rescheduled to 12/20

BACKGROUND FOR DISCUSSION AT IC DIRECTOR'S MEETING, DECEMBER 9, 1999

Update on the status of SNP Research **December 16, 1999**

Number of SNPs to be found

The number of base pairs in human DNA is estimated to be about 3 billion. Of these, about 6 to 10 million sites are expected to harbor common variants among individuals. Some of the variants directly cause functional differences; the others can be used to map those that do.

There are several efforts at NIH and elsewhere to discover large numbers of these variants, most of which are single nucleotide polymorphisms (SNPs). Some of the efforts aim to find SNPs any place in the genome, called random SNPs. Other efforts aim to find SNPs that are in the coding regions of genes (cSNPs) or are near genes, which are more likely than random SNPs to affect the function or regulation of genes.

The table below shows the number of SNPs expected from the currently funded major current efforts to find SNPs and place them in the public domain.

Anticipated number of SNPs

	Random SNPs	cSNPs or SNPs in gene regions
SNP RFA (NIH)	32,000	24,000
SNPs from large-scale sequencing (NHGRI)	100,000	
CGAP (NCI)		30,000
The SNP Consortium	310,000	
Japan		200,000
Total	442,000	254,000

So it is estimated about 700,000 SNPs will be discovered over the next two years. A meeting is planned in March, 2000, jointly sponsored by NIH and TSC, to evaluate current data about linkage disequilibrium and determine whether this number of SNPs is likely to be sufficient to enable whole genome association analysis.

Need for better genotyping technology

Producing the SNP catalog is not the only challenge. In order to use these SNPs to find the genes associated with disease, the technology for genotyping will need to be much cheaper and usable on a larger scale than is currently available.

The SNP Consortium plans to ascertain what will be needed in technologies in order to use SNPs for research on the genetic basis of disease, and then to evaluate the available technologies to see how well they will meet these needs. Examples of technologies to be considered include minisequencing, dHPLC, oligonucleotide arrays with hybridization (chips), beads with hybridization, TaqMan with hybridization, reverse dot-blot arrays, microarrays with primer extension, Invader with primer extension and allele-specific cleavage of a tag, fluorescence polarization with primer extension, primer extension with antibody-based detection

(Genetic Bit Analysis) or with mass spec detection, and template-directed dye-terminator incorporation (TDI).

Proposal for NIH Collaboration with the SNP Consortium

Genotyping Technology Assessment

The ability to genotype many individuals will be important for studies in disease associations and pharmacogenomics. Currently many technologies are being developed for large-scale genotyping, but it is not clear which ones will be most efficient and cost-effective. The SNP Consortium (TSC) plans a project to ascertain the research needs for genotyping technology, and then compare how various current technologies meet those needs.

The assessment will be directed towards the research needs of the pharmaceutical companies and others who aim to do high throughput SNP genotyping. The variables to be considered will include throughput, sensitivity, accuracy, robustness, cost, scalability, multiplexing potential, risk, and the maturity of the technology. There is more than one type of research need. For example, some association studies will require genotyping thousands of individuals for hundreds of thousands of SNPs; some pharmacogenomics studies will require genotyping many individuals for only one or a few SNPs.

Once the needs have been identified, the available technologies will be evaluated on how they meet these needs, or are anticipated to. The comparison will be based on 1) user surveys, 2) information that the genotyping companies provide, and 3) experiments done in an independent laboratory. About 10 – 15 technologies will be compared. The samples used for genotyping will be from the NIH DNA Polymorphism Discovery Resource, so useful frequency information will result. The plan aims to have results in a year, which will be publicly available.

TSC expects the cost will be about $1.4 million. They are asking NIH to contribute $700,000 of this. NIH's involvement in this project will mean that the research needs of the academic community will be considered as well as those of the pharmaceutical industry. In addition, the funding will allow TSC to have much more work done by an independent laboratory comparing the technologies, rather than simply relying on data provided by the genotyping technology companies. TSC will soon be submitting a detailed proposal to NHGRI.

SNP Meeting

In a year or two there will be hundreds of thousands of SNPs found by NIH-funded researchers, TSC, and others. To discuss whether more large-scale SNP discovery will be needed, NHGRI is planning a meeting on March 7-8, 2000, in the DC area, to assess what is known about patterns of human genetic variation and what this tells us about the number, location relative to genes, and allele frequency of SNPs needed to find genes associated with disease. TSC will cosponsor this meeting, and is providing pharmaceutical company expertise on mapping disease genes. The conclusions of this meeting will help NHGRI and other ICs decide whether more SNPs, and of what type – random, near genes, or in coding regions – are needed in order to find genes associated with disease. The conclusions will also help NIH decide how and in what areas it should collaborate with TSC in the future.