Development of low-cost, high-throughput genotyping to complement SNP discovery by individual animal whole-genome sequencing

R. Mark Thallman[1], Warren M. Snelling[1], John D. Curry[2], Timothy P. L. Smith[1], Ted S. Kalbfleisch[3], Larry A. Kuehn[1], Tara G. McDaneld[1], Gary L. Bennett[1], Amanda K. Lindholm-Perry[1], Viacheslav Y. Fofanov[2], Heather Koshinsky[2], and E. John Pollak[1]

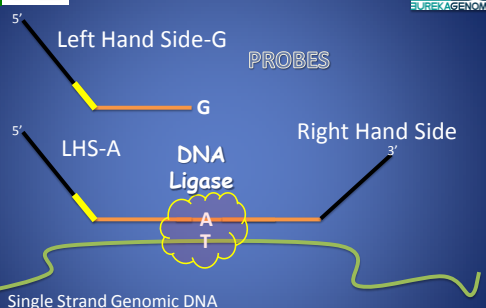
[1]USDA, Agricultural Research Service, U.S. Meat Animal Research Center, Clay Center, NE

[2]Eureka Genomics Inc., Hercules, CA and Sugarland, TX

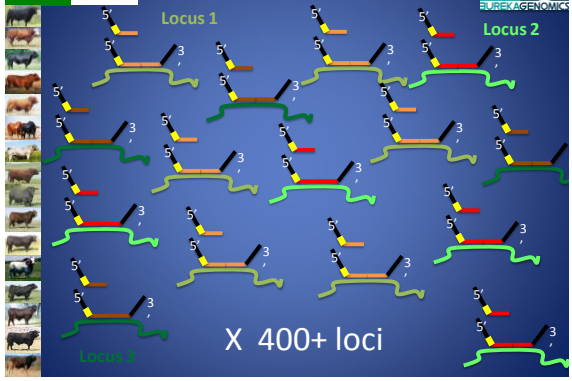
[3]Intrepid Bioinformatics, Louisville, KY

Next Generation Genotyping (**NGG**)

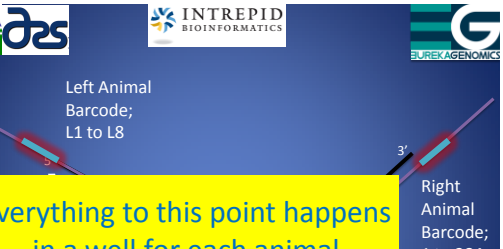
- Low-cost, high-volume genotyping based on Next Generation Sequencing
- Genotyping is focused on specified loci
- Available commercially from our Cooperative Research and Development Agreement (**CRADA**) partner, Eureka Genomics



A vs G Allele Discrimination



X 400+ loci



Everything to this point happens in a well for each animal

Use of two barcodes added by PCR primers allows for identification of a virtually unlimited number of animals

- Pool the reaction wells

- Generate sequence data

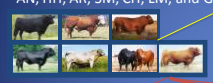
- Use bioinformatics to parse out animal, locus, and allele IDs and call genotypes

Cost per Sample of NGG is Impacted by:

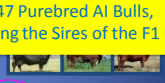
- Cost of DNA extraction and sample handling
- Volume of samples
  - Designed for high volume projects
- Number of loci that can be multiplexed in a well
- Oligo synthesis per locus
  - Most competitive when each locus used for many samples

Germplasm Evaluation Project Cycle VII

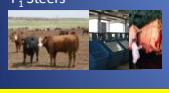
AI Sires:
AN, HH, AR, SM, CH, LM, and GV



Individual Genomic Sequence on 47 Purebred AI Bulls, Including the Sires of the F1 Bulls



F_1 Steers



F_1 Bulls

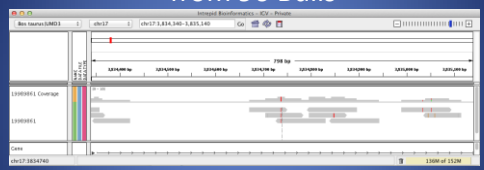
F_1 Heifers

Individual Genomic Sequence on 49 F1 Bulls that Sired the Most Progeny with 50K chip

eers & Heifers

nerations of Extensively Phenotyped Animals

Discovery of Severe Variants from 96 Bulls



545	SPLICE SITE ACCEPTOR
444	SPLICE SITE DONOR
27	START LOST
1	EXON DELETED
1101	FRAME SHIFT
294	STOP GAINED
20	STOP LOST
0	RARE AMINO ACID

Screening Severe Variants

- Variants were ranked based on effects of flanking 770K SNP imputed genotypes on Birth Weight of > 5,000 cattle in the Germplasm Evaluation Population
- Top 222 variants were selected for assay design

40	FRAME SHIFT
26	SPLICE SITE ACCEPTOR
17	SPLICE SITE DONOR
17	STOP GAINED
122	NON-CODING RNA WITH POTENTIAL REGULATORY EFFECTS
222	**TOTAL**

Severe Variant Assay Development

- Eureka developed NGG assays for 185 of the variants and genotyped several external resource populations with them.
  - Various subsets of these assays were used in the analyses of different populations based on call rates and genotype variability in order to assess their predictive ability.

Assessment of Severe Variants

- The variants for which assays were designed were not validated as biologically real until the assessment populations were genotyped, so some were monomorphic and some were all invariant heterozygotes.
- The marker effects were estimated within the assessment populations (in GenSel analyses), so this was not truly independent validation.
  - Therefore, the severe variants were compared with a similar number of parentage markers to evaluate discovery bias.

Assessment of Severe Variants

- Based on deregressed birth weight EPDs of 1019 bulls from a collaborating large commercial ranch genotyped using NGG assays.

# of Variants	Type of Variants	% of Genetic Var.	Increase in Genetic Var.
40	severe	21	12
40	parentage SNP	9	N/A
160	severe	41	29
83	parentage SNP	12	N/A

Assessment of Severe Variants

- Based on deregressed EPDs of 1169 AI Sires.

Breed	# Bulls	# Severe Variants	% Gen. Variance	# Prntg SNP	% Gen. Variance	Increase Gen. Var.
AN	323	117	11	114	30	-18
AR	134	117	38	114	27	11
CH	99	117	62	114	81	-19
GV	108	117	52	114	17	34
HH	246	117	33	114	33	0
LM	98	117	47	114	57	-10
SM	161	117	35	114	62	-27

Conclusions

- Assaying severe variant polymorphisms directly shows promise as a means of developing marker panels that account for greater genetic variance.
- Performing individual sequencing on many influential ancestors of intensively phenotyped and genotyped resource populations has advantages for screening polymorphisms for preliminary evidence of effect prior to investing in assay development for validation.

Future Plans

- Continue sequencing, focusing on individuals and coverage depths that provide the most complete imputation of sequence data across our resource population of >5,000 intensively phenotyped animals.
- Continue to develop NGG assays to assess severe variants in external populations

Questions?

- Special thanks to Steve Kachman, Univ. of Neb. for estimating the proportion of variation accounted for in the commercial ranch data
- For more information on Next Generation Genotyping:
  - See Posters P0525 and P0090.
  - Stop by Eureka Genomics booth #206
  - See www.EurekaGenomics.com