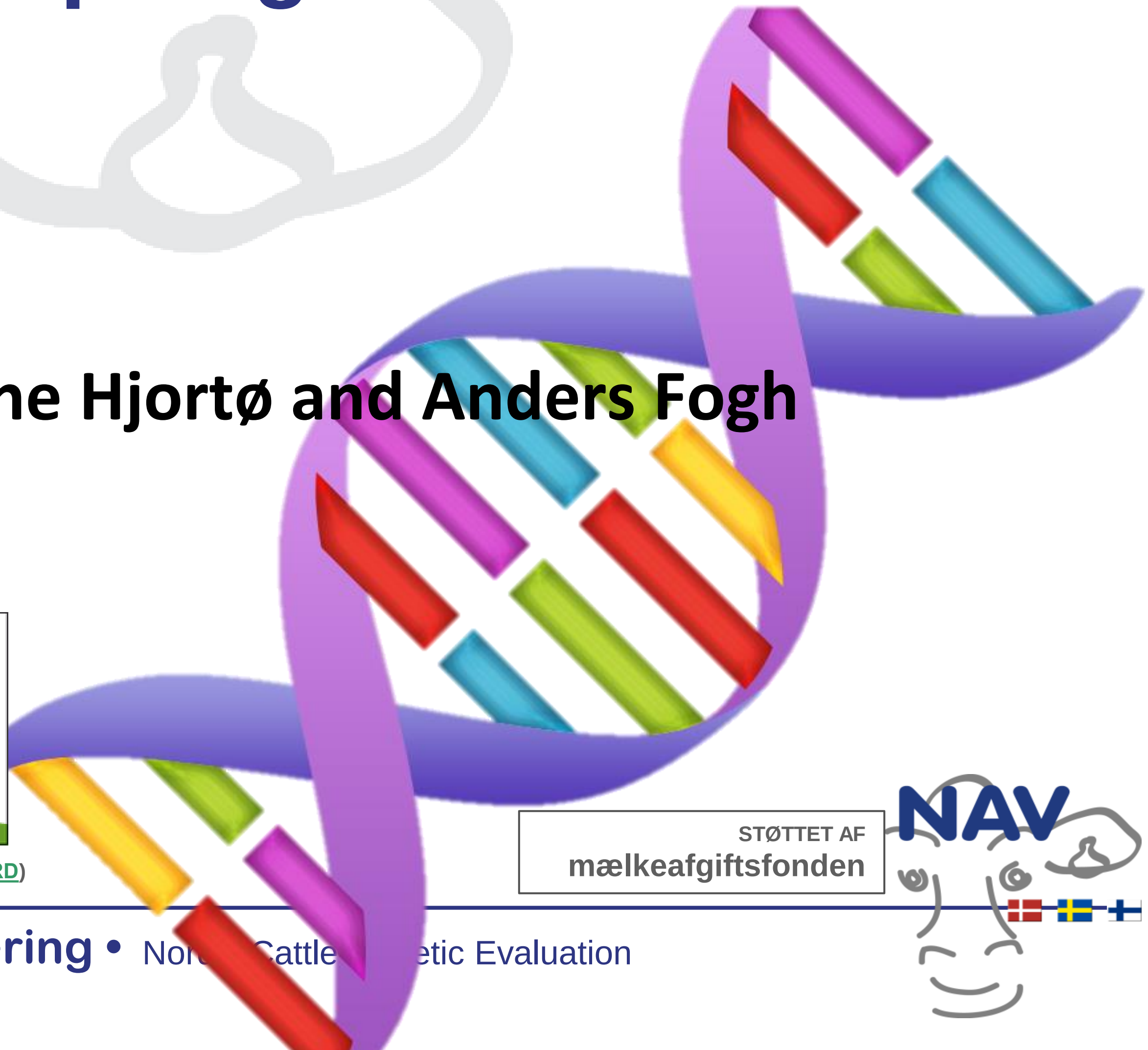
Imputation and pedigree verification



Kevin Byskov, Line Hjortø and Anders Fogh

Ministeriet for Fødevarer,
Landbrug og Fiskeri
Den Europæriske
Landbrugsfond for Udvikling
af Landdistrikterne
Naturerhverv.dk
Danmark og EU investerer i landdistrikterne.

Se "European Agricultural Fund for Rural Development" (EAFRD)

STØTTET AF
mælkeafgiftsfonden

Females tested in the LD-project (spring 2013)

- **7571 animals genotyped**
  - **22 deleted due to low callrate**

Country	JER	RDC	Total
Denmark	3835	532	4367
Sweden	78	1233	1311
Finland	-	1871	1871
Total	3913	3636	7549

How many genotypes are accepted?

- **% approved, rutine: 96-97 %**

Percent approved per breed, LD-project

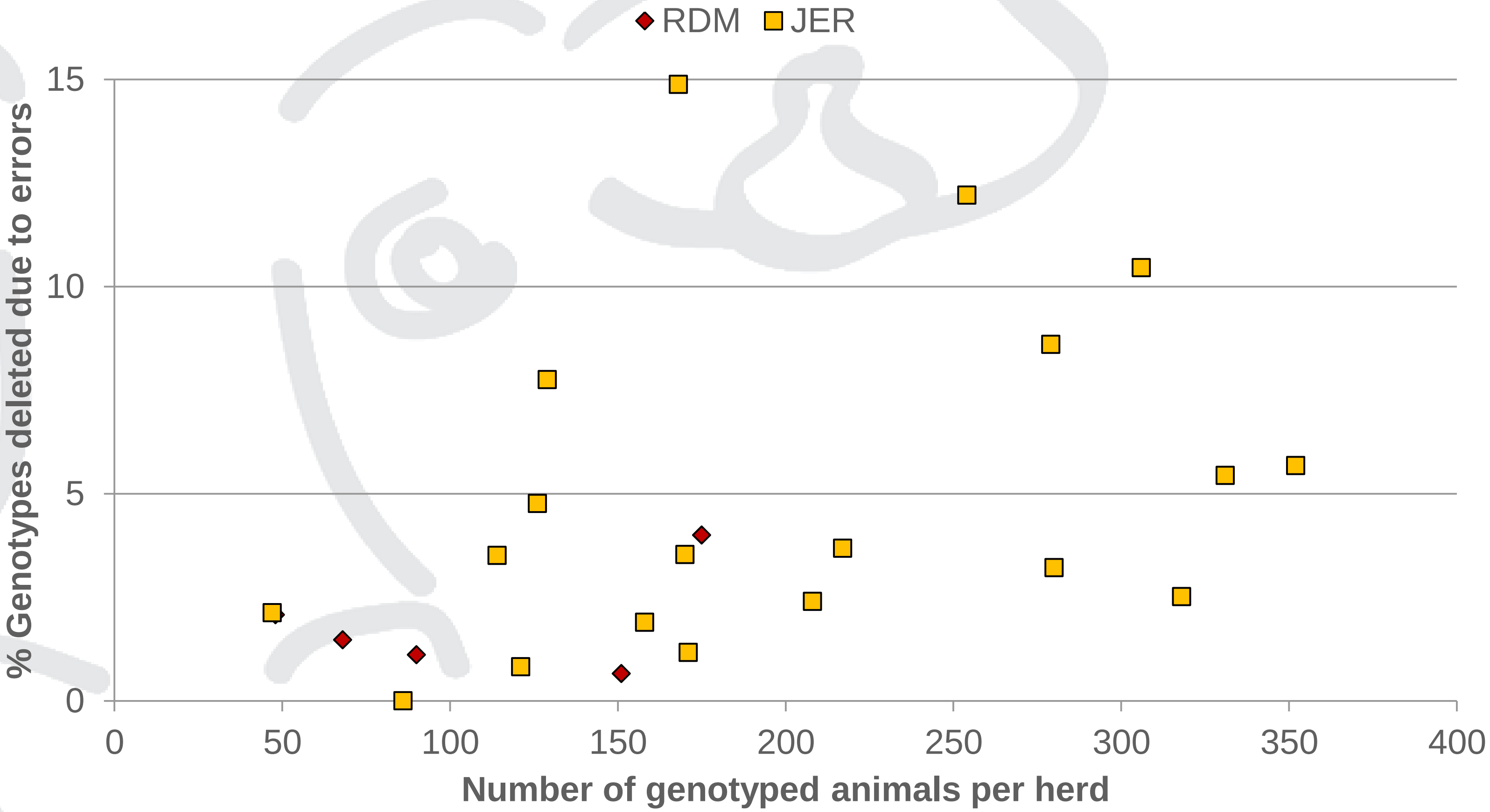
- JER: 94.5 %
- RDC: 97.9 %

Percent approved per country, LD-project

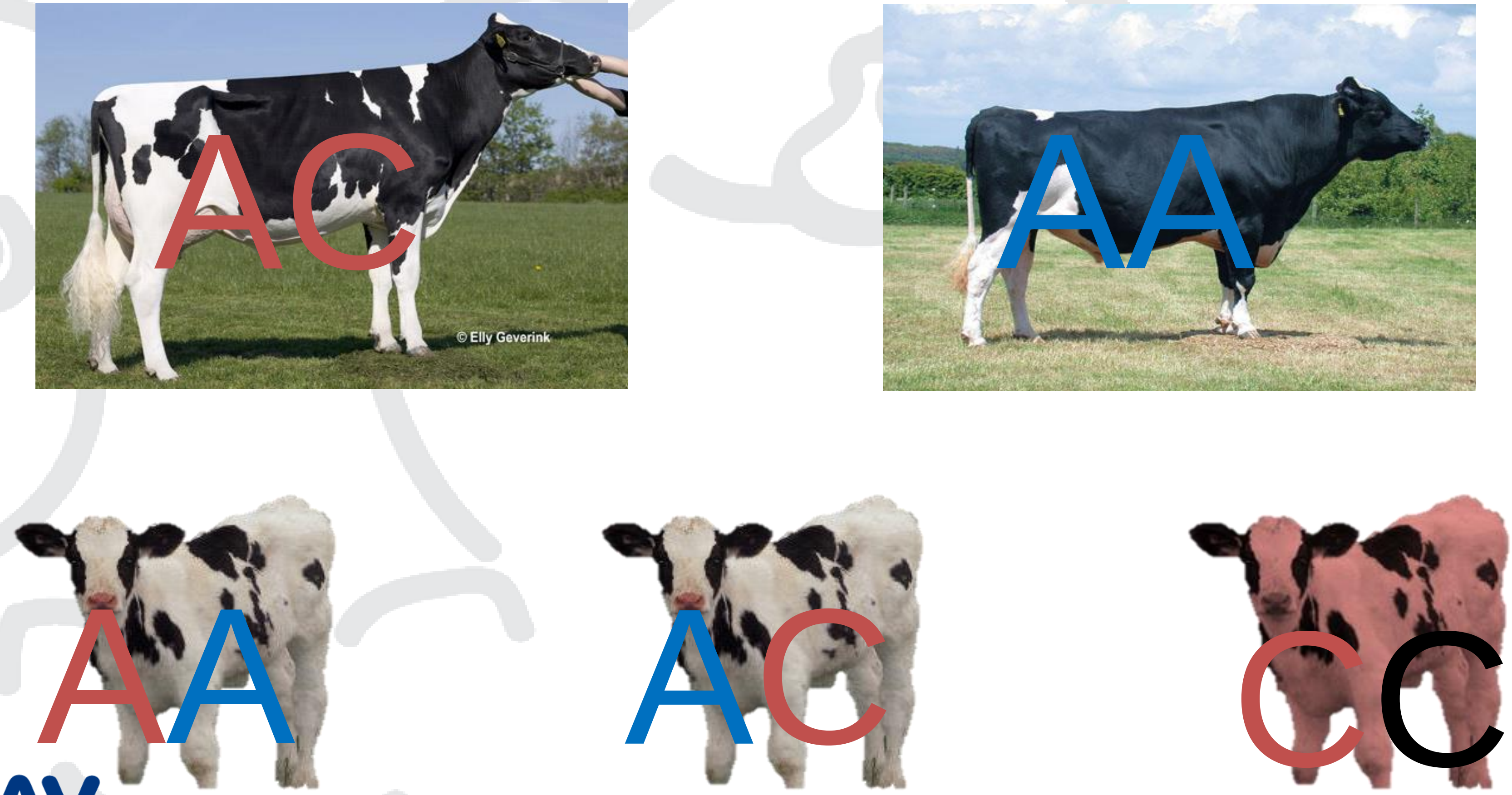
- DNK: 94.9 %
- SWE: 97.6 %
- FIN: 98.0 %

Errors are mainly ME

% deleted genotypes – Danish herds



Principle behind Mendel Error Check



Mendel Error Check - Failed

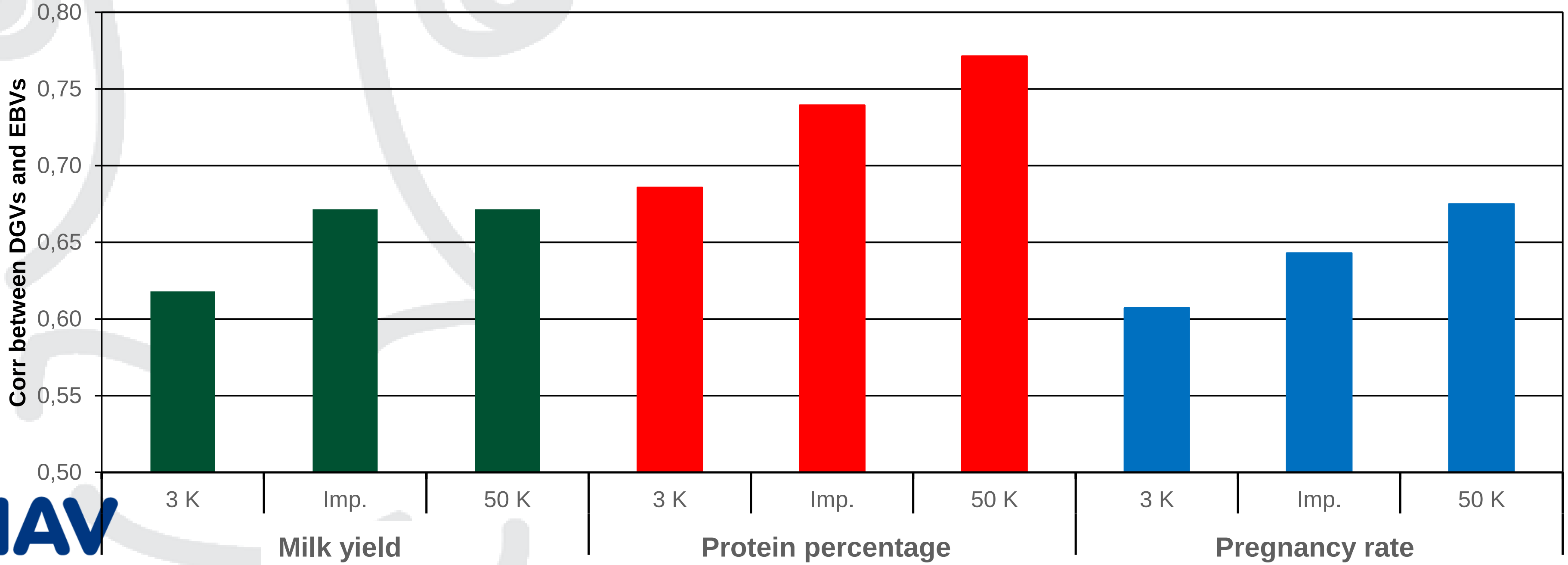
- Rejected pedigree or sire is unknown / not genotyped
  - search for potential parents
- Technique similar as for ME-Check
  - Counting opposite homozygous loci
- Result is returned to VG or Växa
  - When errors are corrected genotype can be included again – farmers responsibility

Imputation

- When pedigree, sex, and ID is verified the genotype can be imputed
- HOL - FImpute
  - Quick & accurate in homogeneous populations
- RDC & JER - Beagle
  - Slow but accurate and robust

Why do we impute?

- Accuracy of GEBV
  - 54K > imputed LD > LD



American Jersey; modified after Weigel et al., 2010

Why do we impute?

- Cows (in referencegroup - RDC and Jersey) and heifers
  - Low density chip
- Illumina Bovine 54K chip
  - Version 1 and 2
- Eurogenomics (Holstein) og GENO (RDC)
  - NLD (HOL) and Geno (RDC) has previously used another chip
- Missing calls

Basics

- SNP = Single Nucleotide Polymorphism
- At animal level

2	2	2	2	2	1	2	0	0	2	1	1	2	2	1	1

  - Homozygote for one allele = 0
  - Heterozygote = 1
  - Homozygote for the other allele = 2

- At chromosome level

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1

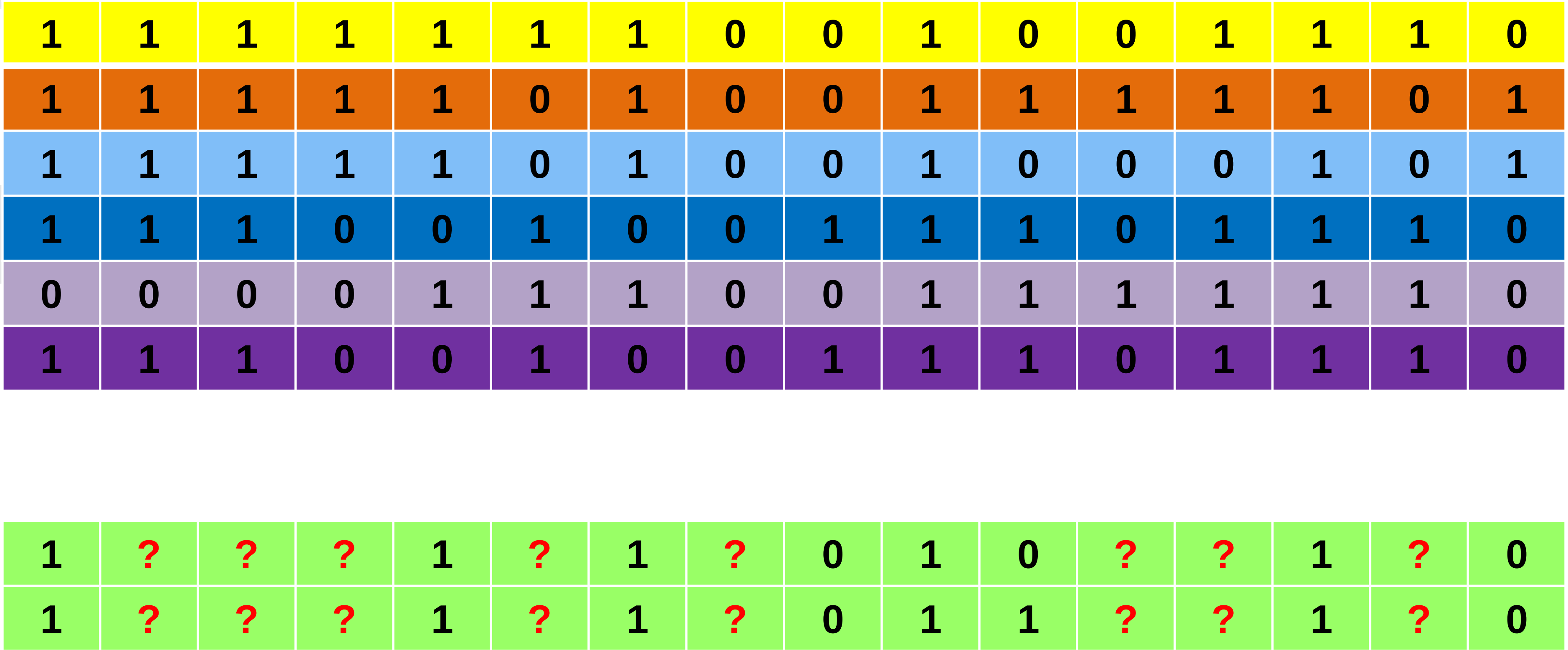
  - Variant A = 0
  - Variant B = 1

SNP Reference animals genotyped with 54K

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

Modified after Marchini & Howie, 2010

SNP candidate animals genotyped with LD

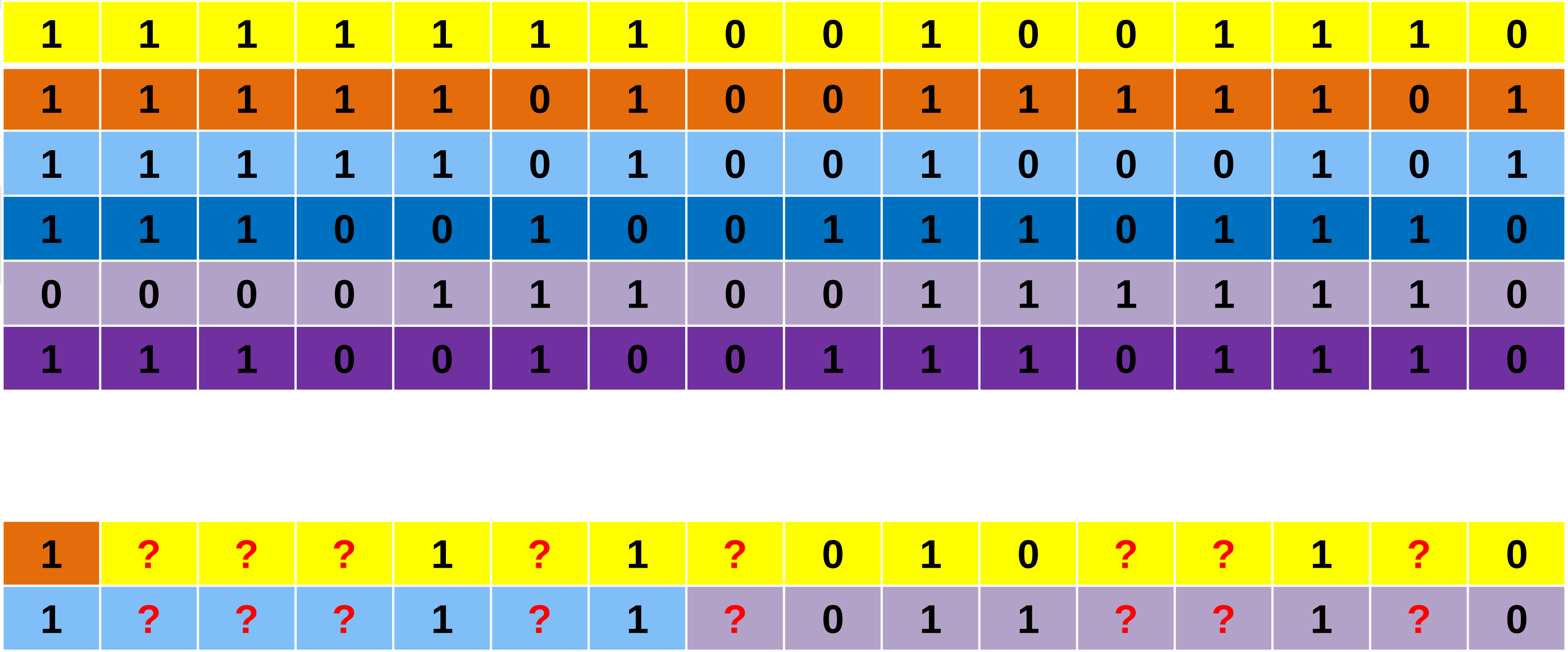


1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

1	?	?	?	1	?	1	?	0	1	0	?	?	1	?	0
1	?	?	?	1	?	1	?	0	1	1	?	?	1	?	0

Modified after Marchini & Howie, 2010

Utilize information from reference

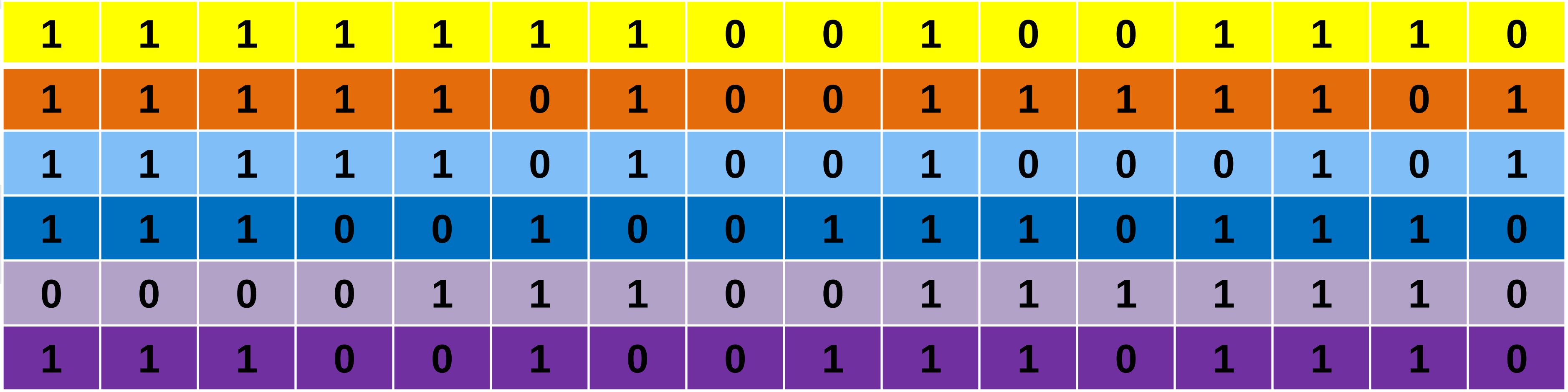


1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

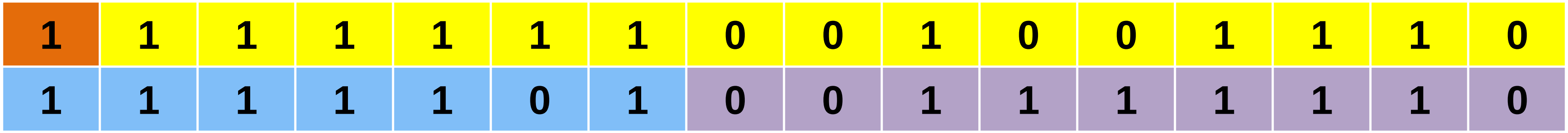
1	?	?	?	1	?	1	?	0	1	0	?	?	1	?	0
1	?	?	?	1	?	1	?	0	1	1	?	?	1	?	0

Modified after Marchini & Howie, 2010

Impute missing values



1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0



1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	0

Modified after Marchini & Howie, 2010

Imputation finished

- Genotypes checked for errors and imputed are then used as input for genomic predictions

Questions?