07/02/99

The
SNP Consortium Ltd.

The SNP Consortium--Program Summary
NIH Discussions 7/8/99

The SNP Consortium (TSC) has been formed to advance the field of medicine and the development of genetic based diagnostics and therapeutics, through the creation of a high density single nucleotide polymorphism (SNP) map of the human genome. This map will be freely available to all parties [members and non-members] at the same time.

Objectives

The [*minimum*] objectives of the Consortium are as follows:

- Identify 300,000 SNPs within two years of launch of the scientific work plan (4/99),
- Map 150,000 of the SNPs over the two year term of the program, and
- Manage publication of the resulting SNP map in a manner intended to maximize the number of SNPs that enter the "public domain" [as that term is understood in the patent law].

Scientific Plan

The TSC's scientific plan will be achieved by contracting with existing academic centers currently involved in genome sequencing and/or mapping of genetic markers. The SNP identification will be divided among the three research centers, including the Whitehead Institute, Washington University and the Sanger Center. Mapping will be conducted at the Stanford Genome Center. Data handling and bioinformatics will be provided by Cold Spring Harbor Laboratory.

SNP Identification will be performed by each center through sequencing of reduced representation fractions of the genome generated by selecting certain size fractions of restriction digests of genomic DNA pooled from 24 diverse individuals [from the NIH panel]. The genome centers have refined this technique through pilot studies and are in the process of scaling up their SNP identification production capacity. The initial SNP identification workload will be allocated equally across the Whitehead Institute, Washington University and the Sanger Center.

Agreements with these centers require they agree to the TSC's strategy for intellectual property handling, achieve quarterly production milestones, and provide for the right to terminate the work if production milestones are not met. The TSC reserves the right to re-allocate work across the three centers if agreed upon quarterly

performance levels are not achieved by an individual center. The individual center budgets have been standardized [to a certain degree] and are "roughly" equivalent at the outset.

The SNP identification "production targets" are summarized below:

Time Period	Target Total SNPs Identified	Center Production Target
1Q99	Pilot Study -None	0
2Q99— Start Period	4,000	1,333
3Q99	18,500	6,167
4Q99	33,750	11,250
1Q00— Year I End	41,250	13,750
2Q00	48,750	16,250
3Q00	48,750	16,250
4Q00	51,750	17,250
1Q01— Year II End	53,250	17,750
Total SNP ID	**300,000**	**100,000**

SNP mapping will be performed in silico to the extent permitted by the draft human genome sequence and is expected to yield approximately 24,000 SNPs from the 97,750 SNPs identified in year one. Additional mapping of 27,000 SNPs [mostly in the first year] will be conducted by the radiation hybrid method at a cost of approximately $193 per SNP ($6.9MM for the project). SNP mapping in year two is expected to be restricted to that which is achievable in silico, with a year two target of 99,000 additional mapped SNPs.

The Scientific Management Committee and the C.E.O. will oversee the progress of the scientific plan and ensure to the maximum extent possible the consortium scientific objectives are achieved.

Intellectual Property

The overall IP objective is to maximize the number of SNPs that (1) enter the public domain at the earliest possible date, and, (2) are free of third-party encumbrances such that the map can be used by all without financial or other IP obligations by all parties. To meet point (2), the TSC intends to withhold public release of identified SNPs until mapping has been achieved to prevent facilitating the patenting of the same SNPs by third parties. Mapped SNPs will be publicly released quarterly [in year

one] and monthly [in year two]. The intellectual property plan is intended to maintain the priority dates of discovery of the unmapped SNPs during the period between identification and release, for use as "prior art".

All parties will receive access to the SNPs at the same time. TSC members and the genome sequencing centers **will not** be granted advanced access to unpublished SNPs.

Current Budget for the TSC SNP Map

The two year budget for the current TSC SNP program is $44MM. The TSC program, as presently defined, is budgeted for $44MM. The budget will be allocated as follows:

- SNP ID=$31.5MM
- SNP mapping=$6.9MM
- Bioinformatics= $1.4MM
- Other R & D= $1.75 MM
- General & Administrative = $1.92MM
- External Advisors = $.73MM

The program has obtain agreements with 10 members and the Wellcome Trust [see the attached 4/99 press release]. Of this amount, the Wellcome Trust will contribute $14MM. Each TSC member will commit to provide a total of $3MM, split as $2MM in year one and $1MM in year two.

The SNP Consortium enthusiastically welcomes the opportunity to expand its public endeavors in conjunction with the NIH. The consortium strongly believes a map of 500,000 mapped SNPs will best serve the research needs of the medical, scientific, and pharmaceutical communities.

Submitted by:

Arthur Holden
Chairman and CEO,
The SNP Consortium

7/1/99

Summary of the SNP Meeting

June 7-8, 1999, Bethesda, MD

The SNP meeting brought together all the PIs from the SNP RFA as well as the principals from the SNP Consortium, to discuss issues related to coordination, SNP quality, resources, and databases. The conclusion of the meeting was that for SNPs to be most useful, several questions need to be addressed and several additional resources need to be provided.

Major scientific questions related to genetic variation that need to be addressed

Patterns of variation: How much variation and linkage disequilibrium (LD) exists, and how does this vary across the genome and by population? How do other factors affect this?

Comparative analyses: How can comparing patterns of variation within and among species, including other primates and mammals, be used to make inferences about function and selection?

The number and frequency of SNPs needed: How many SNPs are needed to address various questions? What allele frequencies do these SNPs need to have?

Function: How important is it to focus on SNPs in functionally important regions? How can variable sites be related to functional differences, particularly when there is LD across many sites?

Assessment of currently available information and resources

Linkage disequilibrium: After more data are generated, in about six months, it would be useful to have a meeting that summarizes what is known about linkage disequilibrium and identifies gaps in our knowledge.

Population samples: It would be useful to find out what population samples being collected by the NIH are available, to see whether they would be informative for a general population sample.

New resources needed

Technology for genotyping: To use SNPs to relate genotypes to phenotypes will require much better technologies than currently exist for cheap and large-scale SNP genotyping. An RFA on novel SNP genotyping technologies would potentially be useful.

Somatic cell hybrids of the DNA Polymorphism Discovery Resource, and complete hydatidiform moles: These lines are useful standards for detecting duplicate genes that are incorrectly assayed as SNPs, for defining haplotypes, and for technology development.

Primate samples: Standard samples of each species and subspecies would be useful for figuring out which human SNP alleles are ancestral, and comparing variation within and among species.

Human samples: Much more discussion is needed of the purposes and types of human samples. Another RFA addressing ELSI issues in defined groups would potentially be useful.

Analytical tools for SNP data: Methods are not yet adequate to analyze the large amount of data soon to be produced. Another RFA on complex trait and SNP analysis would potentially be useful.

SNP quality

Standards: a working group was set up to make recommendations about standard sets of samples, gene regions, and methods to assess SNP quality.

SNP RFA Awards

						Large-scale SNP discovery				
PI	Institution	Title	Regions	Technology development	Aim	Technology used	random	cSNPs	IC primary	IC secondary
Barany	Cornell U	Identifying genome changes in cancer development	Cancer-related genes	Arrays and a mismatch detection enzyme	Technology		100		NCI	
Boyce-Jacino	Orchid	SNPs on chips	Various genes	Arrays	Technology		400		NHGRI	
Chakravarti	Case Western Reserve U	Human genomic polymorphisms -- in populations of different ages	Gene regions related to neuropsychiatric disorders		Discovery	Chips	10,000		NIMH	NIA NIDA NINDS NIAAA NICHD NHGRI
Chee	Illumina	Randomly ordered DNA arrays for SNP discovery and typing		Bead decoding for random arrays	Technology				NHGRI	
Cox	Stanford U	A high resolution SNP map of the human genome	Random		Discovery	Chips	9,100		NHGRI	
Davis Oefner	Stanford U	Large-scale discovery of SNPs	Random and various gene regions		Discovery	dHPLC	4,000	2,000	NHGRI	NIDCD
Geraghty	Hutchinson	SNPs in the MHC	MHC		Discovery	Sequence	5,000		NHLBI	NIDDK
Hogan	Genometrix	The Risk/Tox chip program	Genes involved in toxicology and environmental carcinogenesis	Genotyping chips	Technology				NIEHS	
Kwok	Wash U	New methods for high-throughput genome analysis		Fluorescence polarization	Technology		10		NEI	
Kwok	Wash U	Method for global and targeted discovery of SNP markers	Random and disease genes	MutS protein	Technology		500		NIA	
Lander	MIT	Discovery of SNPs in coding regions of genes	Various genes		Discovery	Chips and dHPLC		20,000	NHGRI	NIA NIDCR
Myers	Stanford U	A mismatch enrichment method for discovering human SNPs	Random	MutS protein	Technology		1,000		NIDCD	
Nickerson	U Washington	Finding and genotyping SNPs by automated sequence analysis	T cell receptor genes	Computer programs to find SNPs in sequence traces	Technology	Sequence	1,200		NIAID	
Nolan	Los Alamos Nat. Lab.	High throughput SNP discovery and scoring using flow cytometry		Flow cytometry	Technology			15	NCRR	
Olson	U Washington	Methods for discovering and scoring SNPs	"Environmental" genes		Discovery	Sequence		1,000	NIEHS	NIAAA
Ton	U Washington	Discovery of cSNPs in nicotinic acetylcholine receptors	Nicotinic acetylcholine receptors		Discovery	Sequence		700	NINDS	
Weber	Marshfield	Human diallelic insertion/deletion polymorphisms	Random	Porous chips	Technology		1,000		NHLBI	
Total							32,310	23,715		

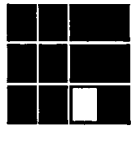
NIH Membership Discussions

"The SNP Consortium"

July 8, 1999

Arthur Holden, Chairman and CEO

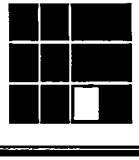
David Wang, M.D., Ph.D., SMC Chairman



The
SNP Consortium Ltd.

The SNP Consortium -An Overview

- Introduction
- Scientific Elements
- Management Elements



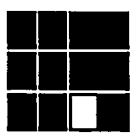
The
SNP Consortium Ltd.

Our Mission

"The SNP Consortium will strive to advance the field of medicine and the development of genetic based diagnostics and therapeutics, through the creation of a high quality, dense single nucleotide polymorphism (SNP) map of the human genome, which will be made available to all parties at no cost. We will emphasize scientific and operational excellence, teamwork and the public good in all of our activities."

Through this mission, we are committed to serving:

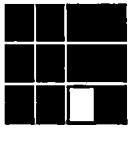
- The Medical Community,
- The Scientific Community, and
- Our Membership



The
SNP Consortium Ltd.

What?

- Non-profit organization [501c3] >> public benefit
- Development of "pre-competitive genomic databases"
- "Diverse" participants>>companies, private foundations, and academic collaborators
- Funded exclusively by "member" contributions



The
SNP Consortium Ltd.

Alternatives?

A variety of alternatives were examined:

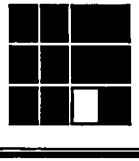
- Wait for "academic developments">>public domain,
- Work to accelerate the "public domain" efforts,
- Leave it to "specialized firms">>license post development

✔ Establish a strong, focused alliance to **collaborate** and **complement** ongoing public efforts.

The
SNP Consortium Ltd.

Why?

- ✔ Strong industry support>>common goal...make it happen.
- ✔ Desired outcome>>international, industrial standard SNP map
- ✔ Quality & timing of existing SNP initiatives>>inadequate
- ✔ Universal access Genomic/SNP maps>>public domain
- ✔ Maximize the total data available>>complement public efforts
- ✔ Economies of scale>>cost & risk sharing
- ✔ SNP Map>>move to other "appropriate" genomic opportunities

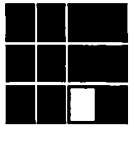


Founding Members

- ✔ AstraZeneca
- ✔ Bayer
- ✔ Bristol Myers Squibb
- ✔ Glaxo Wellcome
- ✔ Hoechst Marion Roussel
- ✔ Hoffman LaRoche
- ✔ Novartis
- ✔ Pfizer
- ✔ Searle
- ✔ SmithKline Beecham

plus:

- ✔ Wellcome Trust

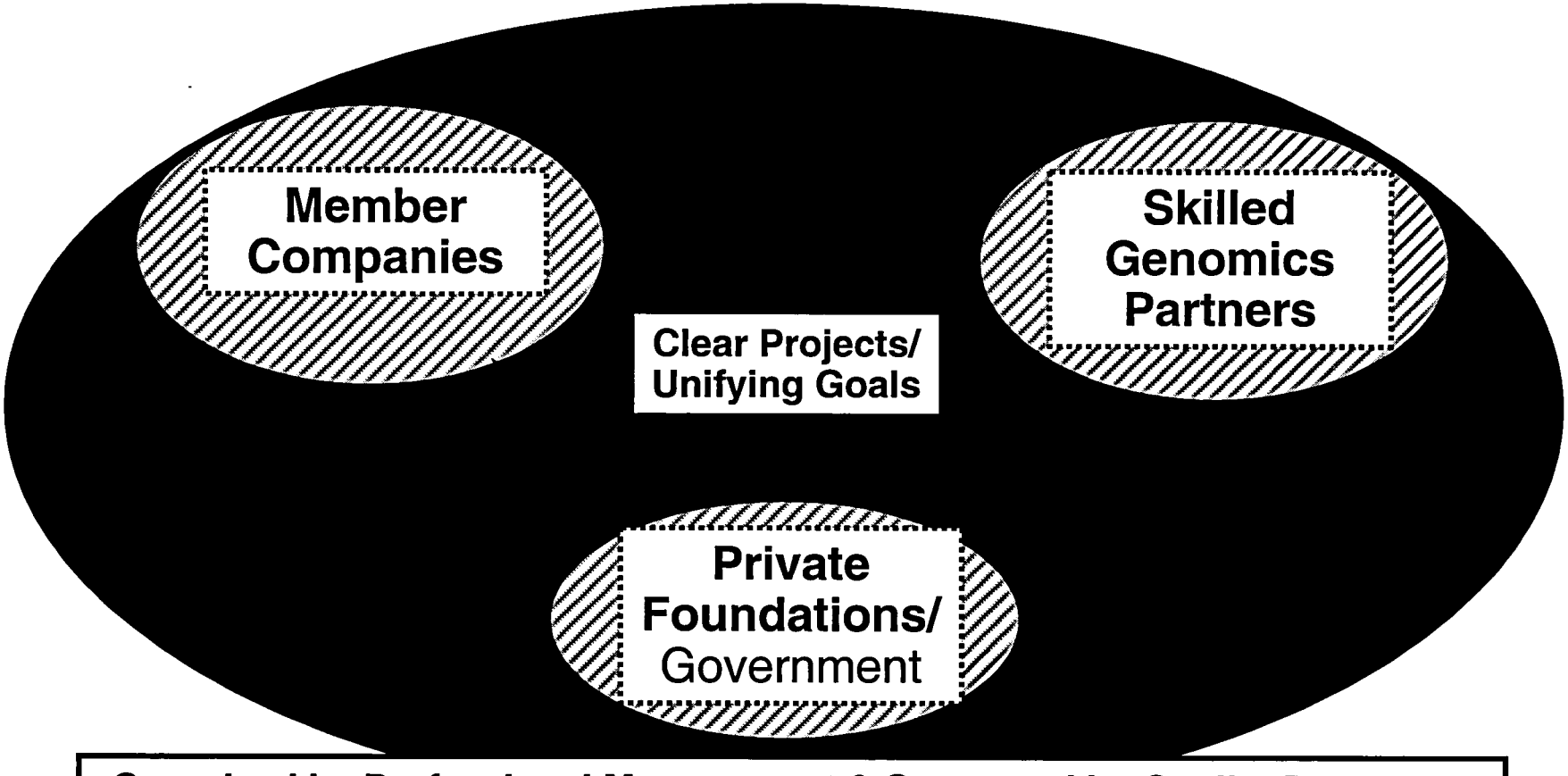


The
SNP Consortium Ltd.

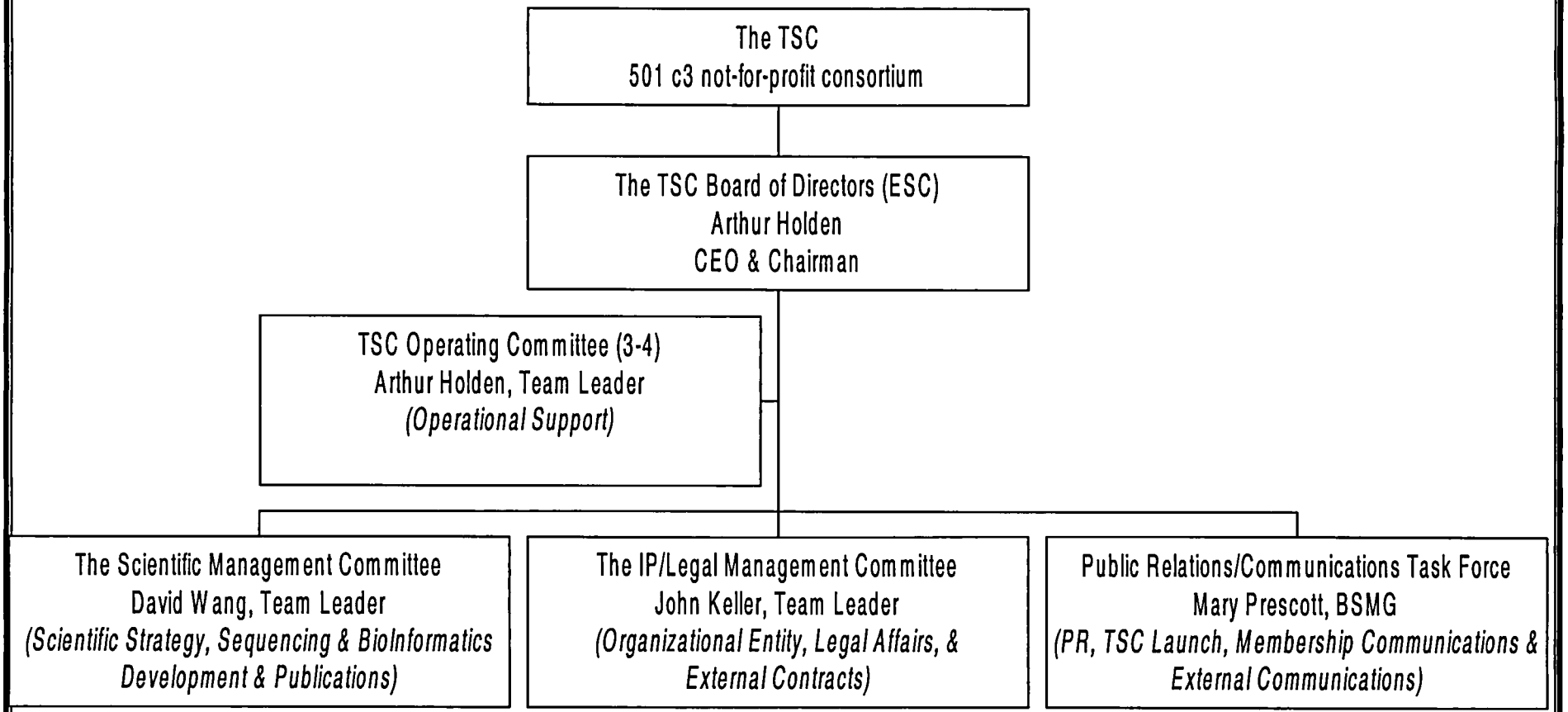
Member Commitments

- Management Time--Research & Governance
- Annual Financial Commitment
  - Companies--Fixed
  - Foundations/Trusts/Other--Variable
- No withdrawal unless "under performance"
- Coordinated "external" communications
- Planning

"Our Model"-Summary

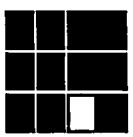


Organization



SNP Map Project-Key Objectives

- Create the highest quality, publicly available SNP map
- Identified 300,000 SNPs [minimum]
- Map at least 150,000 SNPs [broad, evenly spaced]
- Two year production period [4/99-4/01]
- Maximize public accessibility >> effective IP mgnt.
- < $45M -Budget [Phase I]--*[by leveraging existing investments, production scale, and a novel technical approach]*

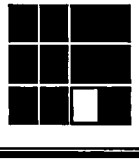


The
SNP Consortium Ltd.

Probable Major Genomic Developments 2000-04

- **early 2001-TSC** *first significant human genome SNP map.*
- **2000**--Working Draft Human Genome Sequence
- **2003**--Full Reference Human Genome Sequence
- **2004**--High Throughput Methods to Analyze Gene Function

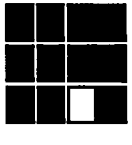
A Quality Human Genome SNP Map (300K+) is a highly significant contribution.



The
SNP Consortium Ltd.

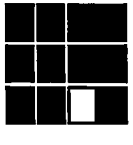
Progress Update

- Established entity & management processes
- Completed 11 F & F agreements
- Completed 5 research agreements
- Completed pilot studies with excellent results.
- Commenced production scale up
- Continuing to expand membership



A High-Density SNP Map of the Human Genome

Scientific Program Overview

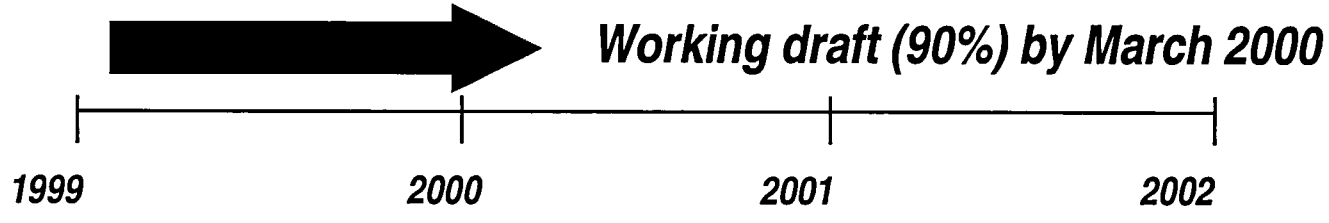


The
SNP Consortium Ltd.

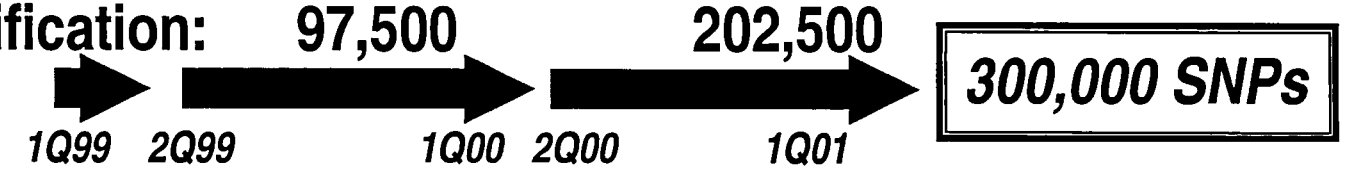
Scientific Plan

Objectives and Timetable:

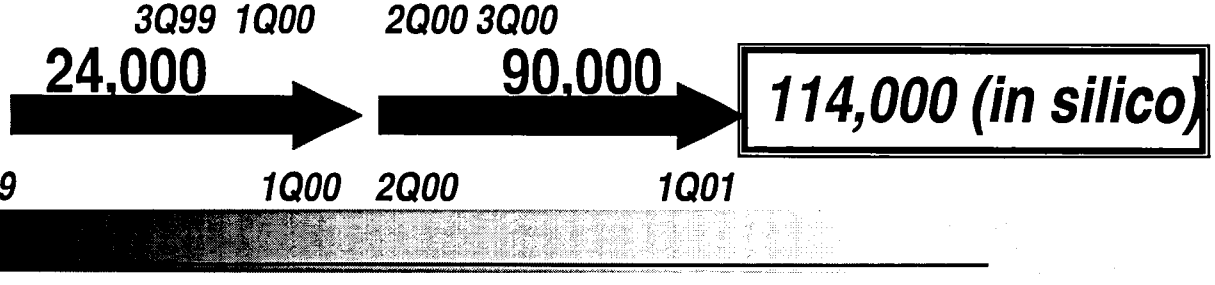
Human Genome Project:



SNP Identification:



SNP Mapping:



The SNP Consortium Ltd.

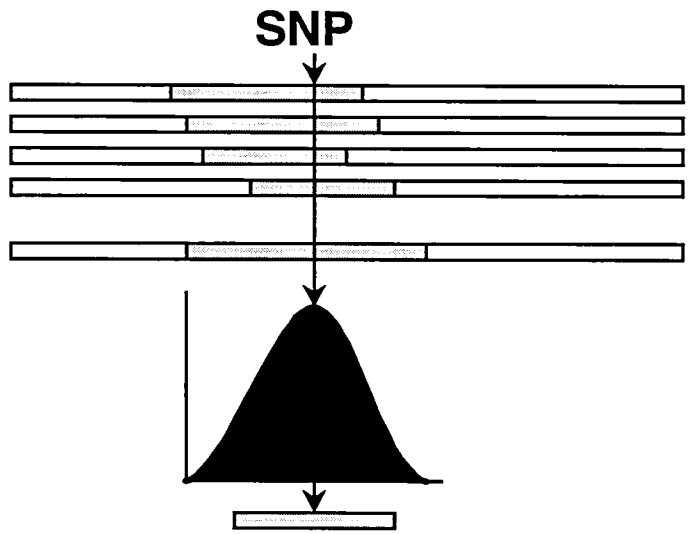
Review Points

- Scientific Approach
- Large-Scale SNP Discovery
- Creating an SNP Map
- Bioinformatics
- Current Status of the Scientific Plan

Number of SNPs Needed-Whole Genome SNP Map

Linkage disequilibrium (LD) in human populations:

Chromosome fragments:



Average size of LD fragment: **5 to 50 kb**

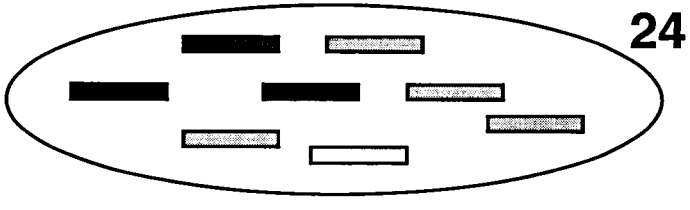
Number of SNPs needed for whole-genome studies:

1 SNP/LD fragment → **60,000 - 600,000 SNPs**

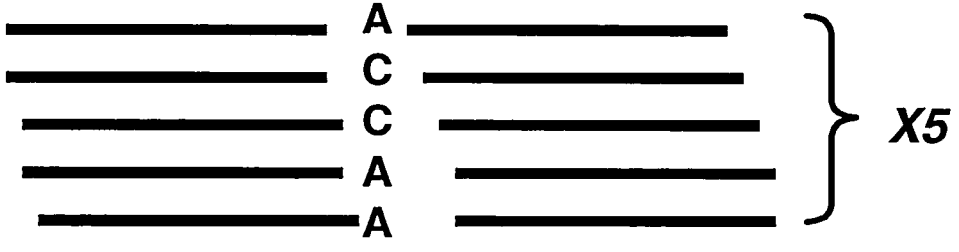
SNP Identification: Experimental Design

- Sequencing ~30% of the human genome 3-5 times

Representative Segments



24 unrelated individuals
- Caucasian
- African
- Asian, and Other



Shotgun Sequencing (x5)

[A/C]

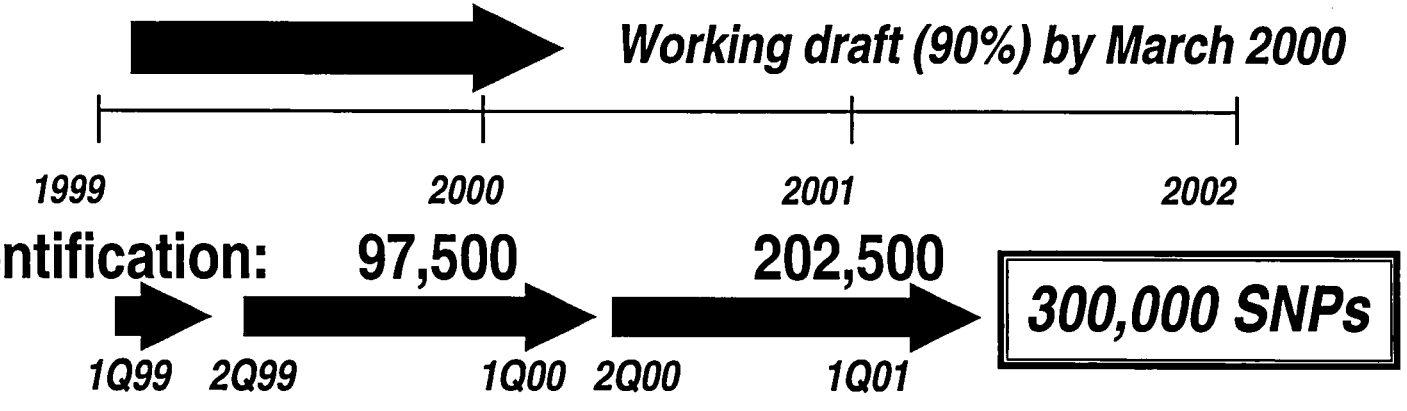
300,000 SNPs

The SNP Consortium Ltd.

Scientific Plan: SNP Identification

Objectives and Timetable:

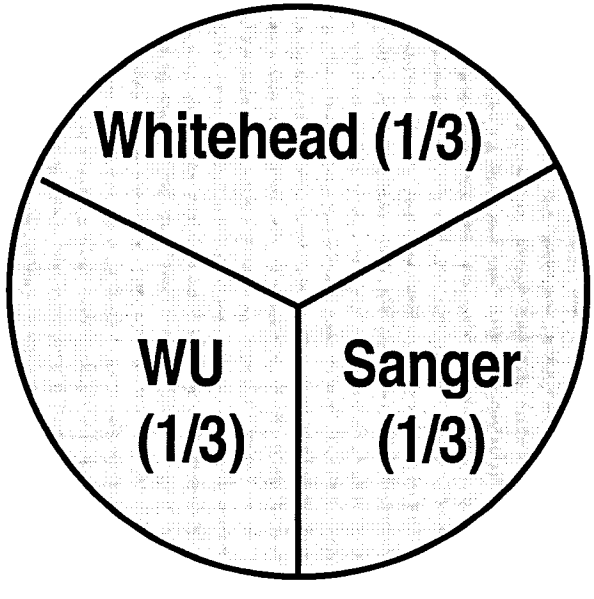
Human Genome Project:



Metrics: Measuring the Productivity & Quality (I) – *SNP Identification*

- **Metrics:**

Quarter	SNPs Identified	
Q1 99	*Pilot*	
Year 1:		
Q2 99	*4,000*	
Q3 99	*18,500*	
Q4 99	*33,750*	
Q1 00	*41,250*	97,500
Year 2:		
Q2 00	*48,750*	
Q3 00	*48,750*	
Q4 00	*51,750*	
Q1 01	*53,250*	202,500
Total		**300,000**



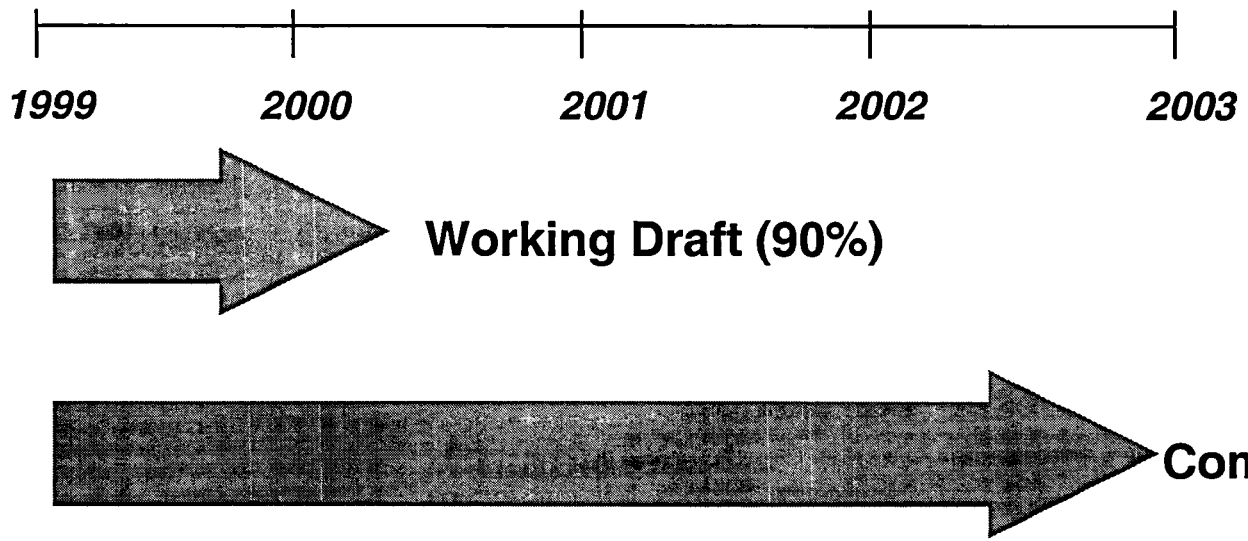
- **Quality: ~ 95% accuracy**

Scientific Plan: SNP Mapping

- Direct RH Mapping [24%]
- Mapped to at least 100kb resolution [>95%c]
- In Silico Mapping [76%]
- In Silico Mapping tied to progress on HGP [25%-Year 1; 45%-Year 2]
- Map assembly>>integrated mapped SNPs and other public SNPs, genes & sequences

SNP Mapping: In Silico and RH mapping

Human Genome Project:



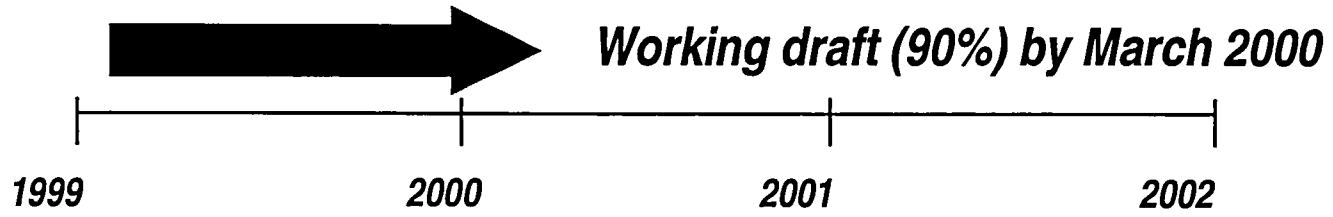
Useful Genomic Sequence Available for Mapping:

10% NOW ⟶ At least 50% in 1 year

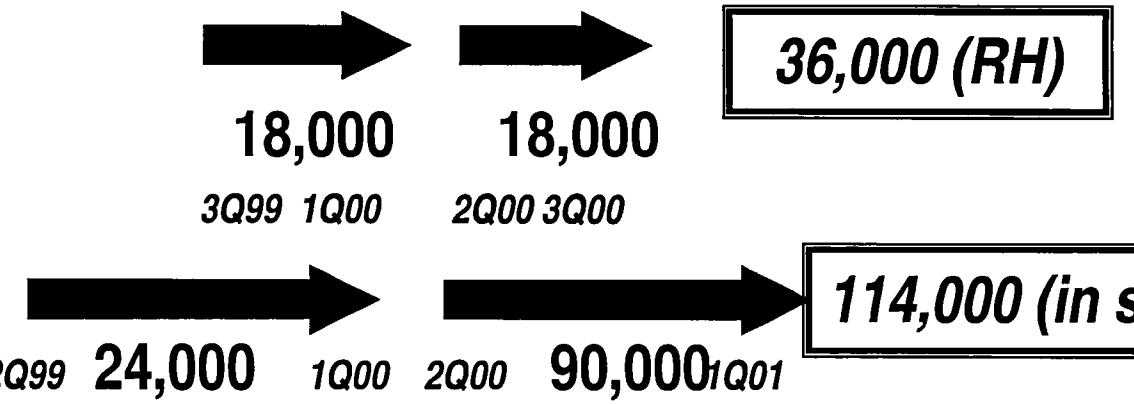
Scientific Plan: SNP Mapping

Objectives and Timetable:

Human Genome Project:



SNP Mapping:



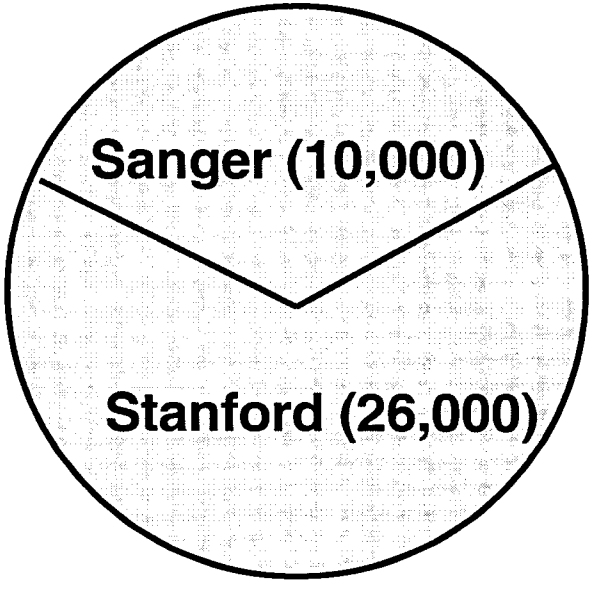
Quarterly targets will be determined by amount of genomic sequence available each quarter

The SNP Consortium Ltd.

Mapping Metrics: Measuring the Productivity & Quality – *RH Mapping*

- **Metrics:**

Quarter	SNPs Mapped	
Year 1:		
Q3 99	*1,000*	
Q4 99	*8,000*	
Q1 00	*9,000*	18,000
Year 2:		
Q2 00	*9,000*	
Q3 00	*9,000*	18,000
Total		**36,000**



- **Quality Checking:**
  1) Remapping loci
  2) Comparing with loci independently mapped
  3) Checking internal consistency

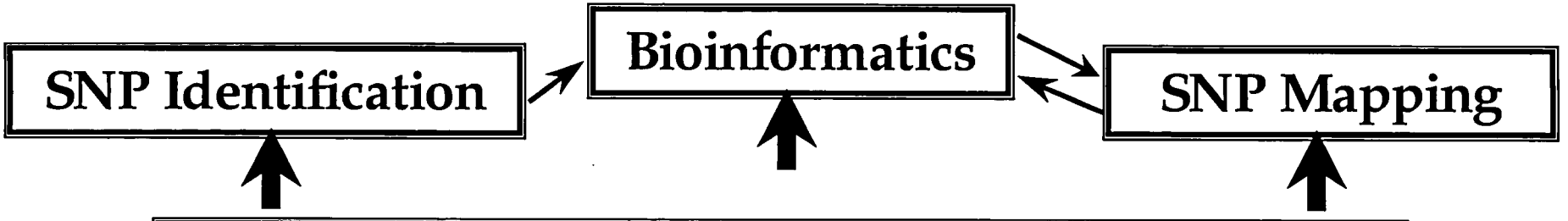
Metrics: Measuring the Productivity

– *In Silico Mapping*

- **Metrics:**

Quarter	SNPs Mapped	
Year 1:		
Q2 99	*adjustable*	
Q3 99	*adjustable*	
Q4 99	*adjustable*	
Q1 00	*adjustable*	24,000 (25%)
Year 2:		
Q2 99	*adjustable*	
Q3 99	*adjustable*	
Q4 99	*adjustable*	
Q1 00	*adjustable*	90,000 (45%)
Total	**114,000**	

TSC Research Organization



TSC Scientific Management Committee

Research Partners

- *David Bentley*
- *David Cox*
- *Eric Lander*
- *Lincoln Stein*
- *Bob Waterston*

TSC OSC

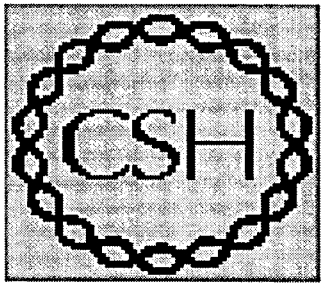
- *Eric Lai, Co-Chair*
- *David Wang, Co-Chair*
- *Arthur Holden*
- *John Keller*
- *Michael Morgan*
- *Michael Silber*

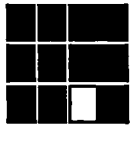
Functions: Scientific Strategy, Productivity, QA/QC and Communications

The Pilot Study to Detect SNPs

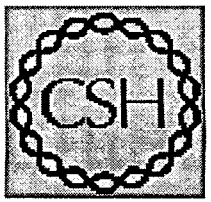
- **Number of Candidate SNPs Detected: ~8,000**
- **Initial False Positive Rate: ~15% >> 5%**
  [Algorithm for Detecting Repeats & Quality Scores]

Data Coordinating Center



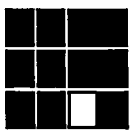


The
SNP Consortium Ltd.

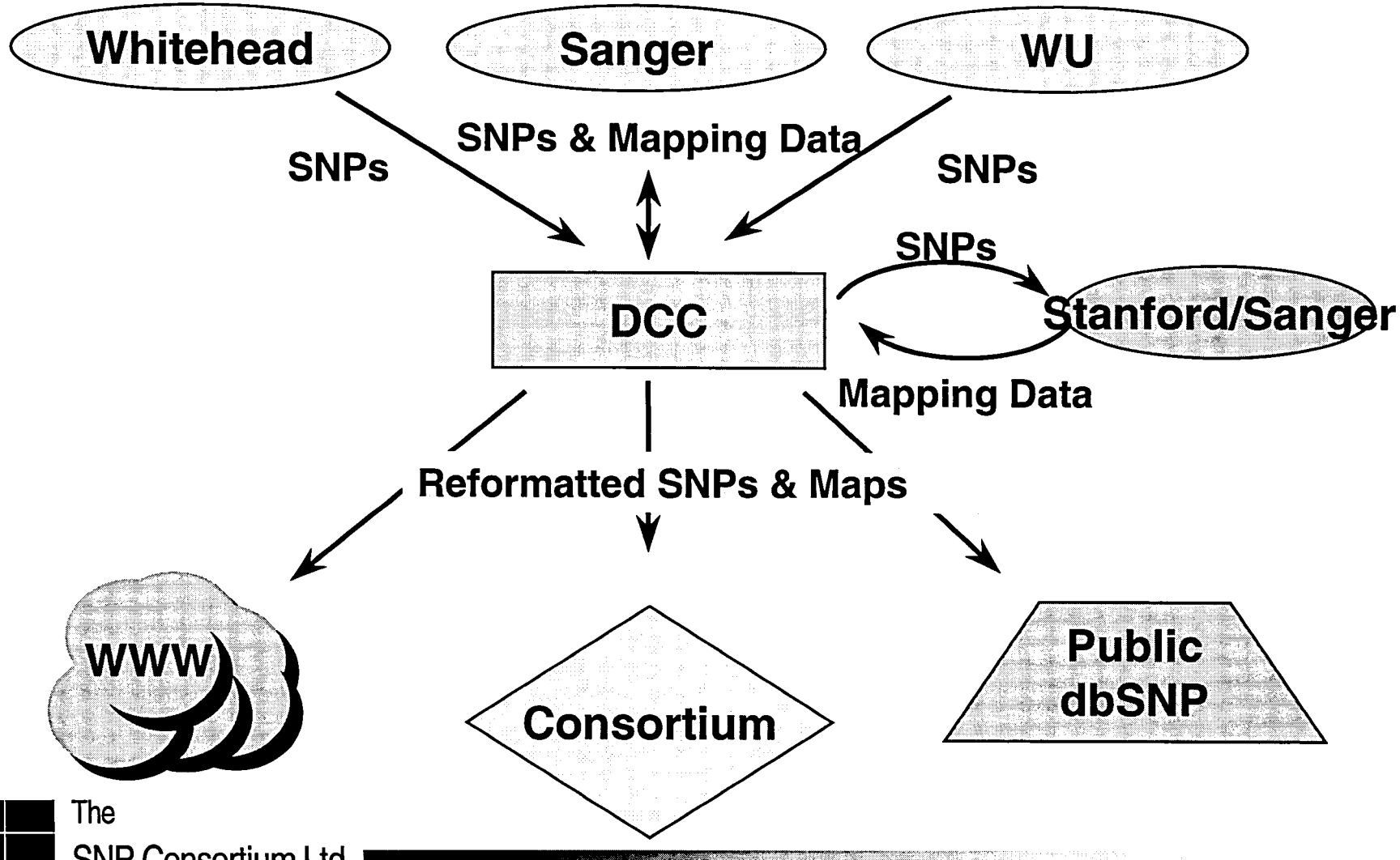


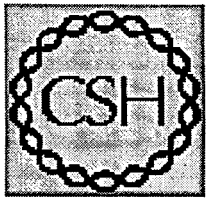
Data Coordinating Center: Key Objectives

- ***Effective SNP Database Maintenance :***
  - **Collect SNP and mapping data from the genome centers into a central, highly usable database**
  - **Facilitate management the overall project's progress**
  - **Coherent data release**
- ***Integrated map of SNPs, genes & genomic sequences***
- ***Systematic & uniform data release [TSC and public]***
- ***Ensure compatibility of data format with public databases* - NIH, EBI, and others**
- ***Provide graphical and programmable interfaces to the SNP database***



Data Coordinating Summary



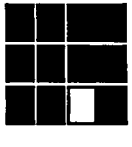


Data Release[s]

- Regular Releases [Quarterly>>Monthly]
- **All** data to be released into public domain
- No preferential access!
- **Outlets**:
  - dbSNP submissions to NCBI
  - Flat files via FTP
  - Browsable WWW interface
  - Relational format to SNC and others
  - Object-oriented ACEDB interface

Other Factors

- Management Process
- Intellectual Property Program
- Public Relations & Communication
- Funding-Sources & Uses
- Membership--Future Directions

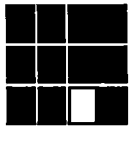


Management Process

Group	Meeting Frequency	Decision Making:
◆ Board of Directors	◆ Quarterly	◆ 2/3 quorum >>majority vote of members.
◆ TSC Operating Committee	◆ Monthly	◆ Consensus
◆ Scientific Mgnt. Committee	◆ Monthly Teleconferences ◆ Quarterly Project Review	◆ Consensus
BOD Committees ◆ Audit ◆ Compensation ◆ Legal/IP	◆ As required, when called by the team leader	◆ Consensus

IP Program

- "Protective" strategy [until public release]
- File provisionals [in bundles] >> priority dating
  - Quarterly>>Monthly
  - All generated SNPs
- Convert all provisionals to utility applications.
- Convert to SIRs as SNPs are mapped and ready for release.
- Research agreements mandate that all IP is managed according to TSC program.
- All parties will have equal data access.

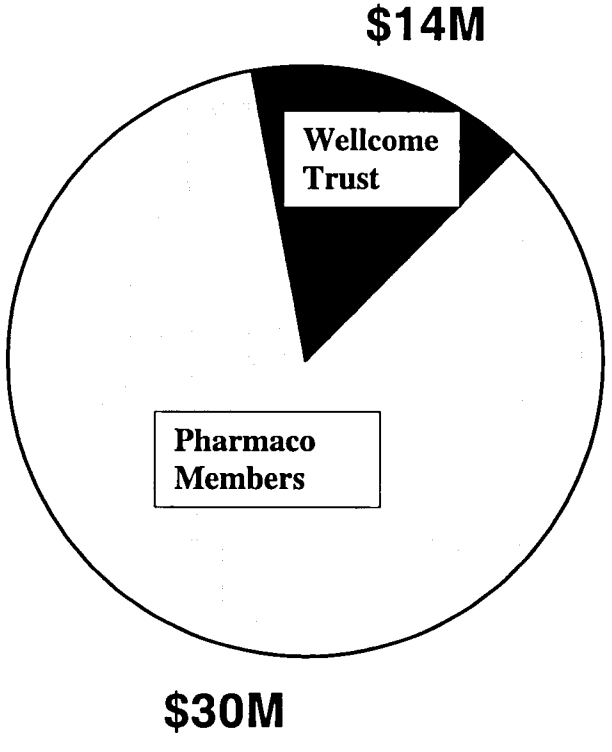


Public Relations

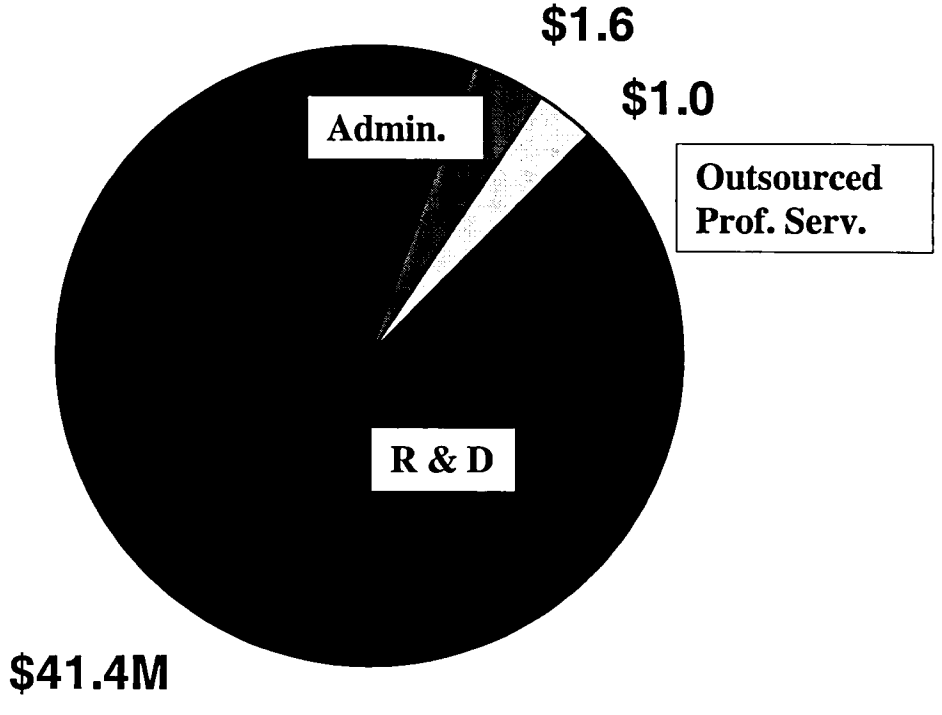
- Goal: Effective & timely communications
- 4/15 PR launch:
  - Well received & covered, especially in US.
  - Covered by most major media [print] outlets
- Continued PR releases and communications:
  - Data releases
  - Significant events & milestones
  - New Members

Sources/Uses of Cash-Phase I

Sources ($44M)

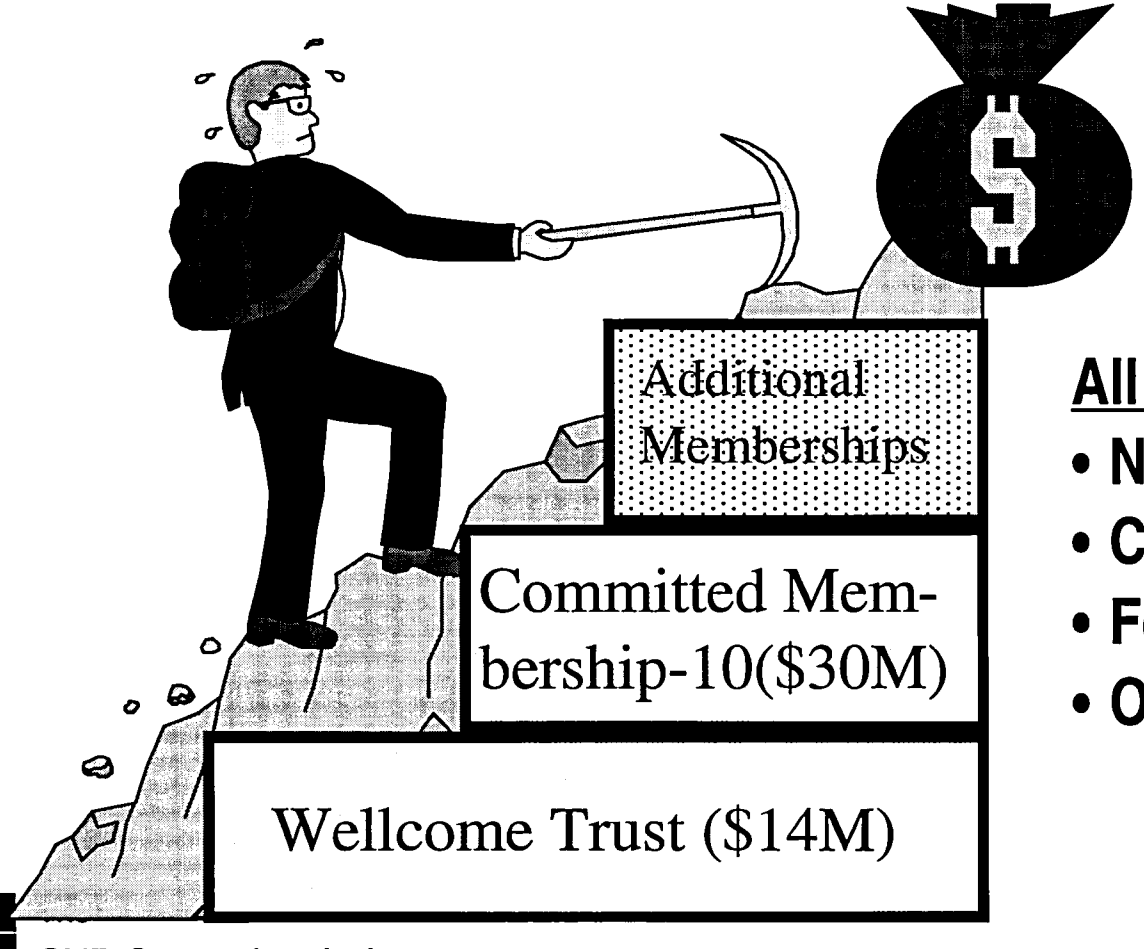


Uses ($44M)



Membership Development

$65M



Continuing efforts

All Members Welcome:

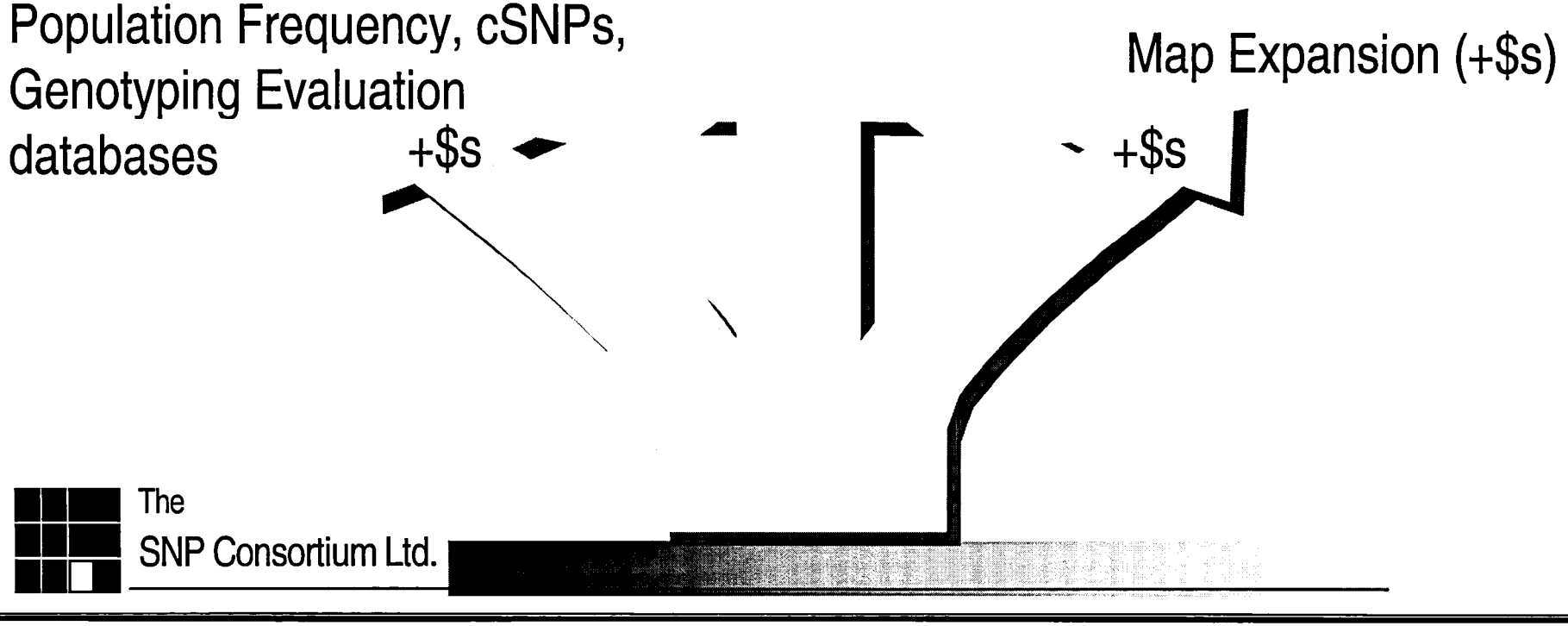
- **NIH**
- **Companies**
- **Foundations**
- **Other Government**

Future Directions

Phase I Ramp Up:

- Full Scale SNP ID production & commence SNP Mapping

Core SNP Map Development



The
SNP Consortium Ltd.

Closing

- Thank you
- Excited about an NIH collaboration
- Next Steps--TSC Proposal & Discussions