

Utilization of low cost multiplex SNP genotyping techniques

Robert A. Welch

Deputy Scientific Director

Core Genotyping Facility, National Cancer Institute

Overview

- Samples (Human)
 - Collection
 - DNA Extraction / Purification
 - Quantification
 - Quality Assessment
- Genotyping Technologies
 - Choice is dependent on the study design / purpose
 - “Single-plex”
 - Medium Multiplex
 - Large Multiplex
 - Whole Genome

Samples pre-genotyping lab

■ Samples

■ Collection and storage

Genotypability ↑

- Immortalized Cell Lines
- Buffy Coat (if prepared correctly)
- Whole Blood
- Buccal cells
- Serum
- Plasma

■ DNA Extraction / Purification (and Aliquoting)

- Extraction and purification method should match sample type (genotyping lab needs this information)
- Aliquots should contain moderate mass, volume & concentration

Samples – genotyping lab

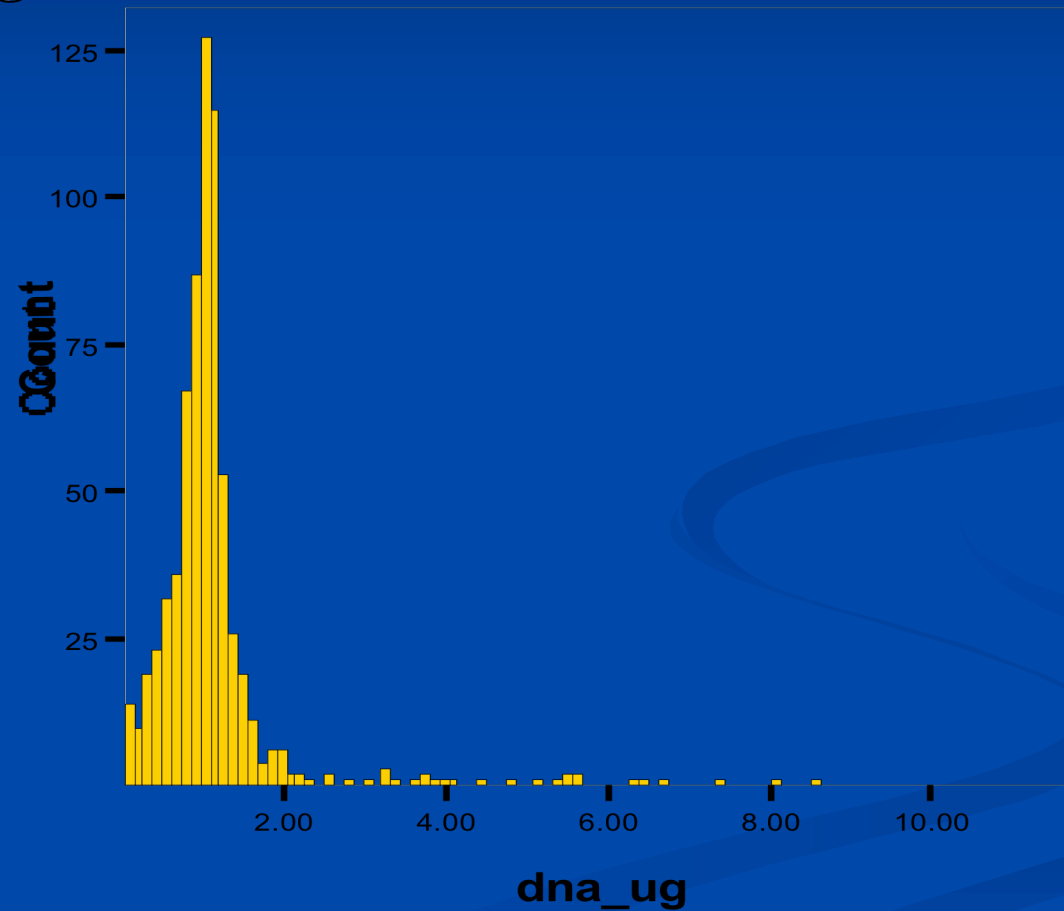
DNA -- Quantity

- Optical Density
 - Measure absorbance of light at 260 and 280 nm
 - Single-stranded nucleic acids and proteins can increase signal
 - Compounds commonly used in the extraction / preparation of nucleic acids absorb at 260 nm leading to abnormally high quantitation levels.
- Fluorescent intercalating dyes (PicoGreen)
 - PicoGreen is an intercalating fluorophore that selectively binds double-stranded DNA (dsDNA)
 - Protein and contaminants will not be detected
 - All dsDNA will be detected (human/non-human)
 - Requires generation of standard curve (standard curves will differ between laboratories λ vs human DNA standards)
- Real Time PCR
 - Semi-quantitative (TaqMan) assay designed for human specific loci
 - Measures amount of human specific DNA participating in PCR
 - Accuracy depends on the generation of a standard curve

Haque et al BMC Biotechnol 2003

DNA Quantitation

Target: 2 ug total mass



Biorepository deliveries of DNA have excess variance

DNA -- Quality

- TaqMan real-time quantification (will the sample 'work' in high throughput genotyping)
- AmpFLSTR® Identifiler®
 - Microsatellite multiplex assay(15 STRs + Amelogenin)
 - Intensity data allows confirmation of quantification.
 - Complex assay provides evidence for future sample success
 - Profile creates genetic fingerprint
 - **Compare determined gender to reported gender**
 - **Provides data to compare sample replacements and check duplicates/replicates (intentional or not)**
 - **Indicator of possible sample contamination**
- Other quality measures ?

Is Whole Genome Amplification (WGA) an alternative?

- Scarce DNA quantity can be recovered

Large scale analysis can proceed to yield 5-40 ug of WGA product from as little 10 ng but with caveats (>50ng preferred)

Performance could be platform specific

High fidelity SNP analysis by TaqMan

Slight loss of accuracy with STRs (longer PCR Products)

Higher throughput technologies could suffer biased completion rate differences.

Input DNA critical -- Range: 5-50 ng (the more the better to a point)

- Performance depends on Input

DNA

Quality of input DNA critical

Source of DNA (contaminants in buccal, etc)

Interrogated region / technique

Suggestion that Sequence context especially GC content is important

Size of Fragment (larger STRs < SNPs)

Methodology some methods work better with WGA products than others

Telomeric position of fragments

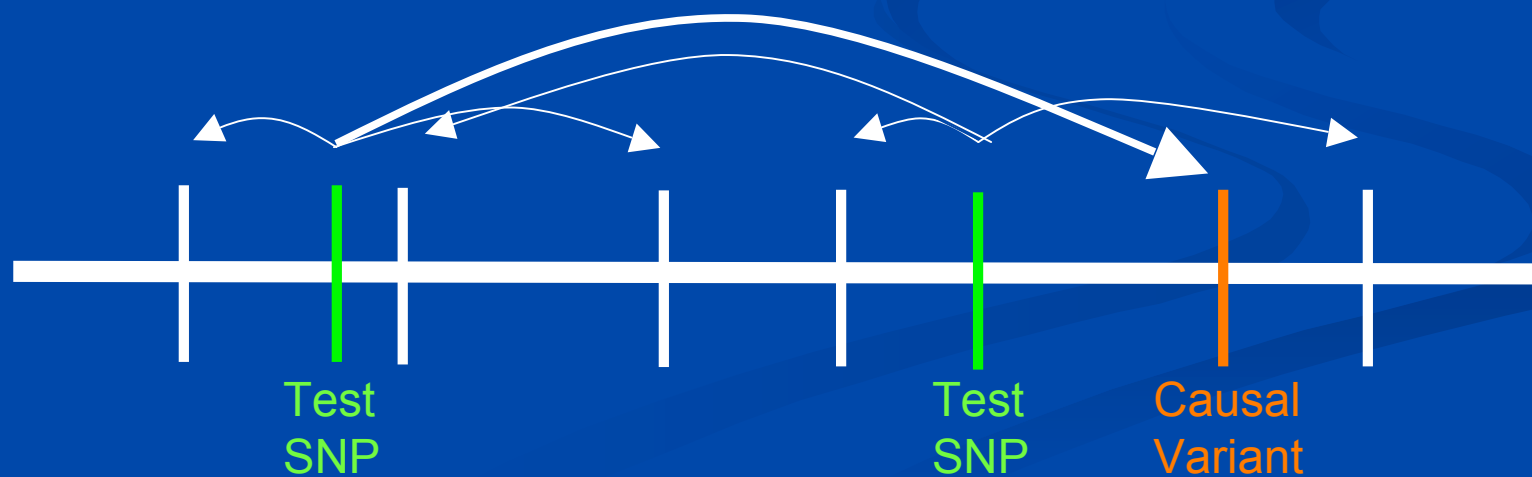
Bergen et al Hum Mut 2005

Bergen et al BMC Biotechnol 2005

Association Studies: Indirect Tests

Indirect approach

- Identity of functional variants not required
- Use underlying pattern of variation to detect SNP Markers
- Test subsets of SNPs that are highly correlated with all unmeasured SNPs (causal alleles)



Genotyping Technologies

- Overview some of current technologies, choice depends on focus of study

Single-plex

TaqMan™ / MGB Eclipse™ / Fragment Sizing (VNTR)

Medium Multiplex

Sequenom iPLEX™ / SNPlex™

Large Multiplex

Illumina GoldenGate™ / Affy/ParAllele™ Targeted Genotyping

Whole Genome

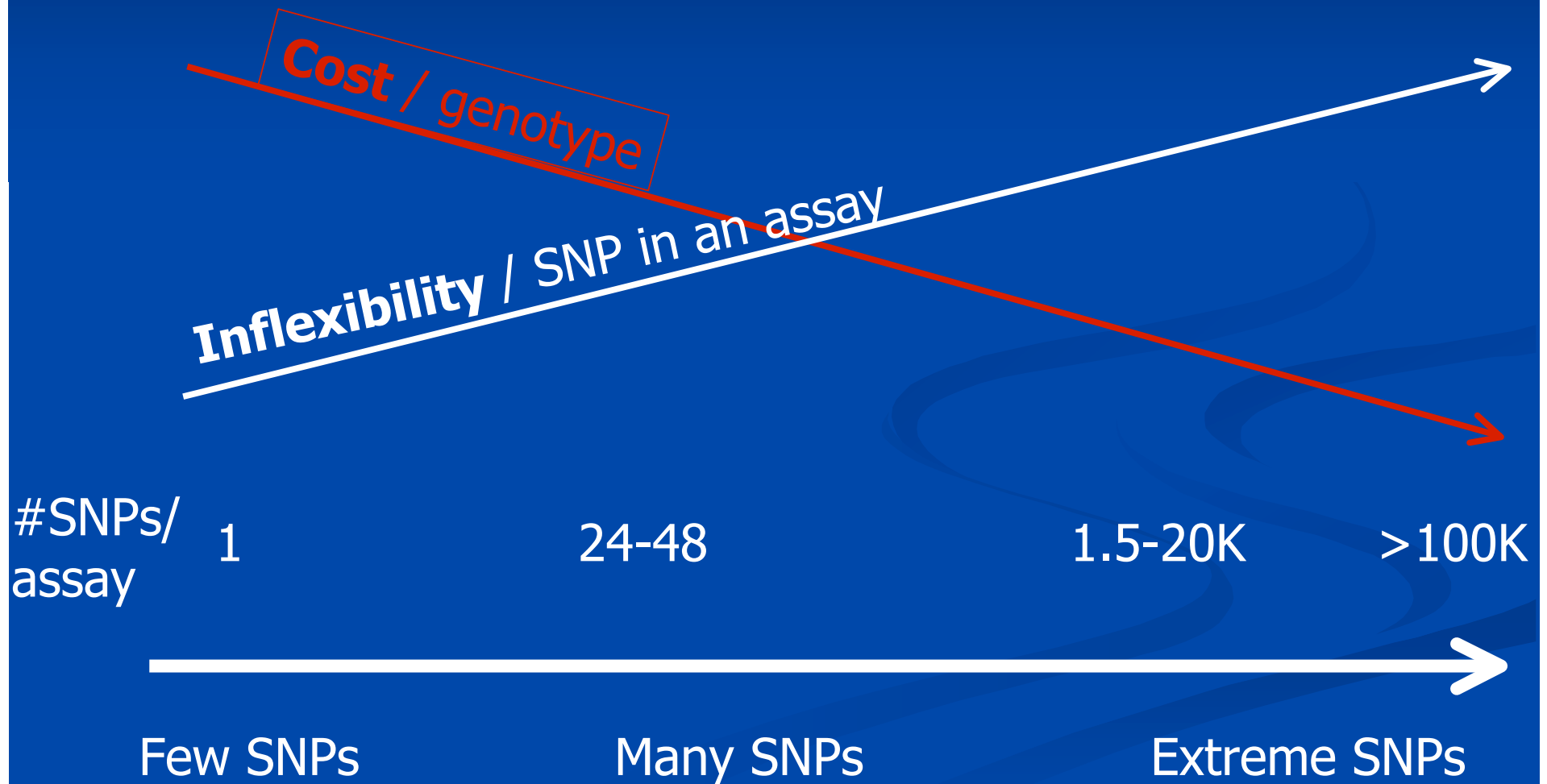
Illumina Infinium™ / Affymetrix Mapping Assays

PCR based cycle sequencing

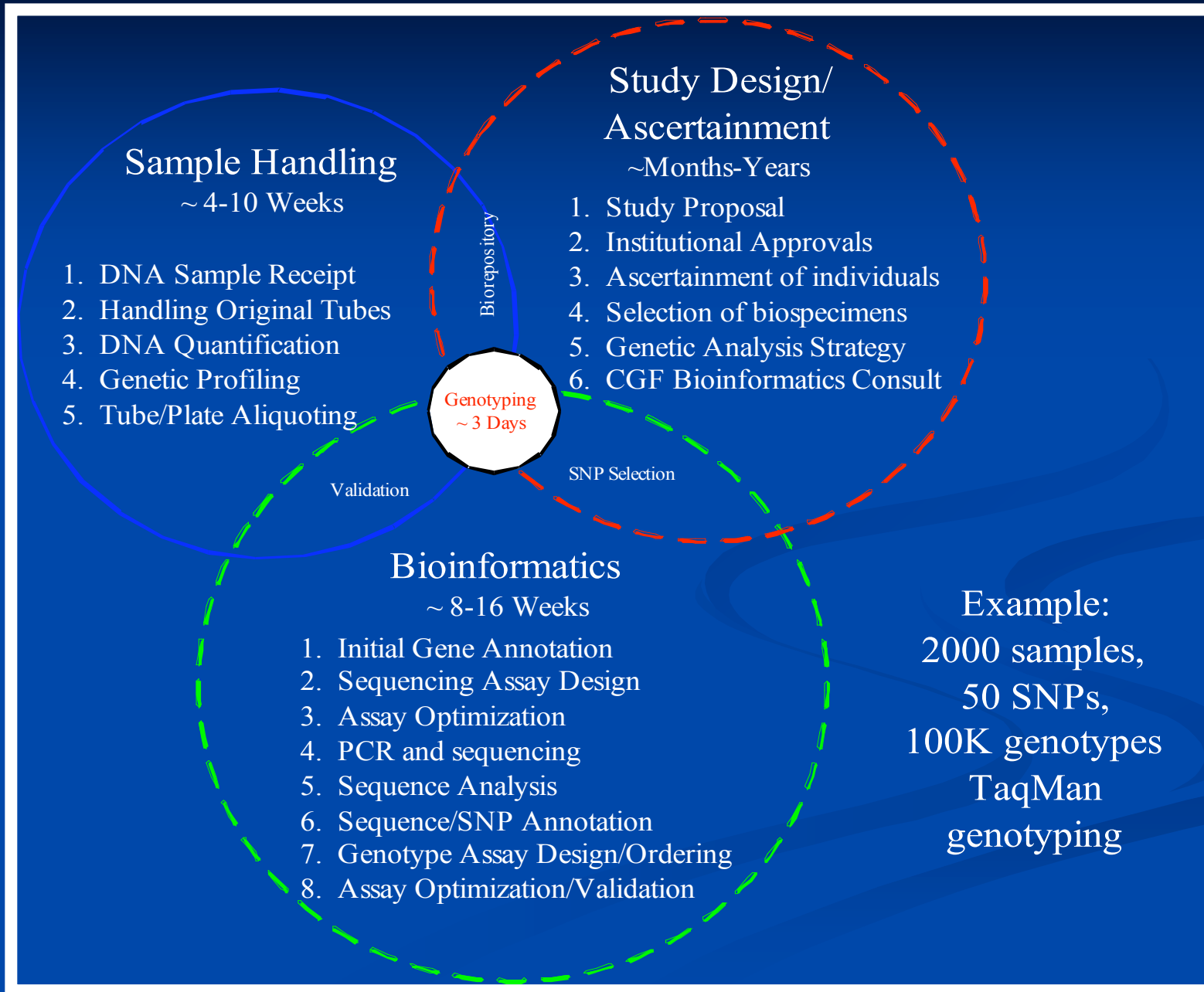
Genotype Platforms

| <i>Intent</i> | <i>Flexibility</i> | <i>Technology</i> | <i>Platforms</i> | <i>~genotypes per day (Samples)</i> |
|---|--------------------|-------------------|--|---|
| High genomic coverage (no customization) | None | Whole Genome | Illumina Infinium Affymetrix Mapping Assays | 12,000,000 (24-36) |
| High-throughput survey large candidate genes or linked region survey | Low | Large Multiplex | Illumina GoldenGate™ Affy/ParAllele Targeted Genotyping | 1,000,000 (50-600) |
| Moderate throughput, single candidate gene or candidate SNPs in pathway | Moderate | Medium Multiplex | Sequenom iPLEX™ SNPlex™ | 200,000 (3000-5000) |
| Low throughput, Candidate SNP or Confirmation | High | Single-plex | TaqMan™ MGB Eclipse™ Fragment Sizing (VNTR) | 100,000 (96 – 100K) |

Genotype Needs



Requirements for High Throughput Genotyping



Economic Considerations

- In general as you increase the number of SNPs per assay you decrease the cost per genotype in that assay, but increase the total cost of the assay per sample.
- For the custom higher plex assays there is a definite economy of scale when designing and typing custom vs fixed panel assays.

This cost is based on the oligo synthesis cost that is amortized over the N of samples genotyped per order of the custom assay.

Fixed panels will not have a such a cost as the N of samples is maximized to the lowest cost by the vendor.

“Single-Plex” Genotyping (TaqMan™ and MGB Eclipse™)

- Assay Design - majority of assays designed by vendors, high assay conversion rates 77% and 95% respectively
- 5 ng of DNA per assay – genotyping completion > 99% (> 10M genotypes)
- Sample quality leads to single assay failure – determined by completion concordance.
- Used to genotype few SNPs (candidate SNP / confirmation) or to fill in where higher multiplexes have failed.
- Easy to implement in most laboratories (design / wet-lab / analysis)
- Extremely flexible

Genotyping using the Sequenom iPLEX™ Assay

- Multiplexing at 20 to 28-plex level
- High multiplex design and conversion rates for individuals assay, with both small and large numbers of input SNPs
- ~ 10 ng of DNA per panel
- Multiplex PCR amplification / Single Base Extension with detection via mass spectrometry
- Utilize a single termination mix / universal protocol
- observe lower completion rates in general when compared to other platforms, due to complex chemistry (non-specific failure, in our data).
- More cost effective when typing larger N of SNPs on smaller N sample sets <800
- Used internally to generate genotypes for control populations used in tagSNP selection.

Genotyping using the SNPLex™ Assay

- Multiplexing at 24 to 48-plex level
- Multiplex design and conversion rates for individuals assays, is lower but assay that are designed and optimized perform well on wide range of sample qualities
- Design and conversion rates greatly increase with large numbers of input SNPs
- Requires ~ 60ng of DNA per attempt (DNA is fragmented at 1st step, WGA samples perform well)
- Oligo ligation / amplification / clean up & labeling AMP products / detection using fluorescent tags via Capillary Electrophoresis
- Uses universal reagents for all steps post ligation
- Automated data analysis / clustering for all samples within a project
- Most cost effective when typing fixed panel of SNPs on large N of sample sets >2500.

Genotyping using the Illumina GoldenGate™ Assay

- Multiplex panel assay -- 96 to 1536 loci per panel (panels must be created in multiples of 96)
- Requires ~ 500ng of high quality DNA per assay
- Oligo extension / ligation / fluorescent labeling / binding to proprietary beads, which are hybridized to matrix, decoded and imaged
- Highly automated data analysis / clustering -- train software where clusters should fall (based on actual data for all loci) and genotype calling is fully automated
- Assay design and conversion rates lower, but once working assays perform quite robustly (however bioinformatics input needed to generate OPA's)
- Most cost effective and efficient when used on large N of samples
- Extremely high completion and concordance rates
- Supports custom and preformatted (tested) panels

Genotyping using the Affymetrix (ParAllele) Targeted Genotyping™

- Multiplex panel assay targeting 1,000 to 25,000 SNPs at a time
- Requires ~ 2-4 ug of high quality DNA per assay (pre-amplification step is being tested)
- Molecular Inversion Probe (MIP) technology for allele specific physical ligation (4 different reactions 1 per allele) / digestion / inversion / universal amplification / hybridization
- Hybridization and detection using Affymetrix GeneChip® Universal Tag Arrays available in 3K, 5K, 10k or 25K configurations and contain novel tag sequences allows non competitive add on panel creation
- Automated data analysis using Affymetrix G-type software (same used for Mapping Assays)
- High reported assay design and conversion rate, with ability to build more flexible / stackable panels.
- Supports custom and preformatted (tested) panels

Whole Genome Genotyping using the Illumina Infinium™

- Two commercially available pre-designed panels
 - 109K Exon Centric content focus
 - 317K tagSNPs from Phase I HapMap Project
- Single tube whole genome amplification allows relatively low DNA consumption (~750ng per sample)
- Single chip / single sample processing
- Whole genome amplification / fragmentation / denature & hybridization to solid state / SNPs discrimination / amplify signal and detect
- Reported high completion and concordance rates >99% >99.96% respectively (38M genotypes compared N=120 samples, 317K SNPs)
- Highly automated data analysis / clustering and genotype calling is fully automated (virtually no user intervention)
- Supports only preformatted panels, custom panels to be released Q2-2006

Whole Genome Genotyping using Affymetrix Mapping Assays

- Three commercially available pre-designed panels
 - 10K, 100K, and 500K
 - 500K SNPs designed as linkage mapping panel (spaced markers across genome = significant redundancy for association based studies)
 - 500K is comprised of two arrays (dual chip processing), which are defined by restriction sites that reduce genome size (~262,000 SNPs for Nsp based arrays and ~238,000 SNPs for Sty based arrays).
- Restriction enzyme digestion / adapter ligation / universal amplification and hybridization to predefined oligonucleotide arrays
- Completion rates ~ 94 % for 500K chips (~470K genotypes per sample)
- Highly automated data analysis all loci per sample (virtually no end user intervention)
- Supports custom panels and preformatted arrays, although custom arrays not cost effective unless very large N is typed.

Points to remember

- When choosing technology, focus on the goals of the study(s)
 - How many samples will you be genotyping (range)
 - What will the sample quality be
 - How many SNPs will be genotyped
 - How important is it that every SNP is genotyped
 - Timeframe to complete (design to genotype reporting)
 - Cost
 - What are the capital costs / can they be amortized
 - What are the M&S costs / are you efficiently planning on the use of a technology
 - How will informatics infrastructure impact cost
 - Labor and QC costs are often hard to pre-plan