



UNIVERSITY OF
GEORGIA

College of Agricultural &
Environmental Sciences

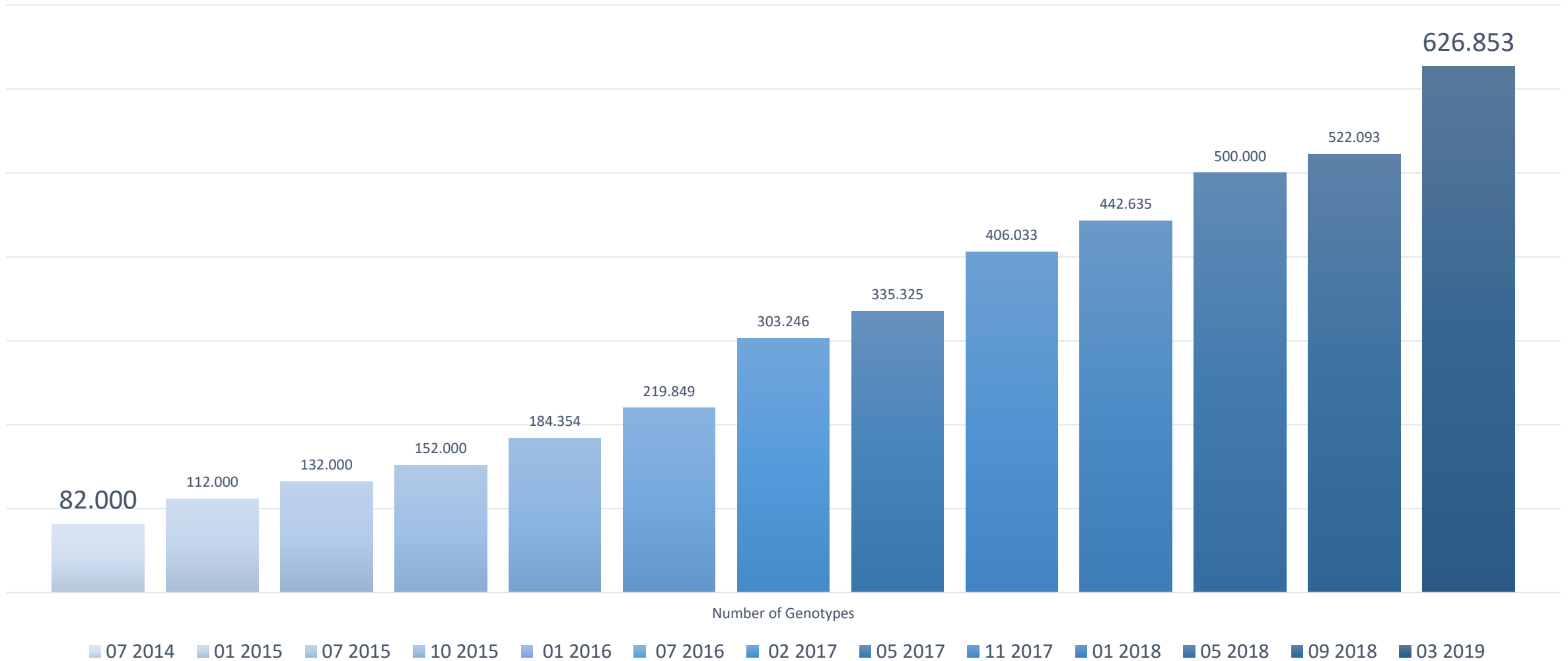
Genomic accuracy for indirect predictions based on SNP effects from ssGBLUP

Daniela Lourenco

I. Aguilar, A. Legarra, A. Garcia, S. Miller, S. Tsuruta, I. Misztal

Ever-increasing number of genotypes

American Angus



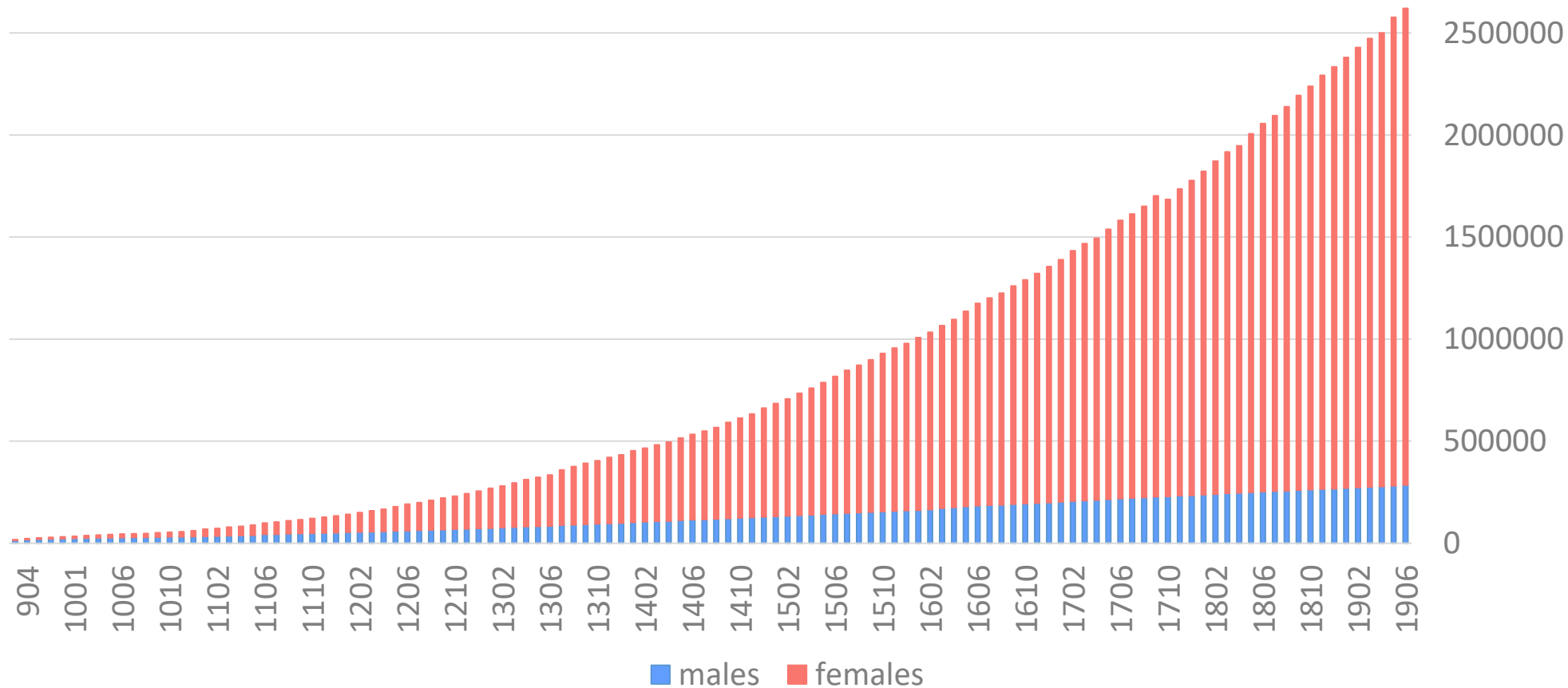
Ever-increasing number of genotypes

US Holsteins

89% Females

20% Phenotypes

16% PED miss



Do we need to include all genotyped animals in the ssGBLUP evaluations?

Indirect Predictions

- GBLUP

- ssGBLUP

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{G}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{H}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

SNP
effects

GEBVs

$$\hat{\mathbf{a}} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \hat{\mathbf{u}}$$

$$\hat{\mathbf{u}}_{ip} = \mathbf{Z}^* \hat{\mathbf{a}}$$

VanRaden 2008


Stranden & Garrick 2009

Wang et al. 2012

When to use indirect predictions

- Not all genotyped animals are in the evaluations
 - Animals with incomplete pedigree increase bias and reduce accuracy
- Interim evaluations
 - Between official runs
- Commercial products
 - e.g. GeneMax for non-registered animals

APY and Indirect Predictions

$$\hat{\mathbf{a}} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \hat{\mathbf{u}}$$


$$\mathbf{G}_{APY}^{-1} = \begin{bmatrix} \mathbf{G}_{cc}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\mathbf{G}_{cc}^{-1} \mathbf{G}_{cn} \\ \mathbf{I} \end{bmatrix} \mathbf{M}_{nn}^{-1} \begin{bmatrix} -\mathbf{G}_{nc} \mathbf{G}_{cc}^{-1} & \mathbf{I} \end{bmatrix}$$

Misztal et al., 2014

$$\hat{\mathbf{a}} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}_{APY}^{-1} \hat{\mathbf{u}}$$


$\text{CORR}(\text{GEBV}, \text{IP}_1) > 0.99$

$\text{CORR}(\text{GEBV}, \text{IP}_2) > 0.99$

$\text{CORR}(\text{IP}_1, \text{IP}_2) > 0.99$

Lourenco et al., 2018

Is it possible to compute accuracies for IP?

How good are the accuracies of IP compared to accuracies of GEBV?

Accuracy of GEBV and IP

- Accuracy of GEBV

- $acc_i = (1 - C_{ii}^{u_2 u_2} \sigma_u^{-2})^{**} 0.5$

└──────────┬──────────> PEV of GEBV

SNP effects

GEBVs

$$\hat{\mathbf{a}} = \lambda \mathbf{D} \mathbf{Z}' \mathbf{G}^{-1} \hat{\mathbf{u}}$$

$$\hat{\mathbf{u}}_{ip} = \mathbf{Z}^* \hat{\mathbf{a}}$$

- Accuracy of IP

- $acc_i = (1 - z_i C^{aa} z_i' \sigma_u^{-2})^{**} 0.5$

Liu et al., 2017

└──────────┬──────────> PEC of SNP

From PEC of GEBV to PEC of SNP

Gualdrón Duarte et al. *BMC Bioinformatics* 2014, **15**:246
<http://www.biomedcentral.com/1471-2105/15/246>



METHODOLOGY ARTICLE

Open Access

Rapid screening for phenotype-genotype associations by linear transformations of genomic evaluations

Jose L Gualdrón Duarte¹, Rodolfo JC Cantet¹, Ronald O Bates², Catherine W Ernst², Nancy E Raney² and Juan P Steibel^{2,3*}

Aguilar et al. *Genet Sel Evol* (2019) **51**:28
<https://doi.org/10.1186/s12711-019-0469-3>



SHORT COMMUNICATION

Open Access

Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle

Ignacio Aguilar¹, Andres Legarra^{2*}, Fernando Cardoso^{3,4}, Yutaka Masuda⁵, Daniela Lourenco⁵ and Ignacy Misztal⁵



doi: 10.1111/age.12378

Meta-analysis of genome-wide association from genomic prediction models

Y. L. Bernal Rubio^{*†}, J. L. Gualdrón Duarte^{*}, R. O. Bates^{*}, C. W. Ernst^{*}, D. Nonneman[†], G. A. Rohrer[†], A. King[†], S. D. Shackelford[†], T. L. Wheeler[†], R. J. C. Cantet^{†§} and J. P. Steibel^{*¶}



Implementation for the BLUPF90 software suite

From PEC of GEBV to PEC of SNP

1) Factorize and Invert LHS of ssGBLUP with YAMS (Masuda et al., 2014)

2) Store LHS⁻¹ into a file

3) Solve the MME for $\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}$ using the sparse Cholesky factor

4) Extract from LHS⁻¹ coefficients for genotyped animals ($\mathbf{C}^{u_2u_2}$)

5) Obtain prediction error covariance of SNPs (Gualdron-Duarte et al., 2014):

$$PEC = C^{aa} = \frac{1}{2\sum pq} \mathbf{Z}' \mathbf{G}^{-1} (\mathbf{G}\sigma_u^2 - \mathbf{C}^{u_2u_2}) \mathbf{G}^{-1} \mathbf{Z} \frac{1}{2\sum pq}$$

6) Backsolve GEBV to SNP effects (\hat{a}): $\hat{a} = \frac{1}{2\sum pq} \mathbf{Z}' \mathbf{G}^{-1} \hat{u}$

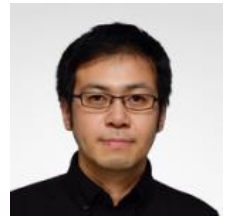
7) Compute IP and accuracy

$$acc_i = (1 - z_i C^{aa} z_i' \sigma_u^{-2})^{**} 0.5 \quad (\text{Liu et al., 2017})$$

blupf90



Ignacio
Aguilar



Yutaka
Masuda

postGSf90



Andres
Legarra

predf90

Data

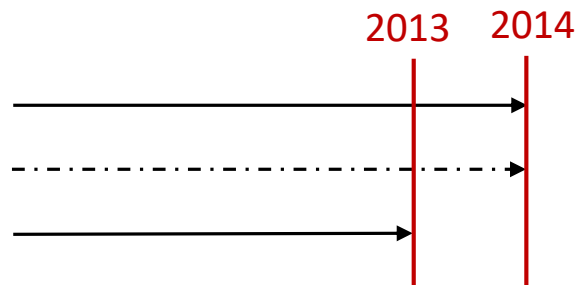
- American Angus Association
- Post-weaning gain

- Complete data

202k Pedigrees

60k Genotypes

35k Phenotypes



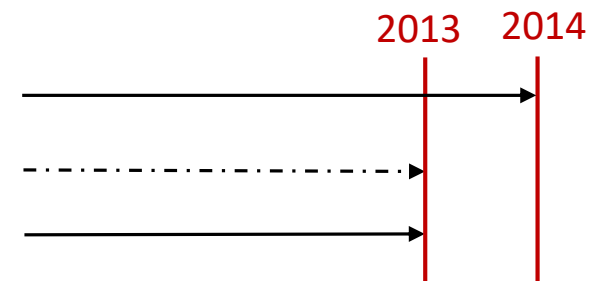
GEBV and accuracy
2k young animals born in 2014

- Reduced data

202k Pedigrees

58k Genotypes

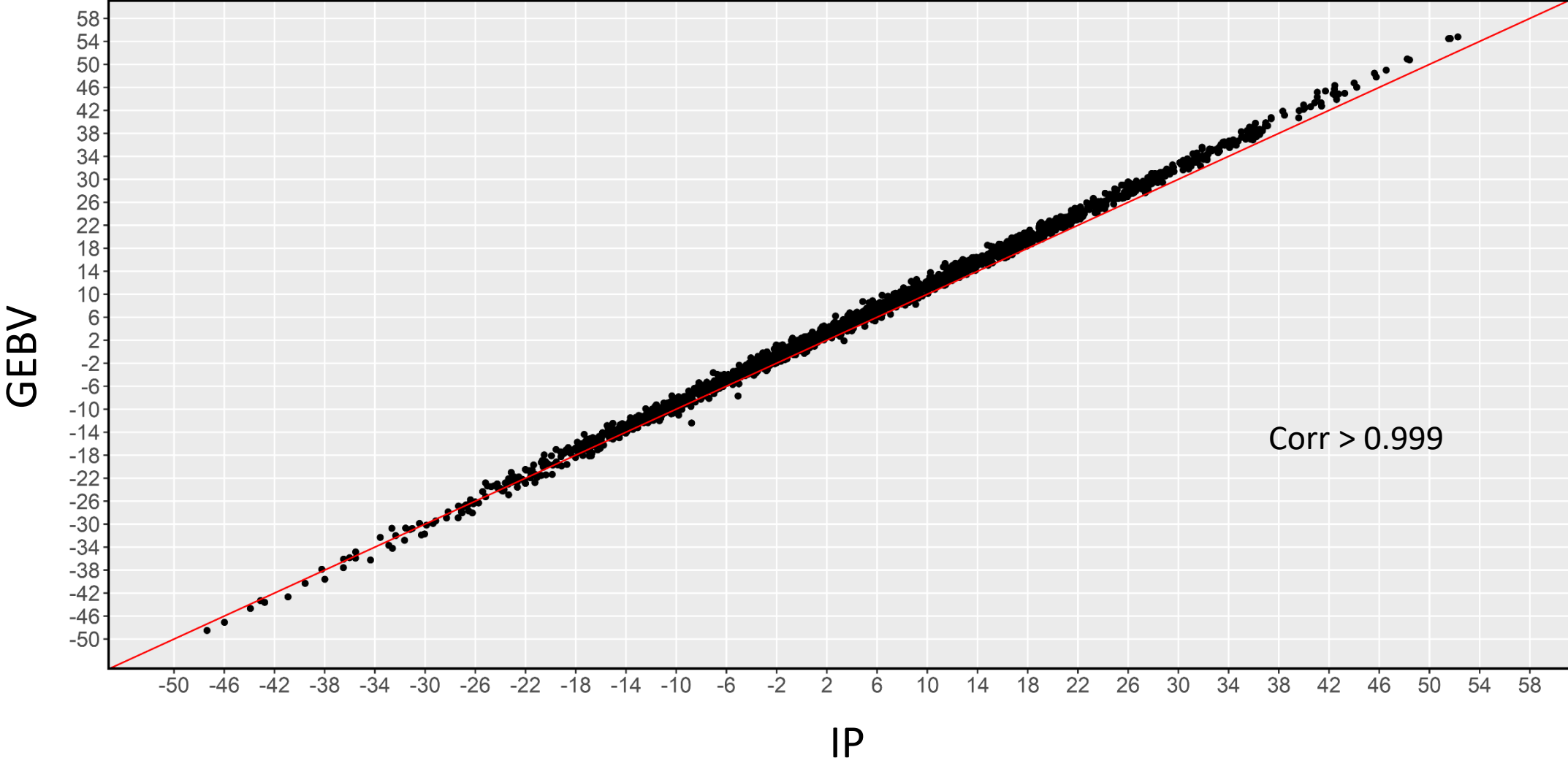
35k Phenotypes



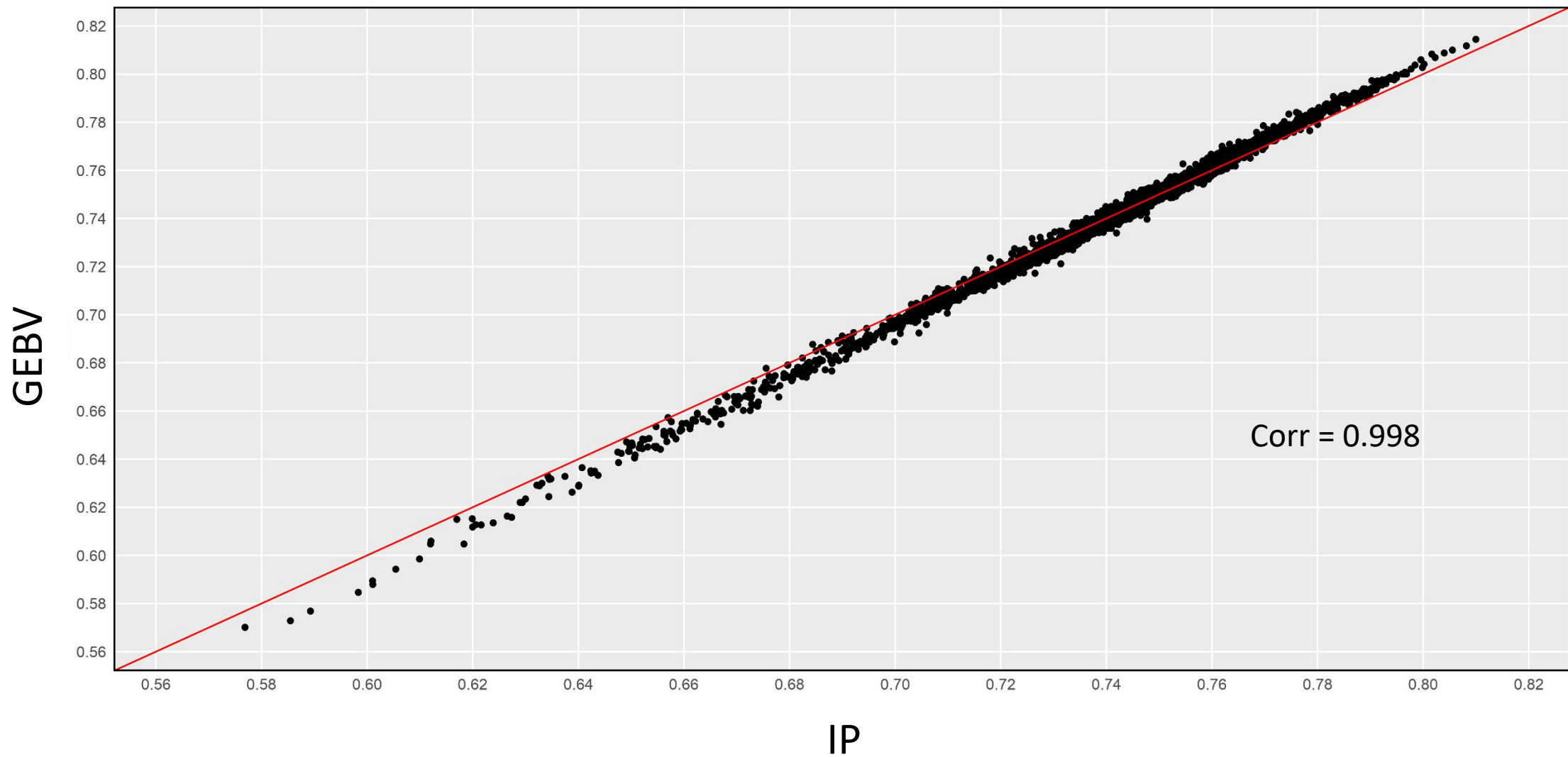
IP and accuracy
2k young animals born in 2014

GEBV and IP

Post Weaning Gain



Accuracy of GEBV and IP



Need to account for blending and tuning

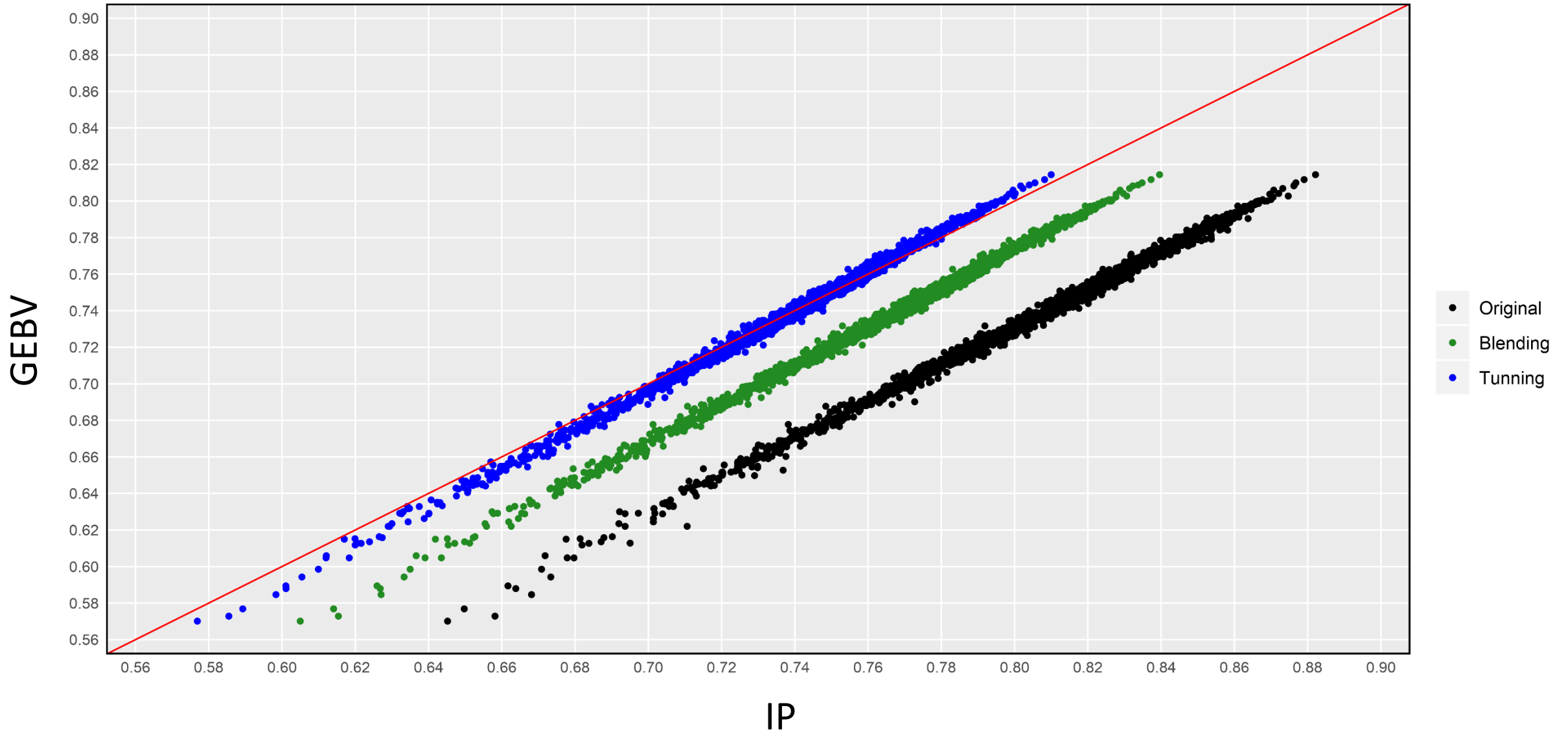
5) Obtain prediction error covariance of SNPs (Gualdron-Duarte et al., 2014)

$$PEC = C^{aa} = \alpha\delta \frac{1}{2\sum pq} \mathbf{Z}' \mathbf{G}^{-1} (\mathbf{G}\sigma_u^2 - \mathbf{C}^{u_2 u_2}) \mathbf{G}^{-1} \mathbf{Z} \frac{1}{2\sum pq} \alpha\delta$$

Tuning (Vitezica et al., 2011)
 $\delta = 1 - FST/2$
 Blending

6) Backsolve GEBV to SNP effects (\hat{a}): $\hat{a} = \alpha\delta \frac{1}{2\sum pq} \mathbf{Z}' \mathbf{G}^{-1} \hat{u}$

Accounting for blending and tuning



Limitations and next step

1) Factorize and Invert LHS of ssGBLUP with YAMS (Masuda et al., 2014)

*Computationally
expensive!!!*

- Limitation (same as REML)
 - Cannot be used for large datasets and large numbers of genotyped animals
 - Need to be used in a subset of animals
- Next step
 - Take advantage of the dimensionality of **G** (Pocrnic et al., 2016)
 - Use information on core animals to obtain LHS^{-1} and PEC

Final Considerations

- Accuracy of indirect predictions can be computed based on ssGBLUP
 - PEC for GEBV can be converted to PEC for SNP
 - Correlations (accuracy IP, accuracy GEBV) > 0.99
- Computations are limited by the amount of data
 - Enable computations of accuracy of IP for large-scale evaluations

Acknowledgements

