

PLEASE FILL OUT AND RETURN THIS FORM TO: Center for Public Genomics, Duke University; c/o Susan Brooks; Center for Genome Ethics, Law, and Policy; 304 Research Drive, Box 90141; Durham, NC, 27708. OR: You may fax it to us at (U.S.) 1-919-668-0799.

Interviewee Information. Please list an address where we can contact you.

Full name: Aristides A.N. Patrinos Date of interview: 9-1-2011
Current institutional affiliation: Synthetic Genomics Inc.
Street Address: 11149 North Torrey Pines Road
Phone: 858-754-2902 Email address: apatrinos@syntheticgenomics.com

Interviewer Information.

Full name(s): _____
Affiliations(s): _____

I, the undersigned, have read the above, and I **AGREE** to release my interview materials, subject to any restrictions listed below:

(A) I place **no restrictions** on my interview materials.

OR

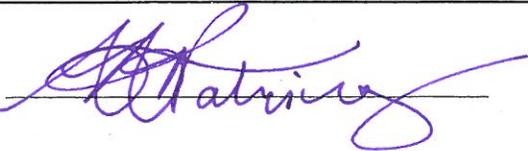
(B) My interview materials may be reviewed, used, and quoted by the researchers affiliated with the Center for Public Genomics, Duke University; *and in addition* (check all that apply):

- Researchers unaffiliated with the Center for Public Genomics may **read** the interview transcript and any related documents only after obtaining my permission.
- Researchers unaffiliated with the Center for Public Genomics may **quote** from the interview only after obtaining my permission.
- Researchers unaffiliated with the Center for Public Genomics **DO NOT HAVE** my permission to **read or quote** from the interview.

Posting interview materials to public digital archives: In spite of any restrictions listed above, I give permission for my interview materials to be made publicly available on the Internet by deposit in an institutionally affiliated archive:

- 1 year from the date of this form
- 5 years from the date of this form
- 10 years from the date of this form
- 25 years from the date of this form
- After my death
- Other: 3 YEARS (please specify a date or condition)
- NEVER: MAY NOT BE DEPOSITED IN A PUBLIC ARCHIVE

Please specify any further restrictions in the space below:

Signature: 

Date: 10-19-2011

Informed consent for: “The ethos and effects of data-sharing rules: Examining the history of the ‘Bermuda principles’ and their effects on 21st century science”

**University of Adelaide
Duke University**

Researchers at the University of Adelaide, Australia, and the IGSP Center for Genome Ethics, Law & Policy, Duke University, are engaged in research on the **Bermuda Principles** for sharing DNA sequence data from high-volume sequencing centers. You have been selected for an interview because we believe that the recollections you may have of your experiences with the International Strategy Meetings for Human Genome Sequencing (1996-1998) will be interesting and helpful for our project.

We expect that interviews will last from 30 minutes to much longer, but you may stop your interview at any time. Your participation is strictly voluntary, and you do not have to answer every question asked.

Your interview is being recorded and we may take written notes during the interview. After your interview, we may prepare a typed transcript of the interview. If we prepare a transcript, you will have an opportunity to review it and to make deletions and corrections.

Unless you indicate otherwise, the *information* that you provide in this interview will be “on the record”—that is, it can be attributed to you in the various articles and chapters that we plan to write, and thus could become public through these channels. If, however, at some point in the interview you want to provide us with information that might be useful for us to know, but which you do not want to have attributed to you, you should tell us that you wish to go “off the record” and we will stop the recording. We will, however, take notes for our own use. When you are ready to go back “on the record,” we will resume recording. Anything you say while “off the record” will not be on the audio recording and therefore will not appear in the transcript.

All *materials* from your interview (audio recording; transcript; interviewer's notes) will be available only to members of the research team affiliated with this project, unless you consent to their wider use, as described in the paragraph below. The digital materials will be maintained in a secure, HIPPA-compliant drive at Duke University. The paper materials will be stored in a locked cabinet.

In addition to the scholarly articles and chapters that we plan to write, we also hope to create a resource for other scholars and members of the public. We plan to post some of our research data to online digital archives. While we will use your “on the record” comments to inform and write our articles, we will not post your interview transcript or audio recording online unless you give us permission to do so, in a separate agreement. At the time we send your transcript to you for review, we will also provide a consent form asking your permission to post your interview transcript and/or audio recording online. The form will provide you with different options for how, when, and with whom the materials may be shared. You will, of course, also have the option not to share the materials beyond the Duke and Adelaide researchers.

One risk of this study is that you may voluntarily disclose identifiable information that later could be requested for legal proceedings, or otherwise be used against you. Please take this into consideration when you are speaking. There may be other risks associated with your “on the record” views being made publicly available, such as having your views mischaracterized or misunderstood.

The main benefit of participating in this study is ensuring that your side of the story is properly portrayed in this history of the Bermuda Principles, which have become a model for open and collaborative research in genomics and other fields.

To help us protect the privacy of those parts of your interview that are not public, we have obtained a Certificate of Confidentiality from the U.S. National Institutes of Health. With this Certificate, we investigators cannot be forced to disclose information that may identify you, even by a court subpoena, in any U.S. federal, state, or local civil, criminal, administrative, legislative, or other proceedings. We researchers can use the Certificate to resist any demands for information that would identify you.

The Certificate cannot be used, however, to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of federally funded projects or for information that must be disclosed in order to meet the requirements of the federal Food and Drug Administration (FDA).

A Certificate of Confidentiality does not prevent you or a member of your family from voluntarily releasing information about yourself or your involvement in this research. If an insurer, employer, or other person or institution obtains your written consent to receive research information, the researchers may not use the Certificate to withhold that information.

Signature 
Printed Name Aristides Patrinos
Date Aug 31, 2011

If you have read this form in its entirety and agree to the interview and its terms, please sign and date above.

Contact information:

Rachel Ankeny, Ph.D. (University of Adelaide)

rachel.ankeney@adelaide.edu.au

+61-8-8303-5570

Kathryn Maxson, B.S. (Duke University)

kat.maxson@duke.edu

(919) 668-0791

Robert Cook-Deegan, MD (Duke University)

bob.cd@duke.edu

(919) 668-0790

*If you have any questions about your rights as a research subject, you may contact the **Duke University Institutional Review Board** at 919-684-3030 or ors-info@duke.edu.*

(2/27/97) Thursday GTN/Bermuda (via JFK)

— went home/packed / called office to tell USB re Rich Stephens
Took metro to DCA / got message that Elbat will arrive tomorrow.

AA to JFK

same flight w. Francis & Elke; they had
just come from the House Approp hearing

— Chatted w. Francis about strategies

~~with~~ he said he wanted us involved in the workshops
that will be organized by Chokrasanti that will lead to the
next 5-year plan ————— NEED to tell DAN

— Also we should coordinate the next set of meetings we're
going to have to deal/agree on standards -

needs attention. *

— At JFK chatted w. Elke / other participants joined

AA to Bermuda

picked up by limo to hotel (Princess)
cocktail party
chat w. some folks.

2/28/97 Friday Bermuda

~~_____~~

- register for Hudson/Hawkins

- Catherine & Victoria, Conference Suite

- Michael Morgan / intro - slides of seq. facilities

- Progress strategies & development

• Cox Chair

• Sulston 228x (1 6 20 other)

PACs / Peter DeSang's libraries

shotgun

"GAP closure" issue / attention to finishing in GC regions

Finish 14.6 Aug. 11.9 Ready for 17.5 (over all chances)

- 34 Mb full finished seq. last year. (total 52.2) (all kinds)

(expected human for this year 30MB-40MB)

expected to be 80MB the following year (response to Venter's question)

(Waterston)

7, 22, x (Sanger)

as Eric Green

3.50 MB in libraries

15.1 in shotgun

- PAC Peter DeSang's libraries / minimal and walking.

- Map status. 600 STSs. sp. 50 MB

128 BAC/PAC

250 kb av. size

32 Mb total

175 clones underway 21 finished

Producing Sequence shotgun / directed
PHRED / PHRAP / -

QA: complete continuity, annotation, $< 10^{-4}$ error rate.

software for human decision making:

data checking (central data base - bar coding)

getlans / pl2 / phred / phrap.

finish - reorganizing.

Technologies

64-72 lanes in 377, gel readers, Amersham dye terminators.

Transposons

Future - 96 lanes on 373, 377, pipetting station, U.W. sequencer (?)

w. Lloyd Smith

Testing of capillary sequencing. —

(clone availability the reason for apparent bottleneck: ... too little finished
sequence)

finishing 1-2 MB per month

Hudson mapping @ Whitehead.

still a lot of "walking" for back of BAC's

high density pooling scheme for BAC's

Hawkins: 5MB by 5/1 goal.

2.1 MB by 1/1

most on 17

expected to go up to 20 MB by next year.

Sequencing-2 in effect.

DA/GC

Resequents

gel-to-gel

projects

auto tied detection

2/28

(Hawkins [continued])

Current issues: Finishing (lab issues, computer issues, automation)
Interpretation. Human-mouse sequencing
Ken Fossma

Finishing should be a production line. (small team to treat unusual clones)

Finishing by numbers

set of methods & landmarks for progress & automation streamlining

Finishing focus.

Lab/Automation (to box - learning process, "Finishtron" ...)

Computer (automated workflow, Post Sequencing Verification)
Big brother?

q. some of finished clones have gaps.

a: 2.1 MB have 11 gaps.

nothing in GenBank yet.

will be submitted as soon as possible

Adams TIGR/Cattech collaboration using LAMB (Mozzic)

16p.

human sequencing catching up w. microbial
also rapidops's.

2.6 MB submitted to GenBank

closure .735 K (5 BACs).

11 MB next year's goal

random ...

robot from SAIC - 3 arm w. temp. control

(closure by individual)
team member.

storage space built in.

12 genes in 16p. 1 gene per 196 kbp (!)

picked 16 because of map quality.

Gibbs 3 MB in Genbank

fully annotated & highly accurate

1.2 (over the last 4 months)

95-96 objectives: reduce redundancy, etc, reduce reaction rates.

reads 16 12.7 for BAC's

Costs (per reaction)

reagents cost per reaction \$1.56 → \$1.32.

new dyes

BODIPY replacing ABI

Completed > 150 DNA's

Human vs. Mouse

97-98 ambition - 15 MB in next year. April to April.

scale to 100 MB/year in 1998 [x, 12, 3(?)]

Introduce Dallas automation

complete 15 MB

Big: ... scale-up questions.

Cox: Stanford human genome centers

ch. 4.

200 MB by ? (2002).

- 100 KB of finished seq. in Genbank.

1.2 MB also in Genbank > 3 kb

5 MB between April to April the goal.

500 kb average resolution - map.

- less-leamed for sequencing.

high resolution maps - BAC / PAC / A / cosmid

quality of libraries - bottleneck / knowing how the good things are up front?

2M over 2 years. 20MB

Fiona Francis German Human Genome Project

6 MB over the next 3 years

ch. 21. to be on web in a couple of months.

PAC's, BAC's later

IS for data quality (?)

Jean Weissenbach

haven't started working

(Centre National de Sequencage)

frame of future working place

\$14M, staff 110-120, Location Evry (near Genetip)

starting in summer 1997

joint venture: Ministry & CNRS

(signed (to be) lease on new building) near Evry.

inhouse & collaborative projects. (less clear about ratio)

Scientific & Steering Committees

issues of data policy & intellectual property

human sequencing not @ the beginning / avoidopsis & model organisms also pathogens.

John Mattick (Australia)

parallels with French situation

just getting started.

political reasons: \$8M to set up national sequencing facility.

30 ABI's w. robotics.

1500 templates/day

funded for equipment & building but not for projects



\$15M to fund projects (proposed)
 human & pathogens intended
 to do sequencing for groups across the country.

Singapore doing some sequencing
 Thailand/new institute. but no people to manage

Break /	96 (2.5 MB completed)
<u>Andre Rosenthal</u>	97 6 MB
97-2000 (April)	98 12 MB
40 MB	99 19 MB
	00 3 MB

Targets \times ²¹ ~~29~~, 7
 also mouse to start this year.

6 production groups 4 informatics 4 library groups
 1 postdoc 5 folks 4 postdoc.
 3 technicians 3 technicians

13 + 144 ADPI

IMB, MPINC, GBF the three groups involved in seq.

year	↓	↓	↓	
1	4	1	1	
2	9	2	2	
3	15	3	3	(MB) sequenc ⁱ →

- collaboration w. Japanese group.

- # disease gene hunting,
also bacterial sequencing

zebra-fish genome program ?

↳ sequencing in Germany not viewed
as meritorious.

Phil Green

Genome Sequence Quality Criteria.

- Fidelity

closer to genome

2x confirmation of every clone to detect small (<1 kb) deletions, duplications,
transposon insertions

- Accuracy

- error rate < 1/10kb

base-specific probabilities submitted w. sequence

independent test of assembly accuracy

- Contiguity

all gap sizes estimated

all sequence contigs should be oriented & ordered within chromosome

UNGC

- MCO mapping (clone validation, better tiling paths, better gap closure..
\$0.05 - \$0.12 per bp. (?)

• Long reads

• Objective finishing criteria

chromosome 7, Human HLA Class I, Mouse T-cell receptor alpha.

↳ collaboration w. Lee

300 kb submitted

bottleneck requiring editing/annotation (0.6 + 0.6 + 0.37)

500 kb needing additional data collection

2

1

1 year goal was 2MB / goal will be met.

Ellson Chen from Applies Biosystems Division.
(no connection w. ABI Marketing)

Collaborations w. ABI R&D, Wash U, Wof AL, Shanghai

rate · 0.25 MB/month human X 1.6 Mb so far goal: 2.5 MB/year

3 MB/year

microbial

evolutionary

modified shotgun / sounds like end-PCR sequencing?

OS advantages

shoulder
fracture

Yoshi Sakaki Fujisawa:

genome program in Japan started as pilot project in 1995

slides not readable

ch. 21 / (working w. Stanford).

3.2 MB per year (?)

- proposal ...

USJ still somewhat reluctant to make big investment in sequencing.

ch. 21 - eventually 11 (as well as more 11)

initial proposal 7 years / budget a problem

because of economic decline.

Look @ home page IS1 ALIS Project
Human Genome Sequencing

also Sakaki's homepage

Evans

11, 15

BAC/PAC end sequencing.
1.6 Mb in Genbank.

finishing the major bottleneck

PAC-end sequencing

ORCA vs: robot

QualPlot (for quality)

propose to coordinate the seq. of 11 (attractive because of the high quality map).

Palazzo.

Drosophila. / LBNL HGC (DOE)

DJGI LLNL, LBNL, LANL.

Producing finished sequence

Totals. 5MB Drosophila.

4 MB Human

800 kb/month rate of sequencing

Strategy

Physical maps

Random digest shotgun

build paths.

Transposon facilitated

quality - all double stranded, redundancy for assembly 1×10^4

Software

path-building suite

space (assembly, Rel, MAR)

VI Hardware

Colony picker

library pooling & replication

oligosynthetizer

Agaros

VII Partnership w. Industry & JGI DOE

Goals ... Precise Goal Definition

Benchmarking, metrical tools, Roadmaps for operation & tech

Implementation + evolution

New Space

metrical tools

Process model, cost model, cost accounting, pick-a-model

/ 9. re applicability
of Motorola approach

Ques:

total seq. submitted to GenB 6 MB (?)

4 MB level (3)

ROI not center.

biggest cost is sequencing what pays the db.

4% overhead @ UoG @ K

(shoot the clean...)

ch. 22

also ch 9.

~~Francis C.~~
Francis C.

half life issue

OK. 34 years.

5-6 year. in trouble

need good numbers. / good accounting.

Quality Issue

what is it we think is an acceptable standard of quality.

10^{-4} sequencing accuracy
most standards for gaps

criticism by Council because how will we know?

Quality Nucleotide-level accuracy

Assembly

Gaps

Fidelity to human sequence

Round-Robin check by the Centers funded by NIH =
random BAC to be sequenced by two other centers for QA.

Software: PHRED/PHRAP

Craig: why 10^{-4}

Subin: do not get too bureaucratic.
need to temperize.

standards of practice rather than report card.

"convey-in-coalmine"

Subin: against two other seq. of a BAC
only once.

~~with~~
Assembly
- accuracy, quality, & GC content

- (1) PIRAD / PHRAP parameters
- (3) Reassembly by another center
- (4) Sample resequencing
- (5) Resequencing by another center
- (2) check original traces against consensus.

Define the method used to define accuracy

to restaurant (Harborfront) w. Elbert ... chat & strategies of
DJGI ... mtg. w. Palazzolo the next day
already had a couple of friendly chats w. Mike P.
... Telever Hawkins interested in LANL job.

Sat @ head table with

David Weatherall, Michael Morgan, Francis, Cox, Sulston,
Waterston, Weisenbach

Lots of interesting discussion about cloning & God (Francis a
bit uncomfortable)

also jokes about Jim Watson

lots of wine (██████████)

- stroll later w. David Cox (good Cuban cigars - his treat)

3/1/97 Saturday Bermuda

~~_____~~

Palozzolo

Costs

Value/Danger

Methods

Validation

Methods ?

Cost Model Extrapolations

Cost Accounting

Cost Models

Top down cost & finance analysis

Cost Model

1. Define product 2. establish process plan model 3..

Buyers "cooperative" issue

USG folks cannot lead effort

Rosenthal: proposal to go around to p discuss \$/base

No objection to audits (unanimous)

Francis:

$$n = \frac{\ln \left[\frac{k c_0 M}{\gamma} + 1 \right]}{k}$$

$$k = \frac{\ln 2}{T_{1/2}}$$

c_0 = cost @ time zero

M = total sequence that must be done (in MB)

γ = \$M/year available for sequence production

n = number of years to finish

finished sequence 44 MB

" " 27 " ~~in Genbank~~

predicted next year

98-99

384 MB

Data Release Policy Discussion

Germany 3 month - delay of raw sequence data.

France situation may be worse

DOE (Elbert) coherent w. NIH but need to make sure rules are upheld

Japan assembled raw data deposited to funding agency (JST) which
controls the further disposition -
however individual PIs can release the data.

break: Elbert & Mike P

Elbert (re Trevor Hankins)

Lipman: re data bases

finished vs unfinished
"in Genbank".

Hagen

next mtg. 2 next year?



2 sometimes concepts re.
during February

~~2/27-8~~

2/19-2/22

Tense finale about declaration on data policy that put
Germans in uncomfortable position. ~~_____~~
somewhat smoothed in the end.

Did not join the final dinner

Went instead to the Red Carpet w. Elbert & Mike P
& had a good mtg. on the Institute's future
perhaps we made some progress w. Mike P.

Will need mtg w. Covino & Mike P. to smooth
ruffled feathers

Impressed on Mike P. the need to work w. Elbert &
me more closely & not to run to Oddone &
Shank that often.

