



IT Solutions for
Animal Production

Dairy Cattle Genetic Evaluation Using Genomic Information

Zengting Liu, F. Seefried, F. Reinhardt, and R. Reents

vit, Germany

<http://www.vit.de/>

OUTLINE

- **Genome and SNPs**
- **vit genomic evaluation system**
 - vit evaluation model
 - System components
- **Validation using simulated data**
- **Routine application issues**
- **Future developments**



Design and Application of the BovineSNP50

Curt Van Tassell

USDA

Agricultural Research Service

curtv@ars.usda.gov

301-504-6501

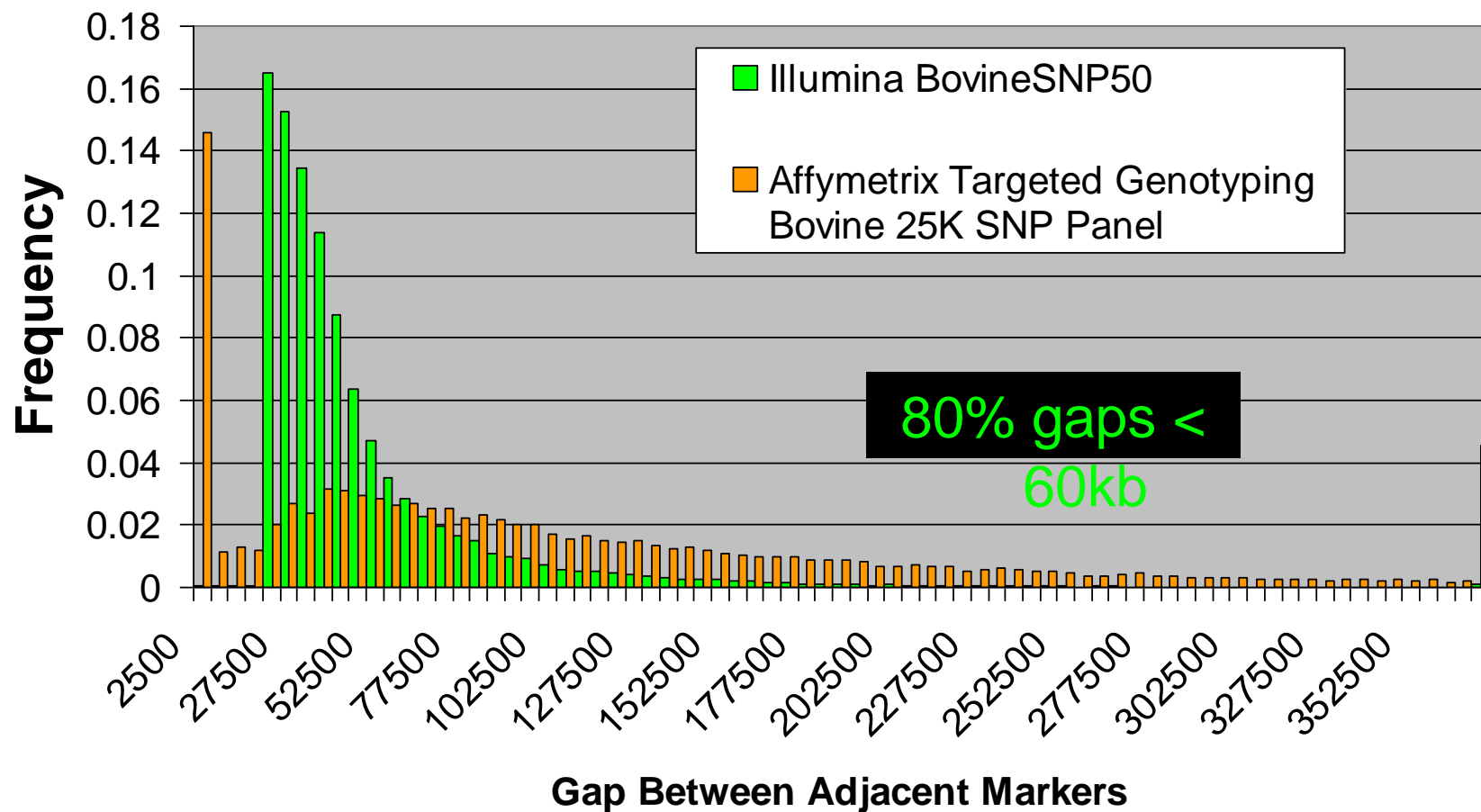


Cattle SNP Collaboration iBMAC

- Develop 60,000 Bead Illumina iSelect® assay
 - USDA-ARS Beltsville Agricultural Research Center: Bovine Functional Genomics Laboratory and Animal Improvement Programs Laboratory
 - University of Missouri
 - University of Alberta
 - USDA-ARS US Meat Animal Research Center
- Starting 60,800 beads – expected 53,000 SNPs to result
- Plan to genotype ~30,000 animals for multiple projects



Gap Distribution



Conventional Genetic Evaluation

- Elements of conventional breeding programmes
 - Pedigree data
 - Performance data: production, type, etc.
 - Artificial insemination
 - Data structure for across herd genetic evaluation
 - Computing power and algorithms
 - BLUP procedures give very reliable EBVs from bulls after a progeny test (95 –99 % reliability)
- Very successful tool in the last 40 years
 - National proofs converted to foreign country scales
- Expensive progeny testing
- EBV of young bulls → parent average (PI) $R^2 \sim 50\%$
- R^2 of cows much lower than bulls



Advance in genotyping technique (SNP)



- Genotype and alleles at a given chromosome location
- State of the art technology
 - ➔ ~ 50.000 SNPs from an individual animal for 200 EUR

Genotype:

Animal 1:	...AGGCACC GCAATCCACG GAGGC	T	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	TT
	...AGGCACC GCAATCCACG GAGGC	T	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	
Animal 2:	...AGGCACC GCAATCCACG GAGGC	A	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	AA
	...AGGCACC GCAATCCACG GAGGC	A	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	
Animal 3:	...AGGCACC GCAATCCACG GAGGC	T	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	AT
	...AGGCACC GCAATCCACG GAGGC	A	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	
Animal n:	...AGGCACC GCAATCCACG GAGGC	A	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	AA
	...AGGCACC GCAATCCACG GAGGC	A	ACGC CCTCACCGGA GGTTTCGCTC TCCACGG...	

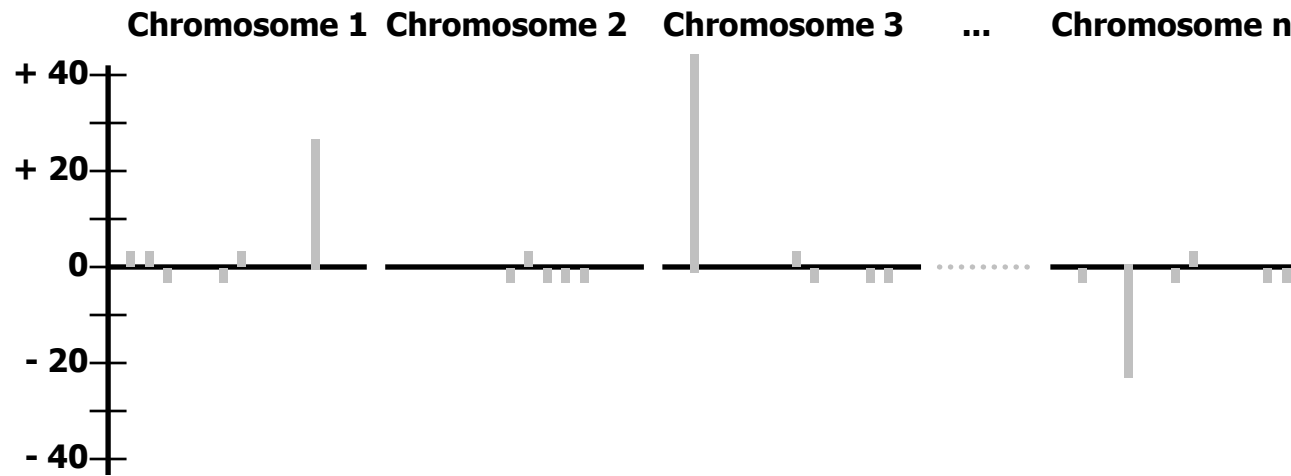
e.g. Chromosome 6, Position # 43.675.239



Genomic Breeding Value



SNP effect



$$1 + 1 - 1 - 1 + 1 + 25 - 1 + 1 - 1 - 1 - 1 - 1 + 42 + 1 - 1 - 1 - 1 - 1 - 1 - 22 - 1 + 1 - 1 - 1 = +38$$



Int'l Development in Genomic Selection

- **Already running projects (AUS, NZL, NLD, CAN/USA, FRA and IRL)**
- **Routine implementation by evaluation centres**
 - **AIPL of USDA (VanRaden 2007)**
 - **LIC of NZL (Harris et al. 2008)**
 - **CR-Delta of NLD (de Roos et al. 2007)**
 - **INRA of FRA**
- **Interbull Task Force for genomic evaluation**
- **Implementation in 2009 by Interbull**



Genomic Selection in Germany

- ***GenoTrack* project among German institutions**
- **vit database project for genomic selection**
- **vit project for genomic enhanced genetic evaluation**
- **Routine genotyping started**



vit Genomic Evaluation Model

$$q_i = \mu + a_i + \sum_{j=1}^p m_j + e_i$$

- **DYD / YD of bull / cow as dependent variable**
 - **Weighted by EDC**
 - **Deregressed proofs for missing DYD**
- **Polygenic effect with full pedigree relationship**
 - **Identical modelling as in conventional evaluation**
 - **Phantom parent groups**
- **Variable marker variances**
 - **All markers included**



Features of vit Genomic Model

- **Deterministic approach (not Gibbs sampling)**
- **Residuals weighted by EDC (cows & bulls jointly)**
- **Polygenic effect and relationship matrix with phantom parent grouping**
- **Various approaches to SNP variance implemented**
 - **A non-linear model (VanRaden 2008)**
 - **Robust regression model (Draper & Smith)**
- **Alternative weighting functions**



Features of vit Genomic Model (cnt'd)

- Ranging from pure marker model to pure polygenic model, depending on weightings
- Special model: equal marker variance (linear BLUP model)
- Special Gauss-Seidel algorithm with periodical residual update (Legarra & Misztal 2008)
- Using pre-defined marker variances, if needed
- Using previous solutions as starting values



Components of vit Genomic System

- **Combining GEBV with conventional EBV**
 - **Reliability for combined EBV**
- **Alternative model** $q_i = \mu + a_i + u_i + e_i$ with $u_i = \sum_j^p m_{ij}$
- **Calculating genomic relationship and inbreeding**
- **Approximating reliabilities of GEBV**
- **Estimating allele frequencies of base population**
- **Deregression procedure for selected traits**
- **Programs for pre- & post-processing**



Alternative Weighting Functions

Model variant	Function
Linear model with equal variance (EQ)	$\sigma_j^2 = \bar{\sigma}^2 \times 1$
Non-linear model, linear weight (LW)	$\sigma_j^2 = \bar{\sigma}^2 \times \hat{s}_j $
exponential weight (E1, USDA)	$\sigma_j^2 = \bar{\sigma}^2 \times 1.12^{ \hat{s}_j }$
exponential weight (E2, USDA)	$\sigma_j^2 = \bar{\sigma}^2 \times 1.25^{ \hat{s}_j }$
quadratic weight (Q1)	$\sigma_j^2 = \bar{\sigma}^2 \times \hat{s}_j^2$
quadratic weight with limits (Q2)	min/ max <i>for</i> $ \hat{s}_j $
Polygenic model (PG)	$\sigma_j^2 \approx \bar{\sigma}^2 \times 0$

\hat{s}_j is standardised effect estimate of marker j



vit Genomic Evaluation Software

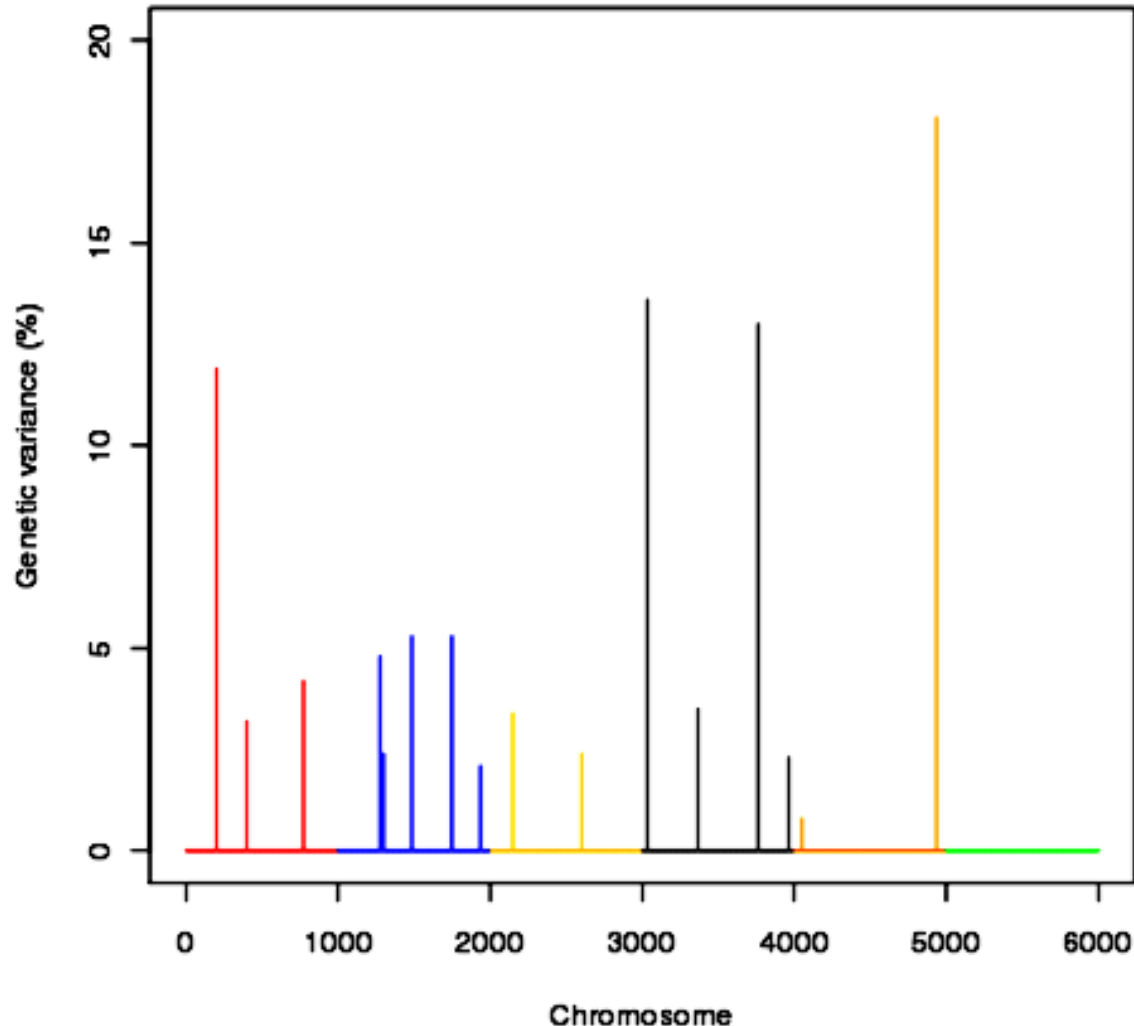
- **Fortran 90 source programs**
 - **Library and object files managed by makefile**
- **Automation with shell scripts**
 - **Including master and sequence scripts**
- **Fast and well-defined convergence**
- **Marker data processing across all traits**
- **Genomic evaluation by trait group**
 - **Difference in conventional GE timing among traits**
 - **Difference in # animals with phenotype & markers**
 - **Bulls or cows requiring different files**

Validation Using a Simulated Data Set

- **Data set from 2008 QTL-MAS Workshop**
- **Simulated marker and phenotypic data**
 - **Swine-type pedigree structure**
 - **6000 SNP markers evenly distributed on 6 chromosomes**
 - **48 QTLs with different variances (none on chr. 6)**
 - **Estimation set: 4665 animals (165/1500/1500/1500)**
 - **Validation set: 1200 animals (400/400/400)**
 - **No fixed and random effects, except QTLs**
 - **45 (450) sires (dams) with 100 (10) progeny**
- **True values: $h^2=0.304$, genetic std dev = 1.17**
- **Model PG (.303, .001), genomic (.001, .303)**



True Variances of Simulated QTLs

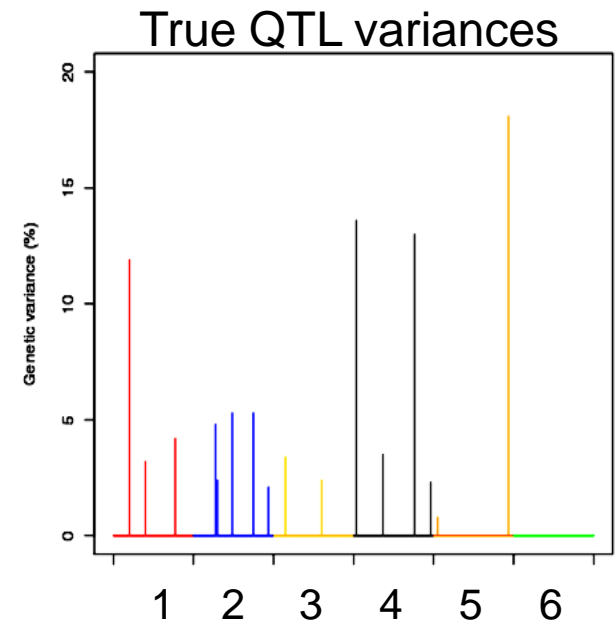
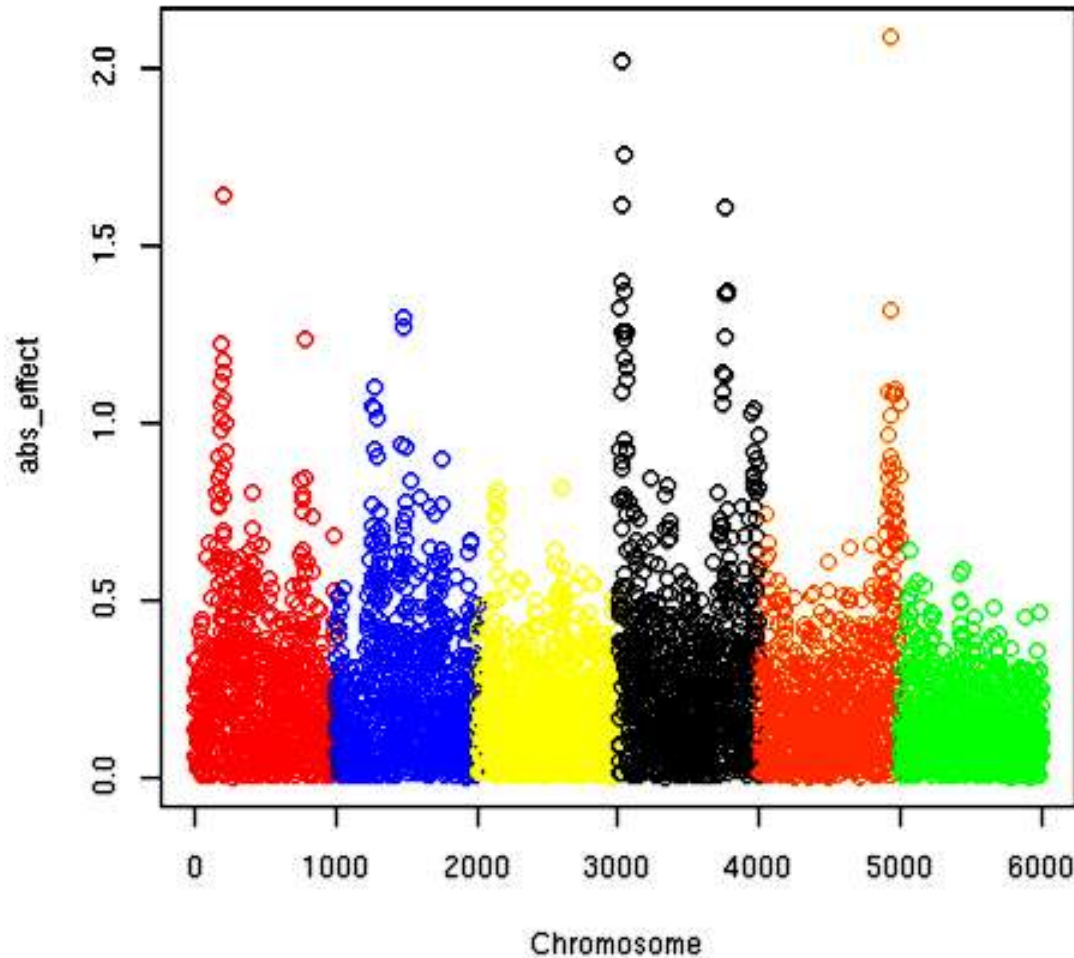


16 (of 48) big SNPs explain 96.3% genetic variance



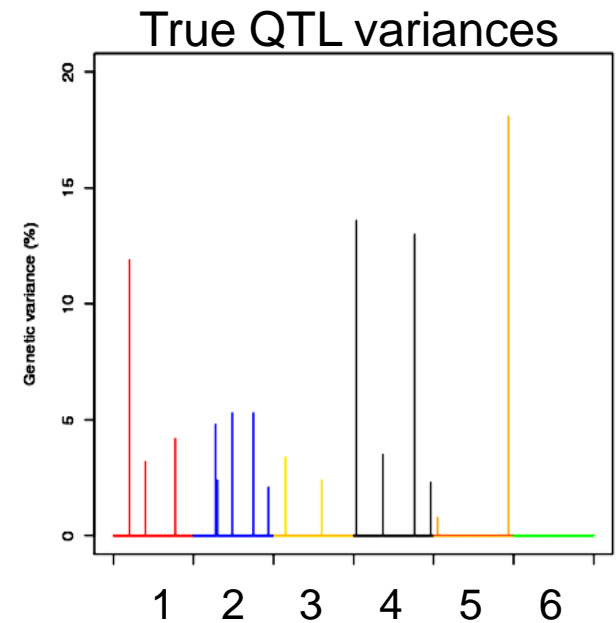
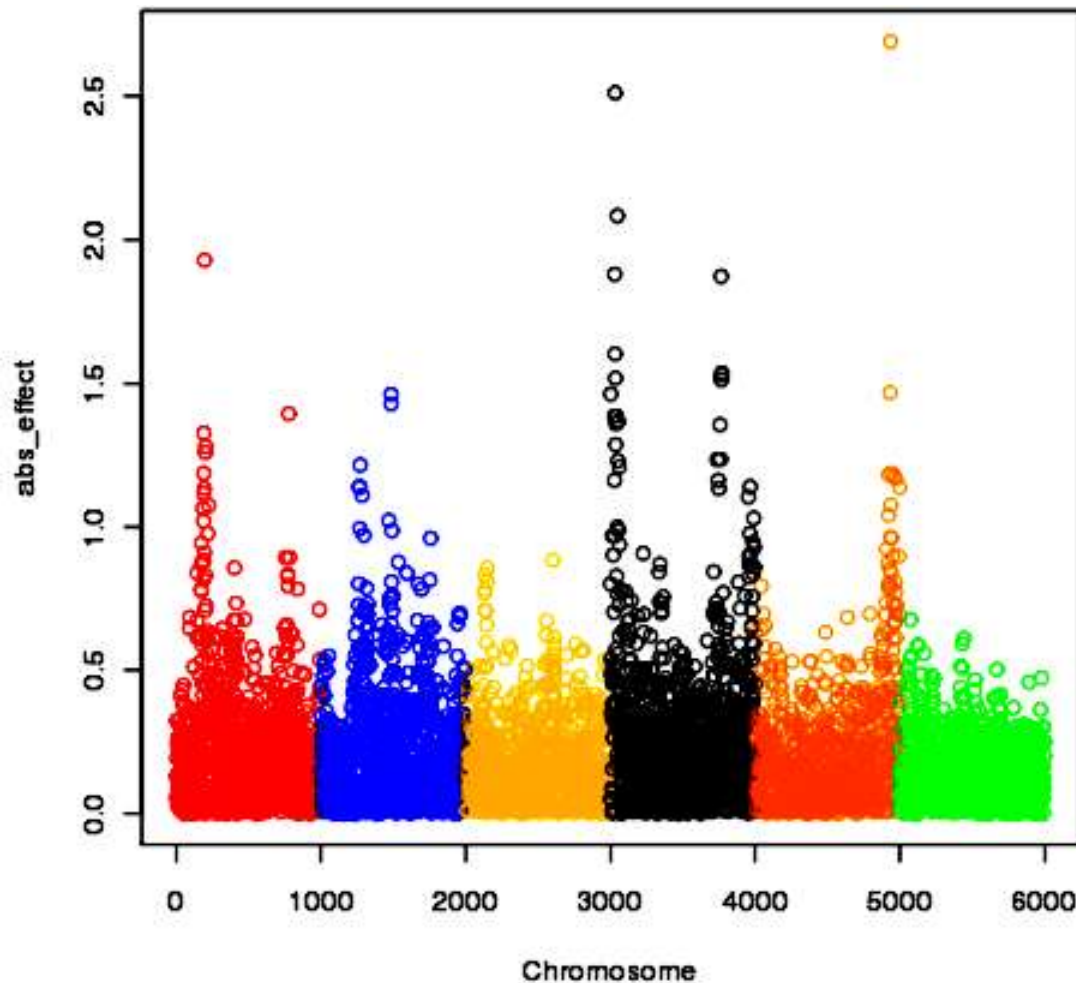
SNP Effect Estimates

Model EQ (linear BLUP model $\sigma_j^2 = \bar{\sigma}^2 \times 1$)



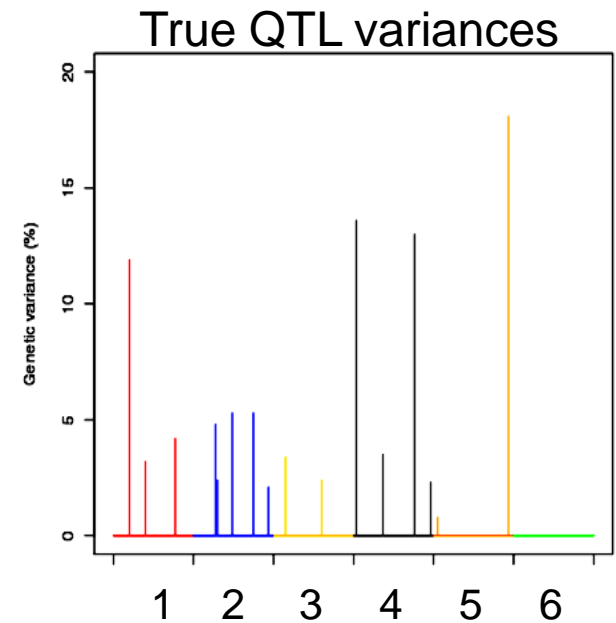
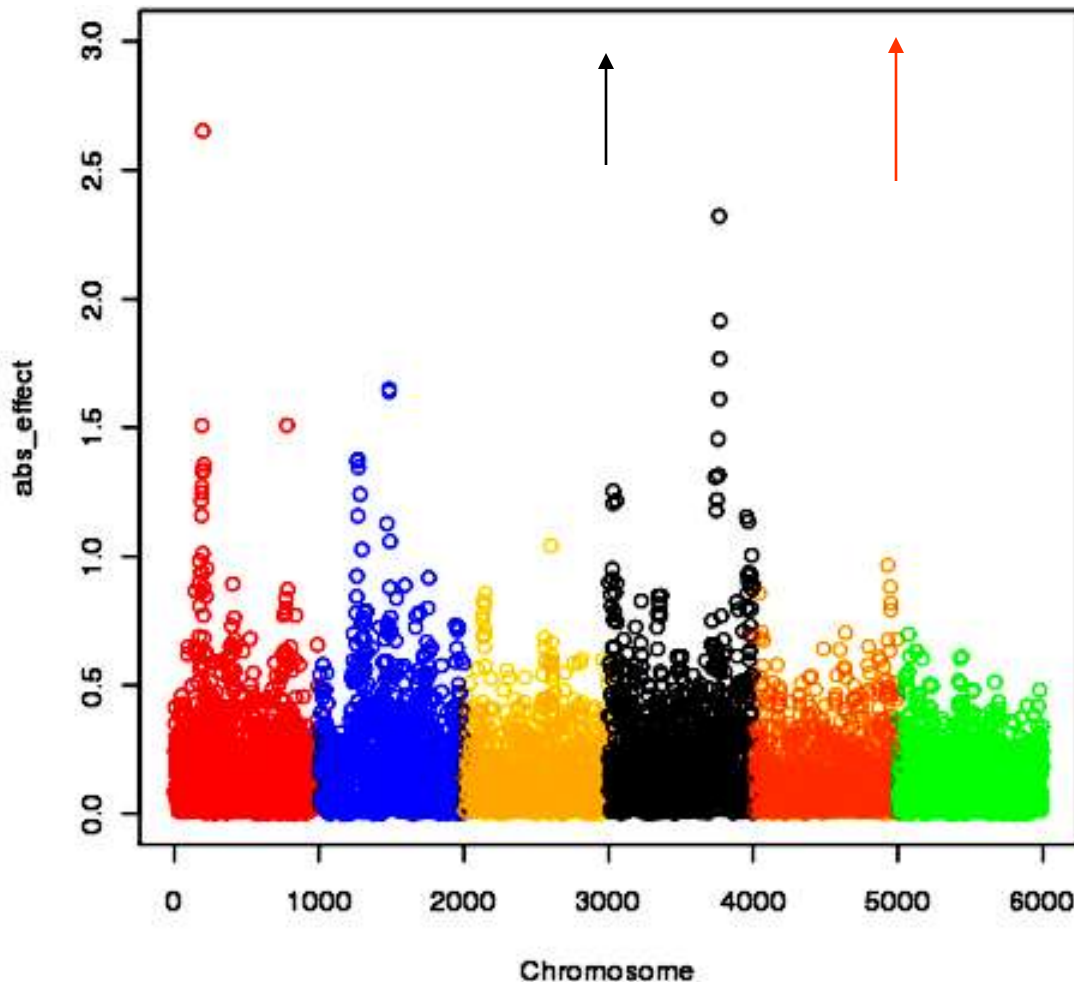
SNP Effect Estimates

Model E1 (non-linear model $\sigma_j^2 = \bar{\sigma}^2 \times 1.12^{|\hat{s}_j|}$) vit 



SNP Effect Estimates

Model E2 (non-linear model $\sigma_j^2 = \bar{\sigma}^2 \times 1.25^{|\hat{s}_j|}$) vit 

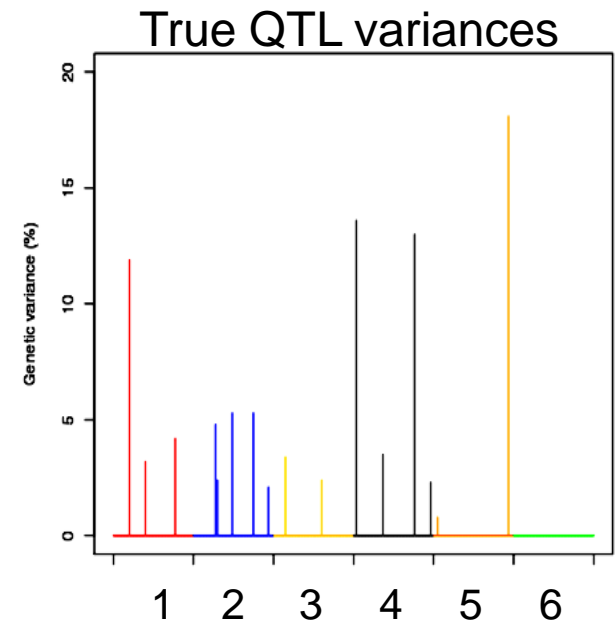
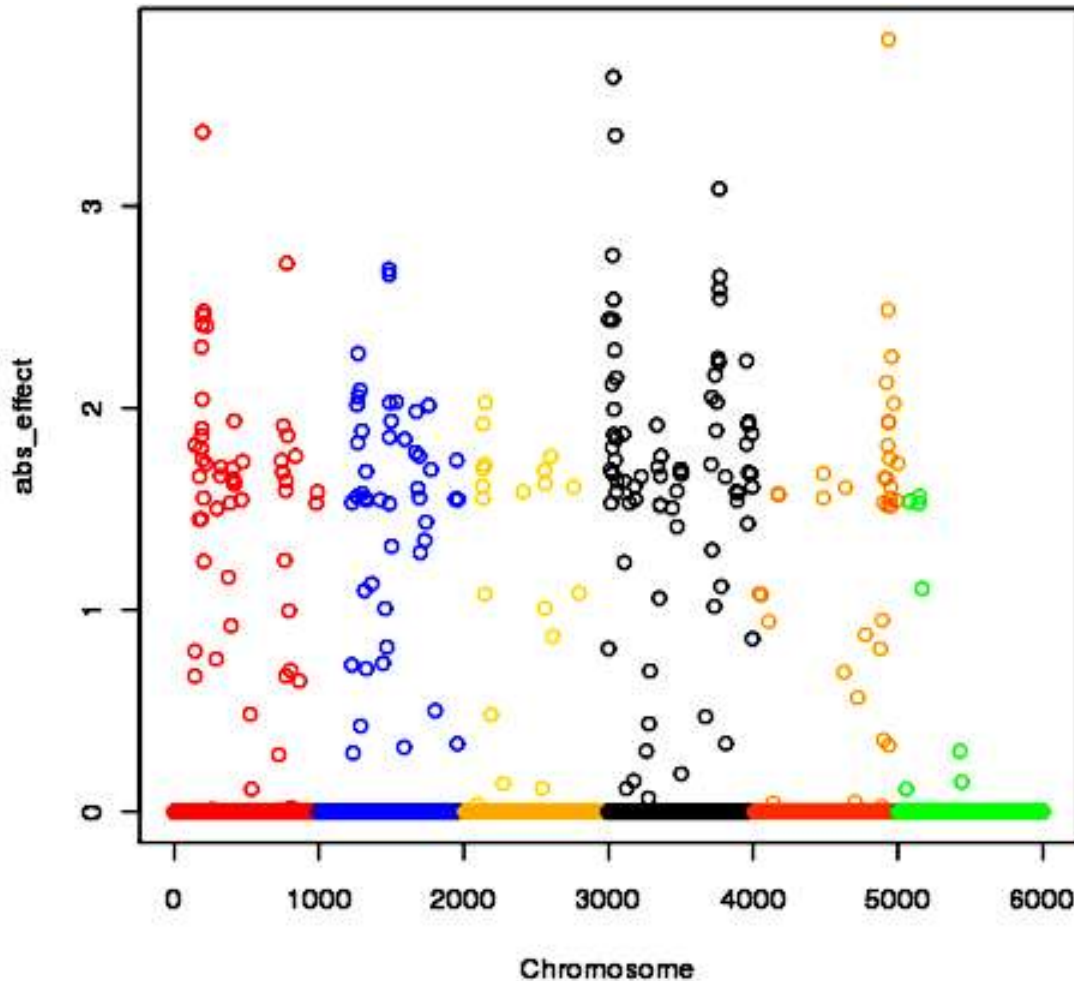


2 big markers on chromosome 4 & 5



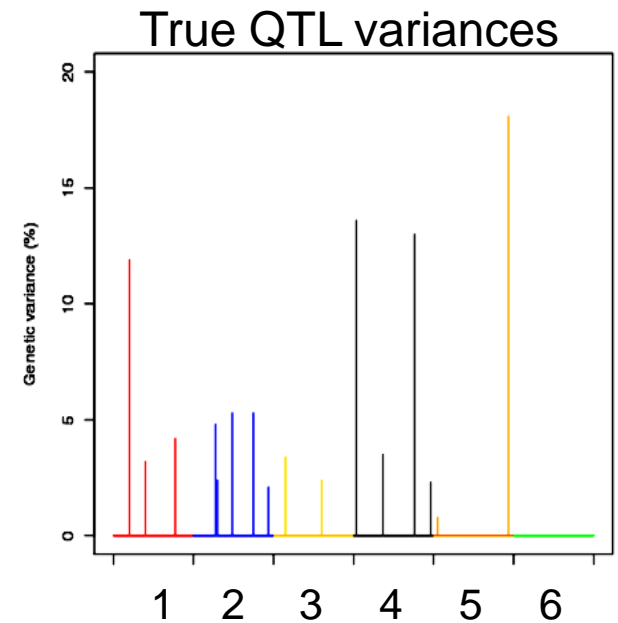
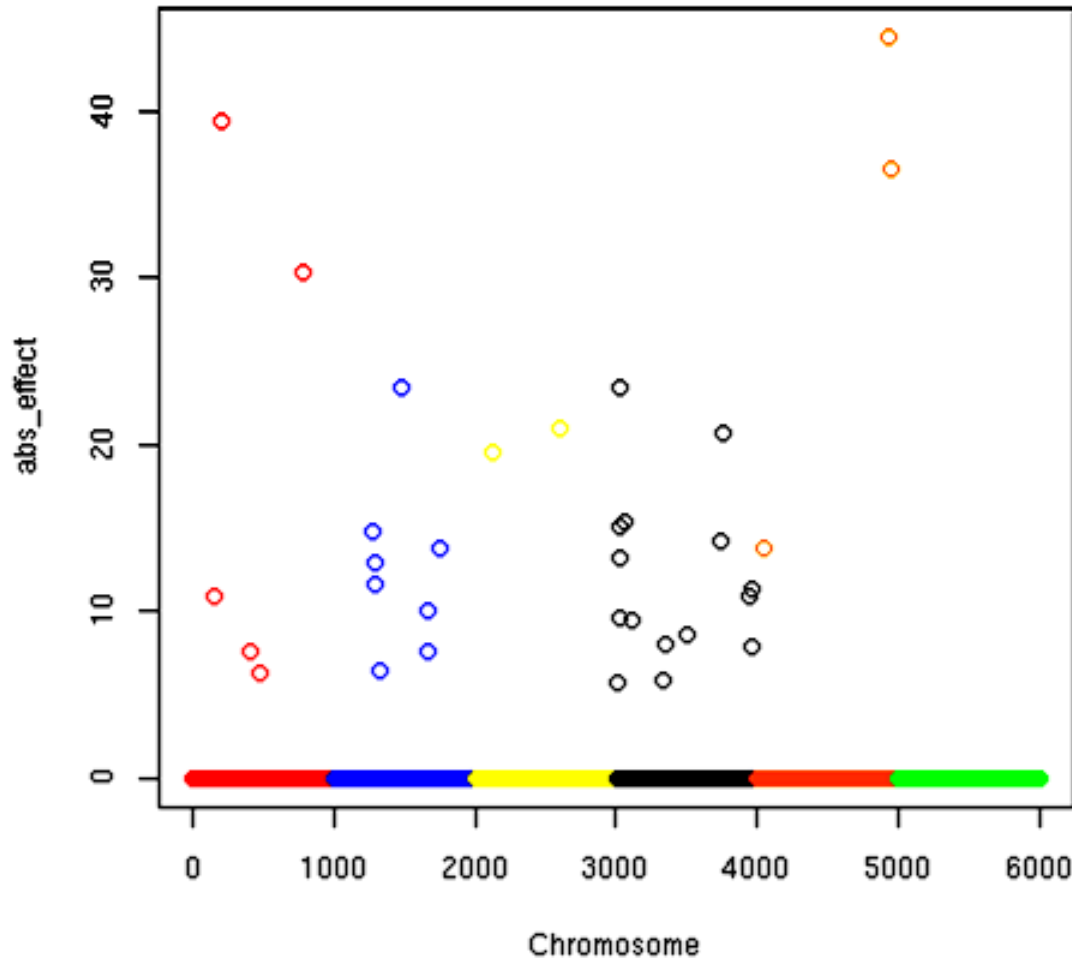
SNP Effect Estimates

Model LW (non-linear model $\sigma_j^2 = \bar{\sigma}^2 \times |\hat{s}_j|$)



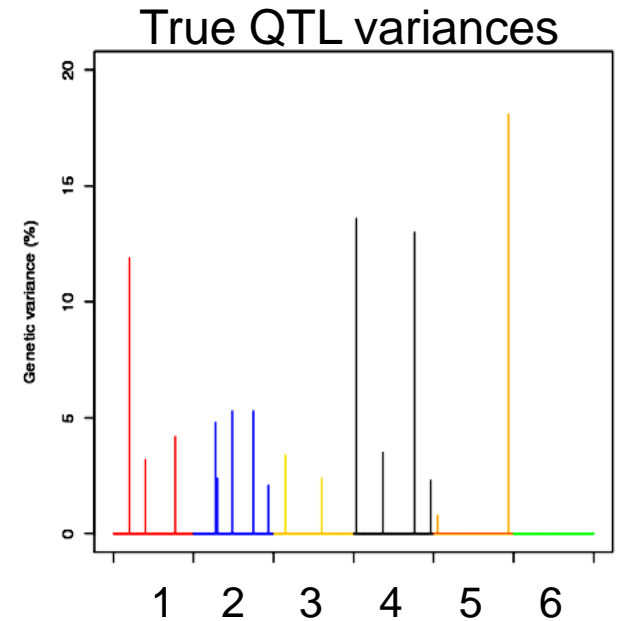
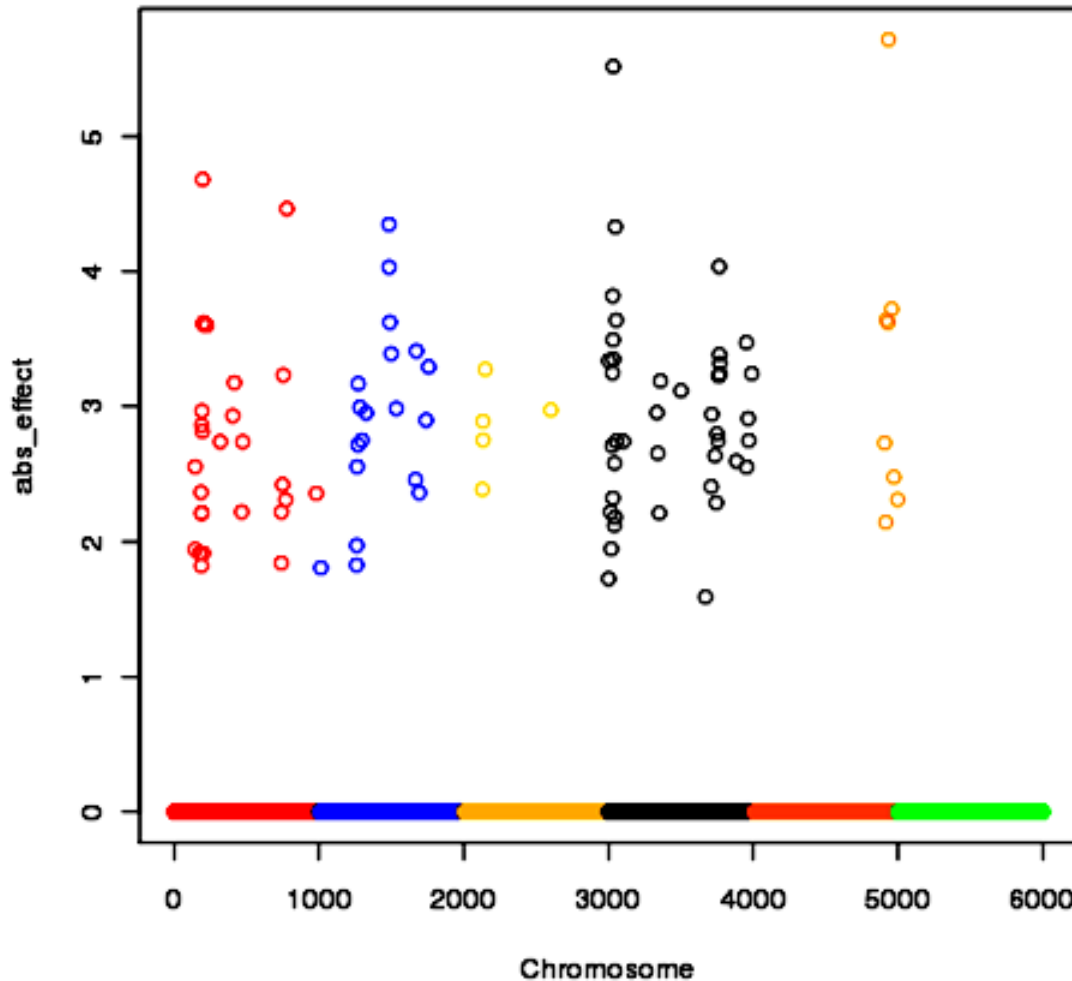
SNP Effect Estimates

Model Q1 (non-linear model $\sigma_j^2 = \bar{\sigma}^2 \times \hat{s}_j^2$)

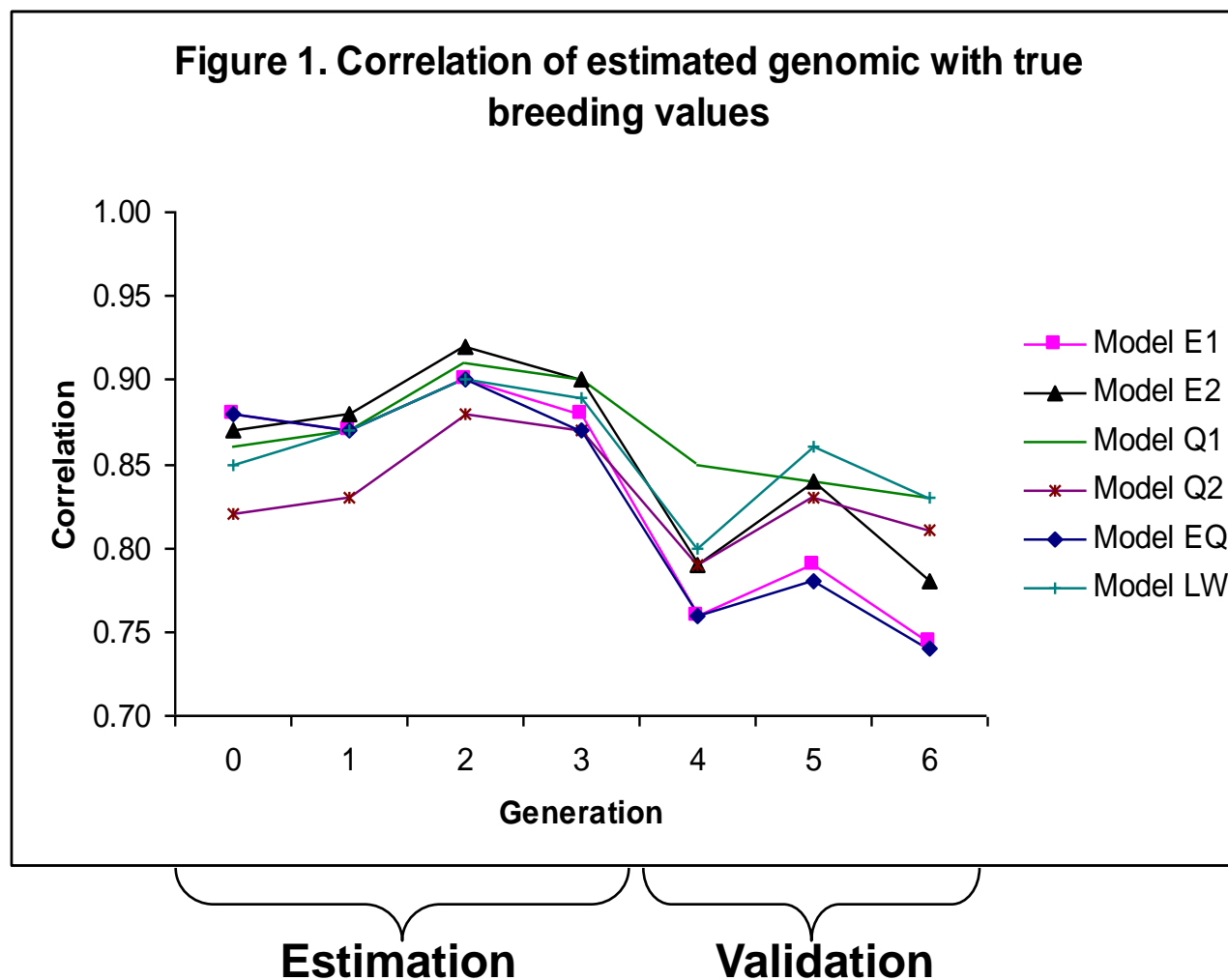


SNP Effect Estimates

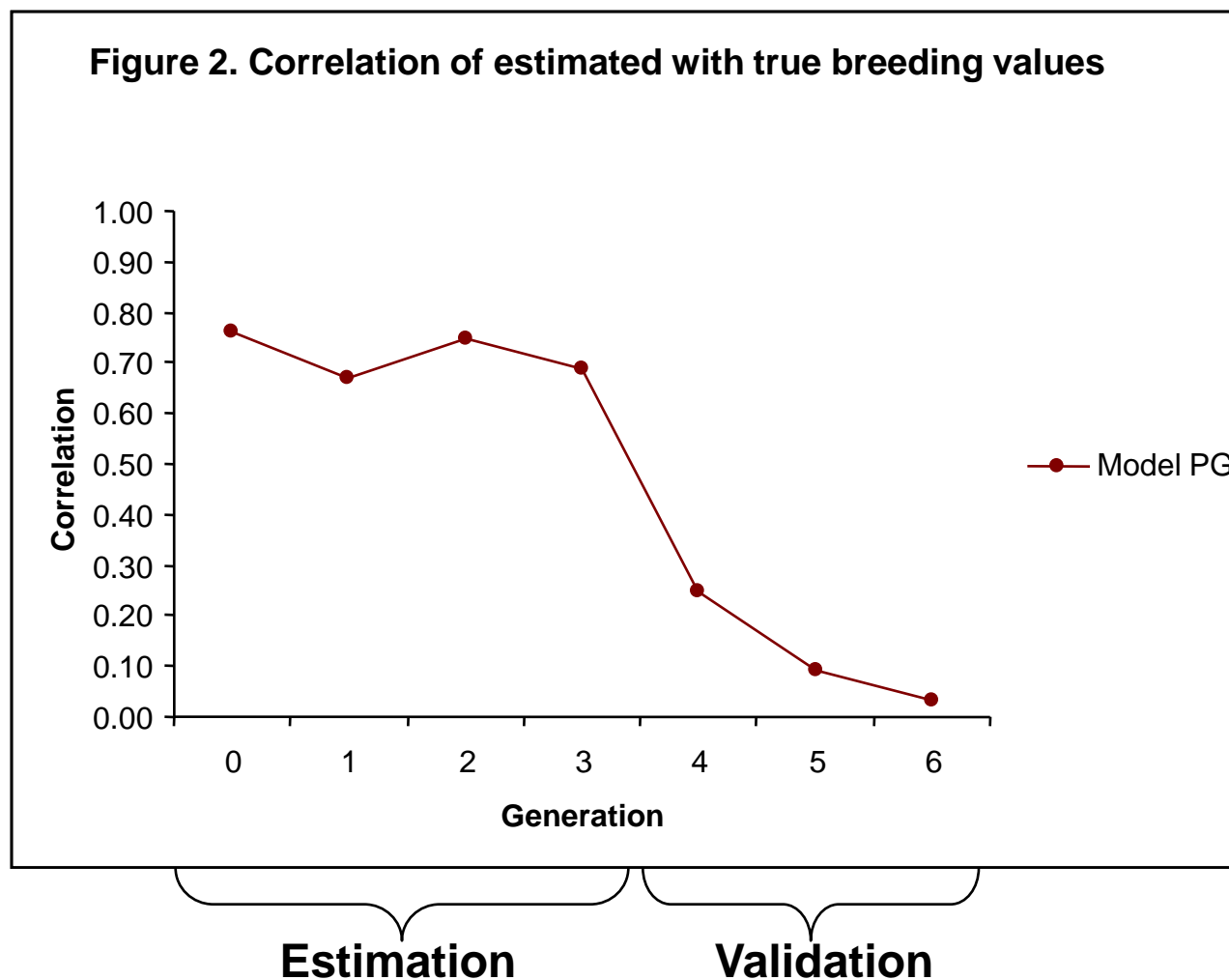
Model Q2 (non-linear model) $\sigma_j^2 = \bar{\sigma}^2 \times \hat{s}_j^2 \min/\max \text{ for } |\hat{s}_j|$



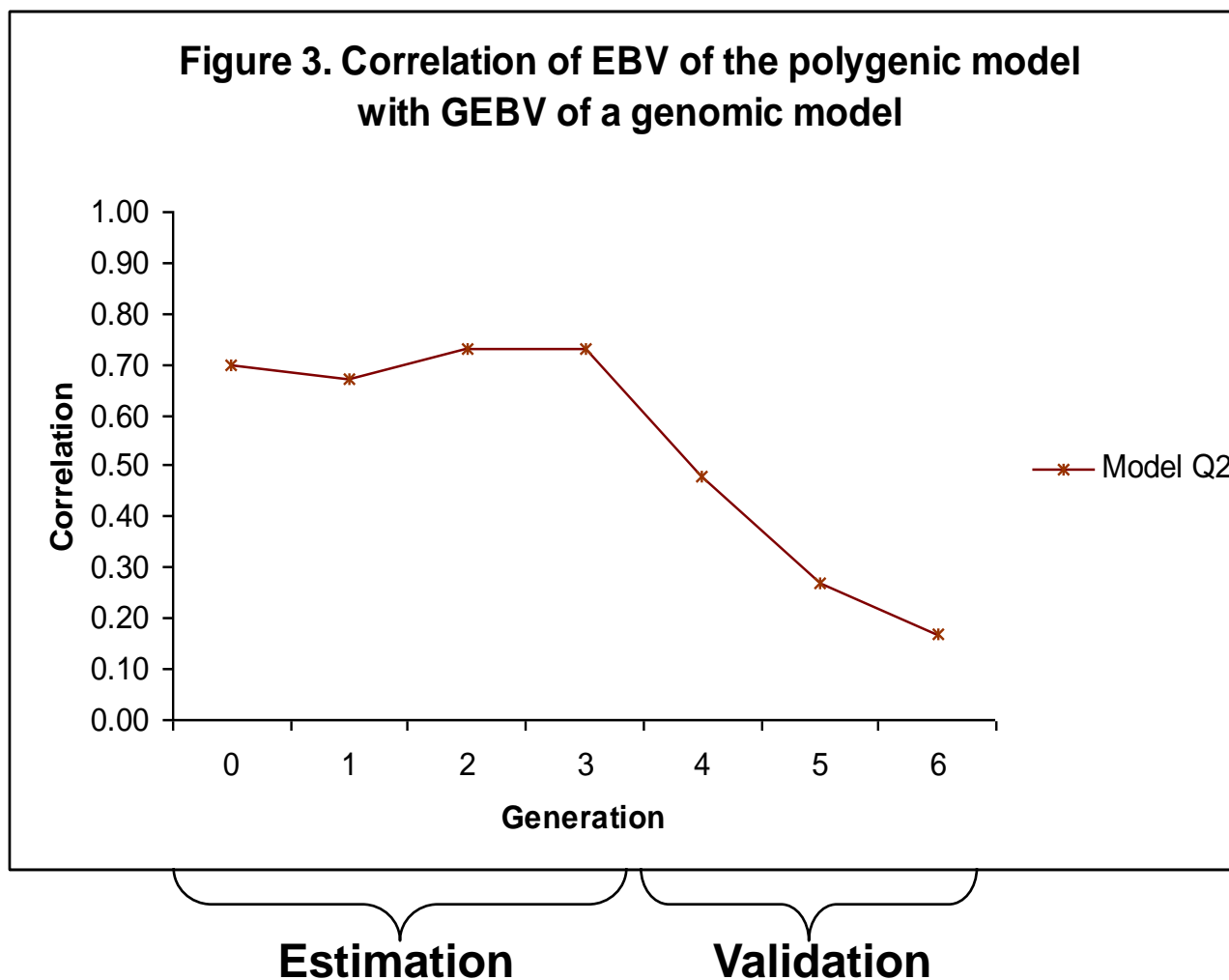
Correlation between Estimated and True Genomic Breeding Values



Correlation between Estimated and True Genomic Breeding Values



Correlation between GEBV & EBV of polygenic model PG



Correlations between GEBV for 45 Sires

Genomic Model	True BV	EBV of Model PG
Model EQ	.95	.96
Model LW	.94	.95
Model E1	.95	.96
Model E2	.96	.96
Model Q1	.96	.94
Model Q2	.93	.93



Correlations between GEBV for 450 Dams

Genomic Model	True BV	EBV of Model PG
Model EQ	.90	.85
Model LW	.88	.80
Model E1	.90	.85
Model E2	.91	.85
Model Q1	.88	.77
Model Q2	.85	.76



Summary of the Results

- **Differences in marker effect estimates**
 - Almost all QTLs detected with several SNP markers each
 - Heavier weights led to fewer big genes
 - Weaker weights led to many small genes
- **All genomic models have high accuracy of GBV**
 - Minor differences between models
- **High correlations between model PG and genomic models for sires and dams**
- **Genomic models predictability better than PG**
- **Genomic model ranking depends on simulation**
- **Genomic model ranking may change for field data**



Routine Application Issues

- **Marker data processing**
 - Editing, etc. (a long list of processes)
- **Phenotype data processing (DYD & EDC)**
 - Production traits (M/F/P), DYD (RRC) (deregression)
 - Udder health (S), DYD (RRC) (deregression)
 - Type traits (DYD for 25 traits, some new traits)
 - Workability (DYD for 3 traits, DMG and/or MBK)
 - Calving (CE & SB for maternal/direct effects, deregres.)
 - Female fertility (DYD for 5 traits)
 - Longevity (DYD as sum of YD of daughters, EDC)
- **Calculation of Indices and reliabilities**
- **Incorporation of all the steps related to genomic selection into a genetic evaluation system**





Birth years	Milk yield			Interval first to last service heifer		
	No. Bulls	Corr. X 100	Difference	No. Bulls	Corr. X 100	Difference
1990	888	99.5	-0.01	856	54.8	-14.55
1991	899	99.6	-0.59	875	56.0	-3.88
1992	962	99.7	-0.36	940	69.6	-10.54
1993	1031	99.7	0.66	1015	70.1	-8.55
1994	1158	99.7	0.19	1145	69.0	-12.08
1995	1237	99.8	-0.16	1222	71.8	-8.58
1996	1283	99.8	0.11	1272	69.2	0.49
1997	1335	99.7	-0.79	1324	69.1	-2.83
1998	1172	99.6	-0.15	1165	72.2	3.80
1999	1140	99.6	-0.18	1130	72.7	3.99
2000	1086	99.7	0.10	1080	75.3	3.46
2001	1091	99.7	-0.10	1087	73.6	7.02
2002	1046	99.5	0.11	1042	70.5	6.42
2003	815	98.5	2.38	811	58.9	-3.77



Routine Application Issues

- **Validation of the genomic models**
 - Young predictee bulls needed
- **Reliability calculation for GEBV**
 - Considering LD loss due to recombination etc.
 - Comparable to conventional reliabilities
- **Impact of strong genomic pre-selection on conventional genetic evaluation**
- **A more realistic simulation study is conducted**
- **Interbull evaluation issues (Task Force)**



Future Developments

- **Single trait genomic model to be extended**
 - Conventional GE models dictate genomic models
 - For maternal/direct effect model (CE & SB)
 - For multiple parity model (heifer & cow fertility)
 - For complementary traits (DMG and MBK)
 - For test-day models (lactations, RRC)

- **Using MT-MACE model from DEU-FRA project**
 - Calculation of multi-trait DYD
 - Approximation of multi-trait EDC
 - Modifying the MT-MACE model (Tarres et al. 2007)

- **Modifying conventional genetic evaluation to consider strong genomic pre-selection**



Future Developments

- **Importance of polygenic effect**
 - Variation in relationships among genotyped animals
 - For low heritability traits, e.g. female fertility
- **Genomic evaluation considering strong selection in the past (e.g. milk yield)**
 - More realistic simulation study
 - Including ungenotyped relatives with phenotypes
- **Extending to new traits, e.g. health traits**

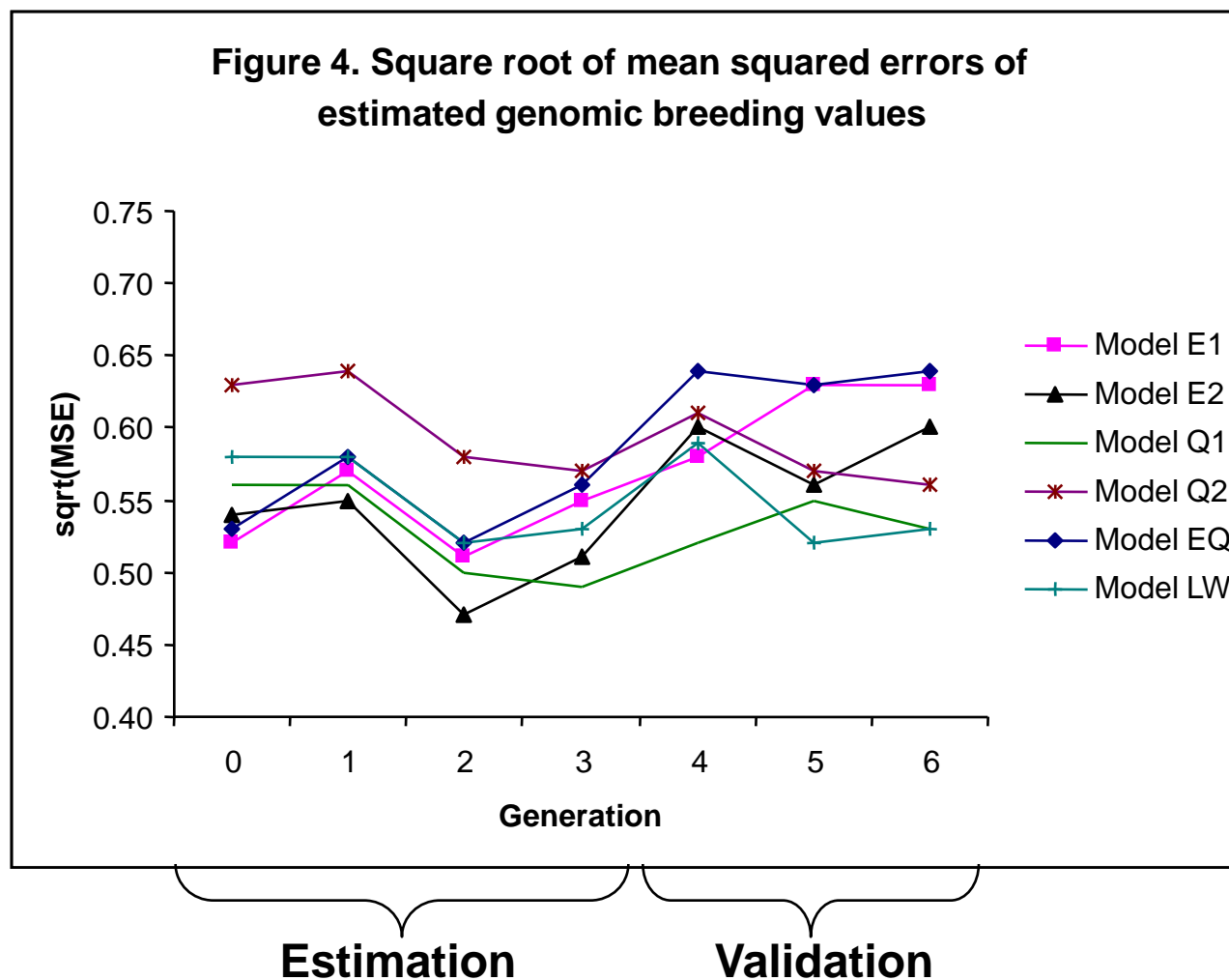


vit

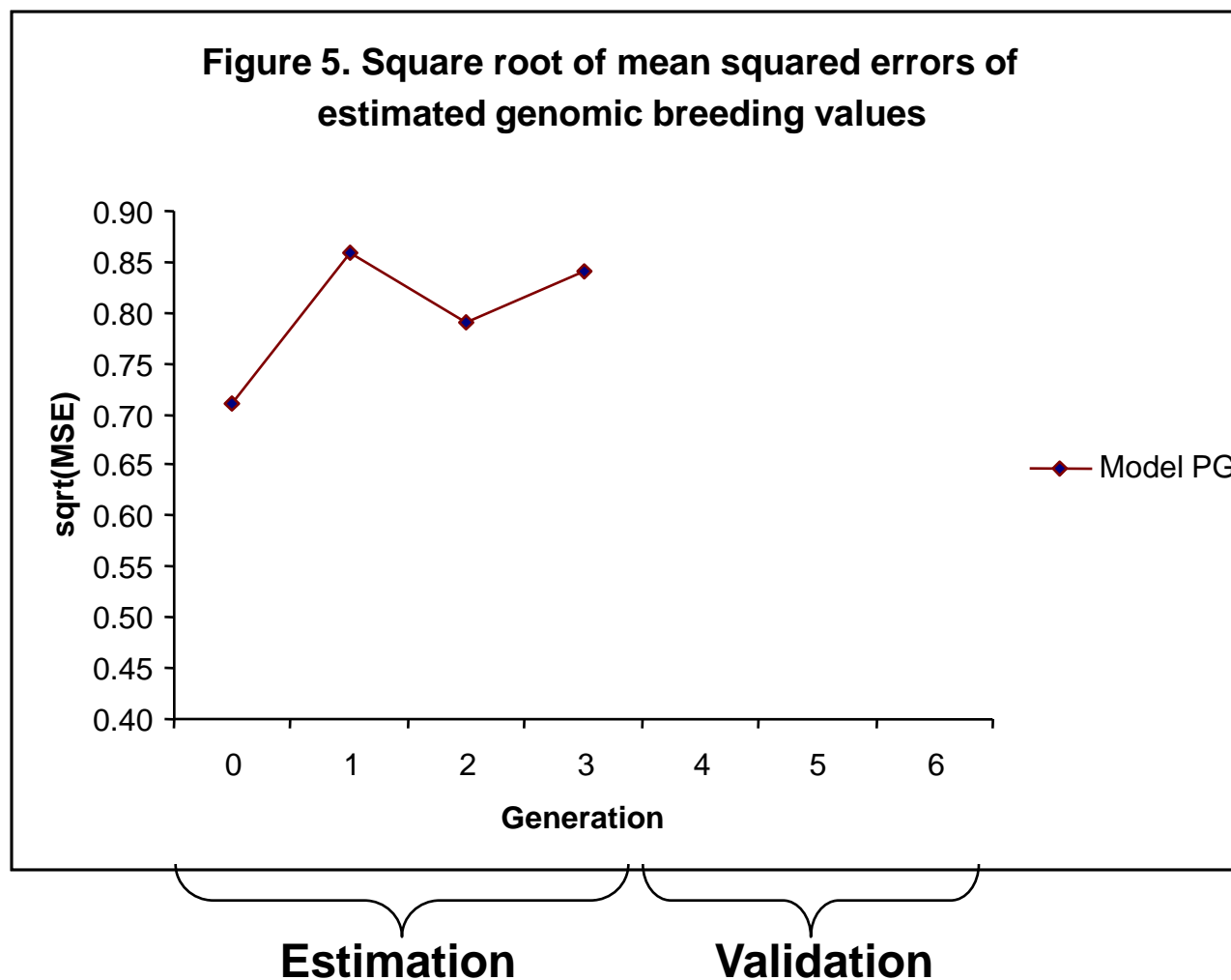


Thank you !

Mean Squared Error of GEBV



Mean Squared Error of GEBV



History

- Goals
 1. Implement whole genome selection
 2. QTL detection
- Challenge
 - Needed at least 30,000 good markers
- Problem
 - Product did not exist

SNP Available for Assay Design

With MAF (18%)

- Next Generation Sequencing 62,042
- Bovine HapMap Consortium 33,836
- DPI, US-MARC,UA 10,574

InSilico SNPs (72%)

- Assembly SNP (Filtered) 278,429
- Baylor Interbreed 123,049
- BAC and BAC-end Derived 89,832
- INRA 764

Total: 598,648 85% Infinium II (1 Bead)

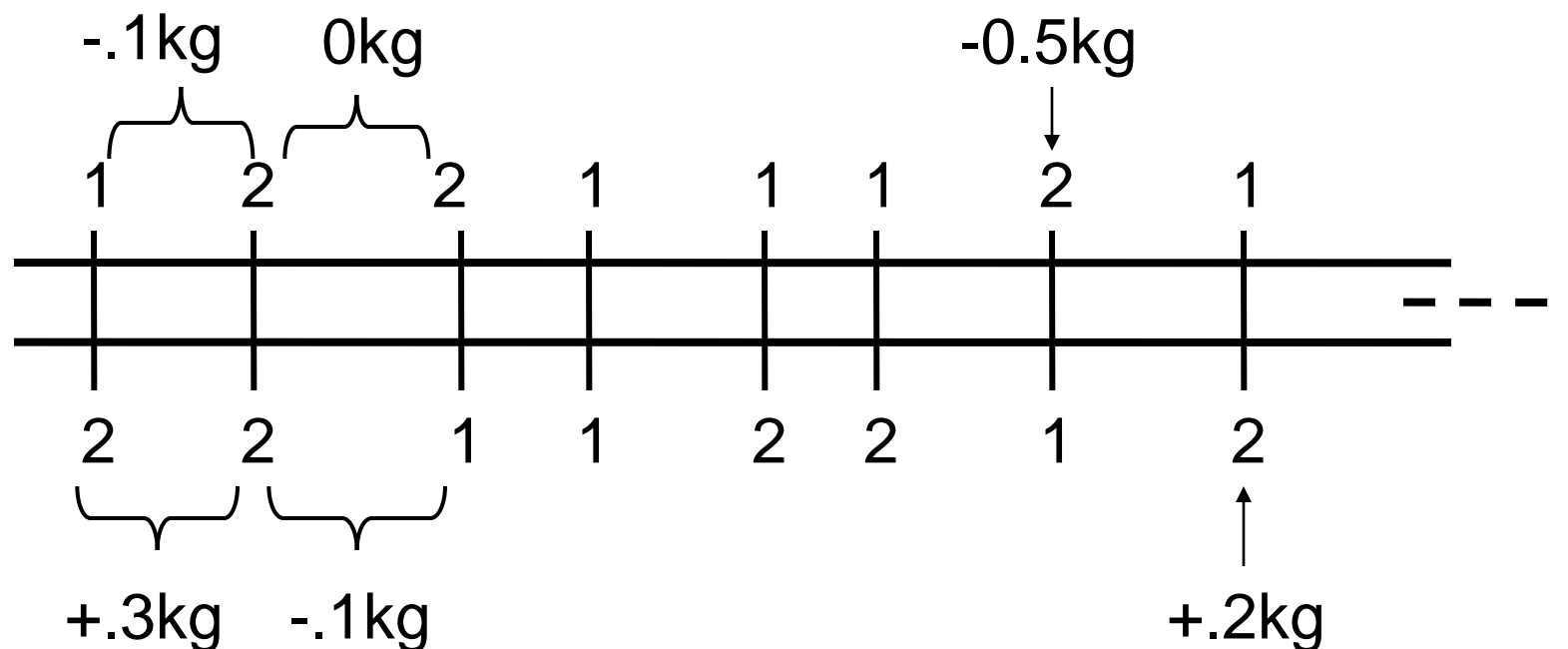
15% Infinium I (2 Beads)



Practical application

- Genotype animals that have reliable EBVs from conventional genetic evaluation
- Calculate regression formulas so that SNPs explain well the conventional EBV
- Use the regression formulas derived by historic data to evaluate young animals
 - r^2 significantly higher than parent average
 - r^2 of cows similar as bulls
- Select these young genotyped animals

Estimating Genomic Breeding Values



50,000+ haplotype or genotype estimates

$$\text{EGBV} = (\text{Sum of estimates} \mid \text{genotypes})$$



Understanding Cancer and Related Topics

Understanding SNPs and Cancer



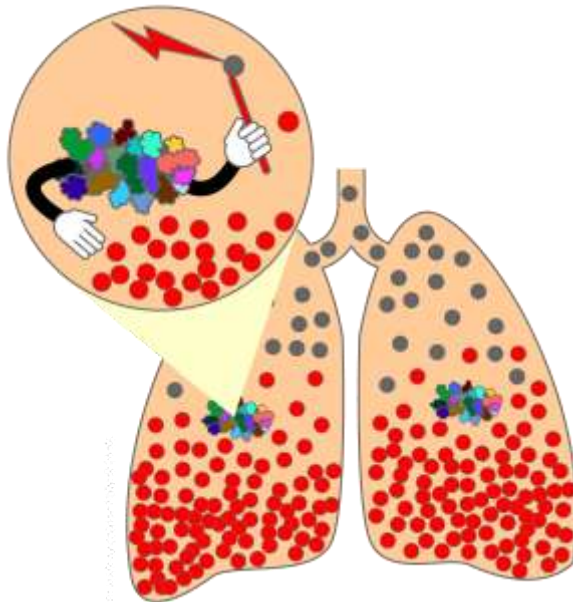
Developed by:

Susan Greenhut, M.S.

Donna Kerrigan, M.S.

Jeanne Kelly

Brian Hollen



Explains tiny variations in the human genome called Single Nucleotide Polymorphisms (SNPs) that can influence a person's health. Shows how SNPs occur in both coding and noncoding regions and can cause silent, harmless, harmful, or latent effects. Shows how SNPs can be markers for cancer. Suggests that SNPs may also be involved in the different levels of individual cancer risk observed. Suggests that, in the future, SNPs databases may be used to improve cancer diagnosis and treatment planning.

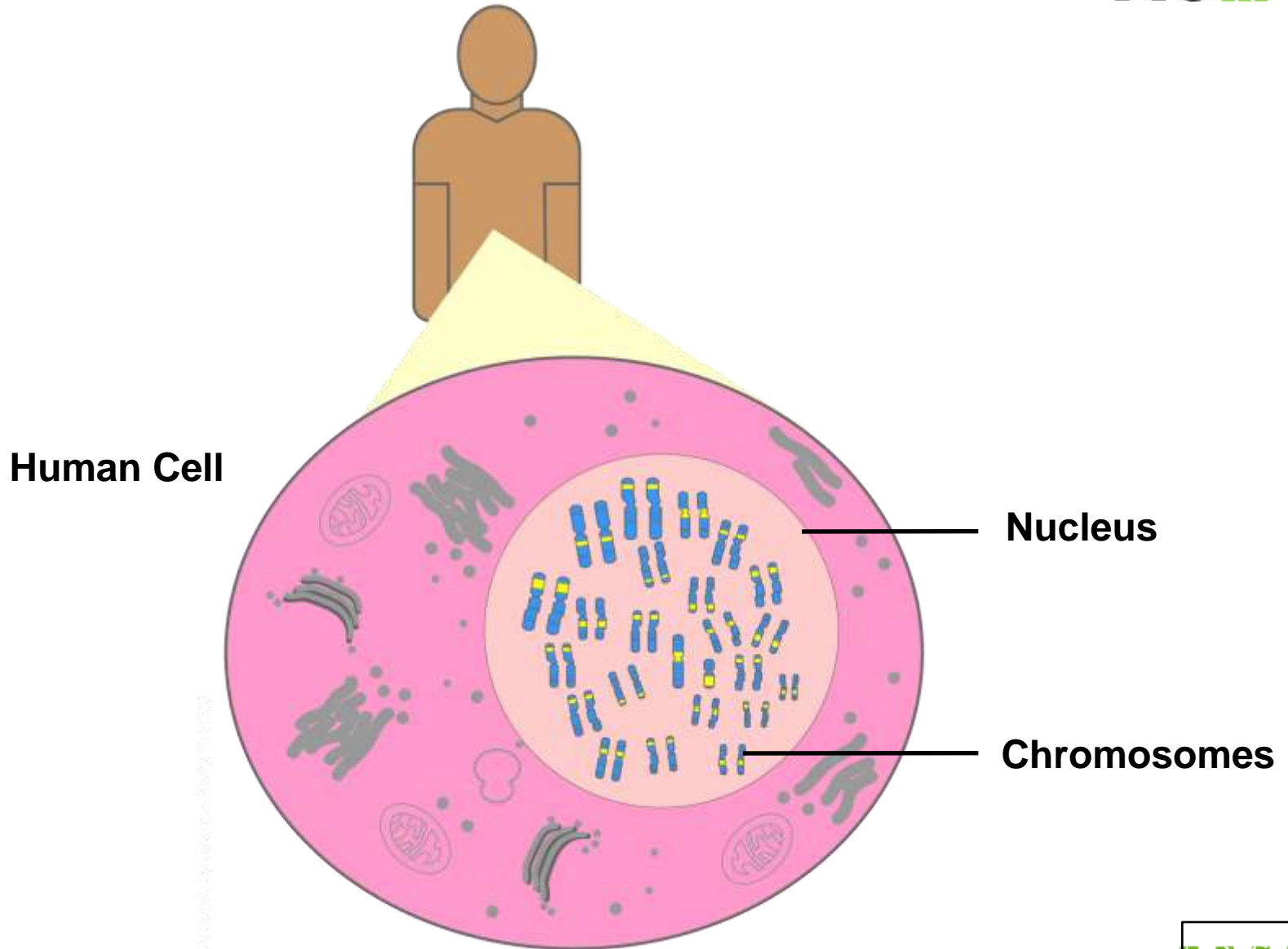
These PowerPoint slides are not locked files. You can mix and match slides from different tutorials as you prepare your own lectures. In the Notes section, you will find explanations of the graphics.

*The art in this tutorial is copyrighted and **may not** be reused for commercial gain.*

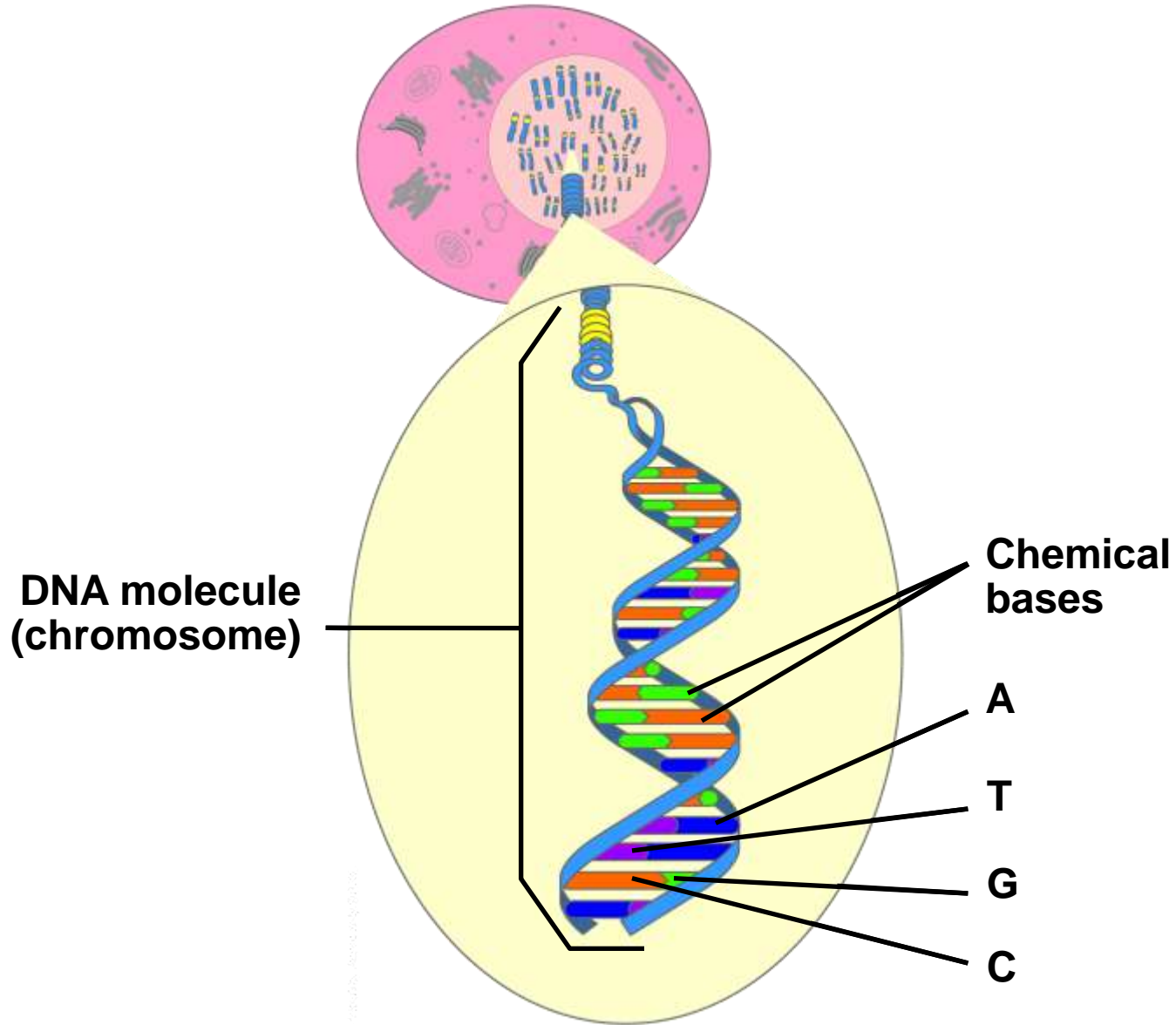
Please do not remove the NCI logo or the copyright mark from any slide.

These tutorials may be copied only if they are distributed free of charge for educational purposes.

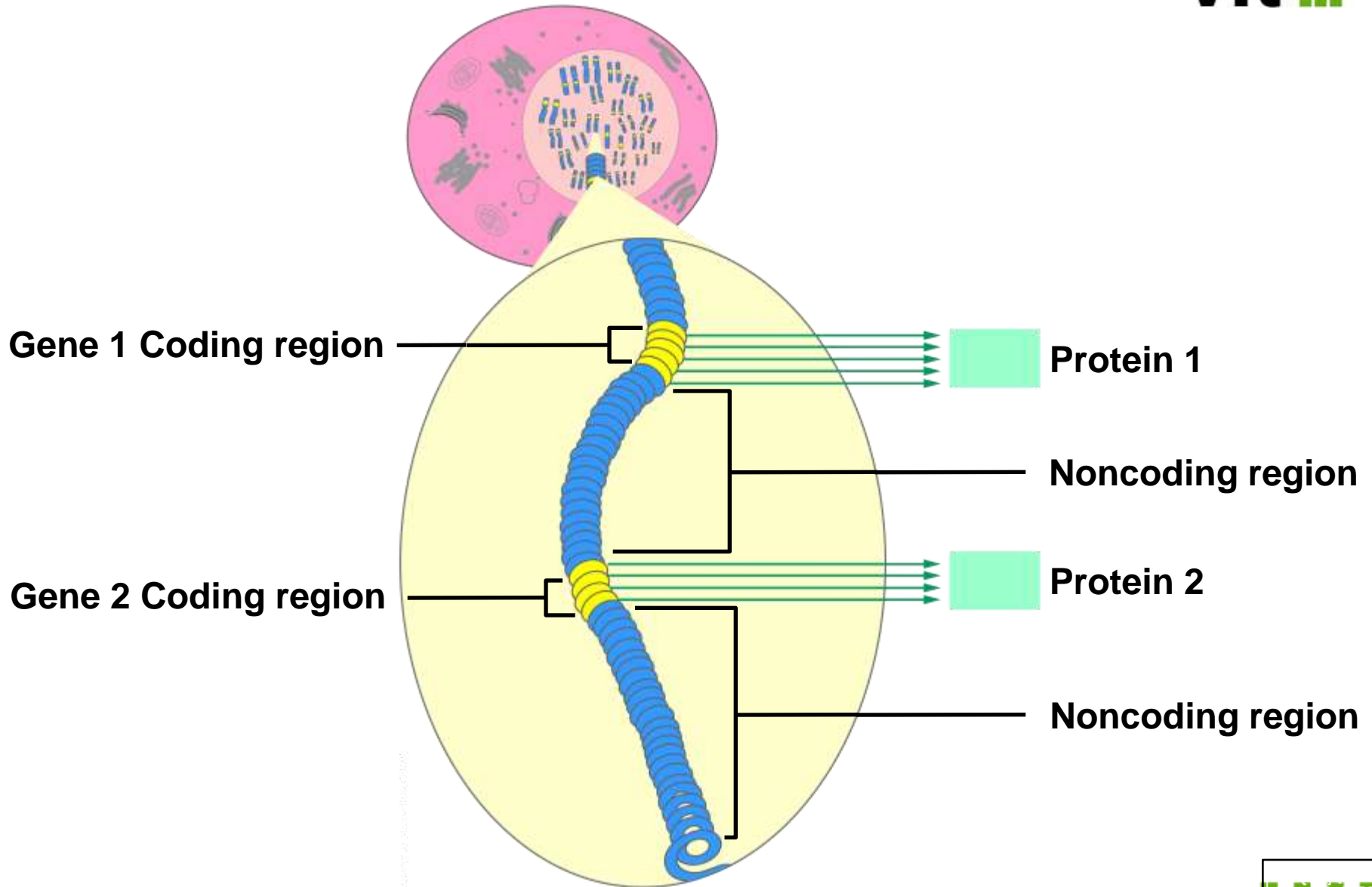
What Is the Human Genome?



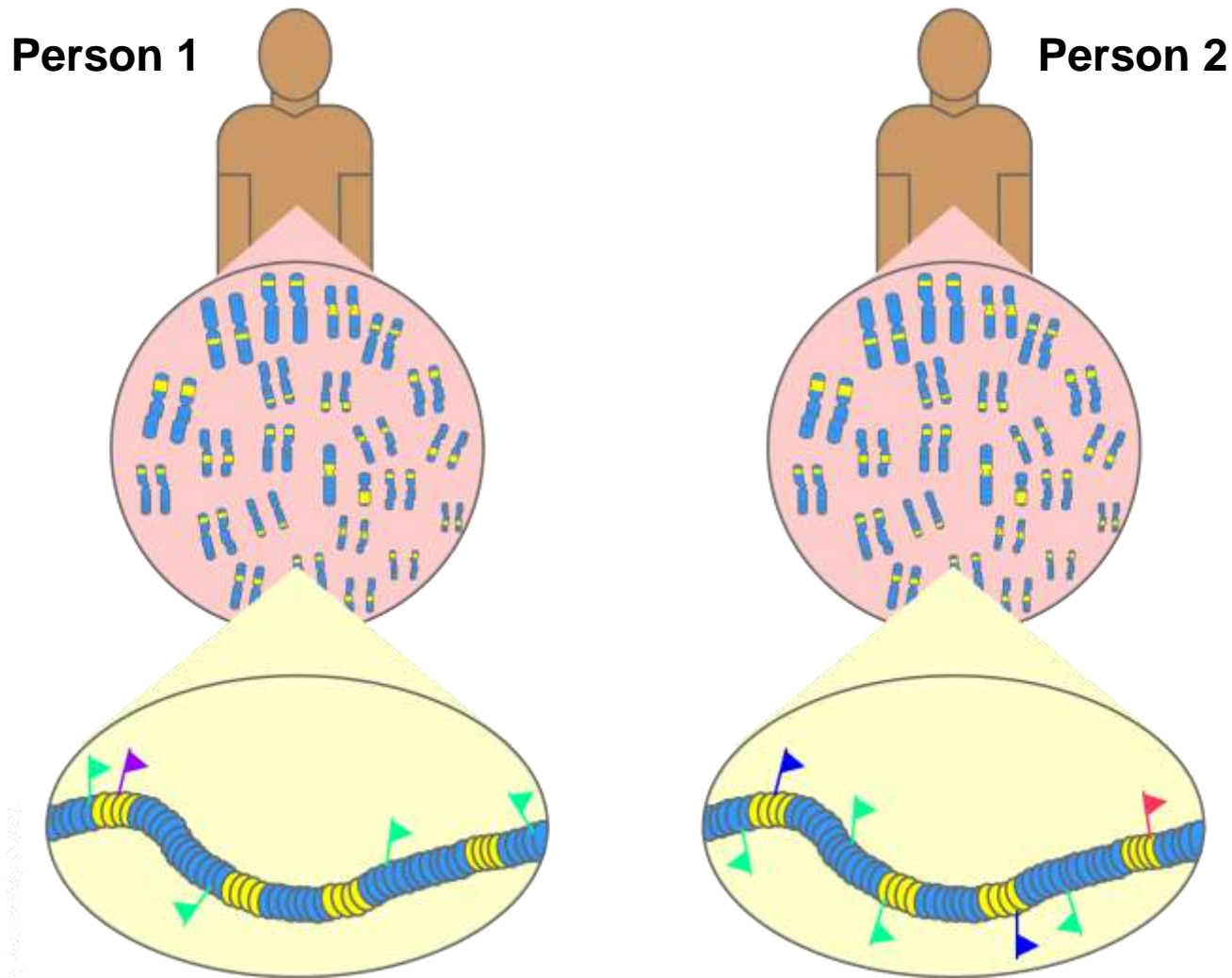
DNA and Chromosome Structure



The Genome Contains Genes



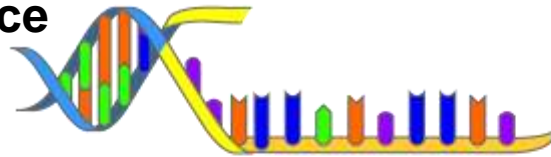
Variation in the Human Genome



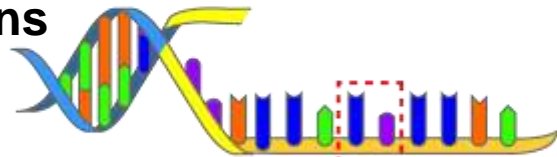
▶▶▶▶ = Variations in DNA

What Is Variation in the Genome?

Common Sequence



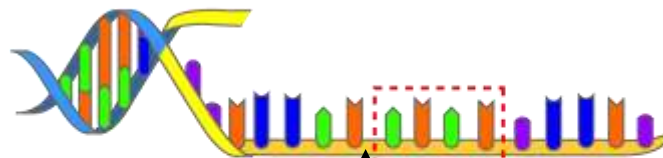
Variations



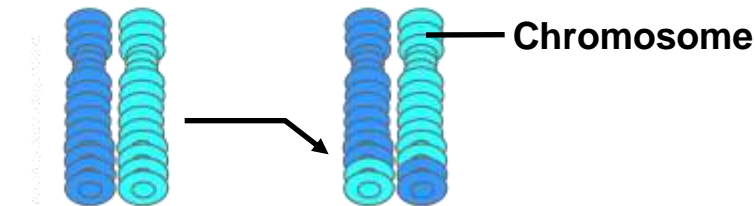
Polymorphism



Deletions

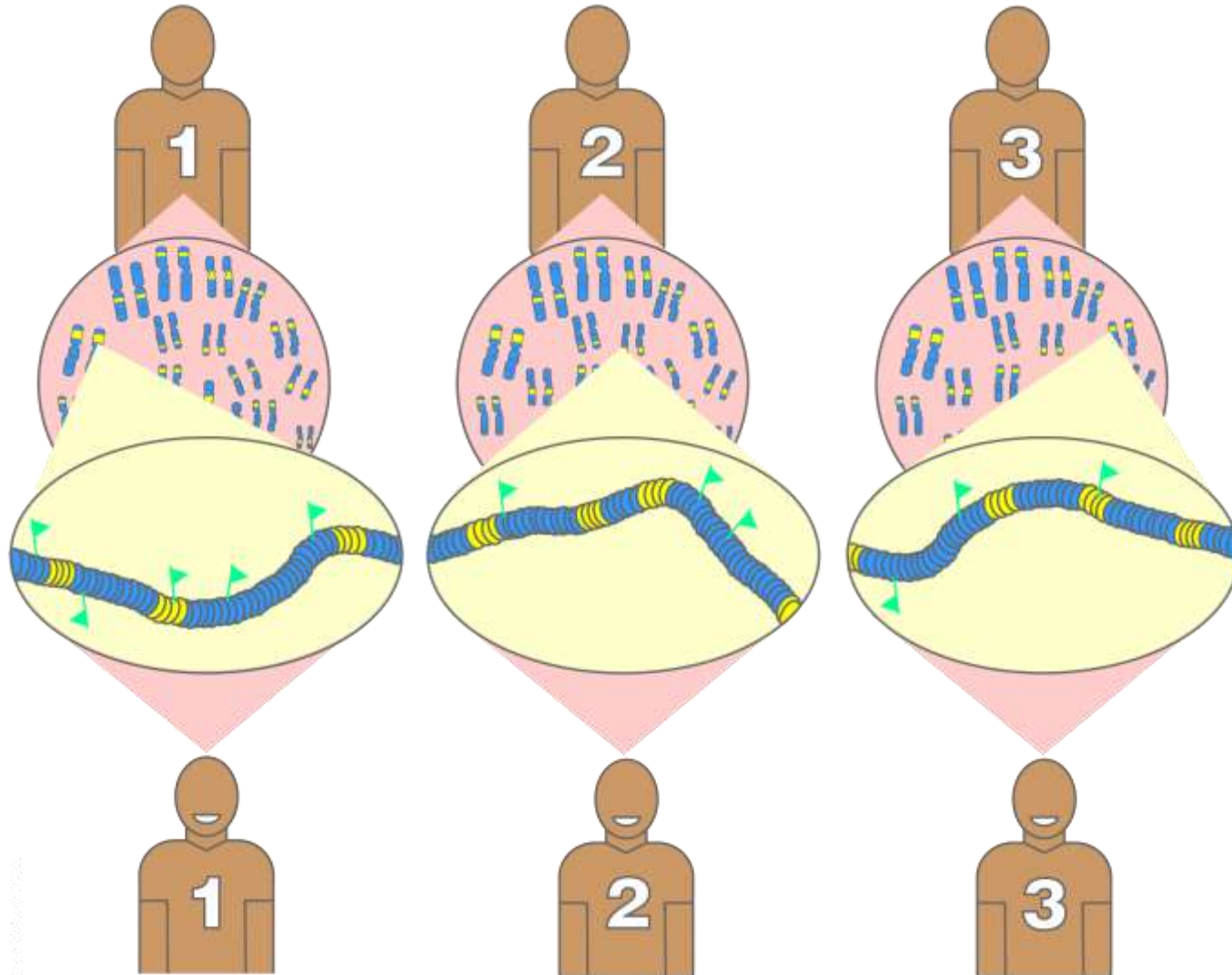


Insertions



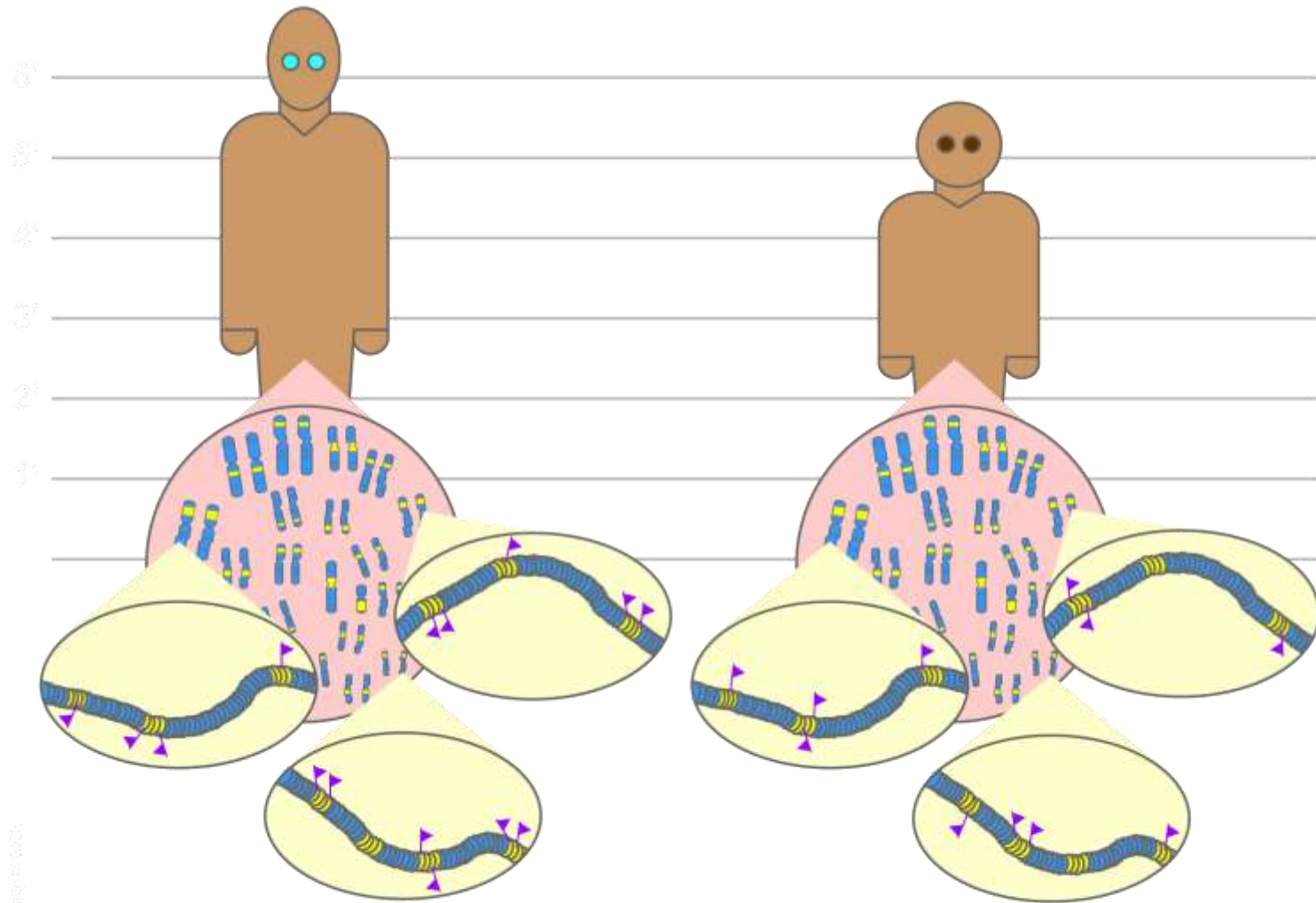
Translocations

Variations Causing No Changes



Variations in DNA that cause no changes

Variations Causing Harmless Changes

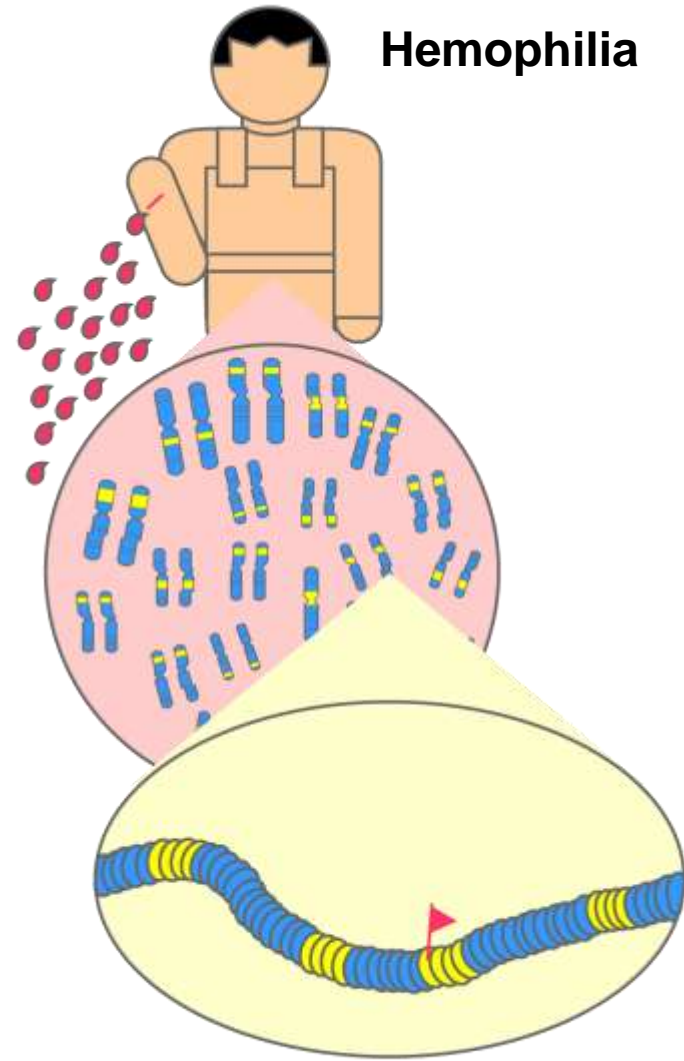
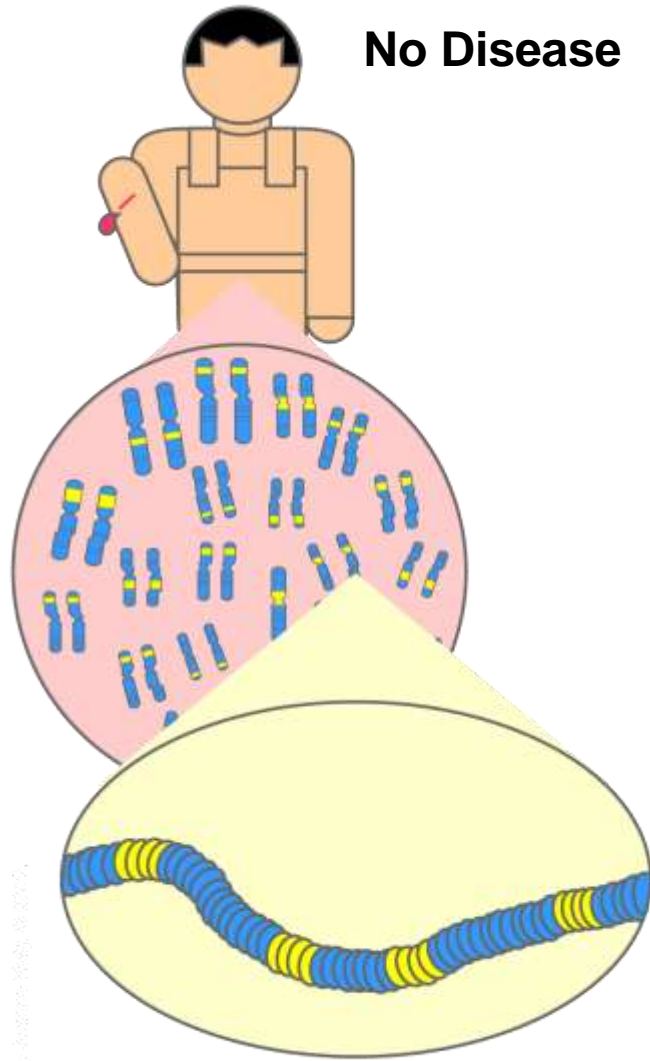


AP01000301, University of Cambridge, Page 16, 02/2008



Variations in DNA that cause harmless changes

Variations Causing Harmful Changes

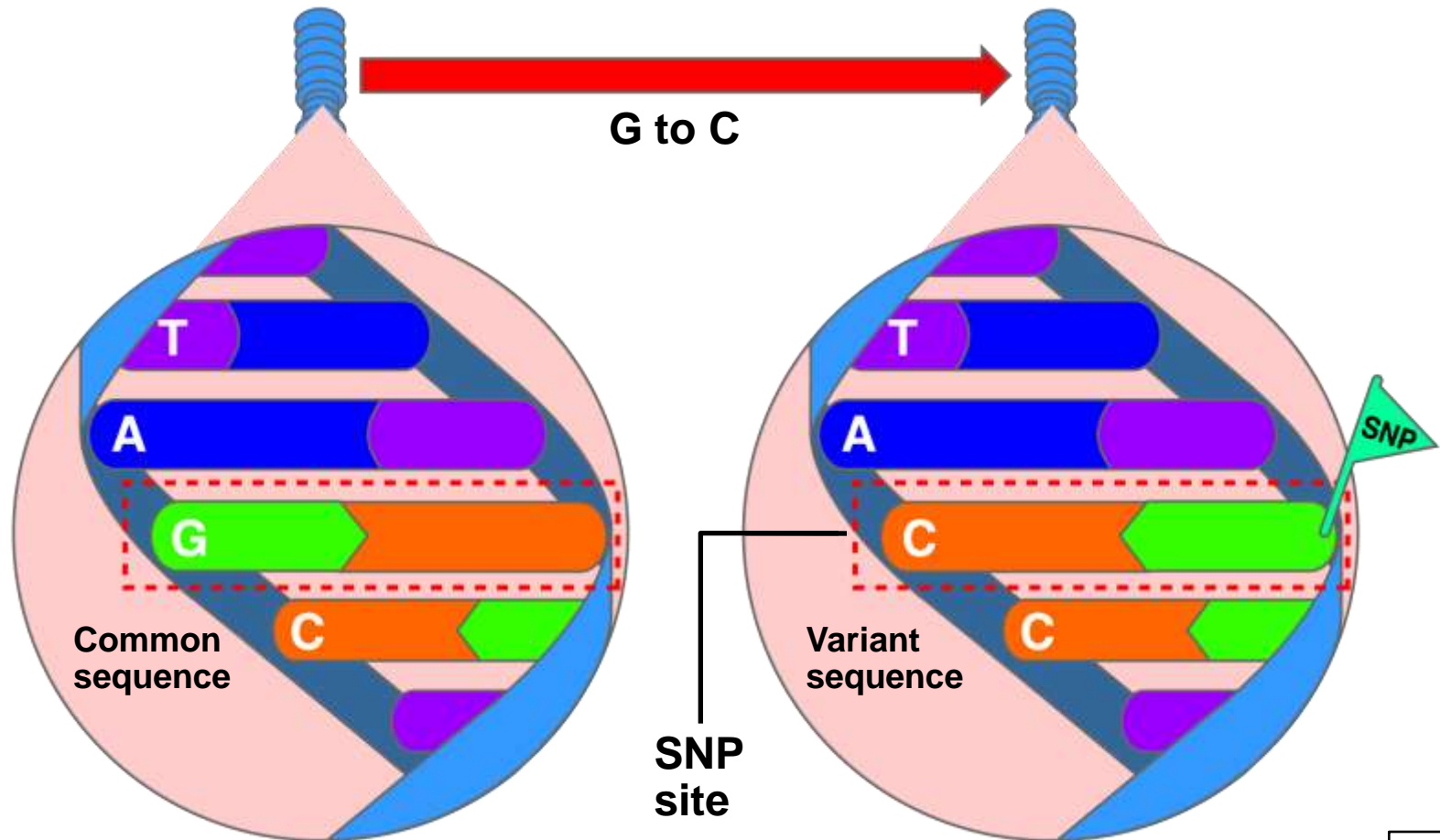


 = Variation in DNA that causes harmful change

SNPs Are the Most Common Type of Variation

Most of the population

At least 1 percent of the population



Approved by the Director, Faculty of Life Sciences