

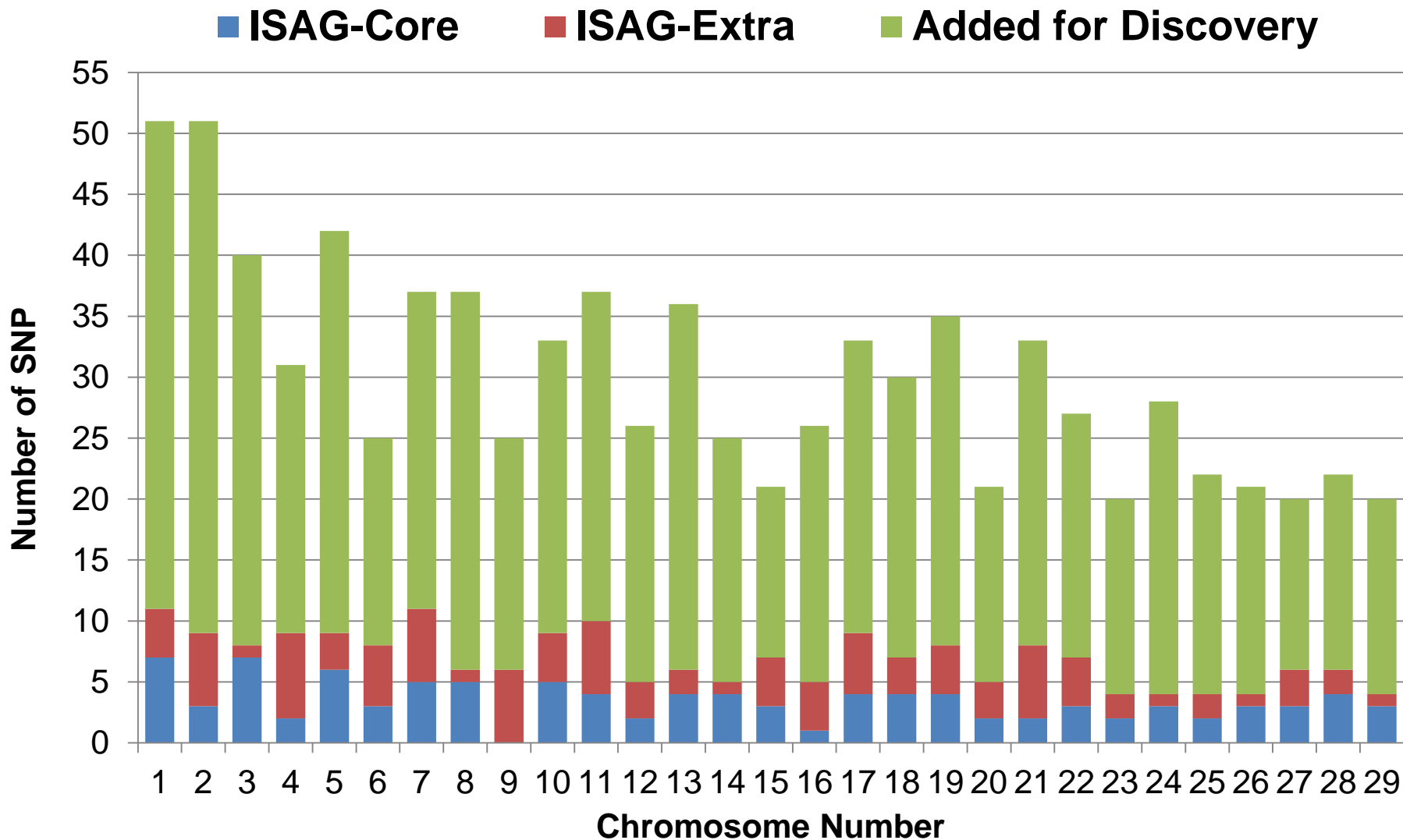
Parentage Analysis Services for Dairy Cattle in Canada

Brian Van Doormaal, Gerrit Kistemaker & Jarmila Johnston
Canadian Dairy Network (CDN)

- **Parentage Verification**
 - 200 SNP recommended by ISAG
 - 100 Core SNP plus 100 Extra SNP
- **Parentage Discovery**
 - 200 ISAG SNP
 - Plus 675 other SNP (total 875) based on research in USA, Ireland and Australia
- **Microsatellite Imputation**
 - Additional 980 SNP deemed most useful for imputation to microsatellite SNP profiles to allow for parentage verification on that basis
 - Overall total of 1,855 SNP for possible exchange

Parentage Analysis

- PSE 875 SNP by Chromosome -



Current Situation in Canada

- **Parentage analysis, both verification and discovery, has been a routine part of genomic evaluation processing at CDN since 2009**
- **CDN has over 1.4M genotypes**
 - 1.2M HO, 163K JE, 29K BS, 6K AY and 3K GU
- **23 different genotype panels**
 - Includes low (3K-30K), medium (44K-140K) and high density (over 600K) genotypes
- **For parentage analysis, a list of 2,683 SNP are used**
 - In common from the 3K and 50K genotype panels
 - 50K panel is basis for national genomic evaluations



Objectives

- 1. Quantify the minor allele frequency (MAF) of each SNP within each of the five dairy breed populations with genomic evaluations in Canada**
- 2. Compare official parentage analysis results at CDN with those that would have resulted based on the GenoEx-PSE SNP**
- 3. Assess the accuracy of imputation to 50K genotypes from GenoEx-PSE SNP as well as the resulting genomic evaluations**

Assessing Minor Allele Frequency

- **MAF is the frequency of the least common allele for each SNP**
 - SNP with low MAF will have more limited value for parentage analysis since only homozygous genotypes in both parent and progeny are informative

Average Minor Allele Frequency (MAF)		
Breed	ISAG 200 SNP for Verification	Added SNP for Discovery
Holstein	0.391	0.381
Ayrshire	0.342	0.343
Jersey	0.291	0.316
Brown Swiss	0.352	0.353
Guernsey	0.322	0.322

ISAG Guidelines for SNP Parentage Verification

- **Minimum requirement is that 95 of the 100 Core SNP are called/available in the genotype**
- **Progeny with 1 parent genotyped:**
 - At least 90 common SNP available in progeny
 - 0-1 mismatches => Parentage accepted
 - 2-3 mismatches => Parentage doubtful (check 200)
 - >3 mismatched => Parentage excluded
- **Progeny with both parents genotyped:**
 - At least 85 common SNP available in progeny
 - 0-2 mismatches => Parentage accepted
 - 3-4 mismatches => Parentage doubtful (check 200)
 - >4 mismatched => Parentage excluded



Comparison Analysis - Parent Verification -

- **Modified CDN process to use GenoEx-PSE SNP**
- **CDN criterion: >2% SNP conflict = Exclude parent**
- **Compared results using CDN 2600, ISAG 200 & PSE 875**

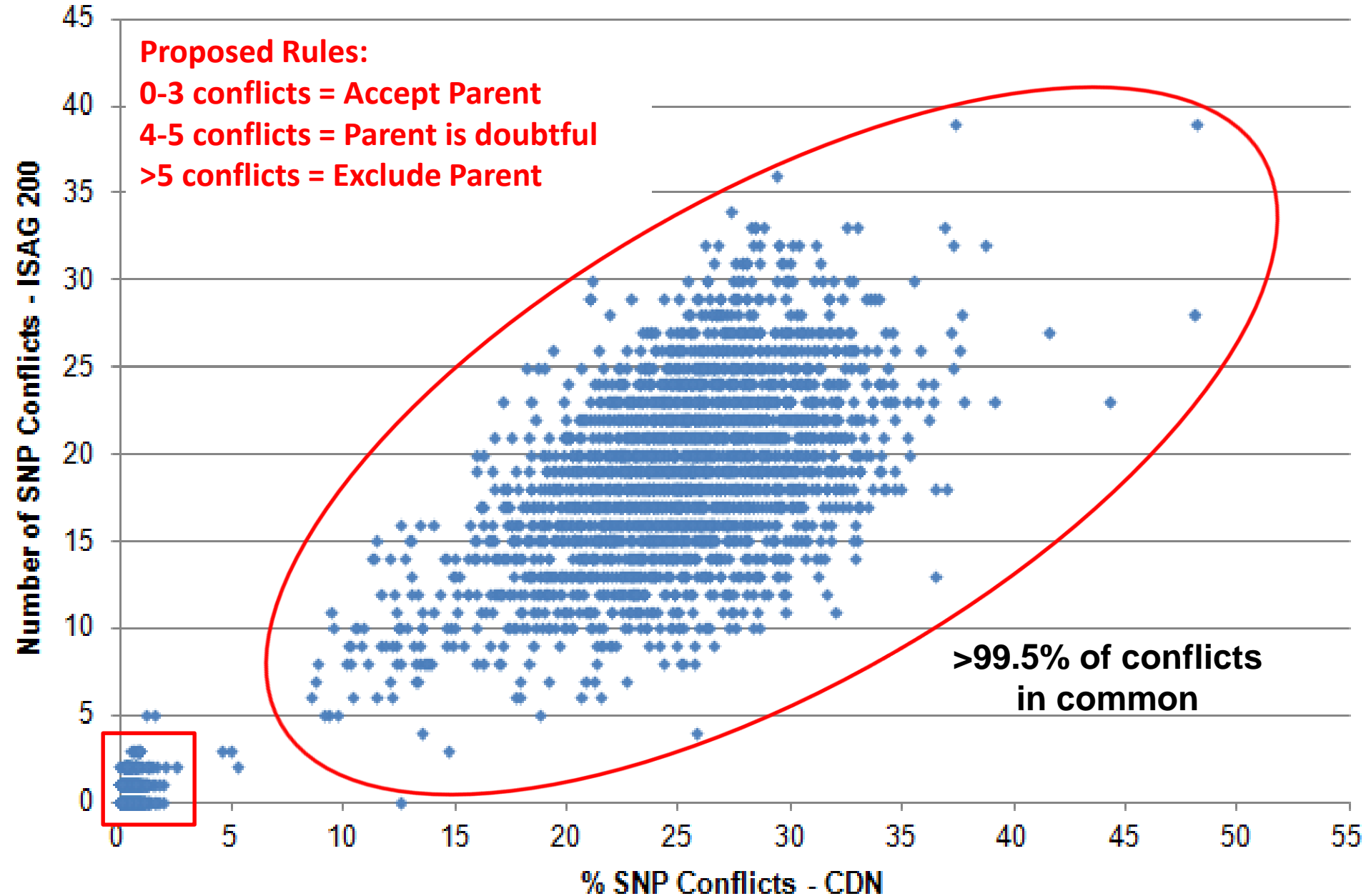
Breed	Total No. Animals	No SNP Conflicts with Any Method	Conflict with at least 1 Method
HO	573,988	97.4%	15,160
JE	63,248	95.0%	3,148
AY	5,372	96.2%	203
BS	23,144	90.9%	2,111
GU	1,771	93.3%	119
All Breeds	667,523	96.9%	20,741

Comparison of SNP Conflicts

- CDN Procedure Versus Using 200 ISAG SNP -

Proposed Rules:

- 0-3 conflicts = Accept Parent
- 4-5 conflicts = Parent is doubtful
- >5 conflicts = Exclude Parent



Comparison of Results - Holstein, ISAG 200 -

ISAG 200 Conflicts	Proposed Status	CDN 2600 Result		
		Conflict	Verified	Total
0	Accepted 15 of 574,000 animals with a different result	1	9,067	9,068
1			3,306	3,306
2		3	126	129
3		3	11	14
4	In Doubt	2		2
5		4	2	6
6	Excluded	8		8
7		9		9
8		22		22
9		27		27
10		52		52
>10		2,517		2,517
Total:		2,648	12,512	15,160

Comparison of Results - Holstein, PSE 875 -

PSE 875 Conflicts	Proposed Status	CDN 2600 Result			
		Conflict	Verified	Total	
0		1	274	275	
1			11,153	11,153	
2			892	892	
3	5 of 574,000 animals with a different result		137	137	
4			41	41	
5		Accept		8	8
6			2	4	6
7					
8		2	1	3	
9			2	2	
10					
>10	Exclude	2,643		2,643	
Total:		2,648	12,512	15,160	

Comparison Analysis - Parentage Discovery -

- **Selected animals born in recent birth years**
 - Removed all known parentage
 - Total of 26,691 animals across five breeds
- **Ran only the discovery of sire for this analysis**
 - “Discovered” = the male with the fewest SNP conflicts and <1% conflict among informative SNP
- **99.865% of discovered sires were the same using CDN 2600 and PSE 875 SNP**
 - For 166, both methods picked an identical twin to the pedigree sire (must be handled by data pre-processing)
 - A few results differed only due to CDN criterion of <1% conflict
- **For 1 animal, CDN 2600 found the pedigree sire and PSE 875 picked a full brother**
 - Really the only important error with PSE 875 (.0037%)



Accuracy of GEBV from PSE 875 Imputation to 50K

- **Reduced genotypes for Holsteins born in 2016 with an official genomic evaluation (mainly females) to include only PSE 875 SNP**
- **Modified current CDN processes to not use SNP from X chromosome since excluded from PSE 875 (both ISAG and discovery)**
- **Used the most updated version of FImpute for imputation to 50K**
- **Examined imputation error**
- **Calculated DGV and GEBV for various traits**
- **Compared results to CDN official evaluations**

Imputation Accuracy

- 875 PSE SNP to 50K Genotype -

Imputation Error	No. Animals	Percentage
< 5%	17,990	65.8%
5 - 9.99%	1,864	6.8%
10 - 14.99%	2,153	7.9%
15 - 19.99%	2,798	10.2%
20 - 24.99%	1,533	5.6%
25 - 29.99%	560	2.0%
30 - 34.99%	332	1.2%
35 - 39.99%	84	0.3%
40 - 44.99%	7	0.0%
45 - 49.99%	3	0.0%
Total	27,324	100.0%

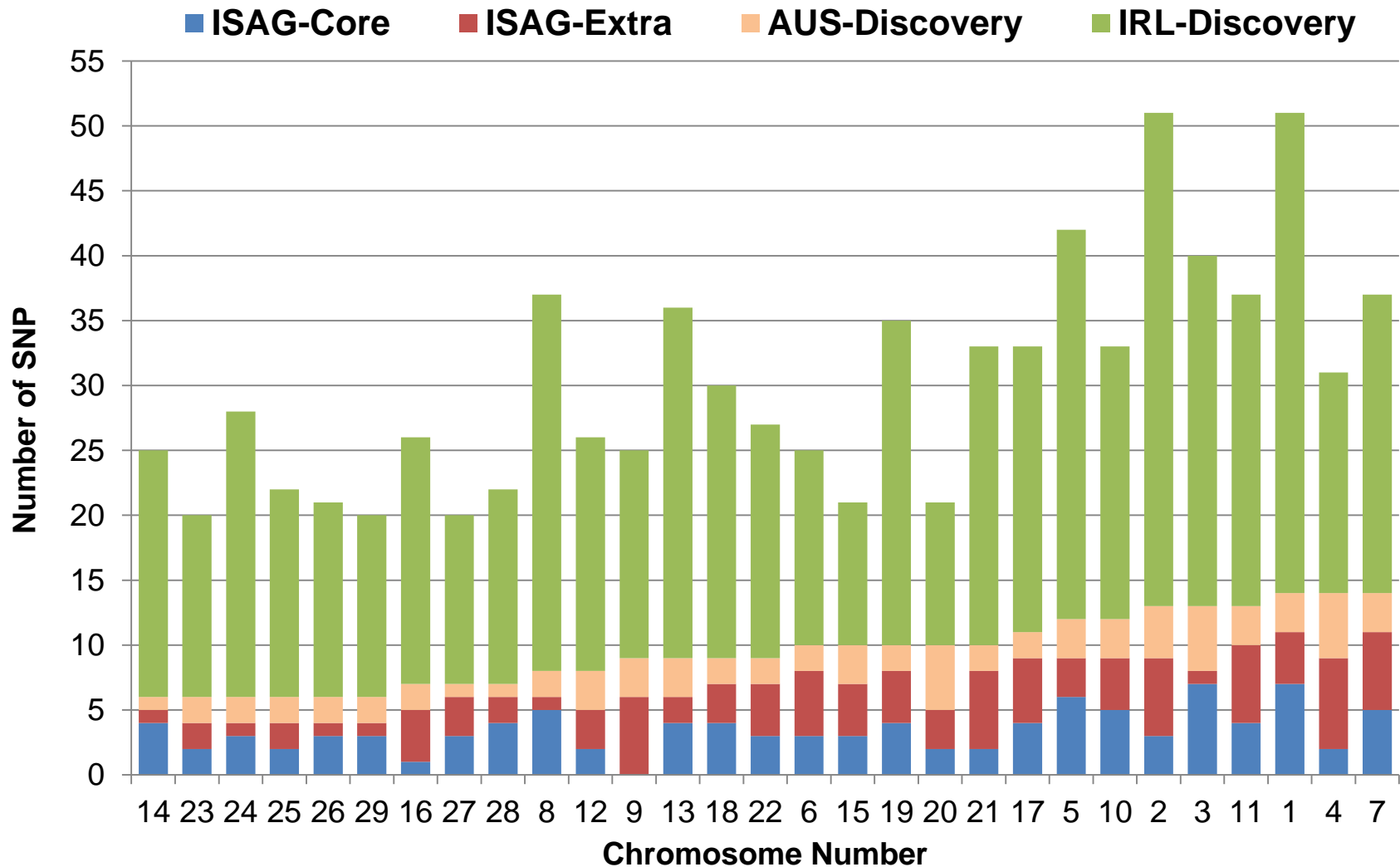
Accuracy of GEBV from PSE 875 Imputation to 50K

- **Very high correlations resulted between evaluations based on imputation of PSE 875 versus use of the full genotype available**
 - >93% for Direct Genomic Value among traits examined
 - >95% for GEBV as a combination of PA and DGV
- **Most animals in this analysis have Sire and MGS genotyped with 50K or higher panel**
 - Gives high accuracy for imputation and DGV estimation
- **High GEBV accuracy from imputation of PSE 875 is not expected to be retained for animals with more generations away from 50K parents**

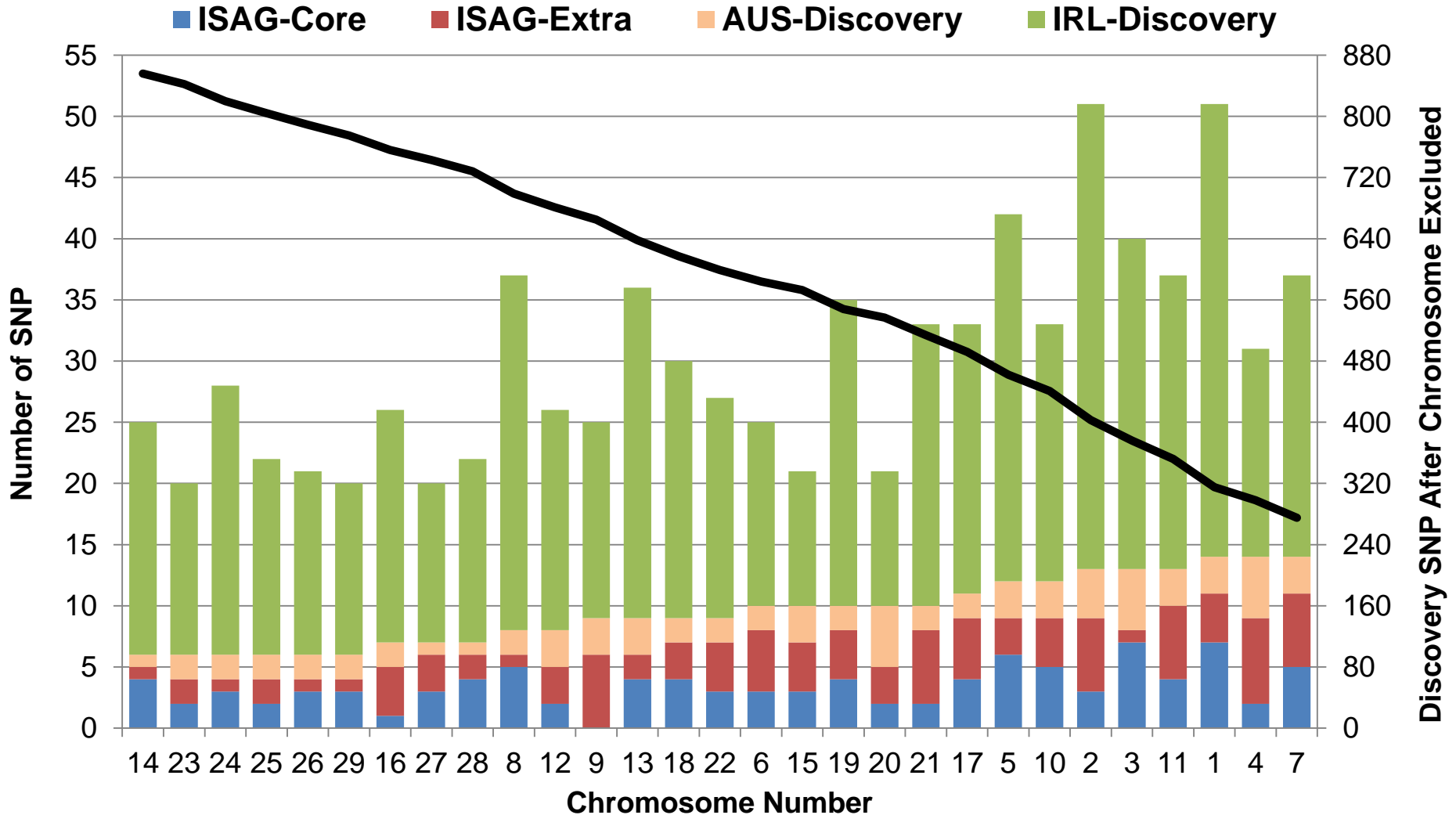
Accuracy of GEBV from PSE 875 Imputation to 50K

- **Given the results of this analysis, GenoEx-PSE should modify list of added SNP for discovery**
- **Total of current 875 SNP proposed include:**
 - 200 ISAG
 - 75 added for discovery based on research in Australia
 - 605 added for discovery based on research in USA and Ireland, of which 5 were in common with Australia
- **GEBV accuracy is attained since PSE 875 are well spaced across all 29 chromosome**
- **Need to modify the strategy of SNP selection to add for parentage discovery**

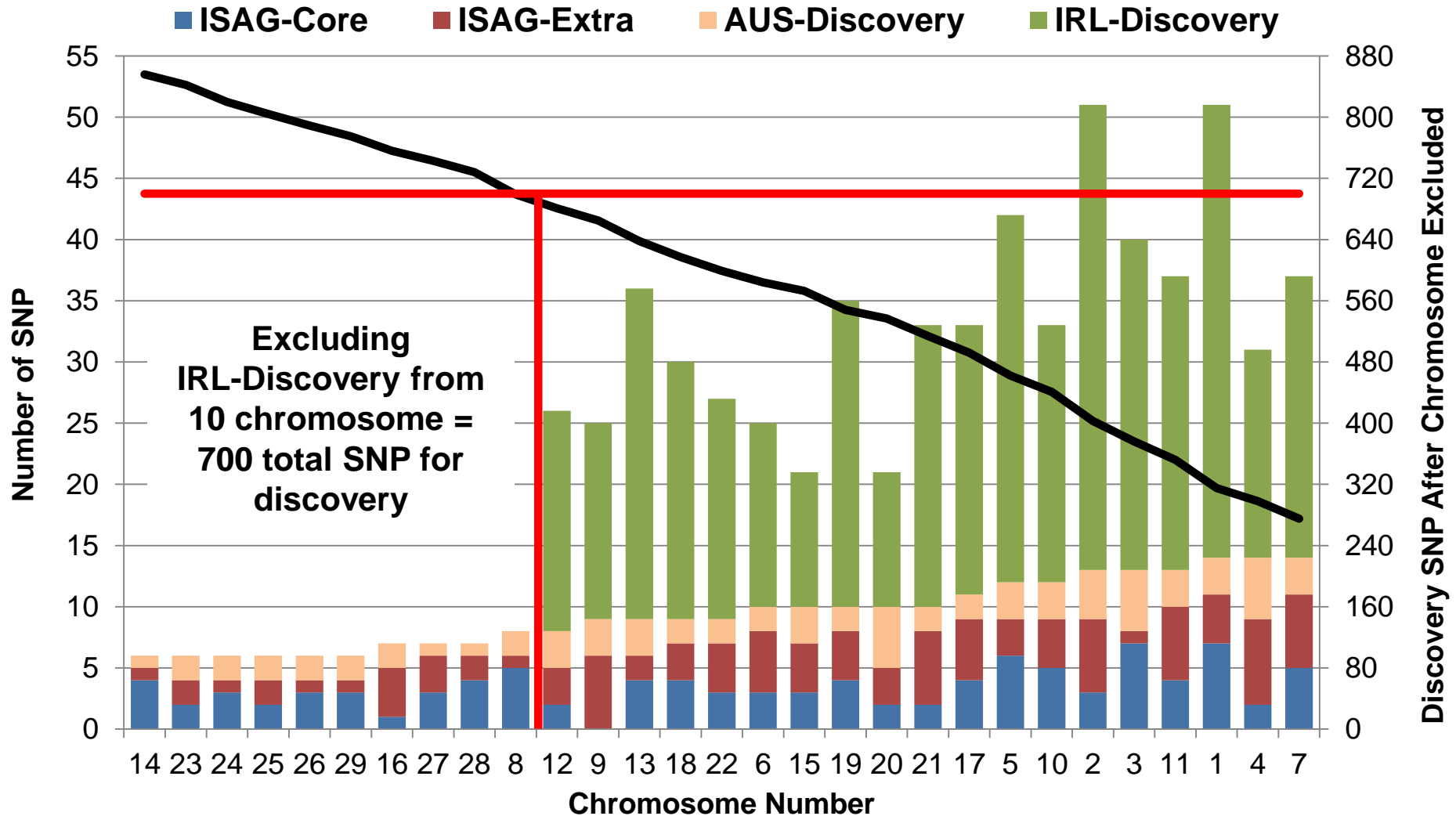
SNP Selection Strategy for GenoEx-PSE Discovery



SNP Selection Strategy for GenoEx-PSE Discovery



SNP Selection Strategy for GenoEx-PSE Discovery





Summary

- Based on Analysis in Canada -

- **SNP recommended by ISAG perform very well for parentage verification**
- **PSE 875 SNP also perform very well for both verification (better than ISAG) and discovery**
- **Imputation from PSE 875 to 50K genotypes can lead to quite accurate DGV and GEBV**
- **Recommend to modify SNP selection added to GenoEx-PSE list for Parentage Discovery**
 - Keep all ISAG 200 and 75 added by Australia (N=275)
 - Select another 500 to 600 SNP by excluding SNP from some of the 29 chromosomes

Réseau laitier canadien



Canadian Dairy Network

Thank you!