

# Copy Number Variant Detection Using Agilent CGH Platforms

**Esteban Braggio**

**Mayo Clinic Arizona, AZ**



**Scottsdale, Arizona**



**Rochester, Minnesota**



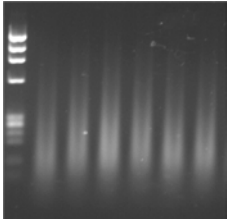
**Jacksonville, Florida**

# Outline

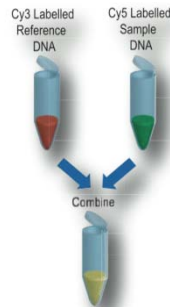
## 2. Genome amplification (Qiagen, Amersham)

## 1. DNA (Storage, QC)

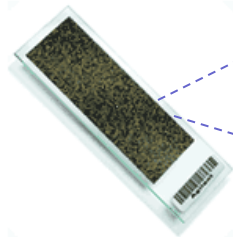
## 3. DNA digestion (Alu\_Rsa, DNase I)



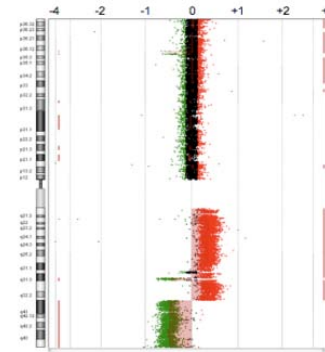
## 4. Sample labeling (Invitrogen + Cy or Alexa dyes, Agilent)



## 5. Platform comparison (Agilent 244K, 1M, Affymetrix SNP6.0)



## 6. Data analysis (CNV identification)



Quantitate / Normalize

Washes/Scanning



# Outline Platform Transition

## ■ Platform Comparison: Agilent 244A, 1M and Affymetrix SNP6.0

*Resolution, Gene and Exon Coverage, Data Compression*

Five samples run in the three platforms

- 1 normal male
  - Test log ratio shift on chromosome X (ie. Amplitude Test)
  - Test CNV detection efficiency across the genome
- 1 human myeloma cell line (Delta47)
  - Has the largest number of abnormalities at 244k resolution
  - Test abnormality detection
- 3 human myeloma patient samples (Random Selection)
  - Test abnormality detection efficiency
  - Test CNV detection efficiency

## ■ CNV identification

*TCAG database*

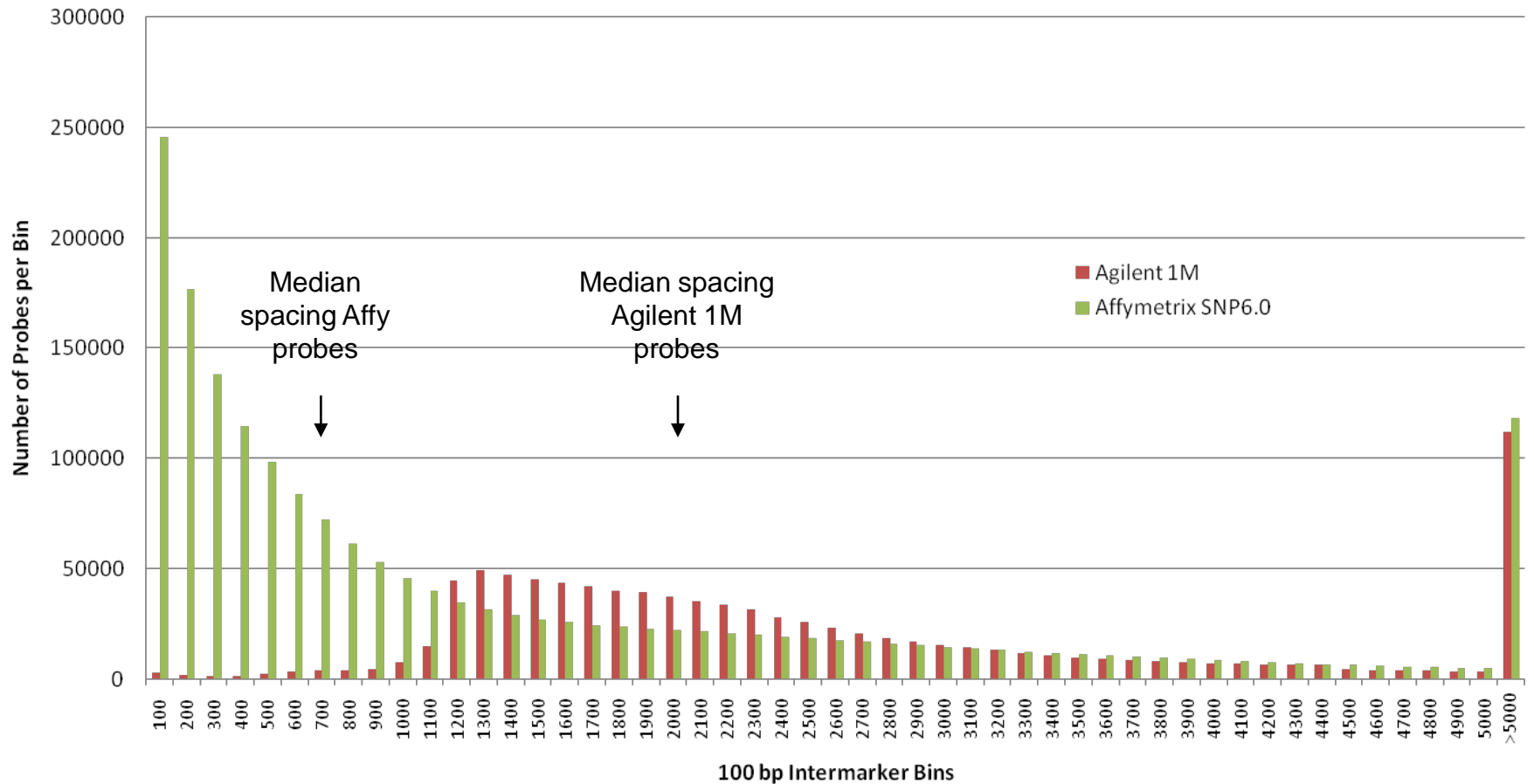
*Custom list*

# Platform Information

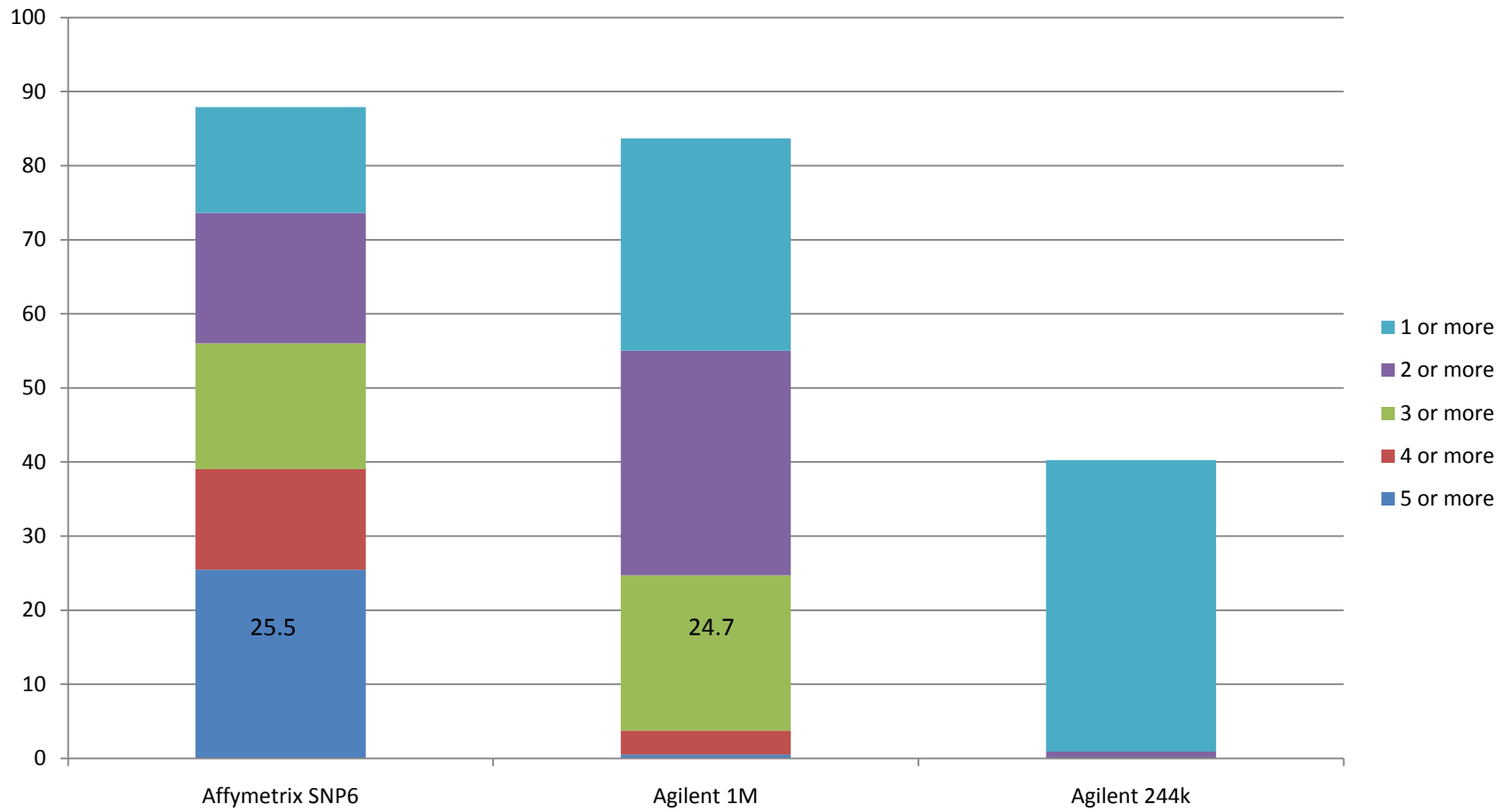
- Agilent Platforms
  - Two-color CGH array
  - Random primer labeling of fragment genomic DNA
- Affymetrix SNP6.0
  - One-color SNP and CN analysis platform
  - Linker-adaptor directed labeling of NspI and Styl fragmented genomic DNA

| Platform          | Mapped Unique Probes | Mean Inter-marker Spacing | Median Inter-marker Spacing | Inter-marker Spacing IQR Range | Inter-marker Spacing IQR Mean | Inter-marker Spacing IQR SD |
|-------------------|----------------------|---------------------------|-----------------------------|--------------------------------|-------------------------------|-----------------------------|
| Affymetrix SNP6.0 | 1,854,093            | 1617                      | 694                         | 228-888                        | 502                           | 186                         |
| Agilent 1M        | 960,431              | 3115                      | 2108                        | 1504-3194                      | 2176                          | 456                         |
| Agilent 244A      | 235,829              | 12,689                    | 8931                        | 6356-13,345                    | 9183                          | 1900                        |
| Agilent 44B       | 42,416               | 70,462                    | 37,690                      | 18,689-81,036                  | 41,212                        | 16,400                      |

# Inter-Probe Distance Comparison

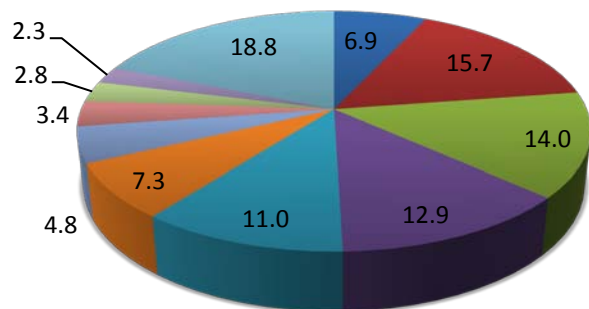


# Percentage of Sequential 5kb Intervals with Specific Probe Coverage Densities

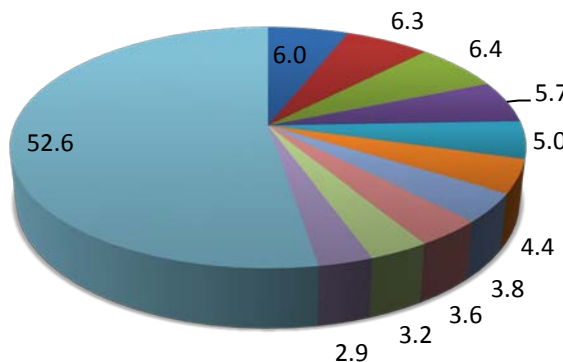


# Percentage of Genes with Coverage

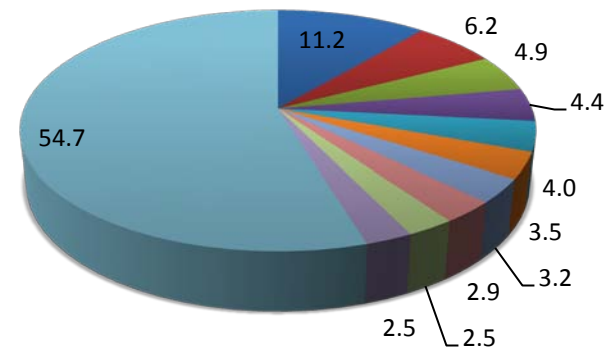
**Agilent 244A**



**Agilent 1M**



**Affymetrix SNP6.0**

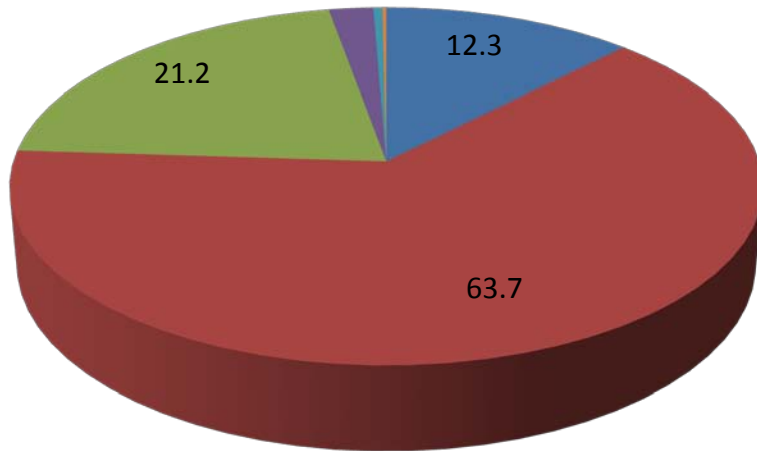


| Probes per Gene | Agilent 244A | Agilent 1M | Affymetrix SNP6 |
|-----------------|--------------|------------|-----------------|
| No Coverage     | 6.9          | 6.0        | 11.2            |
| 1 or more       | 93.1         | 94.0       | 88.8            |
| 2 or more       | 77.4         | 87.7       | 82.7            |
| 3 or more       | 63.3         | 81.2       | 77.7            |
| 4 or more       | 50.5         | 75.6       | 73.3            |
| 5 or more       | 39.5         | 70.6       | 69.3            |

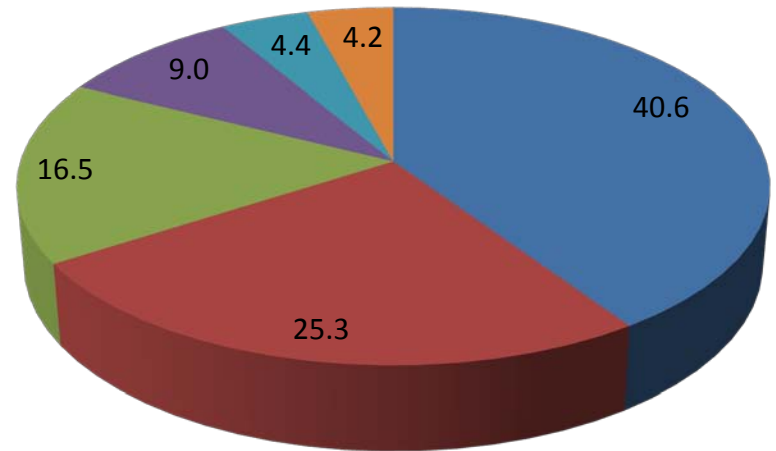


# Exon Level Coverage

**Agilent 1M**



**Affymetrix SNP6.0**



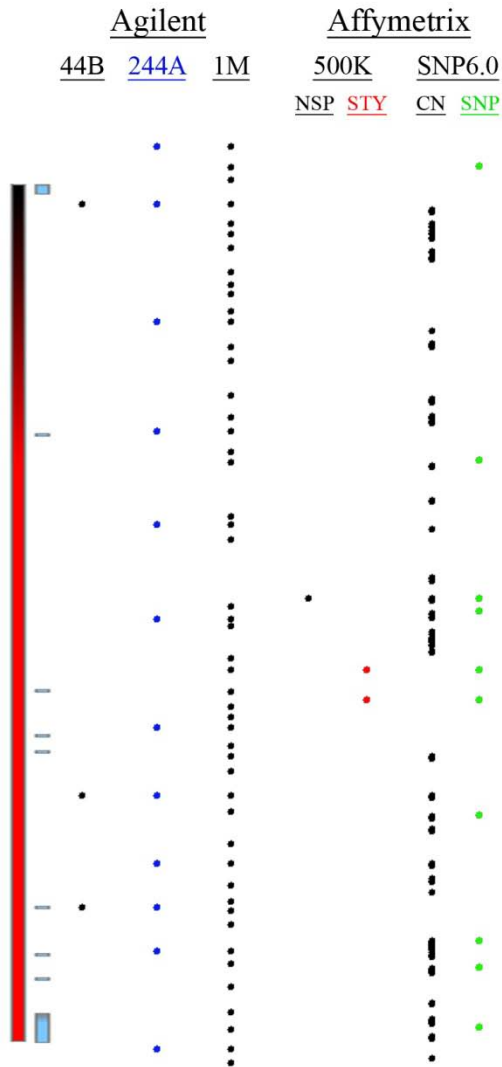
■ No Coverage ■ 1 Probe ■ 2 Probes ■ 3 Probes ■ 4 Probes ■ 5 Probes or more

| Probes per Exon | Agilent 1M | Affymetrix SNP6.0 |
|-----------------|------------|-------------------|
| No Coverage     | 12.3       | 40.6              |
| 1 or more       | 87.7       | 59.4              |
| 2 or more       | 24.0       | 34.1              |
| 3 or more       | 2.8        | 17.6              |
| 4 or more       | 0.6        | 8.6               |
| 5 or more       | 0.2        | 4.2               |

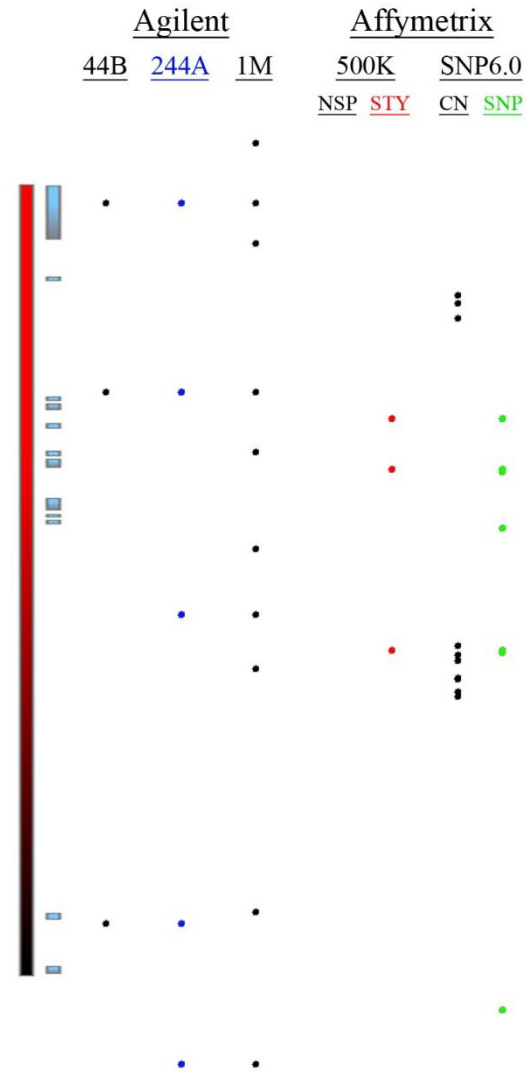


# Probe Distribution

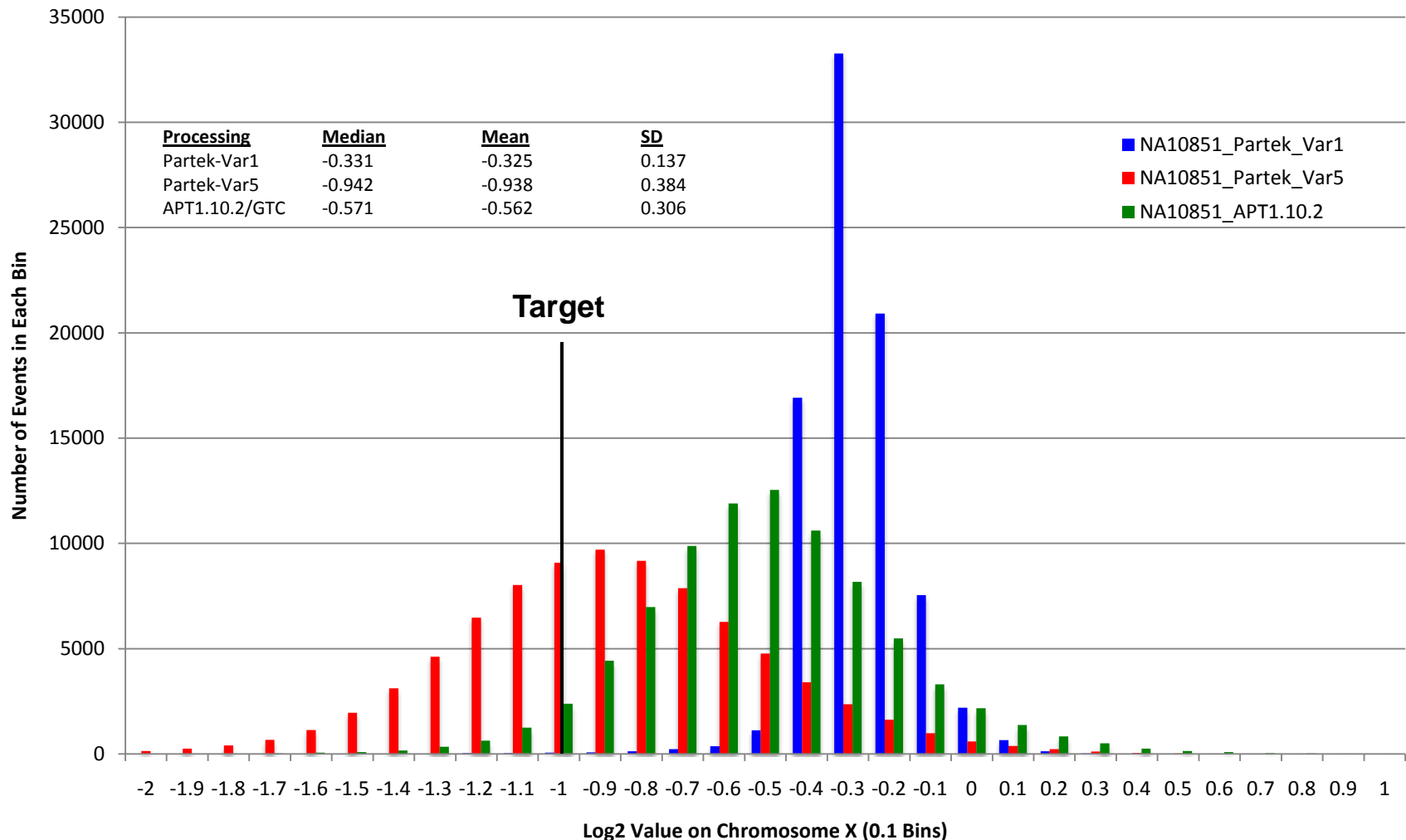
## PTEN



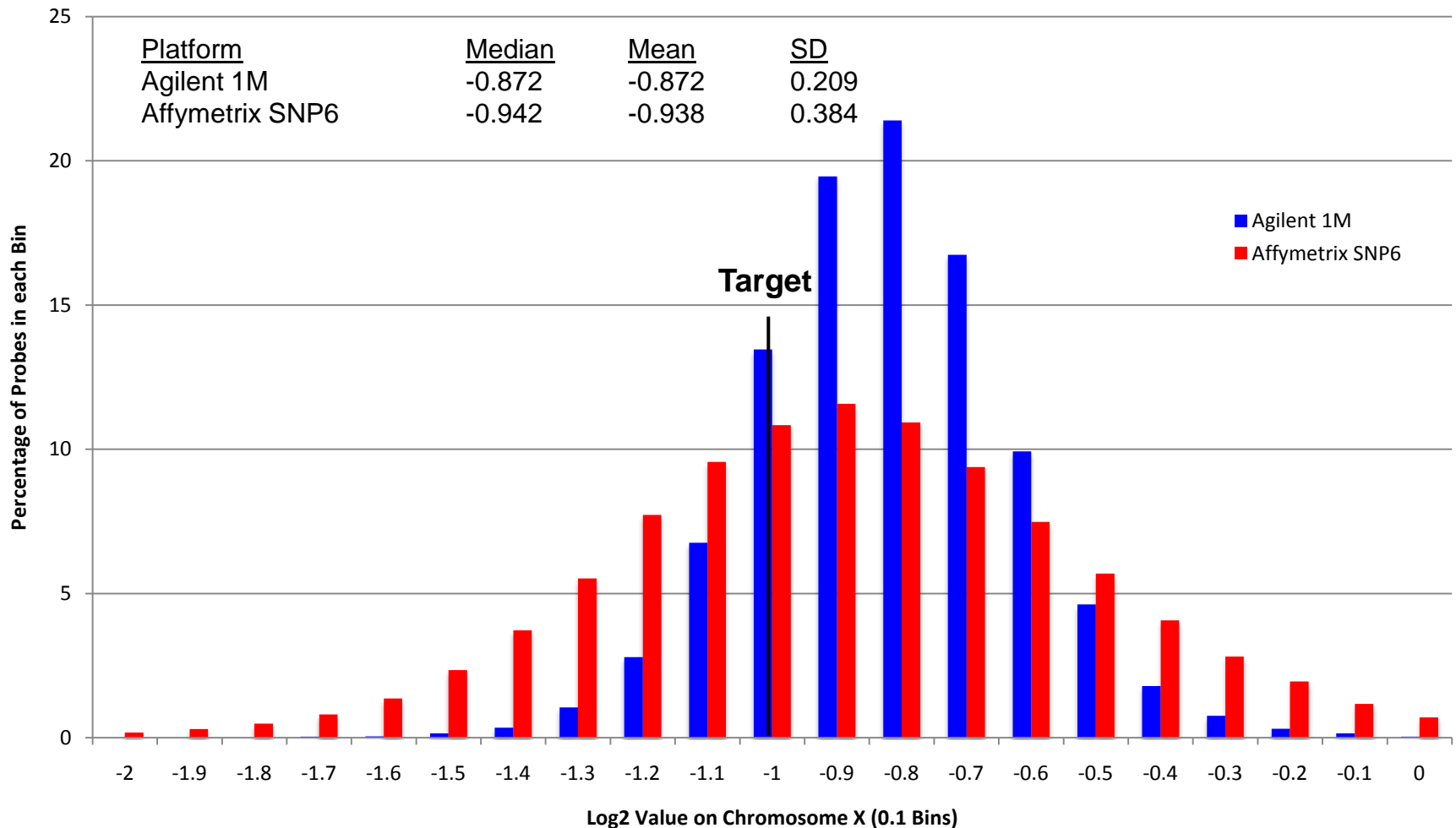
## TP53



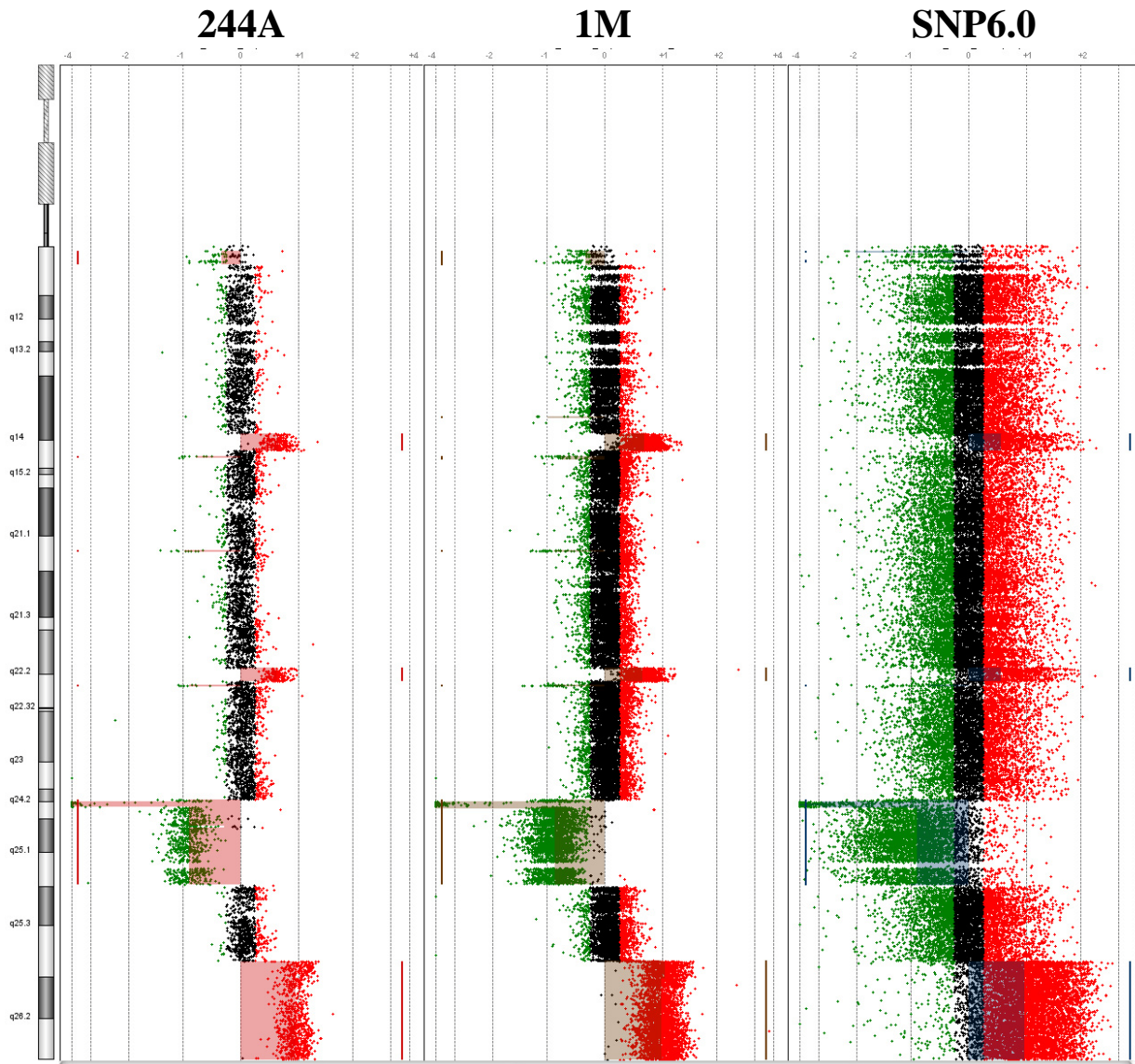
# Amplitude and Noise of Affymetrix Data is Method Dependent



# Comparison of Data Compression



# Visual Comparison of Platforms

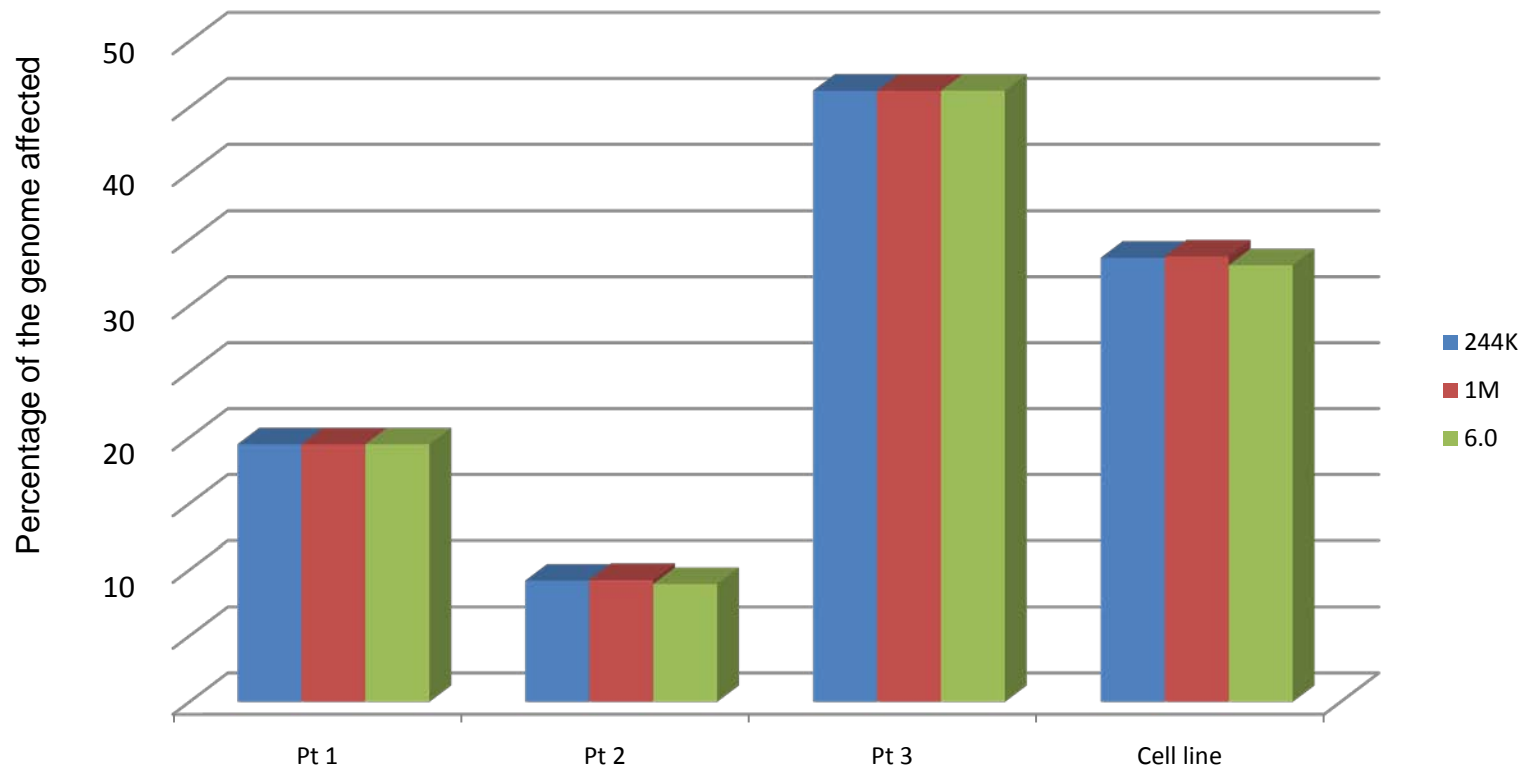


# There are no Significant Differences in the % of Genome Affected in Abnormalities...

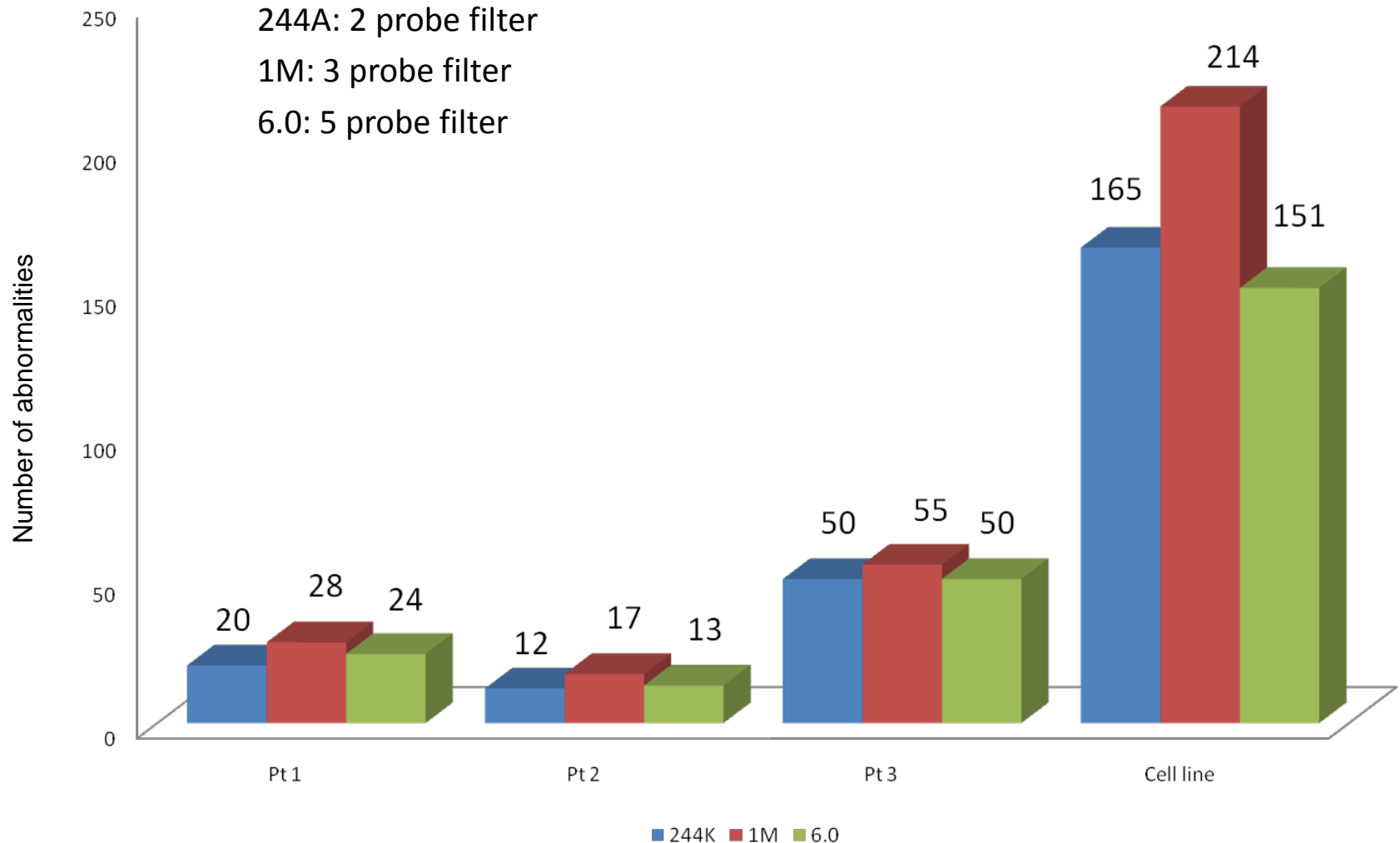
244A: 2 probe filter

1M: 3 probe filter

6.0: 5 probe filter

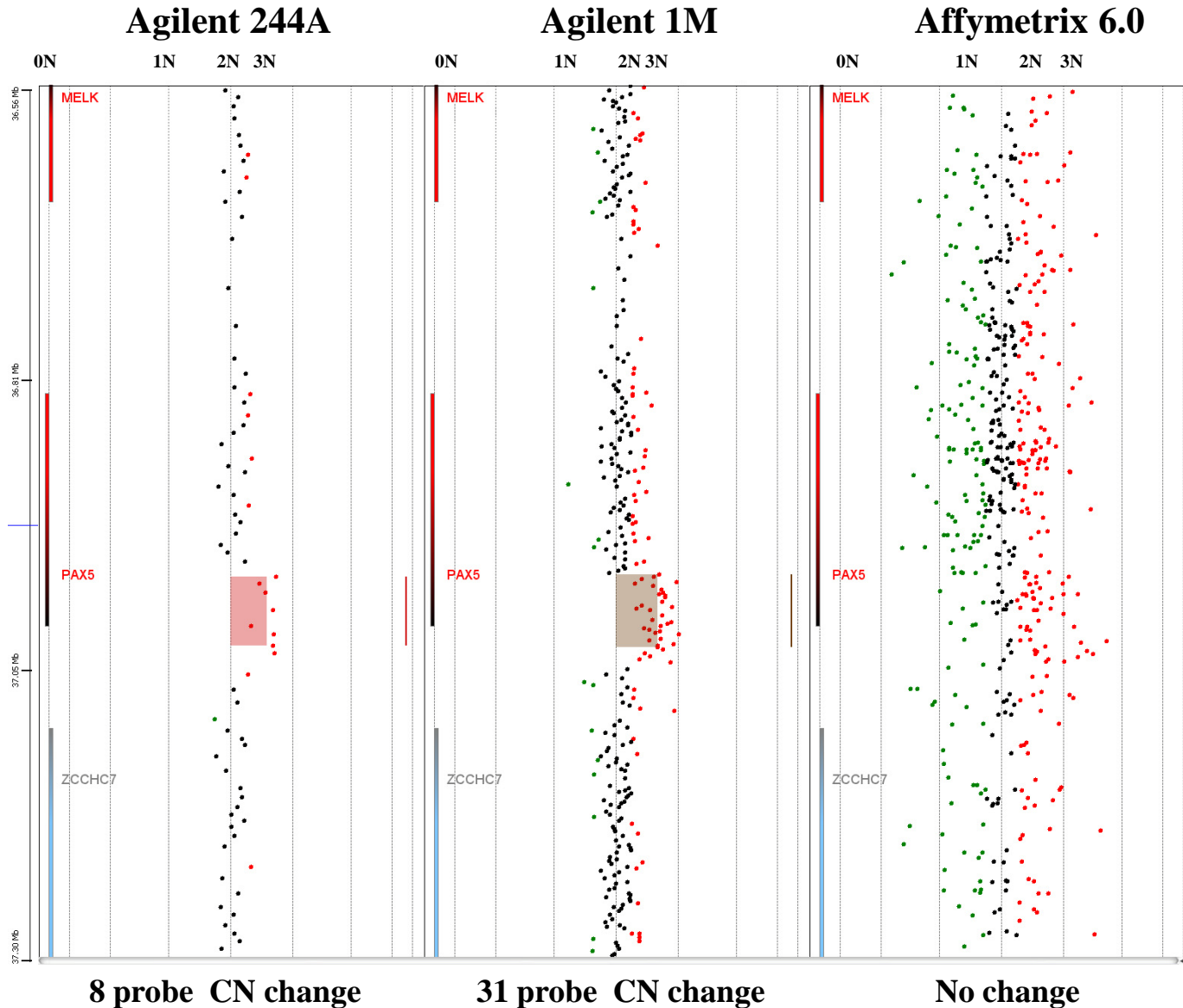


# ...but there are Differences in the Number of Abnormalities Detected



# Gene Level Comparison

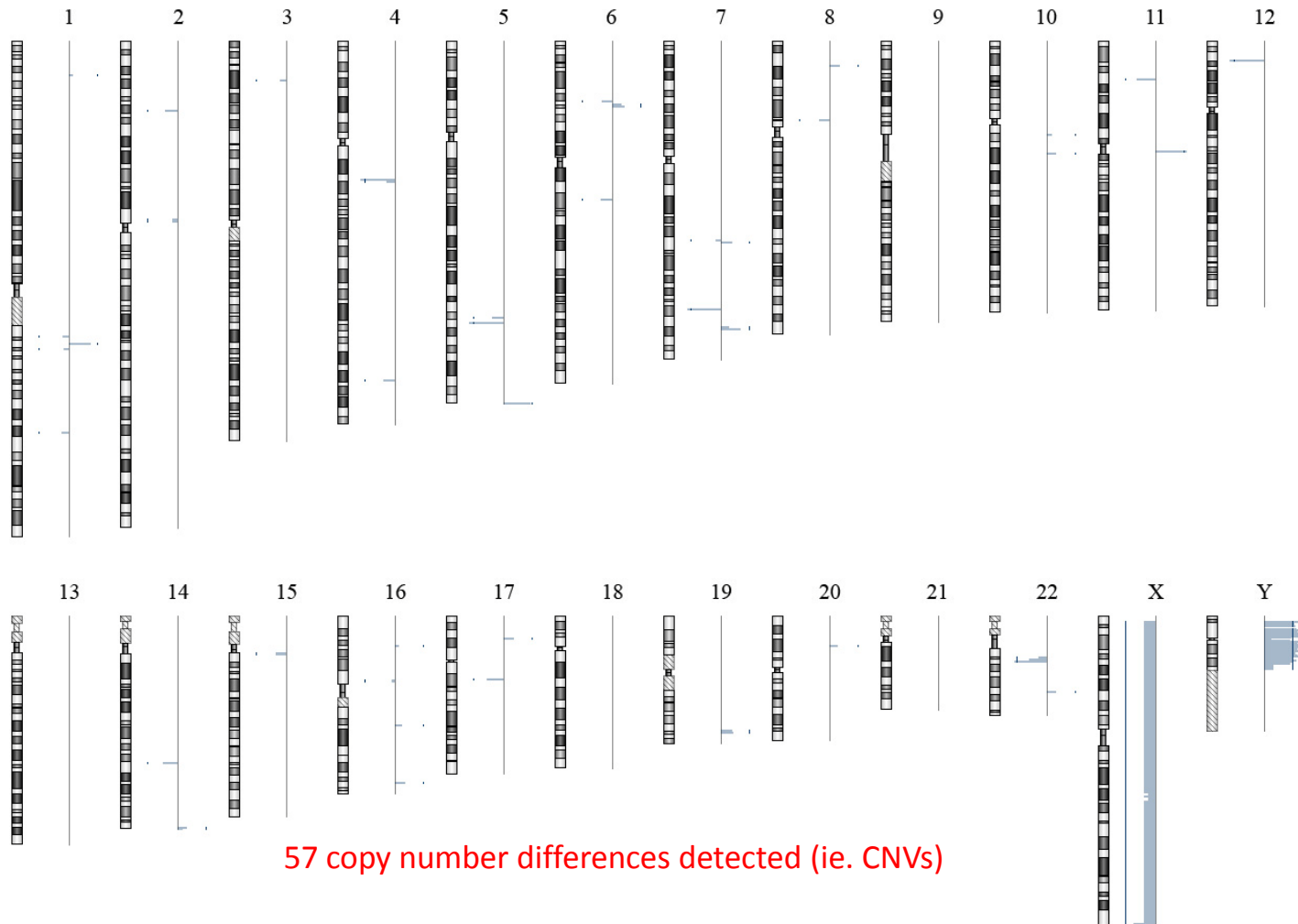
## Delta 47 cell line



# Comparison of Agilent Platforms

More resolution is also associated with more CNVs

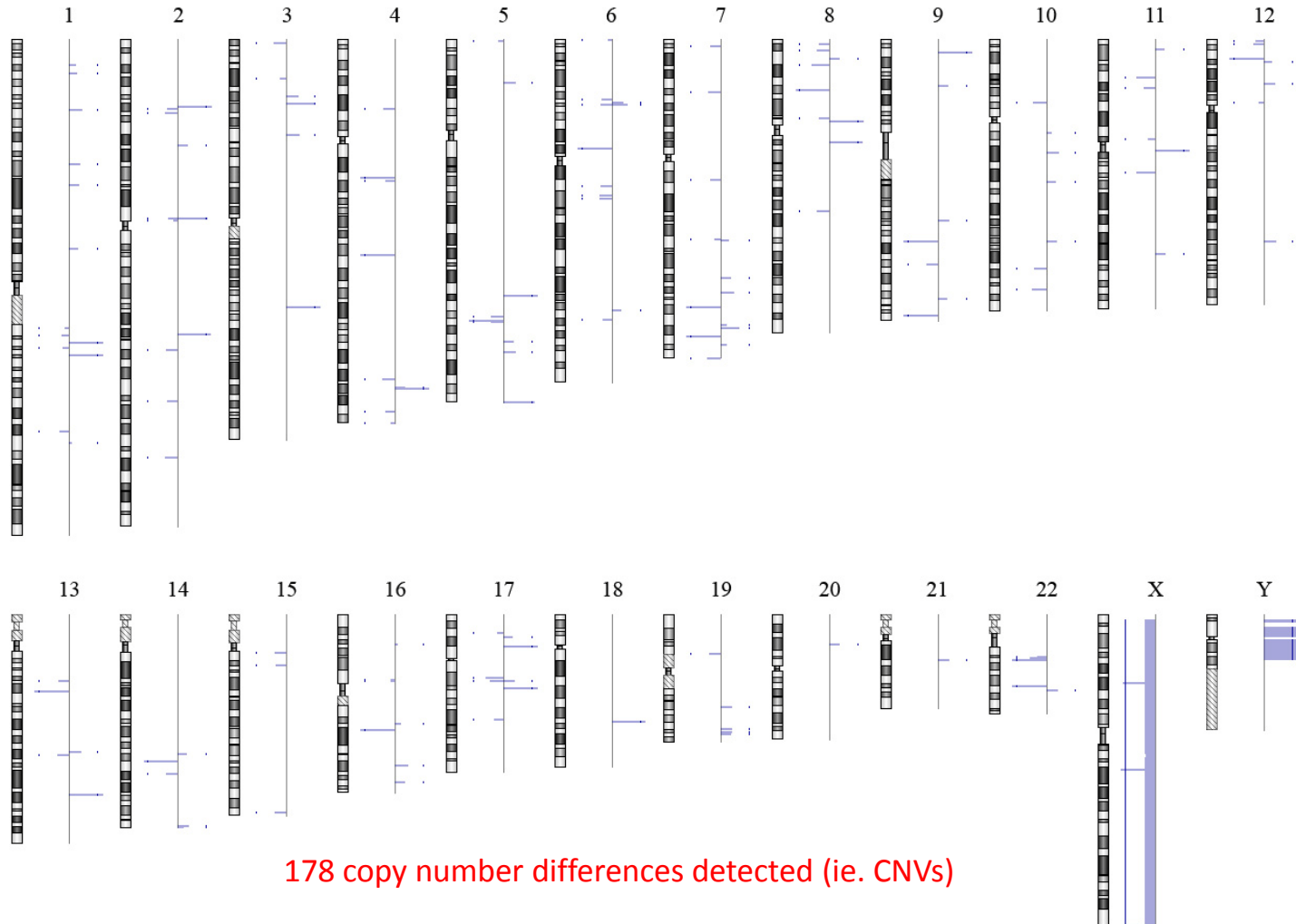
## Agilent 244k NA10851 vs NA15510



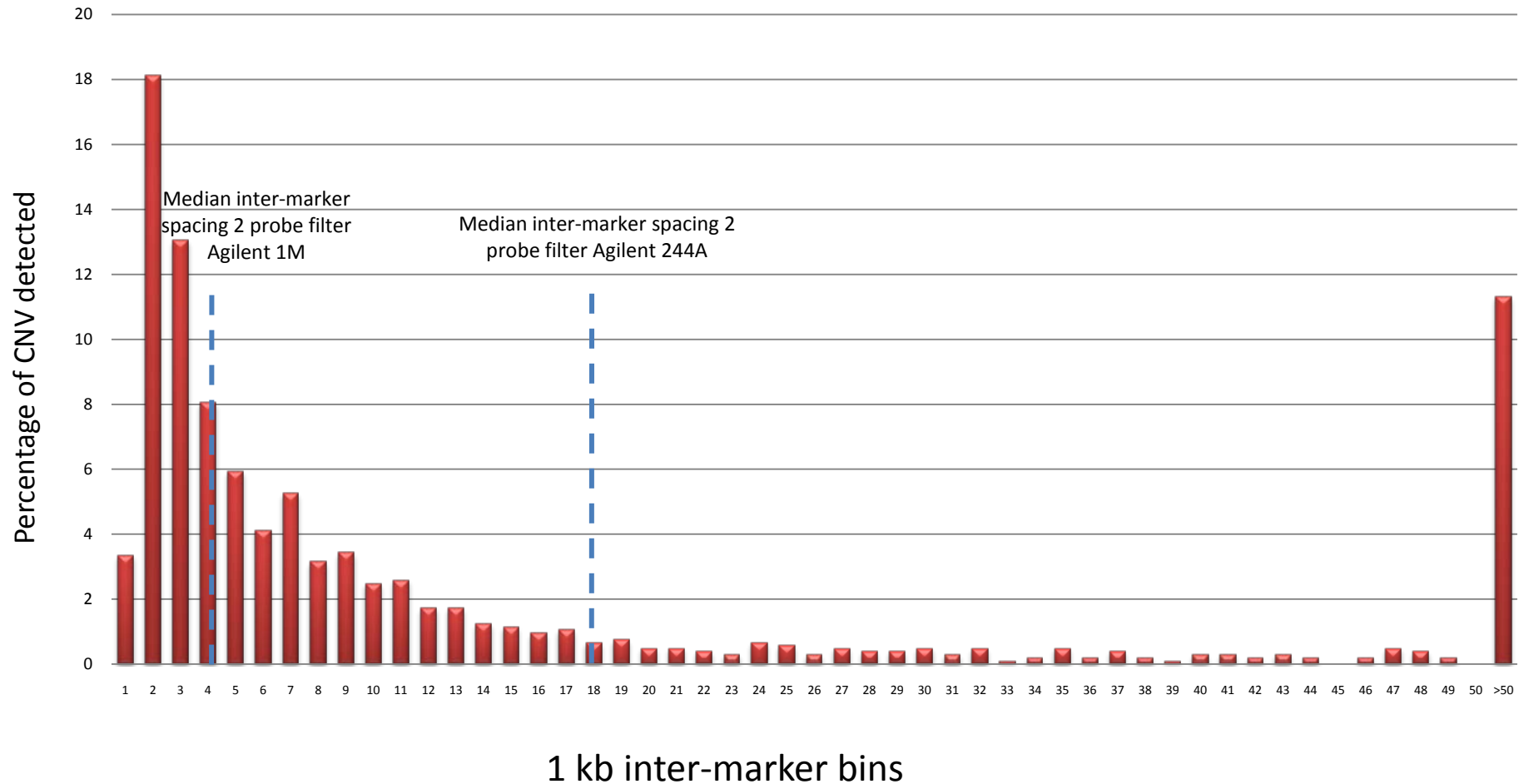


# Comparison of Agilent Platforms

## Agilent 1M NA10851 vs NA15510



# CNV Size Distribution Using Agilent 1M



# Optimizing CNV Identification

## TCAG database

### Database of Genomic Variants

A curated catalogue of structural variation in the human genome

Hosted by:  
The Centre for  
Applied Genomics

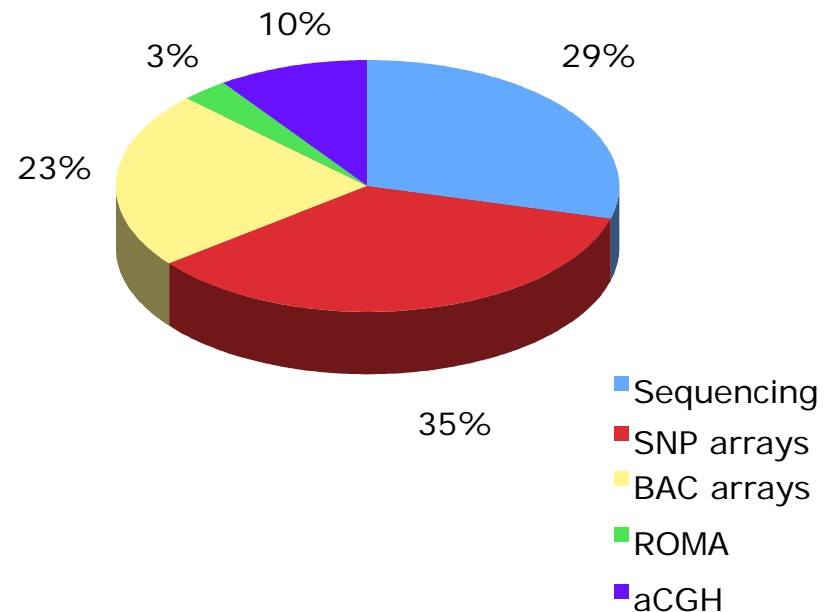


[About The Project](#) | [Genome Browser](#) | [Download](#) | [Links](#) | [Data Submissions](#) | [Email us](#)

Please select genome assembly:

#### Summary Statistics

Total entries: **49988** (hg18)  
CNVs: 29133  
Inversions: 914  
InDels (100bp-1Kb): 19941  
Total CNV loci: 8410  
Articles cited: **35**  
Last updated: Aug 05, 2009  
Join our [mailing list](#)



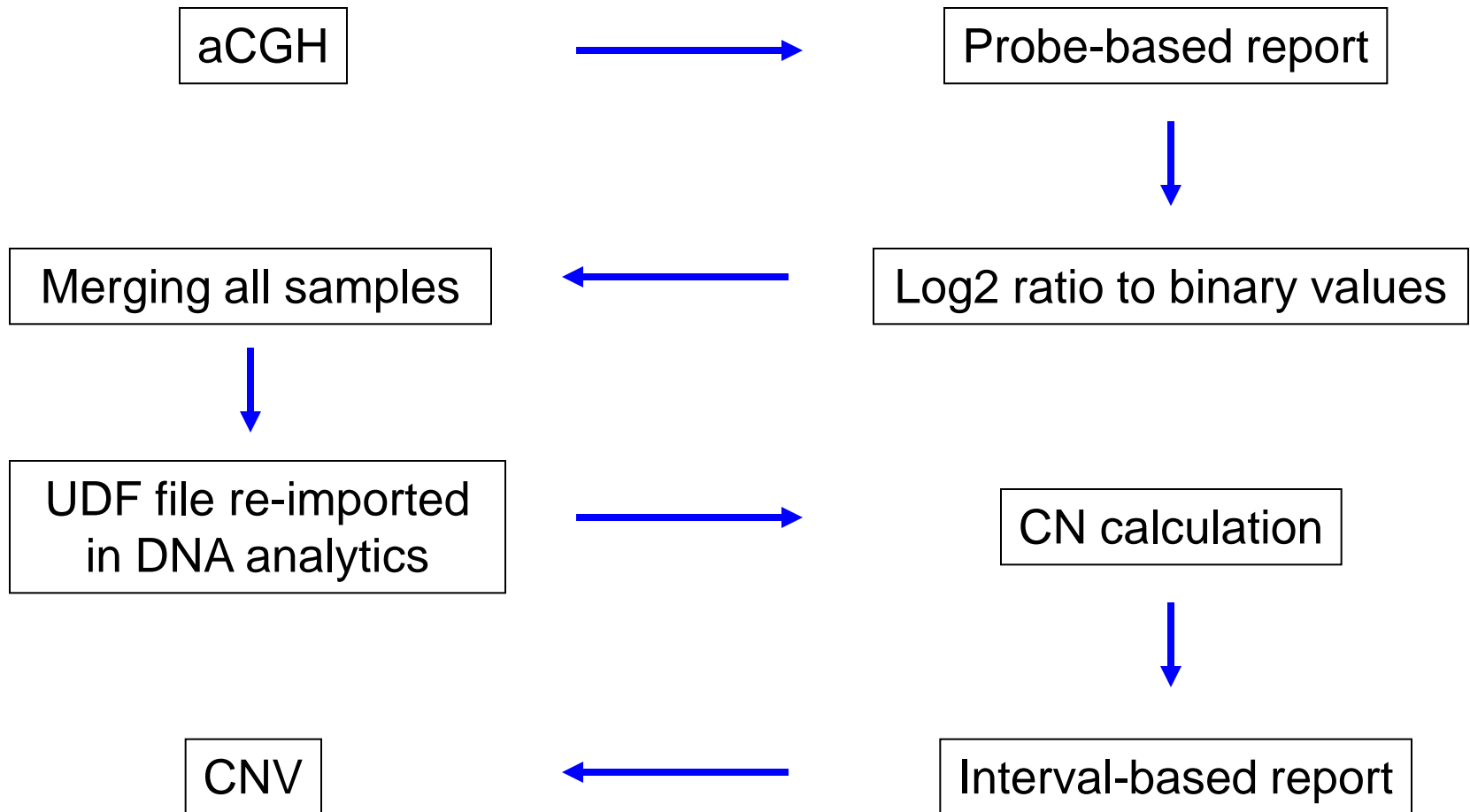
31.9% of the genome excluded from the analysis

# Optimizing CNV Identification

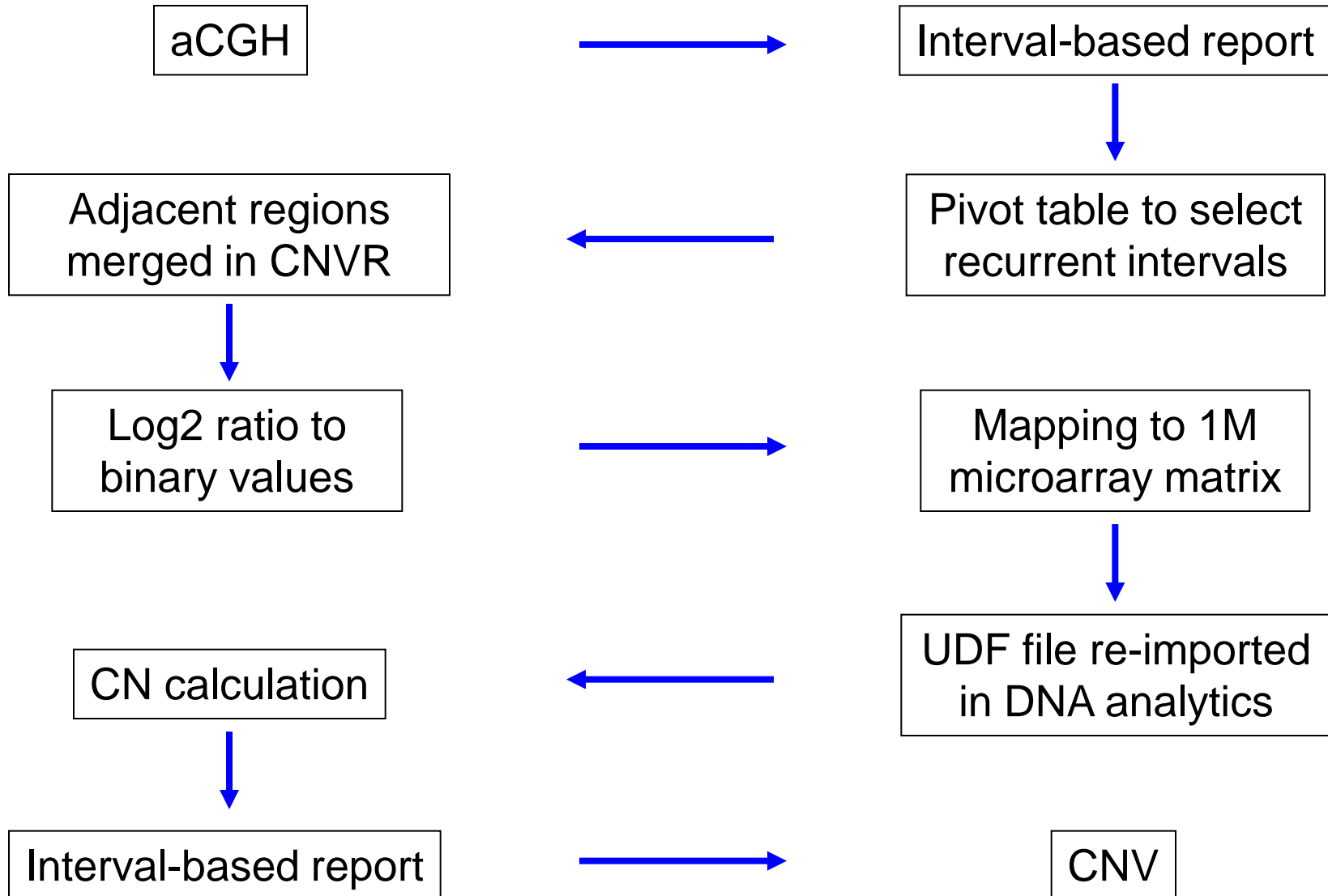
## Mayo Custom CNV list

- 9 HapMap (2 *CEU*, 2 *YRI*, 2 *JPT*, 2 *CHB*, 1 *unknown*) and 109 tumor samples (*CLL*, *MCL*, *TCL*, *MM*, *WM*, *BL*, *Agilent demo data*)
- Reference: mix consisting on 9 female HapMap samples
- Filters: 2 probe\_0.25log2. ADM-2 algorithm threshold of 5.0
- Centralization and fuzzy zero OFF
- Chromosome Y was excluded from the analysis.

# 1. HapMap Samples



## 2. Tumor Samples



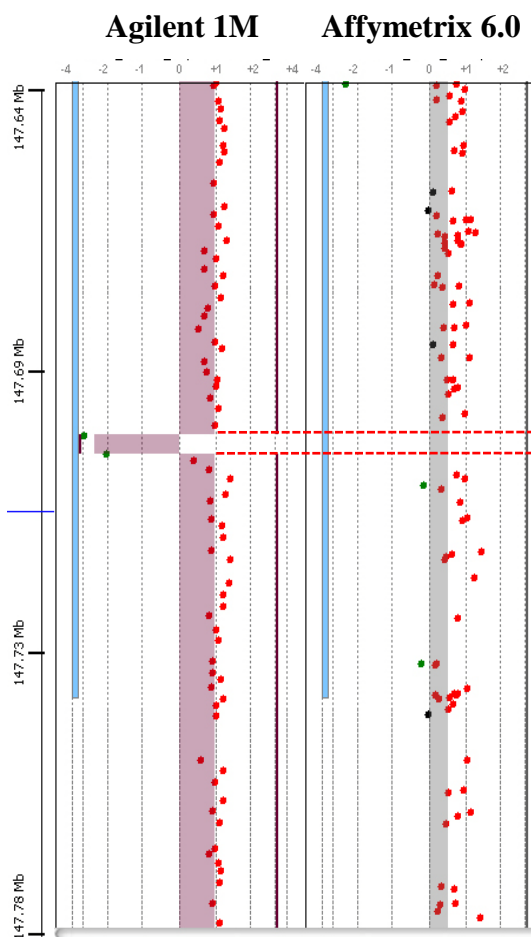
# Merged Master List



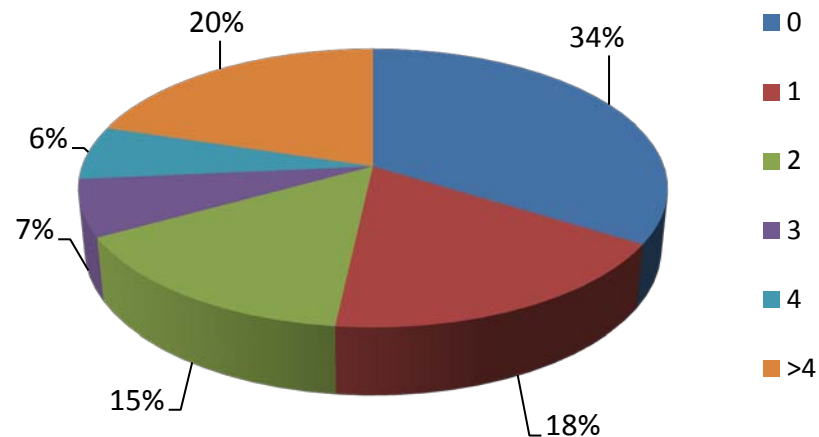
- Median size of 5.3Kb (range 0.05Kb - 2,900Kb)
- 7.3% CNVs larger than 100Kb and 0.8% larger than 1Mb
- 418 genes included in CNVR

# 257 Novel Recurrent CNVs

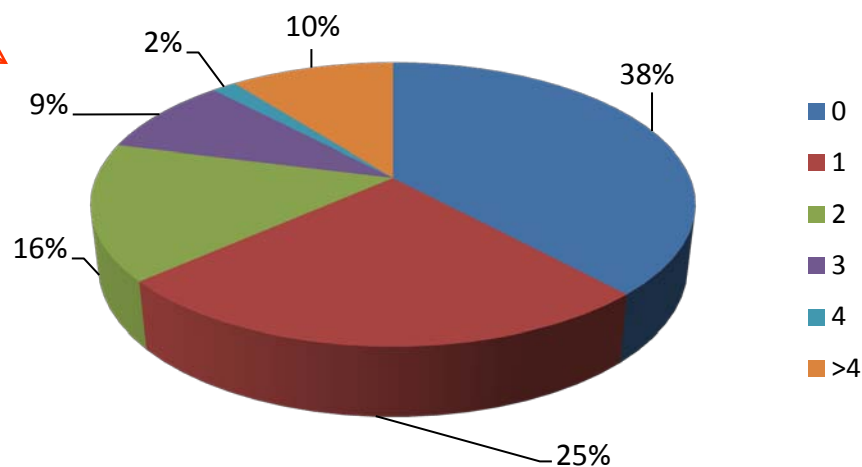
Probe coverage per Agilent-based CNV discovery



**Affy SNP6.0**



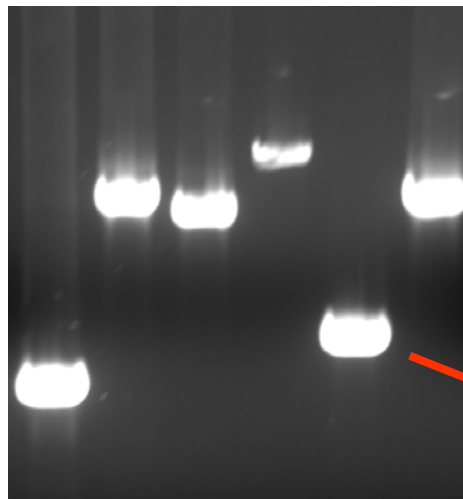
**Illumina 1M**





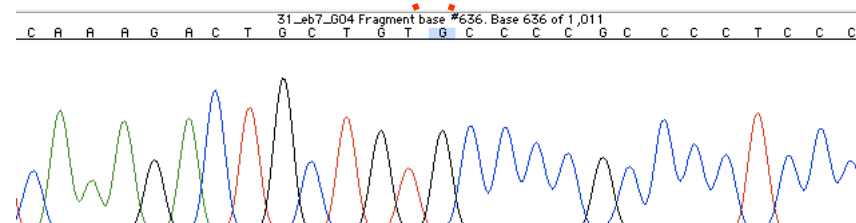
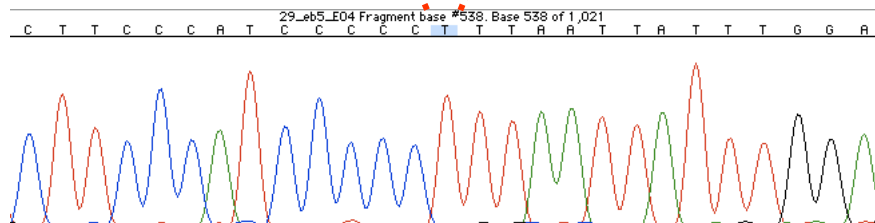
# Validation

CNV\_1090   CNV\_634   CNV\_809

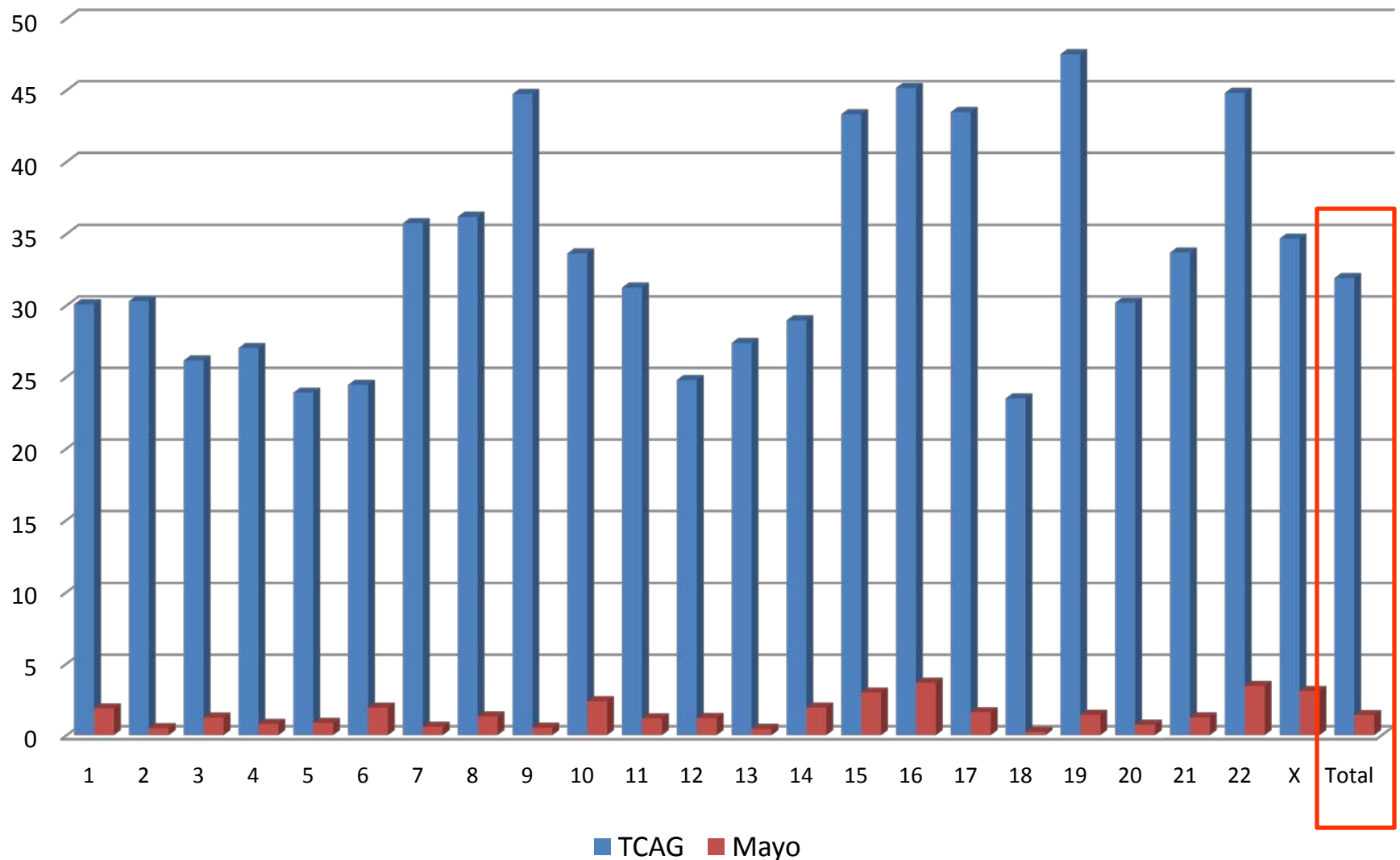


1952 bp deletion

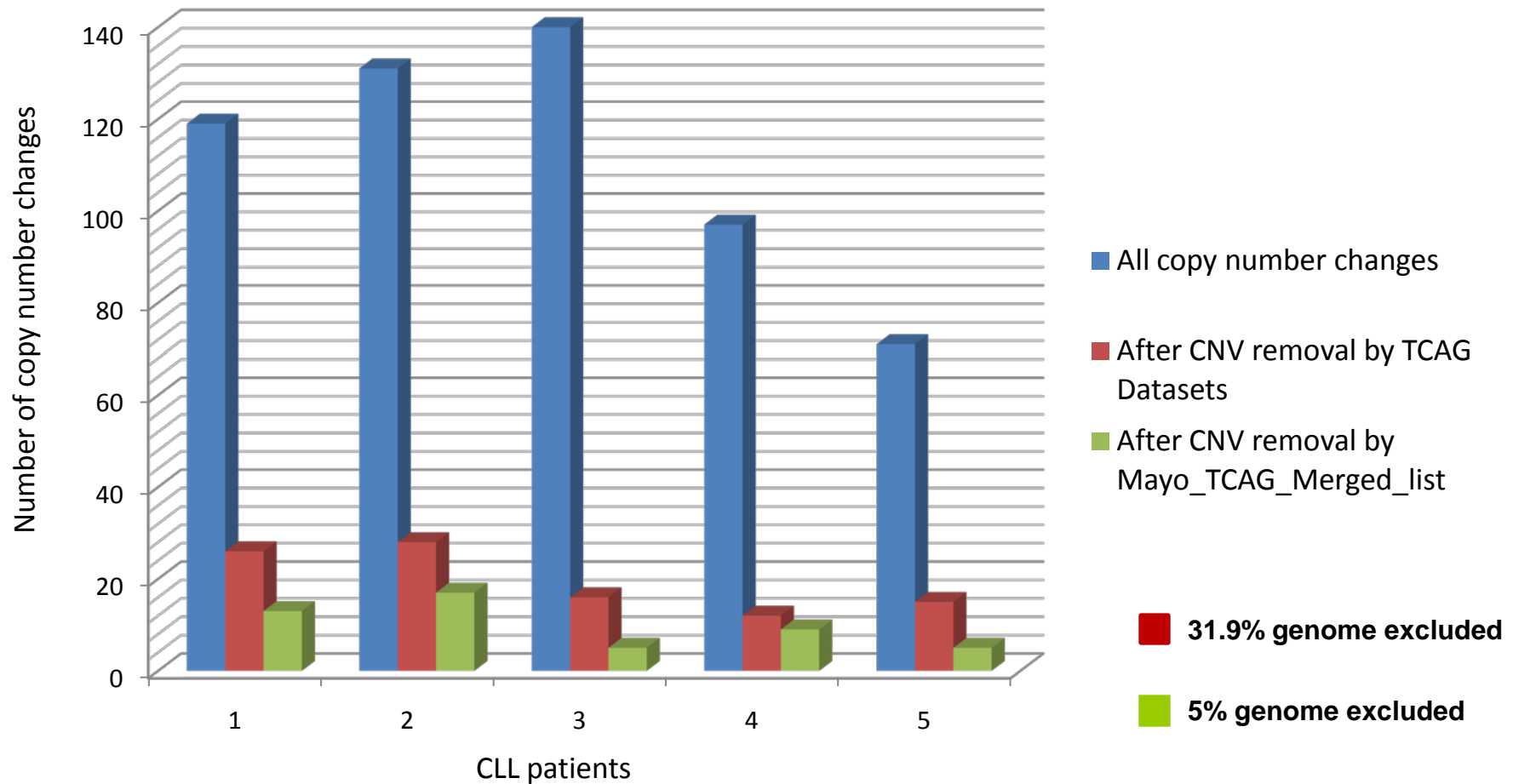
1617 bp deletion



# Percent of Genome Included in CNVs



# Mayo TCAG Merged CNV list\*



\* Mayo CNV list + highest resolution TCAG data (sequencing + SNP6.0 in 270 HapMap samples)

## Concluding Remarks

- Affymetrix versus Agilent platforms: more probe density does not mean more resolution.
- In the transition from Agilent 244K to 1M, the significant increment of CNVs needs to be contemplated and it is crucial the CNV annotation before data analysis.

# Acknowledgements

Rafael Fonseca

Jonathan J. Keats

Mike Barrett

Leif Bergsagel

Scott Van Wier

John Carpten



Affymetrix

Agilent

Invitrogen