

# **The GSID HIV Data Browser:** A Unique Research Tool Providing Access to AIDSVAX Vaccine Trial Data and Specimens

**Faruk Sinangil, PhD**

2nd Annual CAVD Meeting

Lausanne, Switzerland

December 2-5, 2007



# Global Solutions for Infectious Diseases (GSID)

- Co-founded by former Genentech and VaxGen executives Don Francis, Phil Berman and Carter Lee
- A 501(c)(3) tax exempt public charity
- Mission: apply the newest technologies to develop inexpensive public health tools – vaccines and diagnostics – for less developed countries
- Website: [www.gsid.org](http://www.gsid.org)



# Project Goals

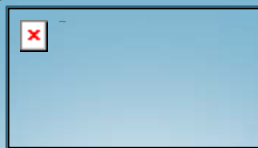
- Acquire and disseminate information that will contribute to the development of a safe and effective HIV vaccine
- Establish a consortium to characterize and evaluate antigenic variation of viruses that mediate new infections

# GSID Consortium



## GENOMA

- Sequence Analysis
- Phylogenetics
- Alignments
  - PI: K. Crandall



- Consortium management
- HIV Data Browser
- Specimen repository
  - PI: F. Sinangil

UC SANTA CRUZ



- Bioinformatics
  - PI: J. Kent
- Protein expression and epitope mapping
  - PI: P. Berman



- Biostatistics and clinical data
  - PI: E. Li



- Pseudotype virus construction
- Evaluation of neutralization sensitivity
  - PI: B. Schweighardt and T. Wrin

# AIDSVAX Phase III Clinical Trials

<i>Design: Randomized, double-blind, placebo- controlled</i>	<b>North America/ Europe VAX004 AIDSVAX B/B</b>	<b>Thailand VAX003 AIDSVAX B/E</b>
<b>rgp120 dose</b>	300 µg (MN)/300 µg (GNE8)	300 µg (MN)/300 µg (A244)
<b>Immunization schedule</b>	0, 1, 6, 12, 18, 24, 30 months	
<b>Transmission</b>	Sexual	Blood borne
<b># of volunteers</b>	5,400	2,500
<b>Annual Infection rate</b>	1.5%	4.0%
<b># of clinical sites</b>	59	17
<b>Start date</b>	June 1998	March 1999
<b>Fully enrolled</b>	October 1999	August 2000
<b>Analysis completed</b>	Q1 2003	Q4 2003



# VAX003 and VAX004 Results

- No efficacy for prevention of acquisition or for modification of infection detected in the overall trial
- Modification of infection not detected in subgroups, but number of infections may be too small
- Possible efficacy from subgroup analyses despite lack of overall efficacy
  - Blacks or non-whites or women:
    - Differential antibody response
    - Genetic factors
    - Non-genetic factors
  - High-risk volunteers:
    - Priming, analogous to highly uninfected cohorts



# Objectives

- I. Transfer custodianship of VAX004 and VAX003 data and clinical materials repository from VaxGen to GSID.
  1. Establish a clinical specimen repository.
  2. Establish a web accessible clinical and sequence database.



# Specimen Inventory

	VAX004		VAX003		Total
<u>Specimen inventory</u>	<u>Collection Time Points</u>	<u>Sample #</u> (tubes)	<u>Collection Time Points</u>	<u>Sample #</u> (tubes)	<u>Sample #</u> (tubes)
<u>Pre-Infection</u>					
Serum	73,137	135,042	34,484	68,966	<b>204,008</b>
<u>Post-infection</u>					
Plasma	2,527	29,282	1,792	15,320	<b>44,602</b>
Serum	2,852	2,852	2,171	4,255	<b>7,107</b>
Cells	46	2,000	1,783	8,974	<b>10,974</b>
<b>Plasmid DNA library</b>					
Full length gp120 plasmids (3 clones per individual sample)		1,050 (350 subjects)		600 (200 subjects)	<b>1,650</b>



# GSID HIV Data Browser

<http://www.gsidhivdatabrowser.org>

- A Unique Research Tool Providing Access to AIDS VAX Trial Data and Specimens

**About the GSID HIV Data Browser Site**

Welcome to the GSID HIV Data Browser, which is hosted by Global Solutions for Infectious Diseases (GSID). The GSID HIV Data Browser is a customized version of the UCSC Genome Browser, which was developed and is maintained by the Genome Bioinformatics Group at the University of California Santa Cruz (UCSCBG), a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE).

Under the guidance of Jim Kent and Fan Hsu, UCSCBG developed a relational database containing the significant clinical data and viral sequence information pertaining to the infected subjects participating in the VAX004 Phase III clinical trial conducted by VaxGen between 1998 and 2003. GSID, through a license and material transfer agreement with VaxGen and with funding provided by the [Bill & Melinda Gates Foundation](#), is making this valuable resource of information and access to serological samples available to the HIV research community.

Three primary views are currently available on the GSID HIV Data Browser. Subject View provides the user with demographic and clinical information pertaining to volunteers who became HIV infected during the VAX004 Phase III clinical trial. Table View provides convenient access to the underlying database by enabling users to view multiple subjects sorted and displayed by the filter controls contained in Select Subjects. Sequence View contains tools and the ability to align sequences with each other, with reference sequences or with consensus sequences.

**GSID Sequence View on HIV (HXB2) Oct. 2002 Assembly**

position search: chr1:6,000-8,000    jump    clear    size 2,001 bp    configure

Conservation: 1.0

move start    Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.    move end

Mapping and Sequencing Tracks

Base Position: VAX004    Short Match

Genes and Gene Prediction Tracks

Genes Region: InterPro

Comparative Genomics

Conservation: full    MSA: hide    Protein Conservation: hide    Protein MSA: hide

**HIV Vaccine Subject View**

search for another subject: [input]    Go

**Demographic Information**

subject ID: GSID1012  
gender: Male    age: 36    risk factor: Low  
race: White/Non-Hispanic    weight(kg): 68    location: Southwest

**Vaccine and HIV Status**

Vaccine-Placebo: Placebo    Days of infection relative to first injection date: 541  
HIV Status: Infected  
Injections: 6

**Clinical Information**

Estimated Study Day of Infection	HIV-1 RNA copies/mL	CD4 count/cell/mm <sup>3</sup>
196	5596	650
217	7567	594
245	53103	645
309	<400	564

**GSID Table View**

subject	group	HIV-1 RNA	CD4	sex	age	race	geography	risk
GSID4382	Vaccine	30434	707	M	30	White/Non-Hispanic	South	Low
GSID4381	Vaccine	147777	635	M	38	White/Non-Hispanic	South	High
GSID4380	Vaccine	3837	654	M	39	White/Non-Hispanic	South	High
GSID4379	Vaccine	1736	950	M	44	White/Non-Hispanic	Midwest	High
GSID4378	Vaccine	338373	443	M	31	White/Non-Hispanic	Midwest	High
GSID4377	Vaccine	103923	375	M	33	White/Non-Hispanic	Midwest	High
GSID4376	Placebo	2807	518	M	31	White/Non-Hispanic	Northeast	High
GSID4375	Vaccine	N/A	N/A	M	22	White/Non-Hispanic	Northeast	High
GSID4374	Vaccine	6037	595	M	36	White/Non-Hispanic	Northeast	High
GSID4373	Vaccine	N/A	N/A	M	35	White/Non-Hispanic	Northeast	High
GSID4372	Vaccine	1096	856	M	32	White/Non-Hispanic	Northeast	High
GSID4371	Placebo	48778	345	M	41	White/Non-Hispanic	Northeast	High
GSID4370	Vaccine	11254	387	M	38	White/Non-Hispanic	Northeast	High
GSID4369	Vaccine	2082	477	M	34	Asian/Pacific Islander	Northeast	High
GSID4368	Vaccine	3812	646	M	36	White/Non-Hispanic	Northeast	High
GSID4367	Vaccine	N/A	N/A	M	33	White/Non-Hispanic	Northeast	High
GSID4366	Vaccine	6495	455	M	35	White/Non-Hispanic	Northeast	High
GSID4365	Vaccine	N/A	N/A	M	33	White/Non-Hispanic	Northeast	High
GSID4364	Placebo	6241	398	M	35	Hispanic	Northeast	High
GSID4363	Vaccine	20585	267	M	25	White/Non-Hispanic	Northeast	High

**Filter Controls for Displayed Columns:**

subject - GSID identification number  
subject search (including \* and ? wildcards): [input]    Include if any words in search terms match.    Limit to items (no wildcards) in list: [input]    [submit]    [clear filter]

group - Immunization Status  
group search (including \* and ? wildcards): [input]    Include if any words in search terms match.    Limit to items (no wildcards) in list: [input]    [submit]    [clear filter]

Available Values:  
Placebo  
Vaccine



# Objectives cont.

- II. Continued molecular sequence and phylogenetic analyses of viruses and virus envelope proteins obtained close to the time of transmission.
1. Cloning and sequencing of transmission viruses' envelope genes.
  2. Sequence analysis of transmission viruses.
  3. Structural analysis of transmission viruses
  4. Phylogenetic and evolutionary analyses of transmission viruses.
  5. Transmission virus database construction and integration with clinical database.



# GSID Consortium: Genoma LLC

## ○ **VAX004**

- HIV genetic diversity remains high even with lower infection rates
- No significant differences in genetic diversity, recombination rates, or levels of selection between vaccinated and placebo patients, gender, geographic region
  - Significant difference in levels of selection between blacks and whites ( $P = 0.023$ )

## ○ **VAX003**

- 185 subtype A/E individuals and 34 subtype B' individuals
- Substitution rates higher in A/E than in B'
- No significant differences in genetic diversity, substitution rates or selection between placebo and vaccinated individuals
- Significantly lower recombination rates in vaccinated individuals compared to placebo



# Objectives cont.

- III. Isolation and characterization of broadly neutralizing antibodies to HIV.
1. Phenosense™ virus neutralization assay.
  2. Construction of pseudotype viruses and evaluation of neutralizing sensitivity.
  3. Identification and characterization of broadly neutralizing antibodies.
  4. Biochemical characterization of VAX004 & VAX003 cross neutralizing sera.



# GSID Consortium: Monogram Biosciences

- Developed a new strategy to define important antigenic regions within HIV-1 gp160
- Utilized the newly available resource of early transmission viruses from the VAX004 trial
- Developed and sequenced a library of functional gp160 clones that differ in neutralization sensitivity to heterologous plasma
- Identifying amino acid changes that are associated with shifts in neutralization sensitivity



# GSID Consortium: Biomolecular Engineering , UCSC

- **Establish panels of envelope glycoproteins, antisera, and pseudotype viruses from envelope glycoproteins of most representative isolates**
  - Bioinformatics and structural analysis to select panel of representative isolates from VAX003 and VAX004, and characterize most common polymorphisms at neutralizing sites
  - Clone and express gp160 genes from panels of representative isolates
  - Establish panel of pseudotype viruses of representative isolates for neutralization studies (collaboration with Monogram Biosciences)
  - Production of antiserum to purified envelope proteins from representative isolates (gp120 and gp160)



# GSID Consortium: Biomolecular Engineering , UCSC

- **Use intra-patient variation in envelope proteins from new infections to characterize epitopes recognized by broadly neutralizing antibodies in HIV+ serum**
  - Compare clones from same individual to identify differences in neutralization sensitivity
  - Characterize minimal sequence differences leading to susceptibility or resistance to virus neutralization
  - In vitro mutagenesis to map sensitivity and resistance to neutralization to single amino acid changes
  - Characterize molecular basis for evolution of neutralization resistance in longitudinal specimens

# GSID HIV Data Browser

<http://www.gsidhivdatabrowser.org>

- A Unique Research Tool Providing Access to AIDS VAX Trial Data and Specimens

**About the GSID HIV Data Browser Site**

Welcome to the GSID HIV Data Browser, which is hosted by Global Solutions for Infectious Diseases (GSID). The GSID HIV Data Browser is a customized version of the UCSC Genome Browser, which was developed and is maintained by the Genome Bioinformatics Group at the University of California Santa Cruz (UCSCBG), a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE).

Under the guidance of Jim Kent and Fan Hsu, UCSCBG developed a relational database containing the significant clinical data and viral sequence information pertaining to the infected subjects participating in the VAX004 Phase III clinical trial conducted by VaxGen between 1998 and 2003. GSID, through a license and material transfer agreement with VaxGen and with funding provided by the [Bill & Melinda Gates Foundation](#), is making this valuable resource of information and access to serological samples available to the HIV research community.

Three primary views are currently available on the GSID HIV Data Browser. Subject View provides the user with demographic and clinical information pertaining to volunteers who became HIV infected during the VAX004 Phase III clinical trial. Table View provides convenient access to the underlying database by enabling users to view multiple subjects sorted and displayed by the filter controls contained in Select Subjects. Sequence View contains tools and the ability to align sequences with each other, with reference sequences or with consensus sequences.

**GSID Sequence View on HIV (HXB2) Oct. 2002 Assembly**

position search: chr1:6,000-8,000    jump    clear    size 2,001 bp    configure

Conservation: 2.0

Mapping and Sequencing Tracks

- Base Position: VAX004
- Genes and Gene Prediction Tracks: InterPro
- Comparative Genomics: Conservation, MSA, Protein Conservation, Protein MSA

**HIV Vaccine Subject View**

search for another subject: [input]    Go

**Demographic Information**

subject ID: GSID1012  
gender: Male    age: 36    risk factor: Low  
race: White/Non-Hispanic    weight(kg): 68    location: Southwest

**Vaccine and HIV Status**

Vaccine/Placebo: Placebo    Days of infection relative to first injection date: 541  
HIV Status: Infected  
Injections: 6

**Clinical Information**

Estimated Study Day of Infection	HIV-1 RNA copies/mL	CD4 cells/micro liter
196	5596	650
217	7567	594
245	53103	645
309	<400	564

**GSID Table View**

subject	group	HIV-1 RNA	CD4	sex	age	race	geography	risk
GSID4382	Vaccine	30434	707	M	30	White/Non-Hispanic	South	Low
GSID4381	Vaccine	147777	635	M	38	White/Non-Hispanic	South	High
GSID4380	Vaccine	3837	654	M	39	White/Non-Hispanic	South	High
GSID4379	Vaccine	1736	950	M	44	White/Non-Hispanic	Midwest	High
GSID4378	Vaccine	338373	443	M	31	White/Non-Hispanic	Midwest	High
GSID4377	Vaccine	103923	375	M	33	White/Non-Hispanic	Midwest	High
GSID4376	Placebo	2807	518	M	31	White/Non-Hispanic	Northeast	High
GSID4375	Vaccine	N/A	N/A	M	22	White/Non-Hispanic	Northeast	High
GSID4374	Vaccine	6037	595	M	36	White/Non-Hispanic	Northeast	High
GSID4373	Vaccine	N/A	N/A	M	36	White/Non-Hispanic	Northeast	High
GSID4372	Vaccine	1096	856	M	32	White/Non-Hispanic	Northeast	High
GSID4371	Placebo	48778	345	M	41	White/Non-Hispanic	Northeast	High
GSID4370	Vaccine	11254	387	M	38	White/Non-Hispanic	Northeast	High
GSID4369	Vaccine	2082	477	M	34	Asian/Pacific Islander	Northeast	High
GSID4368	Vaccine	3812	646	M	36	White/Non-Hispanic	Northeast	High
GSID4367	Vaccine	N/A	N/A	M	33	White/Non-Hispanic	Northeast	High
GSID4366	Vaccine	6495	455	M	35	White/Non-Hispanic	Northeast	High
GSID4365	Vaccine	N/A	N/A	M	33	White/Non-Hispanic	Northeast	High
GSID4364	Placebo	6241	398	M	35	Hispanic	Northeast	High
GSID4363	Vaccine	20585	267	M	25	White/Non-Hispanic	Northeast	High

**Filter Controls for Displayed Columns:**

subject - GSID identification number  
subject search (including \* and ? wildcards): [input]  
Include if any words in search terms match  
Limit to items (no wildcards) in list: [input]    grab list    upload list

group - Immunization Status  
group search (including \* and ? wildcards): [input]  
Include if any words in search terms match  
Limit to items (no wildcards) in list: [input]    grab list    upload list

Available Values:  
Placebo  
Vaccine





# GSID HIV Data Browser

- Relational database encompassing significant AIDS/VAX clinical trial data
  - VAX004 data available 12/10/07 for beta testing
  - VAX003 data to be added in 2008
- Three primary views for accessing data
  - Subject View
  - Table View
  - Sequence View

# Subject View

## Subject View

search for another subject:

### Demographic Information

**subject ID:** gsid4050

**gender:** Male

**age:** 43

**risk factor:** High

**race:** White/Non-Hispanic

**weight(kg):** 86

**location:** Northeast

### Vaccine and HIV Status

**Vaccine/Placebo:** Vaccine

**HIV Status:** Infected

**Injections:** 7

### Clinical Information

Days After Infection	HIV-1 RNA copies/mL	CD4 cells/microliter
114	827376	416
141	70180	501
182	43054	499
616	42759	359

# Table View

Category	Description
Subject	GSID identification number
Group	Vaccine or placebo
Viral load	HIV-1 RNA in copies/mL
CD4	CD4 count in cells/ $\mu$ L
Sex	Gender at birth
Age	Age at enrollment
Weight	Weight in kilograms
Race	Self-reported race
Geography	Geographic region
Risk	Risk group

- Show data from all subjects in database
- Apply filters to select subjects
- View selected subjects in **Table View** and **Sequence View**

# Table View

Home Blat Subject View Sequence View Help

configure filter (now on) display 50 output sequence text

subject	group	HIV-1 RNA	CD4	sex	age	race	geography	risk
GSID4101	Vaccine	47892	219	F	29	Hispanic	South	High
GSID4141	Placebo	10765	548	F	36	Black/Non-Hispanic	Midwest	High
GSID4142	Vaccine	23649	621	F	29	Black/Non-Hispanic	Midwest	High
GSID4143	Placebo	73680	237	F	26	Black/Non-Hispanic	Midwest	High
GSID4144	Placebo	16977	173	F	29	Black/Non-Hispanic	Midwest	High
GSID4145	Placebo	7288	771	F	29	Black/Non-Hispanic	Midwest	High
GSID4244	Vaccine	4178	356	F	43	Other	South	High

Displayed 7 out of 7 subjects passing filter.



# Sequence View

- Rapid visual correlation of nucleotide and protein sequences from VAX004 viruses with multi-sequence alignment (MSA) tools
- DNA and protein conservation from MSA results

# Sequence View

- VAX004 viral sequences from selected subjects, aligned with base genome (here, MN gp120)

Home Sequence View Gateway Select Subjects Blat Table View DNA PDF/PS Help

## GSID Sequence View on HIV MN (GP120) Oct. 2007 Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr1:799-855 jump clear size 57 bp. configure

chr1: | 805 | 810 | 815 | 820 | 825 | 830 | 835 | 840 | 845 | 850 |  
TGTACACATGGGATTAGGCCAGTAGTATCAACTCAACTGCTGTTAAATGGCAGTCTA  
VAX004 HIV-1 gp120 Sequences

VAX004  
gp120  
ENV  
1 -

Conservation  
0 -  
1 -

Protein Conservat  
0 -

Protein Multiple Sequence Alignment

HIV_MN_(GP120)	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-238c1	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-238c2	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-238c3	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-103c1	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-103c2	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-103c3	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-096c5	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-096c6	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-096c7	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-333-2c1	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-333-2c3	C	T	R	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L
U-333-2c8	C	T	H	G	I	R	F	V	V	S	T	Q	L	L	L	L	N	G	S	L

move start Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions. move end

< 2.0 >

default tracks hide all add custom tracks configure refresh

Use drop down controls below and press refresh to alter tracks displayed.  
Tracks with lots of items will automatically be displayed in more compact modes.



# Summary

- Established AIDS-VAX clinical specimen repository at GSID
- Developed web-accessible AIDS-VAX clinical and sequence database that is now ready for beta testing
- Completed preliminary sequence and phylogenetic analysis of early transmission viruses
- Generated preliminary data to define important antigenic regions within HIV-1 gp160
- Initiated efforts to identify and characterize broadly neutralizing antibodies



# Acknowledgements

- **Bill and Melinda Gates Foundation**

- **GSID**

- Carter Lee
- Michael Peterson
- Evie Zaharias

- **Monogram Biosciences**

- Becky Schweighardt
- Terri Wrin
- Yolanda Lie
- Christos Petropoulos

- **Genoma LLC**

- Keith Crandall
- Marcos Pérez-Losada
- David Posada

- **UC Santa Cruz - Bioinformatics**

- Jim Kent
- Fan Hsu
- Ann Zweig
- Robert Kuhn
- Galt Barber
- Erich Weiler

- **UC Santa Cruz – Biomolecular Engineering**

- Phil Berman
- Sarah O'Rourke
- Dora Fonseca
- Bin Yu

- **PharmaStat**

- Elizabeth Li