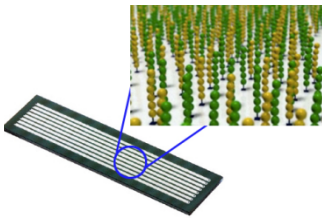
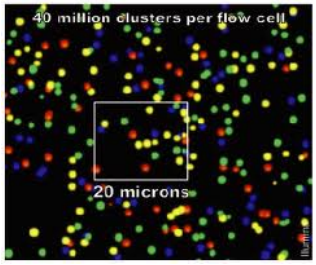
Reduced Cost Genotyping Strategies

by Mark Thallman and Heather Koshinsky

June27, 2012

R. Mark Thallman, PhD
Research Geneticist
U.S. Meat Animal Research Center
P.O. Box 166, Clay Center, NE 68933
Ph: (402) 762-4261
Email: Mark.Thallman@ars.usda.gov





Mass Genotyping by Sequencing Technology (**MGST**) is a method of low-cost, high-volume DNA testing based on recent developments in Next Generation Sequencing (**NGS**). With NGS, it is possible to obtain the DNA sequence of about 30,000 to 240,000 short fragments of DNA for about $1 of sequencing cost. However, it is necessary to run about $1,000 worth of sequencing in each batch to operate the instrument efficiently. Therefore, the enormous amount of data produced by NGS can only be harnessed effectively if many samples need to be genotyped at the same time.

The MGST technology was developed collaboratively by the U.S. Meat Animal Research Center and Eureka Genomics through a Cooperative Research and Development Agreement (**CRADA**). Eureka Genomics is currently commercializing the technology directly and exploring potential to make the technology available through other providers of DNA testing services.

In MGST, DNA from each animal to be tested is labeled with a “DNA barcode” identifying the animal of origin. Then, samples from many different animals are pooled together and run on the NGS machine. Each fragment read by the machine indicates which animal the fragment is from and one copy of the marker possessed by that animal (for one position in the genome). The position in the genome is also indicated by the sequence of the fragment. Both copies of each marker must be read from different fragments in order to determine the “genotype” of the animal. Furthermore, there is substantial variation in the number of fragments read per marker per animal. Therefore, substantially more than two fragments per marker per animal must be read in order to obtain the desired accuracy and call rate.

The chemistry used in MGST allows flexibility in specifying the markers (locations in the genome) to be assayed. Markers can be added to or deleted from an assay much more easily than with the “chip-based” platforms that are currently popular. Generally, between 1,000 and 2,000 animals are genotyped on one “lane” of the sequencing instrument. However, the NGS instrument most commonly used for MGST has eight lanes that must be run in the same way at the same time. A newer NGS machine that is less expensive per read has 16 lanes, each with higher capacity per lane, but it places additional constraints on the runs that can be performed. Therefore, the cost of MGST genotyping (or quantity of markers genotyped per animal) has considerable room for improvement if the volume of testing by MGST increases substantially.

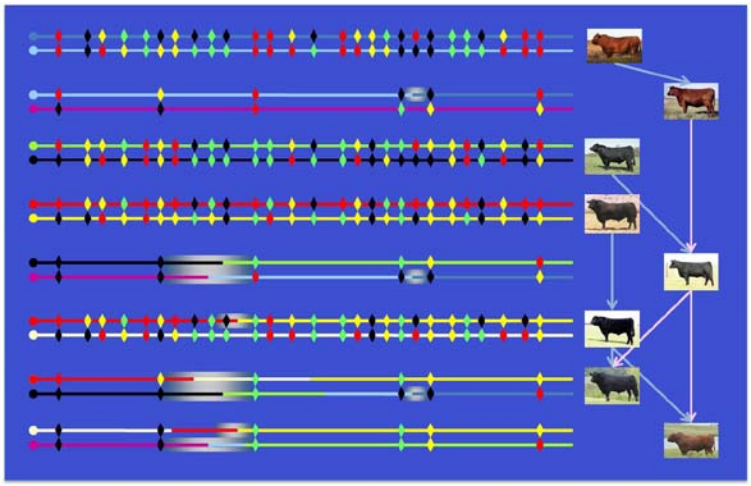
The MGST was validated on 1080 animals genotyped with the Illumina BovineSNP50 Beadchip (**50K**) for 95 single nucleotide polymorphisms (**SNP**) out of Mike Heaton's 112 SNP parentage panel. All genotypes at these SNP for which the 50K chip gave a call were used. The genotype calling algorithm was tuned for the parentage application by emphasizing accuracy over call rate. The results are presented in the tables. The error rate includes any errors in the 50K chip genotypes that were used as the standard.

Rank	Marker	Con-cord-ance	Call Rate
1	AY842472	1	0.99
2	DQ381153	1	0.98
3	DQ786758	1	0.99
4	DQ995976	1	0.99
5	DQ786766	1	0.99
6	AY842474	1	0.99
90	DQ647187	0.97	0.96
91	AY851162	0.97	0.91
92	DQ916057	0.97	0.87
93	DQ647188	0.96	0.82
94	DQ984827	0.96	0.89
95	DQ839235	0.95	0.88
Total		**0.991**	**0.961**

Animal Rank	Con-cord-ance	Call Rate
1 - 100	1.000	0.969
101 - 200	1.000	0.977
201 - 300	1.000	0.977
301 - 400	1.000	0.973
401 - 500	1.000	0.98
501 - 600	1.000	0.98
601 - 700	0.999	0.979
701 - 800	0.989	0.984
801 - 900	0.989	0.962
901 - 950	0.988	0.939
951 - 1000	0.978	0.966
1001 - 1050	0.970	0.928
1051 - 1060	0.951	0.915
1061 - 1070	0.652	0.694
1071 - 1080	0.181	0.304
1 - 1080	**0.991**	**0.961**

Some potential applications of MGST:

- Seedstock
  - Parentage Testing
  - Paternity Determination
  - Genetic Defect Testing
  - Testing for single gene traits
    - Color, polledness, F94L, etc.
  - Prediction of Economically Relevant Traits from Small Panels of SNP
  - New Recessive Defects Discovered by Individual Bull Sequencing
- Feedlot
  - Marker Assisted Management from Small Panels of SNP
- Commercial Ranch Project



Sparse genome scan genotyping is an idea that has been proposed to provide an inexpensive replacement to most other DNA testing that might be done on seedstock cattle (e.g., each of the items listed above under seedstock). Roughly 1,000-5,000 markers spread evenly across the genome would be genotyped to allow tracking inheritance throughout the genome of a pedigreed population as illustrated in the Figure above. But, to be effective, it needs to be applied to a whole herd, or better yet, to an entire breed. Therefore, it will require very inexpensive genotyping of a sufficient number of markers. It is not currently available at a price similar to the smaller panels, but there is substantial effort to make that happen soon.

Acknowledgements: Amanda Lindholm-Perry, Linda Flathman, and John Pollak (USMARC). Heather Koshinsky, John Curry, Paul Dier, Maria Shin, Jessica Nguyen, Viacheslav Fofanov, Nadeem Bulsara, and Jingtao Liu (Eureka Genomics).
For information from Eureka, contact Didier Perez (didier@eurekagenomics.com)