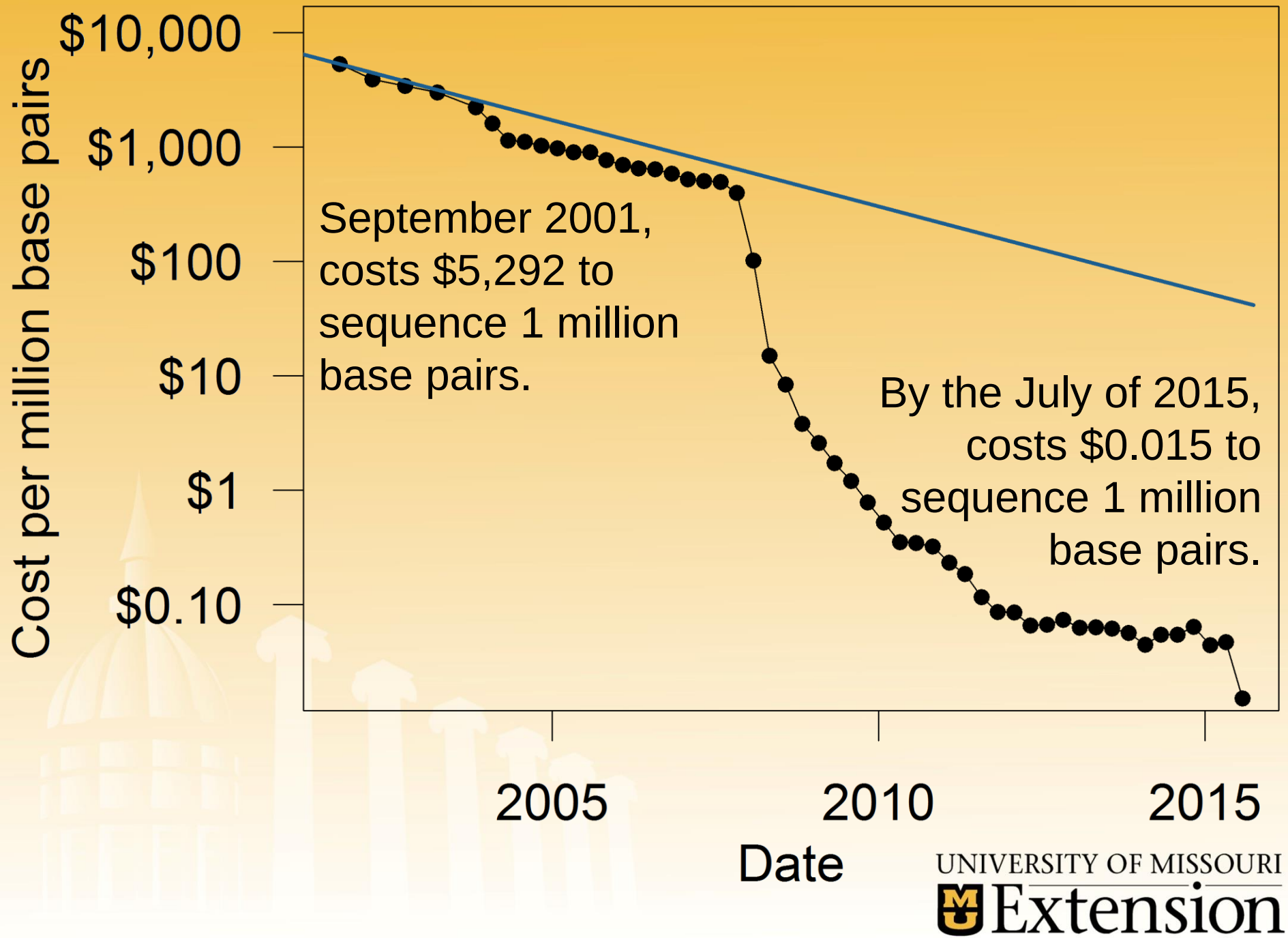
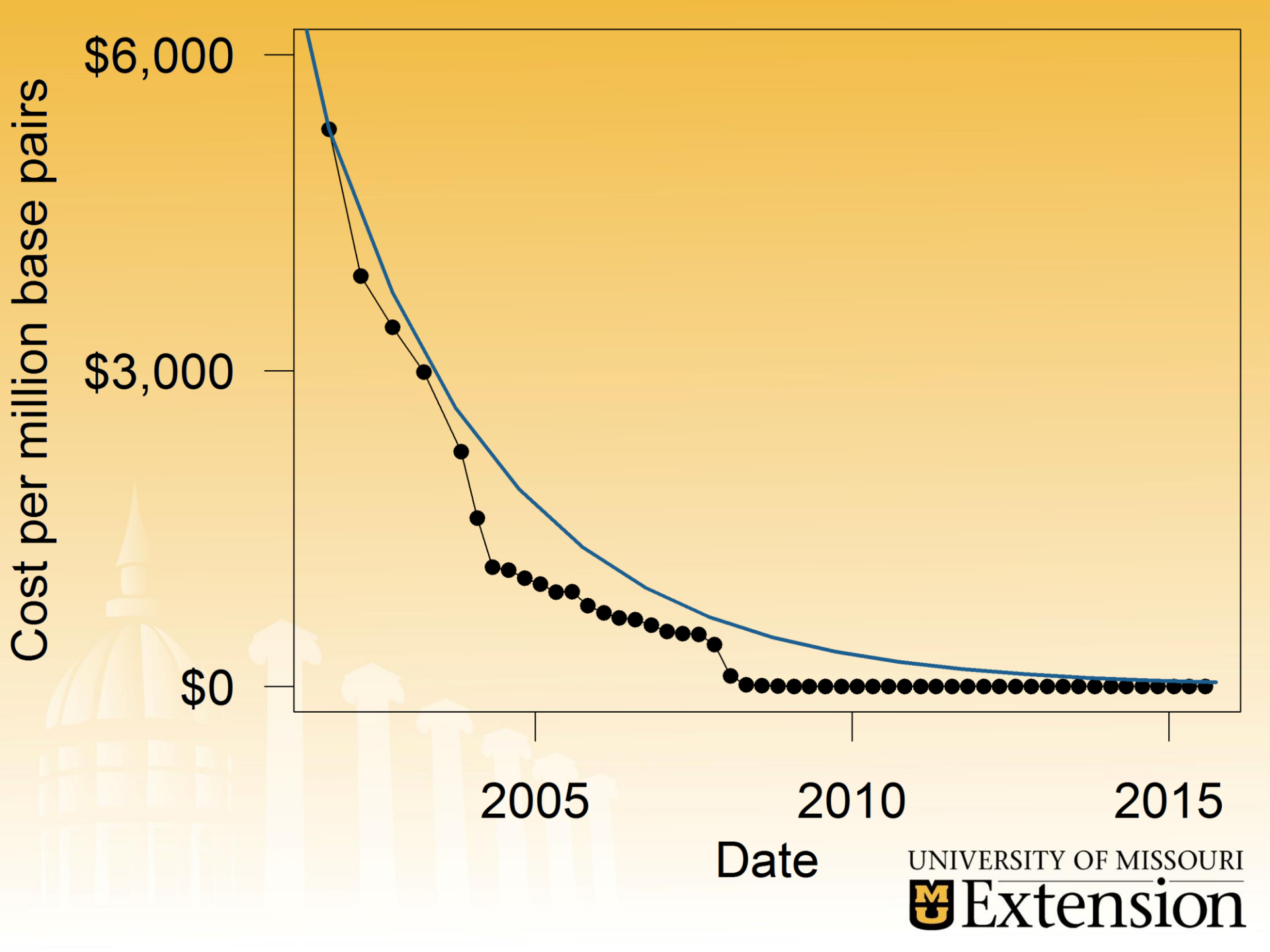
What have we learned from sequencing efforts to date?

Jared Decker
Assistant Professor
Beef Genetics Specialist
Computational Genomics

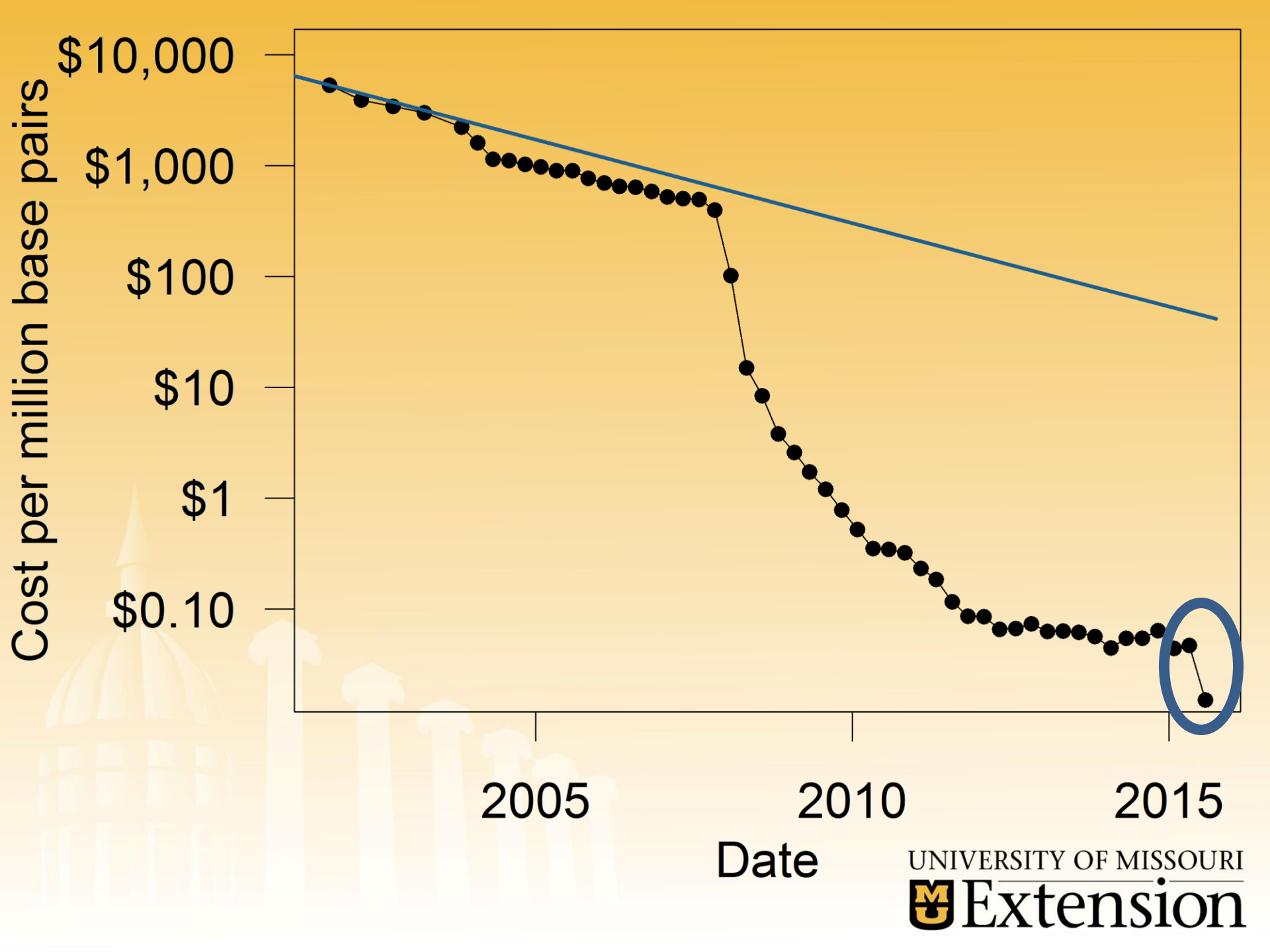
UNIVERSITY OF MISSOURI
MU Extension



UNIVERSITY OF MISSOURI Extension



UNIVERSITY OF MISSOURI Extension

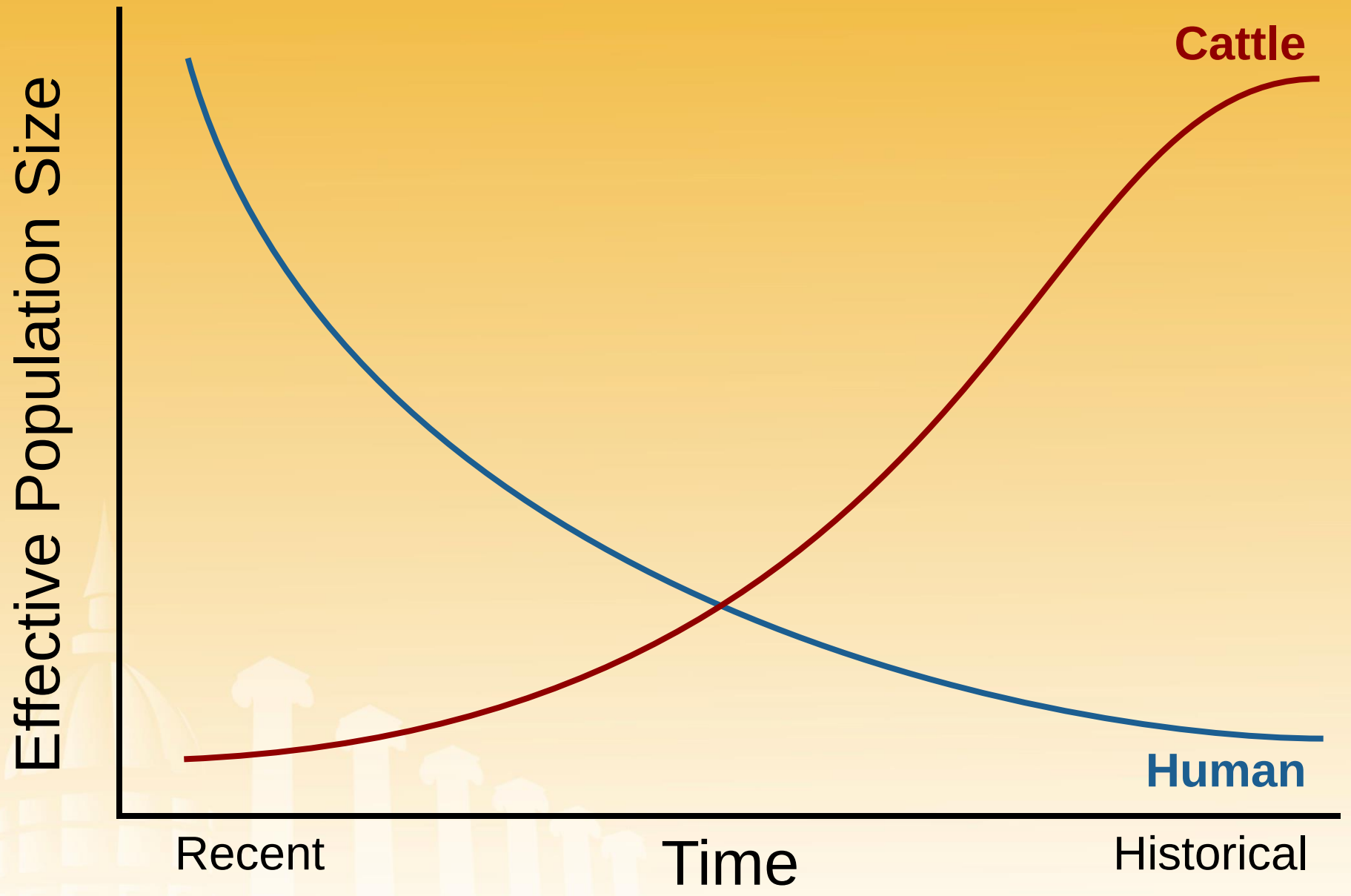


UNIVERSITY OF MISSOURI
Extension

Illumina HiSeq X Ten

UNIVERSITY OF MISSOURI
Extension

UNIVERSITY OF MISSOURI
Extension



UNIVERSITY OF MISSOURI
Extension

Where are we going next?

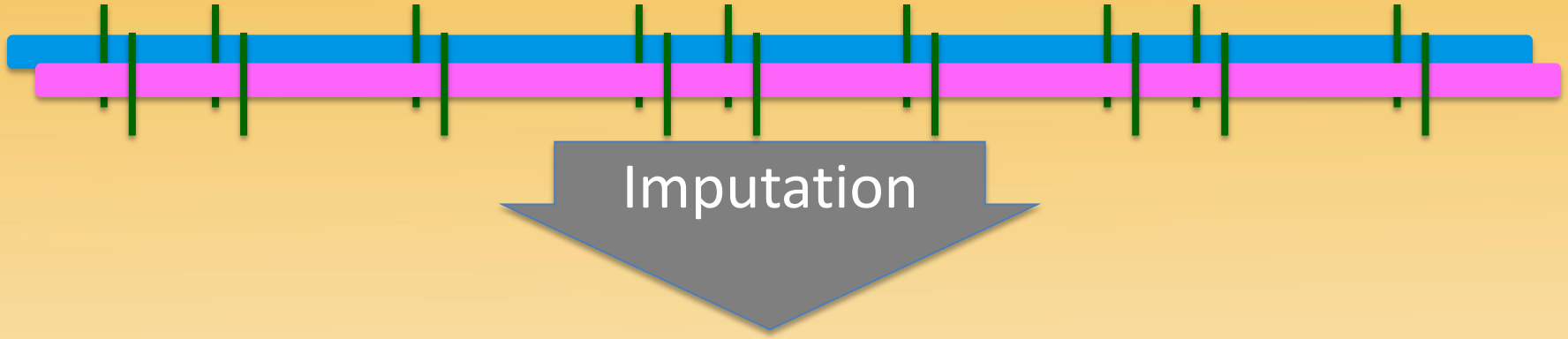
UNIVERSITY OF MISSOURI
Extension

Where are we going next?

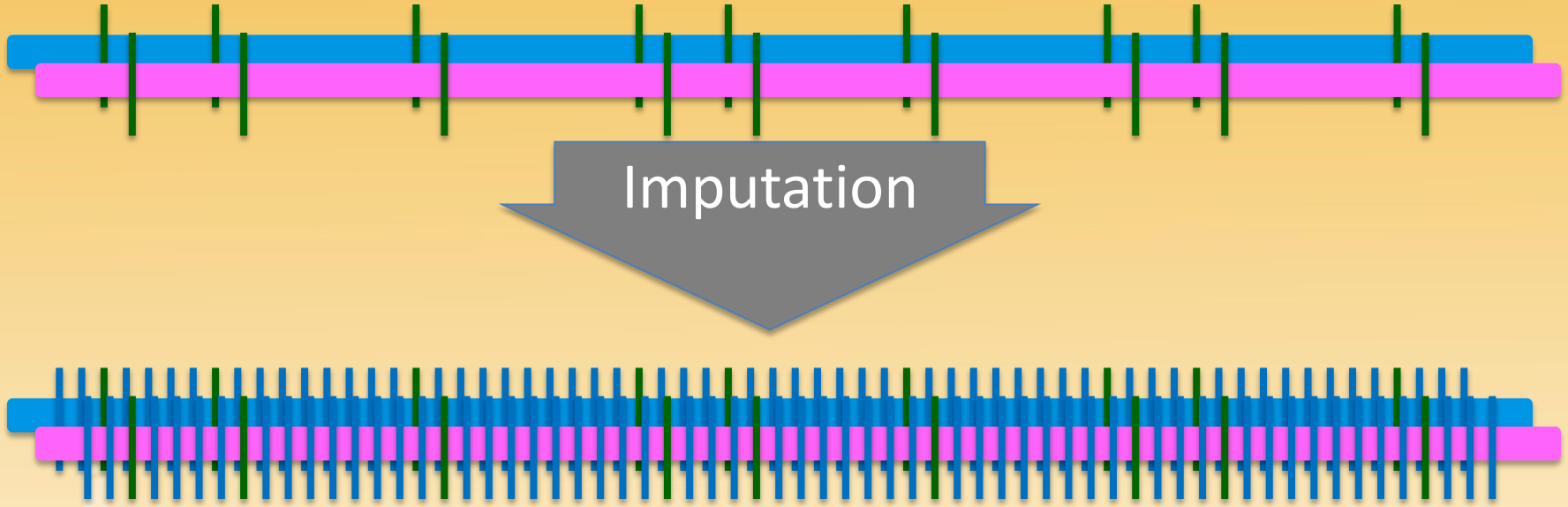
- Imputation


MILE
106

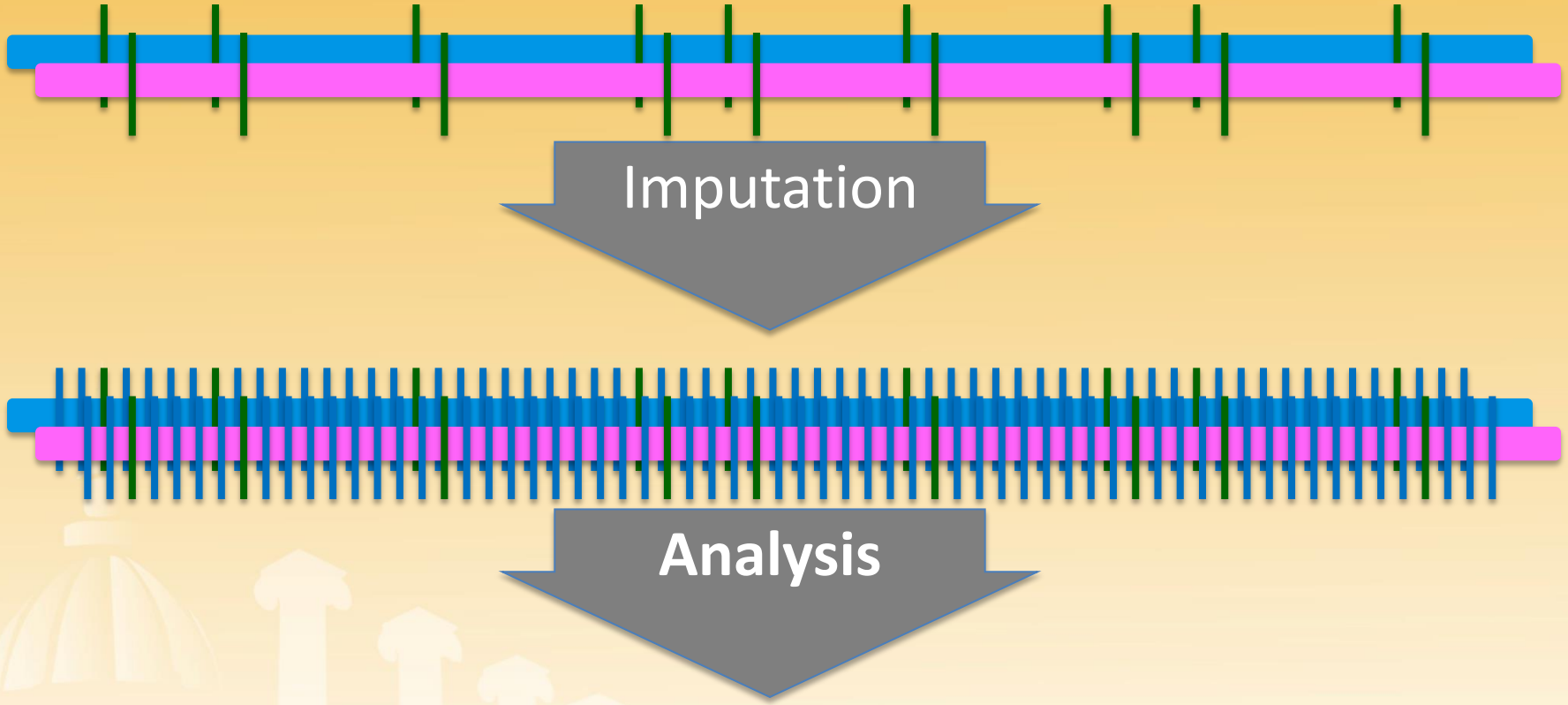
Where are we going next?



Where are we going next?

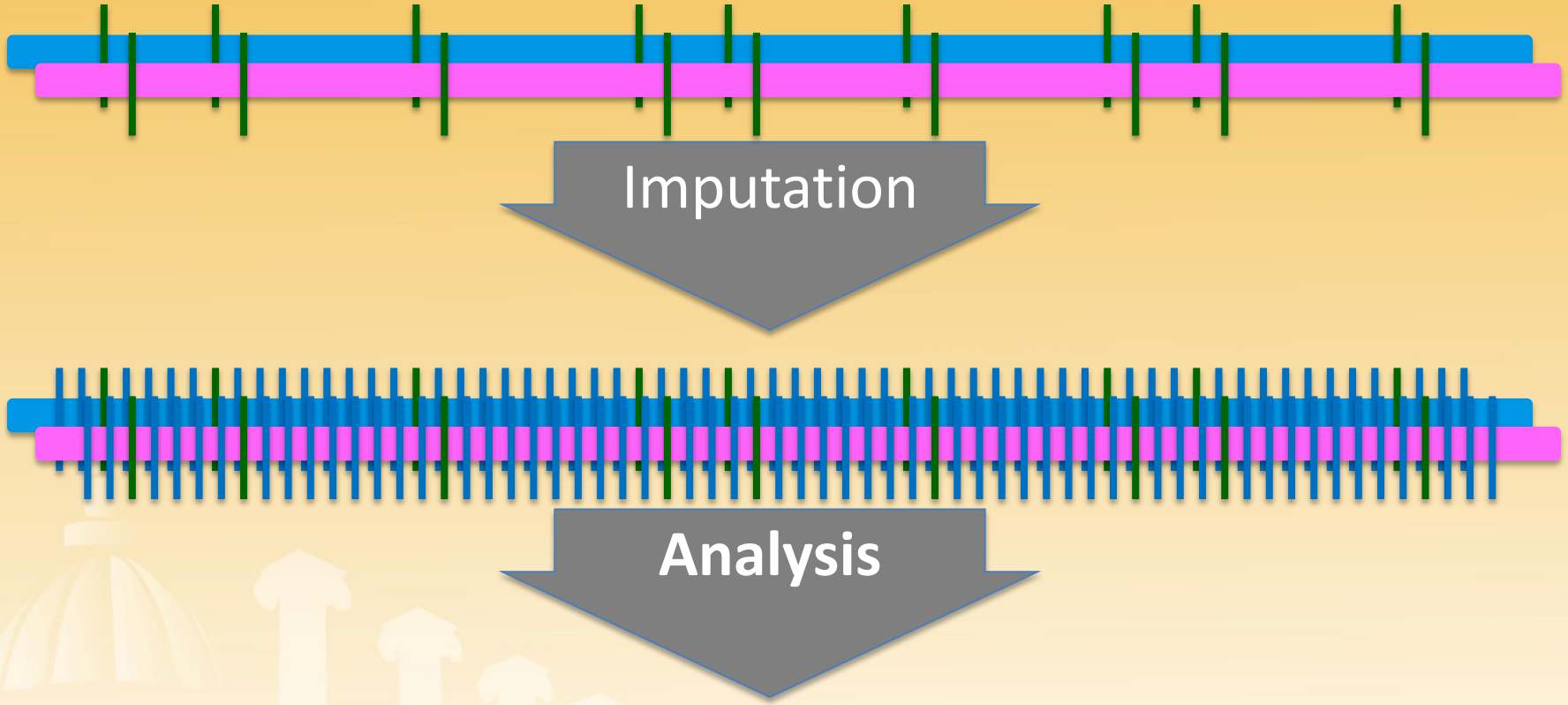


Where are we going next?



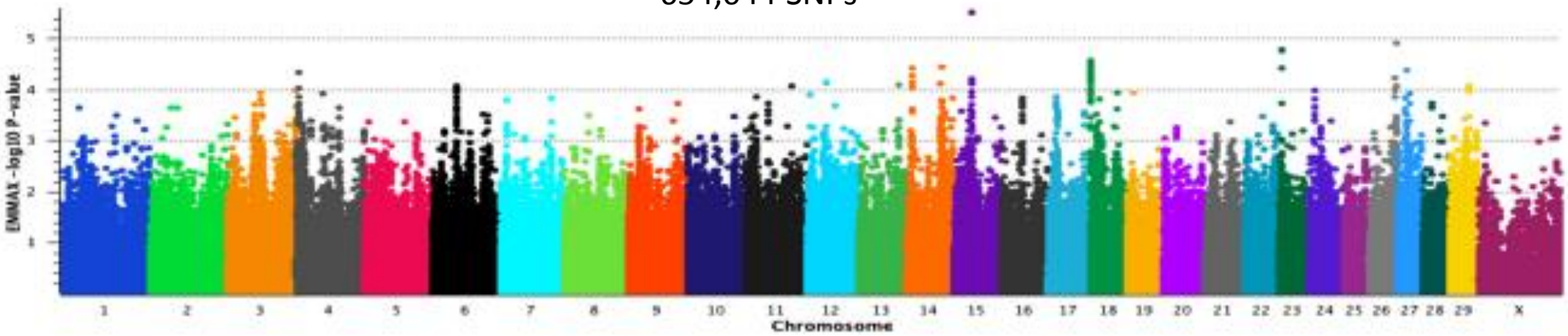
UNIVERSITY OF MISSOURI
Extension

Where are we going next?

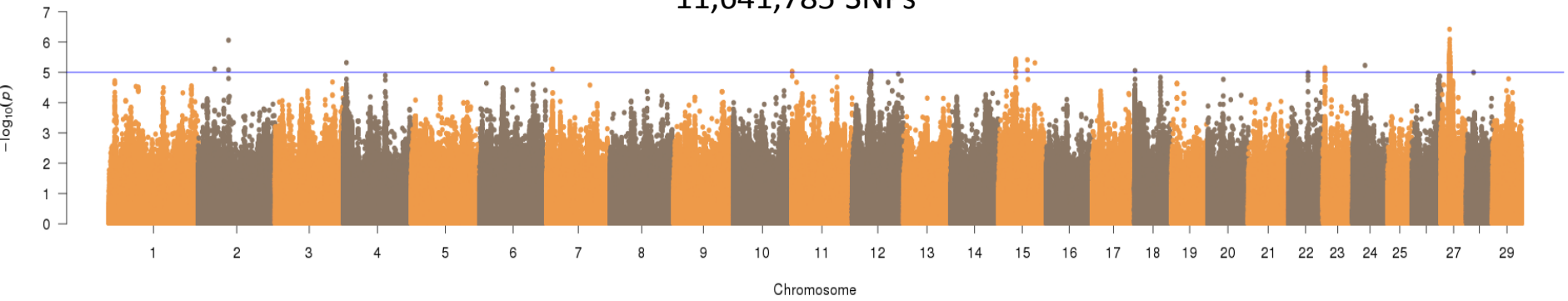


UNIVERSITY OF MISSOURI
Extension

654,044 SNPs



11,641,785 SNPs



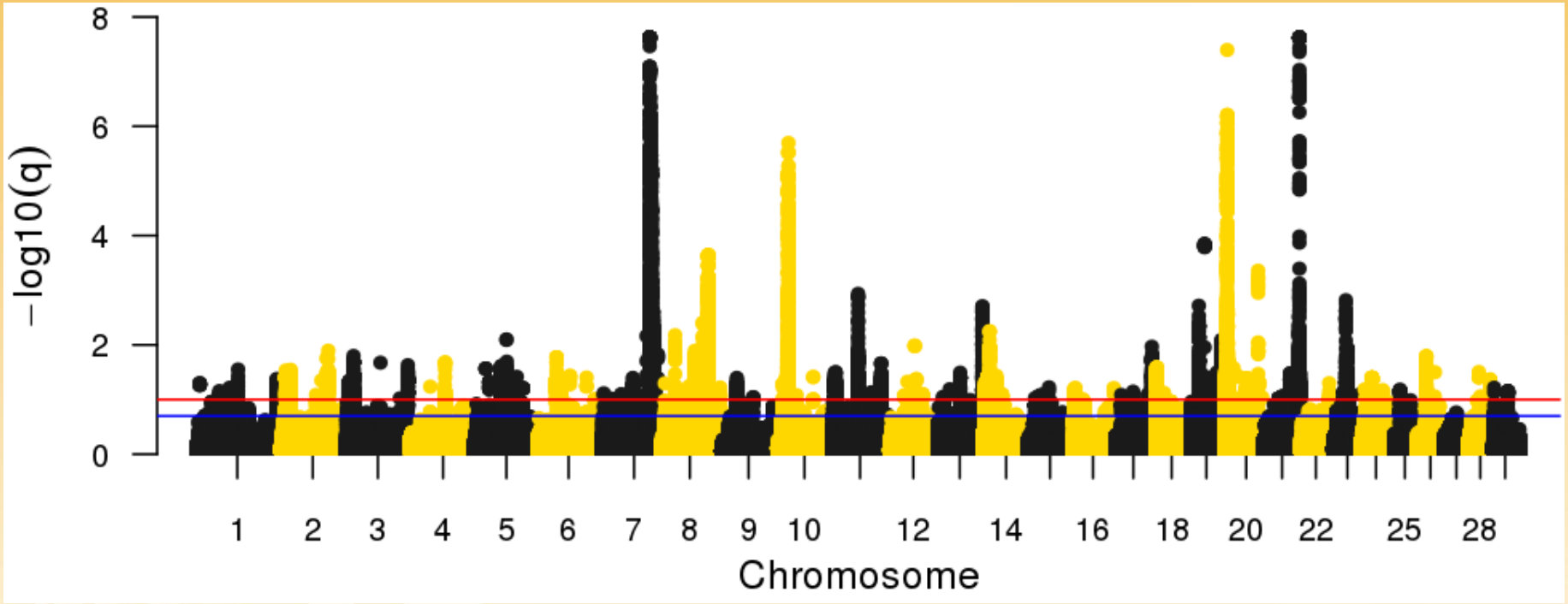
- h^2 increases from 21% to 23%
- A lot more significant associations – new variants are discovered!
- Largest effect QTL changes

GWAS on HD *vs* Imputed Sequence
1978 Holstein calves – BRD Case/Control
Taylor Lab

UNIVERSITY OF MISSOURI Extension

GWAS on imputed data works nicely.

Imputation with multiple breed reference set also works.



Mature Height

- 3,362 Animals
- FDR = 0.1 (——)

UNIVERSITY OF MISSOURI Extension

Where are we going next?

- Back to causal variants
- Led to genes by data not researchers opinion
- Continue to take very large numbers of genes and genomic regions into account

UNIVERSITY OF MISSOURI Extension

Causal Variant Successes

- **Genetic defects**
- Embryonic lethals
- Horned/Polled
- Coat Color

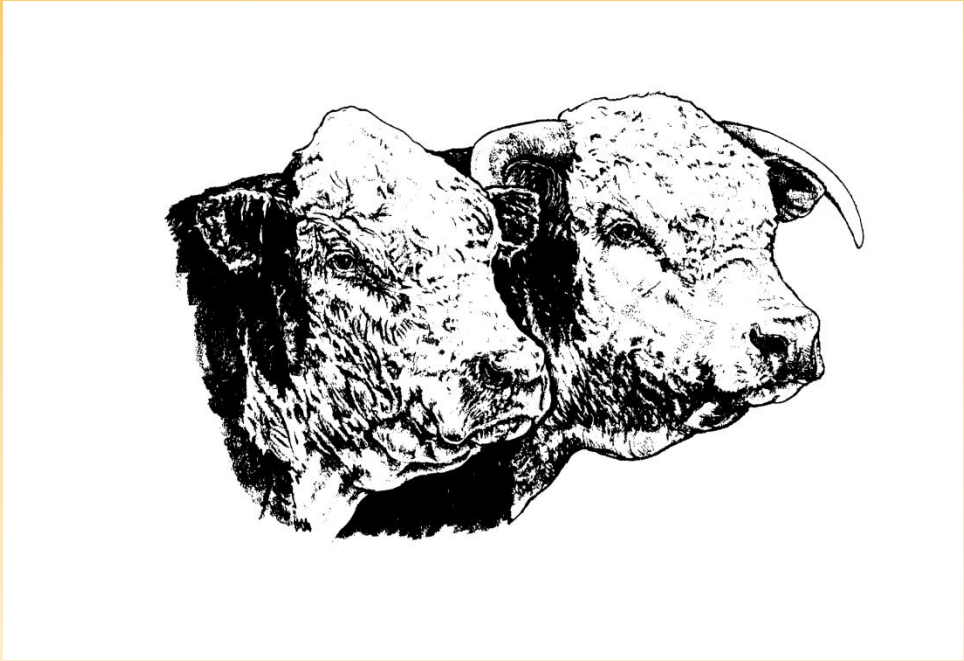
Causal Variant Successes

- Genetic defects
- **Embryonic lethals**
- Horned/Polled
- Coat Color

UNIVERSITY OF MISSOURI Extension

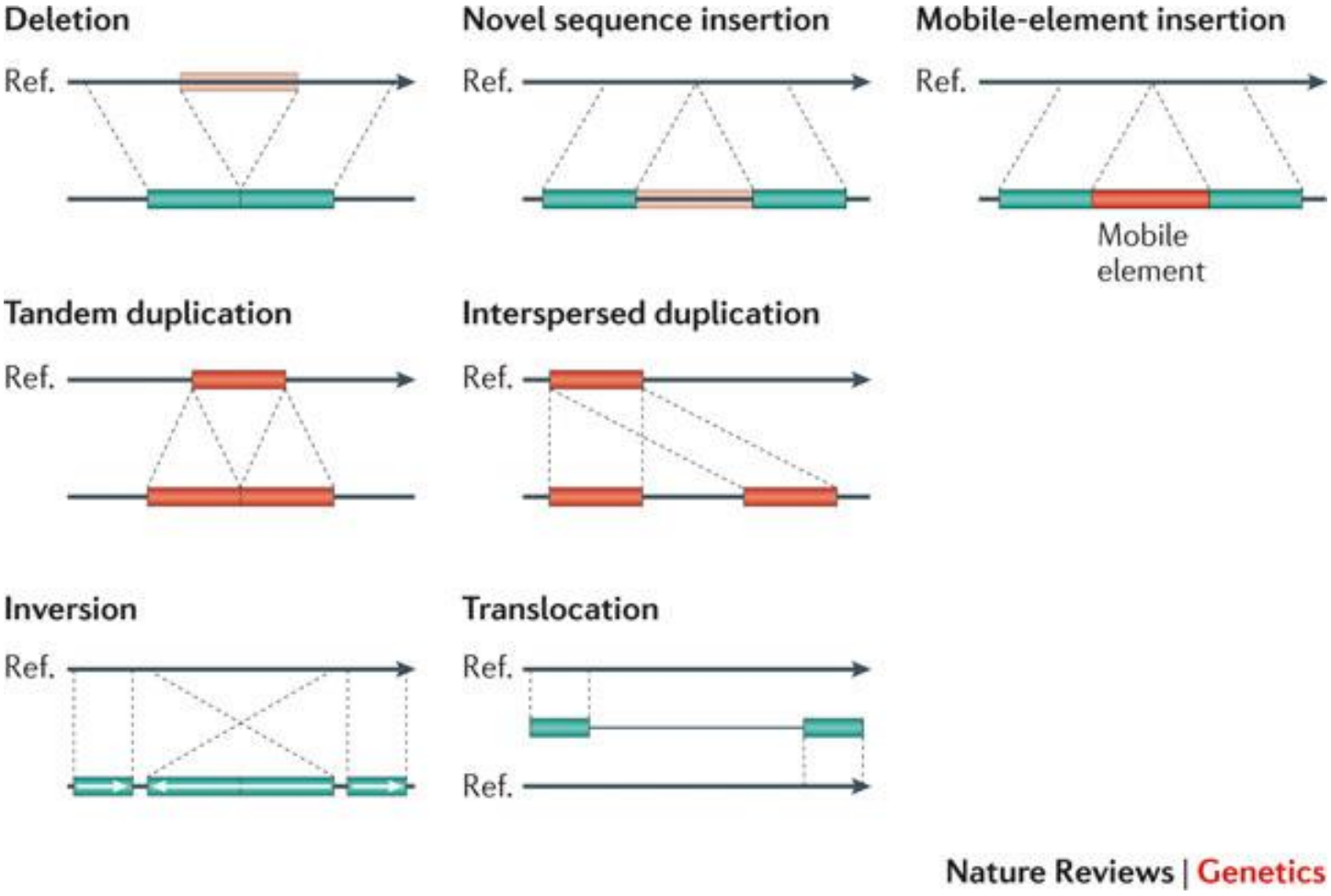
Causal Variant Successes

- Genetic defects
- Embryonic lethals
- **Horned/Polled**
- **Coat Color**



UNIVERSITY OF MISSOURI
MU Extension

Causal Variants
This is not easy!



Genome structural variation discovery and genotyping
Can Alkan, Bradley P. Coe & Evan E. Eichler
Nature Reviews Genetics **12**, 363-376 (May 2011)
doi:10.1038/nrg2958

Description	Number of Variants
Amino Acid Genomic	106,634
Untranslated region Genomic	137,581
Splice Genomic	26,200
No homozygote Genomic	1,016,064
Amino Acid RNA	188,429
Untranslated region RNA	50,061
Splice RNA	279,906
No homozygote RNA	158,999
Amino Acid 1kBulls	32,756
Untranslated region 1kBulls	19,407
Splice 1kBulls	6,820
No homozygote 1kBulls	6,271
	2,029,128

Where are we going next?

- More types of SNP chips/panels
- Higher density panels focusing on functional variants for research
  - GGP-F250
- Lower density panels focusing on functional variants for prediction
- Functional variants -> variants that influence protein sequence or abundance

Description	Number Variants
Imputation content	33,730
Functional content	193,503
Total variants	227,233
from HD as part of imputation content	31,835
from SNP50 as part of imputation content (subset of HD)	22,183
from HD as part of functional content	6,395
from SNP50 as part of functional content (subset of HD)	443
Total HD	38,230
Total SNP50 (subset of HD)	22,626
NS Sift deleterious	22,298
NS Sift tolerated	48,994
NS no sift prediction	49,627
Total Non-synonymous (NS) AA substitutions	120,919
Frameshift indels	1,265
In-Frame indels	585
UTR	20,402
Ensembl ncRNA (snoRNA, miRNA, snRNA, rRNA,Mt_tRNA, Mt_rRNA)	1,573
Conserved non-coding elements	4,081
Splice (not mutually exclusive)	6,378
Genes with at least 1 variant	23,059
Genes with no variant	7,714

Description	Number Variants
Imputation content	**33,730**
Functional content	193,503
Total variants	227,233
from HD as part of imputation content	31,835
from SNP50 as part of imputation content (subset of HD)	22,183
from HD as part of functional content	6,395
from SNP50 as part of functional content (subset of HD)	443
Total HD	38,230
Total SNP50 (subset of HD)	22,626
NS Sift deleterious	22,298
NS Sift tolerated	48,994
NS no sift prediction	49,627
Total Non-synonymous (NS) AA substitutions	120,919
Frameshift indels	1,265
In-Frame indels	585
UTR	20,402
Ensembl ncRNA (snoRNA, miRNA, snRNA, rRNA,Mt_tRNA, Mt_rRNA)	1,573
Conserved non-coding elements	4,081
Splice (not mutually exclusive)	6,378
Genes with at least 1 variant	23,059
Genes with no variant	7,714

Description	Number Variants
Imputation content	33,730
Functional content	**193,503**
Total variants	227,233
from HD as part of imputation content	31,835
from SNP50 as part of imputation content (subset of HD)	22,183
from HD as part of functional content	6,395
from SNP50 as part of functional content (subset of HD)	443
Total HD	38,230
Total SNP50 (subset of HD)	22,626
NS Sift deleterious	22,298
NS Sift tolerated	48,994
NS no sift prediction	49,627
Total Non-synonymous (NS) AA substitutions	120,919
Frameshift indels	1,265
In-Frame indels	585
UTR	20,402
Ensembl ncRNA (snoRNA, miRNA, snRNA, rRNA,Mt_tRNA, Mt_rRNA)	1,573
Conserved non-coding elements	4,081
Splice (not mutually exclusive)	6,378
Genes with at least 1 variant	23,059
Genes with no variant	7,714

Description	Number Variants
Imputation content	33,730
Functional content	193,503
Total variants	227,233
from HD as part of imputation content	31,835
from SNP50 as part of imputation content (subset of HD)	22,183
from HD as part of functional content	6,395
from SNP50 as part of functional content (subset of HD)	443
Total HD	38,230
Total SNP50 (subset of HD)	22,626
NS Sift deleterious	22,298
NS Sift tolerated	48,994
NS no sift prediction	49,627
Total Non-synonymous (NS) AA substitutions	**120,919**
Frameshift indels	1,265
In-Frame indels	585
UTR	20,402
Ensembl ncRNA (snoRNA, miRNA, snRNA, rRNA,Mt_tRNA, Mt_rRNA)	1,573
Conserved non-coding elements	4,081
Splice (not mutually exclusive)	6,378
Genes with at least 1 variant	23,059
Genes with no variant	7,714

Description	Number Variants
Imputation content	33,730
Functional content	193,503
Total variants	227,233
from HD as part of imputation content	31,835
from SNP50 as part of imputation content (subset of HD)	22,183
from HD as part of functional content	6,395
from SNP50 as part of functional content (subset of HD)	443
Total HD	38,230
Total SNP50 (subset of HD)	22,626
NS Sift deleterious	22,298
NS Sift tolerated	48,994
NS no sift prediction	49,627
Total Non-synonymous (NS) AA substitutions	120,919
Frameshift indels	1,265
In-Frame indels	585
UTR	20,402
Ensembl ncRNA (snoRNA, miRNA, snRNA, rRNA,Mt_tRNA, Mt_rRNA)	1,573
Conserved non-coding elements	**4,081**
Splice (not mutually exclusive)	6,378
Genes with at least 1 variant	23,059
Genes with no variant	7,714

Description	Number Variants
Imputation content	33,730
Functional content	193,503
Total variants	227,233
from HD as part of imputation content	31,835
from SNP50 as part of imputation content (subset of HD)	22,183
from HD as part of functional content	6,395
from SNP50 as part of functional content (subset of HD)	443
Total HD	38,230
Total SNP50 (subset of HD)	22,626
NS Sift deleterious	22,298
NS Sift tolerated	48,994
NS no sift prediction	49,627
Total Non-synonymous (NS) AA substitutions	120,919
Frameshift indels	1,265
In-Frame indels	585
UTR	20,402
Ensembl ncRNA (snoRNA, miRNA, snRNA, rRNA,Mt_tRNA, Mt_rRNA)	1,573
Conserved non-coding elements	4,081
Splice (not mutually exclusive)	6,378
Genes with at least 1 variant	**23,059**
Genes with no variant	**7,714**

Where are we going next?

- Then genomic predictions would be based on causal variants or closely linked variants
- Fewer variants on prediction panels leads to lower cost?
- Hope for multiple breed predictions?
- More precise predictions based using causal/closely linked

Where are we going next?

- Sequencing lots of animals!
  - Mostly influential sires
- 1000 Bull Consortium
- Individual groups
  - Support from breed associations

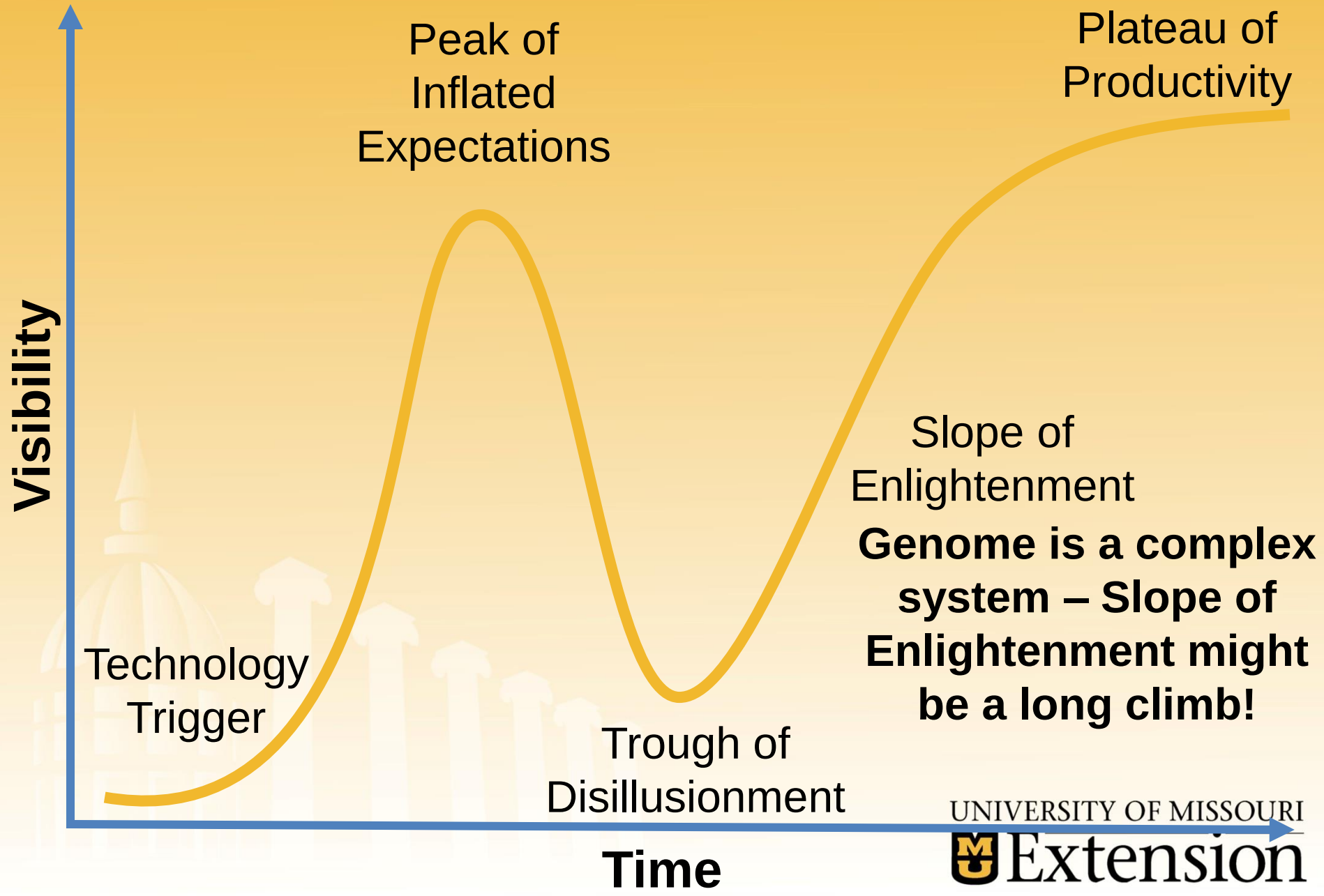
Where are we going next?

- Genomic “surveillance” of influential sires
- Breed associations sequences a sire when it reaches a certain number of progeny equivalents
  - 10,000?
  - 5% of annual registrations?
- Progeny equivalents
  - Progeny are 1 points
  - Grand progeny are 0.5 points
  - Great-grand progeny are 0.25 points

Where are we going next?

- Genomic “surveillance” of influential sires
- Early identification of possible functional variants
- Good and bad
- Variants responsible for embryonic lethals

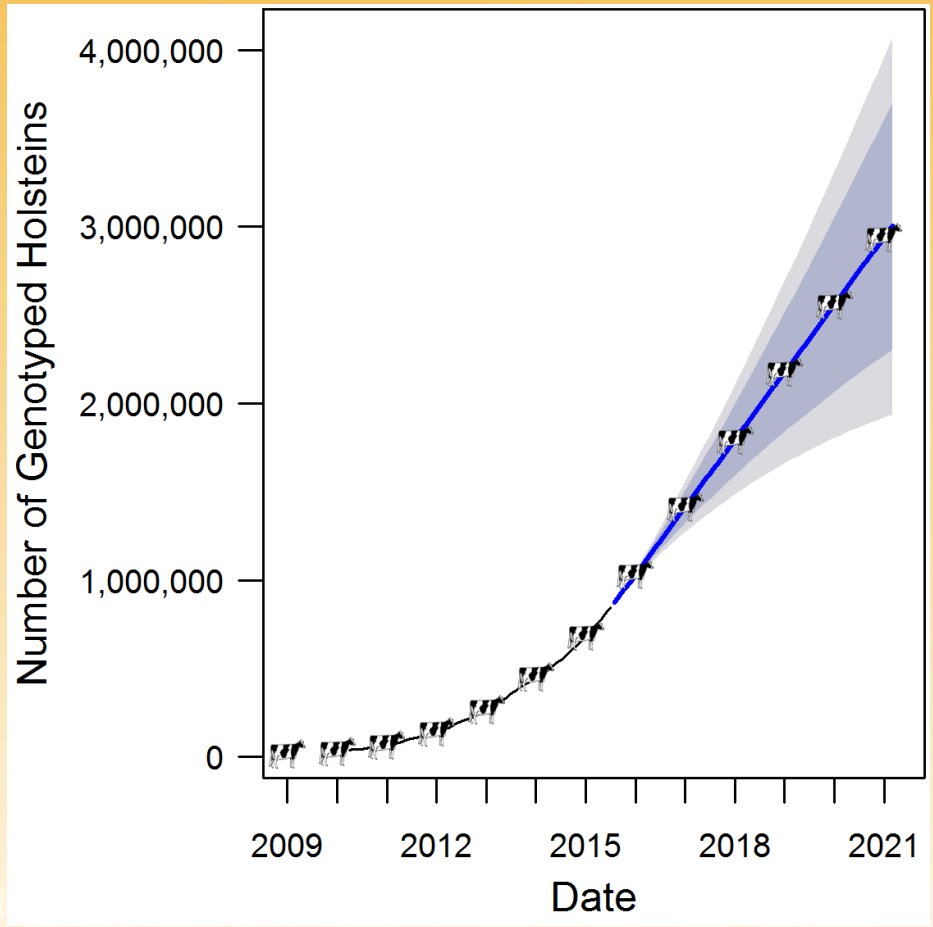
Hype Cycle: DNA in Beef Breeding



Genome is a complex system – Slope of Enlightenment might be a long climb!

UNIVERSITY OF MISSOURI Extension

Questions?



A Steak in Genomics

http://blog.steakgenomics.org

https://www.facebook.com/SteakGenomics

eBEEF.ORG

http://eBEEF.org

UNIVERSITY OF MISSOURI Extension