



Development of an automated quality control pipeline to facilitate the reporting of major gene genotypes

Katie Quigley, Thomas Browne, Karl O' Connell, Paul Flynn, Ross D Evans, Michael P Mullen

ICAR 2023

Date: 24th May 2023



An Roinn Talmhaíochta,
Bia agus Mara
Department of Agriculture,
Food and the Marine

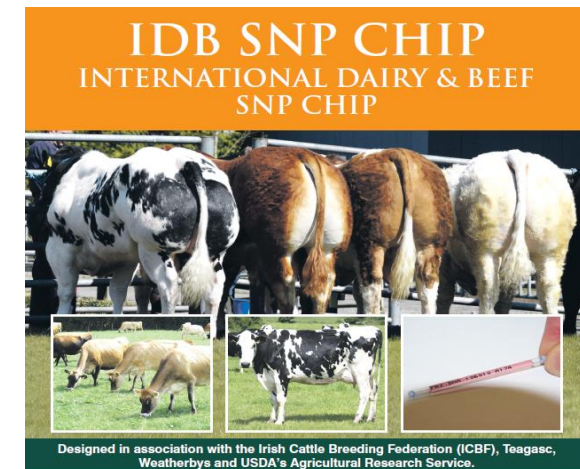


AgTech - it's in our DNA



Background

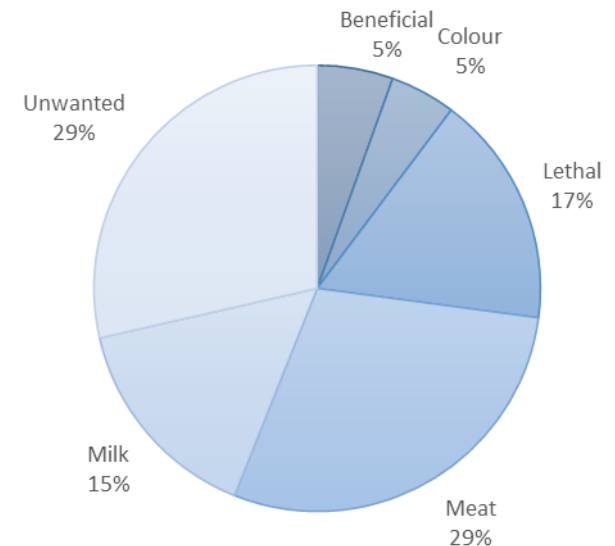
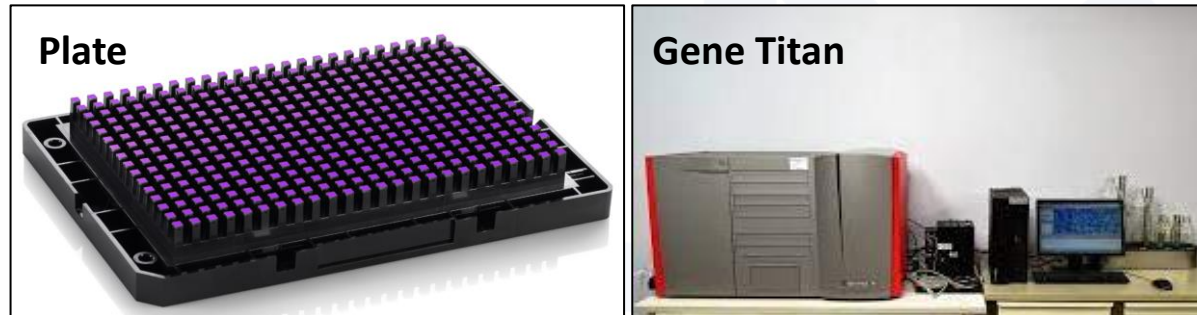
- Irish Cattle Breeding Federation (ICBF) - National cattle database
- >3 million genotypes
 - >2.6M with beef sire
 - >400K with dairy sire
- The International Dairy and Beef SNP Chip (IDB SNP Chip)
 - Parentage verification
 - Genomic evaluations for dairy and beef
 - Research
 - Calling of major genes



International Dairy and Beef SNP Chip (IDB)

- Five iterations to date
- IDBv5 chip produced on Affymetrix/ThermoFisher technology
 - 384 samples processed at a time = plate

IDB version	Provider	SNPs	Genotypes
IDBv1	Illumina Inc.	16k	30k
IDBv2	Illumina Inc.	16k	150k
IDBv3	Illumina Inc.	53k	1.2m
IDBv4	ThermoFisher	52k	400k
IDBv5	ThermoFisher	52k	1.1m



- ~165 major genes (MG) on IDBv5
 - Dairy and beef breeds
 - Classed into major gene categories

Current Process for Reporting Major Genes

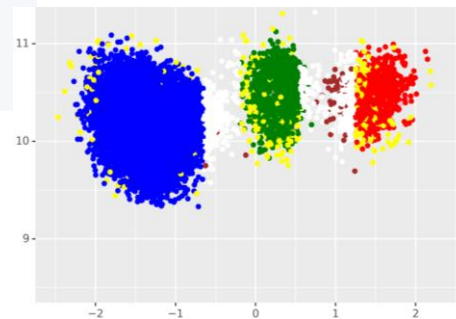
- Commercial service provider (Weatherby's Scientific) currently handles major gene analysis and reporting
- Any AI company, herd-book or individual breeder can submit request
- Manual process, involves checking individual animal on output from genotyping process
- Cost involved per major gene reported irrespective of royalty status



Major Gene Pipeline

Aim: To develop an automated pipeline in ICBF to facilitate large scale routine reporting of major genes

- Focused on IDBv5 genotypes
- QC metrics associated with SNP, genotype and plate
- QC metrics which are informative to aid genotyping process and improve reliability of genotypes
- QC metrics reviewed regularly



Pipeline Quality Control Metrics

- *Generic QC*
- *Thermofisher QC*

- **SNP and genotype specific**
- Contrast intensity thresholds for AA/AB/BB cluster

IDBv5 genotype with CR ≥ 0.97

Clustering Separation

SNP Classifications

- **Plate specific**
- Six SNP classification classes
- Plates in SNP classes 'Call rate below threshold', 'Off target variant' or 'Other' are pushed to no call

- Genotype is compared to that of the sire/dam/trio

Mendelian Check

Confidence Score

X & Y Signal Intensities

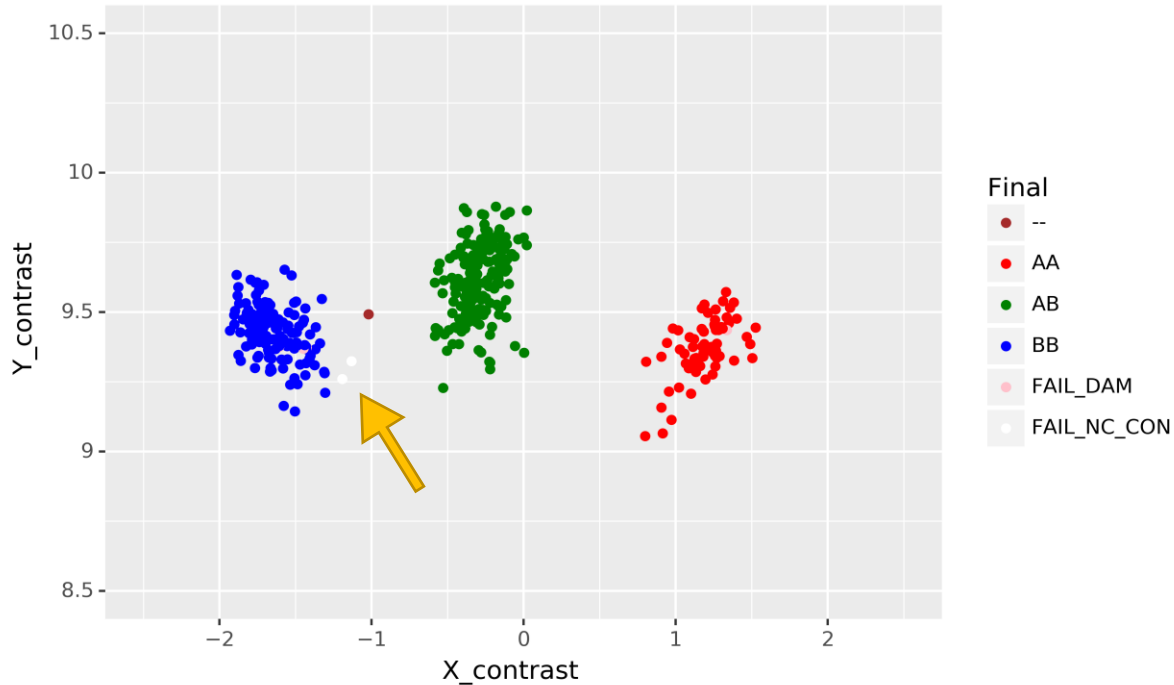
Valid genotype for MG reporting

- **SNP and genotype specific**
- Indicates the confidence of the genotype call

- **SNP and genotype specific**
- Minimum X and Y signal intensity thresholds

Good plate vs bad plate

PASS



FAIL

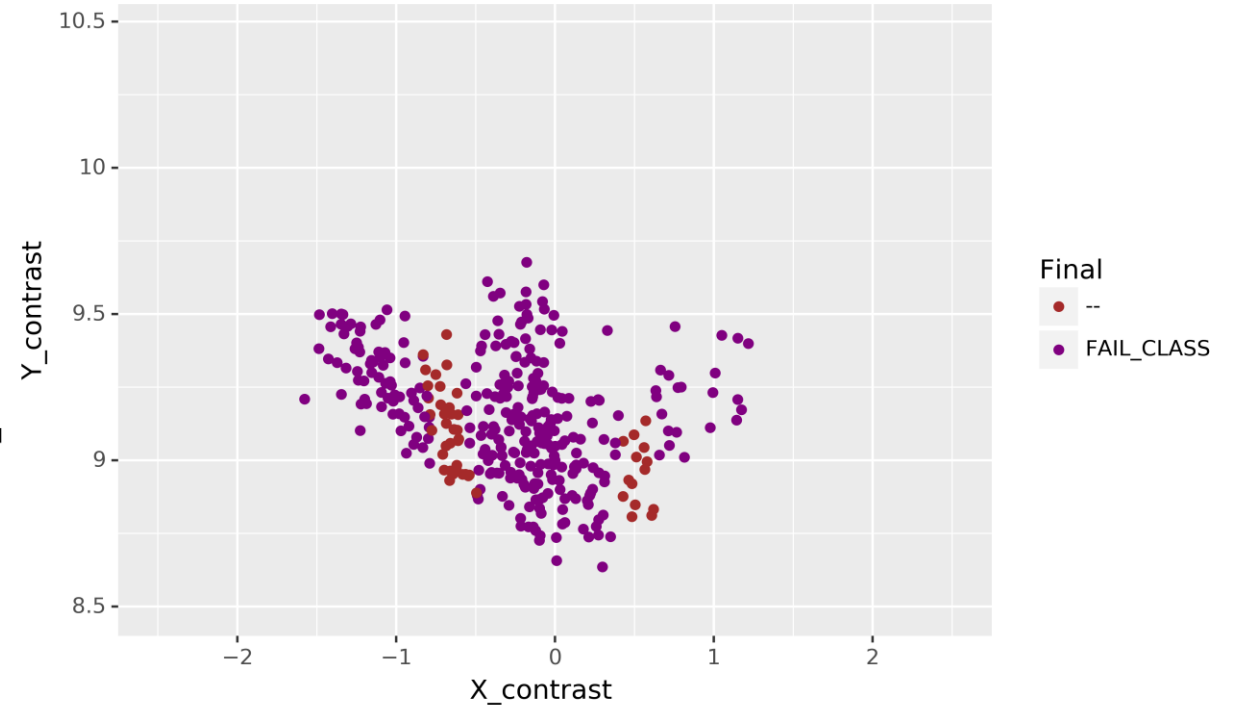


Plate classified as 'PolyHighResolution'
Two samples outside cluster separation thresholds

✓ Well-formed, well-separated, distinct clusters

Plate classified as 'Other'
Poor cluster resolution

✗ Merging, difficult to distinguish between clusters

Reported Major Genes

- Pipeline is live since November 2022
- Live genotypes being released for 9

Myostatin variants and Polled Celtic variant

- ~70 through validation pipeline – final checks before implementation.
- Extended to all non-royalty MG on IDBv5



Locus	Variant	Rs ID	Coordinates	Category
<i>MSTN</i>	<i>L64P</i>	<i>rs449270213</i>	2:6213889	Meat
<i>MSTN</i>	<i>F94L</i>	<i>rs110065568</i>	2:6213980	Meat
<i>MSTN</i>	<i>S105C</i>		2:6214012	Meat
<i>MSTN</i>	<i>nt419</i>		2:6215953	Meat
<i>MSTN</i>	<i>D182N</i>		2:6216072	Meat
<i>MSTN</i>	<i>Q204X</i>	<i>rs110344317</i>	2:6216138	Meat
<i>MSTN</i>	<i>E226X</i>		2:6216204	Meat
<i>MSTN</i>	<i>nt821del11</i>	<i>rs382669990</i>	2:6218379	Meat
<i>MSTN</i>	<i>C313Y</i>		2:6218499	Meat
<i>POLLED</i>	<i>Polled Celtic</i>		1:g.1706051_1706060del-ins1705834_1706045dup	Beneficial

Major Gene Pipeline Results

- >1 million genotypes (n=1,107,481) gone through the pipeline
- Pass Rate (PR) ranges from 91.5% to 99.1%
- Average pass rate for MG released to date is 95.75%

Locus	MG Variant	Total Passed*	Pass Rate (%)
<i>MSTN</i>	L64P	1,059,810	95.7
<i>MSTN</i>	F94L	1,038,833	93.8
<i>MSTN</i>	nt419	1,037,607	93.7
<i>MSTN</i>	S105C	1,012,891	91.5
<i>MSTN</i>	D182N	1,083,609	97.8
<i>MSTN</i>	Q204X	1,091,075	98.5
<i>MSTN</i>	E226X	1,044,293	94.3
<i>MSTN</i>	nt821del11	1,091,882	98.6
<i>MSTN</i>	C313Y	1,097,347	99.1
<i>POLLED</i>	Polled Celtic	1,046,889	94.5

***Total samples through the pipeline to date: 1,107,481**

Breed Frequencies

- Frequencies of released MG in purebred populations

Breed	N	Polled Celtic	Myostatin								
			F94L	nt821	Q204X	C313Y	nt419	L64P	D182N	S105C	E226X
Limousin	7000	0.01	0.90	0.03	0.05						
Angus	5000	1.00		0.03							
Charolais	4900	0.01	0.14		0.14						
Hereford	3100	0.15									
Holstein	2800										
Simmental	1800	0.08			0.01						
Aubrac	800		0.84	0.04							
Dexter	700	0.06									
Saler	600	0.01		0.01							
Irish Moil	300	0.80	0.01								
Shorthorn	300	0.16	0.01	0.05							0.05
Blonde D'Aquitane	200		0.02								
Partenaïse	200			0.91			0.01				0.05
Stabiliser	170	0.74	0.08	0.06							
Jersey	130										
Belgian Blue	120			1.00							

Reporting of MG

2) Herd-profiles on ICBF website

Animal Number:		Genotype Received:	01-JUL-22
Animal Name:		Call Rate:	.99475 ✓
Breed:	LM	Chip Type:	IDBV5 ✓
Birth Date:	30-MAR-22	Genotype Valid:	Yes ✓
Death Date:			
Sire:			
Dam:			

Show 10 rows. Showing 1 to 10 of 10 entries

Hide filters Excel PDF Print

First Previous 1 Next Last

Major Gene Type Code Quality Check Result

Major Gene	Type	Code	Quality Check	Result
Myostatin C313Y	Meat	MYO_C313Y	PASS	NO COPY
Myostatin D182N	Meat	MYO_D182N	PASS	NO COPY
Myostatin E226X	Meat	MYO_E226X	PASS	NO COPY
Myostatin F94L	Meat	MYO_F94L	PASS	SINGLE COPY
Myostatin L64P	Meat	MYO_L64P	PASS	NO COPY
Myostatin NT419	Meat	MYO_NT419	PASS	NO COPY
Myostatin NT821DEL11	Meat	MYO_NT821	PASS	NO COPY
Polled Celtic	Beneficial	POLL_C	PASS	DOUBLE COPY
Myostatin Q204X	Meat	MYO_Q204X	PASS	NO COPY
Myostatin S105C	Meat	MYO_S105C	PASS	NO COPY

Showing 1 to 10 of 10 entries

First Previous 1 Next Last

Analysis Disclaimer: This analysis has been prepared in good faith on the basis of information provided to it.

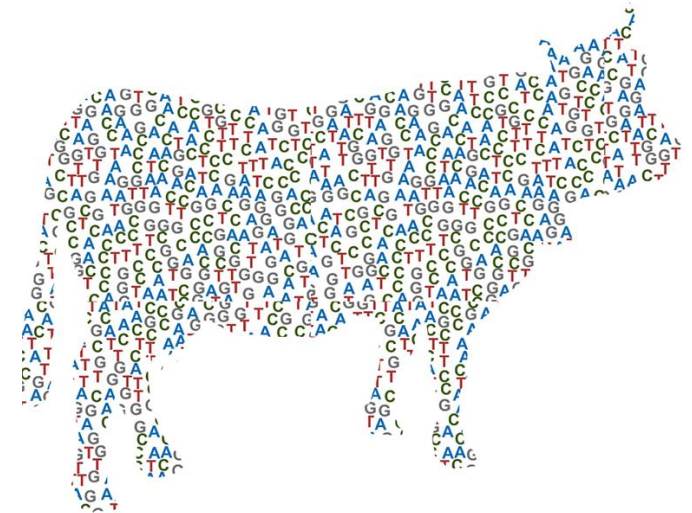
No representation or warranty expressed or implied is made or given as to the accuracy, reliability, completeness or correctness of the analysis.

No liability is accepted for any losses (whether direct or indirect), damages, costs or expenses whatsoever, incurred or arising from any use of or reliance on the analysis or the information contained in it by any person.

Result: The animal will carry 'No Copy', 'Single Copy' or a 'Double Copy' of the gene variant

Summary

- Pipeline overlays additional QC metrics to improve the confidence and reduce errors/miscalling of major gene genotypes
- Why is it of benefits to breeders?
 - Valuable information on both desirable and undesirable genes segregating within herds
 - Informed breeding decisions, carry out selective matings
- Downstream benefits
 - Herd book major gene management
 - Farmers – Mating decision support tool (ICBF - Sire Advice)
 - Genomic evaluations – increased accuracy where dubious SNP calls are censored and imputed



Our Farmer & Government Representation



An Roinn Talmhaíochta,
Bia agus Mara
Department of Agriculture,
Food and the Marine



Our AI & Milk Recording Organisations



Our Herdbooks



Acknowledging Our Members