Common Bean Coordinated Agricultural Project
“BeanCAP”
2013 Advisory Board Meeting at PAG XXI

Marker Development/Genotyping:

Perry Cregan & Qijian (Jian) Song

Presentation Outline

- Summary of research activities 2009 - 2011
- Design of the Illumina Infinium Genechip BARCBEAN6K_3
- The BeanCAP group order of the BARCBEAN6K_3 Genechip
- Analysis of BeanCAP common bean populations using the BARCBEAN6K_3 Genechip
- Illumina GenomeStudio software for allele calling on the data obtained from BARCBEAN6K_3

Illumina GoldenGate
Single Nucleotide Polymorphism (SNP) Analysis

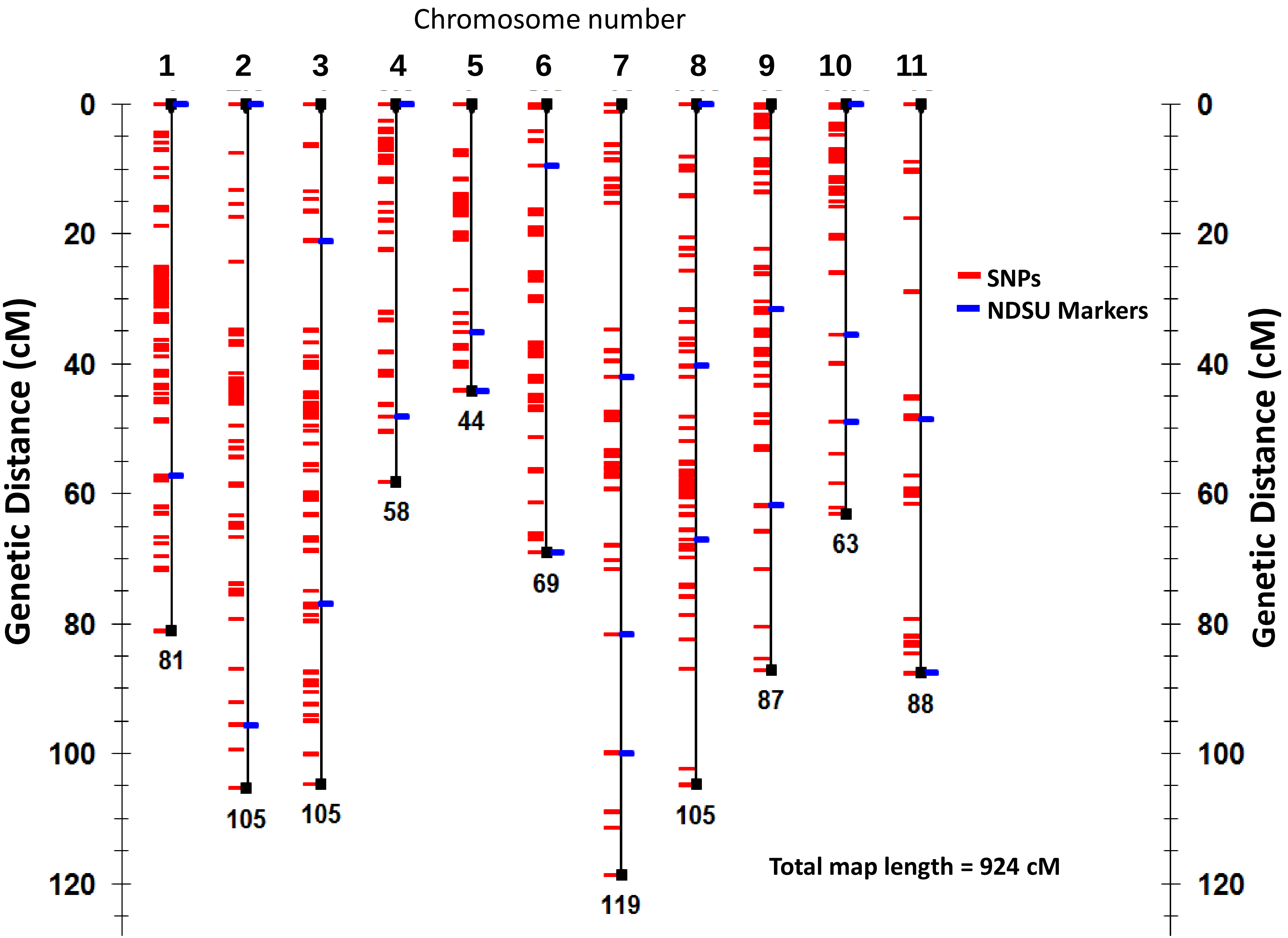
1536 SNPs derived from a number of sources

- 1050 SNPs from a sequence analysis of Jalo EEP558 and BAT93 using "Next generation" sequence analysis (Roche 454 and Illumina Genome Analyzer)★
- 486 SNPs from Sanger sequence analysis of common bean genes and PCR products derived from common bean using primers designed to soybean genes
- Analysis of common bean genotypes
  - 278 F2 lines of the Stampede x Red Hawk mapping population
  - 192 diverse genotypes representing the major market classes

---

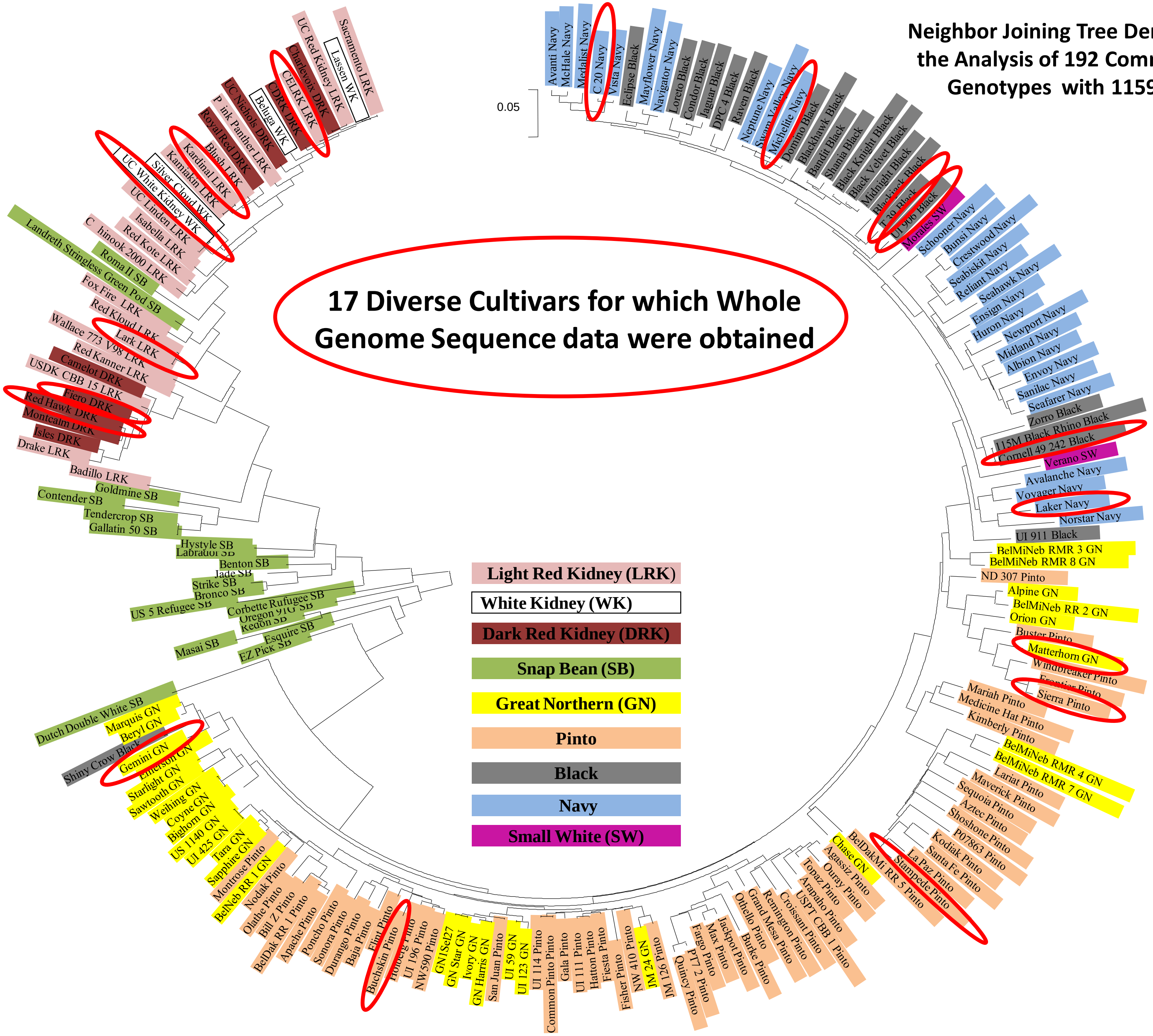
★For details see: Hyten, D.L., et al. 2010. High-throughput SNP discovery and assay development in Common Bean. BMC Genomics 11:475-483.

Stampede x Red Hawk Genetic Map (649 SNPs from PvOPA-1 + 25 NDSU Markers)



Total map length = 924 cM

Neighbor Joining Tree Derived from the Analysis of 192 Common Bean Genotypes with 1159 SNPs



17 Diverse Cultivars for which Whole Genome Sequence data were obtained

Illumina Genome Analyzer Whole Genome Sequence Obtained from 19 Bean Cultivars

Cultivar Name	Market Class
Red Hawk	Dark Red Kidney
Fiero	Dark Red Kidney
California Early LRK	Light Red Kidney
Kardinal	Light Red Kidney
Lark	Light Red Kidney
UC_White_Kidney	White Kidney
Cornell_49_242	Black
T-39	Black
UI_906	Black
Laker	Navy
C-20	Navy
Michelite	Navy
Buckskin	Pinto
Stampede	Pinto
Sierra	Pinto
Gemini	Great Northern
Matterhorn	Great Northern
BAT93	Bayo
Jalo EEP558	Canário

No Snap Beans were included in the whole genome sequence analysis

Initial Results of SNP Discovery Using Whole Genome Sequence Analysis of 19 Cultivars

- Total Illumina Genome Analyzer DNA sequence: 19.6 billion bases
- Total number of high quality SNPs among the 19 bean cultivars:

  1,859,249

- The number of SNPs polymorphic in Mapping Populations:
  - **Stampede X Red Hawk: 241,724**
  - **BAT93 X Jalo EEP558: 275, 131**

Initial Results of SNP Discovery Using Whole Genome Sequence Analysis of 19 Cultivars

- The following are the number of SNPs within each market class:
  - **Kidney: 110,163**
    - **Light Red Kidney: 21,449**
    - **Dark Red Kidney: 7,954**

776 SNPs polymorphic within both LRK & DRK Market classes

  - **Pinto: 213,087**
  - **Navy: 36,310**
  - **Black: 49,655**

687 SNPs polymorphic within all three Market classes

Two Illumina iSelect Genechips for Single Nucleotide Polymorphism (SNP) Analysis

- SNPs were identified from the alignment 19.6 billion bases of sequence data from the 19 diverse common bean genotypes to the developing whole genome sequence of G19833

- BARCBEAN6K_1: Reliable allele call data for 5,188 loci on the BeanCAP genotypes

  BARCBEAN6K_2: Reliable allele call data for 5,265 loci on the BeanCAP genotypes

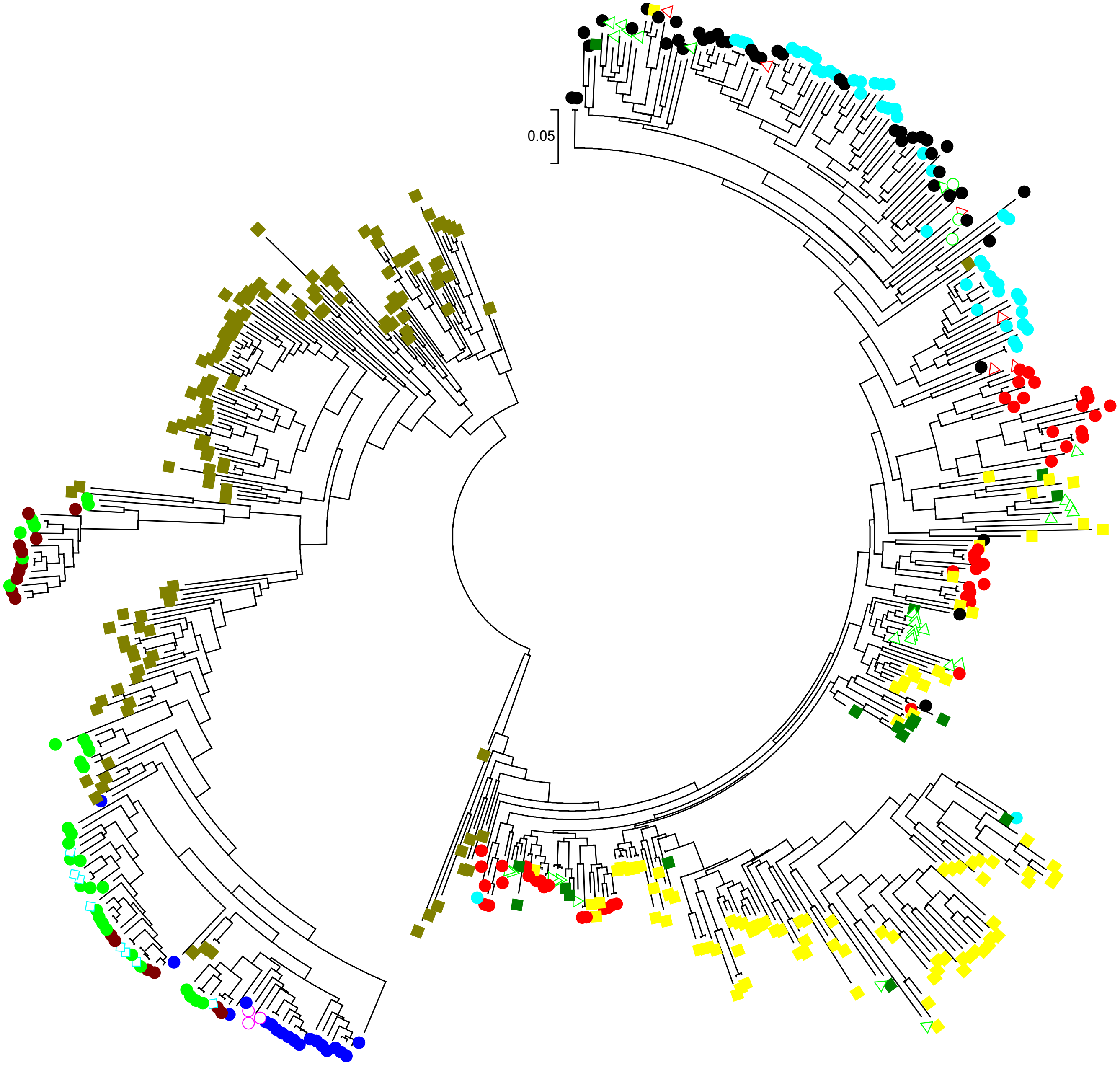
- Between the two Genechips 6,533 SNPs were segregating in the Stampede x Red Hawk population

SNP Analysis of 502 BeanCAP Genotypes

Number of genotypes in each market class and the number and proportion of the 10,453 SNP loci that were polymorphic within the genotypes of each market class

Gene Pool Race Market Class	No. of genotypes	No. of Polymorphic Loci	% of Loci polymorphic
Snap Bean	134	6230	59.6
Middle American			
Durango			
Pinto	91	5258	50.3
Great Northern	48	4338	41.5
Mesoamerican			
Navy	43	3805	36.4
Black	42	3449	33.0
Small Red	29	5091	48.7
Small White	6	4098	39.2
Pink	17	4756	45.5
Andean			
Nueva Granada			
LRK	32	3167	30.3
Cranberry	19	1495	14.3
DRK	15	2488	23.8
WK	8	1840	17.6

Neighbor Joining Tree – 502 BeanCAP Genotypes Analyzed with 10,154 SNPs



Snap Bean (134) 🟫

Pinto (91) 🟨
Gt. Northern (48) 🔴

Navy (43) ●
Black (42) ⚫
Small Red (29) △
Small White (6) △
Pink (17) 🟩

LRK (32) 🟢
DRK (15) 🟤
WK (8) ◇
Cranberry (19) 🔵
Tan (3) ○
Red Mottle (3) ○

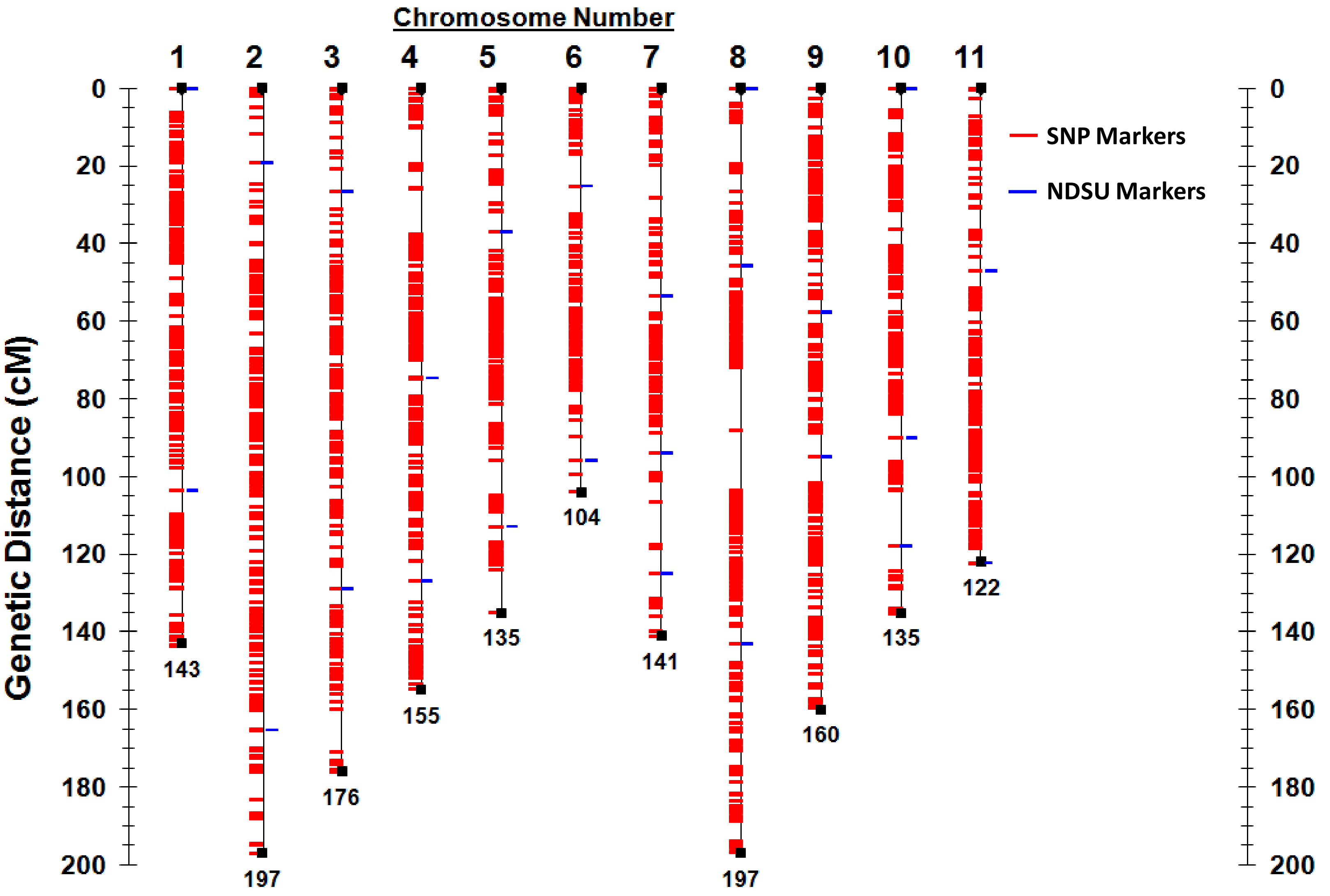
Genetic Mapping

Summary of the current status of the number of SNP markers mapped in the 278 F_2 lines of the NDSU Stampede x Red Hawk mapping population and the length in centiMorgans (cM) of the 11 *Phaseolus vulgaris* linkage groups

Linkage group	SNPs mapped in 2011	Previously mapped SNPs	Total SNPs	LG length in cM
1	760	73	833	84.0
2	518	48	566	127.6
3	531	55	586	116.9
4	708	47	755	94.0
5	666	29	695	82.6
6	199	28	227	70.8
7	192	30	222	107.8
8	824	38	862	114.0
9	449	41	490	94.6
10	763	42	805	60.2
11	923	54	977	81.5
TOTAL	6533	485	7018	1034.0

Stampede x Red Hawk Genetic Map

7018 SNPs from BARCBEAN6K_1 and BARCBEAN6K_2 and OPA-1 + 25 NDSU Markers



Design of the Illumina Infinium Genechip BARCBEAN6K_3

- **6,000 SNPs were selected from the BARCBEAN6K_1 and BARCBEAN6K_2 genechips which were used to analyze the 517 BeanCAP lines**
  - 10,913 SNPs on the BARCBEAN6K_1 and BARCBEAN6K_2 beadchips
  - 7,018 of the 10,453 were mapped in the NDSU Stampede x Red Hawk mapping population
  - 10,039 SNPs with the Minor Allele Frequency (MAF) > 5%
  - SNPs with the same allele calls across all 517 accessions fell into 4,622 groups - 1 SNP selected from each of the 4,622 groups
  - 1179 SNPs selected based upon polymorphism within Durango, Mesoamerican and Guatemalan races

- **6,000 SNPs on the BARCBEAN6K_3 genechip**
  - 4480 polymorphic within the race Mesoamerican genotypes
  - 4795 polymorphic within the race Durango genotypes
  - 4699 polymorphic within the race Guatemalan genotypes
  - 3495 polymorphic within the Andean gene pool
  - 2837 polymorphic within both the Andean and Middle Am. gene pools

The Genechip BARCBEAN6K_3 Genechip

- **5,398 SNPs on the Genechip**
- **3,200 SNPs with genetic positions (Chromosome & cM position) based upon mapping in Stampede x Red Hawk**
- **SNPs are currently named based upon the sequence scaffold in which they reside in the 14X Common bean sequence assembly**
- **We have devised a new naming system based upon the SNP positions in the Common bean Whole Genome Sequence (*Phaseolus vulgaris* - JGI v1.0) at Phytozome (http://www.phytozome.net/)**
- **Example: Pv_1.0_Chr01_283706_C_T**

The BARCBEAN6K_3 Genechip

Distribution of SNPs on the 11 Common bean Chromosomes

Chromosome number	Genetic length (cM)	Physical length (bp)	Markers/ chromosome
Chr01	84.0	52,205,531	504
Chr02	127.6	49,040,938	579
Chr03	116.9	52,284,309	361
Chr04	94.0	45,960,019	534
Chr05	82.6	40,819,286	574
Chr06	70.8	31,977,256	311
Chr07	107.8	51,758,522	458
Chr08	114.0	59,662,532	572
Chr09	94.6	37,469,608	419
Chr10	60.2	43,275,151	451
Chr11	81.5	50,367,376	573
Unassigned			62
TOTAL	1034.0	514,820,528	5398

BeanCAP BARCBEAN6K_3 Genechip Group Order

- Chips to analyze 5760 DNA samples for BeanCAP Principal Investigators
- Chips to analyze 480 DNA samples for BeanCAP Collaborators
- Chips to analyze 576 DNA samples purchased by “Feed the Future” collaborators
- Chips to analyze 480 DNA samples purchased by private industry
- Chips to analyze 960 additional DNA samples purchased by BeanCAP Principal Investigators and Collaborators

8256 Total BARCBEAN6K_3 Samples to be analyzed

BeanCAP BARCBEAN6K_3 Analyses Completed to Date

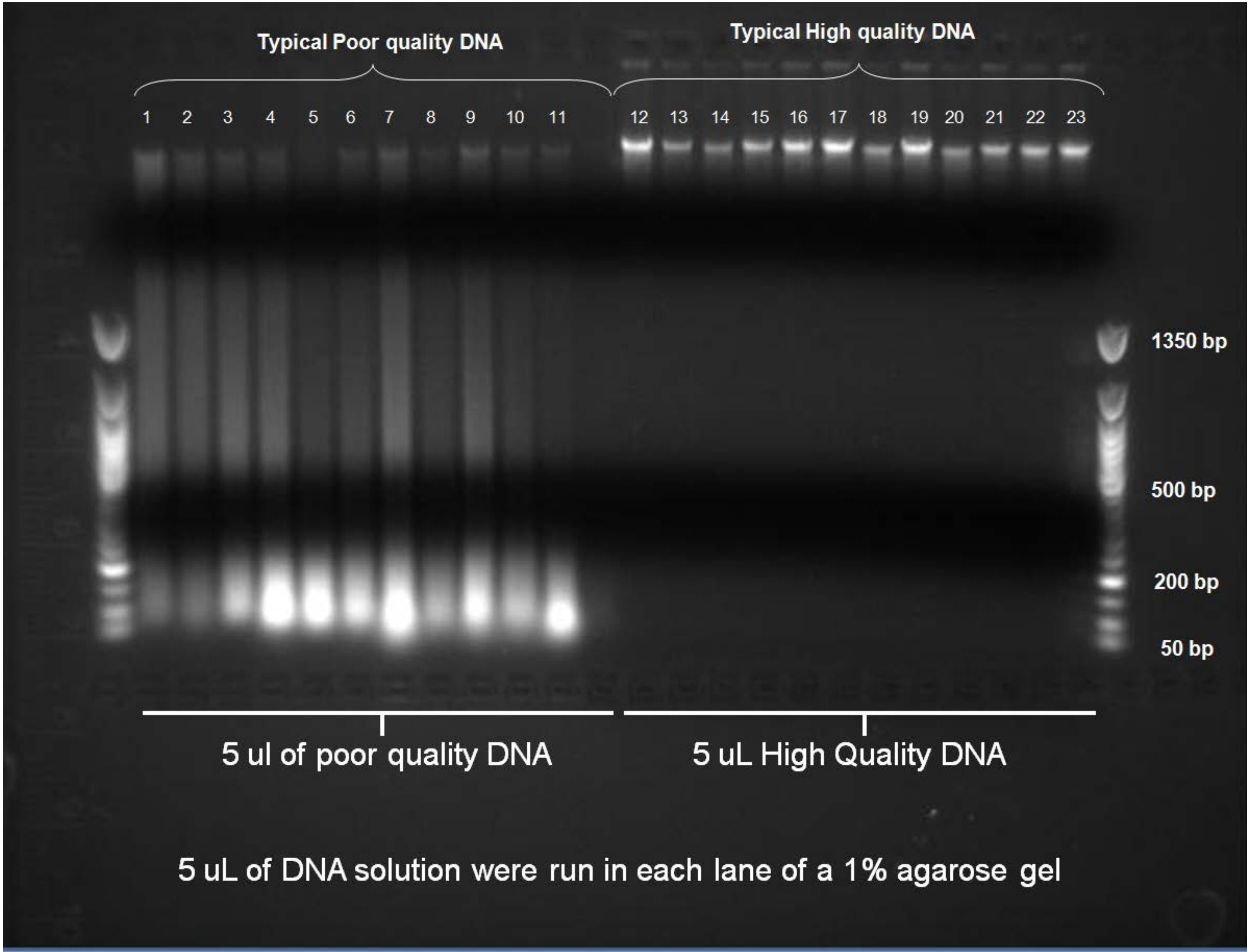
- 192 samples from Dr. Jim Myers, Oregon State
- 192 samples from Dr. Phil Miklas, USDA, Prosser, WA
- 96 samples for Dr. Shree Singh, Univ. of Idaho, Kimberly, ID
- 96 samples for Dr. Talo Pastor Corrales, USDA, Beltsville, MD
- 480 samples from Dr. Jim Kelly, Michigan State

Illumina Infinium SNP Genotyping

- High quality DNA is required with a concentration of at least 100 ng/microliter
- Run 5 uL of the isolated DNA out on a 1% agarose gel to obtain a visual estimate of DNA quality
- Ship 50 uL of each DNA sample in a tightly sealed microtiter plate on dry ice – it should remain frozen
- It is very helpful in the subsequent allele calling analysis that the parents of mapping populations are included in each microtiter plate.

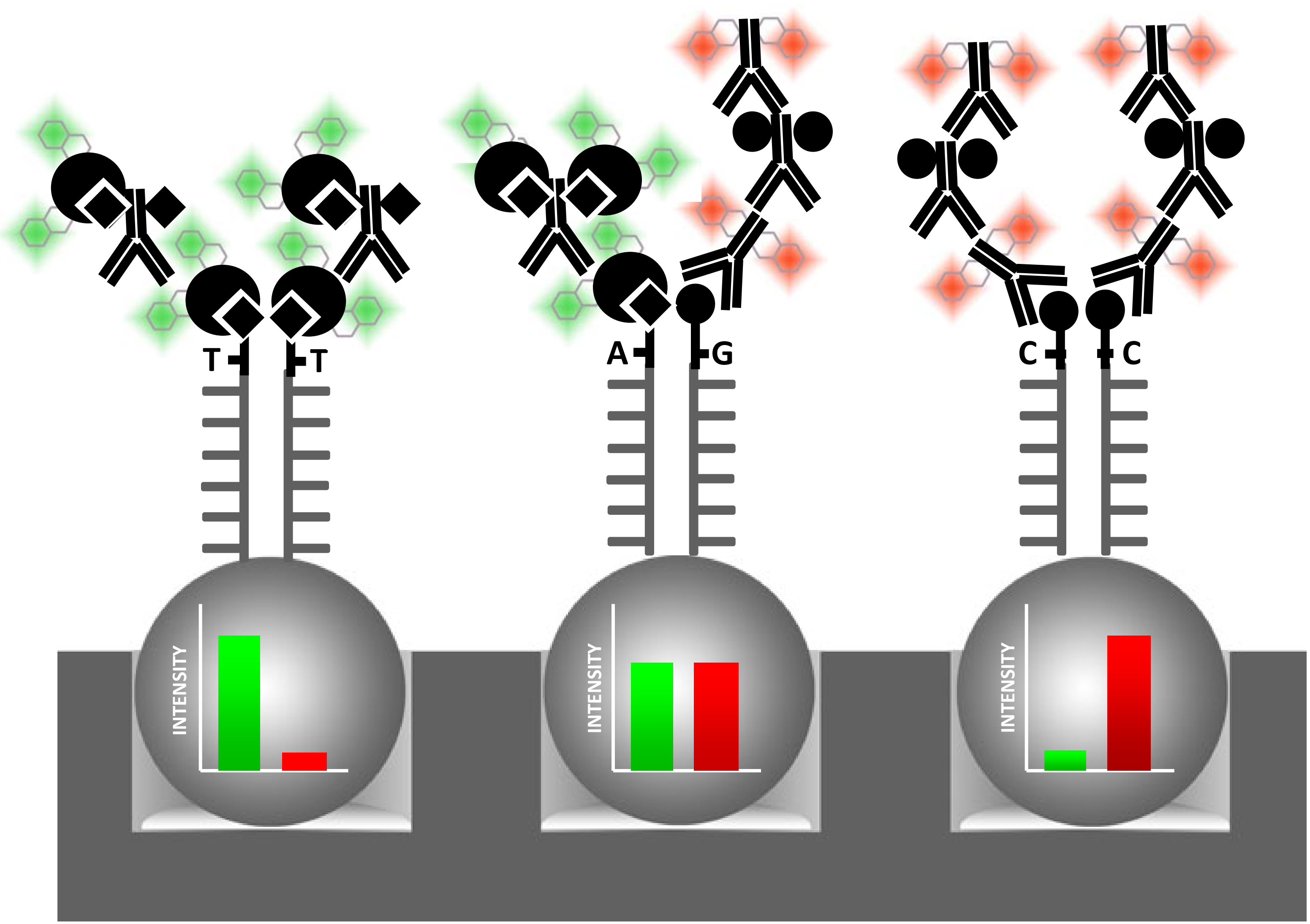
Illumina Infinium Genechip Analysis

** DNA quality is Critical **

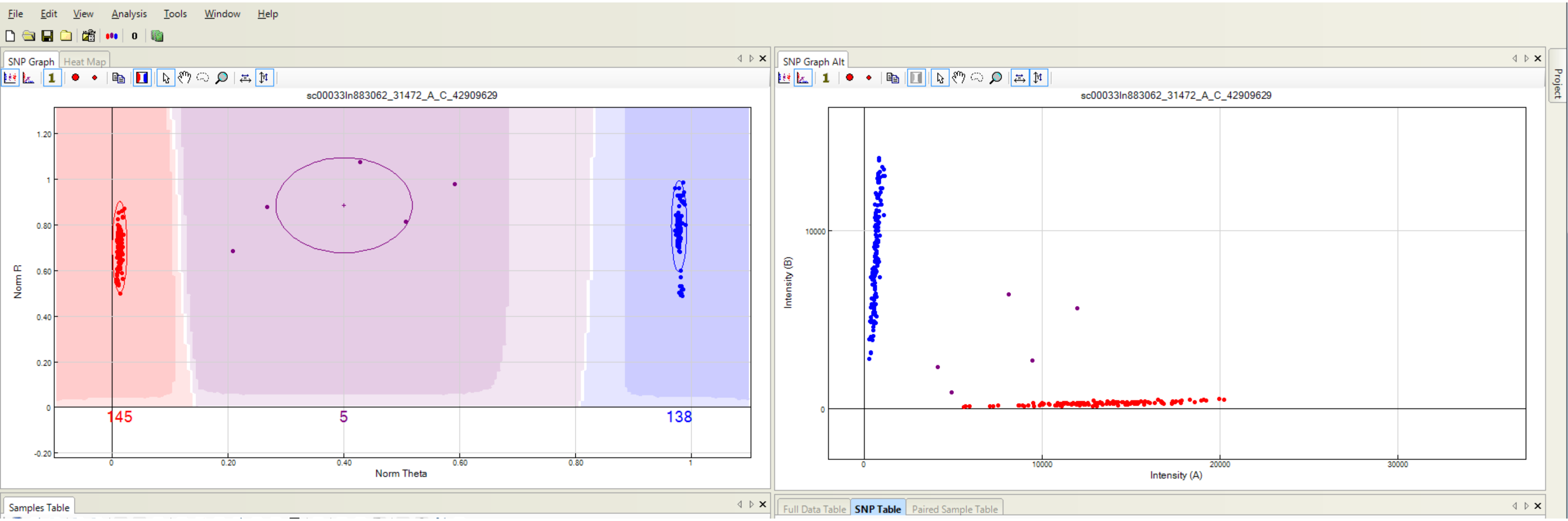


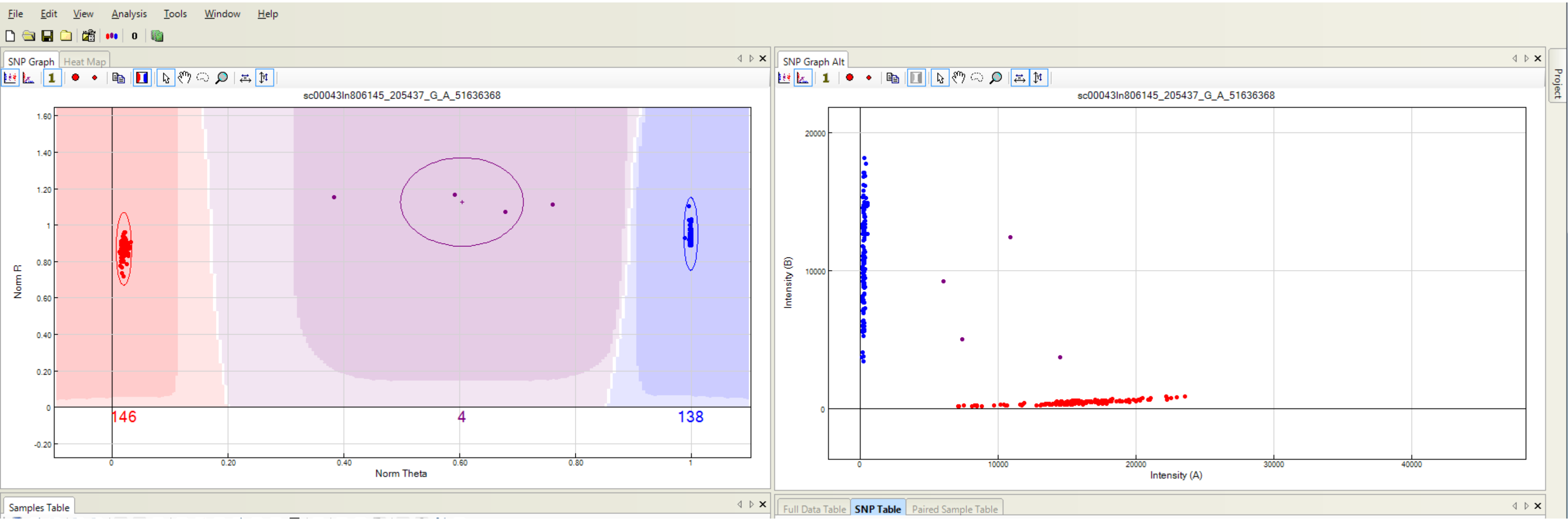
5 uL of DNA solution were run in each lane of a 1% agarose gel

The Illumina HD Infinium Assay

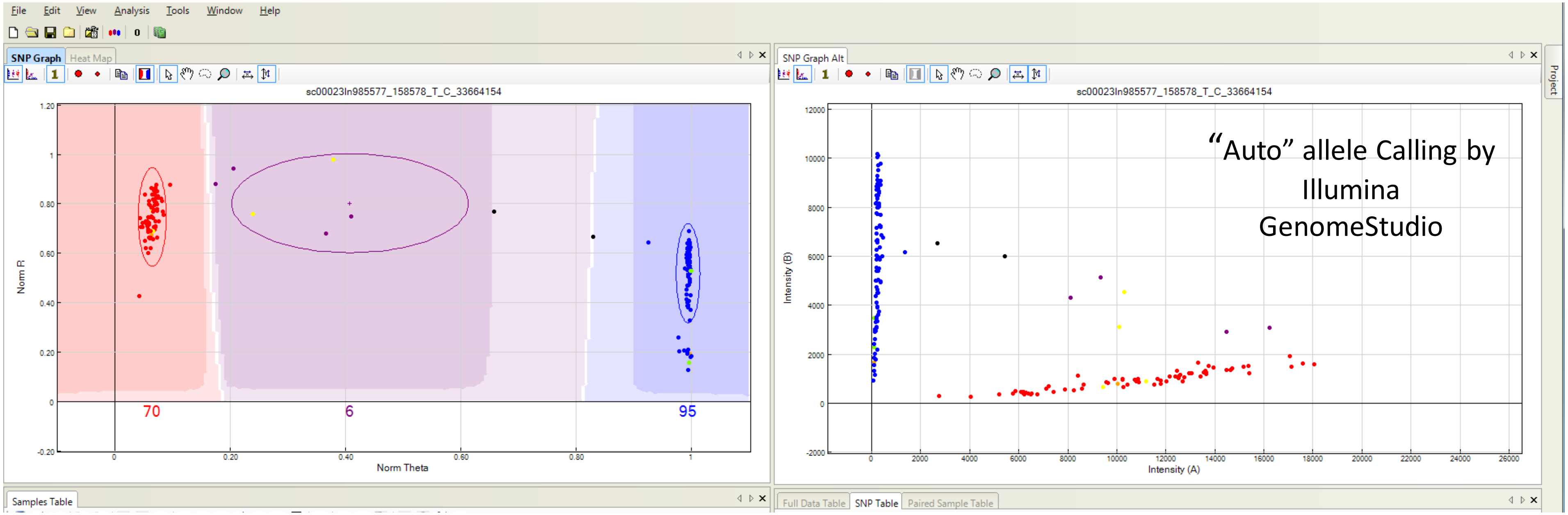


Analysis of Illumina Infinium Genechip Data Using Illumina GenomeStudio Software
- "Auto" Allele Calling by GenomeStudio -

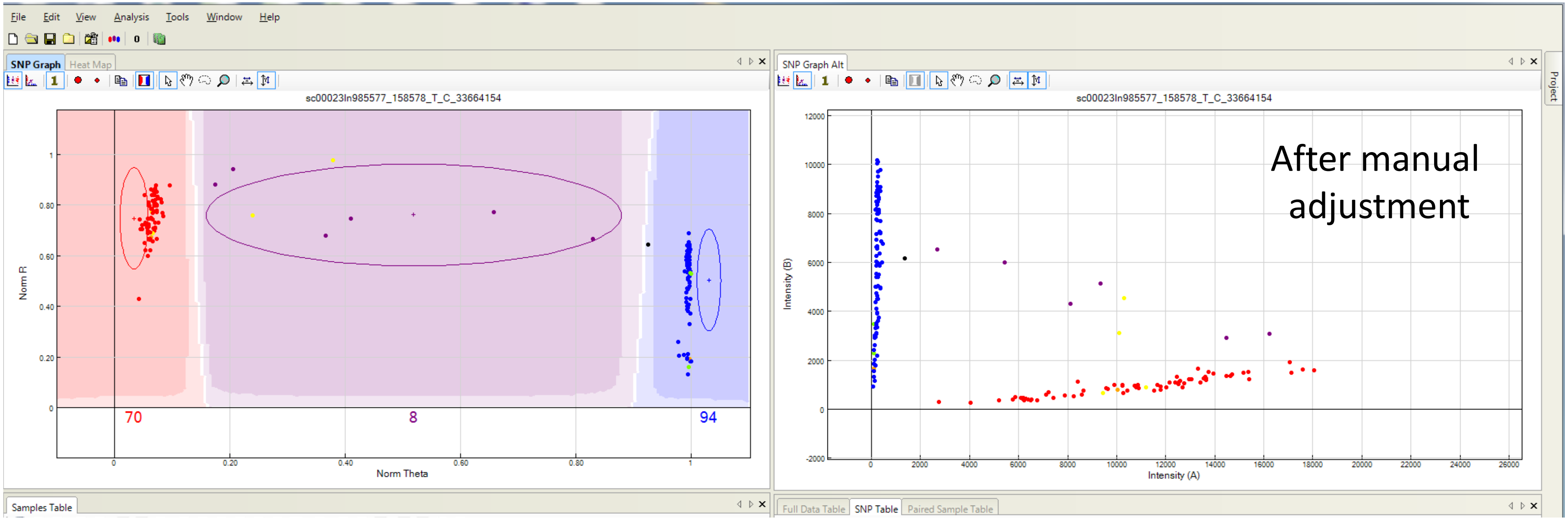




Analysis of Illumina Genechip Data Using Illumina GenomeStudio Software
-- Manual Adjustment of Allele Calls --

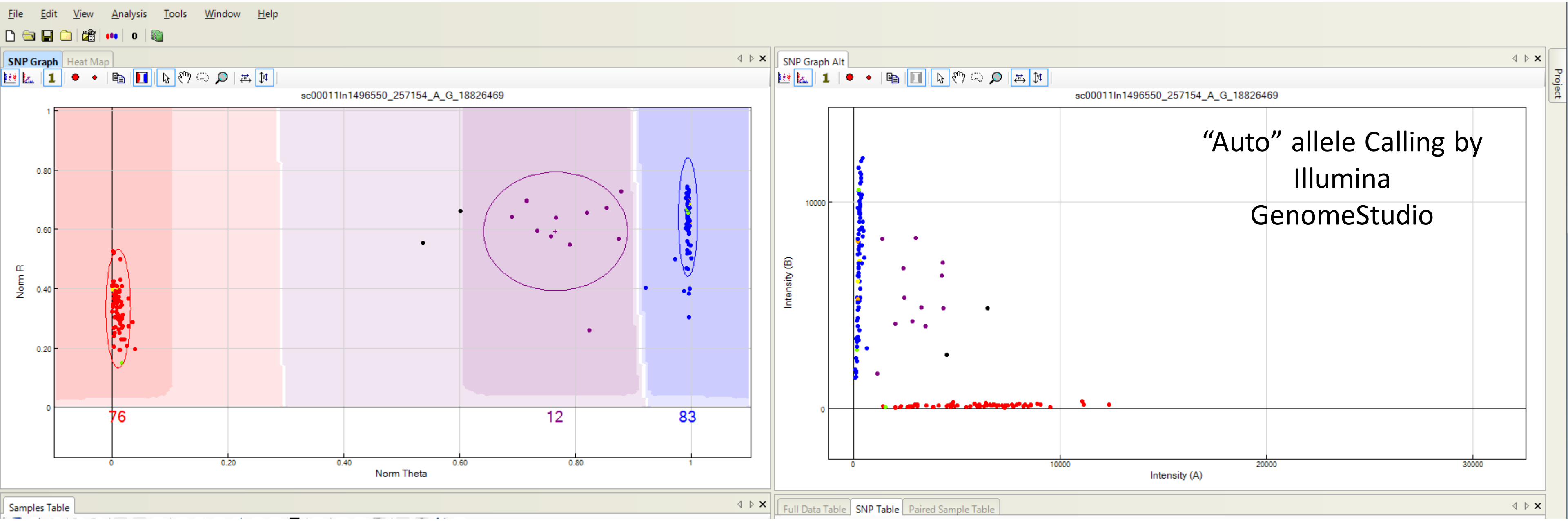


“Auto” allele Calling by Illumina GenomeStudio

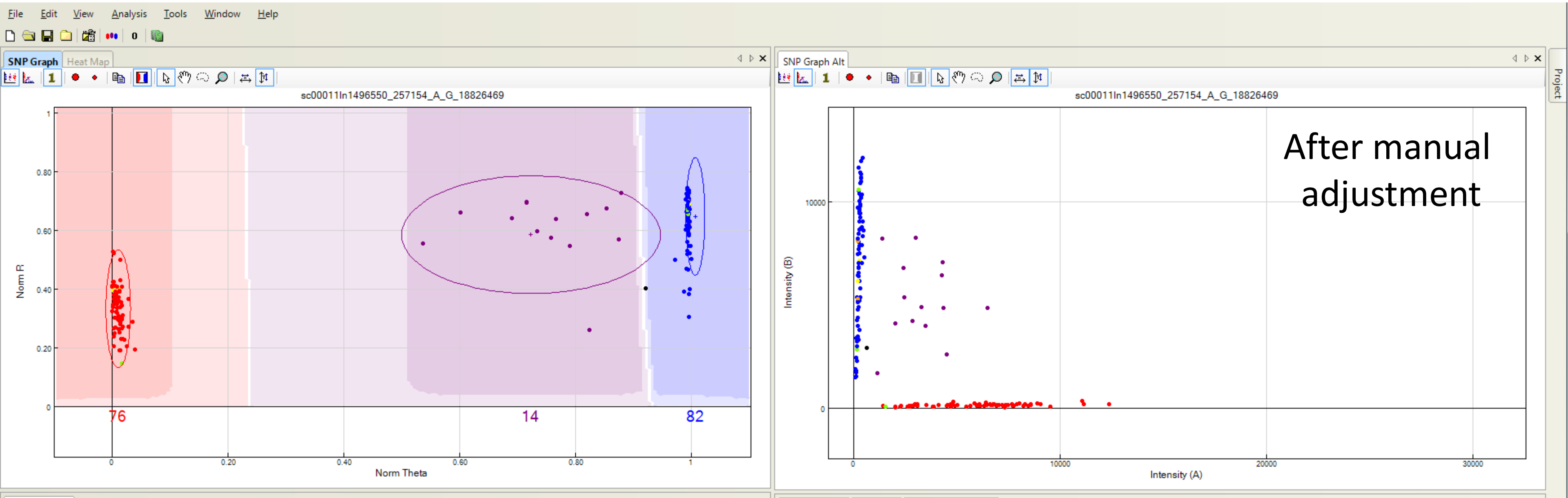


After manual adjustment

Analysis of Illumina Genechip Data Using Illumina GenomeStudio Software
-- Manual Adjustment of Allele Calls --



"Auto" allele Calling by Illumina GenomeStudio



After manual adjustment

Illumina GenomeStudio Software

- We have purchased licenses for 10 copies of GenomeStudio for distribution among the BeanCAP PIs and Collaborators

- We have written up a brief protocol for installation and formatting the data viewing to be similar to the view we have found most useful

- Illumina has agreed to provide a webinar to give basic information on the analysis of genotype data using GenomeStudio