Analytical and statistical consideration on the use of the ICAR-SNP panel for parentage control, genotyped with the Illumina bead chip technology, exemplified on the German Holstein population.

Prof. Dr. Ekkehard Schütz, MD, PhD
Prof. Dr. Dr. Bertram Brenig, DVM, PhD

Center for Molecular Diagnostics
Institute of Veterinary Medicine

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

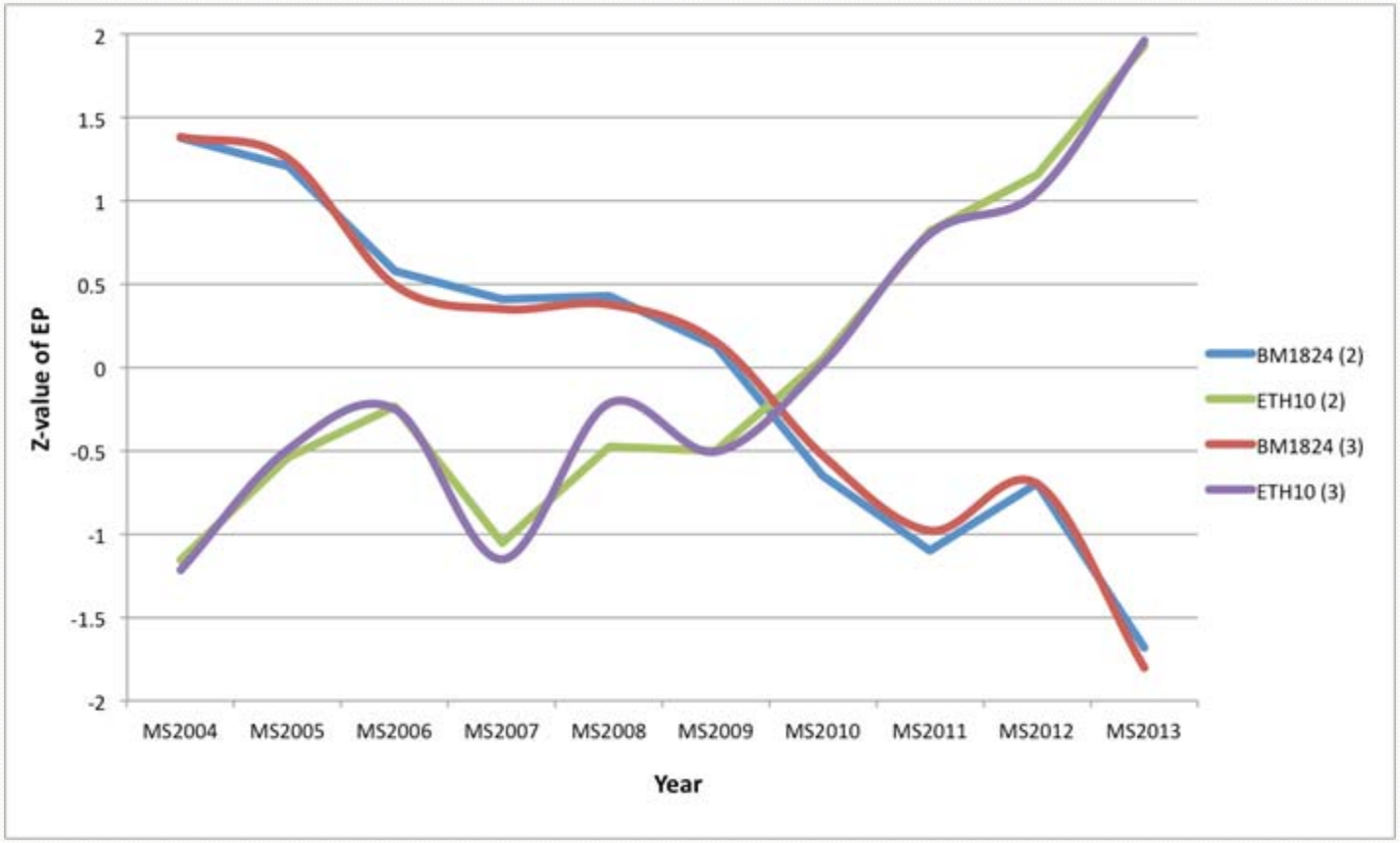
Germany

Background

- **Paternity control is moving from STR to SNP based systems**
  - Cost savings if SNP-genomic selection is done
    - Excl. power in specific lifestock populations needs to be defined
- **Genomic Selection based on thousands of SNP loci (LM)**
  - Tolerant for low confidence calls
- **Pat. Cntrl. based on hundreds (high effect of single values)**
  - Not that tolerant for GC calls
  - Ways of addressing this are desirable

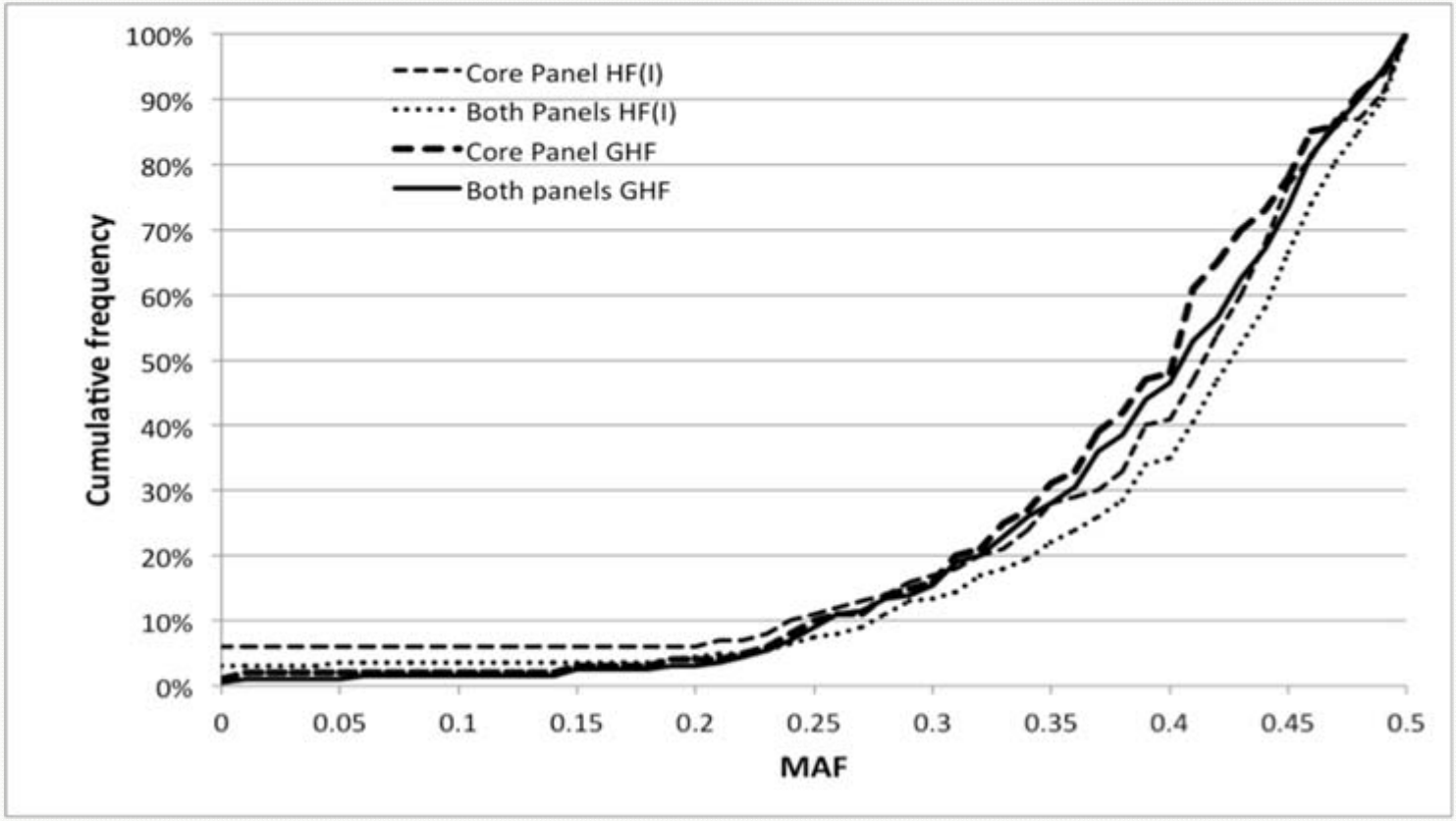
Exclusion Power

- **STRs have shifted over time in German HF**
  - Example for two STRs (~20,000/ year)

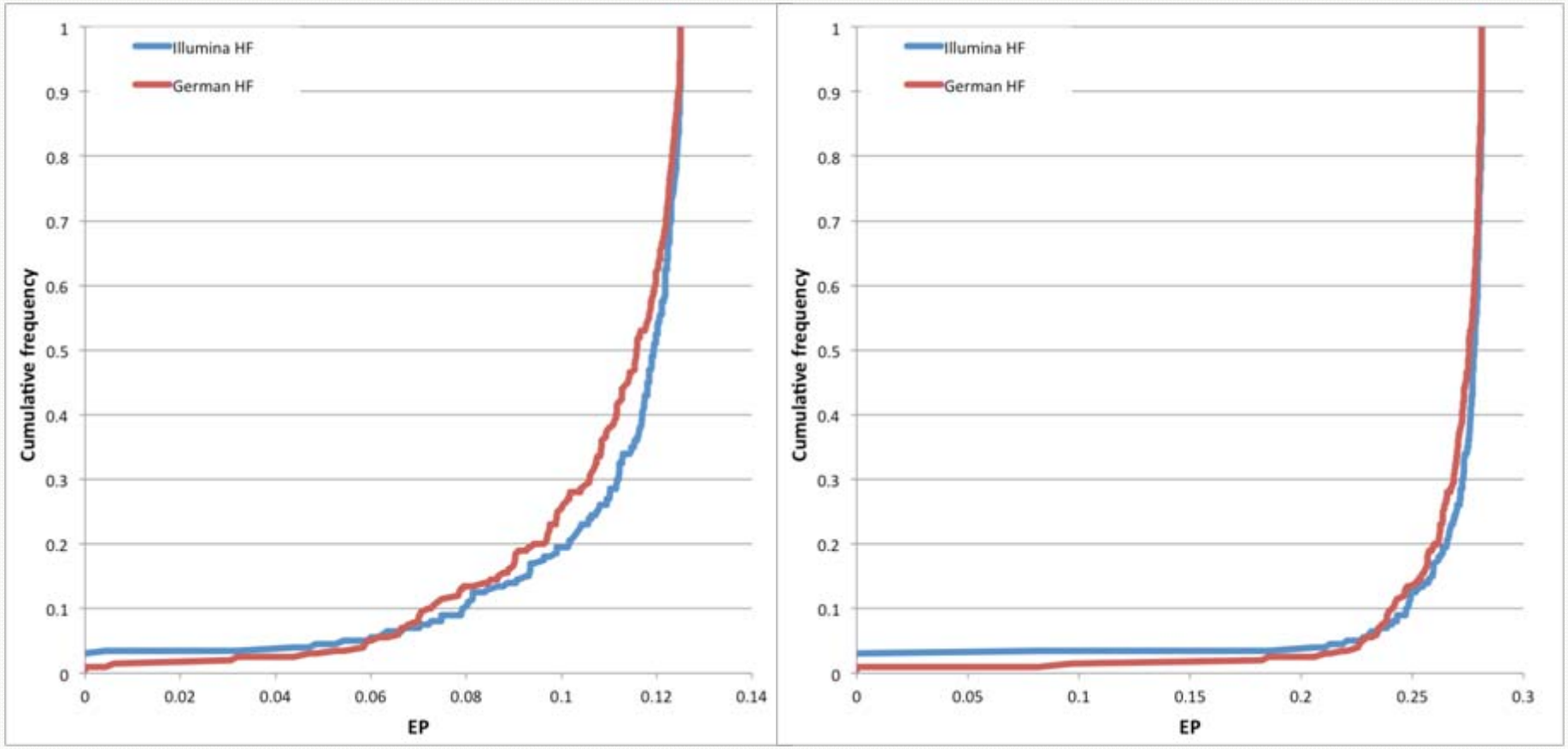


Minor Allelic Frequencies

- **SNP genotyping based on Illumina's Bead-Chip technology**
  - Bov50Kv2 & EuroG10k
- **EP for SNP based on MAF (0 -0.5)**
  - MAF calculated for 200 ISAG SNPs in 25,000 German HF

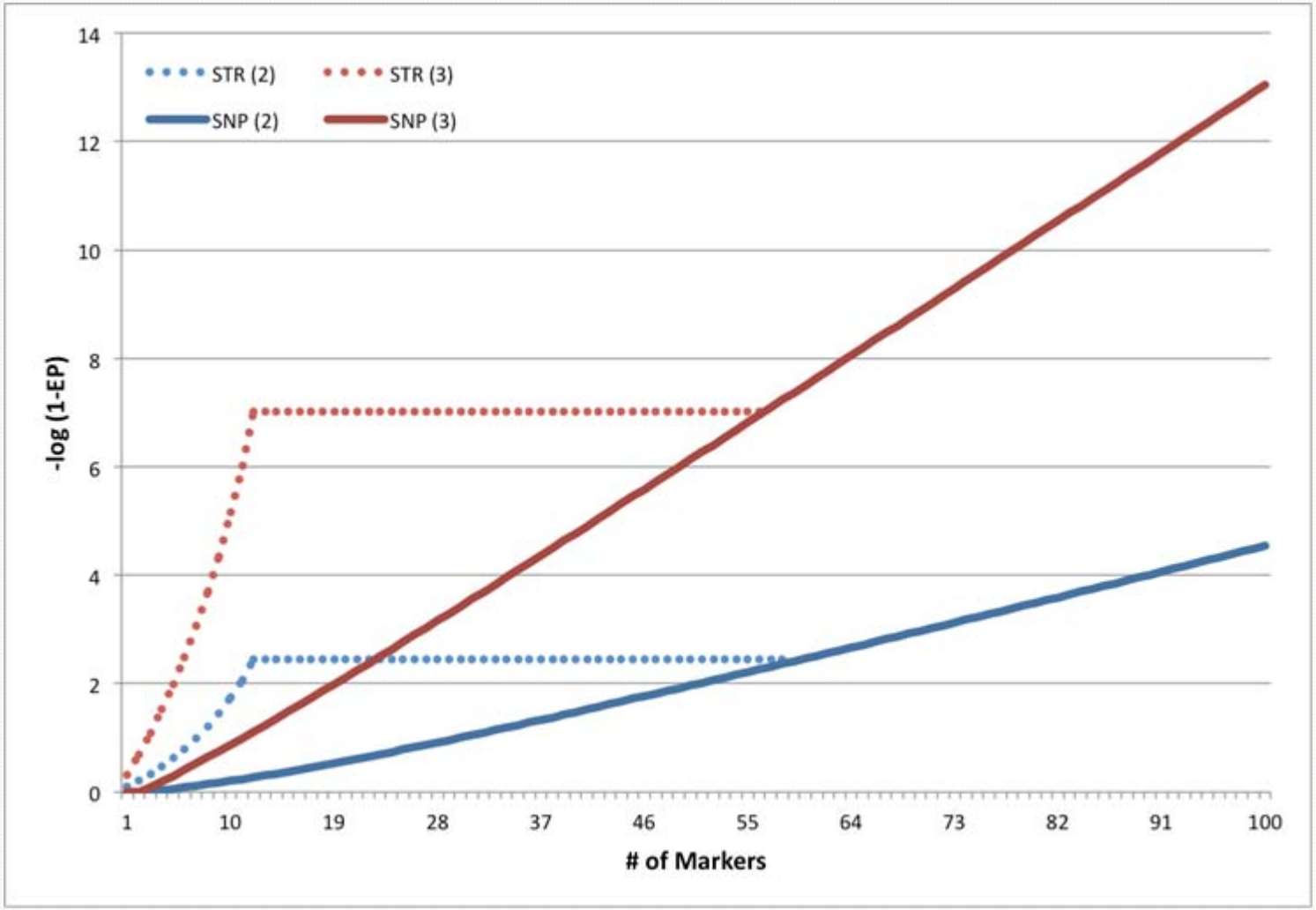


Exclusion Power



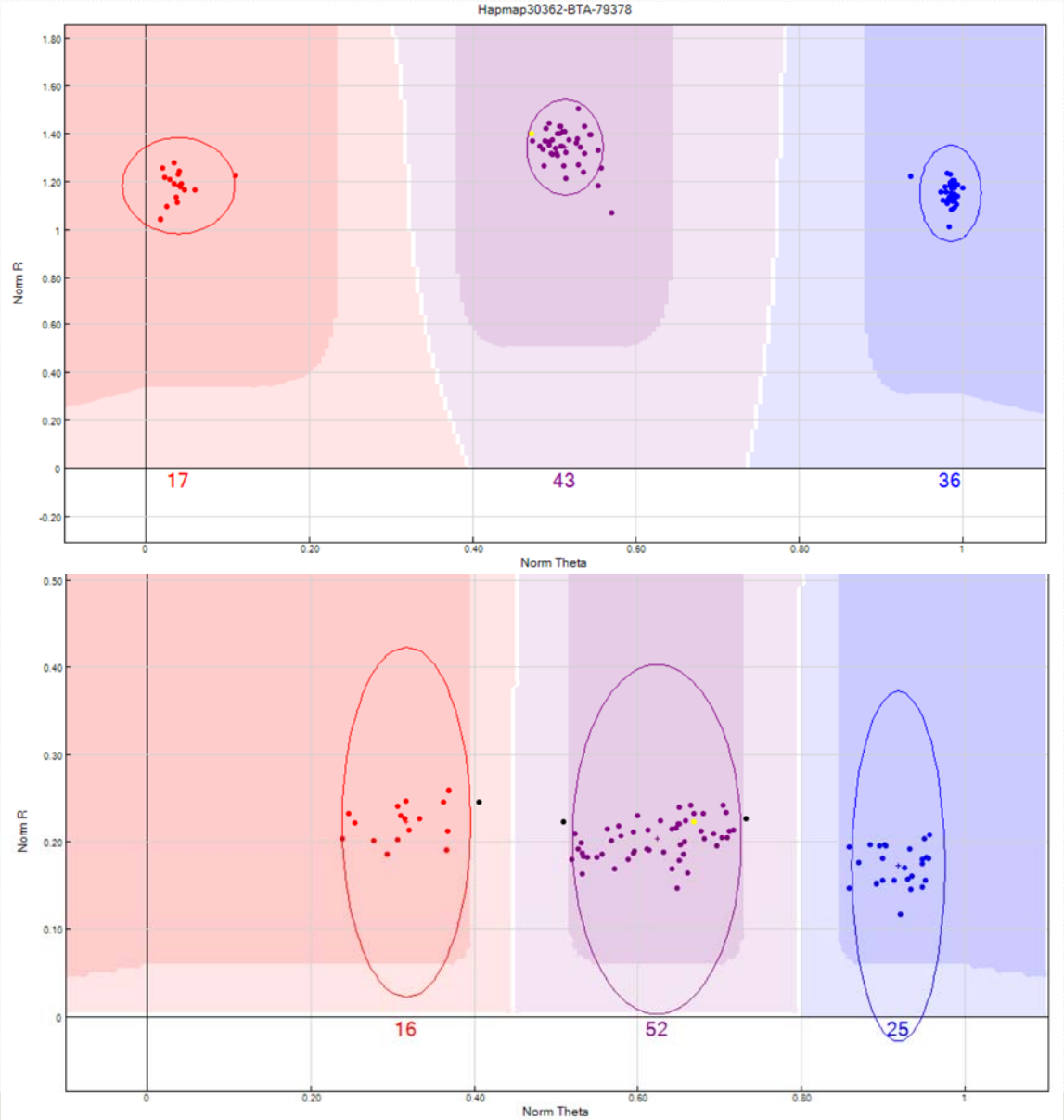
Exclusion Power (SNP vs. STR)

- EP of STRs and (100 core) SNPs compared



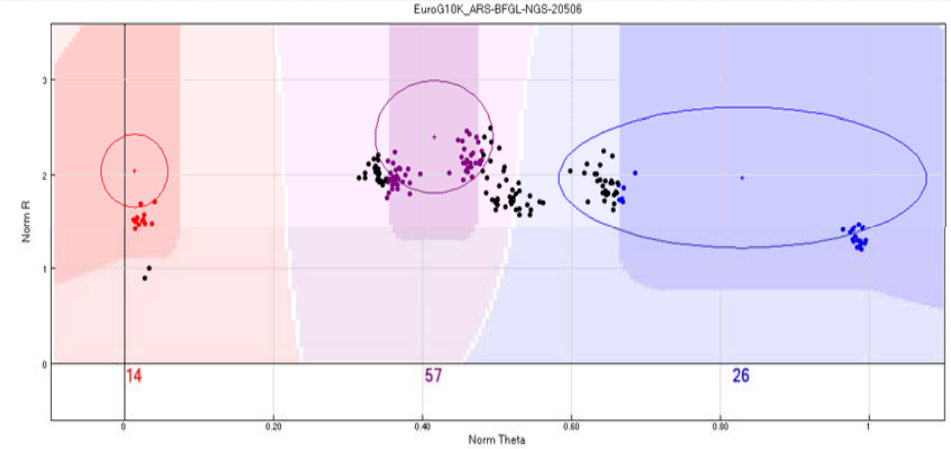
SNP genotyping

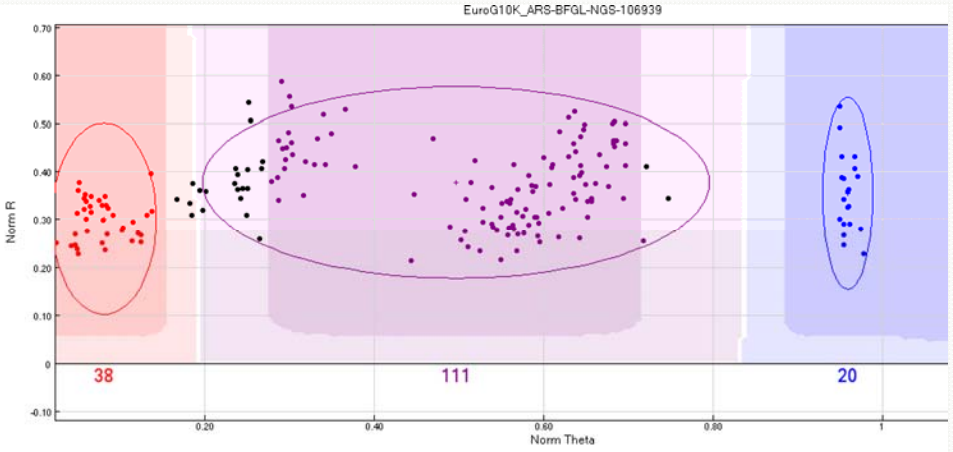
What are we talking about technically?

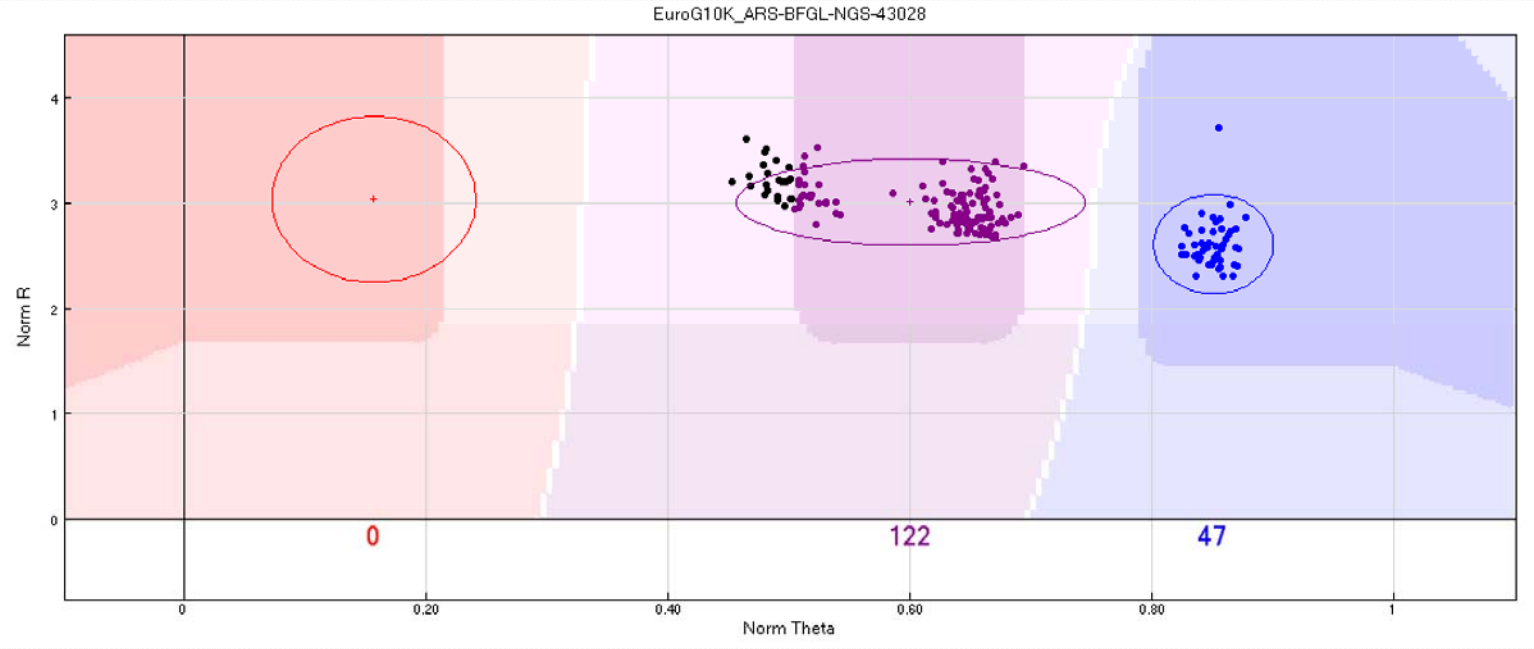


SNP genotyping

What are we talking about technically?







Bead Chip Technology Limitations

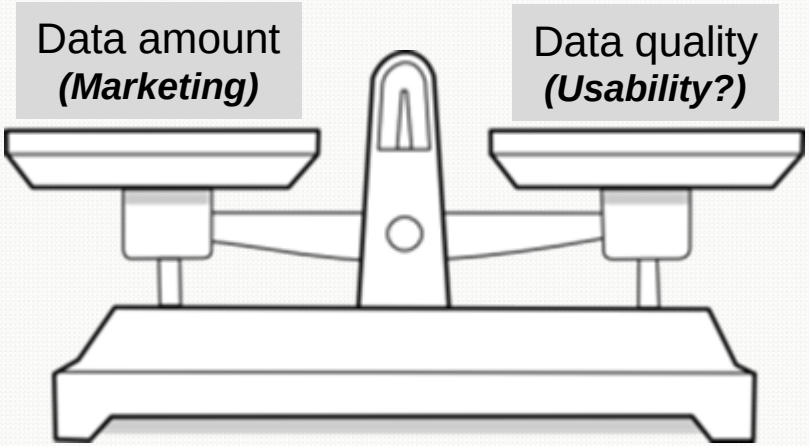
DNA was purified from these cells and was then genotyped at over a million loci using **Illumina's** Omni1-Quad **Beadchip**. Because our assay is particularly sensitive **to false positive homozygous SNP calls** for the recipient, we limited our focus to Beadchip SNPs with high GenCall and Cluster Separation scores, which yielded ~**150,000 usable loci** in each case.

To minimize false positive homozygous recipient SNP calls, only SNP loci with a **GenCall score >0.70** and a Cluster Separation score of 1.00 were considered….

From: Snyder TM, Kush KK, Valentine HA and Quake SR. PNAS 2011: 108; 6229-34

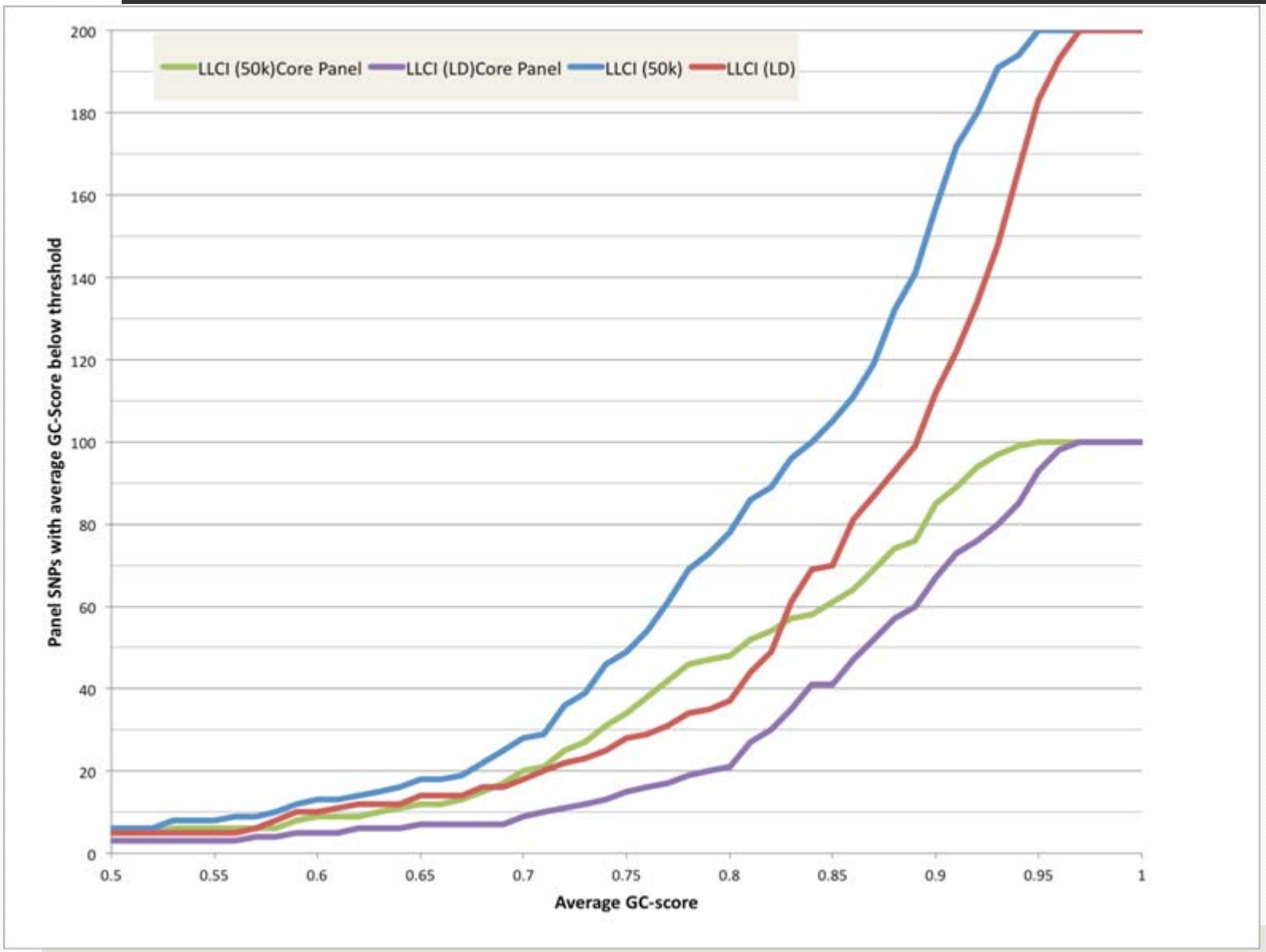
Always a Trade Off

…… during the standard clustering process. This is set by the GenCall threshold for clustering. Illumina recommends that you use a GenCall Score cut-off of **0.15** for Infinium arrays.



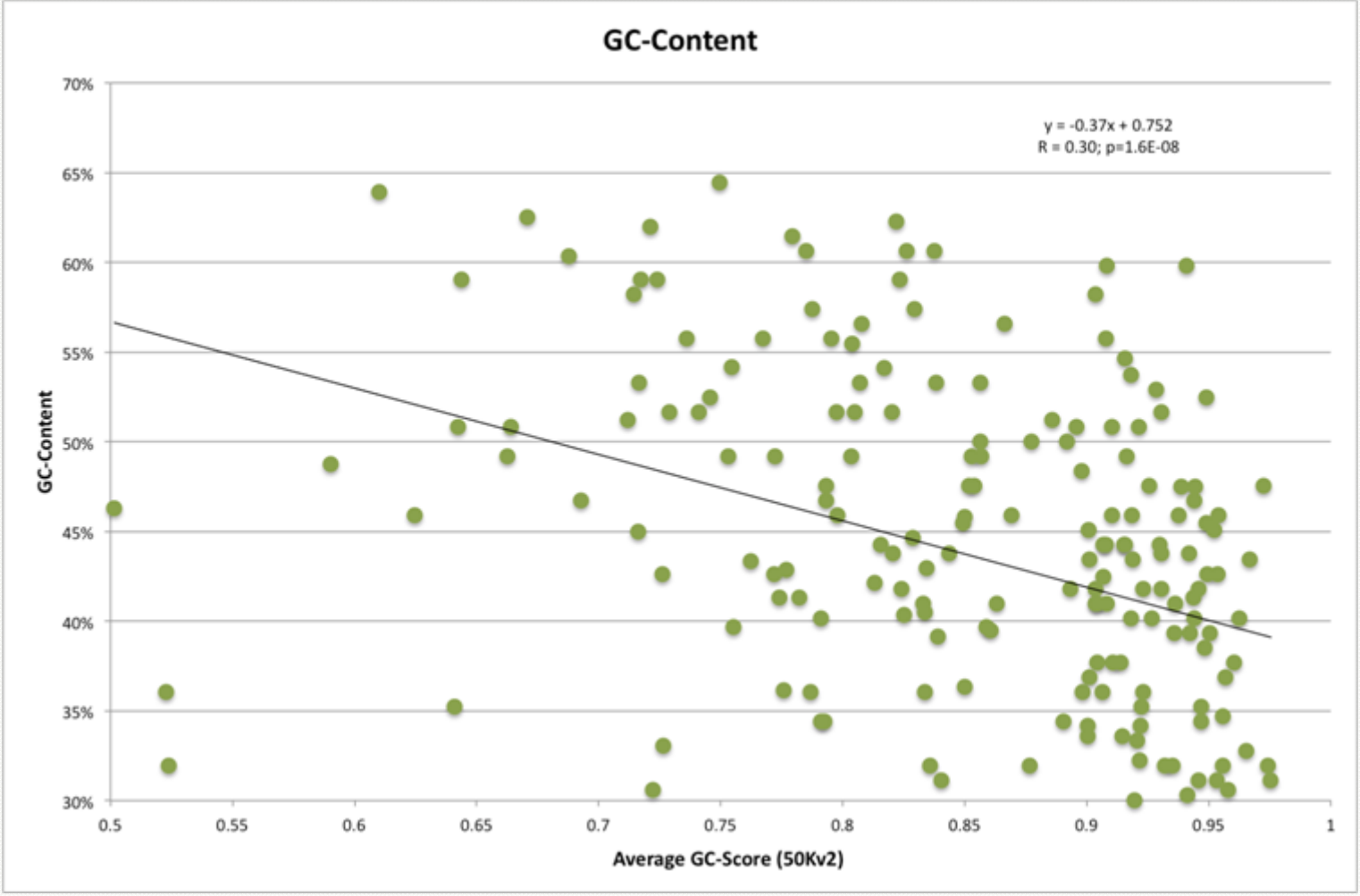
From: Illumina Technical Support; email in April 2014

Goodness of Gene Calls



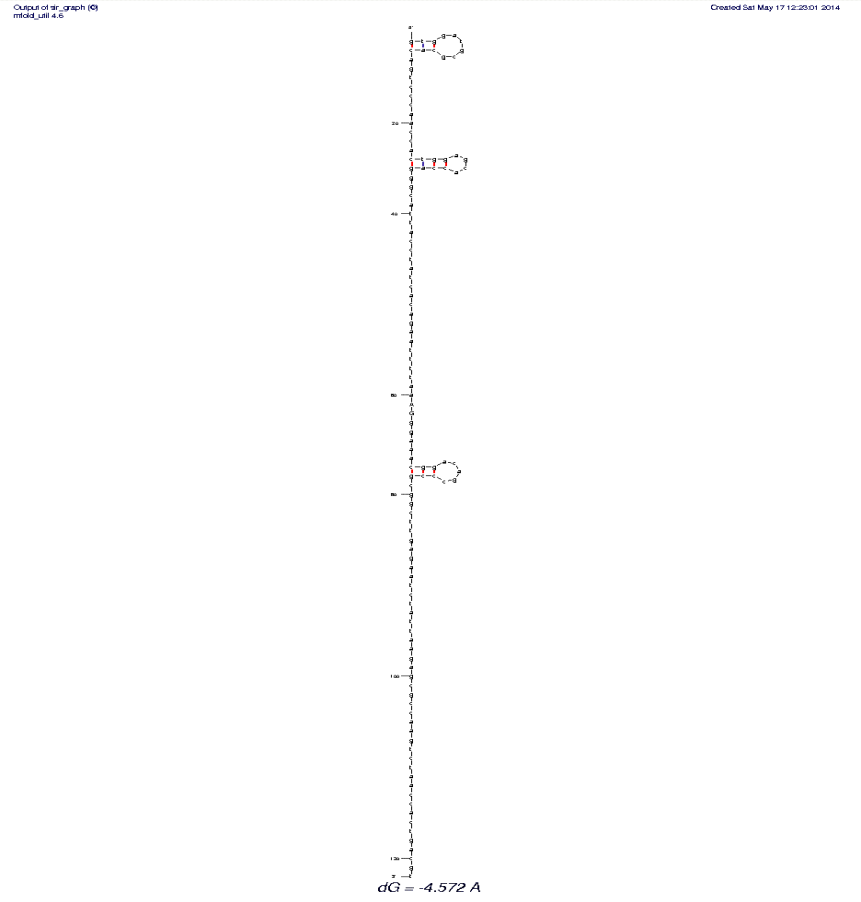
Bead Chip Technology Limitations

Anything we can do about it?

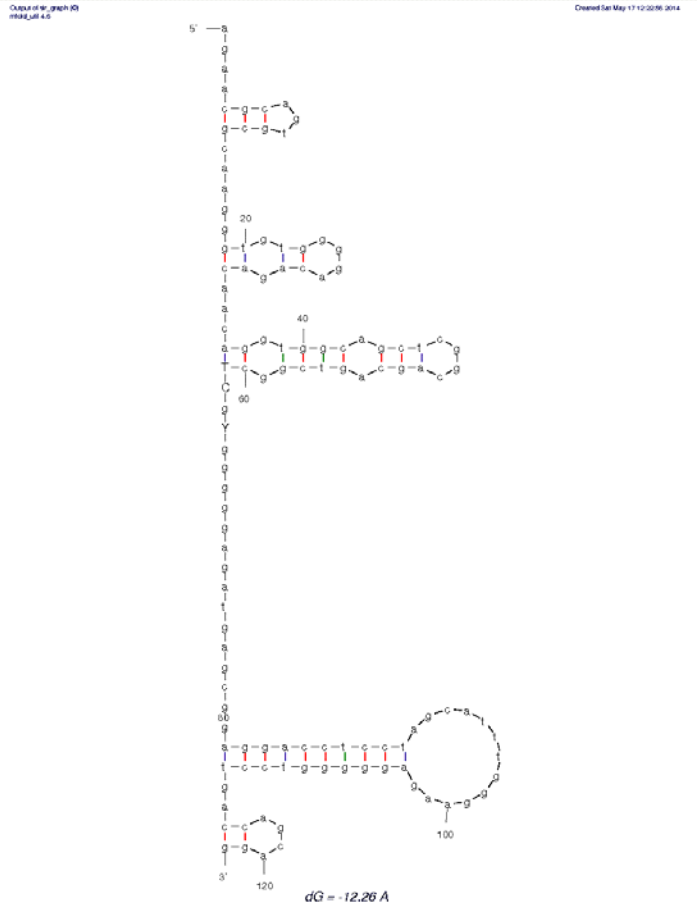


Limitations in Molecular Diagnostic

High GC-Score average



Low GC-Score average



Limitations in Molecular Diagnostic

Very Low GC-Score average (<.6)

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Bos taurus breed Hereford chromosome 12, alternate assembly Btau_4.6.1	99.6	303	100%	1e-19	100%	NC_007310.5
Bos taurus breed Hereford chromosome 16, alternate assembly Btau_4.6.1	32.2	399	98%	19	100%	NC_007314.4
Bos taurus breed Hereford chromosome X, alternate assembly Btau_4.6.1	30.2	227	98%	74	95%	NC_007331.4
Bos taurus breed Hereford chromosome 13, alternate assembly Btau_4.6.1	30.2	264	92%	74	100%	NC_007311.5
Bos taurus breed Hereford chromosome 5, alternate assembly Btau_4.6.1	32.2	443	86%	19	100%	NC_007303.5
Bos taurus breed Hereford chromosome 9, alternate assembly Btau_4.6.1	32.2	403	86%	19	100%	NC_007307.5
Bos taurus breed Hereford chromosome 6, alternate assembly Btau_4.6.1	30.2	516	86%	74	100%	NC_007304.5
Bos taurus breed Hereford chromosome 20, alternate assembly Btau_4.6.1	30.2	284	86%	74	100%	NC_007318.5
Bos taurus breed Hereford chromosome 14, alternate assembly Btau_4.6.1	32.2	467	84%	19	95%	NC_007312.5
Bos taurus breed Hereford chromosome 8, alternate assembly Btau_4.6.1	30.2	371	84%	74	91%	NC_007306.5
Bos taurus breed Hereford chromosome 10, alternate assembly Btau_4.6.1	36.2	437	80%	1.2	100%	NC_007308.5
Bos taurus breed Hereford chromosome 19, alternate assembly Btau_4.6.1	32.2	290	80%	19	100%	NC_007317.5
Bos taurus breed Hereford chromosome 2, alternate assembly Btau_4.6.1	30.2	603	80%	74	100%	NC_007300.5
Bos taurus breed Hereford chromosome 22, alternate assembly Btau_4.6.1	30.2	256	78%	74	100%	NC_007320.5
Bos taurus breed Hereford chromosome 27, alternate assembly Btau_4.6.1	30.2	227	78%	74	100%	NC_007328.4
Bos taurus breed Hereford chromosome 1, alternate assembly Btau_4.6.1	32.2	605	76%	19	100%	NC_007299.5
Bos taurus breed Hereford chromosome 15, alternate assembly Btau_4.6.1	32.2	262	74%	19	100%	NC_007313.5
Bos taurus breed Hereford chromosome 17, alternate assembly Btau_4.6.1	32.2	362	74%	19	95%	NC_007315.5
Bos taurus breed Hereford chromosome 11, alternate assembly Btau_4.6.1	34.2	661	72%	4.7	100%	NC_007309.5

ISAG Recommendations (2012)

GUIDELINES FOR PARENTAGE VERIFICATION BASED ON SNP MARKERS are,
SNP profile: Minimum number of SNPs in panel: 100; Minimum number of SNPs available in profile: 95
(If less than 95 SNPs can be scored, retest the sample or request a new sample).

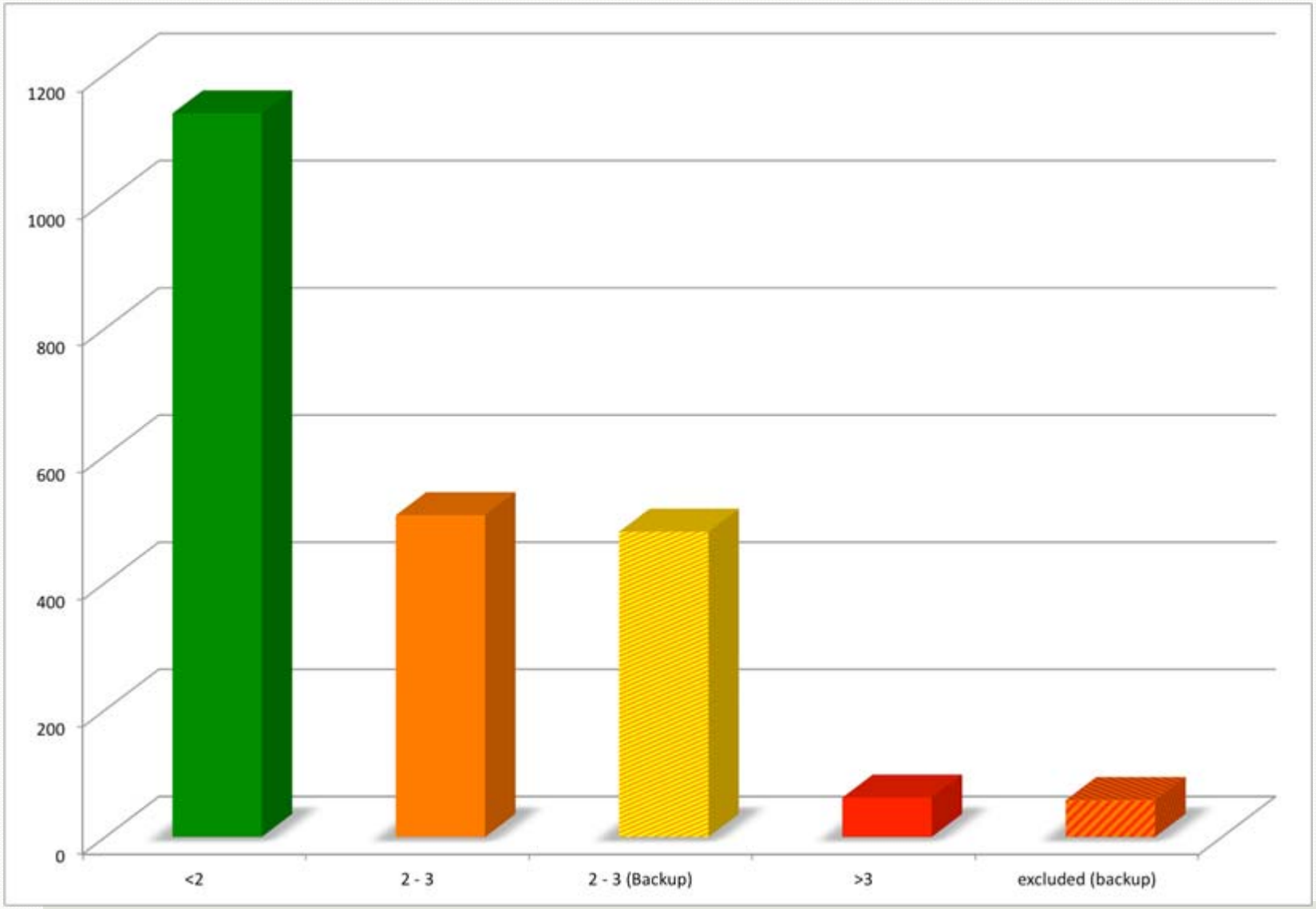
If mismatches occur in a supposed parentage, the general rule first is to retest the samples involved or request new samples to confirm the determined genotypes. If the genotypes are confirmed the following guidelines are suggested.

	one parent	**both parents**	Certify:
Minimum number of SNPs	90	85	
Number of mismatches*:	**0 - 1**	**0 - 2**	parentage accepted
Number of mismatches*:	**2 - 3**	**3 - 4**	parentage doubtful, use additional panel
Number of mismatches*:	**≥ 4**	**≥ 5**	parentage excluded
*Example:	offspring = GG, sire = AA	offspring = AG, sire = AA, dam = AA	

**: When the parentage is doubtful, first genotype the samples with both panels (ISAG panel and the backup) and then, if results remain doubtful, ask customer for other possible parents, if there are no other then qualify the parents. In any case, it is recommended that samples be retested if there is a parentage exclusion with ISAG panel and/or back up panel.

SNP Parentage

- the effect of errors-



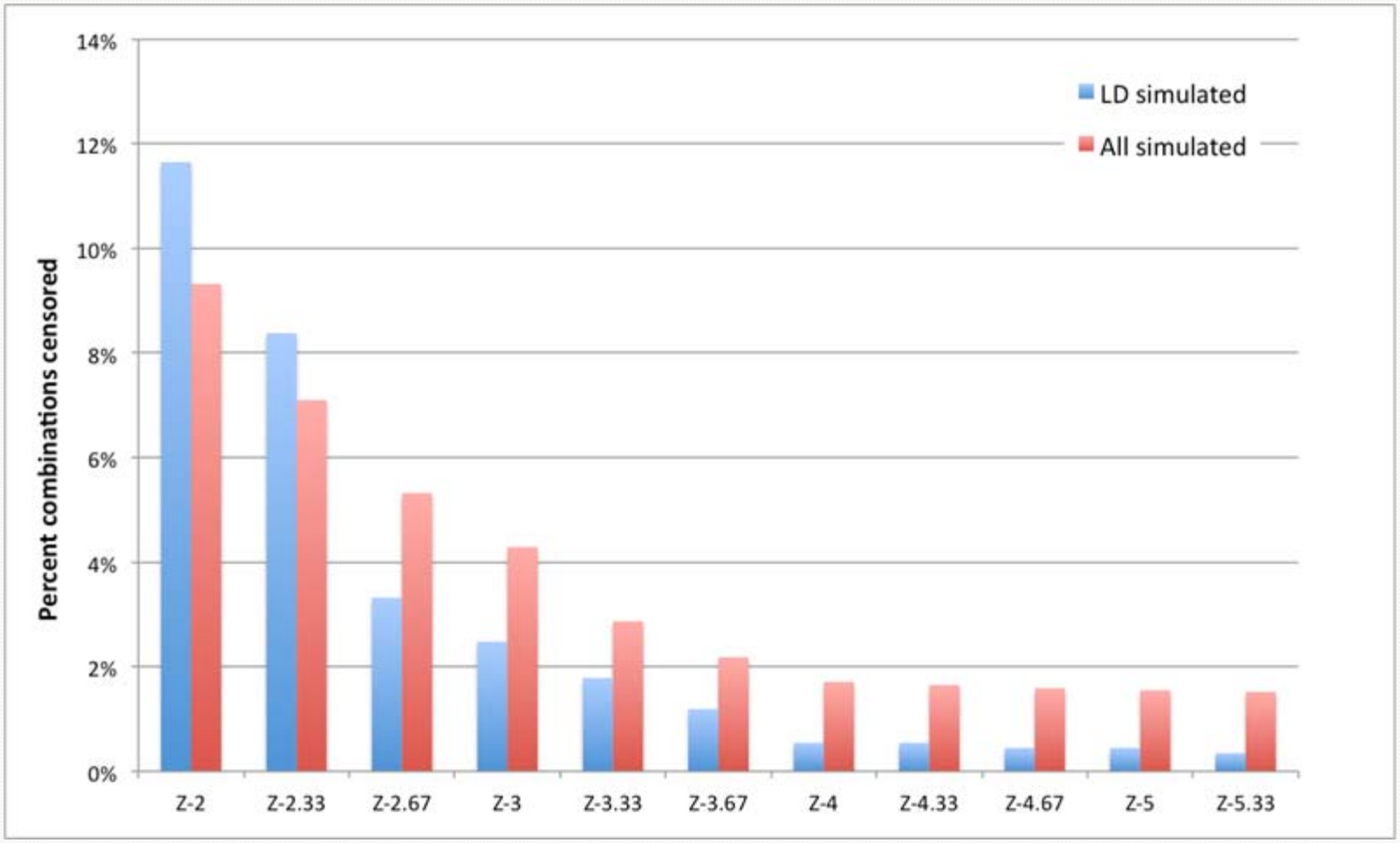
SNP Genotyping based Parentage

- process error control-

SNP Parentage
- process error control-

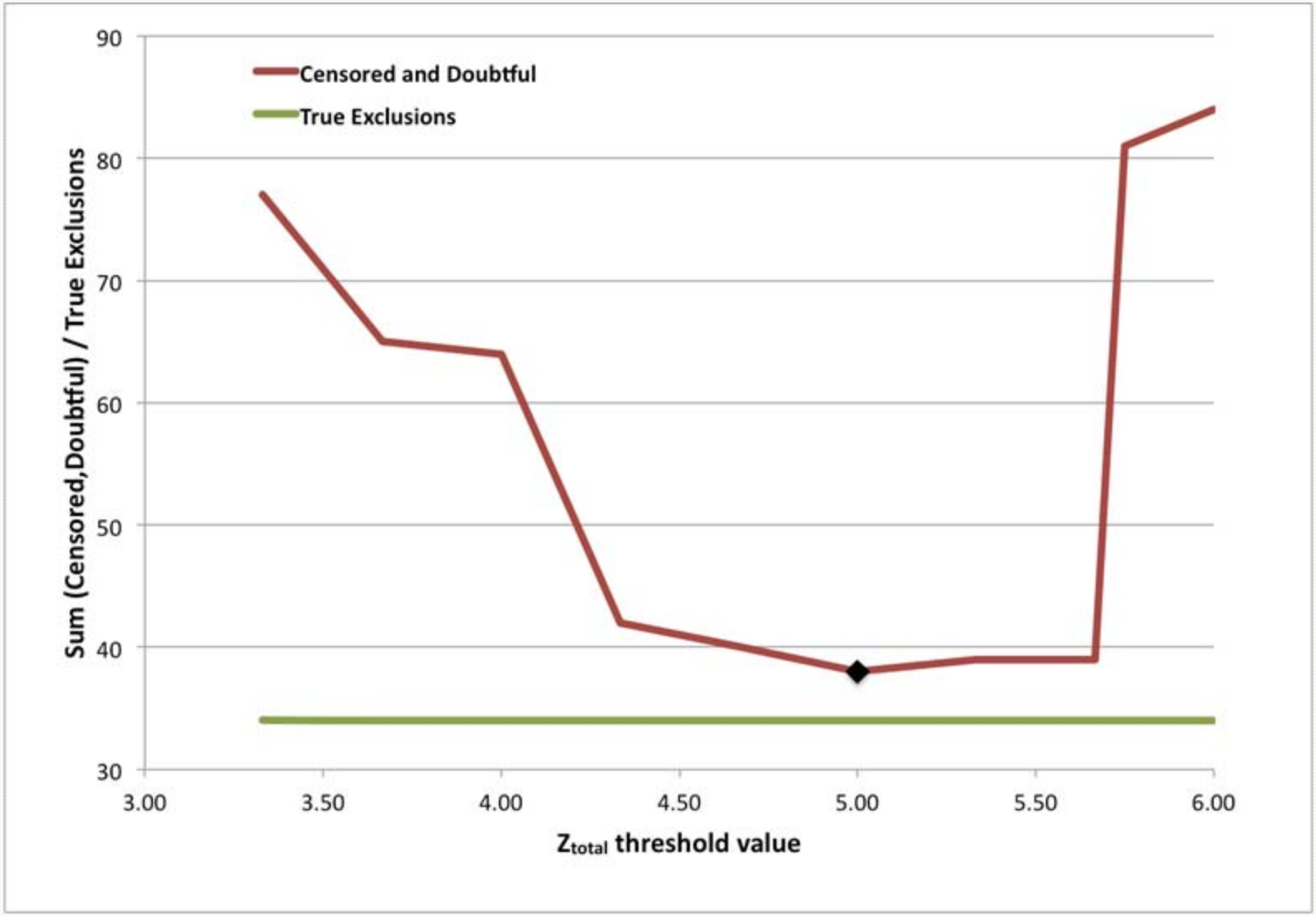
- Using each individual GC-Score as censor:
  - Calculating average and standard deviation for 200 panel SNPs
    - 20,000 bov50kV2
    - 7,000 EuroG10k
  - Defining the error for a given individual SNP call
    - Cross-platform (50k; EuroG10k) / cross-version
    - Transform to Z-Values (per SNP): $\frac{GCs_i - \overline{GCs}}{\sigma(GCs)}$
  - Defining the error for a given parentage (per SNP)
  - Additive EP: $Z_{total} = \sqrt{\sum Z^2}$;Z>1 set to 0
  - Test for effect in simulations
  - Test for performance in randomly chosen field data set

Simulated loss by censoring



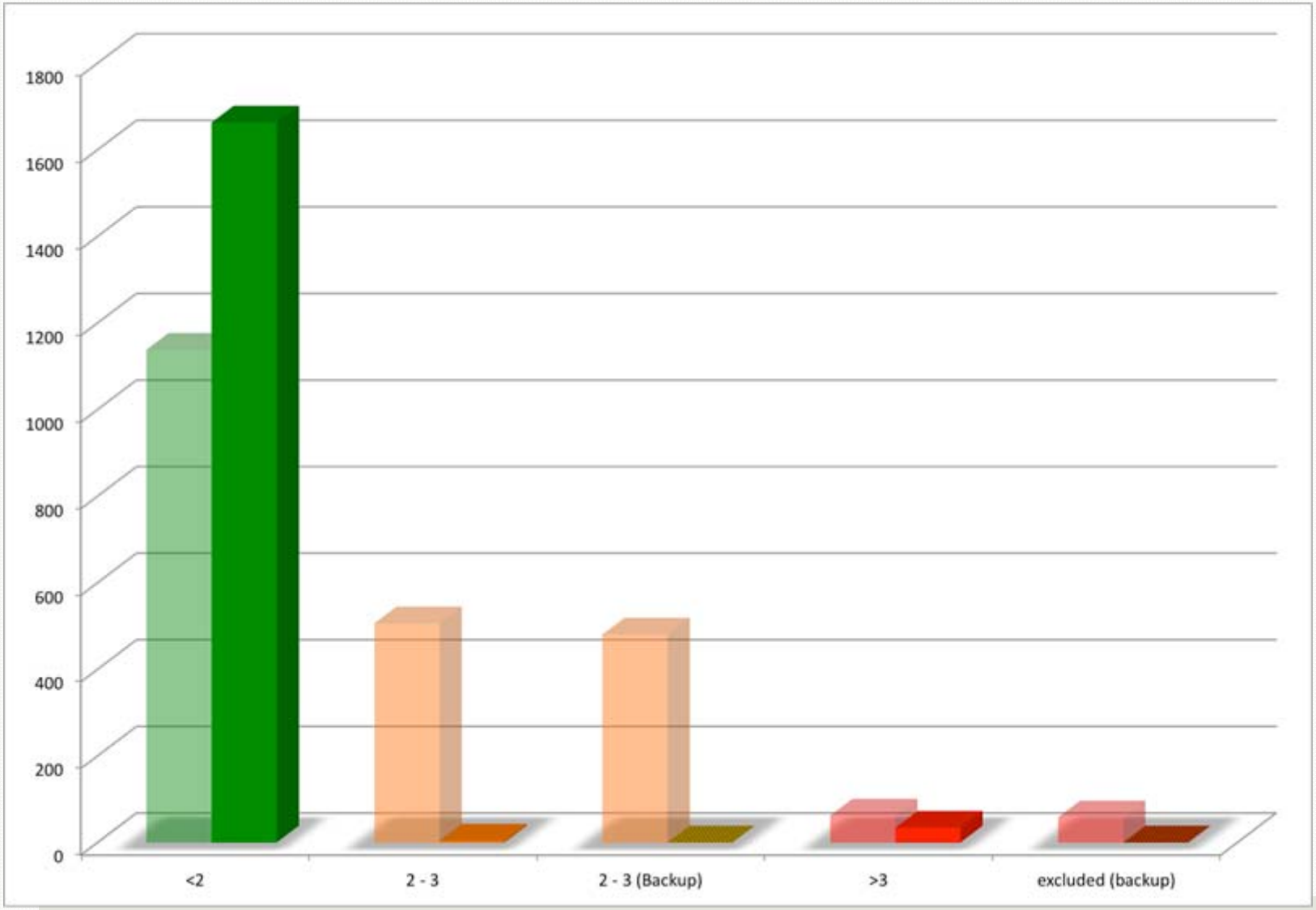
SNP Parentage

- process error control optimization -



SNP Parentage

- process error control final -



Summary

- SNP-based PC has high EP for one and both parents
  - 2.5 and 6 logs better than STRs
- Widely used Illumina Bead-Chips error prone (as any system)
  - e.g. false homozygous calls
- A process error control is proposed that:
  - Minimizes doubtful results (>90%)
  - Avoids false parentage exclusions (100%)
  - Harboring only low censoring frequency (< 3%)
    - *A priori* vs *a posteriori* application?

Discussion points

- Include a goodness of GC measure in DNA certificates?
- Rethinking the 200 SNP Panel
  - Avoid high GC% loci
  - Avoid loci within or adjacent to Repetitive Elements
- Rethinking the Recommendations
  - Do we need 90 / 85 SNPs for Pat. Cntrl. given the EP?

Acknowledgements

- Team of IVM:
  - Melanie Scharfenstein, Susann Loos, Louisa Jüttner
- German Holstein Association (DHV)
  - RUW
  - OHG
  - ZBH
  - VOST
  - WEU
  - RBW
  - SRB