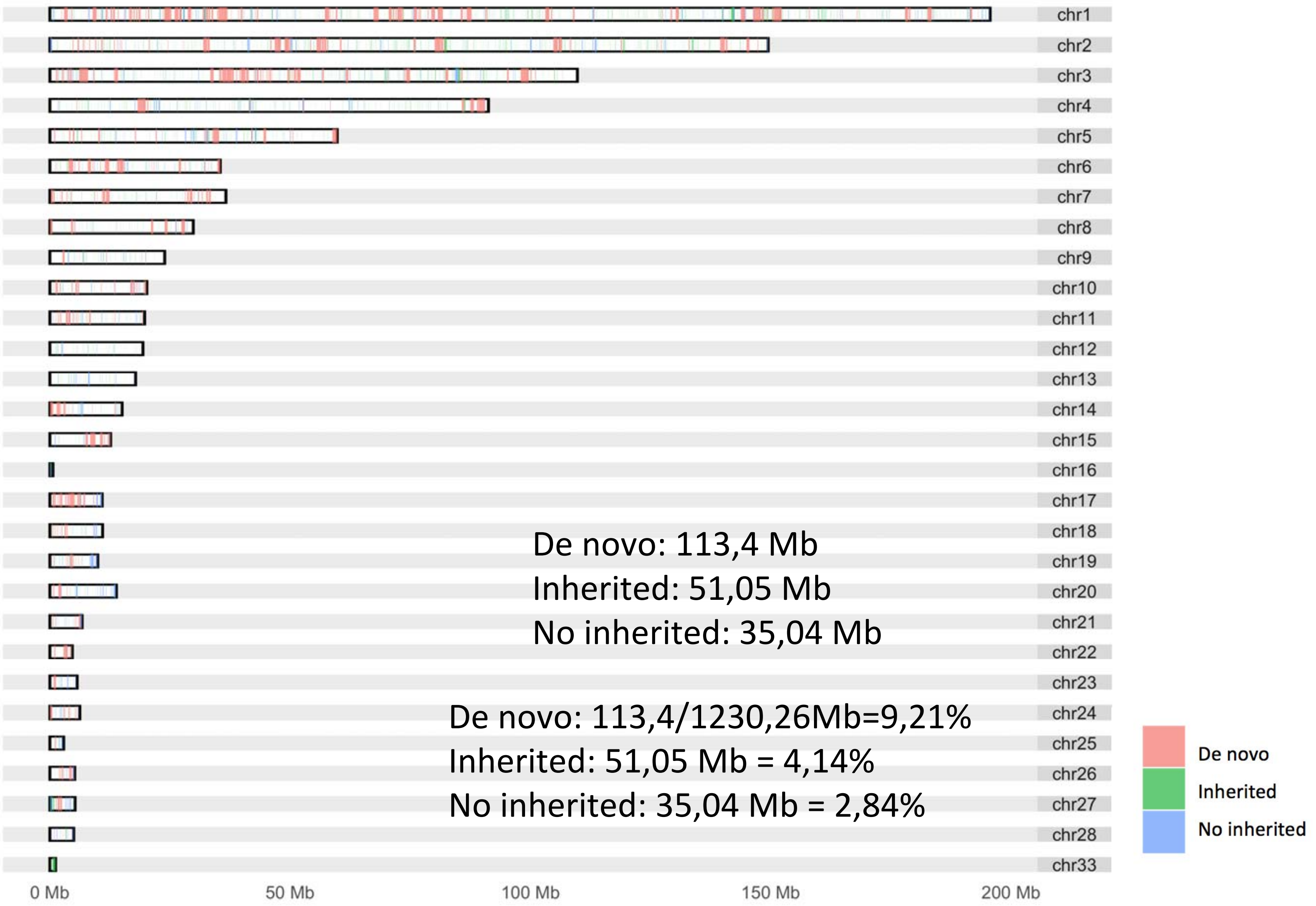
Total: 21,797 CNVs

<u>INHERITED CNVs</u>
16,764 Duplication
1,466 Deletion

18,230 inherited CNVs

2,716 inherited CNVRs

Size of inherited CNVRs
51.05 Mb (4.14% of genome)



De novo: 113,4 Mb
Inherited: 51,05 Mb
No inherited: 35,04 Mb

De novo: 113,4/1230,26Mb=9,21%
Inherited: 51,05 Mb = 4,14%
No inherited: 35,04 Mb = 2,84%

➢ Identification of selection signatures (Fst) located within the genomic windows associated (GGA7).

Clarissa Boschiero
Post-Doc
Noble Research Institute

Boschiero C, Moreira GCM, Gheyas AA, Godoy TF, Gasparin G, Mariani PDSC, et al. *Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines*. **BMC Genomics 2018**; doi: 10.1186/s12864-018-4444-0.

Methods

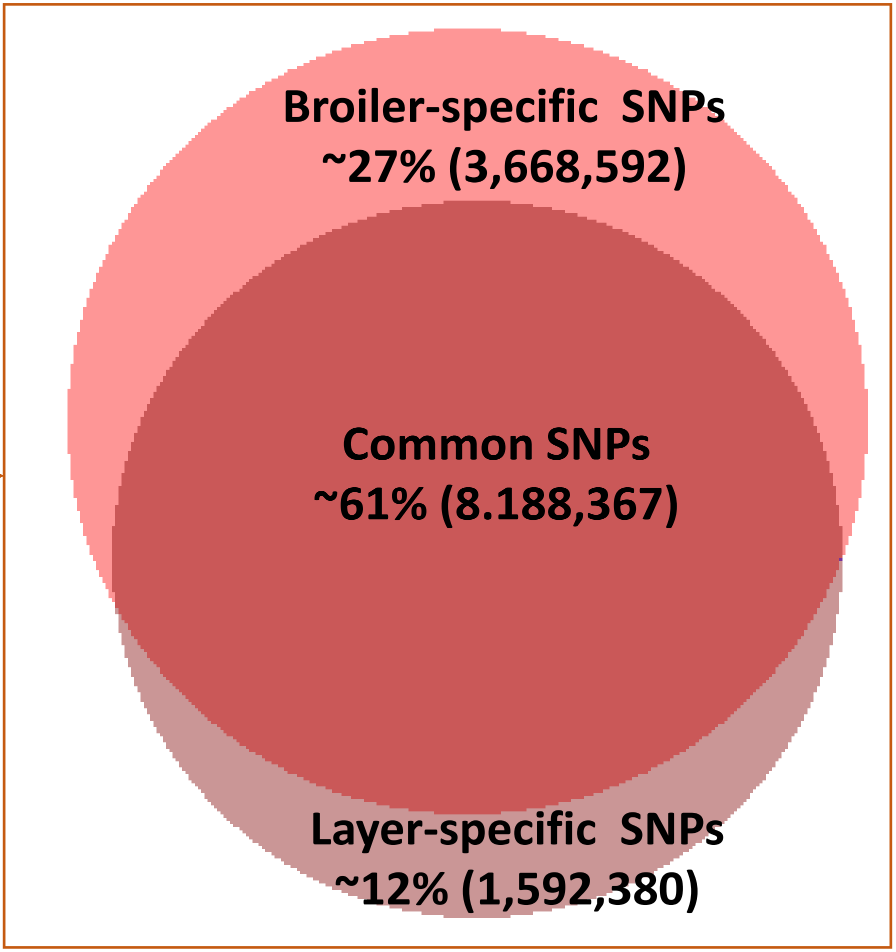
Whole-genome ressequencing
14 CC and 14 TT parental animals

Bowtie2 - alignment

SAMtools - SNP calling

Quality control:
Phred quality score ≥ 40
Depth of coverage ≥ 5
Depth ≤ mean + 3*SD
Forward and reverse strands
Variant in at least 3 reads
SNP clusters (> 10 SNPs in 50 bp)

Total of SNPs = 13,449,339



Methods: Identification of selection signatures (Fst)

13,449,339 (whole-genome)
12,806,643 (GGA1-28)

Weir and Cockerham pairwise Fst (Fixation index) method

Scan for selection signatures - VCFtools software
Overlapping window 20 kb
Step size 10 kb
Windows top 1% Fst – candidates
Windows top 0.1% Fst - candidates with strong evidence

- 92 windows with strong evidence of signature of selection (Fst >= 0.817)
- 916 windows as candidates of signature of selection (Fst >= 0.6718)

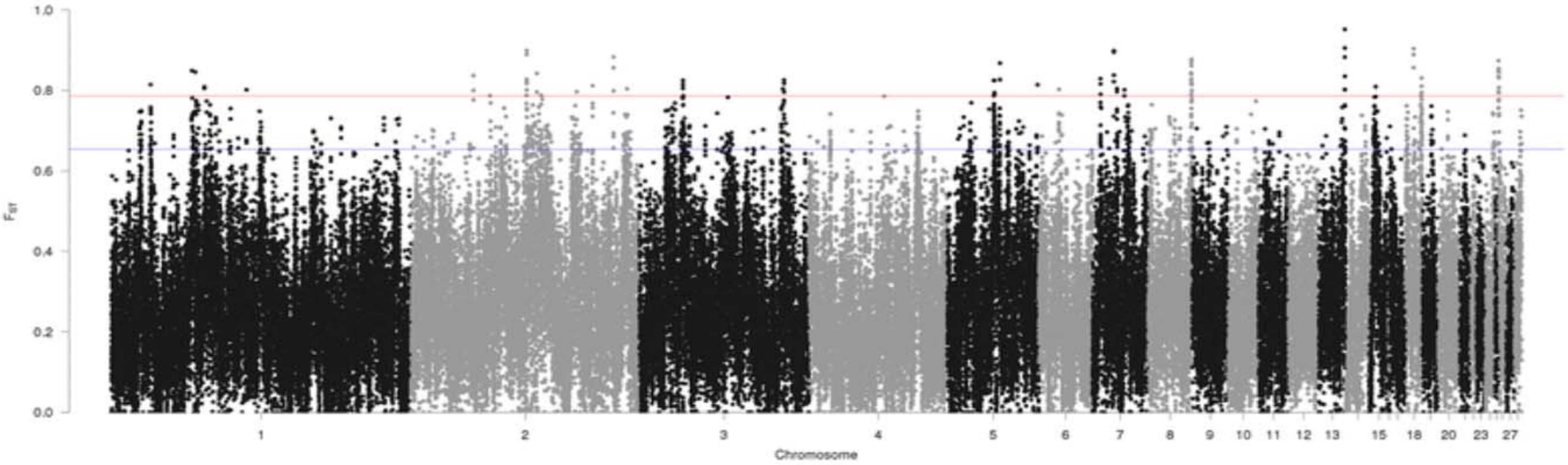
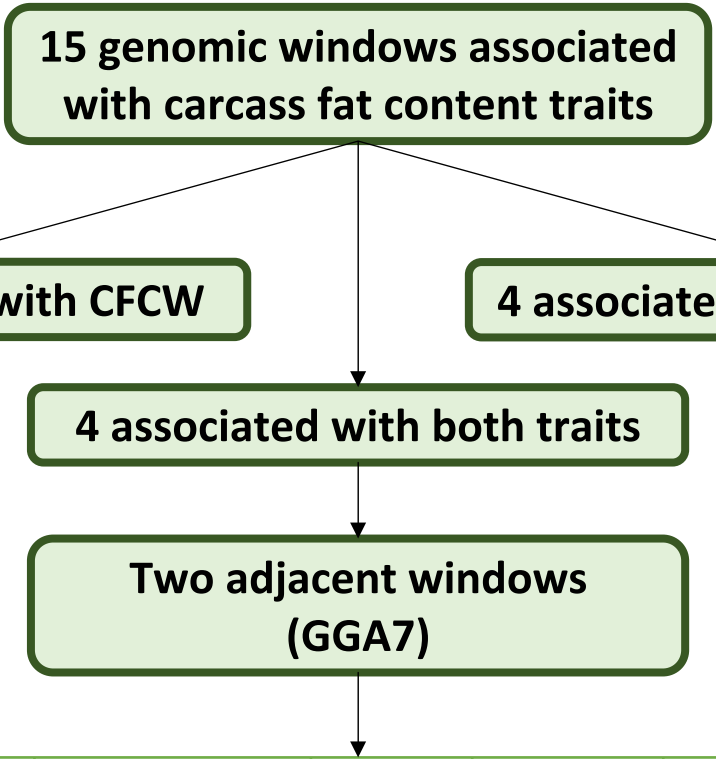


Fig. 5
Manhattan plot of genome-wide putative signature selection regions between broiler and layer lines (INDEL dataset). Blue line represents the top 1% of Fst values and red line represents the 0.1% of the Fst values

- ➢ Integration of GWAS, CNV and selection signature reveals candidate genes for fat regulation in chickens



15 genomic windows associated with carcass fat content traits

7 associated only with CFCW

4 associated only with CFCDM

4 associated with both traits

Two adjacent windows (GGA7)

GGA	Start position[1]	End position[1]	#SNPs	Trait	%Var[2]	QTL overlap[3]
7	35,001,761	36,898,384	643	CFCW	1.31	QTL #2167 Park et al., 2006
				CFCDM	2.18	

[1] Map position based on Gallus_gallus-5.0 NCBI assembly

[2] % of genetic variance explained by the window.

[3] QTL overlap based on Chicken QTL db (accessed in September, 2017).

Integrative results

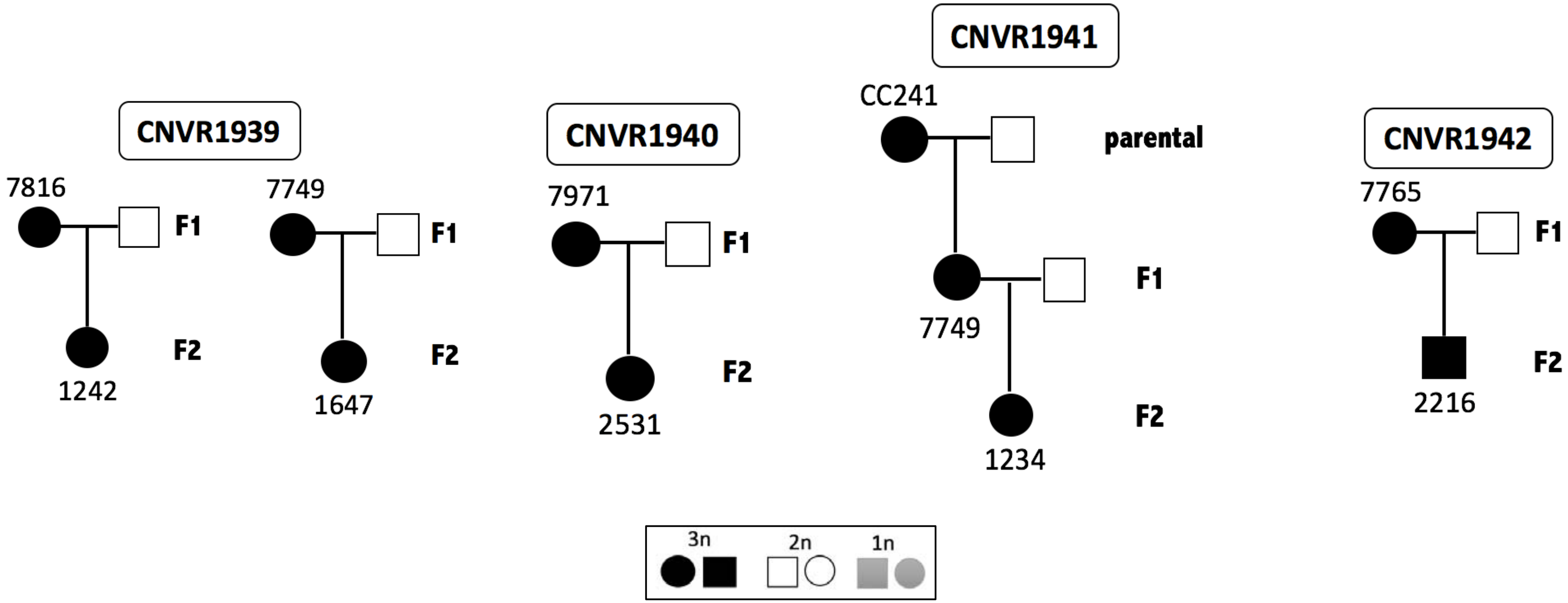
- ➢ Overlap from different approaches

Plot of one region located on GGA7 (~ 30 kb to ~ 37 kb).

GGA7: 35-37 Mb

2,716 Inherited CNVRs

4 inherited CNVRs overlapped with QTL on GGA 7



2 inherited CNVRs overlapped with selection signature and candidate genes

CNVR	GGA	Start position[1]	End position[1]	Size (bp)	State[2]	Families	Gene overlap
CNVR 1939	7	36,214,618	36,229,078	14,460	3n	2	*NR4A2*
CNVR 1940	7	36,273,167	36,283,900	10,733	3n	1	GPD2

[1] Map position based on Gallus_gallus-5.0 NCBI assembly

[2] 3n – partial duplication

[3] Numbers of family

SNP associated

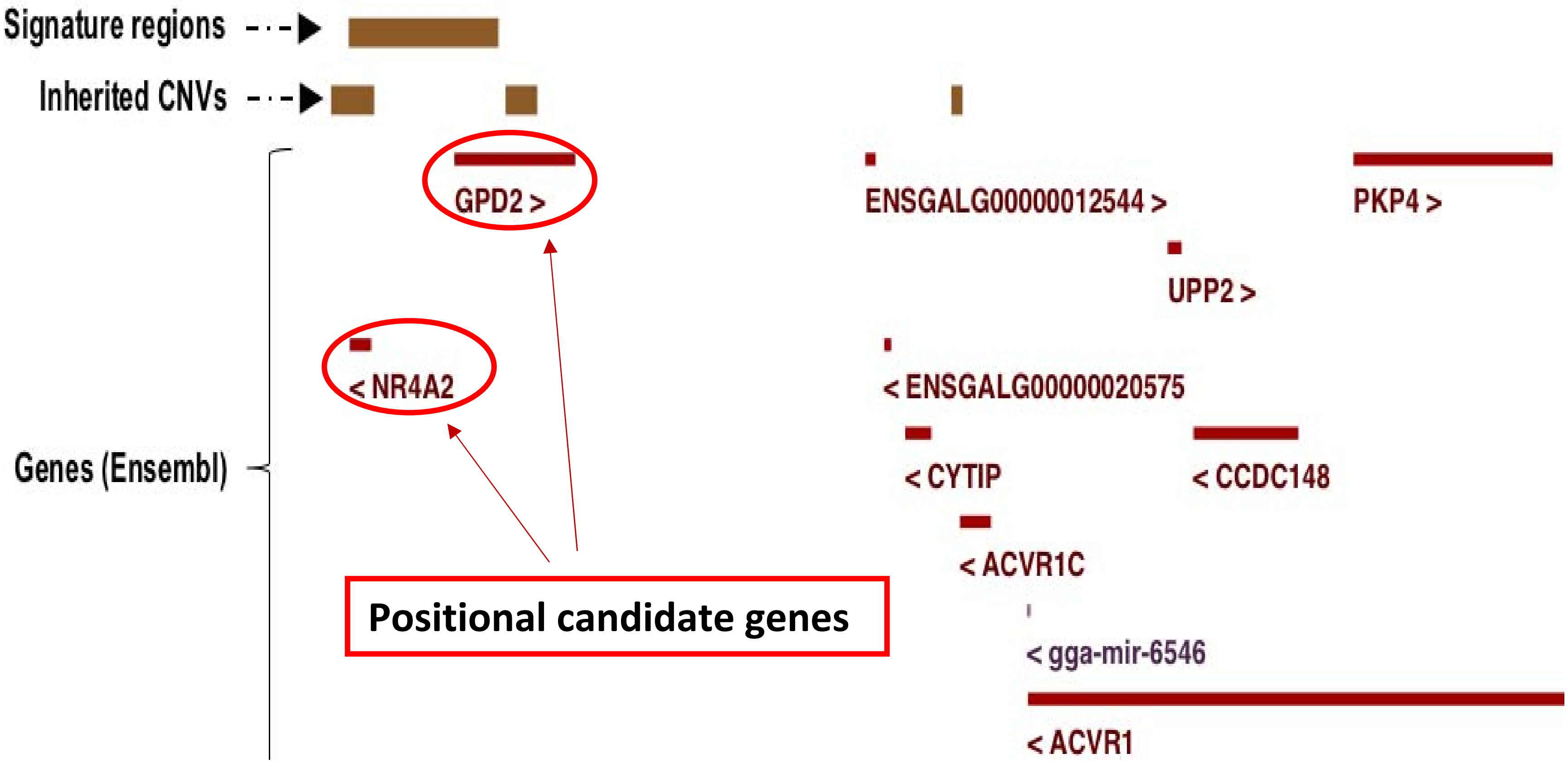
- SNP with the greater model frequency within the QTL on GGA7

Trait	SNP ID	Position	Location	Model freq.	Effect	Genotype (Ref/Alt)
CFCW	rs312848275	36,164,080	Intergenic	0.0158	-0.05	C/T
CFCDM				0.0257	-0.02	

Allele frequency				
Population	**Number of animals**	**SNP ID**	**Genotype (Ref/Alt)**	**B allele frequency**
TT parental line	14	rs312848275	C/T	T:0.615
CC Parental line	14			T.0.072
F2 chicken population	451			**T**:0.146
TT Broiler reference population	1,287			**T**:0.838

Overlap – *GWAS, CNV and selection signature*

QTL region GGA7



Plot of one small interval within the QTL mapped on GGA7 between 35 and 36 Mb showing the overlapping of the signature of selection region, inherited CNVs and putative candidate genes.

Overlap – *GWAS, CNV and selection signature*

- ***NR4A2 (nuclear receptor subfamily 4 group A member 2)***

  Involved in the regulation of hepatic glucose and, consequently, lipid metabolism (Han et al., 2012)

- ***GPD2 (glycerol-3-phosphate dehydrogenase 2)***

  Mitochondrial dehydrogenase that affects gluconeogenesis and glucose homeostasis (Madiraju et al., 2014)

  ***GPD2* knock-out mice causes reduction of 40% in the weight of white adipose tissue (Brown et al., 2002)**

Brown, L. J. et al. Normal thyroid thermogene Brown et al. (2002)** sis but reduced viability and adiposity in mice lacking the mitochondrial glycerol phosphate dehydrogenase. J. Biol. Chem. 2002; 277: 32892-32898.

**Han, Y.-F., & G.-W. Cao, 2012. Role of nuclear receptor NR4A2 in gastrointestinal inflammation and cancers. World Journal of Gastroenterology : WJG, 18(47): 6865–6873.

**Madiraju, A. K. et al. Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. Nature 2014; 510(7506): 542–546.

Search for potentially causative mutations

Whole-genome ressequencing
14 CC and 14 TT parental animals

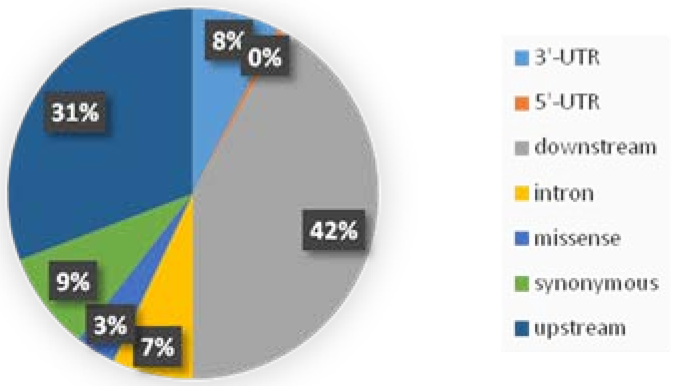
Bowtie2 - alignment

SAMtools - SNP calling

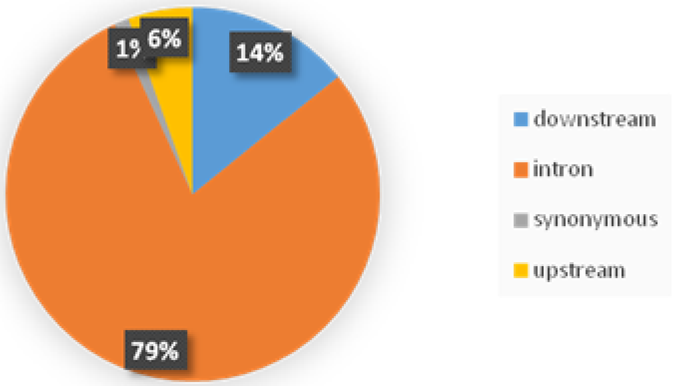
Quality control:

Phred quality score ≥ 40
Depth of coverage ≥ 5
Depth ≤ mean + 3*SD
Forward and reverse strands
Variant in at least 3 reads
SNP clusters (> 10 SNPs in 50 bp)

➢ ***NR4A2* (198 SNPs)**



➢ ***GPD2* (628 SNPs)**



Search for potentially causative mutations

- Variants located in coding regions can lead to phenotypic changes.

- To predict whether SNPs that caused changes in amino acids were tolerant or not (may affect the function of the gene product), we calculated the SIFT score.

Gene Name	**SNP ID**	**GGA**	**Position[1]**	**SIFT score**	**Amino acid changed**
NR4A2	g.36224286>C/T	7	36,224,286	deleterious (0)	Val/Met
	g.36225242>G/T	7	36,225,242	deleterious (0)	Arg/Ser
	g.36225278>C/T	7	36,225,278	deleterious (0.01)	Val/Met

[1] Position based on *Gallus_gallus* 5.0 assembly.

- Further studies may be performed to validate these mutations and check their effect.

Conclusions

- GWAS was successful in identifying regions of the genome that control fat deposition
- CNV and signature of selection complemented the analysis confirming results and revealing new regions
- Integration of the methodologies allowed the identification of candidate gens important for fat deposition in chicken

Acknowledgment

USP ESALQ

Embrapa

ROSLIN

FAPESP

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

IOWA STATE UNIVERSITY

USP
University of São Paulo

CAPES

WAGENINGEN
UNIVERSITY & RESEARCH

Dr. Luiz Lehmann Coutinho
llcoutinho@usp.br
+55 19 3429-4248
http://genfis40.esalq.usp.br/multi/

Final thoughts on gene and animal breeding

- At first, no gene or genome information
- Candidate genes
- Genome selection
- No need for gene information
- Genome editing
- Need for gene information

Research Project - Thematic Grant

Identification of loci of interest for poultry production

Principal Investigator / Grantee: Luiz Lehmann Coutinho

Contributors and associated researchers:

Clarissa Boschiero (The Samuel Roberts Noble Foundation, US)

Mônica C. Ledur (Embrapa Suínos e Aves, Brazil)

Ana Silvia A.M.T. Moura (FMVZ/UNESP, Brazil)

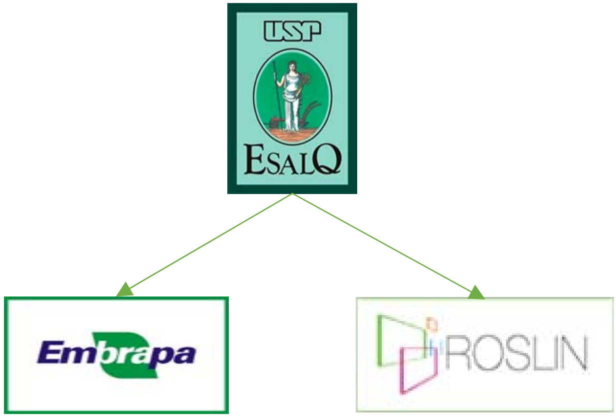
Gabriel R.A. Margarido (ESALQ/USP; Brazil)

David W. Burt (Roslin Institute, UK)

Almas A. Gheyas (Roslin Institute, UK)

Danísio Prado Munari (FCAV/UNESP; Brazil)

James M. Reecy (ISU, United States)



Funding agency: FAPESP (#2014/08704-0)

Research Project - Thematic Grant

Identification of loci of interest for poultry production

Main goal:

The objective of this project is the identification of genomic regions in the chicken that control important economical traits.

Specific goals:

➢ GWAS of 1,000 chickens from Brazilian experimental lines with a dense SNP genotyping array

➢ Genome re-sequencing of 84 chickens from 4 distinct experimental lines (broiler, white egg layer, white egg layer control, and brown egg layer);

➢ Identification of CNVs from SNP genotype array (genotypic data);

➢ Identification, annotation and validation of genetic variations (SNP, INDELs and CNVs) from sequencing data;

➢ Identification of selective sweeps.

➢ Construction of a catalog of genetic variations for the Brazilian experimental lines.

Integration of GWAS, CNV and selection signature reveals candidate genes for fat regulation in chickens

Main goal:

identify genomic windows and putative candidate genes for carcass fat content.

Specific goals:

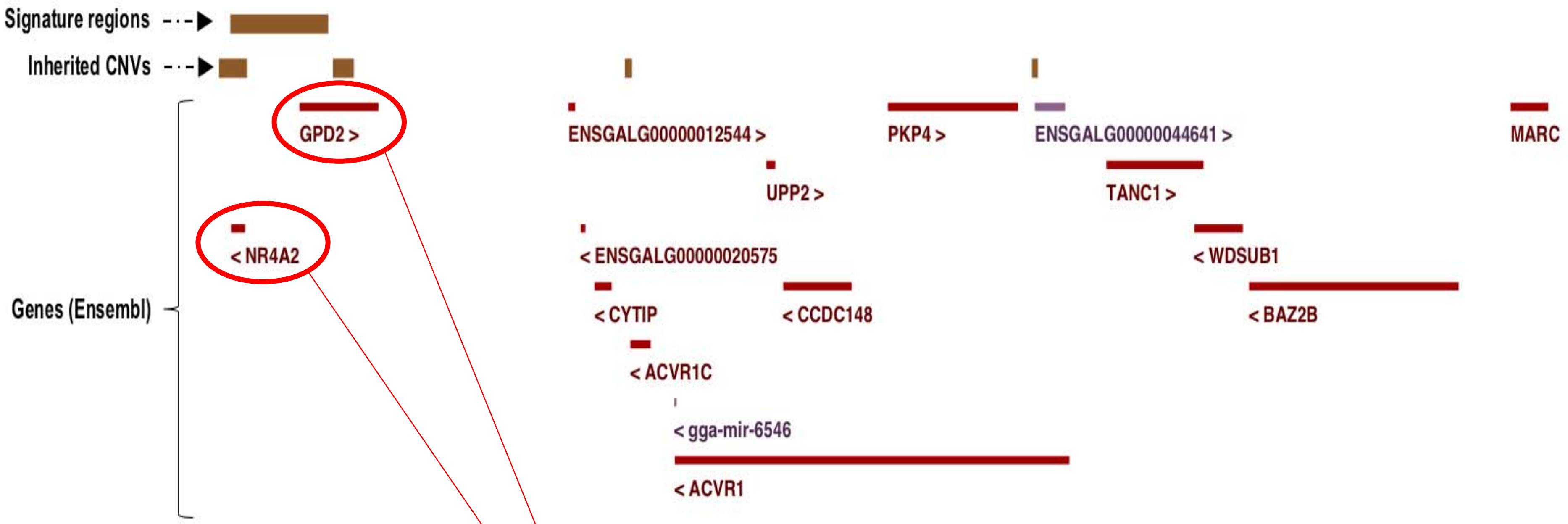
➢ Identification of 1 Mb genomic windows associated with carcass fat content traits in chickens.

➢ Identification of inherited copy number variations (CNVs) located within the genomic windows associated.

➢ Identification of selection signatures (identified by Fst method) located within the genomic windows associated.

➢ Identification of positional candidate genes for carcass fat content in chickens.

Overlap – *GWAS, CNV and selection signature*



Plot of one small interval within the QTL mapped on GGA7 between 35 and 36 Mb showing the overlapping of the signature of selection region, inherited CNVs and putative candidate genes.

Positional candidate genes

2,716 Inherited CNVRs

11 inherited CNVRs overlapped with QTL

2 inherited CNVRs overlapped with genes

CNV	GGA	Start position[1]	End position[1]	Size (bp)	State[2]	Frequency[3]	Gene overlap
CNV 1	7	36,214,618	36,229,078	14,460	3n	2	*NR4A2*
CNV 2	7	36,273,167	36,283,900	10,733	3n	1	GPD2

[1] Map position based on Gallus_gallus-5.0 NCBI assembly

[2] 3n – partial duplication

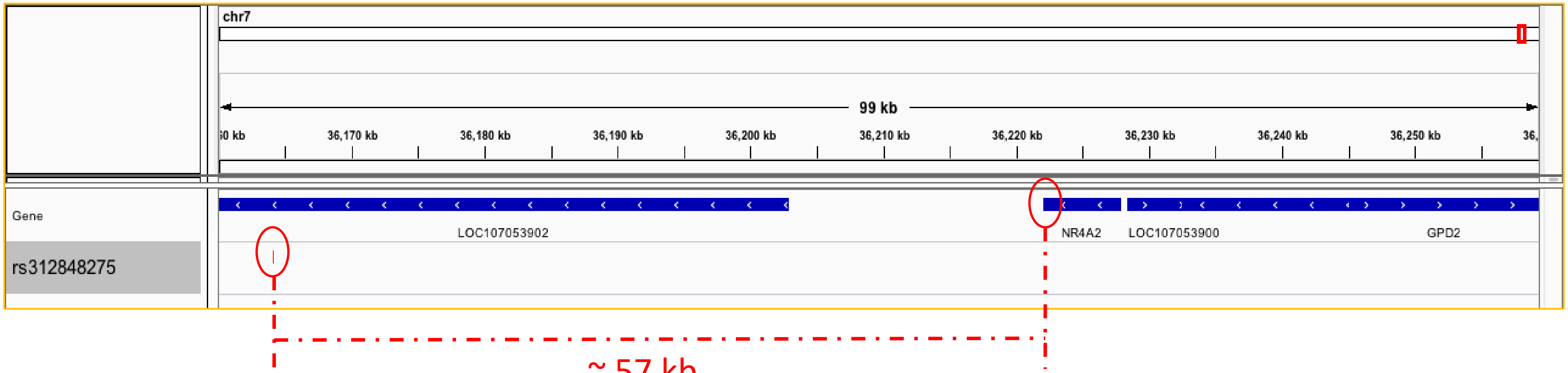
[3] Numbers of family

SNP associated

- ➢ SNP with the greater model frequency within the QTL on GGA7

Trait	SNP ID	Position	Location	Model freq.	Effect	Genotype (Ref/Alt)
CFCW	rs312848275	36,164,080	Intergenic	0.0158	-0.05	C/T
CFCDM				0.0257	-0.02	

Allele frequency				
Population	**Number of animals**	**SNP ID**	**Genotype (Ref/Alt)**	**B allele frequency**
TT parental line	14	rs312848275	C/T	T:0.615
CC Parental line	14			T.0.072
F2 chicken population	451			**T**:0.146
TT Broiler reference population	1,287			**T**:0.838



~ 57 kb