

SNP Application Studies Using OpenArray Genotyping Platform

Kelly Li, Ph.D.

Sr. Staff Scientist

Genomics Assay R&D

Nov. 26, 2009

Shanghai, China





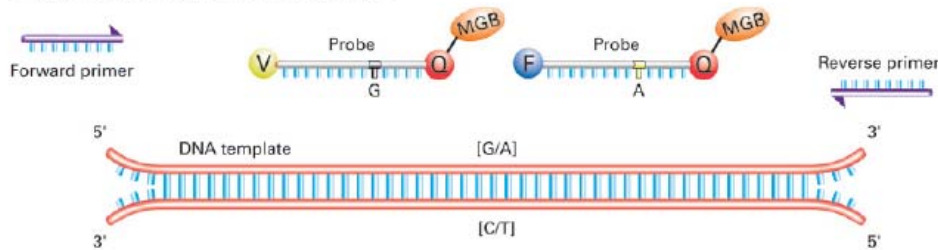
Agenda

- Background
 - TaqMan genotyping assays
 - What is OpenArray?
 - Major features
- Application studies
 - SNP association with heart disease
 - SNP Analysis as a Tool for Microbial Subtyping
 - Forensic application using SNP genotyping

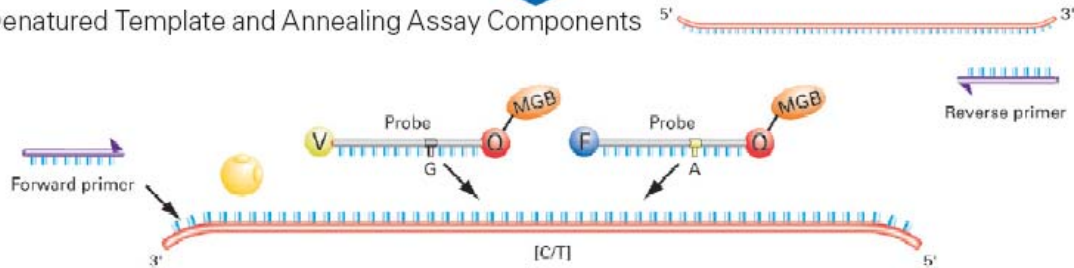
TaqMan[®] SNP Genotyping Assays

assay mechanism for allelic discrimination

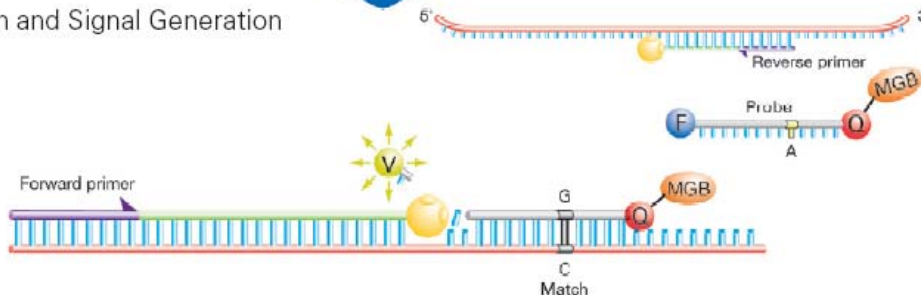
1. Assay Components and DNA Template



2. Denatured Template and Annealing Assay Components



3. Polymerization and Signal Generation



Legend

- V VIC[®] dye
- F FAM[™] dye
- Q Quencher
- MGB Minor Groove Binder
- AmpliAq Gold[®] DNA Polymerase
- Probe
- Primer
- Template
- Extended Primer

Genotyping Throughput Landscape:

AB Has a Broad Range of Options

Taqman® SNP
and CNV
Genotyping
Assays

Sample-to-SNP™

Taqman®
OpenArray®
Genotyping
System

SNiPlex™ System

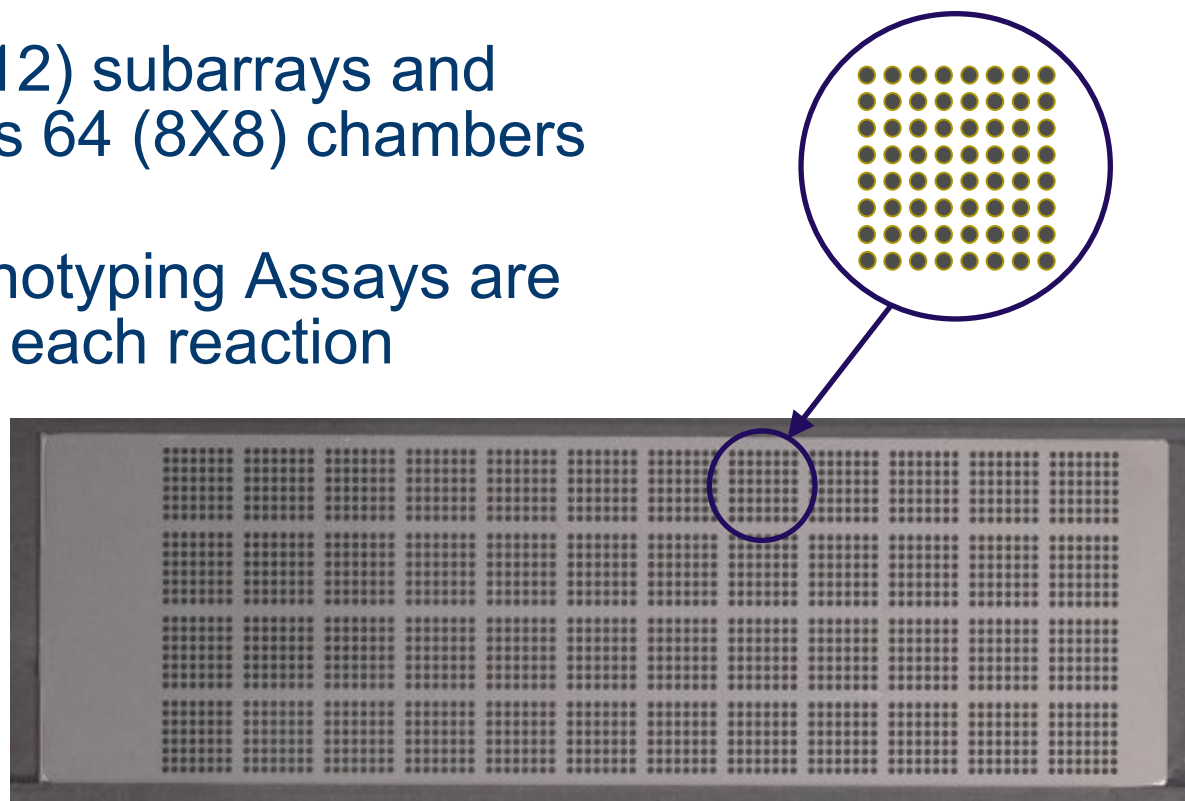
SOLiD™ System



Increasing Throughput
Data points

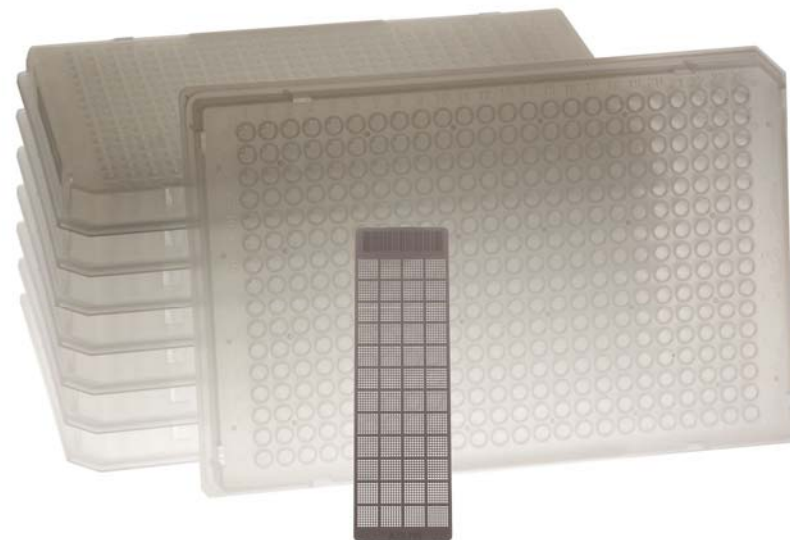
What is OpenArray Platform?

- A new genotyping platform developed by BioTrove
- There are 48 (4X12) subarrays and each subarray has 64 (8X8) chambers
- TaqMan SNP Genotyping Assays are pre-aliquoted into each reaction chamber



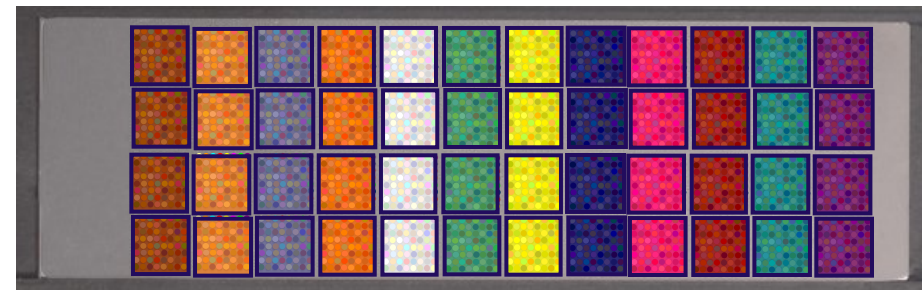
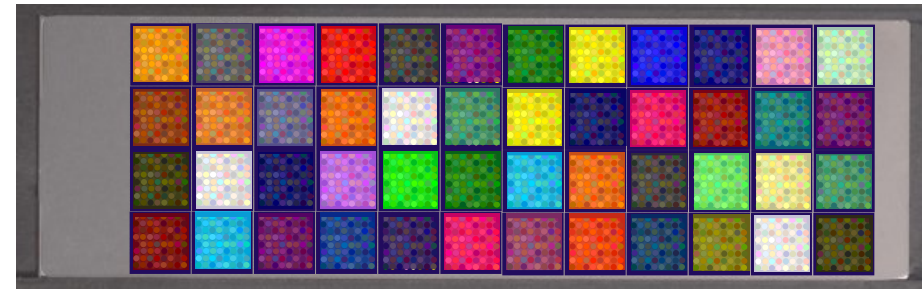
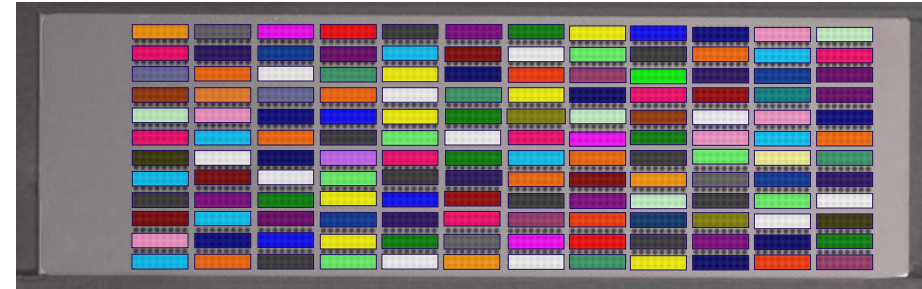
Throughput of OpenArray Platform

- 3072 low-volume reaction chambers (33nl) with preAliquoted TaqMan SNP Assays
- Each OpenArray plate is equivalent to eight 384-well plates
- A wide variety of genomics applications



Flexibility of OpenArray Platform

- Multiple different Assay x Sample Number Combinations
- 16 Assays x 144 Samples
- 32 Assays x 96 Samples
- 64 Assays x 48 Samples
- 128 Assays x 24 Samples
- 256 Assays x 12 Samples



Multiple Independent Genetic Factors at *NOS1AP* Modulate the QT Interval in a Multi-Ethnic Population

Dan E. Arking^{1*}, Amit Khera^{2,3}, Chao Xing⁴, W. H. Linda Kao⁵, Wendy Post⁶, Eric Boerwinkle⁷, Aravinda Chakravarti¹

1 McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **2** Donald W. Reynolds Cardiovascular Clinical Research Center, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **3** Division of Cardiology, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **4** McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America, **5** Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, **6** Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, **7** Human Genetics Center and Division of Epidemiology, The University of Texas, Houston, Texas, United States of America

Abstract

Extremes of electrocardiographic QT interval are associated with increased risk for sudden cardiac death (SCD); thus, identification and characterization of genetic variants that modulate QT interval may elucidate the underlying etiology of SCD. Previous studies have revealed an association between a common genetic variant in *NOS1AP* and QT interval in populations of European ancestry, but this finding has not been extended to other ethnic populations. We sought to characterize the effects of *NOS1AP* genetic variants on QT interval in the multi-ethnic population-based Dallas Heart Study (DHS, $n = 3,072$). The SNP most strongly associated with QT interval in previous samples of European ancestry, rs16847548, was the most strongly associated in White ($P = 0.005$) and Black ($P = 3.6 \times 10^{-5}$) participants, with the same direction of effect in Hispanics ($P = 0.17$), and further showed a significant SNP \times sex-interaction ($P = 0.03$). A second SNP, rs16856785, uncorrelated with rs16847548, was also associated with QT interval in Blacks ($P = 0.01$), with qualitatively similar results in Whites and Hispanics. In a previously genotyped cohort of 14,107 White individuals drawn from the combined Atherosclerotic Risk in Communities (ARIC) and Cardiovascular Health Study (CHS) cohorts, we validated both the second locus at rs16856785 ($P = 7.63 \times 10^{-8}$), as well as the sex-interaction with rs16847548 ($P = 8.68 \times 10^{-6}$). These data extend the association of genetic variants in *NOS1AP* with QT interval to a Black population, with similar trends, though not statistically significant at $P < 0.05$, in Hispanics. In addition, we identify a strong sex-interaction and the presence of a second independent site within *NOS1AP* associated with the QT interval. These results highlight the consistent and complex role of *NOS1AP* genetic variants in modulating QT interval.



- **Institutes**

- **Aravinda Chakravarti, Johns Hopkins University Medical School**
- **Dan Arking & Amit Khera, University of Texas Southwestern Medical Center**

- **Previously known**

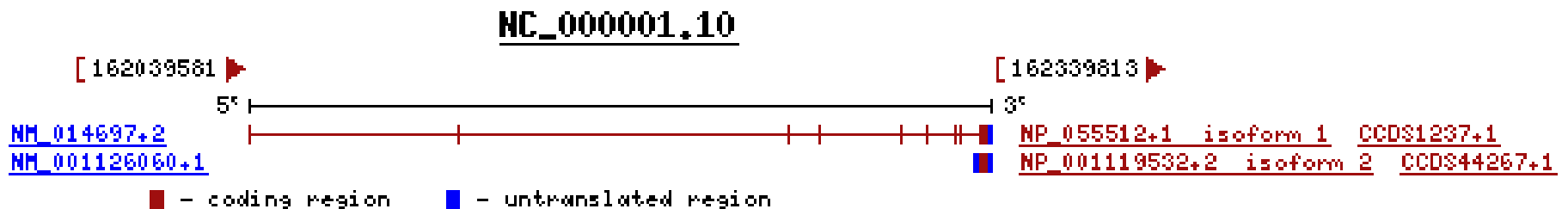
- **QT interval is a measure of cardiac repolarization**
- **QT interval is a biomarker for risk of sudden cardiac death**
- **SNP rs16847548 in the *NOS1AP* locus associate with altered QT interval in European populations.**

- **Questions**

- **Do this SNP (rs16847548) associate with QT interval in other ethnic populations?**
- **Do other SNPs in *NOX1AP* associate with QT interval?**
- **Do other SNPs associate with QT interval in different ethnic populations?**

NOS1AP

- NOS1AP : nitric oxide synthase 1 (neuronal) adaptor protein
- This gene encodes a cytosolic protein that binds to the signaling molecule, neuronal nitric oxide synthase (nNOS).
- As an adapter protein, this protein mediates interactions with nNOS and specific targets including the small monomeric G protein (Dexas1).



Populations and genotyping technology

- **Dallas Heart Study Population**
 - 2,949 Individuals
 - White, Hispanic, Black
 - TaqMan SNP Genotyping Assays
 - 384-well plates
- **Atherosclerotic Risk in Communities Population & Cardiovascular Health Study Population**
 - 14,107 individuals
 - White
 - TaqMan SNP Genotyping Assays
 - OpenArray plates
- **Accuracy of OpenArray genotyping:** determined by comparison to concordance calls generated for 58 samples, genotyped multiple times (350 comparison per SNP)
 - **rs16847548 = 99.7%;**
 - **rs16856785 = 99.4%.**

Association of SNPs with QT interval

SNP	Position	A1	A2	Non-Hispanic Blacks (n = 1,497)				Hispanics (n = 499)				Non-Hispanic Whites (n = 940)			
				A1 Freq	β	SE	P	A1 Freq	β	SE	P	A1 Freq	β	SE	P
rs7539281	158739692	A	G	0.60	0.18	0.63	0.77	0.29	0.66	0.90	0.4639	0.26	1.81	0.66	0.006
rs4657139	158761565	A	T	0.88	1.59	0.90	0.08	0.42	0.98	0.85	0.2476	0.34	1.21	0.60	0.043
rs16847548	158766932	C	T	0.18	3.22	0.78	3.58E-05	0.20	1.47	1.07	0.1706	0.20	2.57	0.72	0.0004
rs12567209	158768137	A	G	0.08	0.08	1.13	0.95	0.12	-1.00	1.23	0.4175	0.08	-0.45	1.09	0.68
rs12576211	158768181	T	G	0.51	1.83	0.61	0.003	0.35	0.01	0.87	0.9919	0.29	1.81	0.63	0.005
rs1415262	158777793	C	G	0.81	1.13	0.76	0.14	0.43	1.11	0.85	0.1931	0.35	0.81	0.60	0.18
rs10494366	158817343	G	T	0.62	1.22	0.63	0.05	0.41	1.52	0.84	0.07228	0.36	0.72	0.60	0.23
rs16856785	158831945	C	G	0.61	1.60	0.62	0.01	0.10	0.58	1.30	0.6573	0.10	0.95	0.97	0.33

Findings:

- SNP **rs16847548** has largest association with QT interval in all populations,
- Allele frequencies for SNP **rs16856785** in Hispanic & white populations are fairly similar to each other, frequencies in black populations are divergent.
- SNP **rs16856785** has a large association in black populations, but in either Hispanic or white the association is not significant ($P > 0.3$), which have lower A1 allele frequencies and smaller samples sizes.

Larger Study & Independent effect of SNPs

Beta = size of effect under additive genetic model

Independent effect assessed by forward stepwise regressions

384-well plate

OpenArray

Model	SNP	A1	A2	Non-Hispanic Blacks (n = 1,497)				Non-Hispanic Whites (n = 940)				ARIC/CHS Whites (n = 14,107)			
				A1 Freq	β	SE	P	A1 Freq	β	SE	P	A1 Freq	β	SE	P
Single SNP	rs16847548	C	T	0.18	3.22	0.78	3.58E-05	0.20	2.57	0.72	0.0004	0.22	2.42	0.22	<2.00E-16
	rs16856785	C	G	0.61	1.60	0.62	0.01	0.10	0.95	0.97	0.33	0.09	2.11	0.32	3.94E-11
Multi-SNP	rs16847548	C	T	0.18	3.22	0.78	3.66E-05	0.20	2.62	0.73	0.0004	0.22	2.22	0.23	<2.00E-16
	rs16856785	C	G	0.61	1.60	0.62	0.01	0.10	0.24	0.99	0.81	0.09	1.74	0.32	7.63E-08

Findings:

- Initial study with 384-well plates showed association with both SNPs for black population (1497 individuals), but could only show a significant association for rs16847548 in white population.
- The two SNPs show independent association in the black population
- Follow up study using OpenArray and a vastly increased sample size of white individuals shows that rs16847548 is also has an association in that population, and that it acts independently in that population as well.



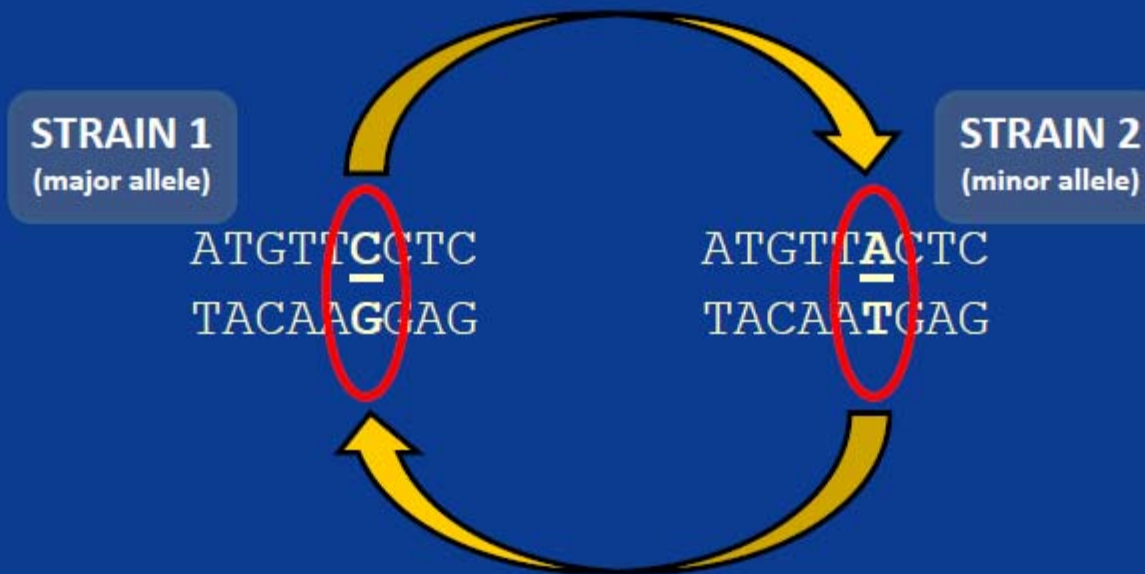
SNP Analysis as a Tool for Microbial Subtyping

Duncan MacCannell Ph.D.

CDC/CCID/NCZVED/DFBMD/EDLB, Atlanta GA

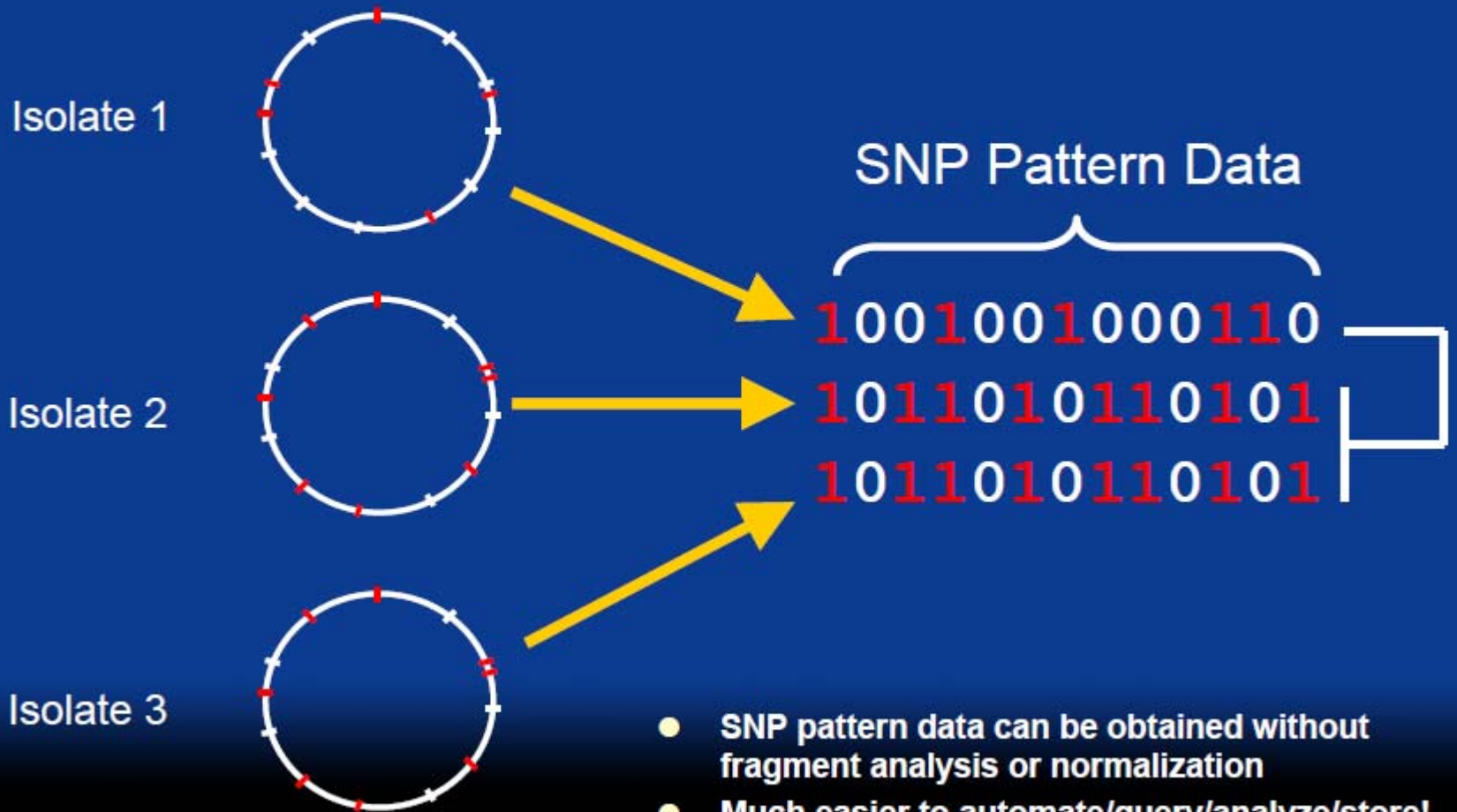
SAFER • HEALTHIER • PEOPLE™

Single Nucleotide Polymorphisms (SNPs)



- SNPs are generally selected so that there are only two possible variants or “alleles” at a given locus (eg: C/A).
- Occur in both coding and non-coding DNA sequences.
- May represent synonymous or non-synonymous differences.

From SNPs to "fingerprints"



- SNP pattern data can be obtained without fragment analysis or normalization
- Much easier to automate/query/analyze/store!

Important considerations for SNP selection

- ***Allelic stability & assay reproducibility.***
 - ◆ Is each SNP stable enough for routine use?
- ***Population-level distribution of SNPs.***
 - ◆ How diverse are the alleles in the wild?
- ***Genomic distribution and biological significance.***
 - ◆ Is the genomic coverage of the panel sufficient?
 - ◆ Are there useful markers that provide secondary information about the organism?
- ***Linkage disequilibrium.***
 - ◆ Many SNPs are coinherited along clonal lines.
 - ◆ Introduces lots of redundancy → tag/htSNPs.

Identifying SNPs in the O157:H7 Genome

- Screened 11 historical, outbreak-associated O157:H7 strains from the PulseNet archives using comparative genomic microarrays (NimbleGen).
- **Genomic coverage (x 11) included:**
 - ◆ 1,199 (~22.1%) ORFs
 - ◆ 1,167,948 bp (21.1%) of the genome.
 - ◆ Complete coverage of pO157.
- **Coding sequences were emphasized:**
 - ◆ 906 SNPs identified from 523 ORFs.
 - ◆ 128 candidate SNPs were selected for 1st-round analysis.

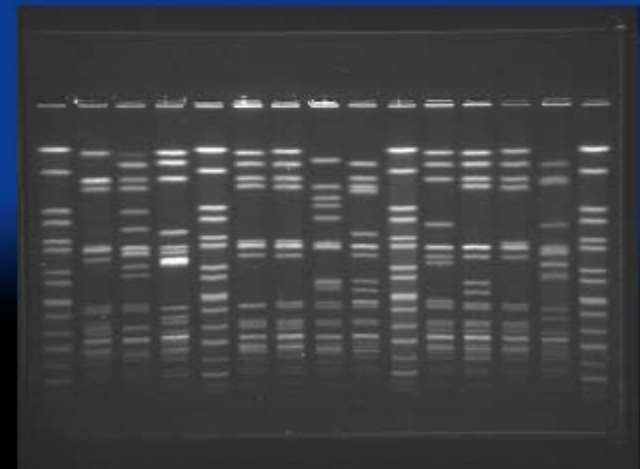
Screening Candidate SNPs

- **288** representative O157:H7 isolates from the CDC collection, including:
 - ◆ 146 isolates from 1982-2006
 - Diverse collection of PFGE types.
 - Historical isolates of note from both sporadic and outbreak cases.
 - For analysis, samples were blind coded, and included 17 identical repeats.
 - ◆ 142 isolates from the 2006 Spinach OB
 - All EXHX01.0124/EXHA26.0015
- **128 SNP x 288 ISOLATES = 36,864 reactions.**
- Needed a platform capable of high-throughput screening of candidate SNPs against a library of reference strains.

Core techniques: PFGE

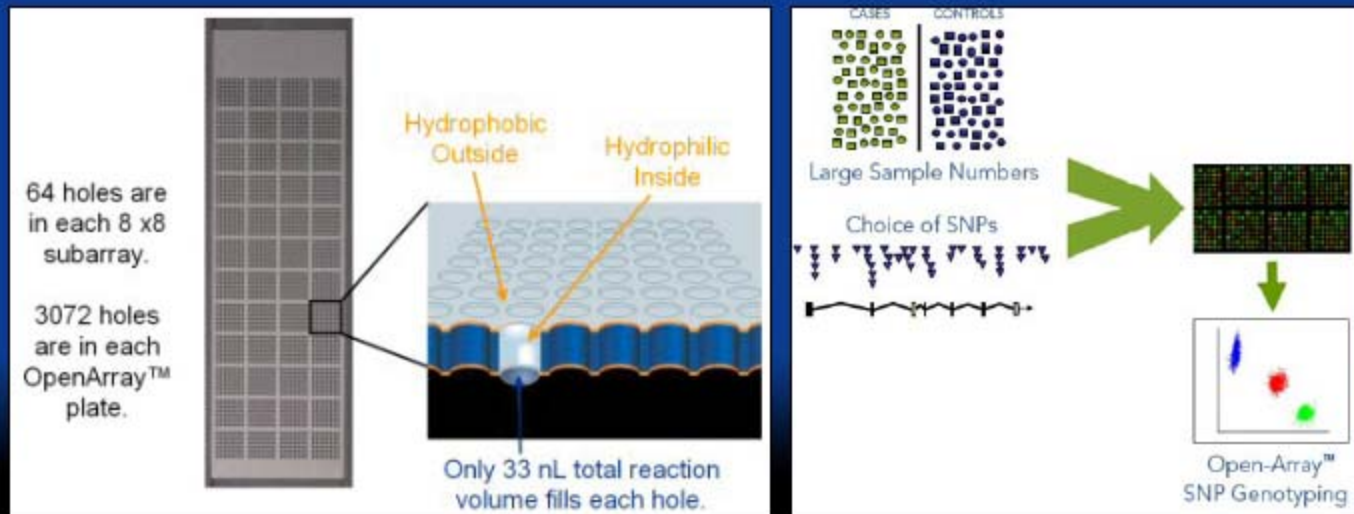
Culture and isolation:	24-48 hours
Preparation of PFGE plugs:	>4 hours
Restriction and gel preparation:	>3 hours
PFGE run time:	17-20 hours
Image acquisition/analysis/upload:	+ <u>0.5 -1 hours</u>
	49-76 hours

Minimum post-culture workup: ~25 h
Total hands on time: ~6 h
Max isolates / gel / enzyme: ~11



High Throughput Screening

- BioTrove OpenArray™ chosen for development.
- HT platform for microfluidic endpoint/realtime PCR.
- 33nL/well with 3,072 (64 x 48) assays per chip.
- Standard TaqMan™ assays with FAM/HEX probes.
- In the present configuration, each chip can screen 128 SNP targets against 24 isolates. 4-6 hours post culture workflow



Assay Reproducibility

- Review of 114 isolates from the August-September 2006 Spinach outbreak.
 - ◆ Average SNP pattern similarity: **99.95%**
 - ◆ All but three isolates were 100% identical across 128 loci.

- 17 Blind coded duplicate isolates in the initial isolate set also demonstrated consistent reproducibility (rng: 98-100%).
 - ◆ Variation was generally limited to loci with high rates of failure.

Optimizing the Panel

- Of 128 SNPs that were assessed:
 - ◆ 66 were of poor quality, were monomorphic or had minor allelic frequencies of less than 20%.
 - ◆ If these are removed, the discriminatory power is largely unaffected.
 - ◆ However, discriminatory power is dependant upon the variability of the loci in the analysis set.
- The panel may be further optimized by careful review of covariant and redundant loci.
 - ◆ Ideal primary typing panel: <50 loci.

The path forward: More SNPs

- To improve discriminatory power, we need to identify as many useful tag-SNPs as possible.
- Assays for 269 additional loci have been designed and synthesized.
- Screening in groups of 64.
- Corroboration with virulence, resistance, serotype markers.
- Ongoing evaluation of the usefulness of each assay set in non-O157 STEC.

Summary Remarks

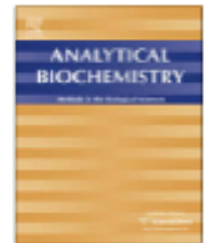
- SNP-based genotyping promises:
 - ◆ rapid, high-throughput subtyping
 - ◆ improved speed and accuracy of outbreak clusters over traditional PFGE.
 - ◆ lab and data analysis components that are both highly amenable to automation.
- Locus selection is critical to resolution.
- Current “snapshot” assay appears to rival the discriminatory power of two-enzyme PFGE.
- While both PFGE and SNP (currently) lack the fine resolution of MLVA, both may be applicable to a wider range of *E. coli* serovars in the future.
- Complimentary subtyping technologies.



Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio



A low-cost, high-throughput, automated single nucleotide polymorphism assay for forensic human DNA applications

Robert Pomeroy^{a,*}, George Duncan^b, Bulbin Sunar-Reeder^c, Elen Ortenberg^d, Melba Ketchum^e, Hannah Wasiluk^e, Dennis Reeder^c

^aDepartment of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093, USA

^bBroward County Sheriff's Office, Fort Lauderdale, FL 33312, USA

^cReeder Analytical Consulting, Frisco, TX 75034, USA

^dBioTrove, Woburn, MA 01801, USA

^eDNA Diagnostics, Timpson, TX 75975, USA

ARTICLE INFO

Article history:

Received 17 June 2009

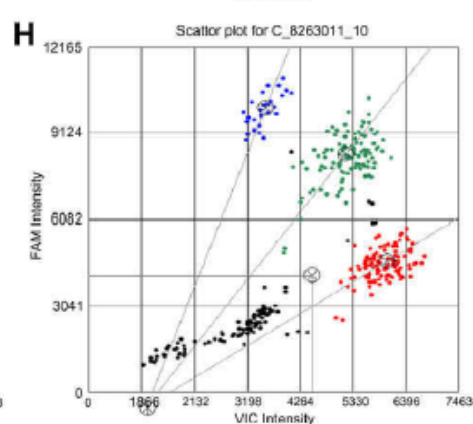
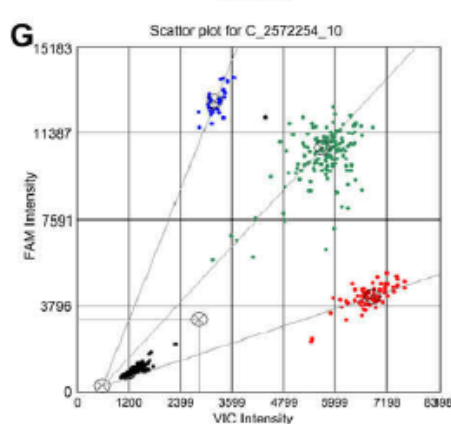
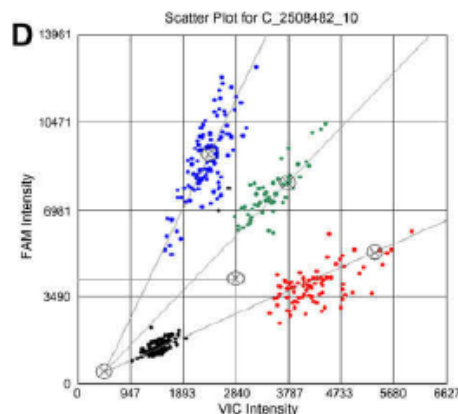
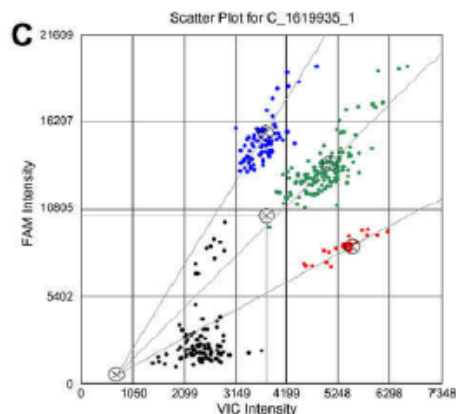
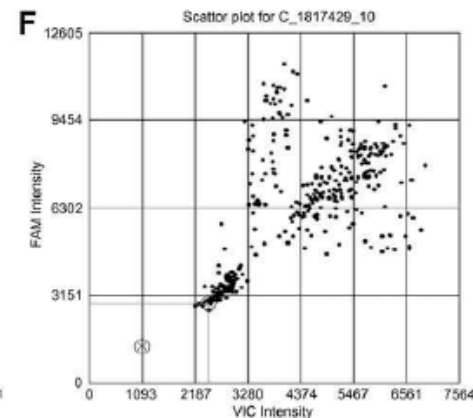
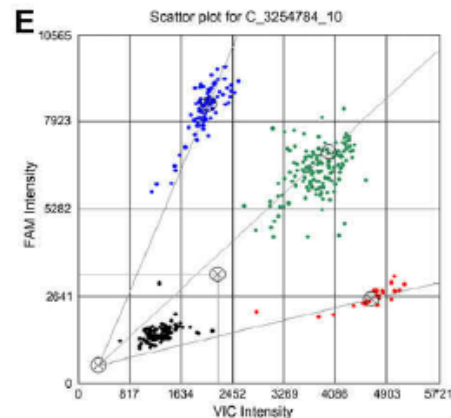
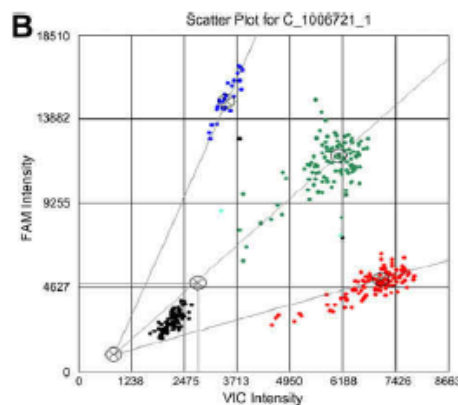
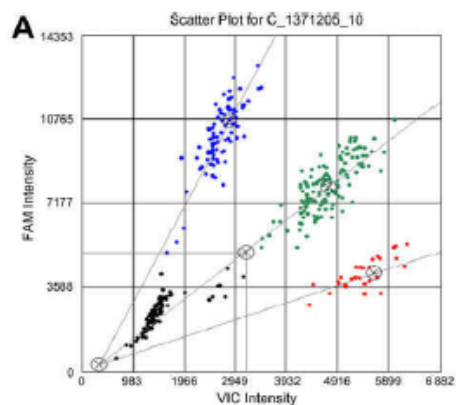
Available online 30 July 2009

A Panel of 8 SNPs for Testing

SNP information from Applied Biosystems' SNPbrowser software.

Celera ID	dBSNP ID	Type	Chromosome	Position
hCV1006721	rs560681	Validated	1	159,053,294
hCV8263011	rs279844	Validated	4	46,024,412
hCV1817429	rs1336071	Validated	6	94,593,976
hCV2572254	rs1019029	Validated	7	13,860,801
hCV3254784	rs740598	Validated	10	118,496,889
hCV1619935	rs1058083	Validated	13	98,836,234
hCV1371205	rs9951171	Validated	18	9,739,879
hCV2508482	rs1523537	Validated	20	50,729,569

Scatter Plots of all DNA Concentrations with Allelic Determination



Population Study with 39 Tamil Samples

- Determination is possible with just 7 SNPs
- **FAM homozygous**
- **VIC homozygous**
- **Heterozygous**
- WGA samples work

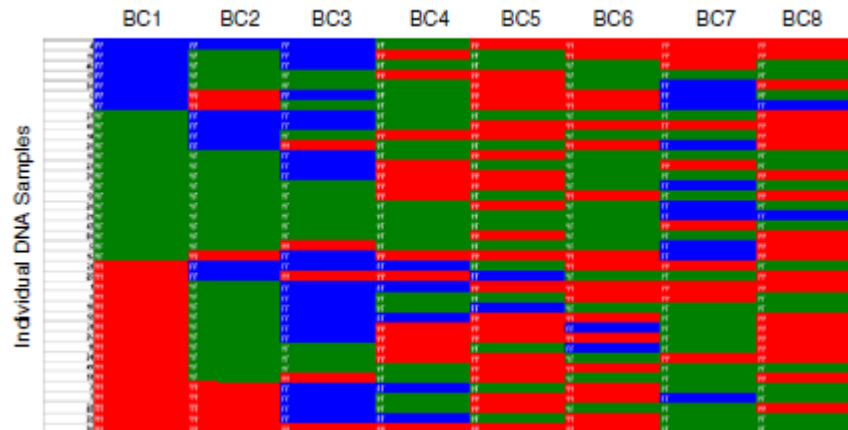


Table 2

WGA amplification with seven SNPs.

Sample description	DNA concentration	SNP assay						
		A	B	C	D	E	G	H
MC-24 original	Too low	VF	NC	NC	NC	NC	NC	NC
MC-24 WGA	7 ng/μl	VF	FF	FF	VF	VV	VF	VF
MC-24 WGA purified	1 ng/μl	VF	FF	FF	VF	VV	VF	VF

Note. SNP assay identification: A = C_1371205; B = C_1006721; C = C_1619935; D = C_2508482; E = C_3254784; F (not shown) = C_1817429; G = C_2572254; H = C_8263011. SNP C_181749 is the problematic assay. Assay F was dropped. Sample MC-24 was initially too low to perform adequately. WGA was performed with and without follow-up purification. Samples are now corrected typed. NC, no call; VV, homozygous; FF, homozygous; VF, heterozygous.

Determine Paternity Samples with 7 SNPs

	C_1371205	C_1006721	C_1619935	C_2508482	C_3254784	C_2572254	C_8263011
F1	VF	VF	VF	VF	VF	FF	VF
M1	FF	VF	VV	VV	VV	VV	VF
C1	VF	VF	VV	VV	VV	VF	FF
F2	FF	VF	FF	VF	VF	VV	VV
M2	VF	VV	FF	VV	VV	VV	VF
C2A	FF	VV	FF	VF	VF	VV	VV
C2B	VF	VV	FF	VF	VF	VV	VV
C2C	VF	VF	FF	VF	VF	VV	VV

