

## Genetic Analysis

# AUTOMATED, SCALABLE SNP ANALYSIS: THE GENOMELAB SNPSTREAM® GENOTYPING SYSTEM

Zhiming Jiang, Keith W. Roby, Dana P. Campbell, Rene M. Oda, Graham J. Threadgill, and Scott K. Boyer  
Beckman Coulter, Inc.

## Introduction

### SNP Analysis

Single nucleotide polymorphism (SNP) is the most common type of genetic variation, occurring every 300 to 1000 bases in the human genome and many other species. Large collections of confirmed and annotated SNPs are available in public and private databases. SNP genotyping is playing an increasingly important role in genetic mapping, disease studies, pharmacogenomics, and drug discovery. To date, genome-wide scans and studies involving thousands of SNPs and samples are hampered by the lack of genotyping systems that provide cost-effective throughput, concordance, and reliability.

### GenomeLab SNPstream Genotyping System

To address this need, an automated, ultra-high-throughput system, the GenomeLab SNPstream® Genotyping System from Beckman Coulter, was developed for the detection of SNPs, utilizing multiplexed PCR\* in conjunction with tagged-array, multiplexed, single-base primer extension technology. The system provides all the necessary components, including SNPware® reagents, software, and instrumentation, for medium- to ultra-high-throughput SNP analysis. At up to 12-plex levels, multiplexed PCR and primer extension are performed in homogeneous reactions. The SNPstream Genotyping System employs an oligonucleotide microarray manufactured in a glass-bottom, 384-well microplate. The system's flexibility enables automated SNP genotyping of 4,600 to 890,000 genotypes in a 24-hour period. Assay results are read by direct two-color fluorescence with the SNPstream Imager.

### CEPH Family Study

In this study, we evaluated the performance of the GenomeLab SNPstream Genotyping System on 180 SNPs at 12-plex levels using 96 DNA samples from 8 members of 12 CEPH/Utah families. Our results demonstrate 87.78% assay conversion at 12-plex, 98.18% genotyping success rate, and data concordance of 99.88%. The SNPstream Genotyping System provides a cost-effective way of performing automated SNP genotyping at medium- to ultra-high-throughput capacity and high data concordance.

## Methods and Results

### Assay Design

*Autoprimer.com* (<http://www.autoprimer.com>) is a web-based primer design tool (Figure 1). The website utilizes tab-delimited text file input for up to 200 sequences. The SNPs are marked in brackets and the sequences are in standard IUPAC format. Each SNP sequence is suggested to be shorter than 2000 bases with sequences of at least 150 bases upstream or downstream of the SNP locus. The design engine analyzes the sequences and designs three primers for each SNP: forward and reverse PCR primers and one tagged SNP primer for primer extension. The primers are designed based on the DNA secondary structure, the melting temperature, and to minimize



Welcome to Autoprimer.com sponsored by Beckman Coulter.

This site is dedicated to providing our customers with the most state-of-the-art comprehensive primer design tools available for optimal PCR and extension primer design using Beckman Coulter's proprietary SNP Identification Technology™ for SNP genotyping.

Registered users click [here](#).

New users click [here](#) to register.

For general questions or additional information regarding Autoprimer.com, please email our [support team](#).

For more information on Beckman Coulter and SNP Identification Technology powered products, visit us at [www.beckmancoulter.com](http://www.beckmancoulter.com)

Note: This application works best with Internet Explorer 4.0 or higher or Netscape 4.7.  
There is a known "bug" within Netscape 6.1 and this application will not function properly using that version.



The highest level of 128-bit digital security encryption secures Autoprimer.com. All information transferred using Autoprimer.com is secure. VeriSign 128-bit SSL Encryption Server ID enables our clients to verify our site's authenticity and to communicate with it securely via state-of-the-art SSL encryption, which protects confidential information from interception and hacking.

Verify our digital certificate by clicking the Verisign logo

Figure 1. Autoprimer.com is a web-based primer design tool.

Table 1. PCR Components and Reaction Setup

Component	Final Concentration	Manual	Beckman Coulter Automation	Other Automation
Genomic DNA (1.00 ng/μL)		2.00 μL		
Volume of PCR mix per well		3.00 μL		
Number of Plates		1	1	1
10X PCR Buffer II	1X	225.00 μL	233.33 μL	608.66 μL
MgCl <sub>2</sub> (25.00 mM)	5.00 mM	450.00 μL	466.67 μL	1217.34 μL
dNTP mix (2.50 mM each)	75.00 μL each	67.50 μL	70.00 μL	182.60 μL
Primer pool (10.00 μM each)	50.00 nM each	11.25 μL	11.67 μL	30.44 μL
Taq Gold	0.10 U/ μL	45.00 μL	46.67 μL	121.74 μL
H <sub>2</sub> O	Add to final volume of 5.00 μL	551.25 μL	571.66 μL	1491.22 μL
Total Volume of PCR mix		1350.00 μL	1400.00 μL	3652.00 μL

**Table 2. The Recommended Thermal Cycling Programs for 12-plex PCR**

Step	Temp	Time
1	94°C	1 min
2	94°C	30 sec
3	55°C	30 sec
4	72°C	1 min
5	Go to step 2 for 34 times	
6	4°C forever	

mispriming. Resulting PCR products typically fall in the range of 90~180 bases in length. The SNP primer may be selected from either DNA strand. To achieve the best results, quality sequences—defined as those that have been BLASTed and repeat masked, contain a GC content between 40% and 65%, and an average PHRED score of 20 or higher (or comparable quality score)—are recommended for assay design. Once primers are designed, they are grouped into panels of 12 by SNP extension type and the melting temperature of each primer. Each panel of 12 SNPs is of the same extension type for processing on SNPstream® since each extension mix contains two labeled terminators (Bodipy-Fluorescein and TAMRA). There are a total of six extension types including G/A, T/C, C/A, T/A, T/G, and C/G. Autoprimer.com optimizes the grouping of SNPs by extension type and assigns 5' tag sequences, complementary to the tags immobilized on the SNPware® array plate, to each SNP primer. An output file is generated to record the SNP panels and the primer sequences that facilitate oligonucleotide ordering. The second part of this output file contains information on these newly designed marker panels that can be uploaded directly into the SNPstream database.

### **Multiplexed PCR and Cleanup**

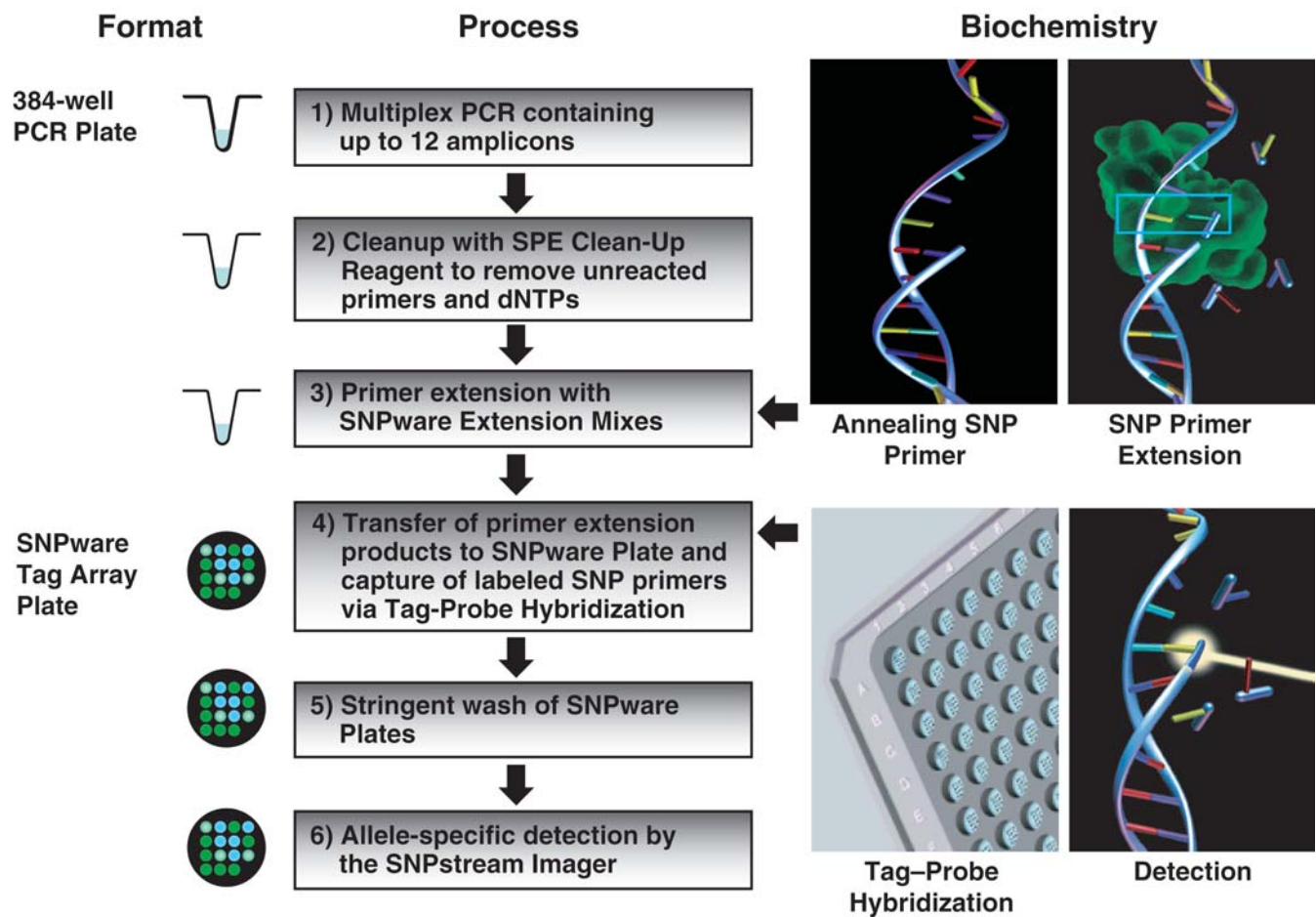
The 12-plex PCR is performed in 384-well plates (MJ Research, part number HSP-3801) at a volume of 5 µL/reaction. As little as 2 ng of genomic DNA are recommended for each reaction. This represents < 0.17 ng/genotype. Table 1 presents the PCR components and the final concentrations. The panel of 24 PCR primers designed by Autoprimer.com are pooled and added at a final concentration of 50 nM. Aliquots of 2 µL of genomic DNA (1 ng/µL) and 3 µL of the PCR mix (Table 1) are added to the corresponding wells. The PCR reaction setup requires the following products: AmpliTaq Gold® supplied with 10X PCR Buffer II and 25 mM MgCl<sub>2</sub> (Applied Biosystems, part number N808-0241) and 25 mM PCR Nucleotide Mix (USB, part number 77119). The nucleotide mix must be diluted to 2.5 mM prior to use. The recommended thermal cycling program is shown in Table 2. After PCR, the following steps are performed: a 1:25 dilution of the SBE Clean-up (USB, part number 78260) using SBE Clean-up Diluent (USB, part number 78261) is made, and 3 µL of the cocktail is added to each well (Table 3). Sealed plates are then incubated at 37°C for 30 minutes in order to degrade the residual PCR primers and dNTP, followed by 10 minutes at 100°C to deactivate the enzymes.

### **SNP Identification Technology**

SNP Identification Technology provides high levels of reproducibility, concordance, simplicity, and versatility in the SNP genotyping process. The technology described here involves a set of biochemical reactions that isolate the precise location of a SNP locus and determine the SNP genotype directly, utilizing the high-fidelity activity of DNA polymerase and terminating nucleotides. As shown in Figure 2, there are three main steps in this biochemical process: single base primer extension, hybridization, and detection.

**Table 3. Calculation for the SBE Clean-Up Reactions**

	Manual		Beckman Coulter Automation		Other Automation	
	2	4	2	4	2	4
Number of Plates	2	4	2	4	2	4
SBE Clean-Up	125.00 µL	250.00 µL	140.00 µL	248.00 µL	200.00 µL	320.00 µL
SBE Clean-Up Diluent	3.00 mL	6.00 mL	3.36 mL	5.95 mL	4.80 mL	7.68 mL
Total Volume	3.13 mL	6.25 mL	3.50 mL	6.20 mL	6.00 mL	8.00 mL



**Figure 2.** SNP Identification Technology is composed of three main steps: primer extension, hybridization, and detection.

### 1. Primer Extension

The 12 SNP primers for each panel are pooled at a concentration of 10  $\mu\text{M}$  each. The extension mix cocktail is prepared using SNPware<sup>®</sup> Reagent Kit A and B (part numbers 10105400 and 10105000,

Table 4). A quantity of 7  $\mu\text{L}$  of the cocktail is then added to the sample wells containing the correspondingly treated PCR products. The plate is spun briefly and incubated following the program shown in Table 5.

**Table 4. Prepare the Primer Extension Pre-Mix Using SNPware Reagent Kits**

	Manual		Beckman Coulter Automation		Other Automation	
	2	4	2	4	2	4
Number of Plates	2	4	2	4	2	4
SNP Primer Pool (10.00 $\mu\text{M}$ each)	27.00 $\mu\text{L}$	54.00 $\mu\text{L}$	28.90 $\mu\text{L}$	52.50 $\mu\text{L}$	35.00 $\mu\text{L}$	58.00 $\mu\text{L}$
SNPware Extension Mix (1 out of the 6 types)	180.00 $\mu\text{L}$	360.00 $\mu\text{L}$	189.80 $\mu\text{L}$	345.00 $\mu\text{L}$	230.00 $\mu\text{L}$	460.00 $\mu\text{L}$
SNPware Extension Dilution Buffer	3385.00 $\mu\text{L}$	6770.00 $\mu\text{L}$	3567.00 $\mu\text{L}$	6486.00 $\mu\text{L}$	4324.00 $\mu\text{L}$	8648.00 $\mu\text{L}$
SNPware DNA Polymerase	18.80 $\mu\text{L}$	37.60 $\mu\text{L}$	19.80 $\mu\text{L}$	36.00 $\mu\text{L}$	24.00 $\mu\text{L}$	48.00 $\mu\text{L}$
H <sub>2</sub> O	2673.00 $\mu\text{L}$	5346.00 $\mu\text{L}$	2835.00 $\mu\text{L}$	5154.00 $\mu\text{L}$	3436.00 $\mu\text{L}$	6872.00 $\mu\text{L}$
Total Volume	6.30 mL	12.60 mL	6.60 mL	12.00 mL	8.00 mL	16.00 mL

**Table 5. Thermal Cycling Conditions for Primer Extension**

Step	Temp	Time
1	96°C	3 min
2	94°C	20 sec
3	40°C	11 sec
4	Go to step 2 for 45 times	
5	4°C forever	

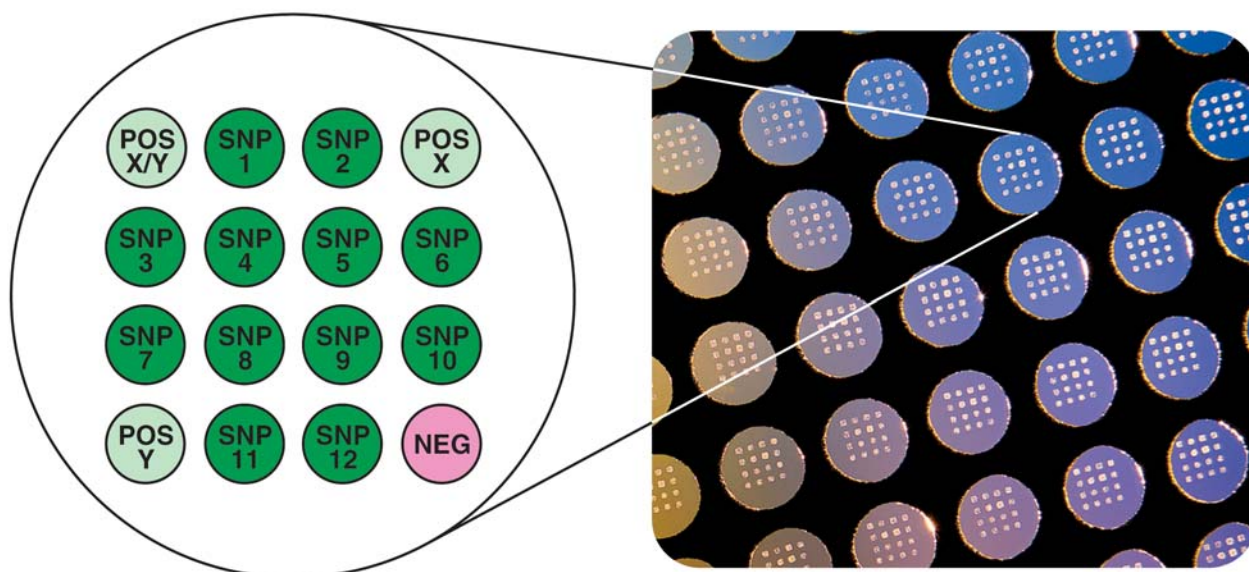
## 2. Hybridization

Each of the SNP primers in the multiplex has a unique 5'-tagged sequence that is spatially resolved by hybridization to the tag array plate. The tag array plate is manufactured in a 384-well plate format on a novel, glass-bottomed plate and has 16 reaction sites per well. Twelve sites contain tag oligonucleotides complementary to the unique tagged sequences on the SNP primers, while the other four sites are control sites. Figure 3 shows the layout of the SNPware® Tag Array Plate (part number 10106400) with positions of positive (XX, XY, and YY) controls and the SNP capture tags. The XX- and YY-positive spots contain tags to capture one of the two allele-specific self-extension control oligonucleotides that are included in the SNPware Extension Mix. The XY-positive spot uses a mixture of both capture tags for X and Y controls. The negative control spot serves as a test for mishybridiza-

tion in the assay. SNP1 to SNP12 indicate the positions of the specific capture tags complementary to the SNP primer tags. Following primer extension, the hybridization mix is prepared (Table 6), 8 µL aliquoted into each sample well, mixed, and 8 µL transferred to the corresponding well on the SNPware Tag Array Plate (which has been washed three times using SNPware Wash Buffer I and vacuum dried). The array plate is incubated at 42°C and 100% humidity for 2 hours ± 15 minutes. Following the incubation, the plate is washed three times using SNPware Wash Buffer II and vacuum dried. The plate is kept in the dark prior to image detection.

## 3. Detection

The SNPstream® Imager was developed and optimized for moderate to ultra-high-throughput array image analysis. The Imager controlling software, Run Manager, allows the operator to image array plates either manually or automatically at a speed of ~7 minutes per plate. A robotic plate handler is used to enable automated loading of plates into the Imager. A bar code reader reads the unique bar code printed at the side of each array plate to enable sample tracking. The Imager is based on a two-laser and two-color approach. Each sample well is first illuminated with a 488-nm (blue) laser beam and subsequently with a 532-nm (green) laser beam. The raw image data is collected and stored in the controller according to the plate bar code. The system is capable of processing up to 72 plates unattended. At a 12-plex level, over 165,000 genotypes can be generated in an 8-hour period. In addition, the sys-



**Figure 3.** The layout of the SNPware Tag Array.

**Table 6. Preparation of the Hybridization Solution**

	Manual		Beckman Coulter Automation		Other Automation	
	2	4	2	4	2	4
Number of Plates	2	4	2	4	2	4
SNPware Hybridization Solution	9.45 mL	18.90 mL	7.10 mL	13.20 mL	18.90 mL	33.10 mL
SNPware Hybridization Additive	550.00 $\mu$ L	1.10 mL	400.00 $\mu$ L	800.00 $\mu$ L	1.10 mL	1.90 mL
Total Volume	10.00 mL	20.00 mL	7.50 mL	14.00 mL	20.00 mL	35.00 mL

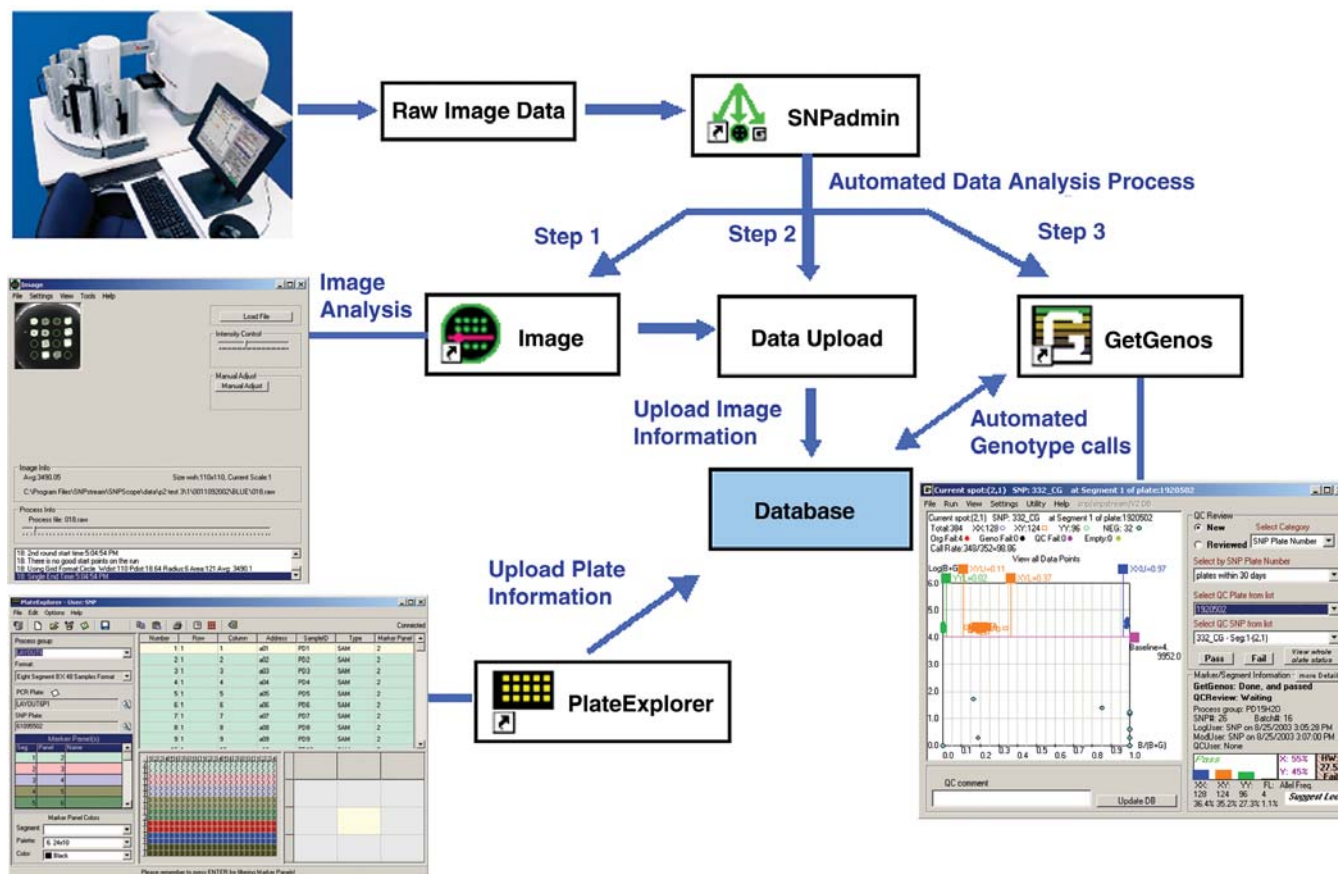
tem is scalable, so that as many as 890,000 genotypes can be scored in 24 hours.

**SNPstream® Software Suite and Data Analysis**

The SNPstream software suite is composed of software modules that work together to provide a seamless workflow from assay design to 384-well plate setup, to image data analysis, and finally to genotype calls (Figure 4). The software modules include *Autoprimer.com* for primer design; *PlateExplorer* for plate layout and experimental set up; *Image* for raw image analysis; *GetGenos™* for automated genotype calls and *QCreview™*; *Edit* for organizing the database; and finally *SNPadmin* for automated data analysis from image data to genotype calls.

As mentioned earlier, the output file from *Autoprimer.com* contains the probe panel information on each newly designed SNP panel in tab-delimited text format. This text file can be uploaded directly to the database using *PlateExplorer*. A unique number is assigned to each SNP panel as the marker panel ID for future reference. *PlateExplorer* is also used to upload the plate layout information, defined as a process group, on sample data, well data, and plate format. During experimental setup, the tag array plate bar code is associated with the process group and marker panel ID in the database prior to data analysis.

Generation of the genotype calls from spots imaged by SNPstream Imager involves three steps, all fully automated through *SNPadmin*. First, the location and intensity of a spot within the well and



**Figure 4. Overview of the SNPstream Software Suite.**

**Table 7. CEPH/Utah Family Pedigrees and DNA Sample IDs**

CEPH Pedigree ID	Paternal Grand-father	Paternal Grand-mother	Maternal Grand-father	Maternal Grand-mother	Father	Mother	Son	Daughter
I334	NAI2144	NAI2145	NAI2146	NAI2239	NAI0846	NAI0847	NAI2138	NAI2139
I333	NA07049	NA07002	NA07017	NA07341	NA07038	NA06987	NA07009	NA07011
I340	NA06994	NA07000	NA07022	NA07056	NA07029	NA07019	NA07040	NA07053
I341	NA07034	NA07055	NA06993	NA06985	NA07048	NA06991	NA07020	NA07006
I347	NAI1879	NAI1880	NAI1881	NAI1882	NAI0858	NAI0859	NAI1878	NAI1870
I362	NAI1992	NAI1993	NAI1994	NAI1995	NAI0860	NAI0861	NAI1984	NAI1986
I416	NAI2248	NAI2249	NAI2250	NAI2251	NAI0835	NAI0834	NAI2245	NAI2246
I350	NAI1829	NAI1830	NAI1831	NAI1832	NAI0856	NAI0855	NAI1822	NAI1824
I421	NAI2282	NAI2283	NAI2286	NAI2287	NAI0841	NAI0842	NAI2281	NAI2280
I424	NAI1930	NAI1931	NAI1932	NAI1933	NAI0845	NAI0844	NAI1925	NAI1926
I458	NAI2840	NAI2841	NAI2842	NAI2843	NAI2831	NAI2832	NAI2836	NAI2837
I582	NAI2921	NAI2922	NAI2923	NAI2924	NAI2911	NAI2912	NAI2916	NAI2918

plate are determined for each wavelength; second, information on the image results is uploaded to the database and associated with the corresponding plate and marker panel information; finally, a genotype call is made based on the relative fluorescent intensities of each spot. Once the genotype call is made, results are written to the database where data can be retrieved easily for viewing. The Image application enables automated image analysis by using a grid-drawing algorithm and converting image spots to numeric values of spot intensities. Genotype calls are made once all spot intensities are loaded into the database. Each SNP is analyzed separately by GetGenos™. The genotypes are based on the data point clusters when plotted on an X-Y graph, where X corresponds to the signal intensity in the 488-nm channel and Y to the 532-nm channel. A sample is failed automatically if the point falls between clusters or signal intensity is too low. Otherwise, it is called XX, XY, or YY, with the X or Y referring to the true allele (A, G, T, or C). The data is presented in both log scatterplot and angle scatterplot formats that allow the user to make a final pass/fail review. Adjustable thresholds are incorporated to accommodate SNP-to-SNP and assay variability. This flexibility helps to maximize concordance and throughput while reducing the potential of failed assays. Once QC review is performed, results are stored in the database by plate bar code number and well and spot location.

### CEPH Family Study

To demonstrate the system performance of SNPstream®, 180 unique SNPs were tested on 96 DNA samples from 8 family members of 12 CEPH/Utah families (Table 7). At 12-plex levels,

these SNPs were divided into 15 marker panels: 2 C/A panels (24 SNPs), 2 T/A panels (24 SNPs), 3 C/G panels (36 SNPs), 2 T/G panels (36 SNPs), 3 G/A panels (36 SNPs), and 3 T/C panels (36 SNPs). Primers were synthesized using standard methods. PCR primers and SNP primers were pooled separately and diluted to a final concentration of 10 µM each. Genomic DNA samples were diluted to 1 ng/µL each. Biochemical assays were performed utilizing Biomek® FX Liquid Handling Systems from Beckman Coulter (Ref. 1) for reaction setup and liquid transfer. Following hybridization, tag array plates were imaged using the SNPstream Imager. The raw image data was collected and analyzed automatically through SNPadmin. The genotype results were summarized to calculate the assay conversion rate and call rate for each panel. Data concordance was calculated by checking Mendelian inheritance within each family and data concordance among sample replicates and multiple instruments. As shown in Table 8, our results demonstrate 87.78% assay conversion at 12-plex, 98.18% genotyping success rate, and data concordance of 99.88%. The high conversion rate and the robustness of the SNP genotyping assays, coupled with the speed and ease of use provided by the GenomeLab SNPstream Genotyping System, make large-scale, ultra-high-throughput SNP screening achievable.

### Conclusion

- The SNPstream Genotyping System is a scalable genotyping system for medium to ultra-high-throughput SNP analysis.

**Table 8. The Performance of the GenomeLab™ SNPstream® Genotyping System on Scoring 15 SNP Panels (180 SNPs) in 12 CEPH/Utah Families**

*Assay conversion rate: passed SNPs/total SNPs; call rate or genotyping success rate: for all the passed SNPs, successfully genotyped samples/total samples; data concordance: for all the successfully genotyped samples, correct genotype calls/total genotype calls.*

Panels	Passed SNPs	Failed SNPs	Conversion Rate	Call Rate	Data Concordance
C/A Panel 1	11	1	91.67%	97.88%	99.97%
C/A Panel 2	12	0	100.00%	99.65%	100.00%
T/A Panel 1	9	3	75.00%	99.17%	100.00%
T/A Panel 2	9	3	75.00%	98.58%	99.51%
C/G Panel 1	9	3	75.00%	99.38%	99.76%
C/G Panel 2	12	0	100.00%	98.86%	99.77%
C/G Panel 3	10	2	83.33%	95.12%	100.00%
T/G Panel 1	11	1	91.67%	98.80%	99.97%
T/G Panel 2	12	0	100.00%	96.70%	99.88%
G/A Panel 1	10	2	83.33%	98.96%	99.89%
G/A Panel 2	12	0	100.00%	99.24%	99.77%
G/A Panel 3	11	1	91.67%	97.54%	99.84%
T/C Panel 1	9	3	75.00%	98.90%	99.88%
T/C Panel 2	10	2	83.33%	98.12%	100.00%
T/C Panel 3	11	1	91.67%	95.83%	100.00%
Average	10.50	1.50	87.78%	98.18%	99.88%

- The complete GenomeLab SNPstream system includes reagents and a hardware and software suite that enables high-quality results.
- The system performs SNP analysis using SNP Identification Technology in a tagged-array format.
- The versatile design of the system allows manual or fully automated processing in assay setup, array imaging, and data analysis.
- The SNPstream Genotyping System successfully scored 180 SNPs in the CEPH/Utah families with high degrees of assay robustness and data concordance.

## Reference

1. Roby, K., Campbell, D., Jiang, Z., Pajak, L., Pittman, C. High Throughput Sample Preparation for the SNPstream® Genotyping System: Automation of Target Amplification and Post-PCR Processing on the Biomek® FX Liquid Handling System. Beckman Coulter *Application Information Bulletin A-1969A* (available 12/03).

\* The PCR process is covered by patents owned by Roche Molecular Systems and F Hoffmann-La Roche, Ltd. AmpliTaq Gold is a registered trademark of Roche Molecular Systems, Inc. All other trademarks are the property of their respective owners.



*Developing innovative solutions in genetic analysis, drug discovery, and instrument systems.*

**Innovate** **Automate**  
SIMPLIFY

Beckman Coulter, Inc. • 4300 N. Harbor Boulevard, Box 3100 • Fullerton, California 92834-3100

Sales: 1-800-742-2345 • Service: 1-800-551-1150 • Telex: 678413 • Fax: 1-800-643-4366 • www.beckmancoulter.com

### Worldwide Biomedical Research Division Offices:

**Australia** (61) 2 9844-6000 **Canada** (905) 819-1234 **China** (86) 10 6515 6028 **Eastern Europe, Middle East, North Africa** (41) 22 994 07 07  
**France** 01 49 90 90 00 **Germany** 49 21 513335 **Hong Kong** (852) 2814 7431 / 2814 0481 **Italy** 02-953921 **Japan** 03-5404-8359  
**Mexico** 525-605-77-70 **Netherlands** 0297-230630 **Singapore** (65) 6339 3633 **South Africa/Sub-Saharan Africa** (27) 11-805-2014/5 **Spain** 91 3836080  
**Sweden** 08-564 85 900 **Switzerland** 0800 850 810 **Taiwan** (886) 2 2378 3456 **Turkey** 90 216 309 1900 **U.K.** 01494 441181 **U.S.A.** 1-800-742-2345