

Impact of the structure of the genomic matrix on accuracy of evaluations for imputed genotypes

Raphael Mrode & Mike Coffey

Introduction



- About 20 chips of varying density are imputed to the 50K Illumina chip in our genomic prediction system
- Imputation of the low density chips to 50K is implemented using the software Findhap
- Aim of the study was to understand the impact of such imputation on the accuracy of genomic evaluations for bulls with imputed genotypes

Introduction



- Initially the accuracy of imputation for each chip type was assessed within four classes of bulls depending on degree of relationship with reference bulls:
 - Those with high (> 80%), medium (25-80%), low (1-24%) and 0% relationship
- Subsequently, the accuracy of evaluations for bulls with imputed genotypes was computed within the same four categories.

Introduction



- Benefits of including a polygenic effect in genomic models to account for genetic variance not captured by markers have been demonstrated
- The question is; does inclusion of polygenic effect provides an additional benefits in accuracy of genomic predictions for bulls with
 - low or no relationship with bulls in the reference population
 - or low relationships with all other bulls in the pedigree used for the imputation?

Accuracy of genomic prediction



- Study therefore examined the impact of various **G** constructed with different weights on the **A** on accuracies of genomic predictions within the 4 classes of bulls.
- Four **G** matrices were:
 - **G** + 0.01 added to diagonal elements
 - computed with a weight of 0.05, 0.10 and 0.20 given **A**
- All evaluations were based on GBLUP

Materials and methods - Imputation



- Genotypes corresponding to 20 different chips were created from bulls genotyped with 50K Illumina chip using their chromosome map file.
- For each low chip type, 1200 bulls with genotypes were generated and then imputed to 50K .
- Reference population consisting of 14,280 bulls
- The accuracy of imputation was computed as the correlation between the imputed and the original 50K genotypes within the 4 classes

Materials and methods – genomic prediction



- Different **G** were computed using both the reference bulls and bulls with imputed genotypes.
- GBLUP was undertaken with bulls imputed regarded as validation bulls for production traits and somatic cell counts
- The accuracy of genomic prediction was computed as the correlation between DGV and the de-regressed proofs for validation animals.
- Relative accuracy of imputed genotypes was computed as ratio of accuracies from using **G** with imputed genotypes to **G** based on original 50 chip.

Description of chip types



Low density chips of size 3k to 15k			Medium density chips of size 17k to 30k		
Chip number	Size	Name	Chip number	Size	Name
3	3000	3K	9	19725	GGP-Super
6	6909	LD	14	26151	GGP V3
7	8762	GGP	100	19720	GGP V1-1
10	11410	ZLD	109	19809	GGP-S+90
12	9072	ELD	110	20074	GGP_SuperLD_ SemiPrivate
13	6912	LD2	111	20077	GGP_SuperLDv1- 1_SemiPrivate
16	14376	IDBV2	112	26359	GGPv3_SemiPrivate_E
			215	17619	ZL2
High density chips of size greater 30k					
8	76999	GHD			
11	56955	ZMD			
15	76883	GGPHD_T_SemiPrivate			
108	77068	GHD +68			
208	76934	GHD T			

Accuracy of imputation



	Low density chips if size 3k to 15k						
AVgRel	3	6	7	10	12	13	16
>80	0.922	0.978	0.981	0.986	0.982	0.979	0.988
25 – 80	0.917	0.974	0.979	0.985	0.981	0.976	0.986
1 - 24	0.913	0.973	0.978	0.985	0.980	0.975	0.985
0	0.899	0.967	0.974	0.982	0.976	0.969	0.981

Accuracy of imputation



Low density chips if size 3k to 15k

AVg Rel (%)	AVgRel I2 (%)	N	3	6	7	10	12	13	16
1-24	>6	427	0.915	0.974	0.979	0.985	0.980	0.976	0.986
	<6	20	0.879	0.941	0.970	0.977	0.971	0.944	0.977
0	>6	207	0.906	0.972	0.977	0.985	0.978	0.974	0.986
	<6	26	0.838	0.931	0.949	0.960	0.958	0.935	0.976

Medium density chips of size 17k to 30k

	Medium density chips of size 17k to 30k							
Avg Rel (%)	9	14	100	109	110	111	112	215
>80	0.982	0.982	0.981	0.983	0.981	0.982	0.983	0.987
25 – 80	0.979	0.981	0.980	0.980	0.980	0.980	0.980	0.986
1 - 24	0.978	0.980	0.980	0.979	0.979	0.980	0.980	0.986
0	0.975	0.976	0.975	0.974	0.975	0.975	0.977	0.985

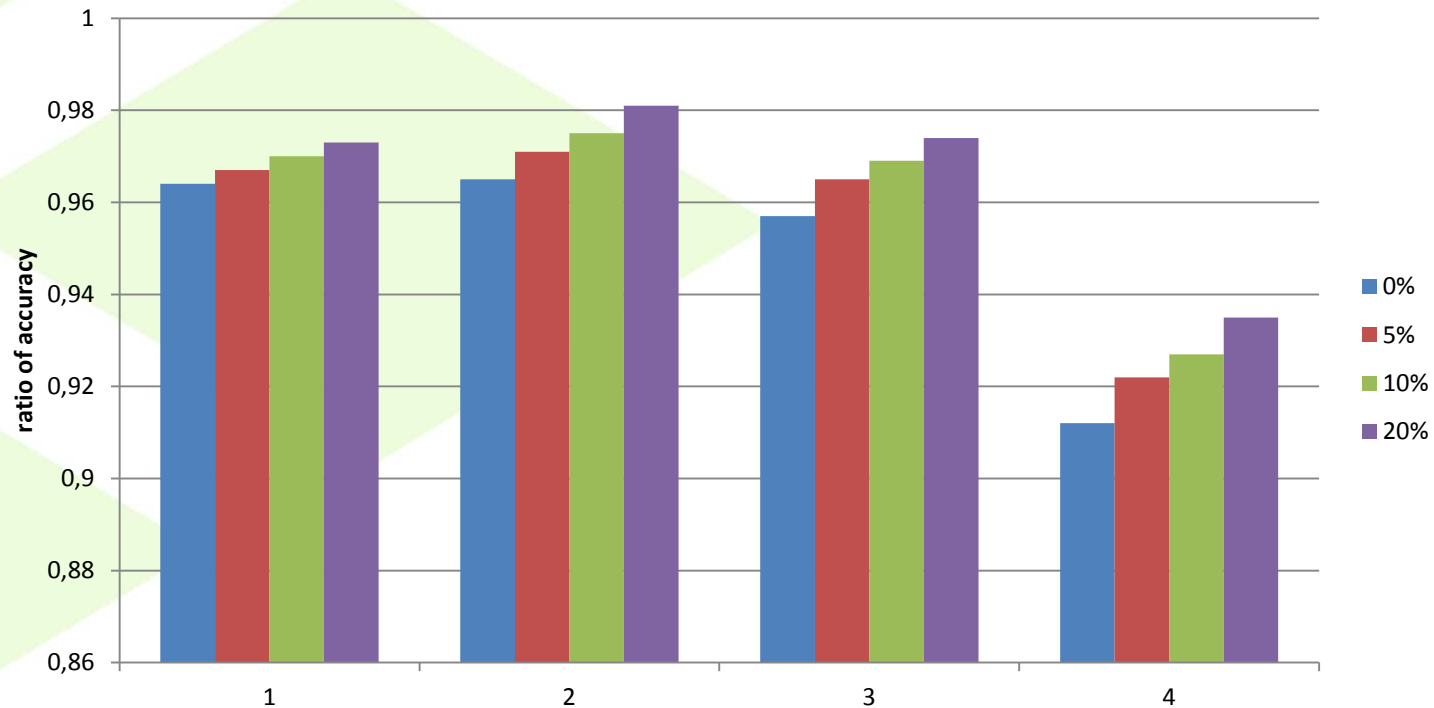
	High density chips of size greater 30k				
AVgRel(%)	8	11	15	108	208
>80	0.996	0.999	0.996	0.996	0.996
25 – 80	0.996	0.999	0.996	0.996	0.996
1 - 24	0.995	0.999	0.995	0.995	0.995
0	0.995	0.999	0.995	0.995	0.995

Relative accuracy for milk yield for chip

3



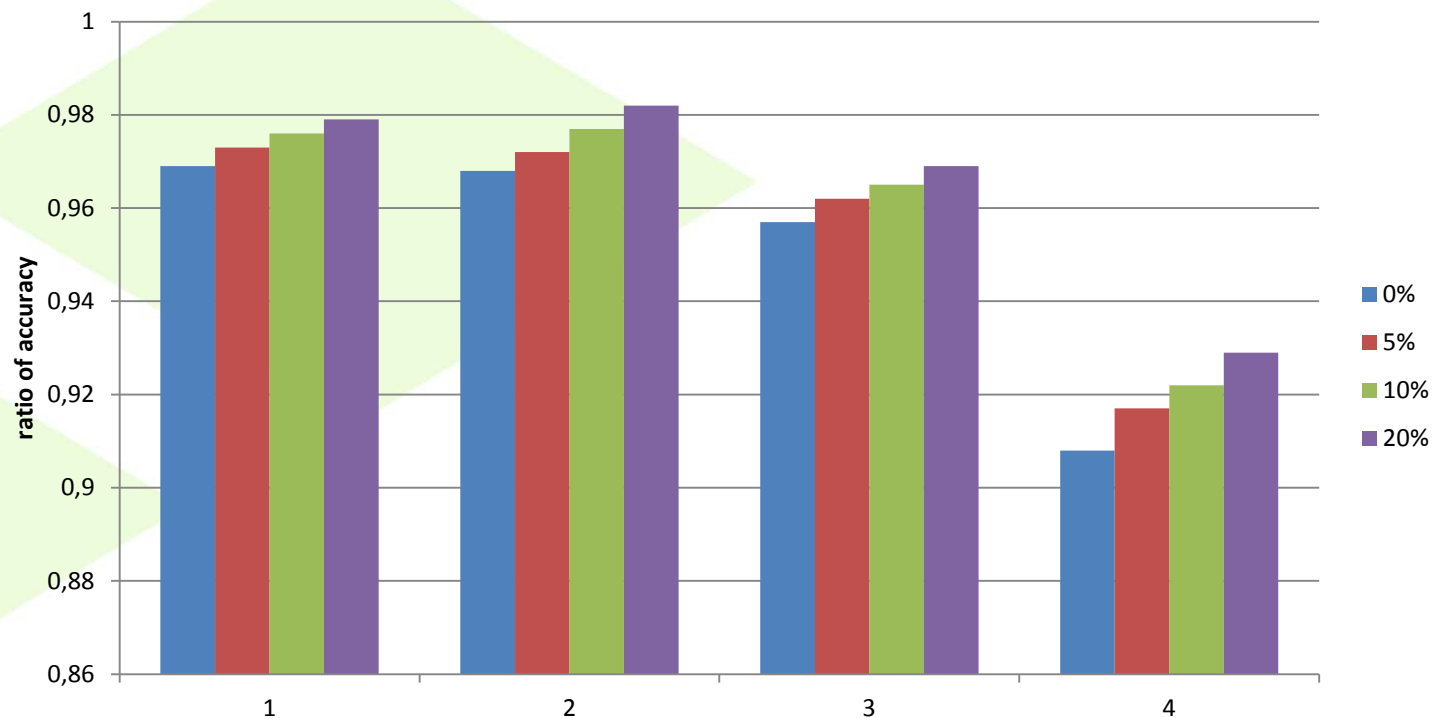
- Milk yield accuracy of predictions from original 50K chip were 0.80, 0.84, 0.80 and 0.80 for the 4 classes of bulls



Relative accuracy for milk yield for chip 6



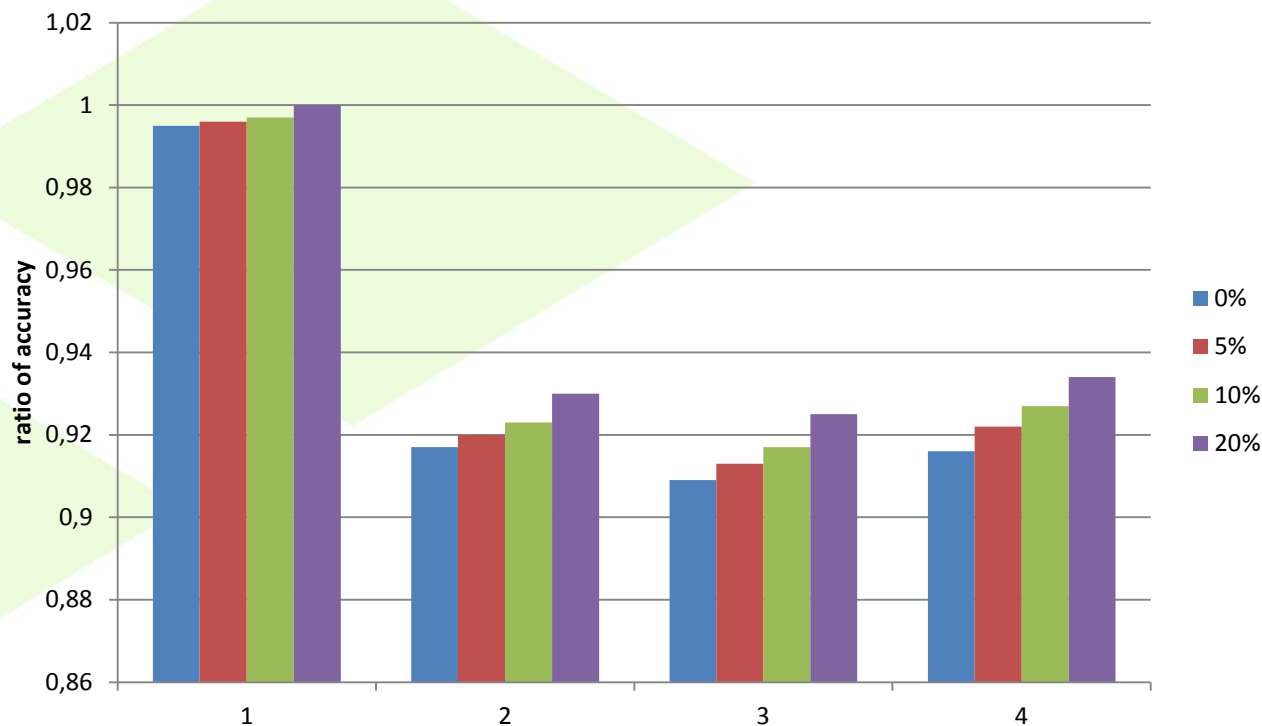
- Milk yield accuracy of predictions from original 50K chip were 0.85, 0.84, 0.80 and 0.78 for the 4 classes of bulls



Relative accuracy for milk yield for chip 9



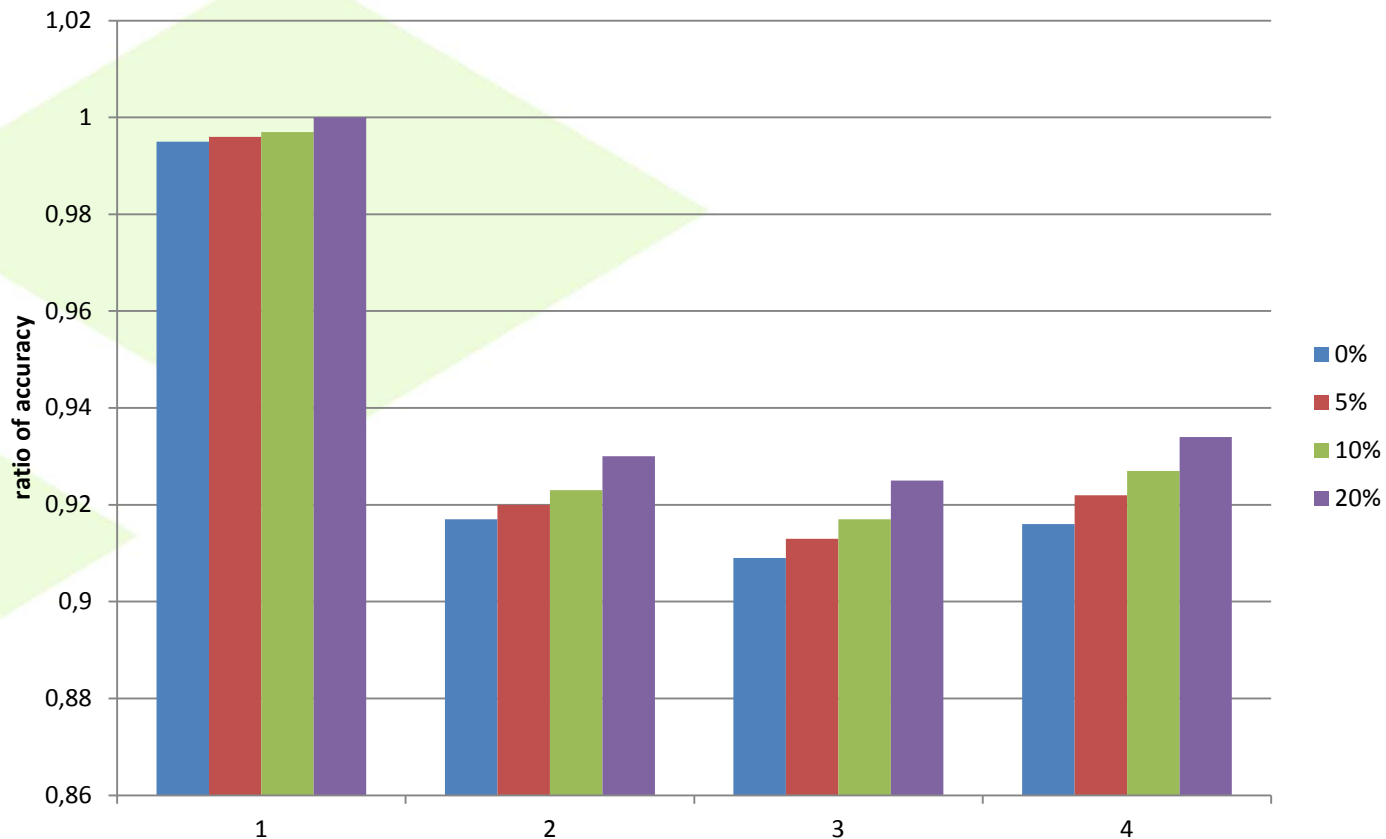
- Milk yield accuracy of predictions from original 50K chip were 0.88, 0.81, 0.81 and 0.75 for the 4 classes of bulls



Relative accuracy for SCC for chip 3



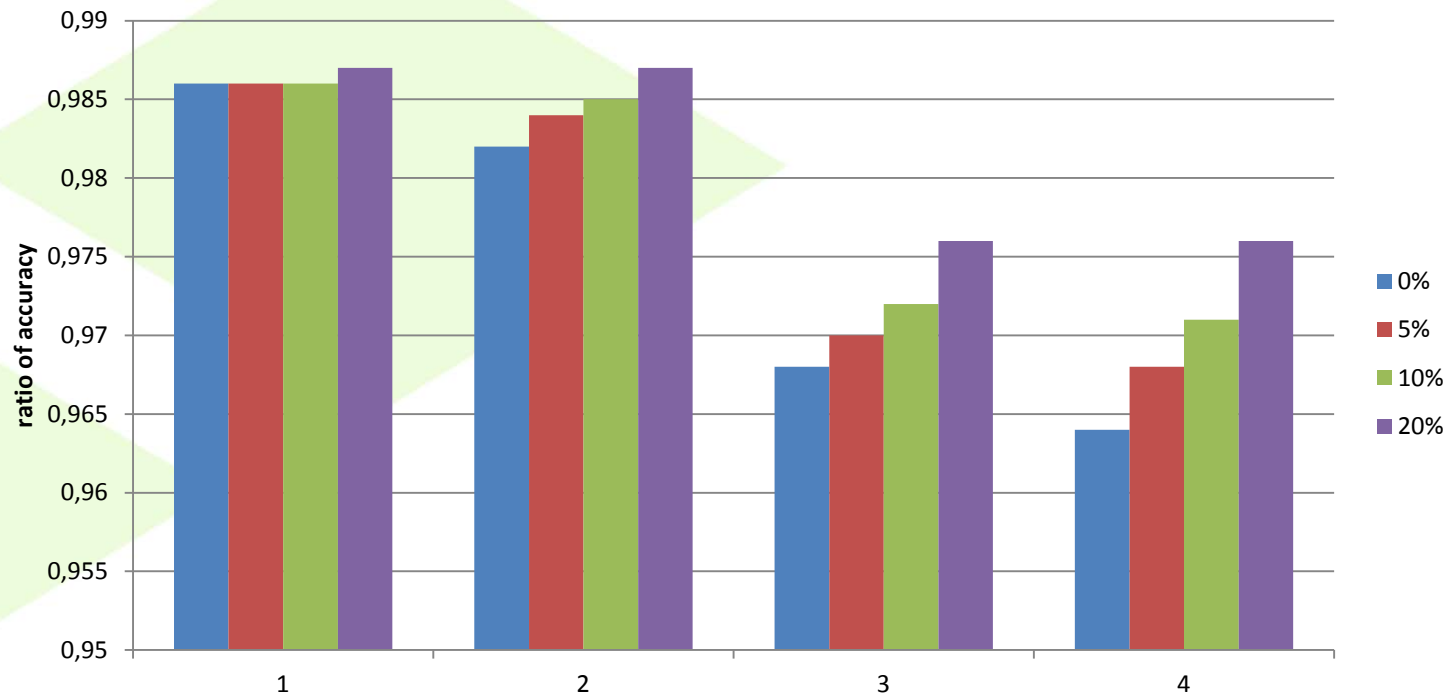
- SCC accuracy of predictions from original 50K chip were 0.73, 0.72, 0.70 and 0.67 for the 4 classes of bulls



Relative accuracy for SCC for chip 6



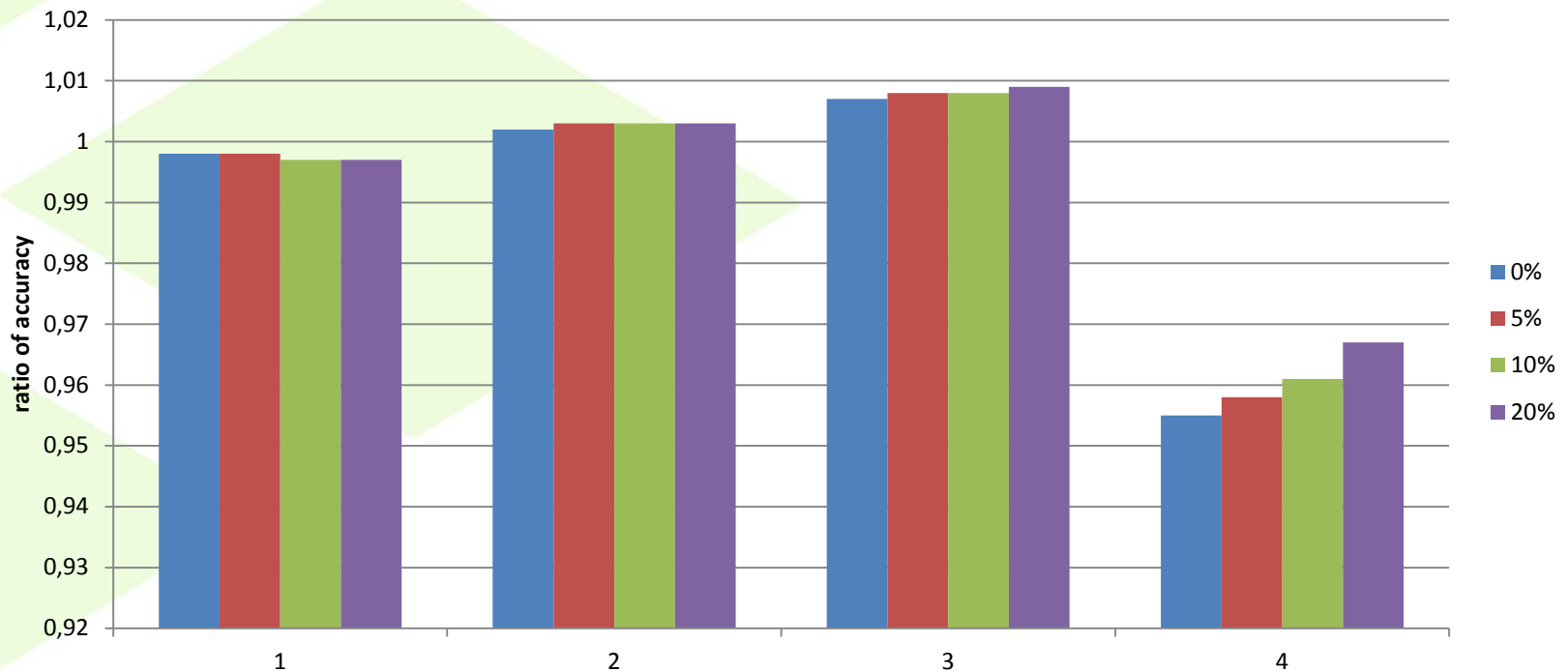
- SCC accuracy of predictions from original 50K chip were 0.80, 0.73, 0.70 and 0.66 for the 4 classes of bulls



Relative accuracy for SCC for chip 9



- SCC accuracy of predictions from original 50K chip were 0.82, 0.75, 0.73 and 0.69 for the 4 classes of bulls



Average relationship from different G for animals with chip 3



	Using Un-imputed genotypes	Using imputed genotype		
AVgRel	G	G	G + 0.05A	G+0.20A
>80	0.047 (0.145)	0.047 (0.140)	0.051 (0.139)	0.063 (0.137)
25-80	0.029 (0.087)	0.029 (0.083)	0.033 (0.082)	0.044 (0.081)
1-24	0.033 (0.085)	0.032 (0.080)	0.035 (0.079)	0.045 (0.078)
0	0.045 (0.108)	0.042 (0.102)	0.044 (0.102)	0.052 (0.100)

Conclusion



- In general for chips of low density the accuracy of imputation is influenced by the degree of their relatedness to reference bulls
 - The accuracy increases as the relatedness increases.
 - However, this tendency decreases in medium sized chips
 - and is non-existent in chips more than 30K.
- For production traits and SCC, the accuracy of genomic predictions similarly increased as bulls are more related to reference bulls.

Conclusion



- In the inclusion of some degree of polygenic effects contributed to improved accuracy in the bulls poorly related to the reference bulls

Acknowledgement



Funding by DairyCo and access to CDDR genotypes gratefully acknowledged

HD Genotypes available as part of the UK Ruminant Genetic improvement Network funded by DEFRA

Thanks for listening

