

The Jennifer Jones Simon Foundation

Altered Gene Expression in Neurological and Neuropsychiatric Disease

January 30-31, 2000. Los Angeles, California

Prepared by Ruth Luthi-Carter

Participants

Arthur Arnold

310-825-2169
310-825-8081 Fax
Email: arnold@ucla.edu
*Department of Physiological Science
University of California Los Angeles
4113 Life Science Building
Mail Code 152706
Los Angeles, California 90095*

Lyle Arnold

510-739-2171
Email: Larnold@incyte.com
*Incyte Microarray Systems
6519 Dumbarton Circle
Fremont, California 94555*

Robert Baughman

301-496-5745
301-402-1501 Fax
Email: rb175y@nih.gov
*National Institute of Neurological Disorders
and Stroke
6001 Executive Boulevard
Room 2N2137
Bethesda, Maryland 20892*

Paul Coleman

716-275-2581
Email: paul_coleman@urmc.rochester.edu
*School of Medicine and Dentistry
University of Rochester
P.O. Box 603
Rochester, New York 14642*

Luda Diachenko

*Clontech Laboratories, Inc.
1020 E. Meadow Circle
Palo Alto, California 94303*

James Eberwine

215-898-0420
215-573-2236 Fax
Email: eberwine@pharm.med.upenn.edu

*Department of Pharmacology and Psychiatry
University of Pennsylvania Medical Center
36th and Hamilton Walk
Philadelphia, Pennsylvania 19104-6084*

David Eisenberg

310-825-3754
310-206-3914 Fax
Email: david@ewald.mbi.ucla.edu
*Molecular Biology Institute
University of California Los Angeles
Box 951570
Los Angeles, California 90095-1570*

Mark Erlander

619-450-2075
619-450-2090 Fax
*R. W. Johnson
Pharmaceutical Research Institute
3535 General Atomics Court, Suite 100
San Diego, California 92121*

Daniel Geschwind

310-206-6814
310-267-2401 Fax
Email: dhg@ucla.edu
*Department of Neurology
University of California Los Angeles
A-251 RNRC
Mail Code 176919
Los Angeles, California 90095*

Olli-P. Kallioniemi

301-435-2896
301-402-7597 Fax
Email: okalli@nhgri.nih.gov
*National Human Genome Research Institute
Building 40, Room 4A24
49 Convent Drive, MSC 4470
Bethesda, Maryland 20892-4470*

Kenneth L. Lange

310-206-8076
310-825-8685 Fax

Email: klange@ucla.edu
*Department of Biomath
University of California Los Angeles
AV-515 CHS
Mail Code 176620
Los Angeles, California 90095*

David Lockhardt
858-812-1000
858-812-1570 Fax
Email: david.lockhardt@nifg.novartis.com
*Novartis Pharmaceuticals
3115 Merry Field Row
San Diego, California 90095*

Ruth Luthi-Carter
617-726-5227
617-724-1480 Fax
Email: carter@helix.mgh.harvard.edu
*Department of Neurology Research
Massachusetts General Hospital
Warren 401
Boston, Massachusetts 02114*

Stanley Nelson
310-794-7981
310-794-1871 Fax
Email: snelson@juno.com
*Human Genetics Department
University of California Los Angeles
695 Charles E. Young Drive South
Box 708822
Los Angeles, California 90095-7088*

James Olson
206-667-7955
206-667-5446 Fax
Email: jimmyo@u.washington.edu
*Fred Hutchinson Center
1100 Fairview Avenue North
Mailstop C3168
Seattle, Washington 98109*

Leena Peltonen
310-794-5631
310-794-5446 Fax

Email: lpeltonen@mednet.ucla.edu
*Human Genetics Department
University of California Los Angeles
Box 708822
Los Angeles, California 90095-7088*

Ethan R. Signer
971-368-5498
971-368-5499 Fax
Email: signer@mit.edu
*7th Floor, 230 Park Avenue
New York, New York 10169*

310-575-9656
310-575-9156 Fax
*Hereditary Disease Foundation
11400 W. Olympic Blvd., Suite 855
Los Angeles, California 90064*

Lawrence Steinman
650-725-6401
650-725-0627 Fax
Email: steinman@stanford.edu
*Department of Neurology and Neurological
Sciences
Stanford University
Beckman Center, B002
Stanford, California 94305-5429*

Michael Sussman
608-262-8608
608-262-6748 Fax
Email: msussman@facstaff.wisc.edu
*Biotechnology Center
University of Wisconsin
1132B Biotechnology Center
Madison, Wisconsin 53706*

Stephen Tapscott
206-667-3399
206-667-6524 Fax
Email: stapscot@fhcrc.org

*Division of Human Biology
Fred Hutchinson Research Center
1100 Fairview Avenue North, C3-168
P.O. Box 19024
Seattle, Washington 98109-1024*

Leslie Thompson

949-824-6756
949-824-2688 Fax
Email: lmthomps@uci.edu
*Department of Biological Chemistry
University of California Irvine
College of Medicine
Room D-240, Medical Sciences 1
Irvine, California 92697*

Allan J. Tobin

310-825-5061
310-267-0341 Fax
Email: atobin@mednet.ucla.edu
*Brain Research Institute
University of California Irvine
College of Medicine
Room D-240, Medical Sciences 1
Irvine, California 92697*

310-575-9656
310-575-9156 Fax
Email: allantobin@hdfoundation.org
*Hereditary Disease Foundation
11400 W. Olympic Blvd., Suite 855
Los Angeles, California 90064*

Jeffrey M. Trent

301-402-2023
301-402-2040 Fax
Email: jtrent@nhgri.nih.gov
*National Human Genome Research Institute
Building 49, Room 4A22
49 Convent Drive, MSC 4470
Bethesda, Maryland 20892-4470*

Nancy S. Wexler

212-543-5667
212-543-6002 Fax
Email: wexlern@pi.cpmc.columbia.edu

*Departments of Neurology and Psychiatry
College of Physicians and Surgeons
Columbia University
1051 Riverside Drive, Unit 6
P.I. Annex 324
New York, NY 10032*

310-575-9656
310-575-9156 Fax
Email: nancywexler@hdfoundation.org
*Hereditary Disease Foundation
11400 W. Olympic Blvd., Suite 855
Los Angeles, California 90064*

Wing Wong

310-206-9475
310-206-6673 Fax
Email: whwong@stat.ucla.edu
*Department of Statistics
University of California Los Angeles
8118 Math Science Building
Mail Code 155404
Los Angeles, California 90095*

Overview

The purpose of this workshop was to consider the utility of new gene expression profiling methodologies in studying neurologic and psychiatric disorders. Discussions focused primarily on the development, application and cost-effectiveness of DNA microarrays as an expression profiling tool. Most of the disease-specific research discussions focused on Huntington's disease as a test case, but the same principles should be applicable to studying other disorders of the nervous system.

Expression profiling - a tool for understanding the workings of the nervous system

Nucleotide microarray technology is revolutionizing the arena of DNA and RNA analysis. This technology has the potential to expedite greatly the search for etiologic mechanisms of neurologic and neuropsychiatric diseases and identify new therapeutic approaches. Systems currently in use allow the simultaneous detection of as many as several thousand mRNAs. This kind of approach can be applied to many aspects of neural function since the "fingerprint" of steady-state RNA populations offers a lot of information about the state of a cell or tissue. It can be particularly powerful in examining mechanisms of transcriptional regulation (Holstege *et al.*, 1998) and cascades of transcriptional activation (Harkin *et al.*, 1999; Iyer *et al.*, 1999). Other studies have used expression profiling to address questions of functional heterogeneity between closely related cells in a particular brain region, disease-state-specific changes in mRNA expression, and the RNA composition of brain lesions such as Alzheimer's disease plaques (Callahan *et al.*, 1994; Callahan *et al.*, 1999; Cheetham *et al.*, 1997).

Microarray Studies of Huntington's Disease - What has been learned thus far?

Max Perutz and colleagues have proposed that polyglutamine stretches may interact via hydrogen bonds to form beta-sheet structures (Perutz, 1999). Consistent with this idea, expanded polyglutamine domains are hypothesized to adopt a unique secondary structure (Persichetti *et al.*, 1999; Sieradzan *et al.*, 1999; Starikov *et al.*, 1999) and have been shown to form homopolymeric and heteropolymeric aggregates with a higher propensity than those associated with wild-type forms of polyglutamine disease-causing proteins (Davies *et al.*, 1997; Georgalis *et al.*, 1998; Scherzinger *et al.*, 1997; Scherzinger *et al.*, 1999).

Huntingtin has been shown to interact with and/or form aggregates containing transcription factor proteins, including TBP, CBP, and the nuclear co-repressor (Boutell *et al.*, 1999; Huang *et al.*, 1998; Steffan *et al.*, 2000). Other mutated polyglutamine proteins demonstrate similar properties (Kazantsev *et al.*, 1999; Wood *et al.*, 1998). Further, previous studies of HD cases and HD models show abnormalities in steady-state mRNA populations (Augood *et al.*, 1997; Augood *et al.*, 1996; Cha *et al.*, 1999; Cha *et al.*, 1998). Thus, transcriptional dysregulation is an important candidate mechanism of huntingtin toxicity.

Other recent studies offer further evidence of mutant-huntingtin-induced transcriptional changes:

A. Mouse HD model:

Ruth Luthi-Carter and Jim Olson discussed their findings of altered gene expression in the R6/2 mouse model of Huntington's disease (Luthi-Carter *et al.*, 2000). Six- and 12-week-old R6/2 mice show decreased expression of signal transduction-, calcium homeostasis- and

retinoid-related mRNAs. Twelve-week-old R6/2 mice show increased expression of interferon-responsive genes and some other signs of a stress/immune response. A subset of the mRNA changes were confirmed using in situ hybridization and northern blotting.

Many of the affected signaling molecules are involved in G-protein-coupled receptor signaling or other well-characterized signaling cascades for which pharmacologic ligands are available. Mark Erlander and the authors of the study suggested that targeting these pathways with new candidate therapeutic agents would be an important next step in the project.

Many genes shown to be differentially expressed in R6/2 animals are known to be regulated by retinoic acid and/or contain retinoic acid response elements in their promoters. Jim Olson has initiated therapeutic trials of all-trans retinoic acid in R6/2 animals. This retinoid delays clasping in R6/2 animals. It impairs rotorod performance in both R6/2 and wild-type animals, however, suggesting that the drug has toxic effects as well. Future studies will test receptor-specific retinoid agents to find an agent with maximal therapeutic benefit and minimal toxicity.

Larry Steinman searched his group's human data for retinoid-related changes in HD cortex (see below), and also discovered decreases in several retinoic acid receptors. Leslie Thompson added that CBP, which interacts with huntingtin, can co-regulate gene expression together with retinoic acid receptors.

It was also mentioned that Marc Diamond at UCSF has shown that glucocorticoids have therapeutic benefit in the R6/2 mouse model (Diamond *et al.*, 2000). This may relate to inhibiting the interferon and stress-response pathways shown to be activated in 12-week R6/2 mice (by expression profiling [discussed

above] and IL-1beta production (Ona *et al.*, 1999).

One of the few "metabolic" genes showing altered expression in R6/2 mice was creatine kinase B. Creatine is currently in pre-clinical trials in HD mouse models, and creatine phosphate is believed to be its active form. Larry Steinman said that a mitochondrial creatine kinase is also decreased in human HD cortex (see below).

B. Human cases of HD and other neurologic disorders:

Larry Steinman discussed his group's data on three neurologic disorders: Multiple Sclerosis (MS, cortex), Parkinson's disease (PD, putamen) and Huntington's disease (HD, cortex). Data from all three disease groups were compared to each other and to region- and age-matched controls. Cluster analysis showed that particular groups of genes correlated with disease phenotype. In MS, for example, glutamate-related signaling pathway mRNAs, including those encoding glutamate receptor subunits, were increased relative to controls and to the other disease states. This is consistent with excitotoxic damage to oligodendrocytes and neuronal axons presumed to occur in the disorder. In MS and PD, evidence of an immune response was observed by virtue of increases in the expression of genes characteristic of activated microglia. Also, in a mouse model of Multiple Sclerosis (EAE), increases in metalloprotease inhibitor and STAT 3.6 are seen. Increased expression of P450 enzymes are also observed in mouse and human PD. In HD, the activation of pathways generally distinct from those seen in PD and MS are observed (Karpuj and Steinman, 1999).

Larry Steinman's group has set out *a priori* to compare human samples of each neurologic disorder they study with an appropriate animal model of the disorder. In cases of MS and PD models, the animal data show strong

correlations with the human data. A study of an HD mouse model (repressible exon 1 transgenics developed by Rene Hen's lab) is underway. Larry suggested that the comparisons between animal and human data and comparisons among a wide range of neurologic disorders would be extremely valuable in painting a complete picture of brain pathology for any given disease.

Microarray technology – What are the options?

Variations in array platforms -
General comparisons:

- A. cDNA versus oligonucleotide
- B. in-house versus commercially manufactured
- C. focused versus expanded
- D. known genes versus ESTs

Strengths and weaknesses of the various platforms:

A. cDNA versus oligonucleotide arrays. In general, cDNA arrays tend to be slightly more sensitive than oligonucleotide arrays; sometimes this can circumvent amplification of the probe, which can introduce new variables into the assay. (For example, Lyle Arnold reported that the sensitivity of Incyte's array is 1 part per 100,000.) The strength of the oligonucleotide arrays is that the probe oligonucleotides can easily be designed to discriminate between closely-related family members (e.g. different splice forms of a single gene or mRNAs transcribed from closely-related genes); thus, cross-hybridization between such sequences is minimized.

B. In-house versus commercially manufactured. The simplest form of the in-house array - a membrane-based "dot-blot" can be made with little or no specialized equipment; thus, the up-front investment is minimal. Some

commercially available cDNA arrays, such as the membrane-based products available from Clontech, are also readily adaptable for use in laboratories that do not want to make a large investment in new equipment for expression profiling studies. (Paul Coleman said that his laboratory had been pleased with the performance of the Clontech arrays that they had used.)

Making cDNA arrays oneself requires the development and maintenance of a clone library. This can be accomplished by purifying and sequencing members of a cDNA library, generating clones of interest by PCR, requesting clones from other investigators, or purchasing clones from a commercial source (such as an I.M.A.G.E. consortium clone distributor). The maintenance of the clone library (including preparing DNA and keeping track of clones) can be a substantial investment of manpower if large numbers of clones and arrays are to be produced. If expanded cDNA microarrays are to be produced in the laboratory, special equipment is needed (array spotter, scanner, powerful computer, etc.), requiring an additional up-front investment (for additional information, see Pat Brown's laboratory web site at <http://cmgm.stanford.edu/pbrown/mguide/index.html>). The same will be true for in-house manufacture of oligonucleotide arrays, although systems to manufacture these are not yet commercially available (see below and Singh-Gasson *et al.*, 1999). Some of the equipment used for the processing of cDNA microarrays can also be used for other applications, so it may have more flexibility than equipment purchased as components of multi-workstation systems. Another flexible aspect of custom array synthesis is that the investigator can choose to represent particular clones of interest in a fashion consistent with the hypotheses of the experiment. Many members of the group, including Mark Erlander, Stan Nelson, Paul Coleman and Art Arnold, were working with some type of in-house cDNA array.

Michael Sussman presented an approach that his lab had taken to produce a custom oligonucleotide array synthesizer using a maskless, light-directed methodology (Singh-Gasson *et al.*, 1999). This method involves the use of a digital micromirror array (from Texas Instruments) to form virtual masks, allowing the synthesis of 480,000 different oligonucleotide sequences in a 10 by 14 mm area. Each array takes approximately 8 hours to synthesize and costs approximately \$100-200. The same principals used in making this device should be transferrable to any photochemistry application. A custom oligonucleotide synthesizer has the distinct advantage of gene flexibility; the sequences can be re-designed and re-programmed for each run.

Commercial array systems employing expanded arrays, such as those of Incyte or Affymetrix, have their own advantages and disadvantages. Processing and analysis systems, which often include a hybridization unit, fluidics station, scanner, computer and software, are expensive and are not readily adaptable to uses other than array processing and analysis. One advantage of these systems is that the quality control of the arrays themselves is high and is generally guaranteed. Another advantage is that no efforts need be spent toward probe design or clone maintenance (although this can also be a disadvantage, since it may be difficult to get information from the company about the exact sequences represented on the array). Still another advantage is that the bioinformatics component of the product is higher since sequence annotation (and possibly other database information) is typically part of the package. Members of the group using commercial array systems included Larry Steinman, Jim Olson, Ruth Luthi-Carter, Leslie Thompson and Paul Coleman.

Some fee-for-service expression profiling services are available, and these may be a

worthwhile option for investigators in small laboratories who want to use the technique for a limited set of experiments.

Decisions between in-house versus commercial options will ultimately depend on the resources of a particular laboratory, e.g. cash, manpower, space, core equipment, etc.

C. Focused versus expanded. In general, the more expanded the array, the higher the cost, irrespective of whether the system is lab-made or commercial, so the nature of the experiment is likely to dictate what is optimal (e.g. in a "screening" type experiment more might be better, whereas in a hypothesis-driven study where regulation of particular gene pathways are suspected, a targeted approach might be more cost-effective).

Some HD researchers said that they could envision a point in the future when custom arrays representing the majority of the gene expression changes discovered in early "screening" experiments would be useful.

D. Known genes versus ESTs. Considerations for screening known genes versus ESTs are fairly obvious; ESTs are less well characterized, so their identities are less certain. In general, studies of differentially detected EST sequences will require a larger investment in confirmatory analyses, for this reason. On the other hand, studying the differential expression of EST sequence offers a higher likelihood of discovering the function of novel molecules.

Software development

The group was in general agreement that array data was easier to acquire than to analyze. Much of the available software is geared for one-to-one comparisons or cluster analysis, without the capacity for quantitative analyses of multiple data sets.

Wing Wong and Kenneth Lange indicated that they and other professors at UCLA have been working with laboratories generating expression profiling data to create new analytical programs. The development of this software requires access to multiple data sets which may be studied to derive algorithms to estimate confidence intervals for particular probe tilings, etc. Paul Coleman's laboratory has been working with a strategy called canonical analysis, which he described as being similar to principal components analysis. He has also used ANOVA and cluster analyses.

Ken Lange announced that Department of Mathematics at UCLA is also preparing to host a quarter-long conference on the analysis of microarray data. Interested parties were asked to contact Mark Greene.

Bioinformatics and database development

Molecular databases:

Incyte is very active in their development of a molecular database. This database contains multiple novel sequences and provides extensive sequence annotation.

Gene Expression Profiling Databases:

Many investigators agreed that mechanisms for archiving expression profiling data were needed. To date, most laboratories have posted their data on a web site as their studies were published. Many investigators at the workshop expressed similar plans for their data. HDF has been contemplating supporting an expression profiling database specifically for investigators studying polyglutamine diseases to use to deposit and share expression profiling data. Paul Coleman and Jim Olson suggested that raw data be accommodated in such a database. Wing Wong agreed, and suggested that, ideally, the raw images should be warehoused. As a second choice, Wing suggested warehousing the PM and MM probe intensity in oligo arrays, and

the separate channel intensity for each spot in the spotter arrays. He recommended that the stored data be raw intensities without scaling and normalization, so that users can apply different processing algorithms of their choice. Wing and his colleagues have learned from experience that the importance of raw data cannot be underestimated. For example, in their experience with dbEST, they discovered that ESTs for which raw trace files were available were much more useful for SNP discovery than those without raw trace files. Since storage is generally not a problem at this time, he sees no reason not to accommodate raw data.

NCBI's Gene Expression Omnibus (GEO) web-accessible database should be fully operational later this year. In addition, information on gene expression in different regions of the brain is available from a growing sequence collection in the Brain Molecular Anatomy Project (BMAP; <http://brainest.eng.uiowa.edu/>).

Software development (see above) will also benefit by increased access to expression profiling data.

Lyle Arnold indicated that Incyte's model for academic partnership was to exchange raw data for the ability to put information in their database.

Standardization and cross-comparisons of microarray platforms

In terms of array to array variability, a number of investigators present said that they thought global scaling of the total hybridization signal was a useful approach that could be applied to the majority of cases.

For any given RNA, one can draw a standard curve by spiking known amounts of RNA into the sample or cRNA/cDNA into the probe pool. This type of approach has been used by many

members of the group including those using Affymetrix, Incyte, Clontech and custom arrays.

Mark Erlander and Paul Coleman reminded the group that if the sample is amplified, such a curve would vary greatly from mRNA to mRNA because the efficiency of amplification is known to be sequence-dependent. In such cases, it is wise to draw curves for a number of mRNAs (and, if possible, include a range of those known to represent a broad range between efficiently and inefficiently amplified).

Another approach for normalizing between samples, and possibly also between platforms, would be to scale a particular kind of sample based on the hybridization signal for a set of “housekeeping genes.”

Lyle Arnold explained that Incyte had considered carefully the issue of how to relate expression array data to biological reality. They do not use amplified samples in their system. Incyte uses many of the approaches mentioned above and is working on improved sets of internal and external standardization controls.

Investigators in the group who were interested in being able to compare data across platforms thought that it would be extremely helpful if a set of common standards could be identified and included in all array platforms. It was recognized that this would require the cooperation of the array manufacturers. It could easily be used to compare data between custom arrays, requiring only that investigators have access to the control probe sequences. Lyle Arnold expressed interest in this and proposed that Incyte would be willing to be involved in the development of common standards.

Steve Tapscott reminded the group that variability and scaling of array data was dependent on the complexity of the sample (e.g. whether multiple cell types were represented). Steve also shared the results of an experiment

that had been done in his laboratory. Three plates of tissue culture cells were transfected with MyoD, and profiled against three control plates using Affymetrix expression arrays (representing approximately 6000 genes). Looking at one comparison, 500 mRNAs received Difference Calls. If data from two comparisons were examined, approximately 200 of the same mRNAs would be Called (in the same direction) in both. If three paired comparisons were examined, only 20 Calls were common to all. Investigators in Tapscott’s laboratory have also used array data to learn about genes. They have also regulated direct and indirect regulation of gene networks by the transcription factor NeuroD2.

Luda Diachenko indicated that Clontech scientists had done studies of their membrane-based arrays and found the following: If one pair of hybridizations was screened for differential expression, approx. 25% of the differentially detected genes were true positives; for two pairs of hybridizations, true positives rose to approx. 50%; for three pairs of hybridizations, it was approximately 75%; and for four pairs of hybridizations, the true positive rate was greater than 90%.

Cost comparisons of various microarray platforms

When considered in terms of price per gene per array, the various commercial platforms were similar in price. Some rough estimates: an Affymetrix Mu 11K array set is approximately \$3400 for approximately 11,000 genes (or 31 cents per gene), an Incyte murine GEM 5200 array is approximately \$2000 for 7854 genes (or 25 cents per gene), and a Clontech Atlas Murine 1.2 cDNA array is approximately \$350 for 1176 genes (or 30 cents per gene). Dan Geshwind points out the estimate of cost per gene as calculated here can be deceiving, in that some array systems allow the hybridization of two

probe samples to a single array simultaneously, while others allow only one probe sample per array. Also, some arrays can be re-probed and others cannot.]

Arrays fabricated in-house may vary considerably in cost. Factors which account for differences in the cost to an individual lab include: whether or not arrays are synthesized by an institutionally-based core facility (and whether or not the facility's operating costs are subsidized directly by the users or indirectly by the organization), whether or not robotic array equipment is used (a robotic arrayer can cost up to \$200,000), whether or not sequence confirmation of clones is included in their price, and whether or not labor costs in the maintenance of clone stocks are calculated as part of the array price.

The following are a few examples of how cDNA array costs vary from laboratory to laboratory and institution to institution:

An estimate by Dan Geshwind, at UCLA: The fabrication of 100-120 high quality glass arrays containing 5000 genes would cost about \$10,000. This would be \$100/slide and less than 10 cents per gene (2 cents per gene/slide). This estimate includes PCR, purification and arraying, but not cost of the original procurement of the clones. If one paid \$20,000 for making or buying a clone library (a high estimate), the cost would be approximately 5 cents per clone. (At UCLA, the library costs are split between several groups.) These cost estimates do not include equipment, maintenance etc., which would be assumed by the core facility.

An estimate by Paul Coleman, at U. Rochester: The Coleman lab approach can be used without a subsidized core facility. They make their own cDNA arrays using a replicator. (One replicator allows the spotting on membranes of 64 cDNAs, each replicated 9 times. Another allows

spotting 382 cDNAs, each spotted twice. The cost of the non-robotic replicators is insignificant.) The most significant cost is associated with collecting the clones and making maxi preps. They get about 30-50 arrays from one maxi prep, estimated to cost approximately \$300 each (counting technician time). For 382 genes this is a total cost of \$114,600. Assuming fifty 382-gene arrays per set of maxi preps, the cost would be \$2292 per array or \$6 per gene/array.

An estimate by Leslie Thompson, at UCI: These costs are all based on the operation of an array facility that is subsidized. For prearrayed human sets (8000 cDNAs and ESTs), the charge is \$210 per slide (less than 3 cents/gene/slide). The charge to develop custom arrays varies, but includes a \$30.00 per run/\$20 set-up per slide flat charge and (depending upon the number of samples per slide) can range from \$70 per slide for 1000 genes up to \$300 per slide for 8500 to 10,000 genes for the spotting of the arrays (with user-provided DNAs). Use of the scanner costs \$20.00 per hour, and there are other charges for preparing cDNAs, making probes, etc.

The in-house oligonucleotide array described in Michael Sussman's presentation (see above and Singh-Gasson *et al.*, 1999) costs approximately \$100-200/array for 12,000 genes, (or 2 cents per gene/array). The resolution of most available scanning/microscopic detection devices limits use to approximately 400 genes at present, but high-resolution (2 μ m) scanners should be available very soon.

Experimental needs and institutional resources will play major roles in determining which system is most cost-effective for a particular laboratory or application.

Limitations of an incomplete set of genomic data

Steve Tapscott and Lena Peltonen reminded the group that although expression profiling is quite useful in its current form, there will be a new phase within the next few years (approximately 2 to 5) when investigators will be able to perform expression profiling experiments that will cover the entire genome. It was hoped that some resources would be reserved for that time. Lyle Arnold indicated that one of the strengths of the Incyte platform was their extensive sequence collection and its representation on their human expression arrays.

Larry Steinman and Michael Sussman added that large advances in proteomics-based methodologies are also underway and the investigators should consider the emerging impact of this approach.

Brain bank issues

Many concerns and considerations of collecting human brain samples were discussed. Collecting intact RNA requires that the tissue be obtained with a relatively short post-mortem interval. Tissue is preferably blocked and flash frozen upon collection. For pathologic workups, parallel samples of fixed tissue should be available. Also, record keeping must be thorough and accurate. Information on cause of death and other antemortem factors is useful in ascertaining whether the sample is likely to be a good representation of HD status.

It was suggested that Deborah Mash at the University of Miami was a good role model for how to collect, organize and keep records on tissue specimens. Also, the pathologist with whom Paul Coleman has worked has been reliable, and might have an example pathology reporting form that could be useful to other members of the group.

The group also considered what kinds of HD brain samples were most valuable to profile. Larry Steinman explained that his group had obtained reproducible and informative data with Grade 3 cortex (study discussed above). Leslie Thompson preferred to consider profiling primarily early grade cases. Mark Erlander cautioned against using the expensive microarray technology to address questions of changing cell populations, when such issues are readily amenable to histologic and other more conventional methods.

The group also discussed how tissue should be shared. Jim Olson suggested that the easiest way to distribute samples to multiple laboratories would be to send it to a central location to extract RNA and then split the RNA between groups. Ruth Luthi-Carter suggested that parts of the tissue blocks also be reserved for histology. It was recognized by all investigators that frozen blocks of early grade cases are extremely rare and valuable. At the same time, Nancy Wexler pointed out, the purpose of collecting these specimens was to study HD, not to have a large collection of material.

Processing/Saving specimens for RNA analysis

To ensure RNA integrity, it is preferable to block and/or dissect samples of fresh tissue and then freeze the sample quickly in liquid nitrogen vapor or dry ice. This is rarely possible with human material, however. In such cases, even the best samples have a post-mortem delay of a few hours and may require further dissection to harvest material from the region of interest (e.g. caudate). Tissue samples and extracted RNA should generally be stored at -70 to -80 deg. C.

The best criteria available to choose human samples prior to RNA extraction are: low post-

mortem delay, flash-frozen as blocks, tissue pH and, if possible, prior success in RNA studies. RNA can theoretically be harvested from fixed tissue, but the recovery is variable and generally less efficient than from fresh tissue. Hybridization to test arrays can be used as a tool to determine the integrity of a probe prior to its hybridization to an expensive analysis array.

Paul Coleman's lab has had good success with Ambion's product RNA Later. This is a solution into which one can drop a sample and store it at 4 degrees C for several months. (The protein in the samples gets denatured, however, making it unsuitable for use in preserving samples from which both RNA and proteins are to be measured, or for histologic studies.)

For animal studies done in collaboration with Jim Olson, dissection methods have been carefully standardized so as to minimize variability of the cell types represented. Jim Olson and Ruth Luthi-Carter offered to give their protocol to anyone who might be interested.

Microdissection approaches to RNA profiling

Mark Erlander and colleagues published a study last year which combined the laser capture microdissection (LCM) technique with microarray-based expression profiling (Luo *et al.*, 1999). Erlander also shared that his group was working on an expression profile "map" of the basal ganglia. They are dissecting various basal ganglia structures and comparing their expression profiles to what is already known about the circuitry of the system.

Paul Coleman mentioned that his group plans to look at cell type specific expression profiles of a full-length cDNA model of HD developed in Dan Tagle's laboratory (Reddy *et al.*, 1998).

Pre-clinical and clinical trials

Microarray data is likely to suggest new therapeutic approaches to treating HD, and it could become difficult to determine which of these leads to take into pre-clinical and clinical trials. Considerations for pre-clinical studies include which models to use, when to start treatment and what to use as a "readout" or "endpoint" for the analysis. For clinical trials, the possibility of treating presymptomatic individuals, the placebo effect, the difficulties of assigning HD patients to placebo groups, and the use of the clinical scale as an endpoint were all issues being explored. Jim Olson suggested that modification of future clinical research trials might avoid a placebo group by comparing various treatments against each other. Olson suggested that some up-and-coming therapeutics to consider taking to clinical trials might be coenzyme Q, creatine and retinoic acid analogs. Bob Baughman mentioned that idebenone, a coenzyme Q analog developed by Takata, was in clinical trials for the treatment of Frederich's Ataxia. Ethan Signer indicated an interest in exploring whether the compound might be available for use in preclinical trials in HD models. Although, there has been one previous trial of idebenone in HD that failed to show a significant effect (Ranen *et al.*, 1996).]

Mark Erlander suggested that the use of surrogate markers for clinical status (e.g. in blood) be explored.

Where do microarray studies fit into the grand scheme of HD research?

Screening method. All investigators agreed that microarray gene expression profiling was a good screening tool. A large number of mRNAs can be measured at once and this lends insight into the mutant-length polyglutamine effects on the brain, including providing a window into the dysregulation of gene

expression that may be caused by specific interactions with transcription factor proteins. The number of studies required to complete a sufficiently large screen was a matter of debate, however. At one extreme, profiling multiple animal, cell and other models in addition to several independent studies of human HD brains could be envisioned. At the other end of the spectrum, it was suggested that: a) the studies already completed would suffice, or b) further studies be postponed until a full set of genomic data was available. In the middle ground, the profiling of a subset of the available models together with a few independent screens in human cases was envisioned to be a good (and cost-effective) representation.

Jim Olson and Mark Erlander proposed a model whereby a wide screen using expanded arrays could identify a set of differentially expressed genes and then the subset of candidate genes could be used to produce less-expensive custom arrays. Lyle Arnold mentioned that Incyte might be able to produce custom arrays for members of the polyglutamine research community.

Monitoring tool. Jim Olson indicated that his group intended to use microarray profiling as an outcome measure for preclinical trials of retinoic acid ligands. The rationale for using arrays in this way is that retinoic acid receptors (which appear to be affected in HD) are themselves transcription factors, and therefore manipulation of retinoid systems would be expected to show consequences at the level of mRNA populations.

Some members of the group expressed opposition to using microarray studies as a monitoring tool because of their expense. Others saw merit in the approach, particularly as it related to changes in the cell which were thought to impact transcription.

Functional studies were thought to be an important component of therapeutic monitoring. Functional studies are not exclusive of microarray measures, however (the two can be used in parallel).

Other uses and other considerations. Larry Steinman suggested that profiles of HD cases and models would be most valuable when studied in the context of other neurologic diseases.

Jim Olson and Ethan Signer also discussed the possibility of determining the range of effects of neurologically active drugs already in use by employing expression profiling in experimental animals. This would have broad application to many neurologic and psychiatric illnesses.

Art Arnold and Luda Diachenko brought up the point that variations from a simple one-sample or two-sample hybridization can be used, such as a subtractive approach. Luda Diachenko also elaborated on ways that arrays can be used in combination with other differential detection approaches; Clonetech is actively working in this area.

What is needed to support microarray Studies of HD?

Funding. The HDF was interested in suggestions for how best to use their resources to collect an informative set of expression profiling data. One suggestion was to award small grants (approx. \$50,000) to a number of groups (approx. 10) who would conduct independent studies. Mark Erlander suggested that available funds be given to the groups who had already established a productive record in HD profiling studies. It seemed to be the consensus of the group that for the first phase of array experiments there was no need to invest in a core facility to make arrays. However, a second phase of expression studies might

incorporate a custom “HD array” which could be centrally manufactured and distributed to a number of laboratories. This array would be comprised of tilings for mRNAs found to be differentially expressed in the first phase of array studies.

Bob Baughman also indicated that NIH is looking for ways to support microarray-based gene expression studies. Ideas being contemplated are: A) large grant support of array manufacturing core facilities which would supply academic investigators with arrays for at minimal cost, B) support of array-based projects (e.g. by an R21 mechanism), and C) support of individual institution or laboratory core array facilities. Some gene-expression RFA (R21) programs have already been initiated.

Tissue. Carefully collected HD tissue, and particularly that from low-grade and presymptomatic cases, is extremely rare. The HDF has been collecting tissue for many years and its collections retain many ideal specimens. Thus, one major potential contribution from HDF to investigators engaged in microarray profiling might be access to these very valuable cases.

Given that various platforms have unique strengths and weaknesses, it would be useful to analyze some samples on more than one platform. One collaboration between the investigators at MGH, FHCRC and UC Irvine has already been planned to compare human cases. Mechanisms for sharing rare HD tissues among multiple investigators were discussed (“see Brain Bank” above).

Choice of Models.

Mouse: The general consensus of the group was that several mouse models should be profiled so they could be compared to each other, as well as to human. It was suggested that large-scale profiling efforts should include at

least one inducible model and one full-length model. It was suggested by Jim Olson that lower-cost custom arrays could be used to expand expression profiling studies to other models in a cost-effective way.

Cells: Cell models were also considered to be of interest. Leslie Thompson mentioned that her group would profile their inducible, stably-transfected huntingtin PC12 cell lines. It was recognized that cell models would be particularly useful for mechanistic studies and the rapid screening of potential therapeutic agents.

Human: The value of profiling human samples was debated. On one hand, human samples are highly heterogeneous (because of genetic background), and this would theoretically require the expense of profiling a large number of cases. On the other side, Larry Steinman’s group has shown that by profiling several Grade-matched cases and comparing them to controls and to cases of other neurologic disorders, many common changes can be identified. (See also “Microarray Studies” above.)

Some investigators thought that additional human profiling were central to the mission of understanding gene regulation in HD (see also “Brain Bank issues,” above). Some thought profiling human cases would only be valuable if a sufficient number of presymptomatic/early grade cases could be identified.

Other approaches to using human samples were also discussed. These included using human cases only to verify gene-expression changes identified in model systems, since this could be accomplished using methodologies which are more forgiving of differences in sample collection and preparation (e.g. to use in situ hybridization histochemistry or to quantitate at the protein level). Another idea was to use the data already collected by the Steinman group as

the human standard (these data will be made available on the web when the study is published).

Choice of Platforms. The two HD profiling studies discussed earlier in the meeting (by the Steinman and Olson groups) both used Affymetrix arrays. Additional studies with custom cDNA and Incyte cDNA arrays were already underway, however. Moreover, not all laboratories have equivalent access to Incyte or Affymetrix systems (some simply because of cost). Thus, the group consensus appeared to be that choosing a "preferred" platform for future studies was not practical (a point made individually by Bob Baughman and Steve Tapscott).

In order to promote cross-comparison of data sets from different polyglutamine disease samples, however, Jim Olson has initiated the formation of a consortium of investigators who will share data generated on the Affymetrix platform. Olson indicated that the consortium was by no means an exclusive group, and he invited all of the workshop attendees to participate.

Data sharing. It seemed to be the consensus of the group by the end of the meeting that it would be logistically difficult to try to conduct all HD studies on a common platform. Therefore, the highest priority would be given to creating a forum for data sharing. One major way to promote data sharing would be to sponsor a web-based data repository that would contain not only the investigators' analyses and interpretations of the data, but also the raw data. (The latter was considered to be extremely important.)

In order to make this information widely usable, it would be best to have it available in a form that could be handled by various software platforms. Lyle Arnold mentioned that Incyte's query into this issue led them to decide to use

SML. Others agreed that this would be a reasonable solution. Lyle Arnold also offered to put investigators in touch with the software specialists to help them with their own software "plug-in" issues.

References

- Augood S. J., Faull R. L., Emson P. C. (1997) Dopamine D1 and D2 receptor gene expression in the striatum in Huntington's disease. *Ann Neurol* **42**, 215-221.
- Augood S. J., Faull R. L., Love D. R., et al. (1996) Reduction in enkephalin and substance P messenger RNA in the striatum of early grade Huntington's disease: a detailed cellular in situ hybridization study. *Neuroscience* **72**, 1023-1036.
- Boutell J. M., Thomas P., Neal J. W., et al. (1999) Aberrant interactions of transcriptional repressor proteins with the Huntington's disease gene product, huntingtin. *Hum Mol Genet* **8**, 1647-1655.
- Callahan L. M., Selski D. J., Martzen M. R., et al. (1994) Preliminary evidence: decreased GAP-43 message in tangle-bearing neurons relative to adjacent tangle-free neurons in Alzheimer's disease parahippocampal gyrus. *Neurobiol Aging* **15**, 381-386.
- Callahan L. M., Vaules W. A., Coleman P. D. (1999) Quantitative decrease in synaptophysin message expression and increase in cathepsin D message expression in Alzheimer disease neurons containing neurofibrillary tangles. *J Neuropathol Exp Neurol* **58**, 275-287.
- Cha J. H., Frey A. S., Alsdorf S. A., et al. (1999) Altered neurotransmitter receptor expression in transgenic mouse models of Huntington's disease. *Philos Trans R Soc Lond B Biol Sci* **354**, 981-989.
- Cha J. H., Kosinski C. M., Kerner J. A., et al. (1998) Altered brain neurotransmitter receptors in transgenic mice expressing a portion of an abnormal human huntington disease gene. *Proc Natl Acad Sci U S A* **95**, 6480-6485.
- Cheetham J. E., Coleman P. D., Chow N. (1997) Isolation of single immunohistochemically identified whole neuronal cell bodies from post-mortem human brain for simultaneous analysis of multiple gene expression. *J Neurosci Methods* **77**, 43-48.
- Davies S. W., Turmaine M., Cozens B. A., et al. (1997) Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **90**, 537-548.
- Diamond M. I., Robinson M. R., Yamamoto K. R. (2000) Regulation of expanded polyglutamine protein aggregation and nuclear localization by the glucocorticoid receptor. *Proc Natl Acad Sci U S A* **97**, 657-661.
- Georgalis Y., Starikov E. B., Hollenbach B., et al. (1998) Huntingtin aggregation monitored by dynamic light scattering. *Proc Natl Acad Sci U S A* **95**, 6118-6121.
- Harkin D. P., Bean J. M., Miklos D., et al. (1999) Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. *Cell* **97**, 575-586.

- Holstege F. C., Jennings E. G., Wyrick J. J., et al. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717-728.
- Huang C. C., Faber P. W., Persichetti F., et al. (1998) Amyloid formation by mutant huntingtin: threshold, progressivity and recruitment of normal polyglutamine proteins. *Somat Cell Mol Genet* **24**, 217-233.
- Iyer V. R., Eisen M. B., Ross D. T., et al. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83-87.
- Karpuj M. V., Steinman L. (1999) Transcriptional profile of Huntington disease brain with gene microarrays. *Soc. Neurosci. Abstr.* **25**, 541.
- Kazantsev A., Preisinger E., Dranovsky A., et al. (1999) Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells. *Proc Natl Acad Sci U S A* **96**, 11404-11409.
- Luo L., Salunga R. C., Guo H., et al. (1999) Gene expression profiles of laser-captured adjacent neuronal subtypes [published erratum appears in *Nat Med* 1999 Mar;5(3):355]. *Nat Med* **5**, 117-122.
- Luthi-Carter R., Strand A., Peters N. L., et al. (2000) Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. *Human Mol Genet*, in press.
- Ona V. O., Li M., Vonsattel J. P., et al. (1999) Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease [see comments]. *Nature* **399**, 263-267.
- Persichetti F., Trettel F., Huang C. C., et al. (1999) Mutant huntingtin forms in vivo complexes with distinct context-dependent conformations of the polyglutamine segment. *Neurobiol Dis* **6**, 364-375.
- Perutz M. F. (1999) Glutamine repeats and neurodegenerative diseases: molecular aspects. *Trends Biochem Sci* **24**, 58-63.
- Ranen N. G., Peyser C. E., Coyle J. T., et al. (1996) A controlled trial of idebenone in Huntington's disease. *Mov Disord* **11**, 549-554.
- Reddy P. H., Williams M., Charles V., et al. (1998) Behavioural abnormalities and selective neuronal loss in HD transgenic mice expressing mutated full-length HD cDNA. *Nat Genet* **20**, 198-202.
- Scherzinger E., Lurz R., Turmaine M., et al. (1997) Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell* **90**, 549-558.
- Scherzinger E., Sittler A., Schweiger K., et al. (1999) Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc Natl Acad Sci U S A* **96**, 4604-4609.

Sieradzan K. A., Mehan A. O., Jones L., et al. (1999) Huntington's disease intranuclear inclusions contain truncated, ubiquitinated huntingtin protein. *Exp Neurol* **156**, 92-99.

Singh-Gasson S., Green R. D., Yue Y., et al. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array [see comments]. *Nat Biotechnol* **17**, 974-978.

Starikov E. B., Lehrach H., Wanker E. E. (1999) Folding of oligoglutamines: a theoretical approach based upon thermodynamics and molecular mechanics. *J Biomol Struct Dyn* **17**, 409-427.

Steffan J. S., Kazantsev A., Spasic-Boskovic O., et al. (2000) The Huntington's disease protein interacts with p53 and CBP and represses transcription *Proc Natl Acad Sci U S A* in press.

Wood J. D., Yuan J., Margolis R. L., et al. (1998) Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins. *Mol Cell Neurosci* **11**, 149-160.