Considerations and Strategies for Imputation Reference Panel Construction

Beef Improvement Federation

Advancements in Genomics and Genetic Prediction

June 2, 2022

Troy Rowan Ph.D.

University of Tennessee

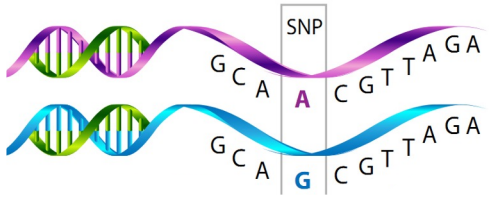
Real. Life. Solutions.™

UTEXTENSION INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE

A big thank-you

Bob Schnabel

Mizzou



Ranch Tested. Rancher Trusted.

Red Angus

Red Angus Association of America

Genomic Data Report

Reg#: Prefix: ID: Birthdate: 3/15/2012 Sex: C

Name:

Genomic Test Performed: GBL/HD50K Angus/01 Job Number: Report Date: 12/7/2012

		CED	BW	WW	YW	Milk	TM	ME	HPG	CEM	STAY	MARB	YG	CW	REA	FAT
Current (With PF50K Genomic Data)	EPD		-1.8	69	95	28	63	5	11		9	0.31	0.62	23	0.1	0.22
	ACC		34	23	26	20		10	10		10	24	24	23	24	35
	Top%		26	7	18	2	1	73	32		44	60	99	28	59	99

		CED	BW	WW	YW	Milk	TM	ME	HPG	CEM	STAY	MARB	YG	CW	REA	FAT
Previous (Without PF50K Genomic Data)	EPD		0	69	102	23	58	5	11		9	0.33	-0.08	27	-0.08	-0.03
	ACC		P+	P+	P+	P		P	10		P	P	P	P	P	P
	Top%		65	7	10	11	2	73	32		44	55	35	19	94	4

Where will future improvements to genomics prediction come from?

Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

EDITORIAL

Is genomic selection now a mature technology?

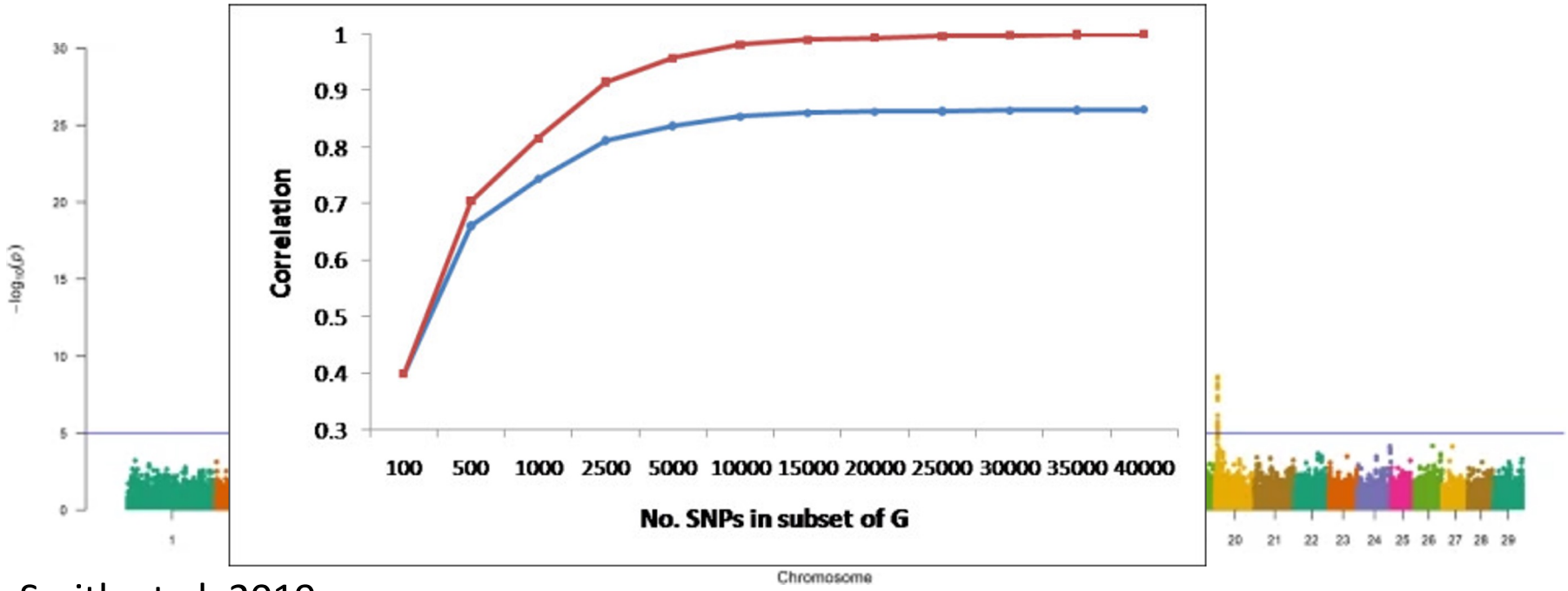
I. Misztal
Subject Editor for JABG
University of Georgia
E-mail: ignacy@uga.edu

UTCVM INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE

UTAGRESEARCH INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}\lambda_G \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

Genomic Prediction



Rolf et al. 2012, Smith et al. 2019

Complex Trait Mapping

PNAS

Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits

Ruidong Xiang[a,b,1], Irene van den Berg[a,b], Iona M. MacLeod[b], Benjamin J. Hayes[b,c], Claire P. Prowse-Wilkins[a,b], Min Wang[b,d], Sunduimijid Bolormaa[b], Zhiqian Liu[b], Simone J. Rochfort[b,d], Coralie M. Reich[b], Brett A. Mason[b], Christy J. Vander Jagt[b], Hans D. Daetwyler[b,d], Mogens S. Lund[e], Amanda J. Chamberlain[b], and Michael E. Goddard[a,b]

[a]Faculty of Veterinary & Agricultural Science, The University of Melbourne, Parkville, VIC 3052, Australia; [b]Agriculture Victoria, AgriBio, Centre for AgriBiosciences, Bundoora, VIC 3083, Australia; [c]Centre for Animal Science, The University of Queensland, St. Lucia, QLD 4067, Australia; [d]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia; and [e]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-21001-0 OPEN

Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations

Ruidong Xiang[1,2✉], Iona M. MacLeod[2], Hans D. Daetwyler[2,3], Gerben de Jong[4], Erin O'Connor[5], Chris Schrooten[6], Amanda J. Chamberlain[2] & Michael E. Goddard[1,2]

Real. Life. Solutions.™

Pausch et al. Genet Sel Evol (2017) 49:24
DOI 10.1186/s12711-017-0301-x

GSE Genetics Selection Evolution

RESEARCH ARTICLE Open Access

Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle

Hubert Pausch[1*], Iona M. MacLeod[1], Ruedi Fries[2], Reiner Emmerling[3], Phil J. Bowman[1,4], Hans D. Daetwyler[1,4] and Michael E. Goddard[1,5]

frontiers in Genetics

ORIGINAL RESEARCH
published: 21 January 2021
doi: 10.3389/fgene.2020.603822

Genomic Prediction Based on SNP Functional Annotation Using Imputed Whole-Genome Sequence Data in Korean Hanwoo Cattle

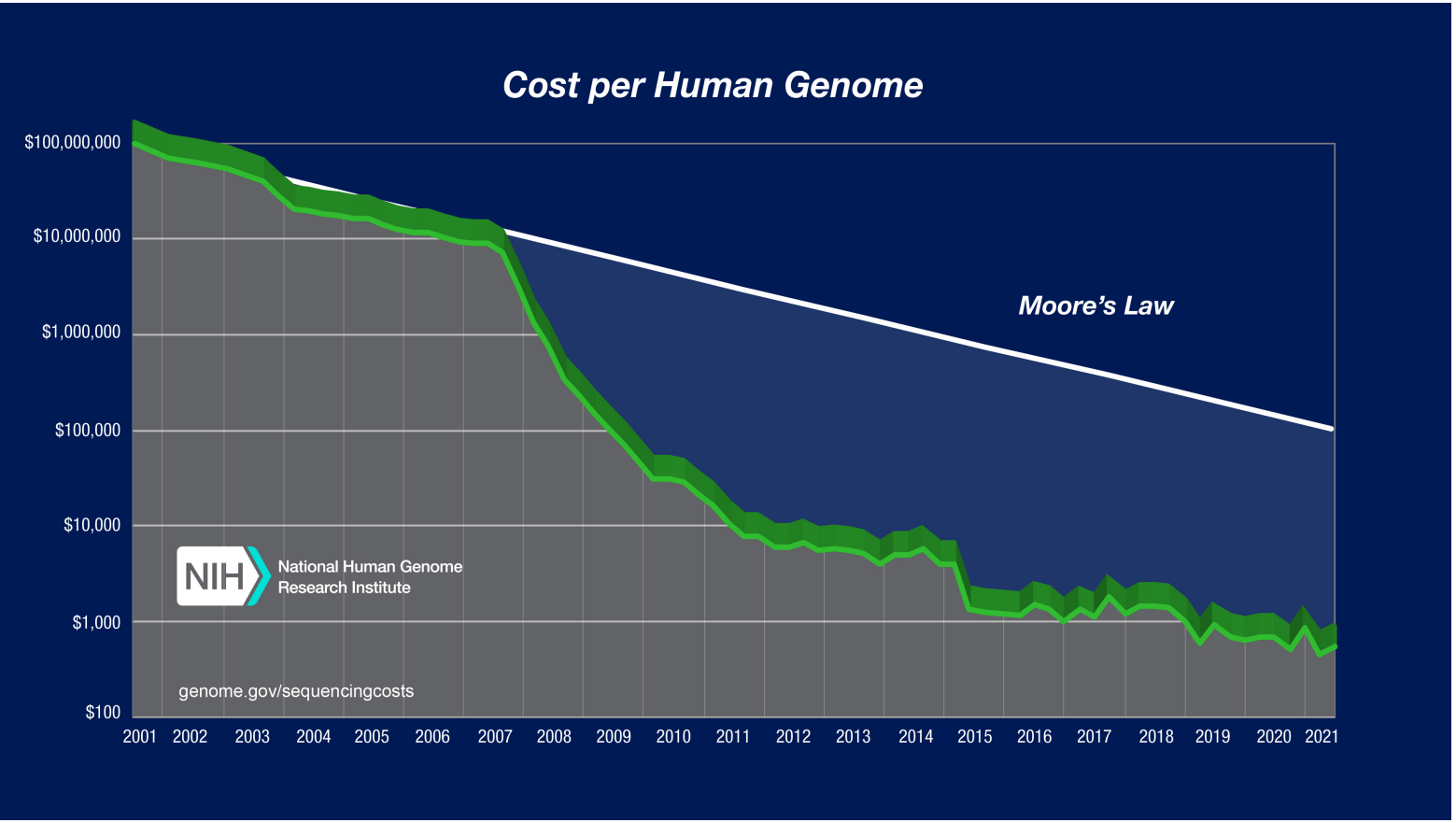
Bryan Irvine M. Lopez[1†], Narae An[1†], Krishnamoorthy Srikanth[1], Seunghwan Lee[2], Jae-Don Oh[3], Dong-Hyun Shin[4], Woncheoul Park[1], Han-Ha Chai[1], Jong-Eun Park[1] and Dajeong Lim[1]*

[1] Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Wanju, South Korea, [2] Department of Animal Science and Biotechnology, Chungnam National University, Daejeon, South Korea, [3] Department of Animal Biotechnology, Chonbuk National University, Jeonju, South Korea, [4] Department of Agricultural Convergence Technology, Chonbuk National University, Jeonju, South Korea

UTCVM INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE

UTAGRESEARCH INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE

Sequencing opportunities



- Declining costs of sequencing
- Evolving tech landscape
- Low-pass
- Imputation

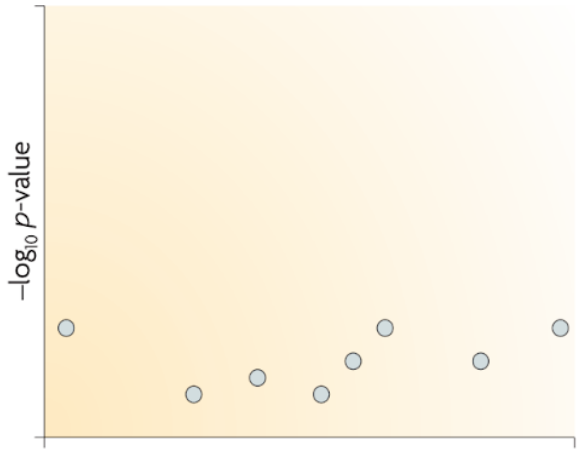
Future utility of sequence information in genetic evaluations

- ***Orders of magnitude*** more information!
- Leveraging functional annotations
- Causal variant discovery & implementation
- Feature selection genomic prediction
- Integration… is another issue

a Genotype data with missing data at untyped SNPs (grey question marks)

1	?	?	?	1	?	1	?	0	2	2	?	?	2	?	0
0	?	?	?	2	?	2	?	0	2	2	?	?	2	?	0
1	?	?	?	2	?	2	?	0	2	1	?	?	2	?	0
1	?	?	?	2	?	1	?	1	2	2	?	?	2	?	0
2	?	?	?	2	?	2	?	1	2	1	?	?	2	?	0
1	?	?	?	1	?	1	?	1	2	2	?	?	2	?	0
1	?	?	?	2	?	2	?	0	2	1	?	?	2	?	1
2	?	?	?	1	?	1	?	1	2	1	?	?	2	?	1
1	?	?	?	0	?	0	?	2	2	2	?	?	2	?	0

b Testing association at typed SNPs may not lead to a clear signal



Imputation visualized

49.2 miles

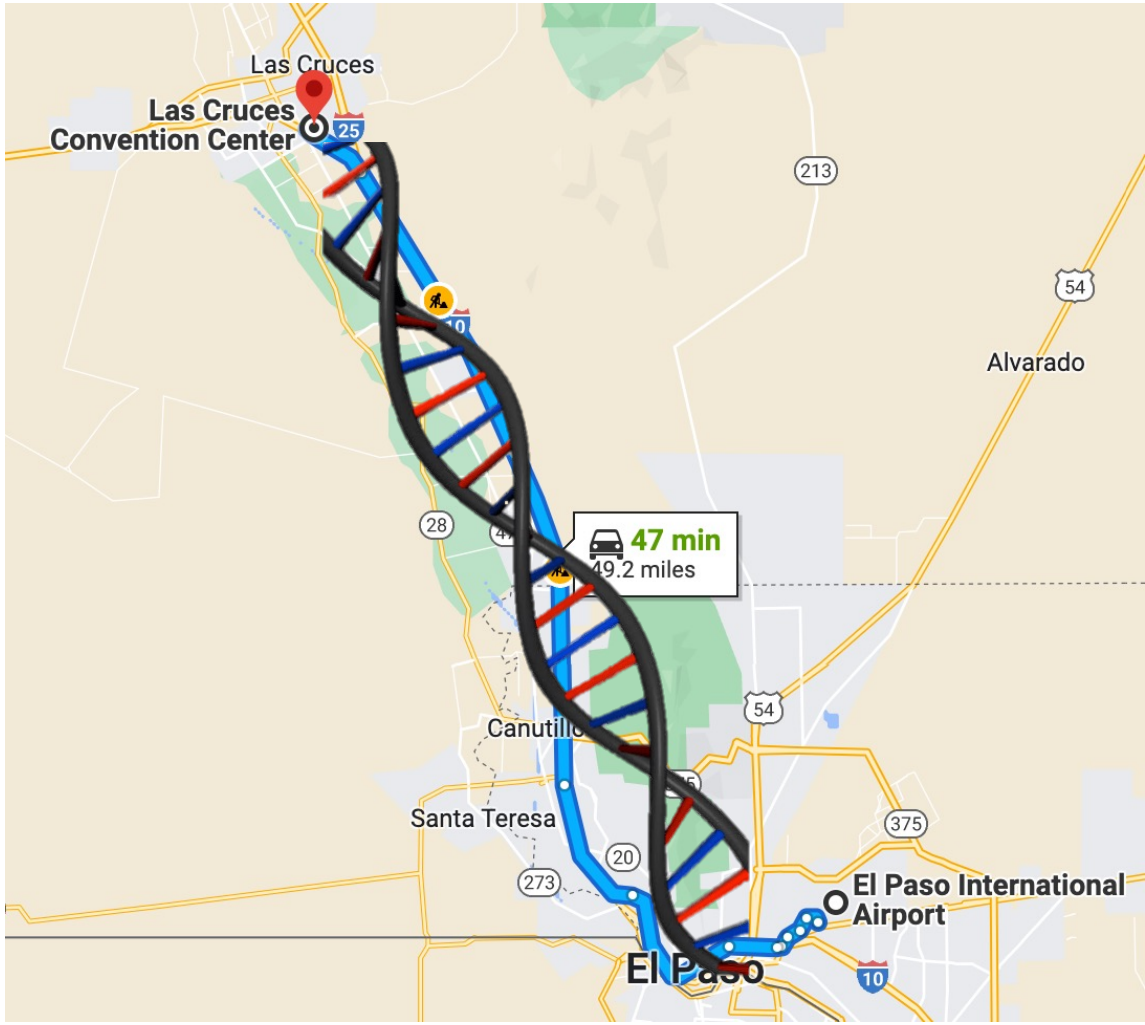
47 minutes

Unimputed chip genotypes:

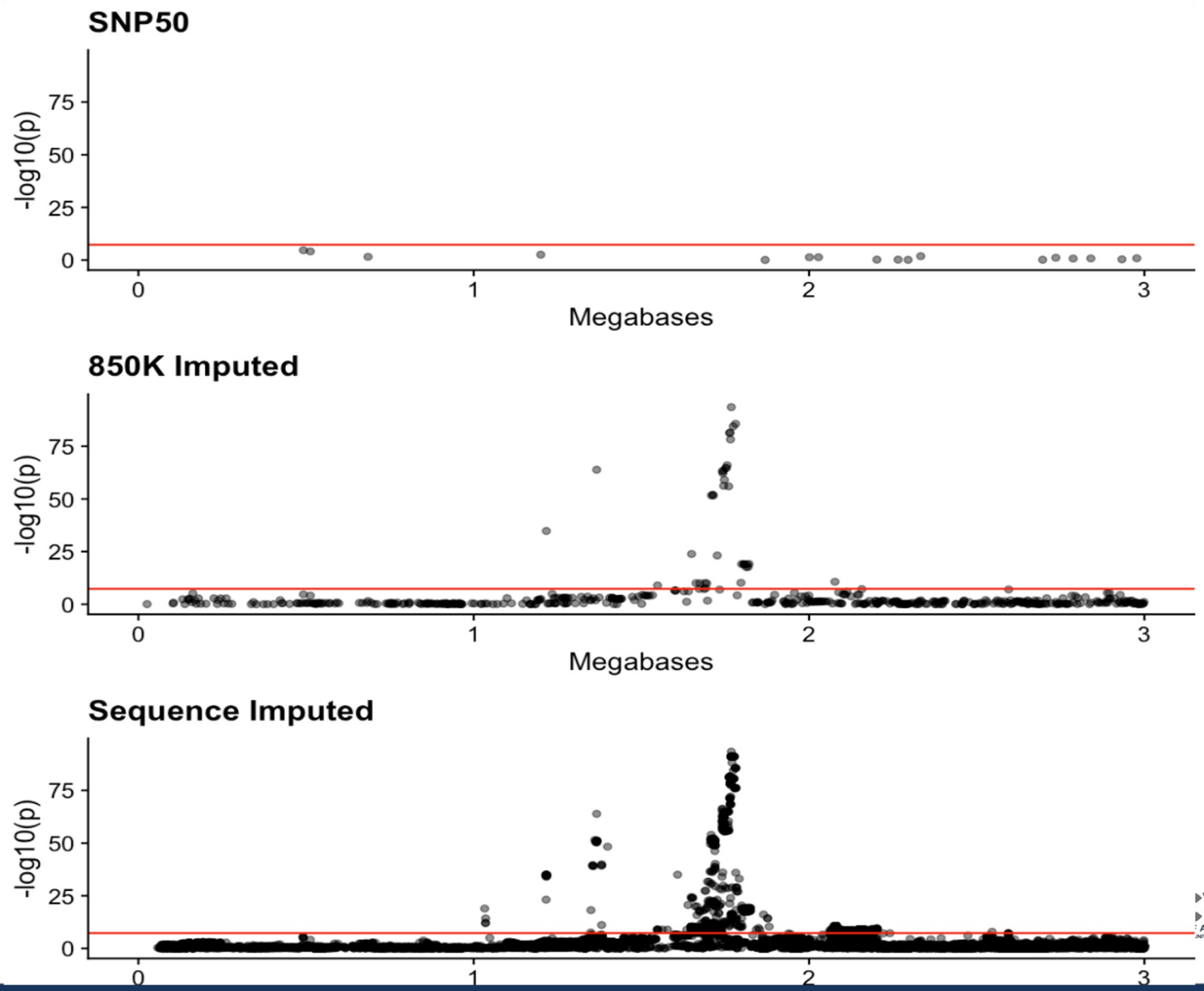
Look for 1-second every ~5 minutes!

Sequence-level genotypes:

A 1-second look every ~5 seconds!

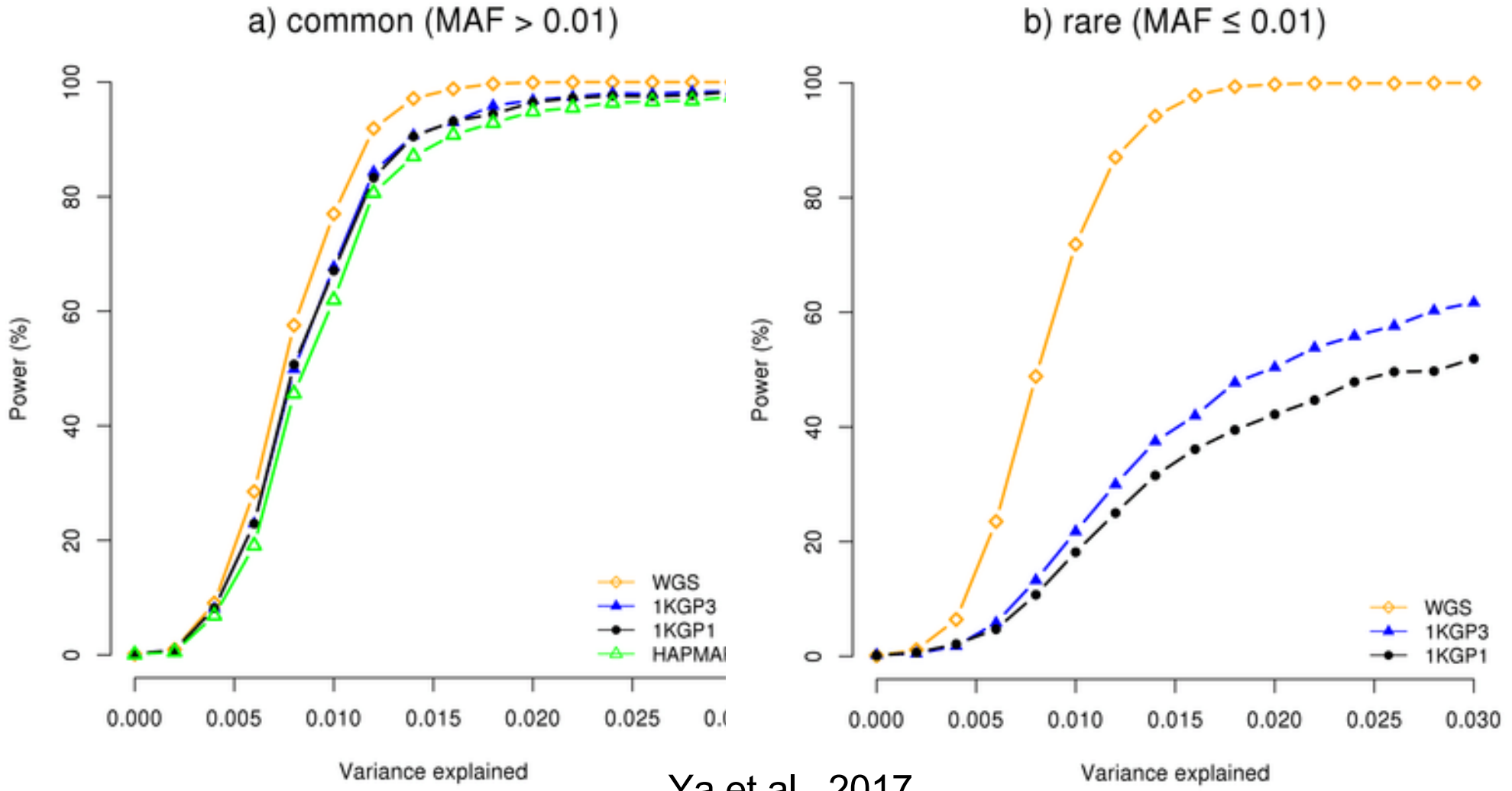


The Power of Imputation



Rowan et al. 2021

If our goal is to identify causal mutations, Imputation quality matters!



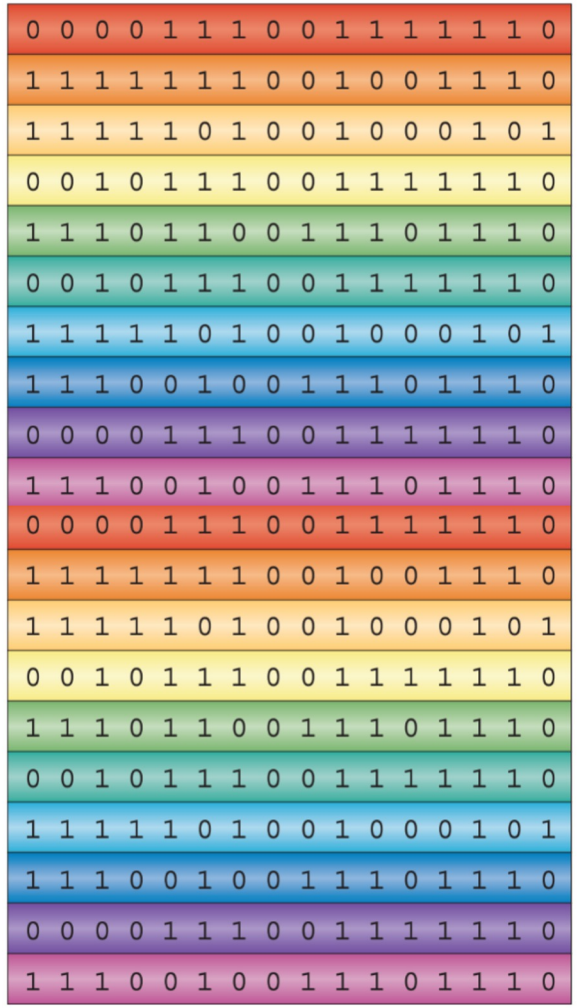
Ya et al., 2017

UTAGRESEARCH
INSTITUTE OF AGRICULTURE
THE UNIVERSITY OF TENNESSEE

What does accurate imputation need?

- **A large set of accurate reference haplotypes**
  - Higher density chip or sequence
  - More is always better
  - Crap in crap out

- Good reference genome
- Recombination map



Marchini et al. 2010

Coverage considerations

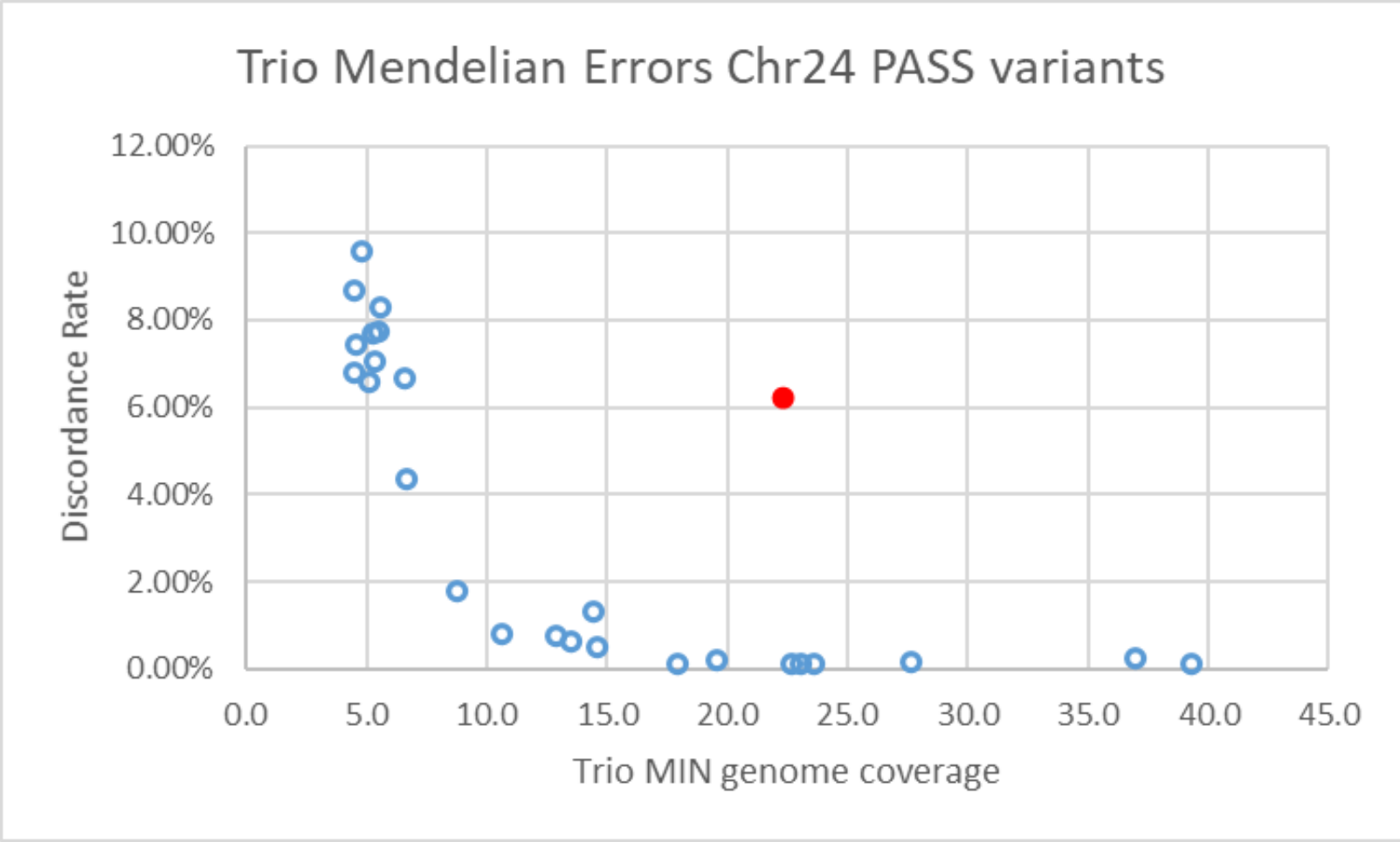


Figure courtesy of Bob Schnabel (Mizzou)

Evaluating Imputation Accuracy

- Metrics
  - Concordance: There's a SNP here. Did I impute the right genotype?
    - Deceiving at low minor allele frequencies
  - Correlation
  - Imputation quality statistic

- Truth set validation vs. internal software values

How to build a reference panel?

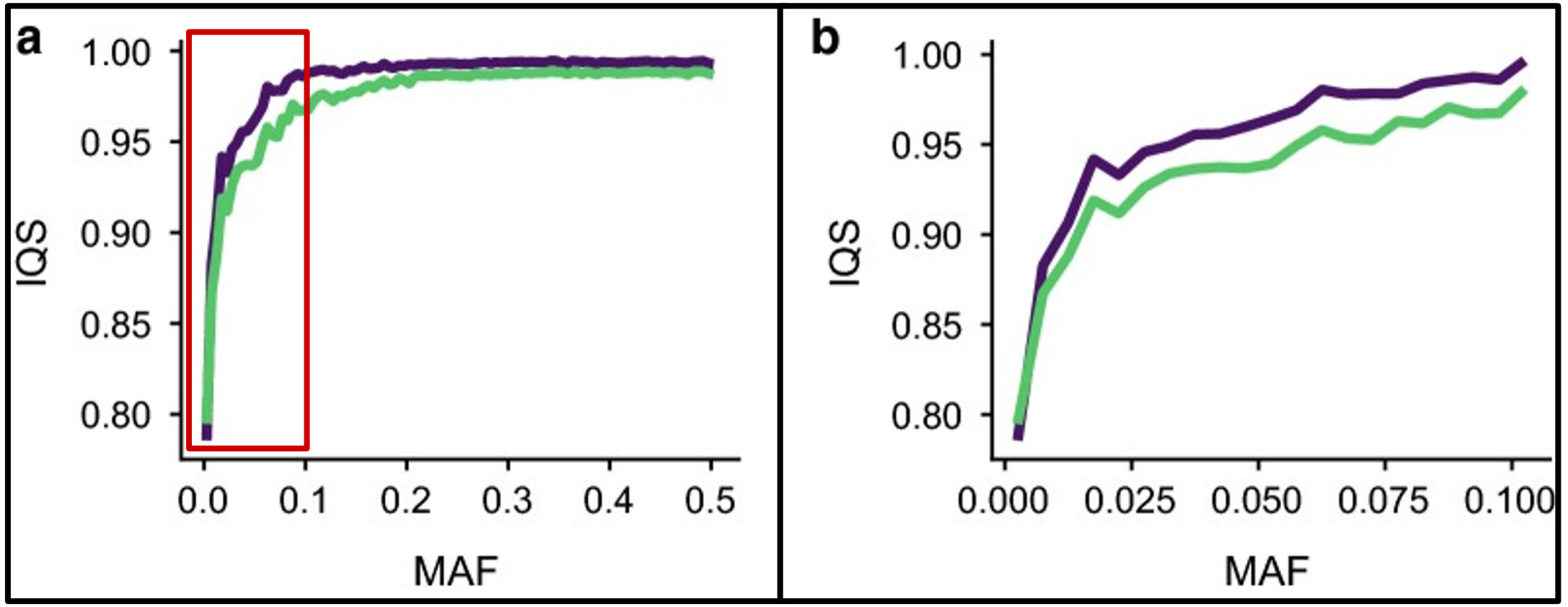
Breed Specific

Multi-breed

1) It’s really hard to impute a haplotype that you don’t observe!

2) A breed label doesn’t guarantee share (or not shared) haplotypes

Diverse samples need a diverse imputation reference panel



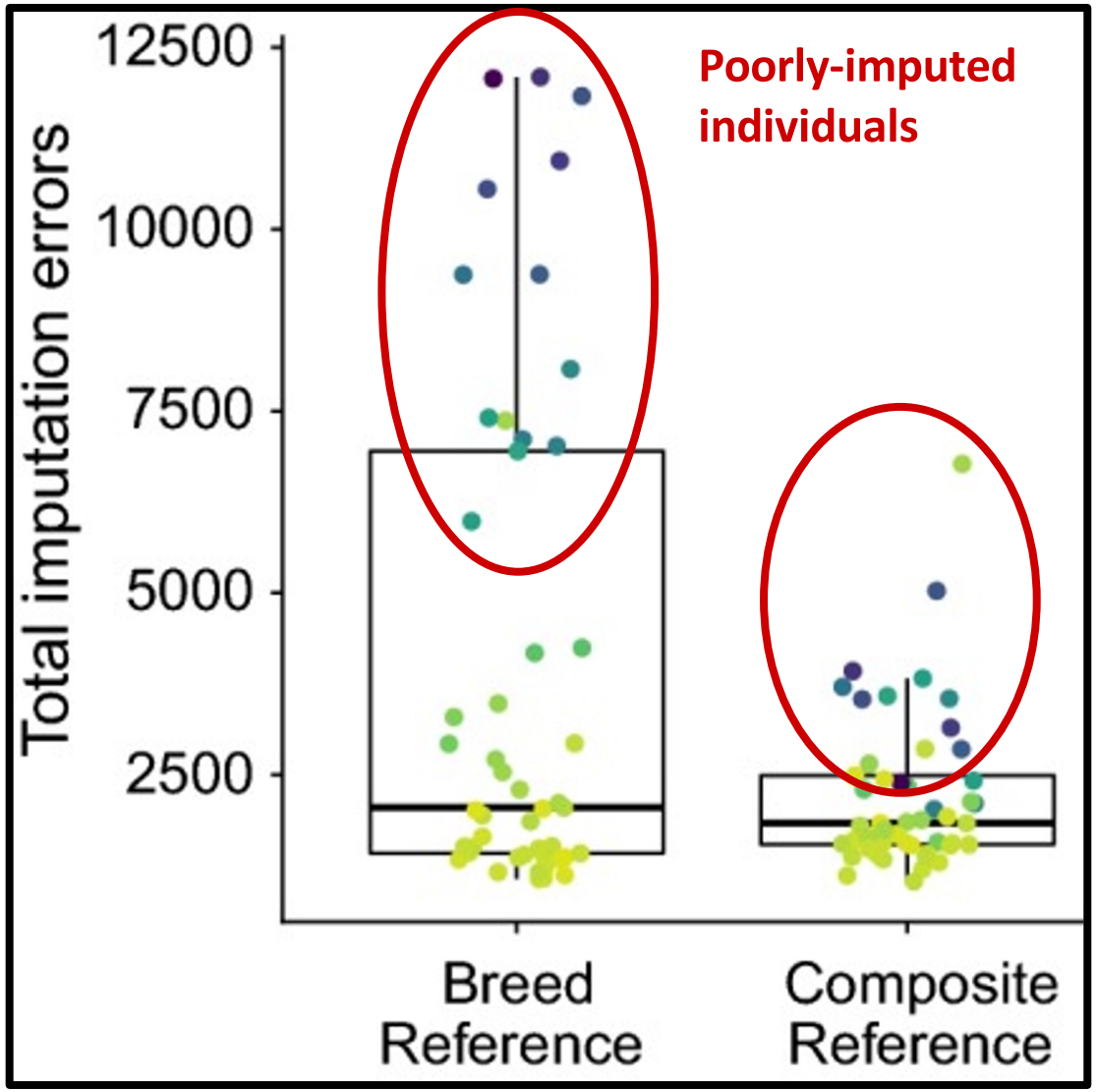
Multi-breed reference

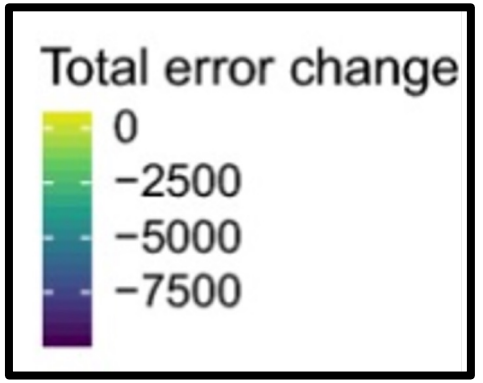
Within-breed reference

Rowan et al. 2019

Imputation is just pattern matching!

n =50 Gelbvieh



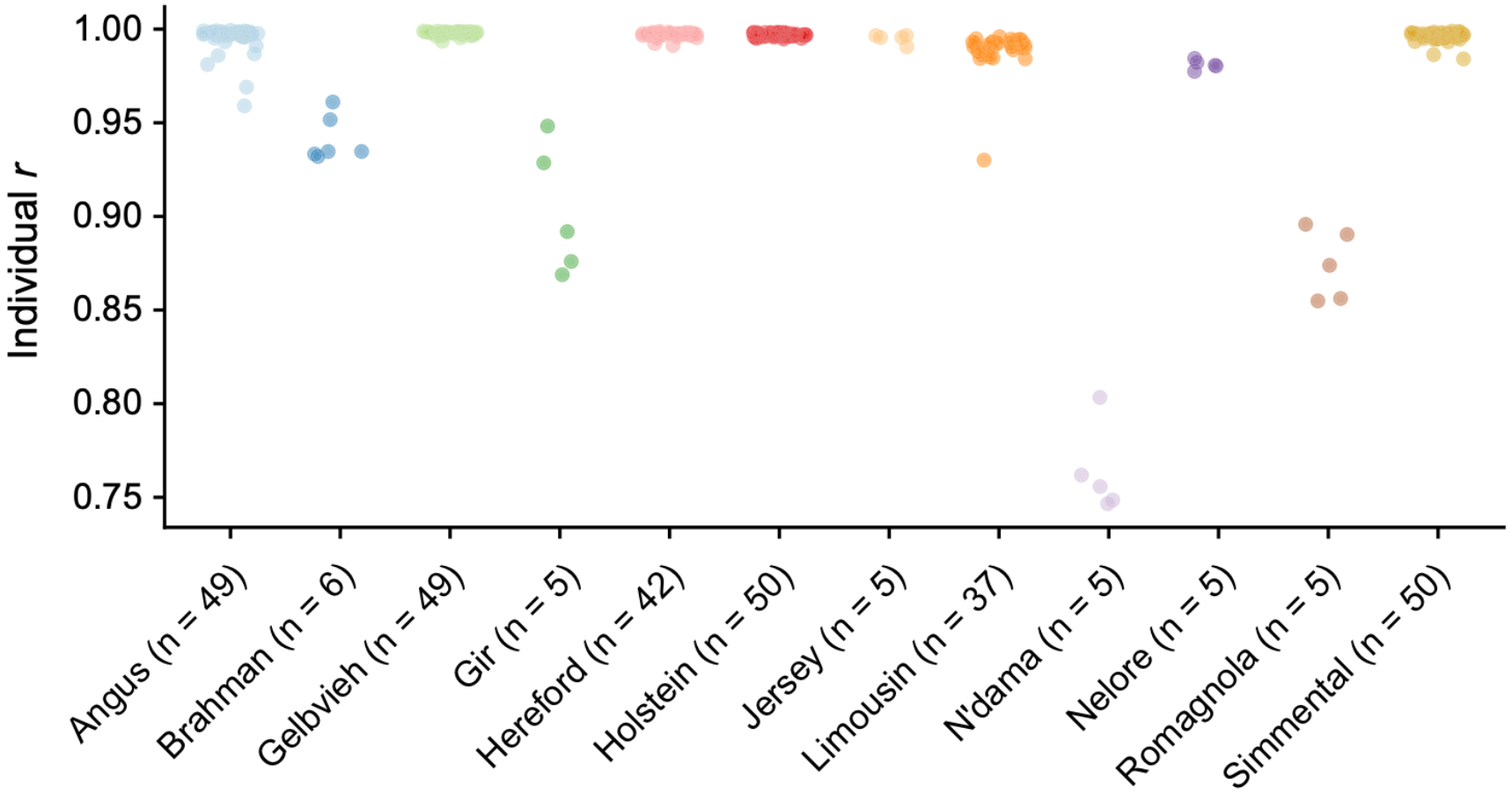


Real. Life. Solutions.™

UTCVM INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE

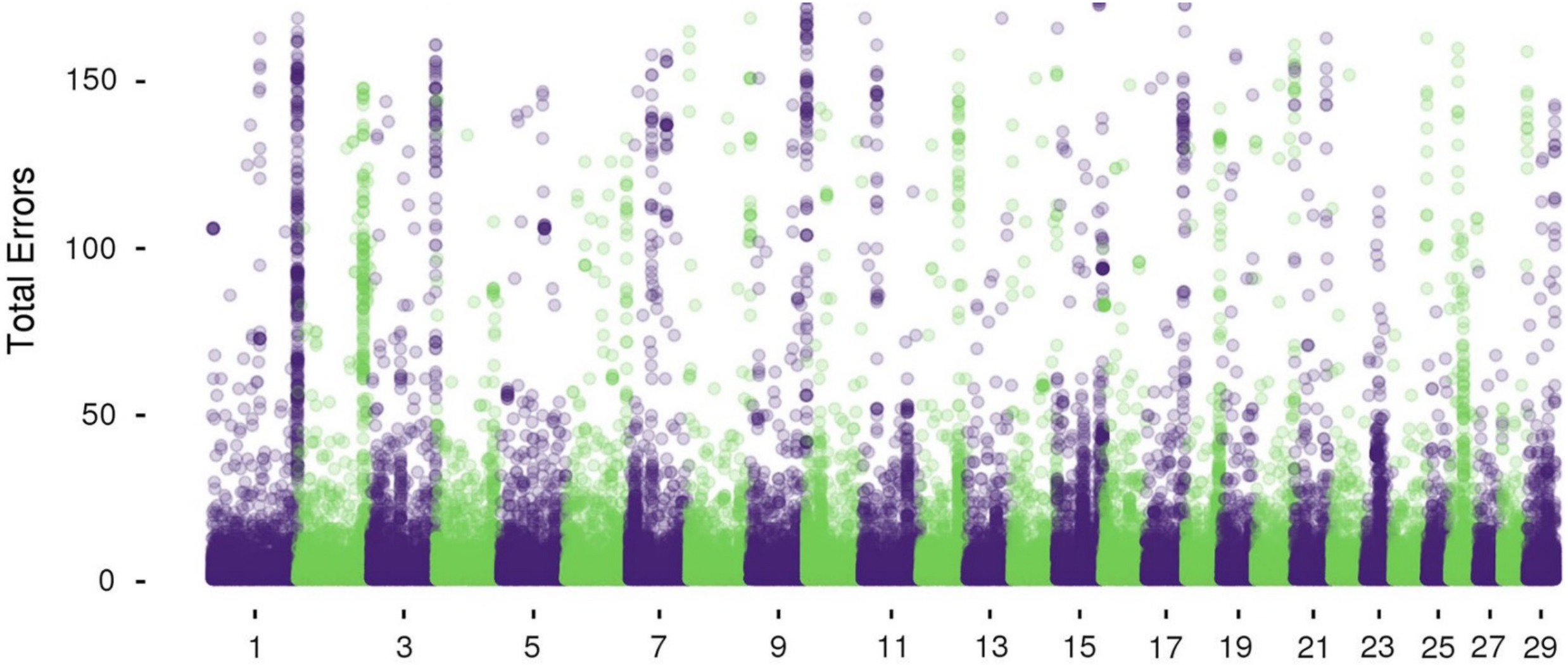
UTAGRESEARCH INSTITUTE OF AGRICULTURE THE UNIVERSITY OF TENNESSEE

Representation matters!



Breed	Mean	Min	Max
Gelbvieh	0.998	0.994	0.999
Hereford	0.997	0.991	0.999
Holstein	0.997	0.995	0.998
Simmental	0.996	0.984	0.999
Angus	0.995	0.959	0.999
Jersey	0.995	0.991	0.997
Limousin	0.989	0.930	0.996
Nelore	0.981	0.977	0.984
Brahman	0.941	0.932	0.961
Gir	0.903	0.869	0.948
Romagnola	0.874	0.855	0.896
N'Dama	0.763	0.747	0.803

Room for improvement



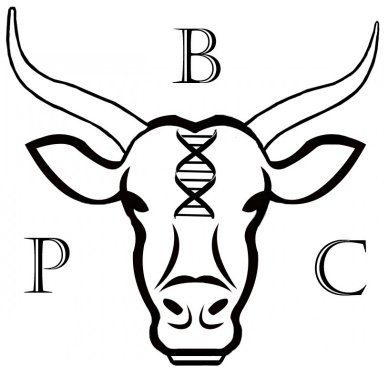
Rowan et al. 2019

Choosing animals for a reference

- Highest progeny number?
  - I'd argue we should sequence these regardless
- Sire/son sequencing = wasted resources
- Let genotype data drive sequencing

Future opportunities

- Pangenome-aware imputation
  - Imputing structural variation
  - Further breed diversity
- Software developments
  - Genotype callers
  - Imputation algorithms
- Rare alleles



Troy's recommendations to take or leave

- Sequence more animals!
  - Be intentional about representing novel haplotypes
  - And make the data public!
- Minimum of 10X coverage
- Don't assume: Evaluate imputation accuracy
- Get creative with the extra ~30 million SNPs

Takeaways

Future improvements to genetic evaluations will rely on sequence-level genotype information

Whether low-pass or using chips, imputation to sequence level can add value to genomic resources

Imputation is pattern matching. Relies on high-quality reference.

Continually improving references is essential. Minimum of 10X, choose animals with a plan

trowan@utk.edu
(865) 974-3190
Twitter: @TroyNRowan