Genomic Selection in Sheep: where to now?

John McEwan, Ken Dodds, Suzanne Rowe, Rudi Brauning, Shannon Clarke

AgResearch, Invermay, New Zealand

EAAP Belfast 29th August 2016

agresearch

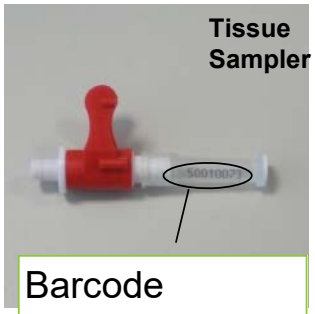
Overview

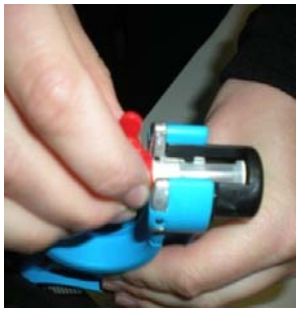
- Tools: Chips and GBS & genomic selection (GBLUP) issues
- NZ industry at present
- Example for meat eating quality
- Better phenotypes & shorter generation interval
- FAANG and precision genomics

agresearch

From tissue sample to breeding values

1





Barcode Number = EID

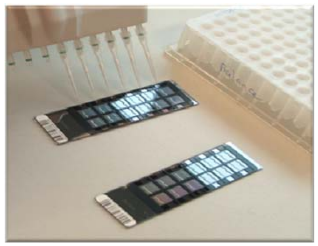
Tag number

2

Ear punch

High throughput DNA extraction

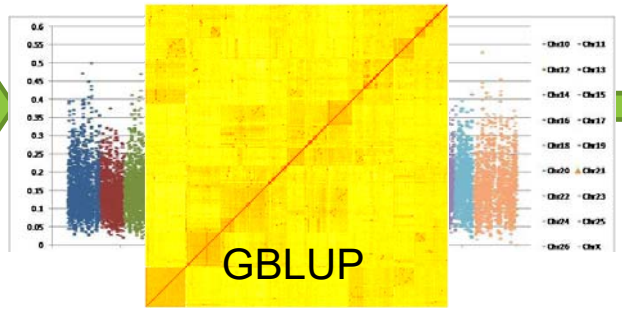
High molecular weight
260/280 > 1.8



SNP chip

or

3



BV

BV Report sent to breeder for selection decisions

What genomic tools are at our disposal?

- 50K SNP chip — 23,000 samples
- HD chip ~600K — 21,000 samples
- LD chip(s) 5, 6,15K — 47,000 samples

- *Enhanced parentage* — *new*

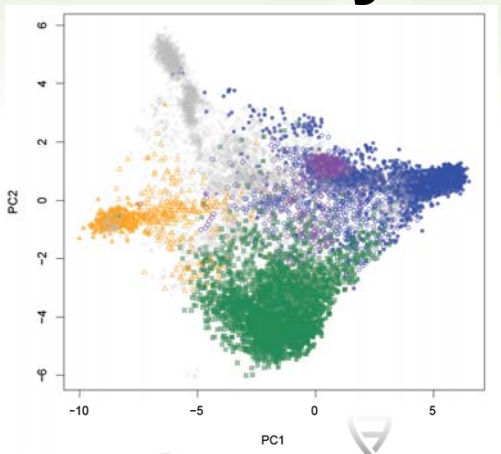
- *GBS (~80 & 561K)* — *5500 samples*

- *Genomic selection same DNA parentage €12-15*
- *Eliminate 2 stage selection*

FARMIQ SYSTEMS LIMITED | ISGC | beef+lamb new zealand GENETICS | Ministry of Business, Innovation & Employment | agresearch

Genomic selection in NZ industry: maternal breeds

- 2 stage selection
- Multibreed training population
- ~6K-15K impute to ~40K
- 22 traits, accuracy breed specific
- no PCA/breed adjustment mBVs
- Blending (Harris & Johnson) moving to single step
- Monthly …. weekly
- Chip dependent
- Most benefit: sex limited, late life, kill, … **20-60% ↑ΔG index**



Progeny equivalent numbers				
Trait	Rom	Coop	Peren	Comp
WWT	14	17	12	6
WWTM	10	9	7	6
LW8	5	5	3	2
LW12	4	3	3	2
CW	6	7	3	3
EWT	3	3	2	2
FATY	3	9	2	2
HQLY	2	6	2	3
LNLY	3	7	2	4
SHLY	4	8	3	4
LEANY	3	7	2	3
FW12	4	9	3	4
NLB	27	16	9	11
GGT21	6	1	2	1
LDAG	3	7	1	6
ADAG	4	6	1	5
FEC1	14	20	9	14
FEC2	7	6	4	4
AFEC	4	4	2	3

Most traits published: Auvray et al, Pickering et al, Phua et al, Lee et al

GBS restriction enzyme based

1. High throughput DNA extraction from ear/fin clip tissue punch

- High molecular weight DNA
- 260/280 > 1.8
- Consistent amount of DNA extracted (CV <20%)

Allflex TSU

Clarke et al., 2014 PLoS ONE

Step 2. GBS Library preparation and purification

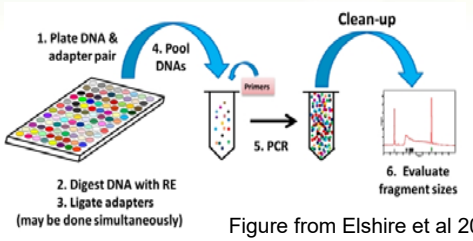


Figure from Elshire et al 2011

- Utilises Elshire *et al* 2011 GBS method with the addition of a library purification step utilising the Pippin Prep (SAGE Science) to further size select DNA sequencing library.
- Accurate nano- robotic systems employed throughout.

Step 3. Sequencing

- HiSeq 2500 V4 chemistry
- Single end reads (1x100)

QC

+ve/-ve controls

Species

Reads/bar code

Step 4. Bioinformatic and statistical analysis

agresearch

Sampling…

Allele count ☺

Traditional

Actual

T/C 1/1 T/C

…..ACG**T**ACTG……

…..ACG**C**ACTG……

T/T 0/0 */*

C/C 1/0 C/C

…..ACGCACTG……

T/C 1/0 C/C

…..ACGCACTG……

T/C 2/0 C/C

…..ACGCACTG……

…..ACGCACTG……

T/T 0/3 T/T

…..ACGTACTG……

…..ACGTACTG……

…..ACGTACTG……

T/C 0/1 T/T

…..ACGTACTG……

C/C 2/0 C/C

…..ACGCACTG……

…..ACGCACTG……

T/C 2/1 T/C

…..ACGTACTG……

…..ACGCACTG……

…..ACGCACTG……

agresearch

Dodds *et al. BMC Genomics* (2015) 16:1047
DOI 10.1186/s12864-015-2252-3

BMC Genomics

METHODOLOGY ARTICLE Open Access

Construction of relatedness matrices using genotyping-by-sequencing data

Ken G. Dodds[1*], John C. McEwan[1], Rudiger Brauning[1], Rayna M. Anderson[1], Tracey C. van Stijn[1], Theodor Kristjánsson[2] and Shannon M. Clarke[1]

- unbiased estimates of relatedness via method 1 of VanRaden (2008) adjusted to account for sequence read depth at each individual SNP location including SNPs with zero/missing reads **KGD**
- allows GBS to be applied at read depths which can be chosen to optimise the information obtained (~2 reads/SNP)
- SNPs with excess heterozygosity, often due to (partial) polyploidy or other duplications can be filtered based on a simple graphical method.

agresearch

Van Raden 2008 J Dairy Sci. 91:4414-23

KGD method - description

- VanRaden method 1

$$G = \frac{QQ'}{2\sum p_j(1-p_j)}$$

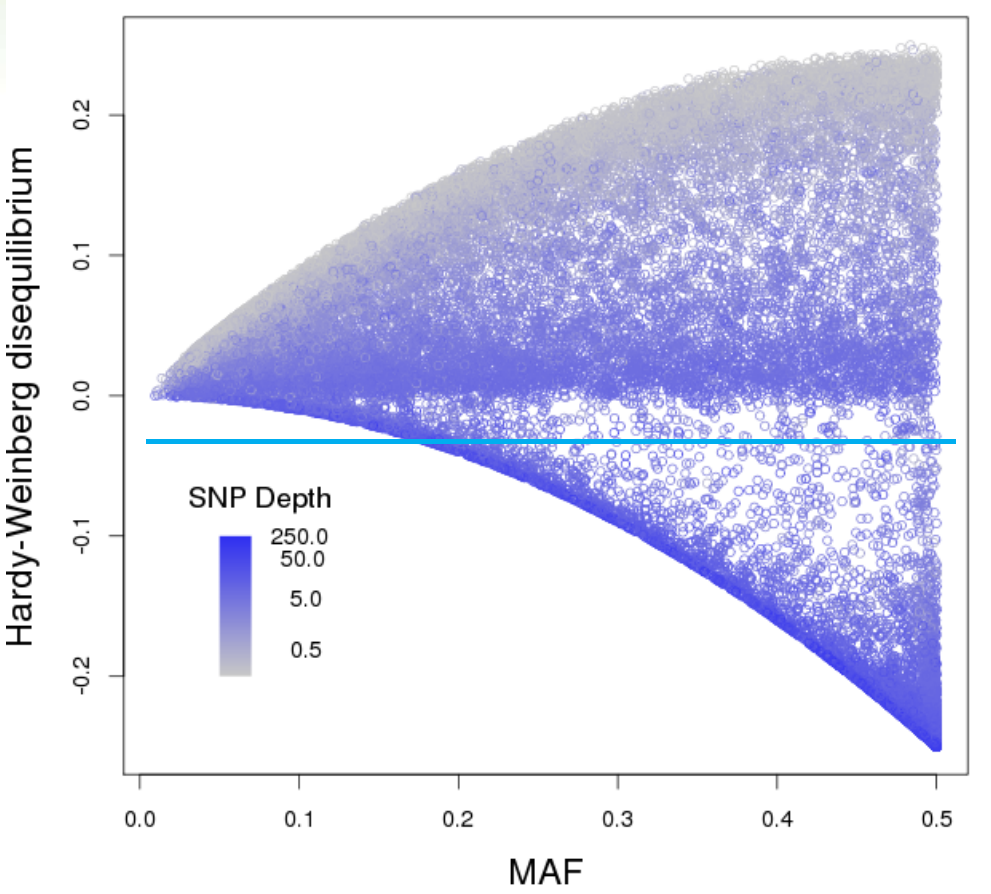
Geno - types

- Off-diagonal elements
  - Don't depend on depth
  - Σs over SNPs where both individuals scored
  - Only need 1 allele from each to estimate relatedness
- Diagonal elements
  - Need a depth correction
  - Uses SNPs with depth ≥ 2
  - Need to see both alleles for inbreeding

GBS data

agresearch

Filter SNPs use “Fin plot “- relationship of Hardy-Weinberg disequilibrium, MAF & SNP depth



- Salmon partial tetraploid
- Filter based on this plot
- Produces expected relationship

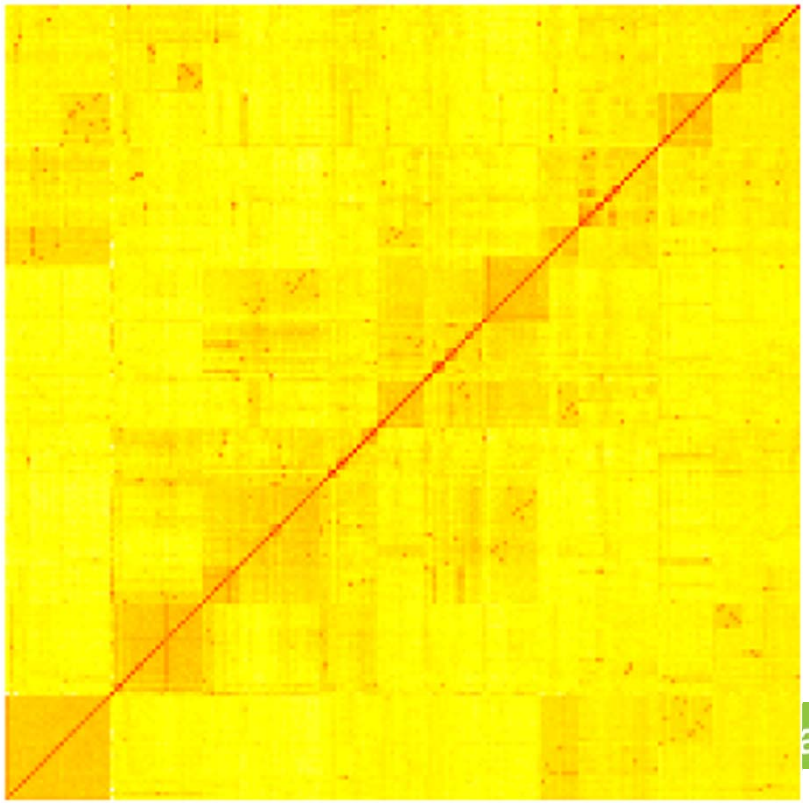
agresearch

GRM uses

- Parentage
- Inbreeding
- Breed prediction
  - & breed composition
- Co-ancestry

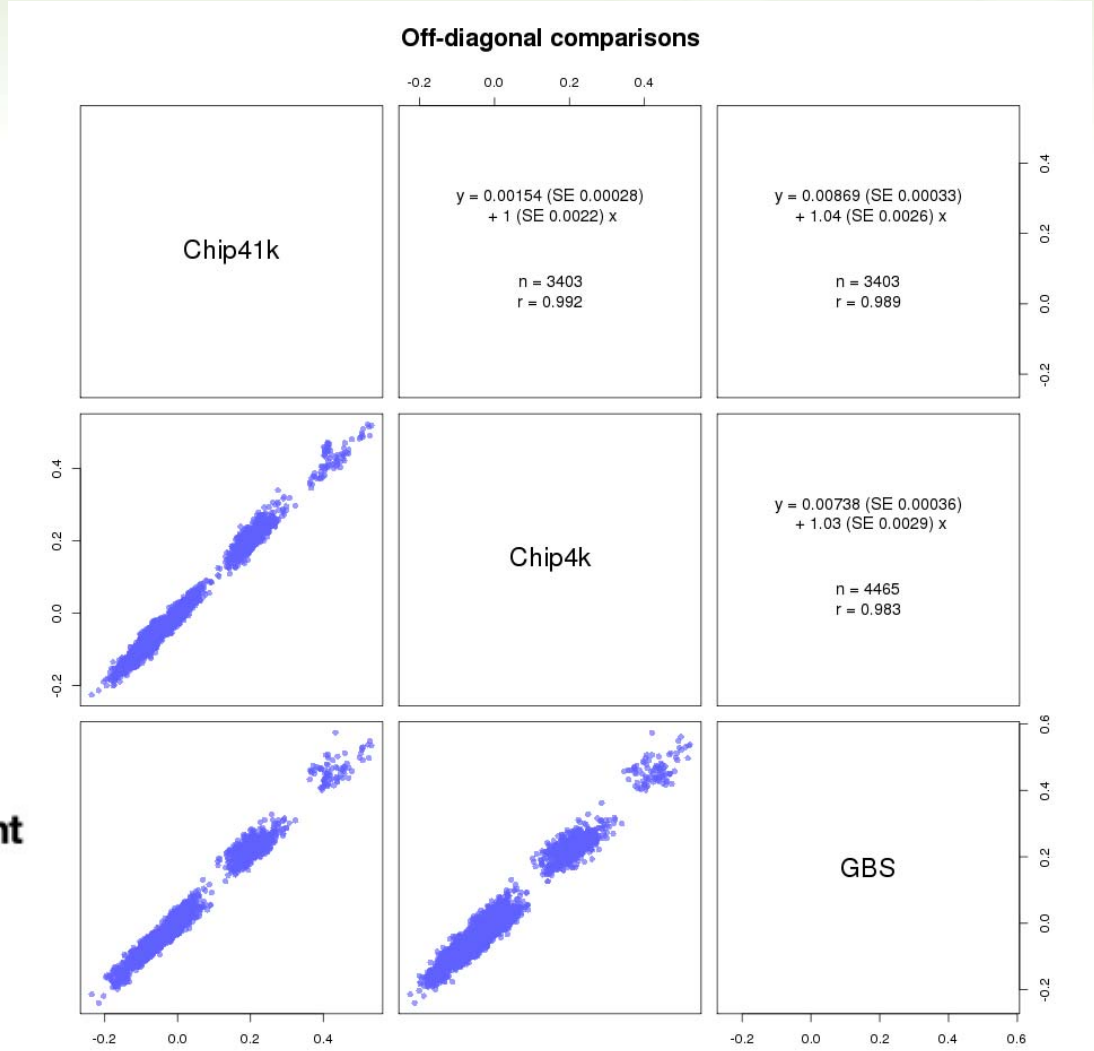
- Estimate h^2 without pedigrees
- Calculate mBVs GBLUP

- Estimate genetic diversity
  - E.g. PCA plots via GRM



agresearch

GBS vs SNP chip: sheep



GBS comparison

- 3400-4400 animals
- HD chip sub samples
- Slope =1
- Intercept =0
- R ~0.99!!!!!!!

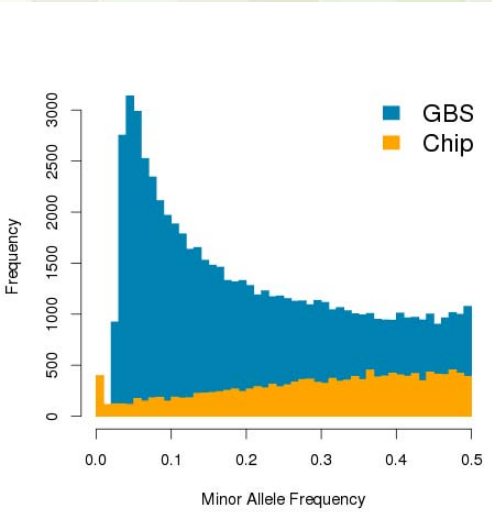
Ministry of Business, Innovation & Employment

FARMIQ SYSTEMS LIMITED

agresearch

Example-Dairy Sheep

- new dairy sheep flock with an East Friesian genetic base
- phenotypic records on 3000 ewes but no pedigree information.
- 300 ram lambs and 50 older rams available for selection candidates – which to breed from?



→ Genotype ewes and rams, generate GRM to estimate ram breeding values

- assess GBS by comparing GRMs from GBS and ISGC15k SNP chip & subsequent eBVs.

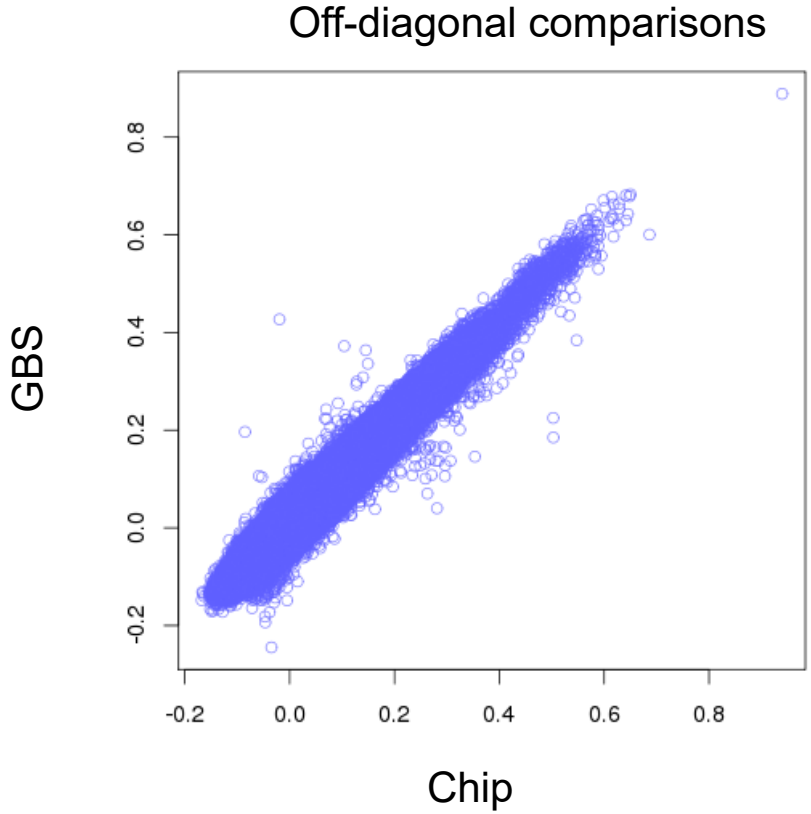
SPRING SHEEP
– MILK CO. –
New Zealand

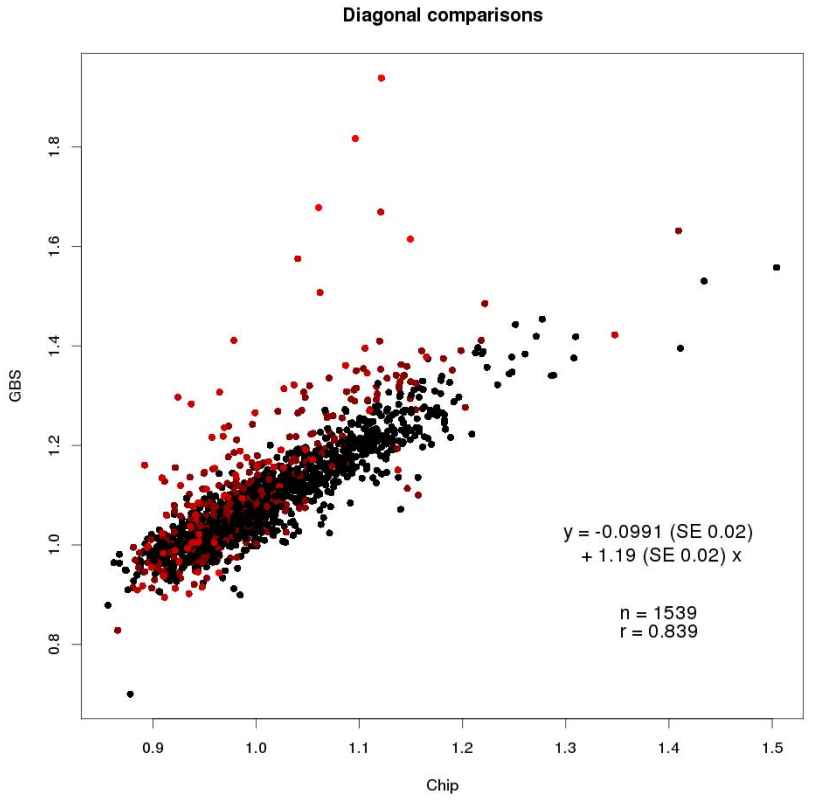
agresearch

Example: GRM

Dairy Sheep
Also 15k chip genotyped

Data courtesy of Suzanne Rowe, AgResearch



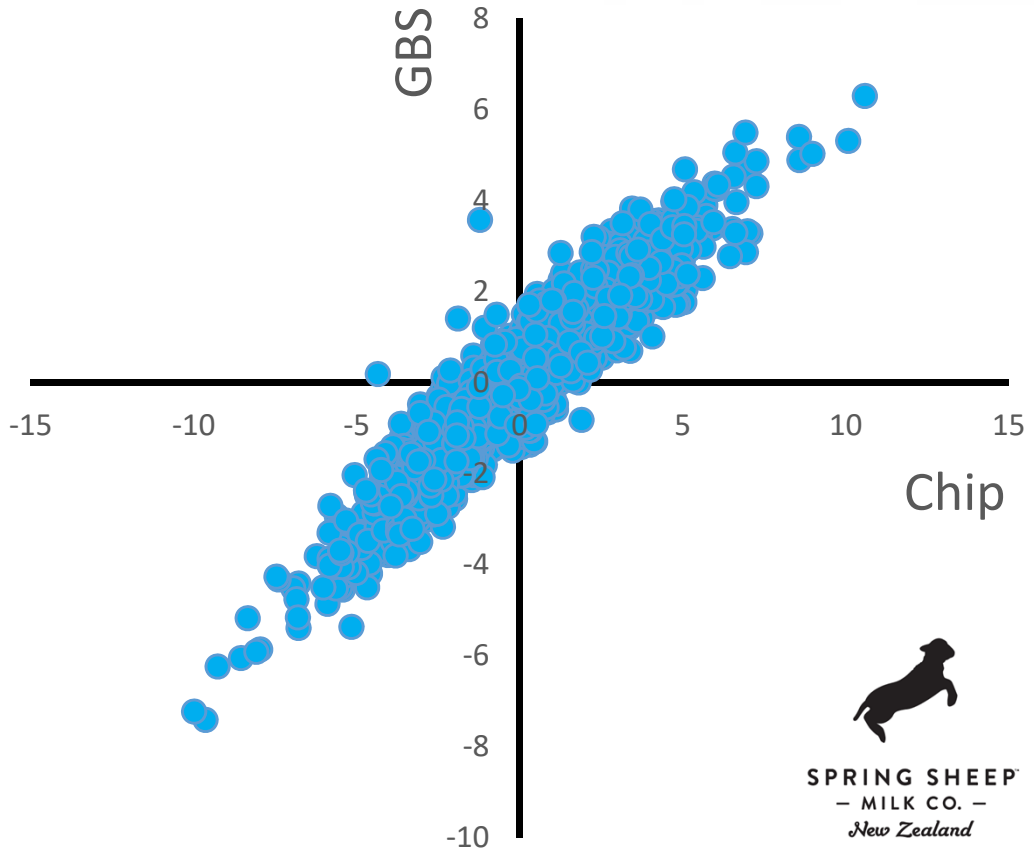


Depth
- ≤ 0.3
- 0.3 - 1
- 1 - 2
- >2

SPRING SHEEP
— MILK CO. —
New Zealand

Example: Genomic Prediction

- GBLUP Breeding values
- Milk (kg)



SPRING SHEEP
— MILK CO. —
New Zealand

agresearch

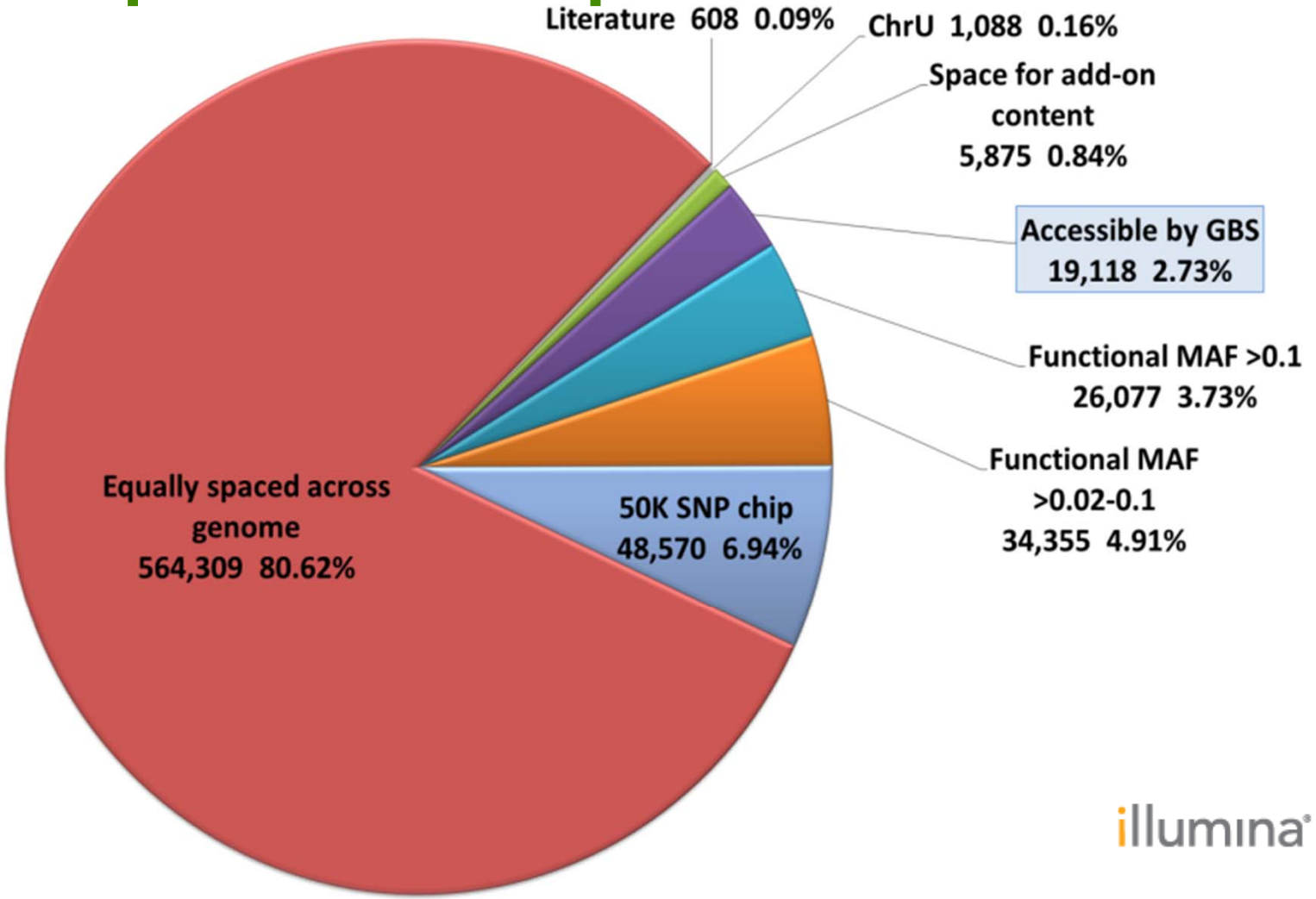
Example: lamb eating quality

- **Which traits are indicators of eating quality?**

FARMIQ SYSTEMS LIMITED

agresearch

Sheep HD SNP chip



illumina

FARMIQ SYSTEMS LIMITED

ISGC

Genetic resources:

Primera & Suftex **Lamb Supreme**

Texel and Other

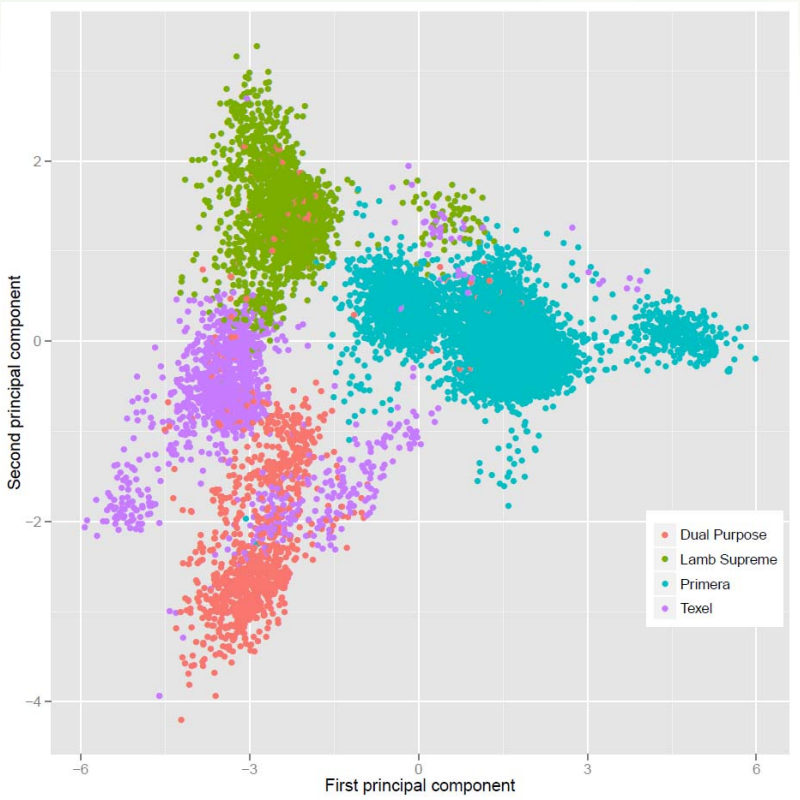
Maternal/Dual Purpose breeds

FARMIQ SYSTEMS LIMITED

agresearch

agresearch

Genomic predictions



Ne ~950

FARMIQ SYSTEMS LIMITED

agresearch

agresearch

Genomic Predictions

Eating quality traits

Trait	h^2	T set	V set	GB0	Eacc
LPH	0.14	7,867	2,213	0.33	0.33
LPHad	0.13	7,856	2,188	0.33	0.32
LMARB	0.31	8,173	2,441	0.52	0.47
LMARBad	0.31	8,160	2,391	0.52	0.47
SHF	0.26	7,740	2,409	0.28	0.43
SHFad	0.27	7,727	2,360	0.30	0.44
A24	0.17	8,167	2,301	0.31	0.36
A24ad	0.16	7,856	2,187	0.34	0.35
B24	0.14	8,160	2,282	0.29	0.33
B24ad	0.07	7,849	2,168	0.15	0.24
L24	0.18	8,052	2,079	0.32	0.37
L24ad	0.21	7,800	2,035	0.34	0.39
Average	**0.17**	**7,890**	**2,212**	**0.29**	**0.34**

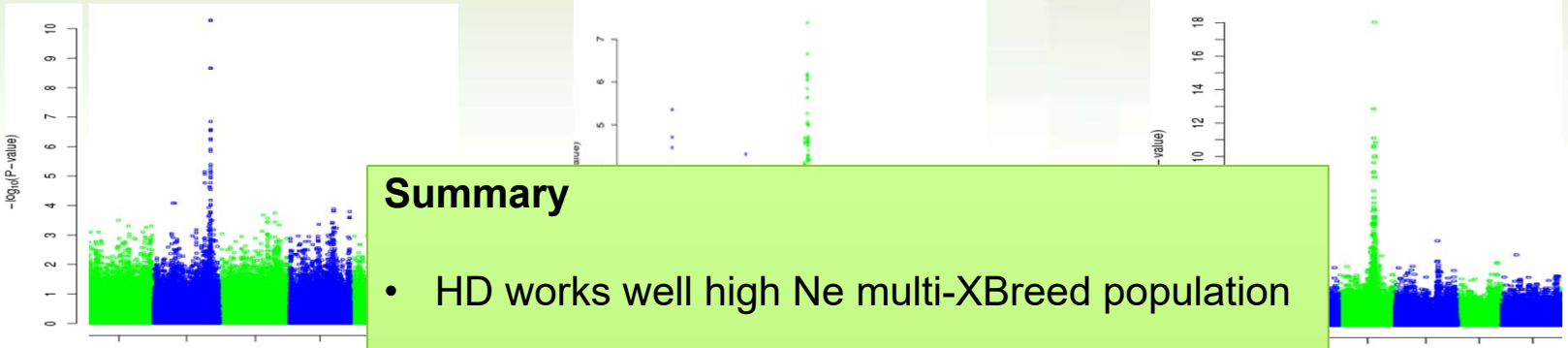
FARMIQ SYSTEMS LIMITED

agresearch

Tenderness (KgF)

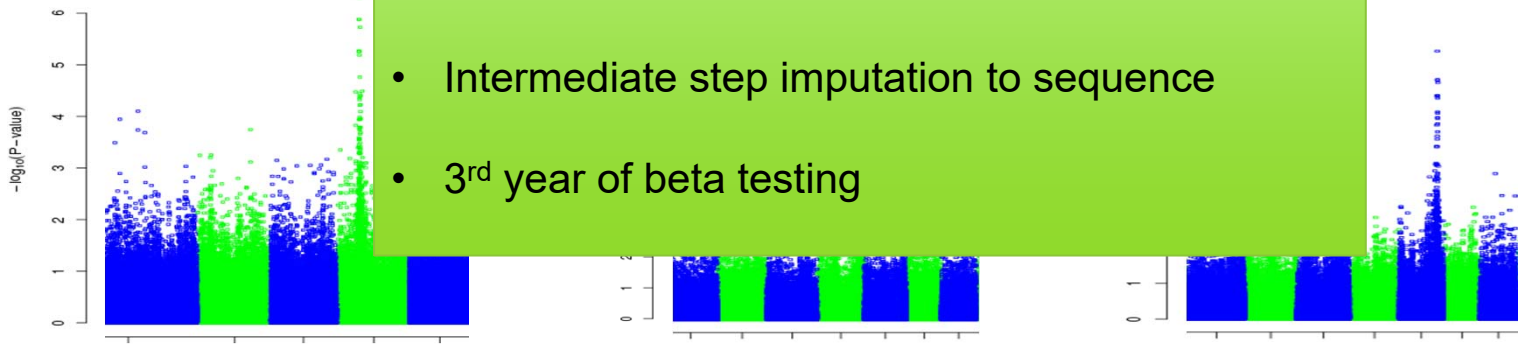
Marbling-visual

Colour a (8 weeks 48 h)



Fat depth

eg length



Summary

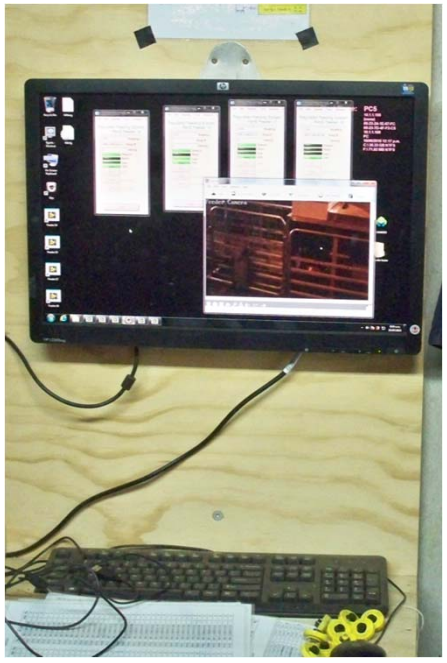
- HD works well high Ne multi-XBreed population
- Benefit mainly in exquisite identification of QTL
- Created 15K chip high imputation accuracy Ventura et al 2015
- Intermediate step imputation to sequence
- 3rd year of beta testing

FARMIQ SYSTEMS LIMITED

agresearch

Feed intake and efficiency

- Est. 2015
- 5 pens, 4 feeders per pen, 40 sheep per pen, 200 sheep
- ~1000 sheep over several years and linked to Australia
- Genomic selection only viable option to extend to industry



Portable accumulation chambers (PAC)

- Measure CH4, CO2 and O2 fluxes ~1hr off pasture
- CH4 and (CH4+CO2): heritable, repeatable, high rg across lifetime
  - Even after adjustment for liveweight
  - Predict methane emissions and feed intake cheaply ~10 Euro ~80/day
  - Measure several times
  - >7000 measurements, >2500 animals genomic predictions (50K +HD)

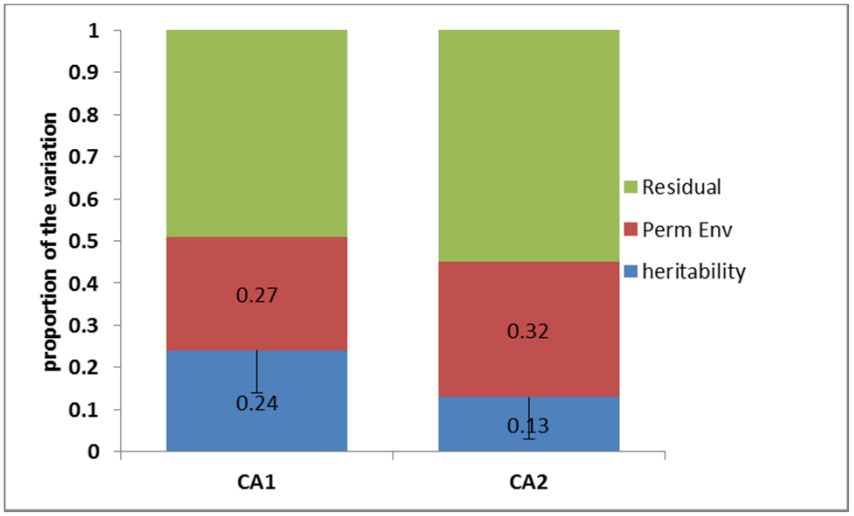
agresearch

Rumen microbial composition as a predictor

- Rumen samples collected 18 hr after last feed
- Via stomach tube
- Bacterial & archaeal 16S rRNA
- **Experiment 1**
  - 236 samples 118 animals
  - Extremes of 340 animals
  - Peter Janssen presented
- **Experiment 2**
  - B2012 & 2013 selection lines
  - 520 samples 260 animals unselected

Kittelmann et al PLoS One 2014 & Rowe et al 2015

Genetic Parameters: RMC



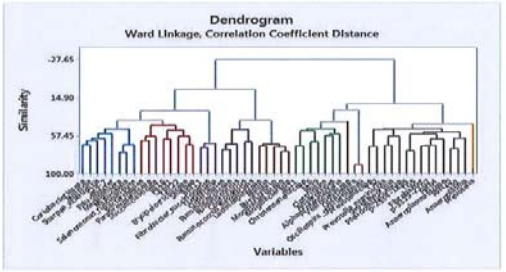
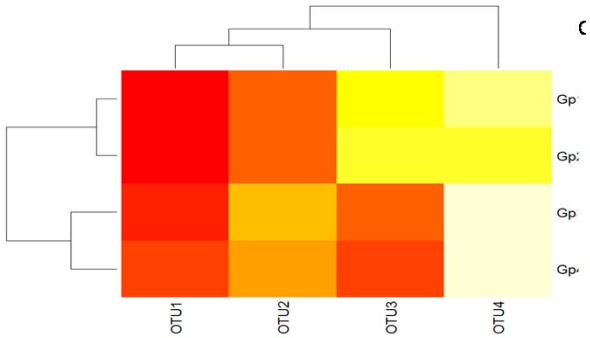
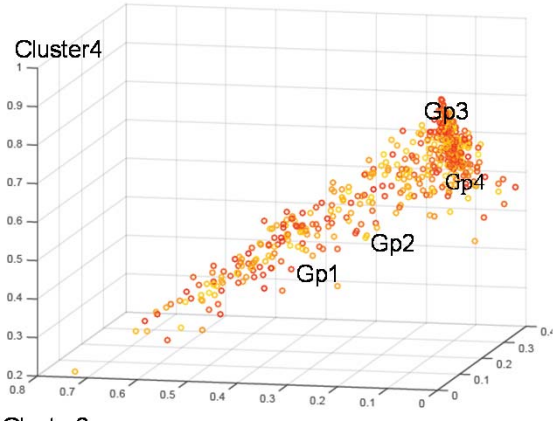


Figure 5: Clustering of the 54 bacterial species into ten groups, based on the profile of their contributions to the bacterial communities of the sheep.

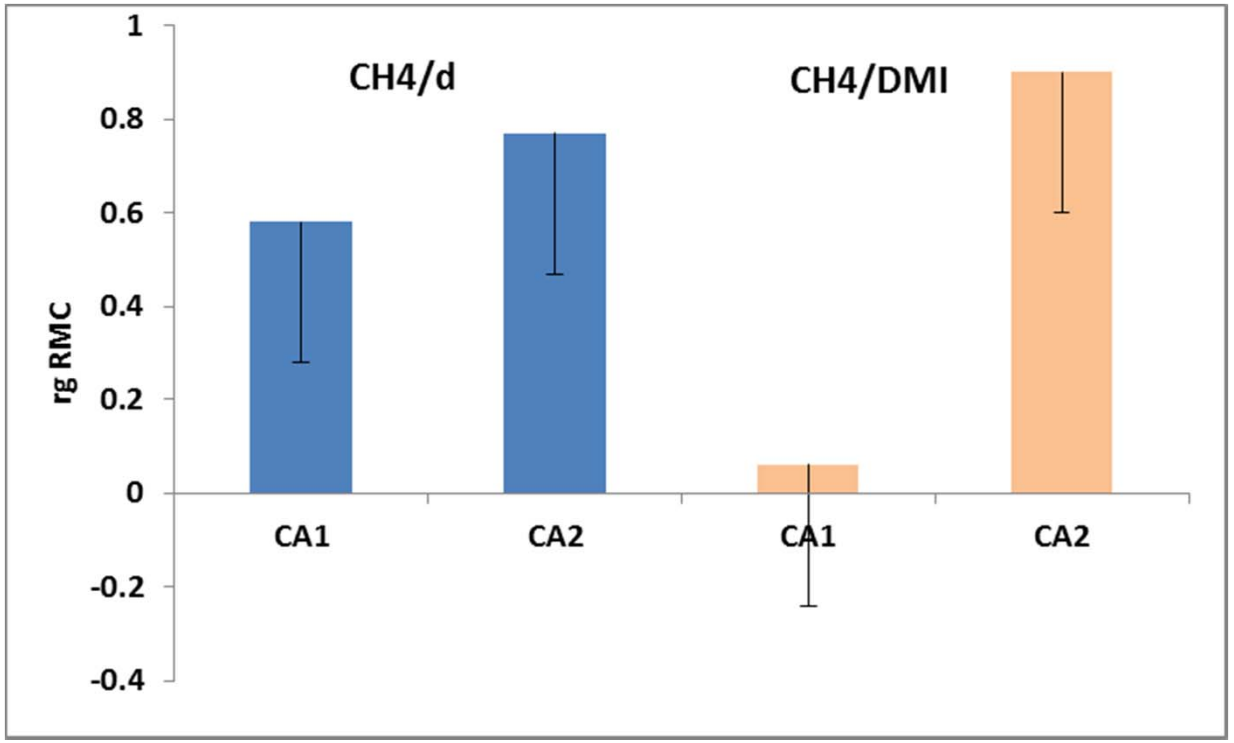




- N = 260 x 2 measures
- 54 “groups” of bacteria (grass -> VFA +H2)
- Most variation is in 2 dimensions (CA1, CA2)
- This variation is repeatable and heritable

Rowe et al 2015 submitted

RMC Genetic Relationship with Methane



Evidence suggests genetic relationship RMC and methane traits

Rowe et al 2015

Genomic selection?

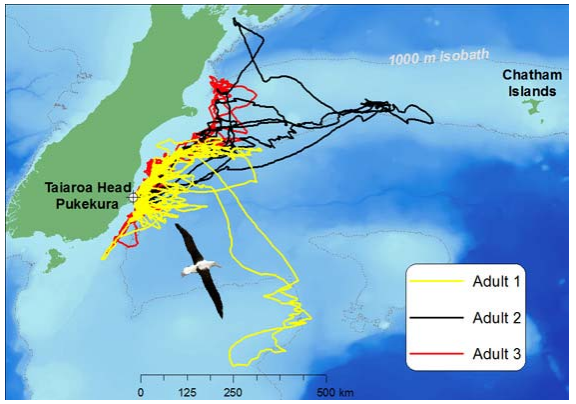
- 1659 animals RC measurements
- 3505 animals observed or imputed 606K genotypes
- 70,401 animals in extended pedigree
- mBV Accuracy and equivalent respiration chamber measurements
  - CH_4 0.50 1
  - g CH_4/ kg DMI 0.46 4
  - Liveweight 0.53

- Valuable for use in industry where young animals routinely genotyped

Pickering et al submitted

Precision phenotyping

- Phenotyping now largest cost & animals in hill country
- Difficult to measure: mating & lambing dates, lambing & grazing behaviour
- Take advantage of electronics and miniaturisation
- Basically put cell phone “fit bit” sensors (GPS, proximity, tilt …) into a tag
  - Communicate via distributed wi-fi
  - Amplify benefits via genomic selection
  - Power/weight is the current problem



agresearch

Reduce the generation length!!!

- Genomic selection only increases accuracy
- However available at a young age!!!
  - Potentially before implantation
  - Take any opportunity to reduce L

- In sheep L= 1 year potentially with JIVET
- Use more ram lambs
- Use SID AI & “mix of the day”
- Better oestrus detection extensive farms
- Potential for more than doubling ΔG
- Synergistic with precision genomics

$$\Delta G_{year} = \frac{i \; r_{AI} \; \sigma_A}{L}$$

i = Selection intensity

r_{AI} = Accuracy

σ_A = Genetic standard deviation

L = Generation interval

agresearch

Precision genomics …. Gene editing

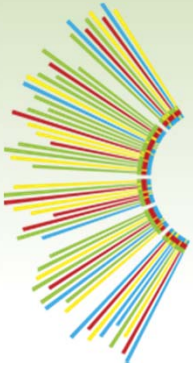
- Its here now!!!
- Initially mine known mutations (GDF8, ASIP, BCO2, GDF9, PrP, TMEM154)
- To make full use: need to know DNA function at a base pair level
  - Need better genome assemblies
  - Need classification of known variants …. 1000 genomes
- Need annotation of all functional elements not just gene coding regions
- Industry structures: need better integration of reproductive technologies

Multiplex gene editing via CRISPR/Cas9 exhibits desirable muscle hypertrophy without detectable off-target effects in sheep

Xiaolong Wang, Yiyuan Niu, Jiankui Zhou, Honghao Yu, Qifang Kou, Anmin Lei, Xiaoe Zhao, Hailong Yan, Bei Cai, Qiaoyan Shen, Shiwei Zhou, Haijing Zhu, Guangxian Zhou, Wenzhi Niu, Jinlian Hua, Yu Jiang, Xingxu Huang ✉, Baohua Ma ✉ & Yulin Chen ✉

Scientific Reports **6**, Article number: 32271 (2016)
doi:10.1038/srep32271

Received: 25 April 2016
Accepted: 04 August 2016
Published online: 26 August 2016



FAANG

Functional Annotation of Animal Genomes

- A coordinated international action to accelerate genome to phenome
- Precision genomics
- Needs better genome assemblies
- Need annotation of all functional elements not just gene coding regions

SheepGenomesDB

Resources for the Sheep Genomics Community

ISGC

- Imputation to sequence first step
- Phase 1 complete: 463 sheep 46.4M variants
- Need high density genotypes for confirmation/imputation
- HD chip & GBS

agresearch

Methylation patterns

- Methylation % correlated with stress, age and environmental exposure
- Heritable and correlated with nearby DNA variants
- Obvious target as “intermediate phenotype” for difficult traits (longevity)

- Problem sheep no industry tools: need methyl arrays and/or epi-GBS
  - epi-GBS compatible with GBS
  - Thought needed on DNA sampling and extraction

Genome-Wide DNA Methylation Patterns and Transcription Analysis in Sheep Muscle

Christine Couldrey, Rudiger Brauning, Jeremy Bracegirdle, Paul Maclean, Harold V. Henderson, John C. McEwan

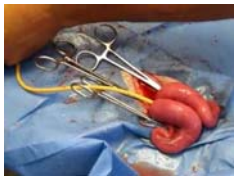
Published: July 10, 2014 • http://dx.doi.org/10.1371/journal.pone.0101853

agresearch

Conclusions

- Shift from DNA parentage to genomic selection **for all the ram breeding tier**
  - Requires <15 Euro/sample for >10,000 SNPs
  - 2 strategies:
    - High throughput low density arrays (+/- imputation)
    - Genotyping by sequencing no imputation ($ already there for sheep)
- Have to rethink how and what extra traits to measure for maximum benefit
  - Includes sex limited, late in life, post slaughter & expensive traits
    - E.g. maternal weaning weight, ewe longevity, disease resistance, meat eating quality, feed intake, methane
    - Sustainability and efficiency focus
- ↓generation interval: e.g. JIVET & increased SID AI in stud breeding tier
- Precision genomics (aka gene editing) requires fully annotated genome assemblies.
  - Hence the FAANG initiative
  - Interim step is probing intermediate phenotypes via GWAS: epi-GBS and RNAseq
  - 1000 sheep genomes

KGD



FAANG
Functional Annotation of Animal Genomes

agresearch

Acknowledgements

Genomics for Production & Security in a Biological Economy

ISGC

Ministry of Business, Innovation & Employment

Patricia Johnson (RFI and taggles)
Peter Janssen Graham Wood (rumen genomics)
L Brito, M Lee (meat quality)

beef+lamb new zealand GENETICS

SF ICELAND STOFNFISKUR

NEW ZEALAND AGRICULTURAL GREENHOUSE GAS RESEARCH CENTRE PGgRc NATIONAL GERMPLASM RESOURCES LABORATORY WORKING TOGETHER

FARM IQ SYSTEMS LIMITED

agresearch

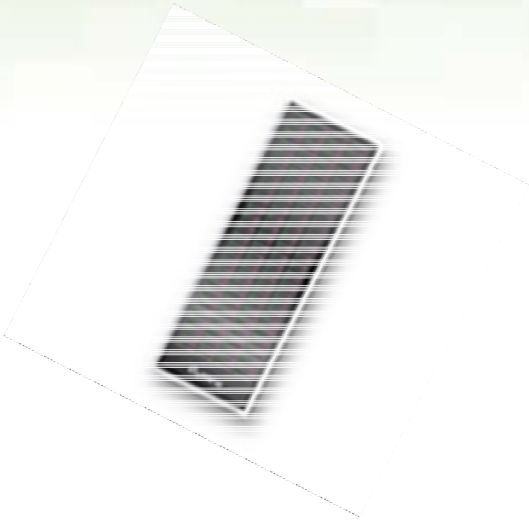
Example-Dairy Sheep

	GBS	15k Chip
Genome Required	No	Yes
Development Costs	Minimal ~US$3k	Moderate >US$60k
Species/Population Dependent	No	Yes
DNA quality and Quantitation	High ~4 days	Moderate ~2 days
SNPs called	Untargeted	Fixed & reliable
Throughput	188/376 per lane	24 per chip
Cost per sample	US$16/25	US$35
Number of SNPs	>60,000 (20-30/Mbp) ave. depth 3.4 reads @ 188 per lane	15,000 (~5/Mbp)
Minor allele frequency	Greater range	Biased

agresearch

Back to Sheep…

“Enhanced” parentage chip



Illumina Infinium XT

- 96-sample BeadChip

- multispecies
- ~800 Parentage SNPs from the 15K chip
- Literature SNPs
- Enhanced imputation SNPs

Aim is to have all animals in breeding tier genotyped

Contact:
shannon.clarke@agresearch.co.nz
john.mcewan@agresearch.co.nz

agresearch