Better Crops with Arrays and Sequencing from Illumina

AEIC, April 2013

Chandrasen Soans, Ph.D, MBA

Product Manager, Agrigenomics

illumina®

Focus Areas for Today's Presentation

- Illumina-Our Vision
- The Toolset
- New Products and Services
- Trends and the Future

Illumina at a Glance

- Founded in 1998
- IPO July 27, 2000
- Headquarters in San Diego, CA
- More than 850,000 sq. ft.
- Facilities in 7 countries
- Over 2,500 employees
  - >1200 R&D staff
  - >400 support personnel
- IP portfolio of >235 issued patents and 168 pending applications
- Added to the NASDAQ-100 listing in 2008
- Forbes “Fastest Growing Technology Company” 2007 & 2009

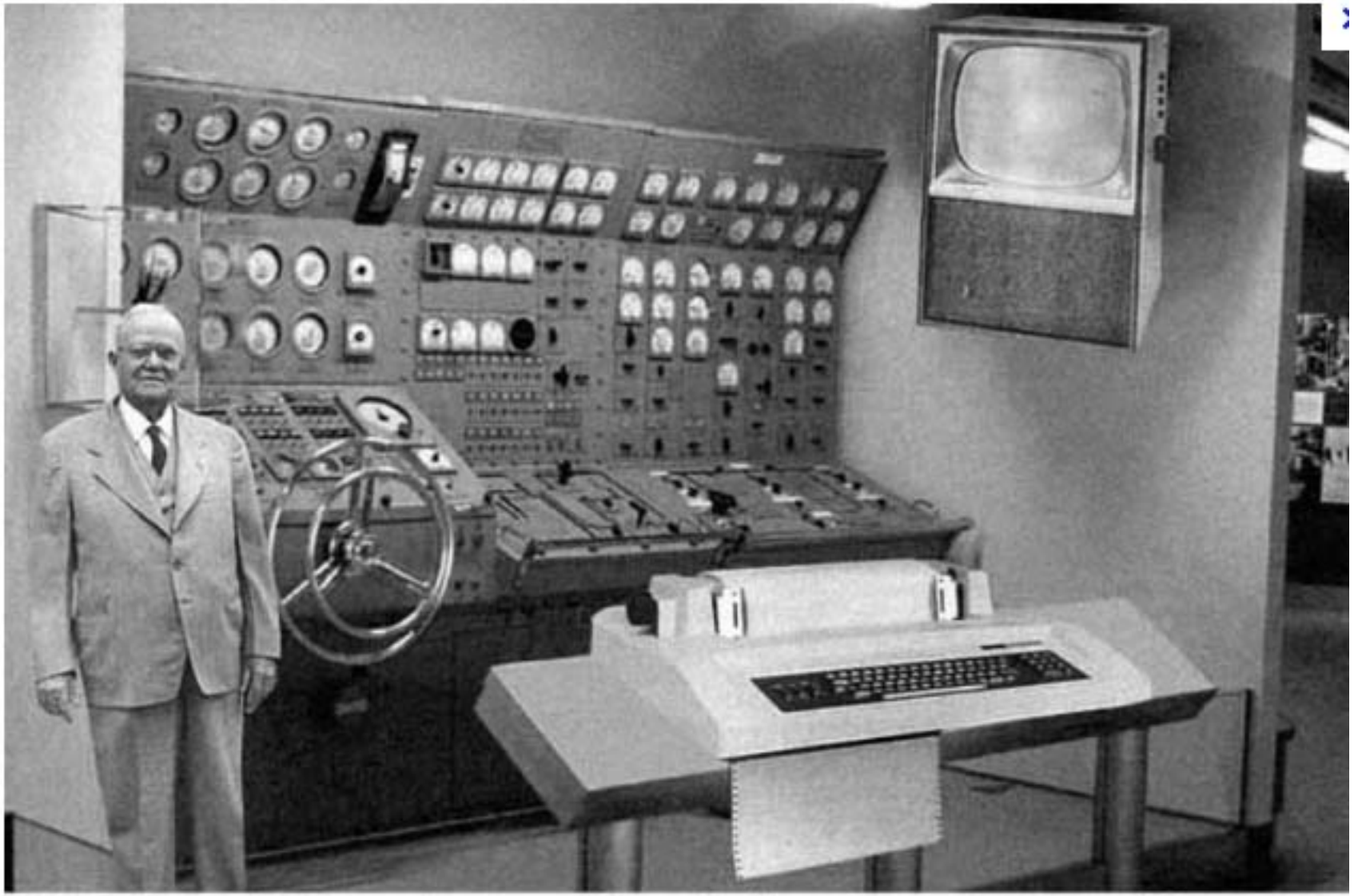
We are a Global Organization

Expanded Manufacturing, R&D, Sales, Service & Support

Over 2500 Employees
>1200 R&D staff
>400 Support Personnel

illumina®

Vision Drives Change…Reality Exceeds Vision



Scientists from the RAND Corporation have created this model to illustrate how a "home computer" could look like in the year 2004. However the needed technology will not be economically feasible for the average home. Also the scientists readily admit that the computer will require not yet invented technology to actually work, but 50 years from now scientific progress is expected to solve these problems. With teletype interface and the Fortran language, the computer will be easy to use.

UNLOCKING THE POWER OF EVERY GENOME

Accelerating genetic gain through predictive breeding

From Genome Wide Discovery…

To Targeted Validation and Beyond…

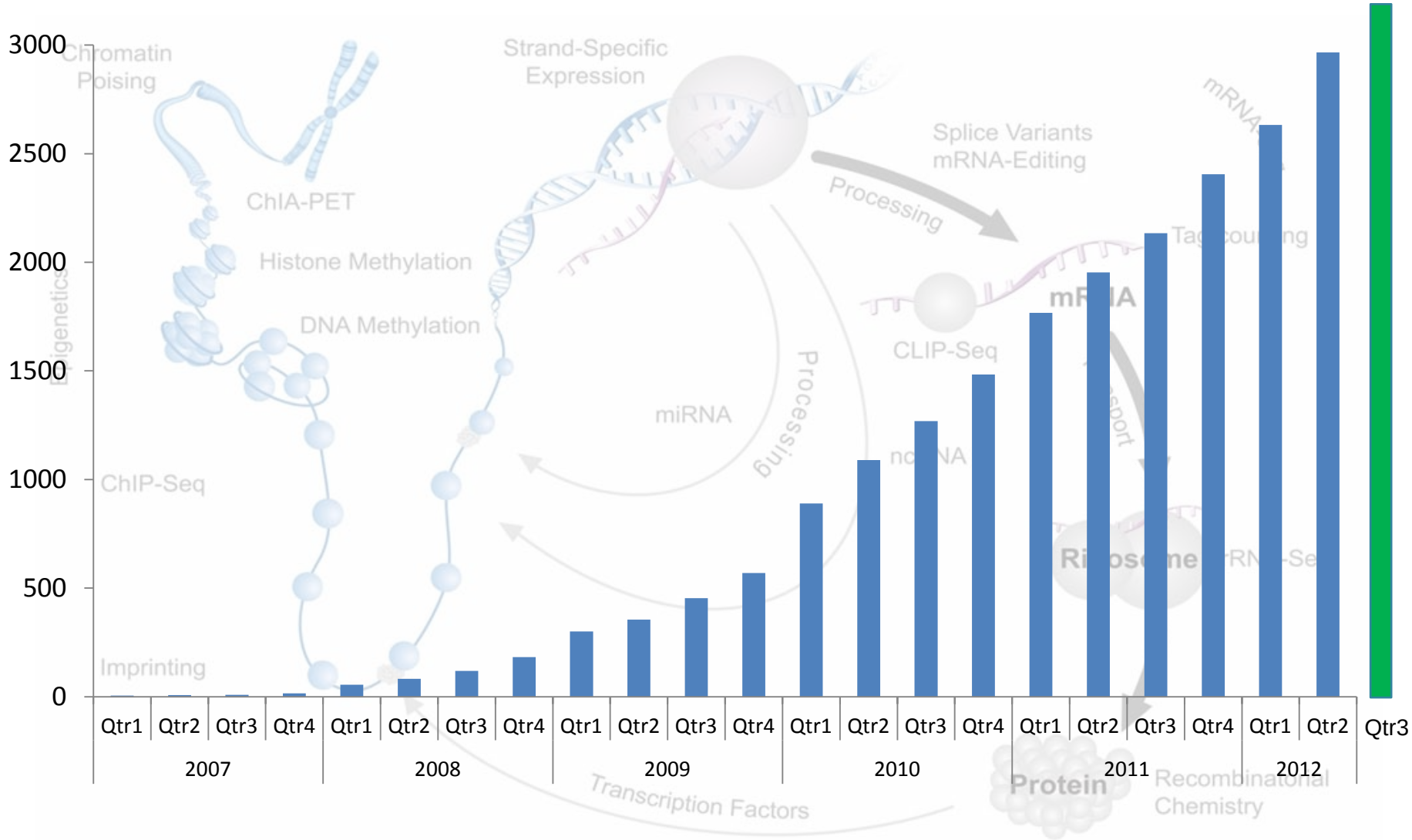
Improving human health through agriculture

Having a Tangible Impact on Reduction Of Poverty And Hunger

- Improving rice productivity in South East Asia: IRRI
- Improving goat productivity in Africa: USDA & partners
- Developing Baobab tree genomics tools: Danforth Center, Monsanto & partners
- Improving Dairy productivity in Africa: University of New England (Australia)
- Improving productivity of pidgeonpea in India & Africa: ICRISAT
- Improving productivity of cassava in Africa: BecA

We Have Seen an Unprecedented Publication Rate

3739 peer-reviewed sequencing publications as of April 10, 2013



illumina®

Enabling an Agrigenomics Revolution

Explosion Of Genomes Available At NCBI

http://www.ncbi.nlm.nih.gov/genome/browse/

Eukaryotes ***(3064)***

- Animals ***(1089)***
  Mammals, fish, insects, roundworms, birds, reptiles, other animals
- Plants ***(710)***
  Land plants, green algea, other plants

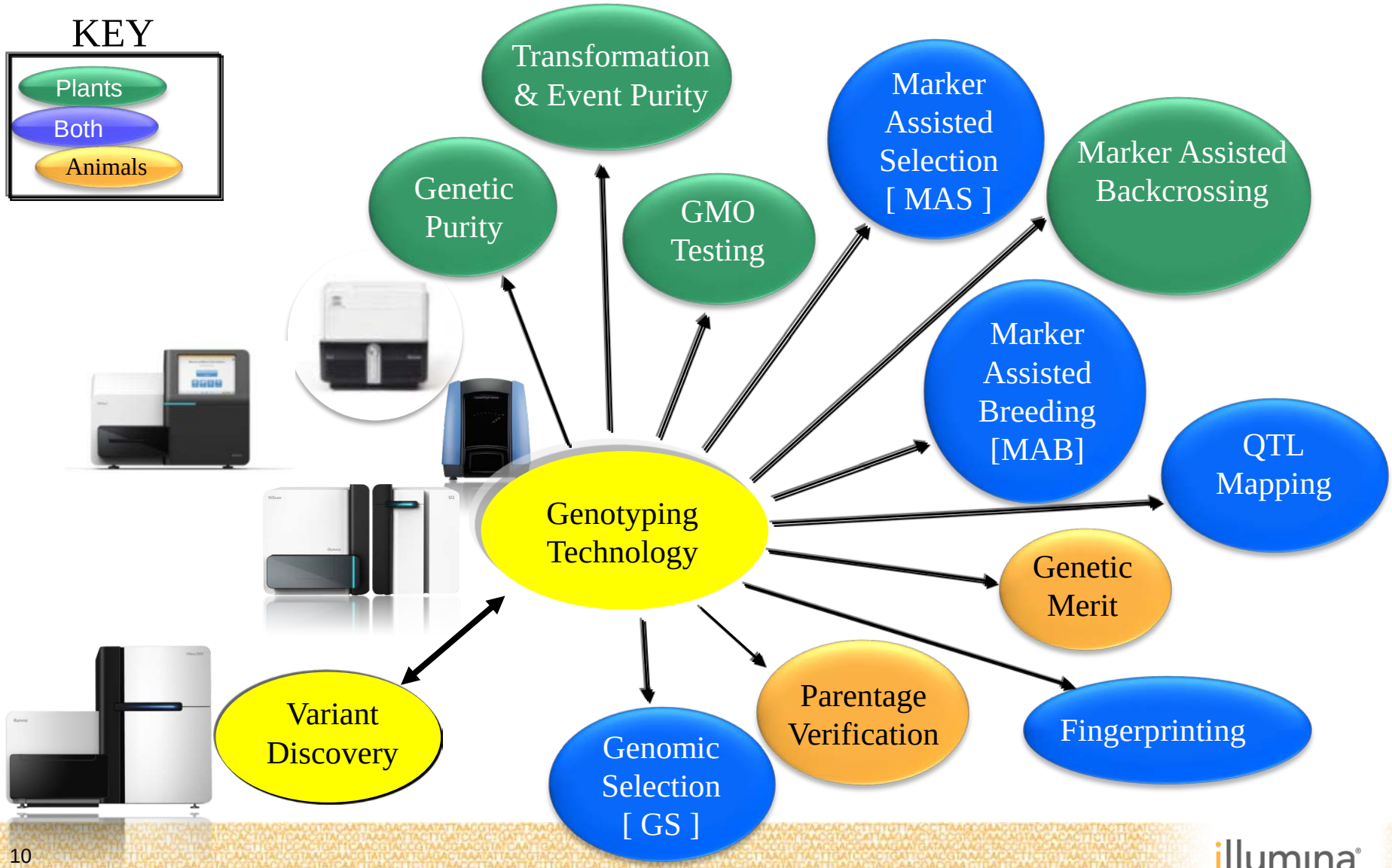
Prokaryotes ***(19405)***

- Bacteria ***(4184)***
- Viruses ***(3629)***

Updated 08April2013

Discover. Develop. Deploy.

Access to multiple applications



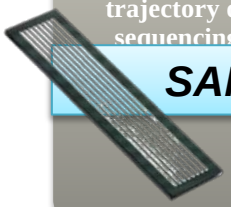
Quick Portfolio Review

Illumina Portfolio Overview

iNNOVATION is in our DNA

From Genome-Wide Discovery to Targeted Validation and Screening

Sequencing			Arrays			qPCR
						
HiSeq 2000/2500 Redefining the trajectory of sequencing	**HiSeq 1000/1500** Powerful, Scalable, Flexible	**MiSeq** Rapid, Integrated, Affordable	**HiScanSQ** Unique combination of sequencing and arrays	**iScan** Speed, quality and versatility for arrays	**BeadXpress** Accuracy, versatility and flexibility for molecular testing	**Eco** Gold-standard qPCR made accessible



SAMPLE PREP & SEQUENCING CHEMISTRY

INFINIUM ASSAY

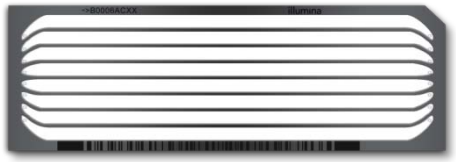
GOLDENGATE ASSAY

NuPCR

HiSeq 2500

HIGH OUTPUT

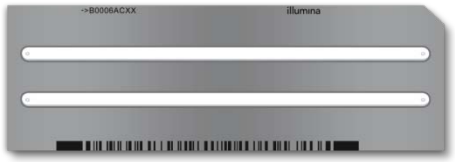
50 – 600Gb
2 – 11 days
2 x 100bp max



Larger projects, fewer runs

RAPID RUN

10 – 180Gb
7 – 40 hours
2 x 150bp max

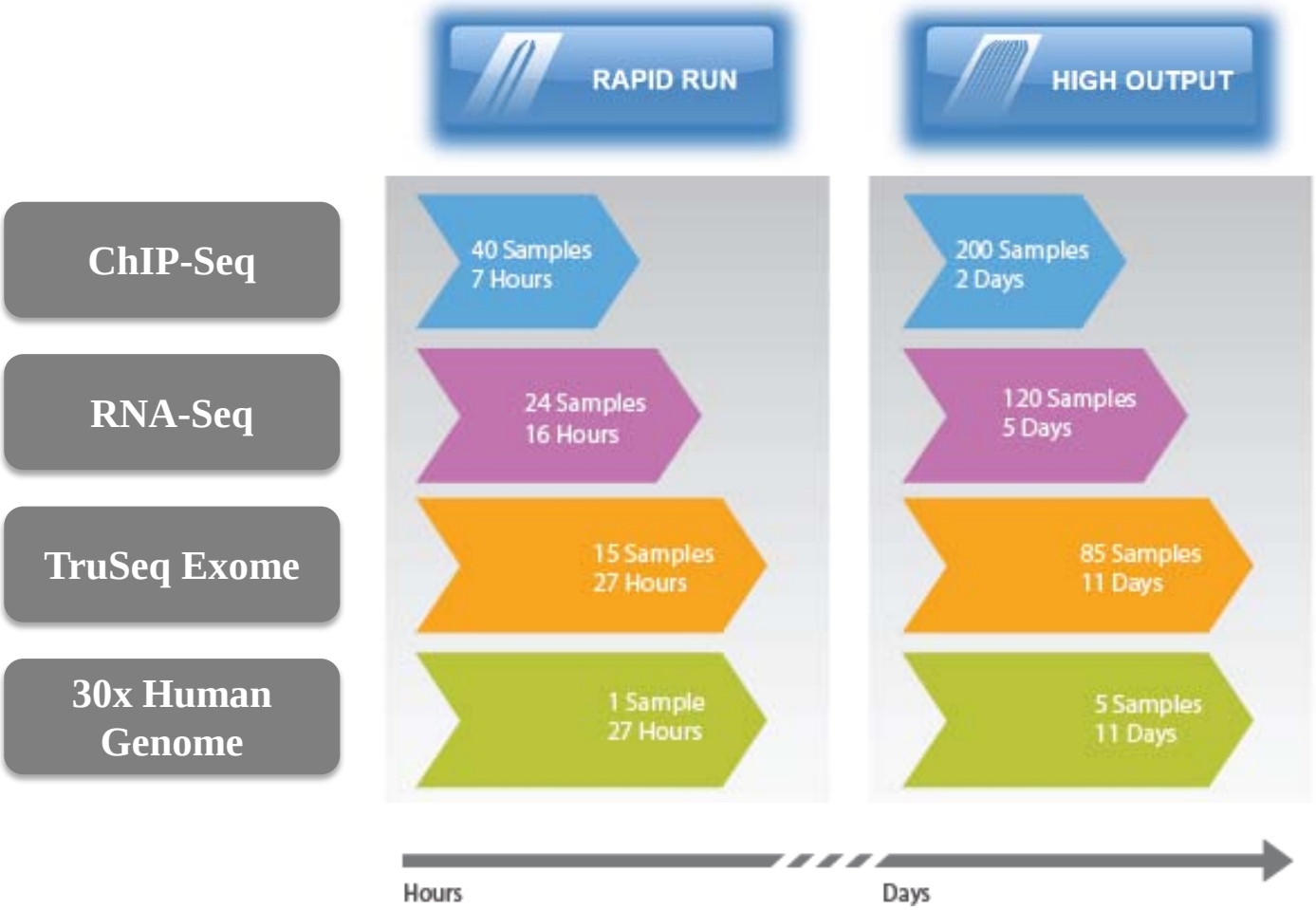


Smaller projects, quick results

HiSeq 2500

Broadest Applications Flexibility

Any application, any sample size. Operational efficiency.



illumina

MiSeq – Continuous Performance Improvements

Path towards 15Gb per run; enabling broader range of applications

Output	>1.5 Gb
Clusters	~7M
Read length	2 x 150 bp

Output	>8 Gb
Clusters	>15M
Read length	2 x 250 bp

Output	15 Gb
Clusters	25M
Read length	2 x 300 bp

Faster chemistry
Dual surface imaging

Plan to commercialize in 2H 2013

Output - Gb

20

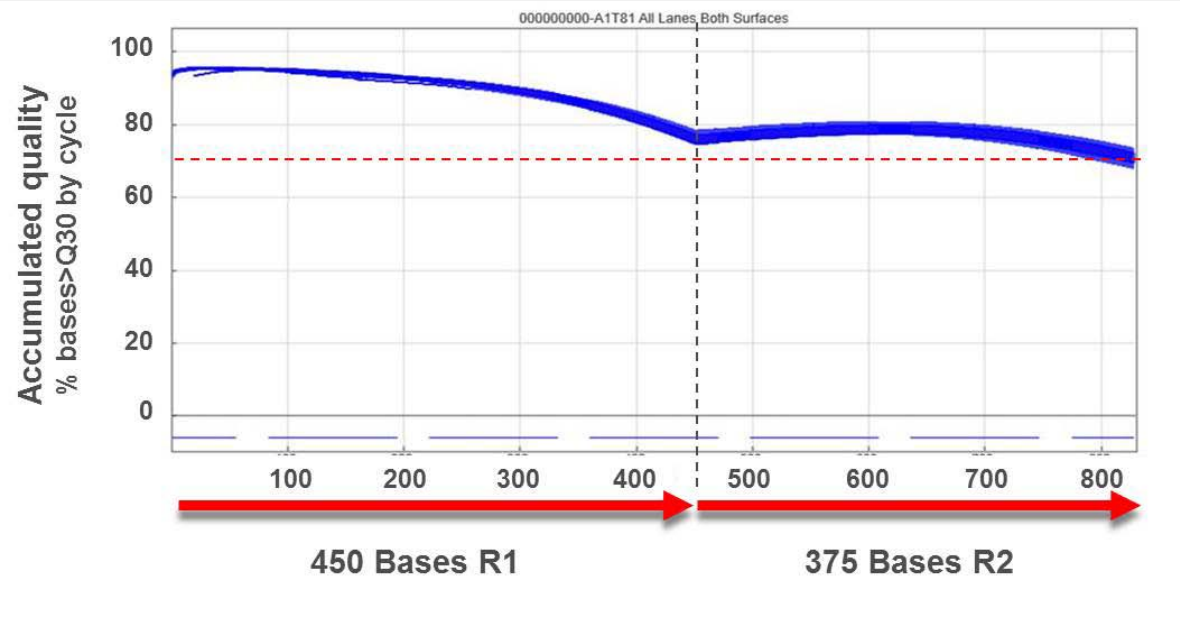
10

1

2Q11 3Q11 4Q11 1Q12 2Q12 3Q12 4Q12 1Q13 2Q13 3Q13 4Q13

MiSeq – R&D Demonstrated Scalability

Long read runs in excess of 20Gb



Output	22 Gb
Clusters	26.5M
Read length	450 (R1), 375 (R2)
Quality	71% ≥Q30
MR	1.7% (R1), 2.4% (R2)

Demonstrated by Illumina R&D

iScan™ System

POWERFUL & ECONOMICAL MICROARRAY PLATFORM

- Semi-confocal laser scanning system
- Resolution, 0.5 micron
- Dual line - lasers at 532 & 658nm
- Fully automated operation
- Compact bench top system

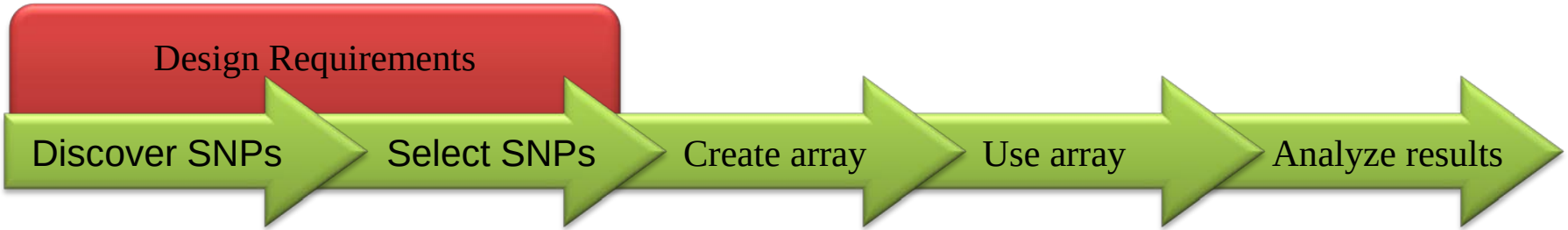
Add-on Content for Commercial Ag Genotyping Arrays

Product	Current BeadChip Format	Additional # Att BTs that can be supported on current BeadChip format	Maximum # of Beadtypes allowable on current substrate
BovineHD	8-sample	85,000	900,000
BovineSNP50v2	24-sample	31,000	90,000
CanineHD	12-sample	58,000	250,000
MaizeSNP50	24-sample	29,000	90,000
OvineSNP50	12-sample	183,00	250,000
PorcineSNP60v2	24-sample	32,000	90,000
BovineLD	24-sample	80,000	90,000

Min # ABT: 1,000

Min # samples: 1,152

Steps in Building Genotyping Tools



-Public/Private
-Genome Sequencing
-Genotyping by Sequencing
-Mapping?

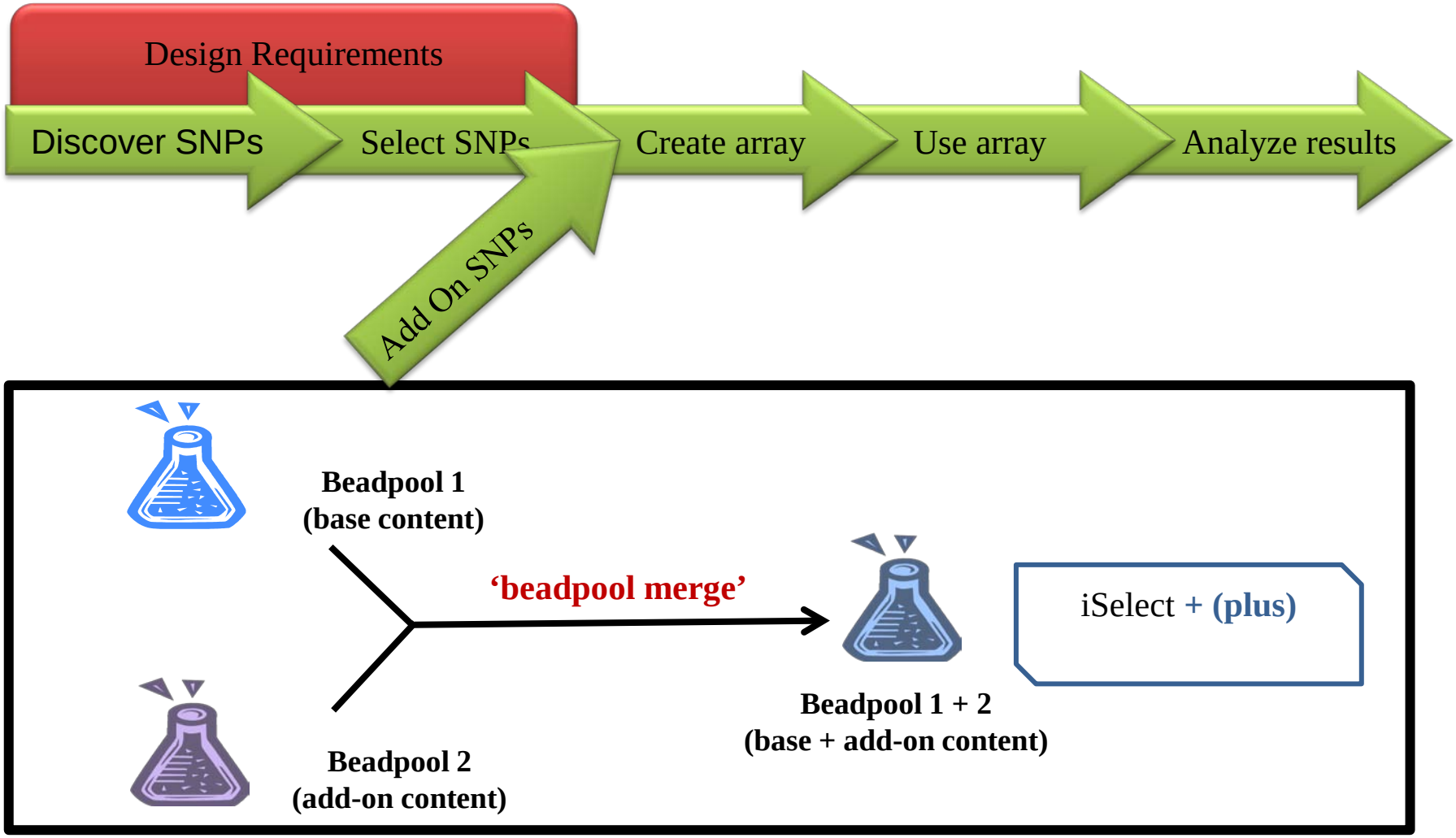
Parameters:
-MAF
-Genic
-Density
-Distribution
Assay Design Tool:
-Design score
-Infinium I vs. II

-Synthesize oligo
-Create beadpool
-Assemble chips
-Decode/QC
-Ship Chips

-Collect/gather Sample
-DNA extraction
-Infinium assay
-iScan/HiScan or Service Provider

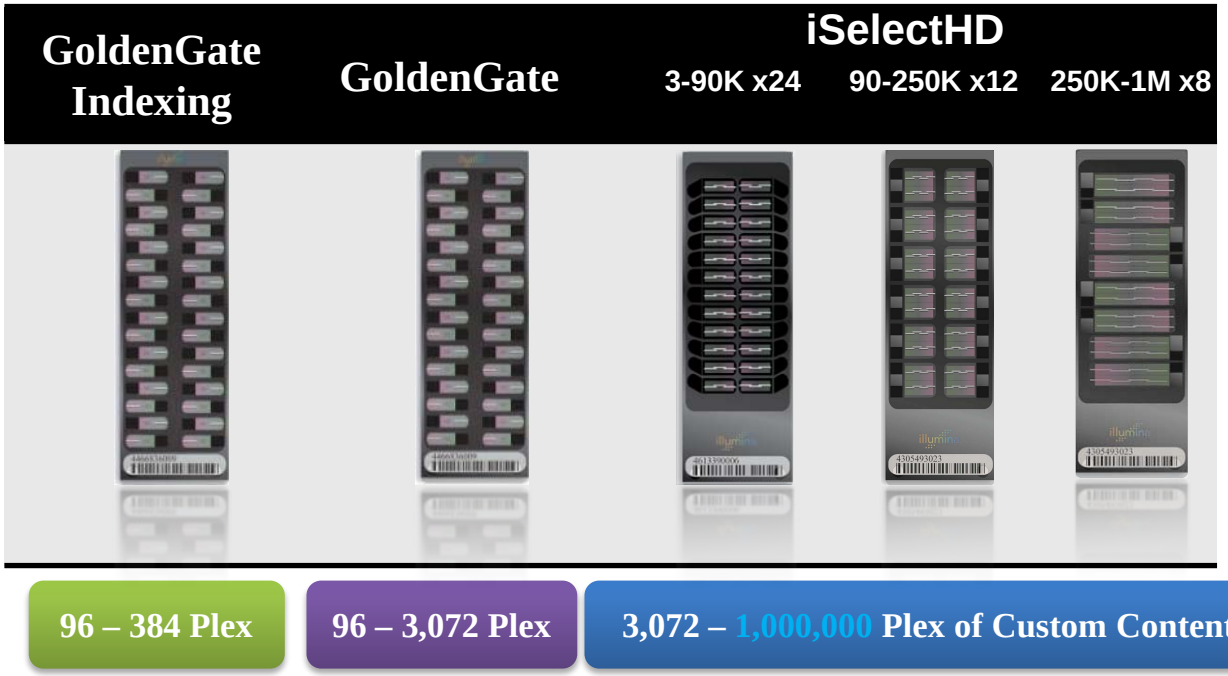
-Many, Many Genotyping Applications

Keeping Updated and/or Private Content



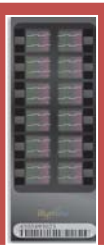
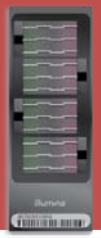
Expanded Custom Genotyping Options

Driving the Cycle of Discovery and Validation



- Ability to add on supplemental content

iSelect BeadChip Formats and Supported Marker Densities

			
BeadChip Format	24x1 2um	12x1 2um	4x1 2um
Number of Attempted Beadtypes (ABT)	3,072 – 90,000	90,001 – 250,000	250,001 – 1M
Add-on content range* (ABTs)	Up to limit of 90,000	Up to limit of 250,000	Up to limit of 1M
Instrumentation Needed	iScan, HiScan,		

What is an Ag Consortium?

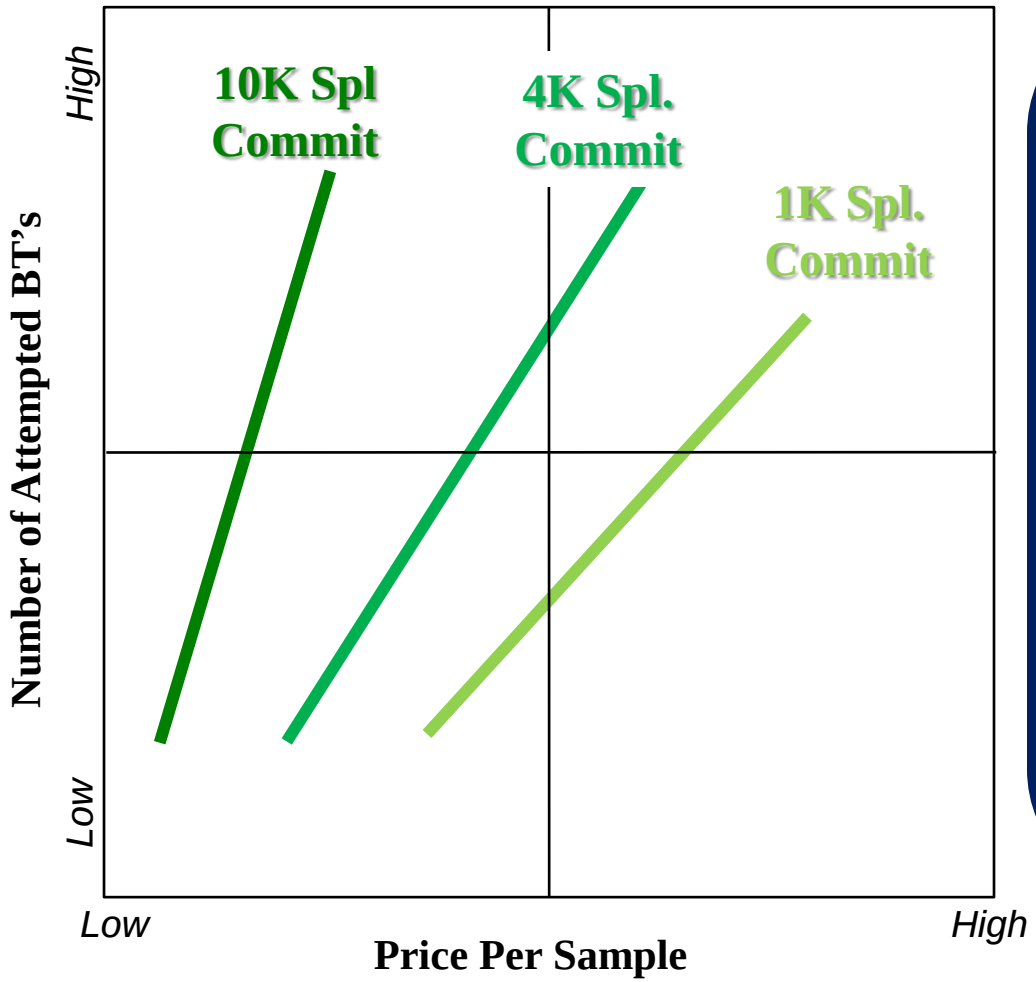
- "A consortium is an association of two or more individuals, companies, organizations or governments (or any combination of these entities) with the objective of participating in a common activity or pooling their resources for achieving a common goal."^
- A group of people or organizations with a common interest in advancing the collective understanding of the genetics of an agriculturally important species.
- The consortium members work together to enable actions that promote that interest

Consortia are helping advance genotyping in agriculture species around the world.

^ *Source: Wikipedia*

Benefits Exist to Pooling Samples

Economies of scale, also known as more for less!



When groups pool resources to create a genotyping tool, it lowers the barrier to entry by spreading the cost of synthesis of the beadpool over more samples. This was the case with the OvineSNP50 and BovineSNP50 commercial development projects. This was also the case with RosBREED Apple, Peach, Cherry and SolCAP Potato and Tomato genotyping beadchips.

Public Consortia-Developed Agriculture Arrays

Species	Identification	Classification	Provider	Consortium	# SNPs
Potato	Potato	Public	Illumina	SolCAP	8,303
Tomato	Tomato	Public	Illumina	SolCAP	7,720
Apple	Apple	Public	Illumina	RosBREED	8,788
Peach	Peach	Public	Illumina	RosBREED	8,144
Cherry	Cherry	Public	Illumina	RosBREED	5,696
Maize	MaizeSNP50	Commercial	Illumina	Commercial	56,110
Chicken	Chicken	Private: Public Sale	Illumina	Cobb Vantress-Hendrix-USDA	57,636
Cat	Feline	Private: Public Sale	Illumina	Morris Animal Foundation	62,897
Horse	Equine	Private: Public Sale	Illumina	Neogen (GeneSeek)	65,157
SheepHD	Ovine	Private: Public Sale	Illumina	AgResearch	680,000
Cattle	BovineHD	Commercial	Illumina	Various	777,962
Cattle	BovineSNP50v2	Commercial	Illumina	Various	54,609
Sheep	OvineSNP50	Commercial	Illumina	Various	52,241
Cattle	BovineLD	Commercial	Illumina	Various	6,909
Pig	PorcineSNP60	Commercial	Illumina	Various	62,163
Dog	CanineHD	Commercial	Illumina	Various	173,662
Wheat	Wheat 9k and 90K	Public	Illumina	Australia/US Wheat Consortium	8,632/81,587
Brassica	Brassica 50k	Public	Illumina	Intl Brassica SNP Consortium	52,157
Grape	Grape	Public	Illumina	ReSeq	18,071
Goat	Goat	Public	Illumina	Int'l Goat Genome	53,347

New Products

moleculo

Illumina Acquires Moleculo Inc. for Longer Reads

January 8, 2013 by nextgenseek · 3 Comments

Illumina announced at the JP Morgan Healthcare Conference (JMP2013) that it has acquired San Francisco based Mocleculo. Moleculo, the sequencing startup company (with not much information in the web), is a product of Stanford University's sequencing pioneer Stephen Quake.

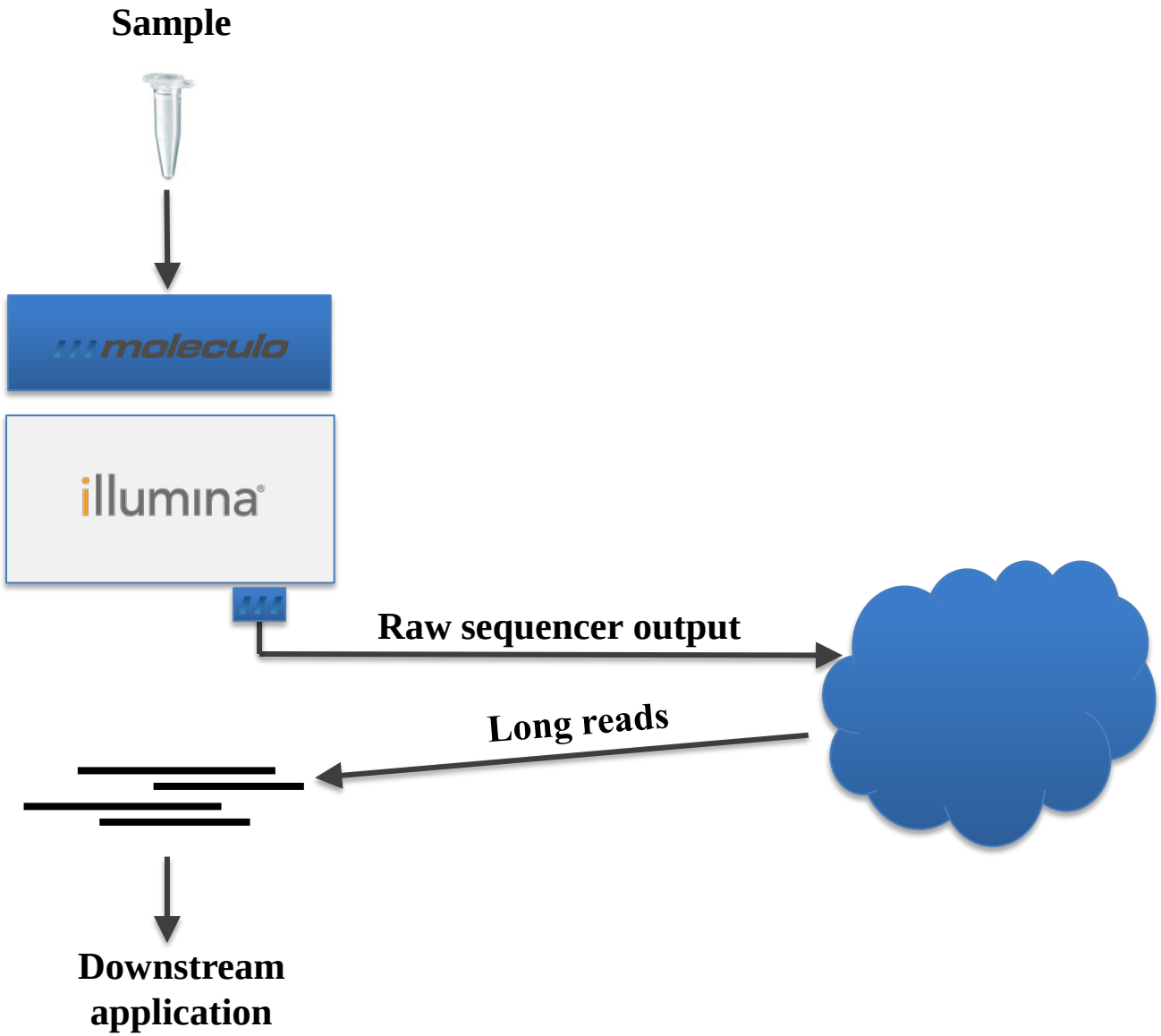
Production of Long (1.5kb – 15.0kb), Accurate, DNA Sequencing Reads Using an Illumina HiSeq2000 to Support De novo Genome Assembly

Date: Saturday, January 12, 2013
Time: 3:50 PM
Room: Town and Country

***Geoff Waldbieser** , USDA-ARS, Stoneville, MS*
Michael Kertesz , Moleculo, Inc., San Francisco, CA
Dmitry Pushkarev , Moleculo, Inc., San Francisco, CA
Tim Blauwkamp , Moleculo, Inc., San Francisco, CA
John Liu , Auburn University, Auburn, AL

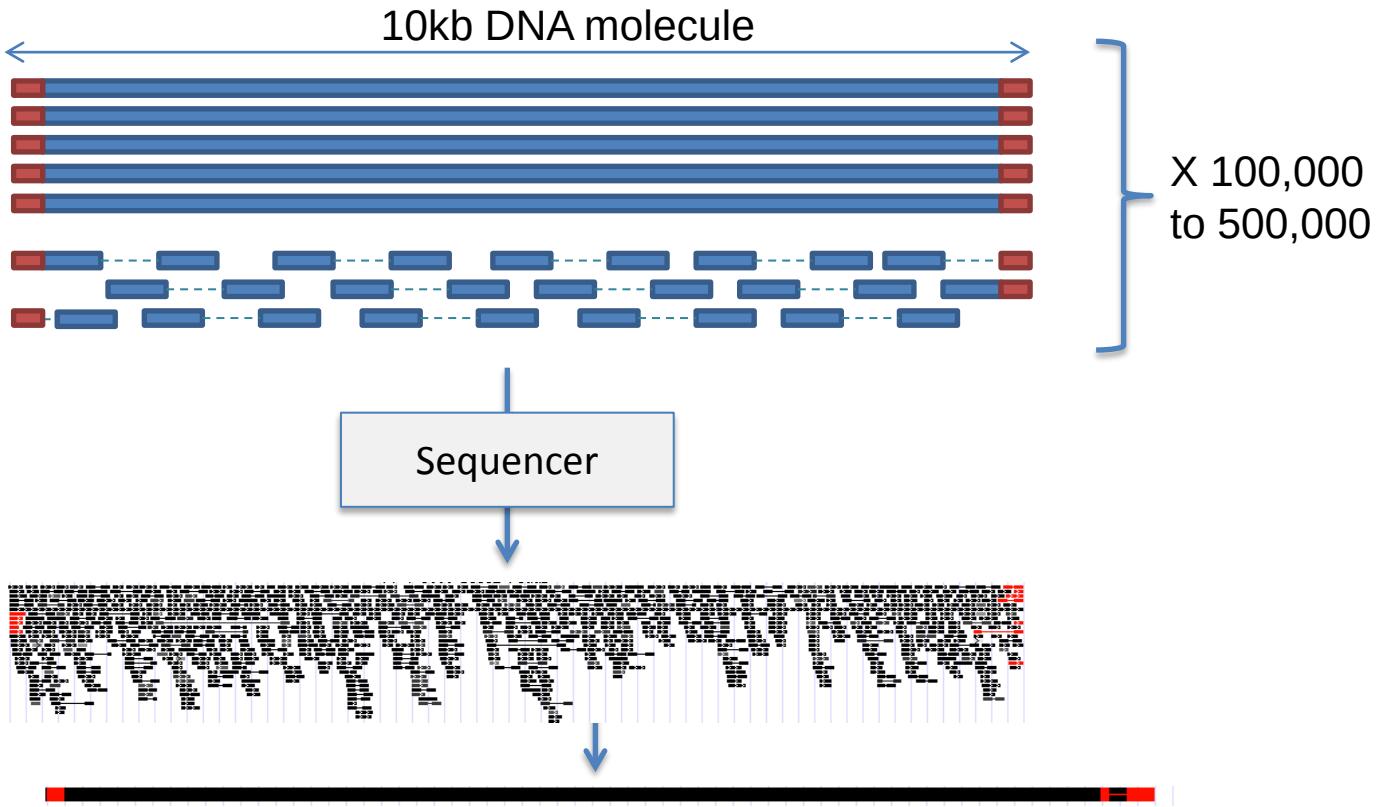
Interspersed repeat sequences such as transposons and short tandem repeats shatter de novo genome sequence assemblies because short DNA sequences cannot span the repeat sequence. In order to produce a more contiguous genome assembly for the blue catfish, Ictalurus furcatus, we have used Moleculo's Long Reads product to generate extremely long and accurate reads through Illumina-based sequencing of libraries produced from long genomic fragments. To date, 201,508 long reads (933Mb total) have been produced from two libraries, ranged in length from 1.5kb to 15.8kb, and 83% of the total sequence was found in 129,554 long reads of at least 3.0kb. Pairwise alignments revealed 145,189 long reads contained only one or no mismatched bases along lengths of 400bp to 15,887bp. Preliminary assembly of only the long reads, using 99% sequence overlap identity, produced 46,098 contigs with an N50 length of 12.9kb and N80 length of 8.5kb. A further 42,141 long reads remained singlets (N80 = 4.6kb, N50 = 7.0kb). The long reads were aligned with the 1.6kb Tip1 and 1.0kb Tip2 transposons of channel catfish. Twenty one long reads contained an average 2.1kb (minimum 200bp) of sequence flanking the Tip1 orthologs. Similarly, 226 long reads contained an average 3.0kb of sequence flanking the Tip2 orthologs. The initial results demonstrate the utility of long and accurate DNA reads in bridging repetitive regions of the genome that cannot be otherwise resolved. Four additional Moleculo Long Read libraries are currently being sequenced to produce an additional ~2Gb of sequence in long reads.

moleculo



illumina

Synthetic Long Reads



Key components:

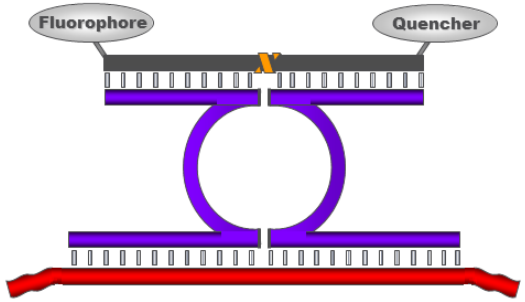
- **Method** for accurate amplification
- **Kit** for highly-parallel library prep
- **Algorithm** for long fragment reconstruction

NūPCR –Product Overview

NūPCR (Gene Expression and Quantification)

Offerings: custom designed probe-based assays + Master Mix

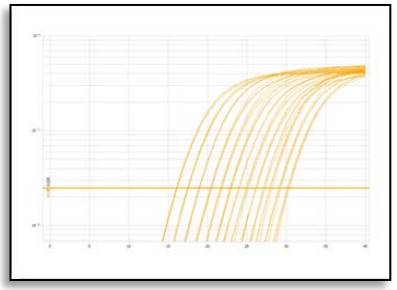
- Illumina developed customizable assays and Master Mix
- **DesignStudio** custom primer and probe design



qPCR DNA Binding Dye Assays (Gene Expression and Quantification)

Offerings: custom designed primers

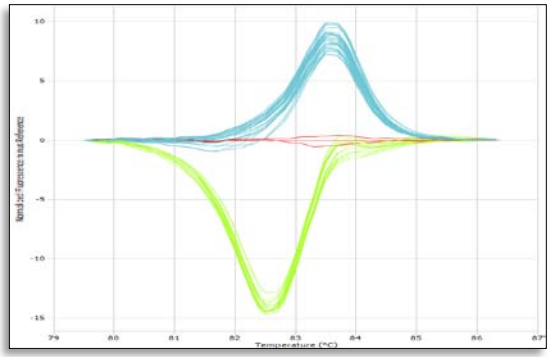
- Ready-to-use custom qPCR assays
- **DesignStudio** custom primer design



qPCR High Resolution Melt Assays (Genotyping)

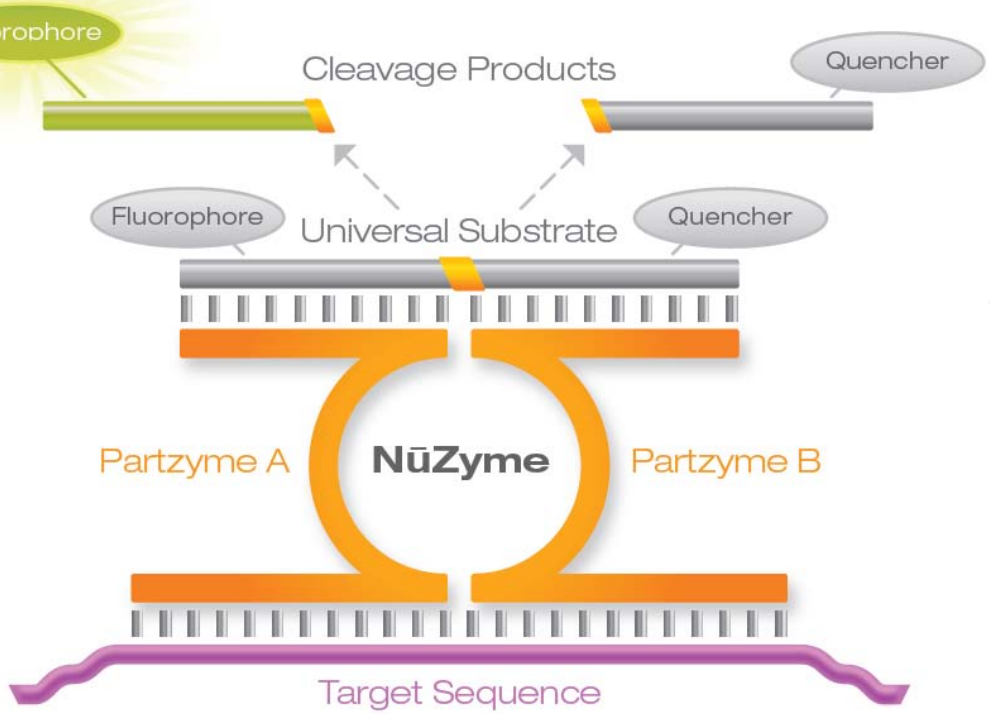
Offerings: custom designed primers

- Ready-to-use custom HRM assays
- **DesignStudio** custom primer design



illumina®

What is NūPCR?



- NūPCR is a probe-based qPCR technology that utilizes Illumina's unique NūZyme chemistry NūZyme is a multipart nucleic acid enzyme (DNAzyme).
- All oligos must assemble on the target sequence in order for fluorescence to be generated (high specificity). (5 oligos including primers)
- Universal substrate results in cost savings
- 4 dyes to choose from- allows multiplexing

Illumina Sample Prep Solutions

Integrated workflows from sample to analyzed data

New - Nextera XT
Fast Amplicons, Small Genomes, Plasmids

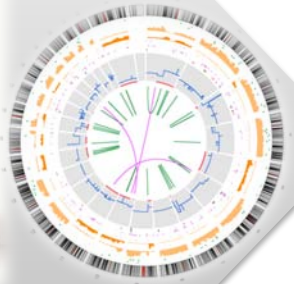
Nextera
Low input, fast, versatile

TruSeq DNA LT & HT
Simple, scalable, and cost effective

Coming soon-TruSeq stranded RNA
For plants

TruSeq Small RNA
High-throughput miRNA discovery & profiling

TruSeq ChIP, FFPE & Stranded RNA



BaseSpace
Data storage, analysis

TruSeq Chemistry
Clustering & Sequencing

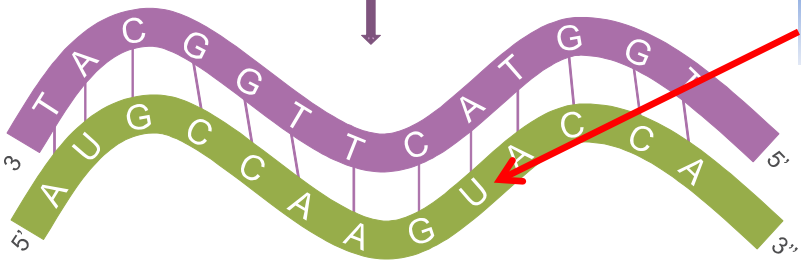
Incorporation of dUTP in Second Strand Synthesis

First strand cDNA synthesis (w random hex)

RNA

First Strand cDNA

Second strand cDNA synthesis (w random hex)



Incorporates dUTP instead of dTTP

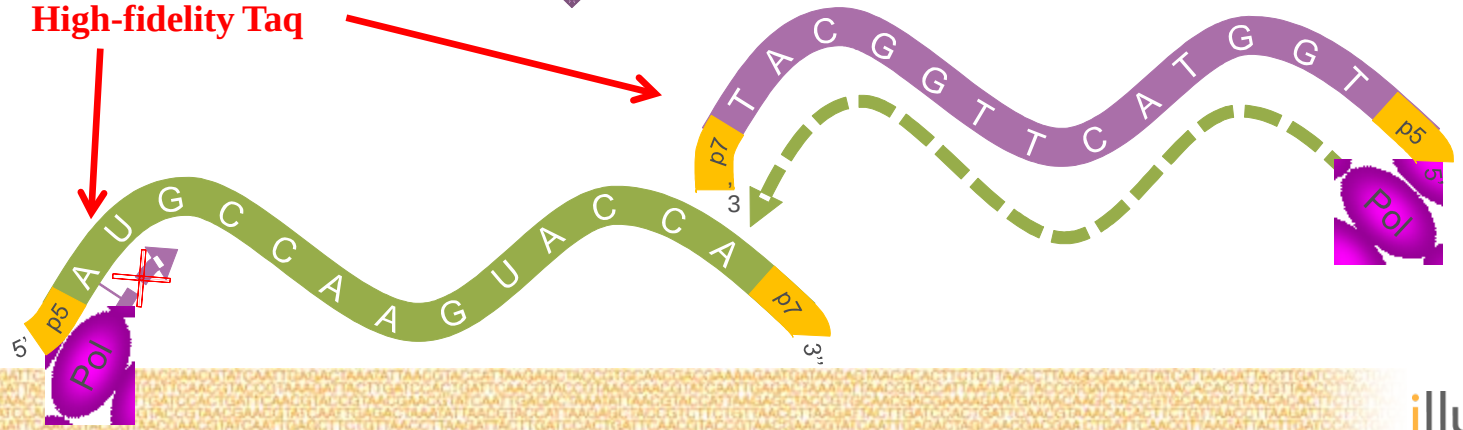
First Strand cDNA

Second Strand cDNA

Ligate Adapters

Enrich DNA fragments / Amplify

High-fidelity Taq



Trends and the Future

Sequence-Based Genotyping
Next-Generation Genotyping
Genotyping by Sequencing
Genotyping using Next-Generation Sequencing

Sequence-Based Genotyping

- Skim sequencing without reduced representation
- PCR and hybridization based methods of target enrichment
  - PCR based methods
  - Solid/Liquid phase hybridization methods
- Restriction enzyme methods of reducing representation
  - RADseq, GBS from Cornell and subsequent Cornell method

Genotyping by Sequencing

Review of currently published methods

Application Spotlight: Analyzing Genetic Variation

illumina®

Agrigenomics Genotyping Decisions Reach a Crossroads

For some applications, sequence-based genotyping provides a lower cost alternative to microarrays in performing genetic variation studies.

Introduction

Today's agrigenomics researchers have a wide variety of technologies at their disposal for collecting genetic information. Array-based methods of SNP screening have been the method of choice in analyzing and associating traits with regions of the genome for many plants and animals. As sequencing costs continue to drop, new approaches that leverage next-generation sequencing (NGS) technology are being developed to perform genotyping studies. We use the term next-generation sequencing-based genotyping (NGG) to encompass genotyping methods that leverage NGS technology. NGG includes targeted, reduced representation and hybridization-based approaches to discover and genotype SNPs, often simultaneously in many individuals or specimens. This application spotlight provides insight into different NGG methods, their benefits, and the role conventional array technology will play in the future.

Arrays Pave the Way in Agrigenomic Genotyping

In the late 1980s, researchers began identifying specific regions of DNA that influenced phenotypic traits in certain species. Efforts soon turned to the development of accurate and cost-effective genetic populations[4]. By leveraging genetic screening, farmers and livestock breeders could gain immediate feedback, supporting better informed breeding decisions and accelerating their return on investment (ROI). Genotyping tools with a lower cost per sample could enable genetic screening to be performed routinely on large populations, with an attractive ROI offsetting the implementation cost of the technology.

Sequencing Advances Can Deliver More Cost-Effective Genotyping

The rapid evolution of sequencing technology has resulted in higher throughput and a lower cost per sample, often positioning NGG as a cost-effective and efficient agrigenomics tool for genotype screening, genetic mapping, purity testing, screening backcross lines, constructing haplotype maps, and performing association mapping and genomic selection[5,6,7]. The number of NGG methods continues to grow, with each offering the fundamental benefits sequencing provides, including reduced ascertainment bias, identification of variants other than SNPs (small insertions, deletions, and micro-satelites), and an ability to perform comparative analysis across samples in the absence of a reference genome (Table 1).

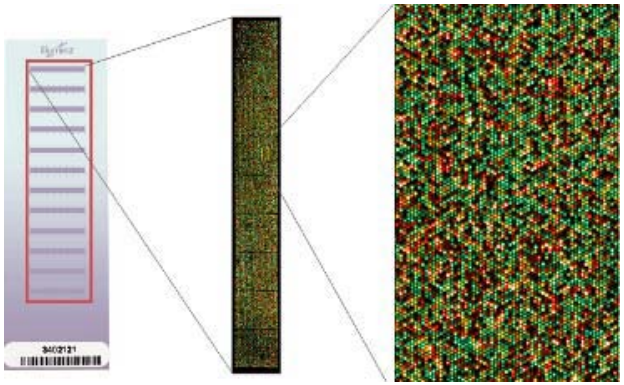
Advantages of Sequencing-based Genotyping

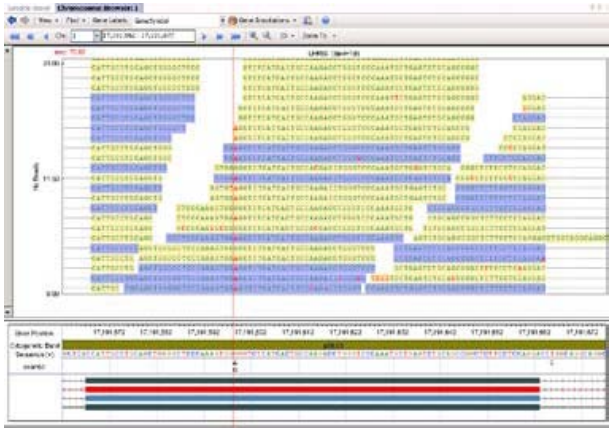
- Doesn’t need a reference genome and can be used with an uncharacterized species
- Simultaneous identification and genotyping of SNPs
- Simple library prep needing low DNA input (100- 500 ng) and is generic
- High multiplexing lowers sequencing costs

Points to Consider

- Complex bioinformatics
- Missing data and therefore need to impute

Providing a Complete Solution with Arrays and Sequencing





Illumina is best positioned to provide you with the best tools, be it microarrays or sequencing

The Future

- Accessibility to genomic tools will increase
- All economically impactful agricultural species, subspecies and their pathogens will be sequenced
- Field-based genome-wide diagnostics tools will be routine for pathogen and variety identification
- Genomic selection will surpass conventional methods as the dominant breeding paradigm

Questions?

csoans@illumina.com