ROSLIN

THE UNIVERSITY of EDINBURGH
Royal (Dick) School of Veterinary Studies

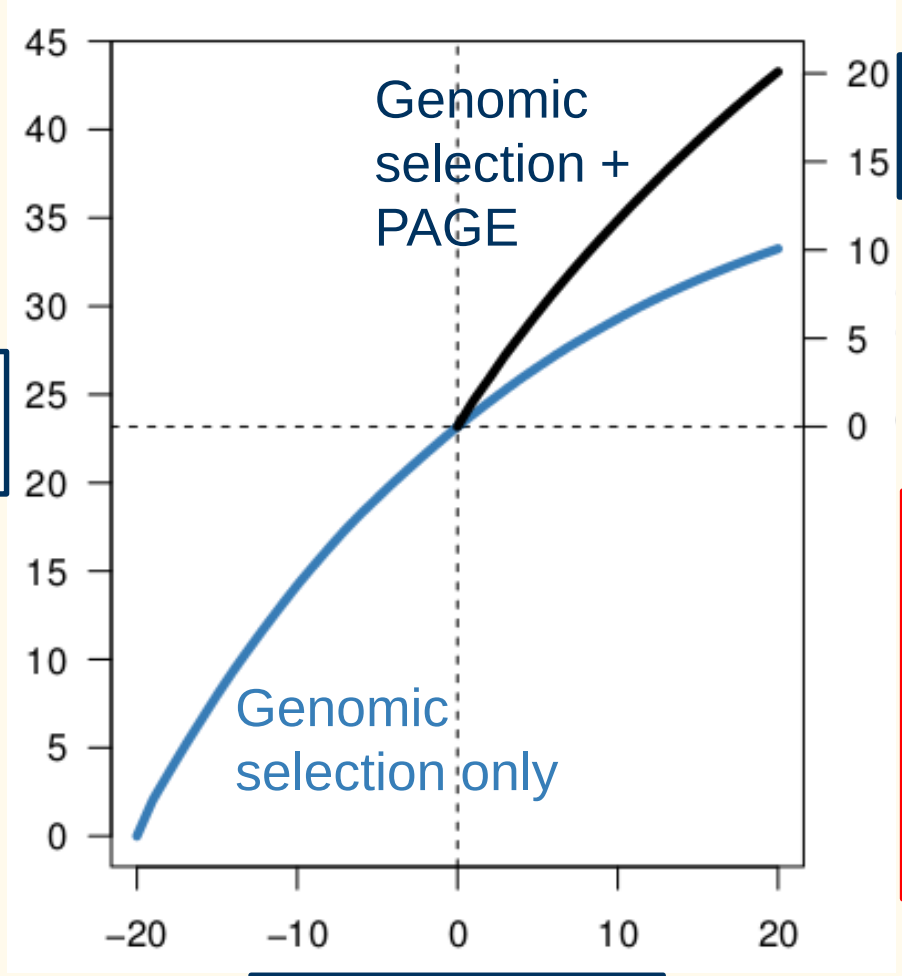
Identification of causal variants using one million individuals with whole-genome sequence information

Janez Jenko, Andrew Whalen, R. Chris Gaynor, Christos Dadousis, Gregor Gorjanc and John M. Hickey

BBSRC
bioscience for the future

PAGE at WCGALP 2014

Genetic Gain
(since generation -20)



Genetic Gain
(since generation 0)

Generations

PAGE can potentially double genetic gain!

How can we achieve this?

- Over the 20 years we edited ~300 distinct causal variants
  - They explain 36% of genic variance
  - 3% of all the causal variants
  - 15 variants per year

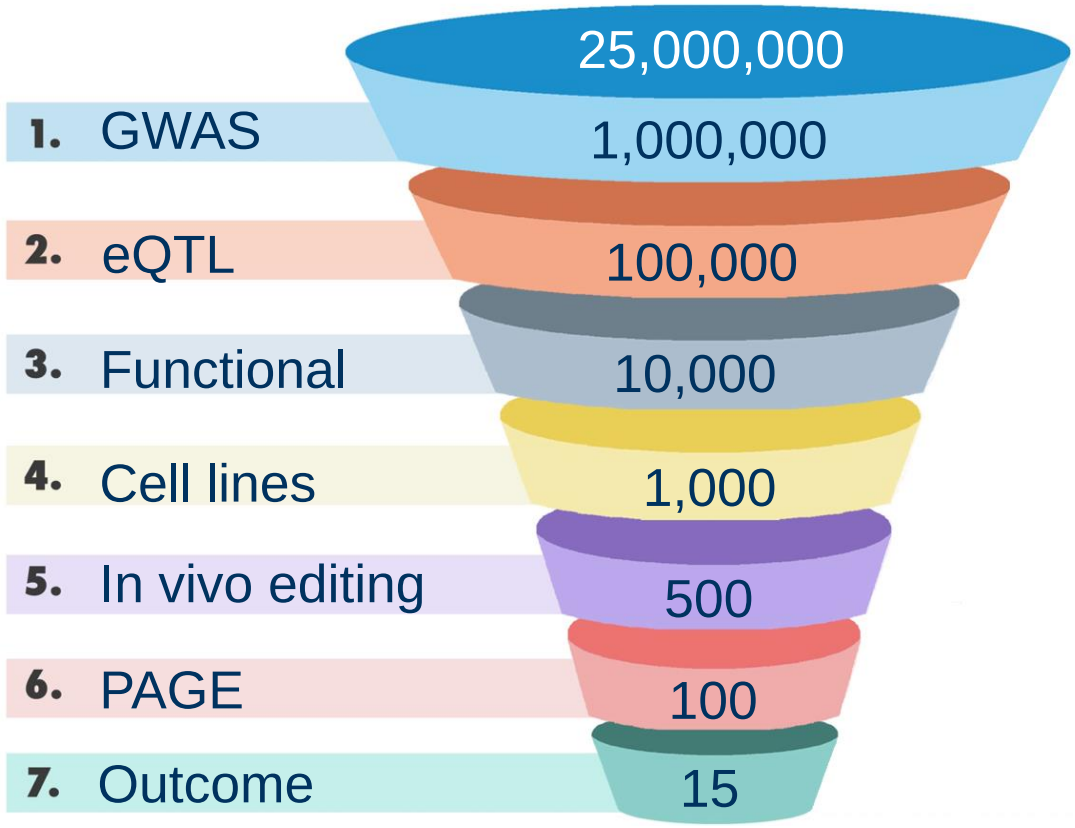
- Old approach to variant discovery will not work

- Allele testing approach

Allele testing scheme

A process to game the odds

~25 million segregating sites
~10,000 affect the trait
~1,000 work in a simple additive way How to find 15 of these?



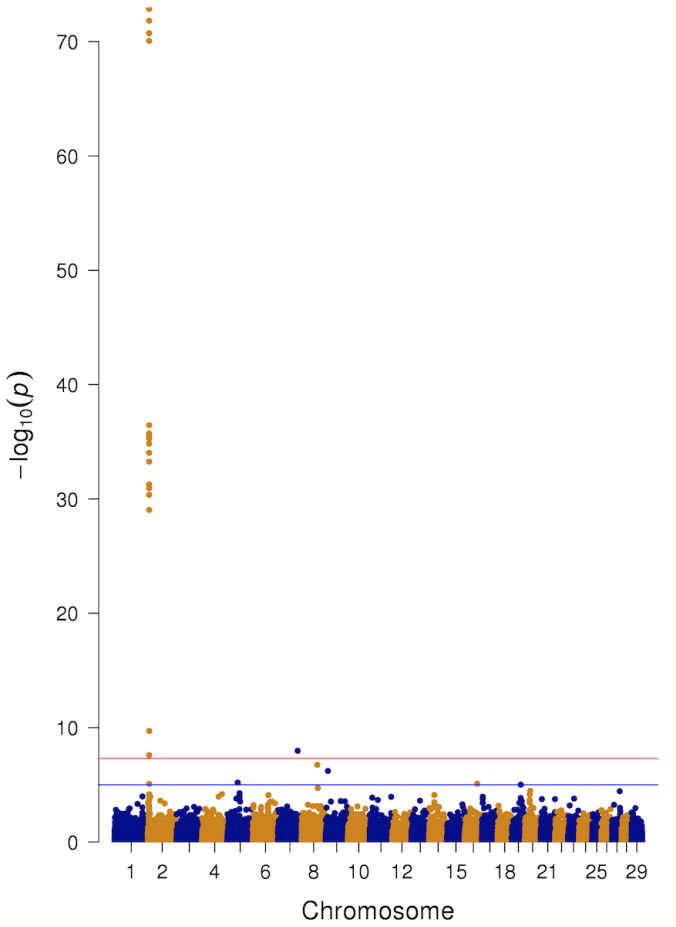
Increase ratio of causal variants in subset	Increase probability that variant is NOT highly deleterious
0.0000006	0.0
0.000015	?
0.00015	?
0.0015	?
0.015	0.1
0.03	1.0
0.15	1.0
1.0	1.0

Aim of current study

Million animals

WGS

Change in the ratio of causal variants in subset



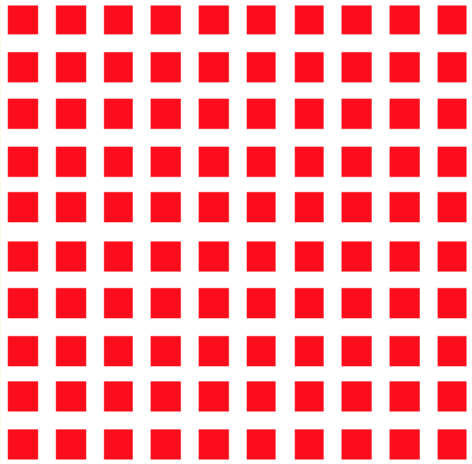
ROSLIN

THE UNIVERSITY of EDINBURGH
Royal (Dick) School of Veterinary Studies

BBSRC
bioscience for the future

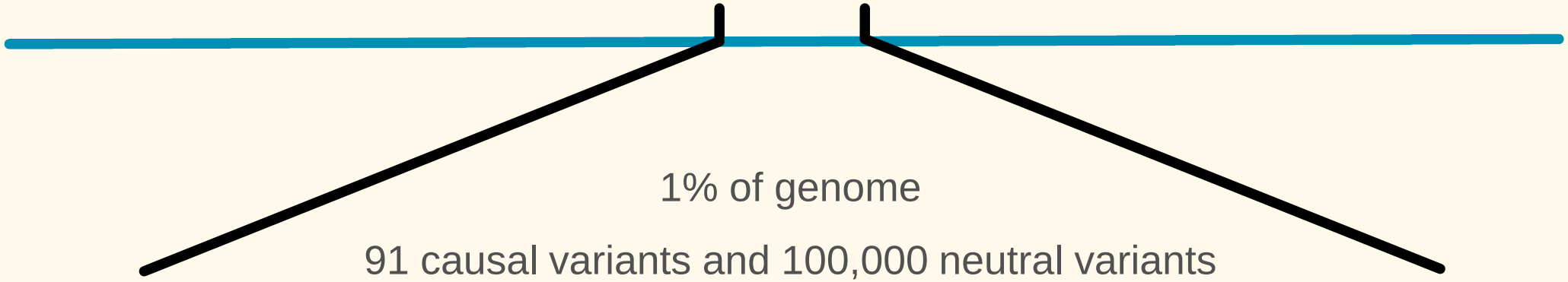
Simulating 1 million animals

- Historical sequences for 10 related populations
- 1 million animals (10 populations with 10 generations)
- Polygenic trait with 10,000 causal variants
- Phenotype with 0.3 heritability



Facilitating simulations

- 9 chromosomes with SNP information
- 1 chromosome with WGS information



1% of genome

91 causal variants and 100,000 neutral variants

Ranked 44th, 420th, 574th... across the whole on the effect size

Single SNP regression model

$$y = \mu + \mathbf{X}\beta + g + e$$

y - vector of phenotypes

μ - mean

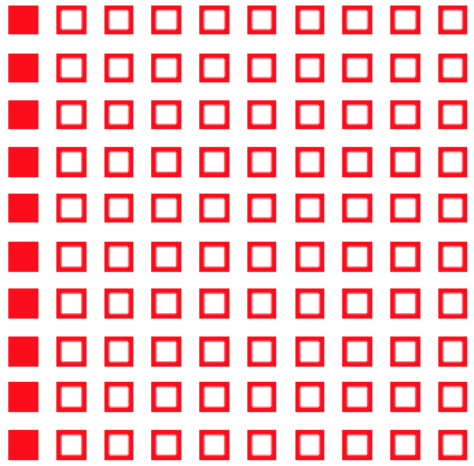
$\mathbf{X}$ - incidence matrix

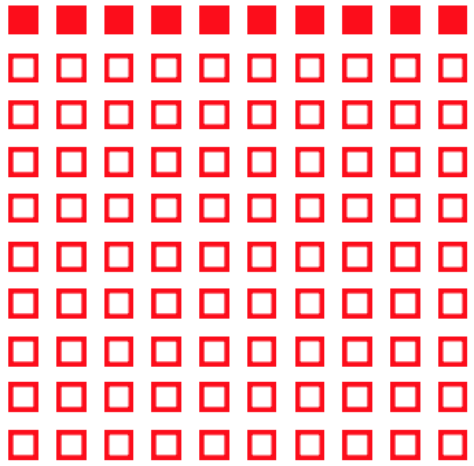
β - fixed effects

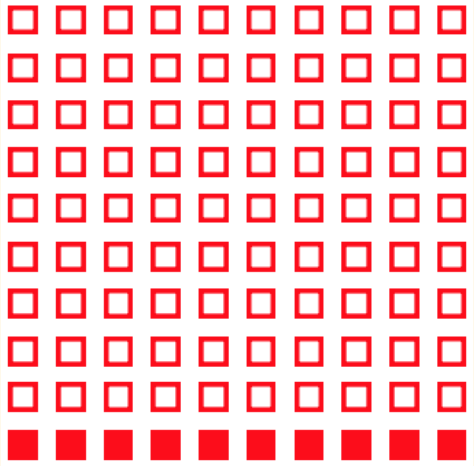
g - random genetic effect $N(0, \mathbf{G}\boldsymbol{\sigma}_{\boldsymbol{g}}^{\mathbf{2}})$

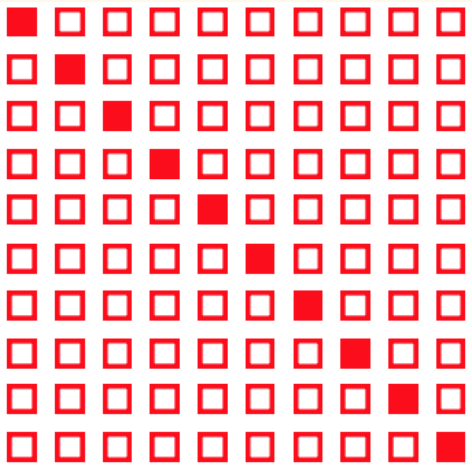
e - residual $N(0, \boldsymbol{I}\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}})$

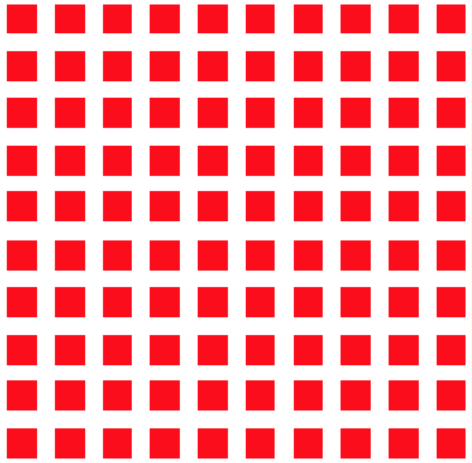
Analysed scenarios



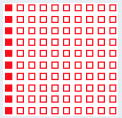
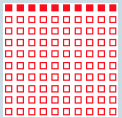
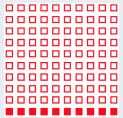
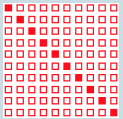
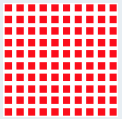




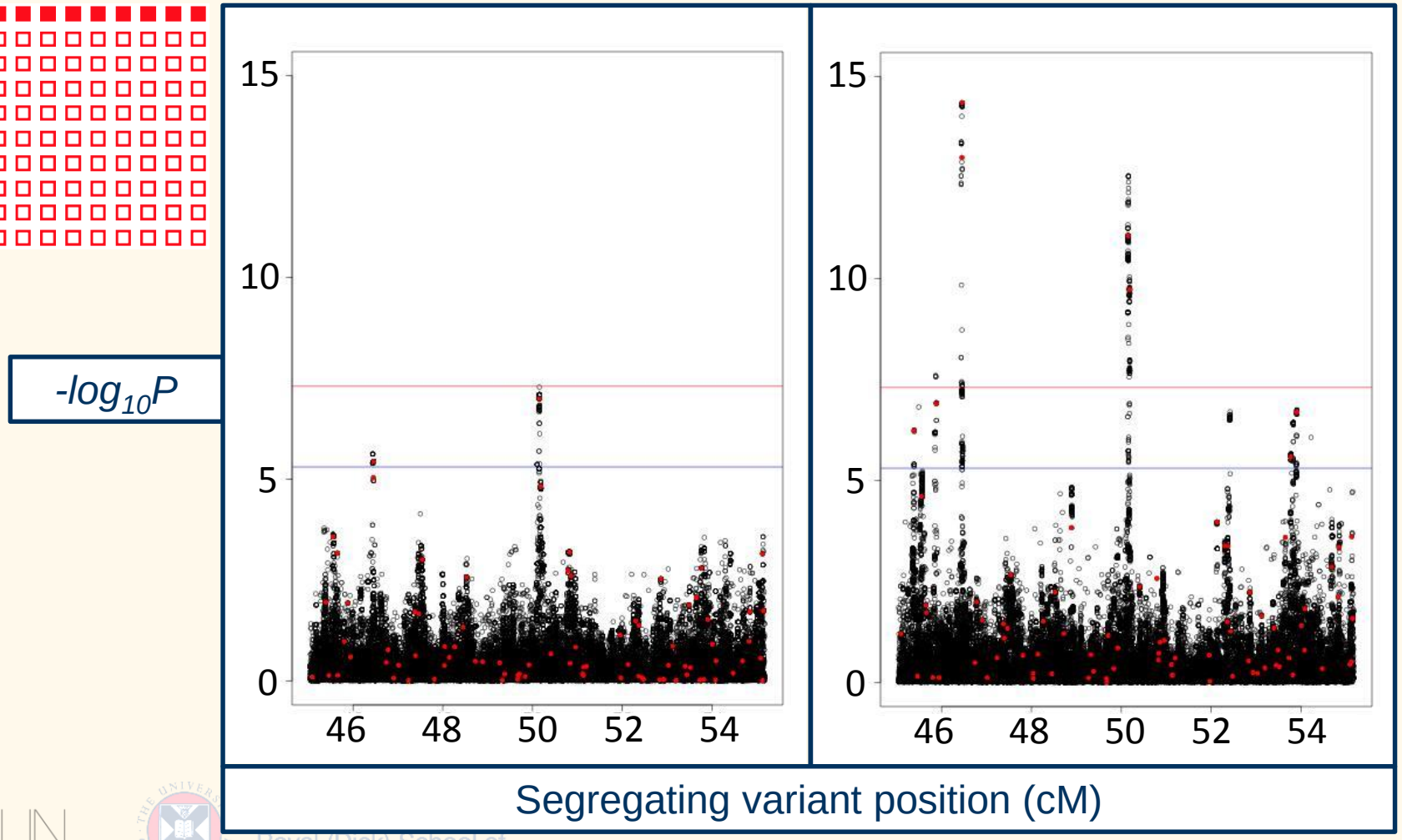




Causal and neutral variants

Data set	Number of causal variants		Number of neutral variants	
	Analysed region	Whole genome approximation	Analysed region	Whole genome approximation
	69	6,900	70,819	7,081,900
	84	8,400	85,438	8,543,800
	79	7,900	83,696	8,369,600
	67	6,700	70,885	7,088,500
	84	8,400	85,435	8,543,500

Manhattan plots

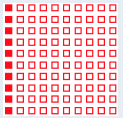
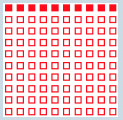
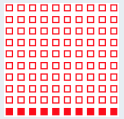
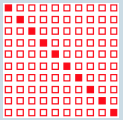
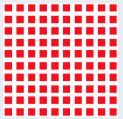


ROSLIN

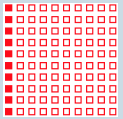
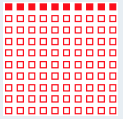
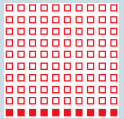
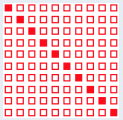
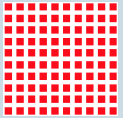
Royal (Dick) School of Veterinary Studies

BBSRC bioscience for the future

Significant variants statistics I.

Data set	Number of causal variants		Number of neutral variants	
	Analysed region	Whole genome approximation	Analysed region	Whole genome approximation
	0	0	0	0
	0	0	0	0
	2	200	176	17,600
	0	0	0	0
	4	400	256	25,600

Significant variants statistics II.

Data set	Genetic variance explained (%)	Correlations between the causal variant effect and $-\log_{10}\mathrm{P}$ value
	0	0.32
	0	0.46
	21.3	0.51
	0	0.51
	22.9	0.68

Change in the ratio of causal variants in the subset

- Before GWAS: 1 causal variant out of 1018 variants (84/85,519)
- After GWAS: 1 causal variant out of 64 variants (4/260)

GWAS increased the ratio of causal variants in the subset for ~**16** times

Conclusions

- GWAS is effective first step in allele testing scheme
- GWAS discovered ~400 causal variants
- ~25,000 false positives
- The next steps in allele testing will be to reduce these false positives to 3000

Acknowledgements

John Hickey, Gregor Gorjanc, Andrew Whalen, Chris Gaynor, Christian Werner, Christos Dadousis, Daniel Money, David Wilson, Jaap Buntjer, Janez Jenko, Joanna Warner, Jon Bancic, Lorena Batista, Martin Johnsson, Owen Powell, Roberto Antolin, Roger Ros Freixedes, Serap Gonen, Stefan Hoj-Edwards.

Aviagen
Bill & Melinda Gates foundation
CIMMYT
DairyCo
Driscoll's Only the Finest Berries
Elsoms
geno
Genus
ICBF.com
ICRISAT Science with a human face
illumina
Innovate UK
KWS
Limagrain de la terre à la vie
MONSANTO
MRC Medical Research Council
R·A·G·T
ROSLIN
THE UNIVERSITY of EDINBURGH Royal (Dick) School of Veterinary Studies
BBSRC bioscience for the future