Introduction

The HuSNP™ mapping assay contains 1,494 single nucleotide polymorphism markers originally discovered in collaboration with the Whitehead Institute Center for Genome Research. There have been a number of collaborations and in-house projects that have attempted to map the entire set of HuSNP™ markers on the human genome. These efforts have resulted in multiple maps, that although all partial, are of great value since they provide redundant and concordant map data for a large proportion of HuSNP™ markers.

This technical note describes an updated composite HuSNP™ WIAF marker map with chromosomal assignments for 1,465 of the 1,494 markers, and centiMorgan (cM) positions for 1,308 of the markers. The composite map provides some map information for approximately 98% and complete map information for 88% of the HuSNP™ WIAF markers. This map adds great value to the genotype data obtained from the HuSNP™ assay allowing users to better utilize and gain more information from their assay results. Described here are also the various maps built for the HuSNP™ marker set over the past three years, as well as data comparing quality and concordance among these maps. The information contained in all of the maps described here is available on the Affymetrix website: www.affymetrix.com

HuSNP™ Individual Maps

GB4 - GenBridge 4 Radiation Hybrid Map

The first Radiation Hybrid (RH) Map for HuSNP™ markers was built by the Whitehead Institute Center for Genome Research utilizing the GenBridge 4 RH panel (GB4), in conjunction with a skeletal framework of Genethon microsatellite markers. Initial mapping efforts at Whitehead Institute resulted in obtaining chromosomal assignments and Estimated Genetic Distances (in centiMorgans) for about two-thirds of the HuSNP™ WIAF markers. This map was used in the HuSNP™ analysis output using the Affymetrix GeneChip® Analysis package, version 3.2 and 3.3.

TNG - The Next Generation Radiation Hybrid Maps

Another collaboration with the Stanford Human Genome Center (SHGC) using the The Next Generation (TNG) high-resolution RH panel, resulted in a second RH map. This map was included in the HuSNP™ analysis output by the Affymetrix® Microarray Suite version 4.0 software.

GP1/GP2 - Golden Path

At Affymetrix, we attempted to map the HuSNP™ markers electronically by BLASTing the WIAF SNP sequences against the Human Sequence Draft available through the UCSC site (Golden Path). There were two iterations of this sequence BLAST mapping approach at varying degrees of stringency, resulting in two maps: GP1 at high stringency, and GP2, at lower stringency. The GP1 and GP2 electronic maps provided chromosomal assignments and kilobase positions for 1,052 and 1,278 markers respectively.

<table>
<thead>
<tr>
<th>Map Source</th>
<th>Date</th>
<th>Map Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>WI: GenBridge 4 RH panel map</td>
<td>5/99</td>
<td>GB4</td>
</tr>
<tr>
<td>SHGC: TNG RH panels last build</td>
<td>7/00</td>
<td>TNG</td>
</tr>
<tr>
<td>Affymetrix: Golden Path BLAST build 1 (high stringency)</td>
<td>4/01</td>
<td>GP1</td>
</tr>
<tr>
<td>Affymetrix: Golden Path BLAST build 2 (low stringency)</td>
<td>4/01</td>
<td>GP2</td>
</tr>
<tr>
<td>JHMI: NCBI MapView BLAST</td>
<td>7/01</td>
<td>JHMI</td>
</tr>
<tr>
<td>JHMI: Genetic map</td>
<td>9/00</td>
<td>GM</td>
</tr>
</tbody>
</table>

Table 1. Individual HuSNP™ WIAF marker maps and sources.
We expected varying levels of concordance due to experimental variations in comparing any two of these maps. Most importantly, these maps fall into two main methodology types:

- The Radiation Hybrid maps, GB4 and TNG were obtained by experimental interrogation of hamster-human hybrid DNA panels to assign chromosome locations and centiRays positions. Based on comparisons to genetically mapped skeleton microsatellite marker sets, cM distances are then determined by interpolation.

- GP1, GP2 and the JHMI, were obtained electronically by BLAST homology searches of the SNP region sequences against the human sequence draft. These searches resulted in a high concordance rate of over 97% in all pair-wise comparisons both in inter- and intra-type map comparisons (see Figure 1).

We next compared the marker positions from the JHMI sequence-based map to the TNG RH map to estimate the level of concordance. In our pair wise comparisons, these two maps showed the highest markers in union, but the lowest level of chromosomal assignment.
concordance of 97%. Since these two maps were created using very different methodologies, we expected a moderate level of discordance between the kilobase and cM values. However, as shown in Figure 2, a relatively high correlation coefficient of 0.884 was observed, with a scattering of approximately 30 outlying markers. Overall, most of these markers show tight correlation, although some non-linearity of one method relative to the other is evident in the slightly sigmoidal deflection of the correlation plot. This high level of concordance between maps obtained through very different experimental approaches demonstrates a high level of confidence regarding the reproducibility and accuracy of the HuSNP™ map data.

**HuSNP™ Update Composite Maps**

Although valuable individually, none of these individual maps provide information for all HuSNP™ markers. In order to provide a more comprehensive and informative map, we have created an updated composite map that provides chromosomal assignments for 1,465 markers and Estimated Genetic Distances (cM) positions for 1,308 of the 1,494 HuSNP™ markers. This map incorporates chromosomal assignment data derived from a number of map sources, (see Figure 3). Below are the steps used in the creation of this map:

- Chromosomal assignments for an additional 157 markers were added based on concordant map information among the other four available maps, GP1, GP2, TNG and GB4. However, no Estimated Genetic Distances (cM) were included from these other maps so as to eliminate the cM distance non-linearity problems that are inevitable in a composite map.
- Due to discordance among maps, there are 29 of the 1,494 HuSNP™ markers that have no chromosome or cM information assigned in the updated composite map.

We attempted to determine the extent of linearity in the composite map between the kilobase positions obtained through the electronic map and the cM positions for 1,308 SNPs obtained through validation with microsatellite and SNP genetic maps (Kashuk, et al.). We observed a correlation coefficient of 0.945 for all 1,308 markers (see Figure 4). The same comparison is shown on a chromosome-by-chromosome basis in Figure 5 demonstrating the extensive coverage of the entire genome by the HuSNP™ WIAF markers. This correlation is highly linear, though to different extents among chromosomes.

*Figure 3. Map information source for the HuSNP™ Composite Map is shown for both chromosome assignments and cM distances. The JHMI map provides approximately 88% of chromosomal assignments and 100% of cM positions for the resulting composite map. The GB4, GP1, GP2 and the TNG maps account for 5.2% to 1.3% of the chromosomal assignment information.*

*Figure 4. CentiMorgan positions are plotted for 1,308 markers against the Kb positions for the Composite Map, showing a correlation coefficient of 0.945.*
Marker Gap Analysis

Marker distribution and gap sizes are important criteria to consider for any set of mapping makers. To illustrate the HuSNP™ Mapping Assay's marker distributions, we performed gap analysis on the 1,308 SNPs that have both chromosome and cM position assignments. We observed that over 95% of the gaps are 10 cM or less, while approximately 62% of gaps are 2 cM or less in size. The median gap size is 1.19 cM and the average gap size is 2.57 cM (See Figure 6). The distribution of gaps is fairly uniform across the genome both in number and size (See Figure 7). Gaps of very small sizes show dense distribution along each chromosome and no consecutive large gaps are present in any one region.

It is important to note that this analysis represents an overestimation of gap sizes, since the remaining 186 markers, once mapped with cM position assignments, are likely to reduce gap sizes further.

Figure 5. Comparisons of cM vs. Kb positions on a chromosome-by-chromosome basis. Positions for 1,308 markers are plotted after sorting the markers -- first by chromosome, then by cM from 0 at the short arm telomer to the end of the long arm telomer. Chromosomes 1 through X are listed from left to right.

Figure 6. Composite Map - Gap Frequency. All 1,308 markers are used to determine 1,285 gap differences. Gap bin sizes of 2 cM are plotted along the x axis, within each bin the number of gaps (in green, histogram graph) and percentage of gaps (in black, line graph) are plotted on the second y axis.
Acknowledgements:

We would like to acknowledge and thank the many individuals involved in our collaborations who have greatly contributed to the mapping efforts described here. The following lists our collaborators and the websites where most of the map data and results reported here have been made available.

The information contained in all of the maps described here is also available on the Affymetrix website.

www.affymetrix.com

**WI:**
Whitehead Institute for Biomedical Research / MIT Center for Genome Research: Dr. Eric Lander, Dr. Kerstin Lindblad-Toh, (GB4)
http://carbon.wi.mit.edu:8000/cgi-bin/SNP/human/SNP_map

**SHGC:**
Stanford Human Genome Center: Dr. David Cox, Dr. Michael Olivier (TNG)
http://shgc.stanford.edu/

**JHMI:**
Johns Hopkins Medical Institute, Genetic Medicine Institute: Dr. Aravinda Chakravarti, Dr. Audrey Lin, Carl Kashuk (JHMI)
UCSC: http://genome.ucsc.edu/

**NCBI mapview file:**

**Affymetrix:** Alan Williams, Brant Wong, Ray Wheeler, and David Kulp (GP1, GP2)

---

**Summary**

Our current composite map provides the most complete map data to date, allowing users to take better advantage of their HuSNP™ genotype data. The updated map provides some map information for 98% of the 1,494 markers. With median marker gap size of only 1.2 cM, the 1,308 fully mapped markers are evenly distributed across the genome with the exception of less dense representation on the X chromosome.

Map updates to the 1,494 WIAF HuSNP™ Mapping Assay markers have been on-going projects and will likely continue to be performed through work at Affymetrix, as well as through collaborations and publications resulting from mapping projects utilizing the HuSNP™ Mapping Assay. In addition, as the Human Draft Sequence becomes more accurately annotated and sequence gaps are eliminated, updates to the HuSNP™ map can provide important information regarding proximity of these markers to mapped genes.

**Figure 7.** Gap size and distribution is shown for 1,308 markers sorted by chromosome then by cM, where chromosomes 1 through X are listed from left to right.