

**DRAFT COPY – THE FINAL DRAFT OF THIS REPORT WILL BE POSTED WHEN  
AVAILABLE.**

**HEREDITARY DISEASE FOUNDATION**

**“Microarrays, Models, and Mechanisms”**

January 26-27, 2001  
Fred Hutchinson Cancer Research Center  
1100 Fairview Avenue North  
Seattle, Washington 98109

Prepared by Stephen Hart

**Participants**

**Barbara Apostol**  
949-824-1010  
949-824-2577  
E-mail: blaposto@uci.edu

Department of Psychiatry  
University of California, Irvine  
Gillespie 2216  
Irvine, California 92697

**Neil Aronin**  
508-856-3239  
508-856-6950 Fax  
E-mail: neil.aronin@banyan.ummed.edu

Department of Medicine  
University of Massachusetts Medical School  
55 Lake Avenue North  
Worcester, Massachusetts 01655

**Sarah Augood**  
617-726-5570  
617-724-1480 Fax  
E-mail: augood@helix.mgh.harvard.edu

Neurology Research, WRN 408  
Massachusetts General Hospital  
32 Fruit Street  
Boston, Massachusetts 02114

**Donald Bergstrom**  
206-667-4286  
206-667-6524 Fax  
E-mail: dbergstr@fhcrc.org

Fred Hutchinson Cancer Research Center  
1100 Fairview Avenue North, C3-168  
PO Box 19024  
Seattle, Washington 98109-1024

**Elena Cattaneo**  
+39 02 20488349  
+39 02 029404961 Fax  
E-mail: elena.cattaneo@unimi.it

University of Milano  
Department of Pharmacological Sciences  
Via Balzaretti 9, 20133  
Milano, Italy

**Jang-Ho Cha**  
617-724-1481  
617-724-1480 Fax  
E-mail: cha@helix.mgh.harvard.edu

Neurological Research, WRN 408  
Massachusetts General Hospital  
Fruit Street  
Boston, Massachusetts 02114

**Edmond Chan**  
604-875-3809  
604-875-3819 Fax  
E-mail: echan@cmmt.ubc.ca

Department of Medical Genetics  
University of British Columbia  
Room 3024, Building 950  
West 28<sup>th</sup> Street  
Vancouver, British Columbia V5Z 4H4  
Canada

**Chun Cheng**  
206-667-7256  
206-667-7004 Fax  
E-mail: ccheng@fhcrc.org

**William Dauer**  
212-543-5309  
212-543-5074 Fax  
E-mail: WTD3@columbia.edu

**Jeffrey Delrow**  
206-667-2763  
206-667-4029  
E-mail: jdelrow@fhcrc.org

**Barrett Eynon**  
650-568-3057  
650-365-1735 Fax  
E-mail: barry@sigenetics.com

**Nathan Goodman**  
617-755-4131  
617-441-3535 Fax  
E-mail: ngoodman@3rdmill.com

**Stephen Hart**  
360-452-5544  
954-301-5773 Fax

**Robert Hughes**  
206-616-4522  
206-543-0754 Fax  
E-mail: rehughes@u.washington.edu

**Carl Johnson**  
608-767-2193  
608-767-3732 Fax  
E-mail: carl\_johnson@axyspharm.com

**Marcela Karpuj**  
415-502-8831  
415-476-8386 Fax  
E-mail: marcela@neurology.stanford.edu

**Tim-Rasmus Kiehl**  
310-423-7611  
310-423-0299  
E-mail: kiehlt@cshs.org

**Ilya Kupersmidt**  
650-591-4459

Fred Hutchinson Cancer Research Center  
1100 Fairview Avenue North, MW-805  
PO Box 19024  
Seattle, Washington 98109-1024

Department of Neurology  
Columbia Presbyterian Medical Center  
Neurological Institute  
710 W. 168<sup>th</sup> Street, Box 204  
New York, New York 10032

Fred Hutchinson Cancer Research Center  
1100 Fairview Avenue North, D4-100  
PO Box 19024  
Seattle, Washington 98109-1024

Silicon Genetics  
2601 Spring Street  
Redwood City, California 94063

3<sup>rd</sup> Millennium, Inc.  
125 Cambridge Park Drive  
Cambridge, Massachusetts 02140

Writer and Editor  
266 Heather Park Road  
Port Angeles, Washington 98362

Department of Genetics  
University of Washington  
J-205 Health Sciences Building  
Box 357360  
Seattle, Washington 98195-7360

4246 Blue Mounds Trail  
Black Earth, Wisconsin 53515

Department of Neurology and Neurological Services  
Stanford University  
Beckman Center, B002  
Stanford, California 94305-5429

UCLA School of Medicine – Cedars Sinai  
8361 West 3<sup>rd</sup> Street, #1145E  
Los Angeles, California 90048

Silicon Genetics  
2601 Spring Street

650-365-1735 Fax  
ilya@sigenetics.com

Redwood City, California 94063

**Andrew Lieberman**  
301-435-9288  
301-480-3365 Fax  
E-mail: lieberman@ninds.nih.gov

Neurogenetics Branch  
NINDS, NIH  
Building 10, Room 3B14  
MSC 1250  
Bethesda, Maryland 20892-1250

**Ruth Luthi-Carter**  
617-726-5227  
617-724-1480 Fax  
E-mail: carter@helix.mgh.harvard.edu

Department of Neurology Research  
Massachusetts General Hospital  
Warren 401  
Boston, Massachusetts 02114

**James Olson**  
206-667-7955  
206-667-2917 Fax  
E-mail: jolson@fhcrc.org

Fred Hutchinson Cancer Research Center  
1100 Fairview Avenue North, D4-100  
PO Box 19024  
Seattle, Washington 98109-1024

**Christopher Ross**  
410-614-0011  
410-614-0013 Fax  
E-mail: caross@jhu.edu

Laboratory of Molecular Neurobiology  
Ross Building, Room 618  
Johns Hopkins University  
720 Rutland Avenue  
Baltimore, Maryland 21205-2196

**Erik Schweitzer**  
310-206-4084  
310-206-5855 Fax  
E-mail: schweitzer@mednet.ucla.edu

Brain Research Institute  
University of California Los Angeles  
2506 Gonda Center  
Box 951761  
Los Angeles, CA 90095-1761

**Daniel Scoles**  
310-423-7610  
310-423-0299 Fax  
E-mail: scolesd@cshs.org

Neurogenetics Laboratory, Room D2091  
Cedars-Sinai Medical Center  
8700 Beverly Blvd.  
Los Angeles, California 90048

**Steven Self**  
206-667-4944  
206-667-4812 Fax  
E-mail: sgs@hivnet.fhcrc.org

Fred Hutchinson Cancer Research Center  
1100 Fairview Avenue North, MW-500  
PO Box 19024  
Seattle, Washington 98109-1024

**Simonetta Sipione**  
+39 02 20488353  
+39 02 029404961 Fax  
E-mail: ssipione@hotmail.com

University of Milano  
Via Balzaretti 9  
Milano 20133  
Italy

**Lawrence Steinman**  
650-725-6401  
650-725-0627 Fax  
E-mail: steinman@stanford.edu

Department of Neurology and Neurological Sciences  
Stanford University  
Beckman Center, B002  
Stanford, California 94305-5429

**Andrew Strand**  
206-667-5066  
206-667-2917 Fax

Fred Hutchinson Cancer Research Center  
1100 Fairview Avenue North, D4-100  
PO Box 19024

E-mail: astrand@fhcrc.org

Seattle, Washington 98109-1024

**Stephen Tapscott**

206-667-4499

206-667-6524 Fax

E-mail: stapscot@fhcrc.org

Fred Hutchinson Cancer Research Center

1100 Fairview Avenue North, C3-168

PO Box 19024

Seattle, Washington 98109-1024

**Leslie Thompson**

949-824-6756

949-824-2577 Fax

E-mail: lmthomps@uci.edu

Departments of Psychiatry and Biological Chemistry

University of California Irvine

College of Medicine

Room D-240, Medical Sciences 1 Gillespie 2121

Irvine, California 92697-4260

**Allan Tobin**

310-825-5061

310-267-0341 Fax

E-mail: atobin@mednet.ucla.edu

Brain Research Institute

University of California Los Angeles

2506 Gonda Center

Box 951761

Los Angeles, CA 90095-1761

310-575-9656

310-575-9156 Fax

Hereditary Disease Foundation

11400 West Olympic Blvd., Suite 855

Los Angeles, CA 90064

**Nancy Wexler**

212-543-5667

212-543-6002 Fax

E-mail: 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, New York 10032

310-575-9656

310-575-9156 Fax

Hereditary Disease Foundation

11400 West Olympic Blvd., Suite 855

Los Angeles, CA 90064

**Sara Winokur**

949-824-2750

949-824-2688 Fax

E-mail: steinoku@uci.edu

Department of Biological Chemistry

University of California, Irvine

Medical Sciences 1

Irvine, California 92697

**Wing Wong**

310 206-9475

310 206-6673 Fax

E-mail: whwong@stat.ucla.edu

Department of Statistics

University of California Los Angeles

8118 Math Science Building

Mail Code 155404

Los Angeles, California 90095

**Ai Yamamoto**

212-543-5569

212-543-5074 Fax

E-mail: ay46@columbia.edu

Center for Neurobiology and Behavior

College of Physicians and Surgeons

Columbia University

722 West 168<sup>th</sup> Street

P.I. Annex 725

New York, New York 10032

**Anne Young**

617-726-2385

Neurology Service, VBK 915

Massachusetts General Hospital

617-726-2353 Fax  
E-mail: young@helix.mgh.harvard.edu

32 Fruit Street  
Boston, Massachusetts 02114

Microarray technology promises to provide gene expression data quickly on a large number of genes. Ideally, microarray studies could provide insight into the effect the expansion of the Huntington gene exerts on the expression of a wide variety of other genes. With that picture in hand, researchers could focus on genes that affect the phenotype of Huntington's Disease.

The Hereditary Disease Array Group (HDAG) is a consortium of approximately 50 scientists from 18 laboratories that is conducting microarray-based experiments on neurodegenerative disease models.

Thirty-four members of HDAG gathered in Seattle and Boston—linked by videoconferencing technology—on January 26-27, 2001 in a round-table discussion on the promises and realities of microarray technology as applied to models of Huntington's disease and other polyglutamine diseases.

The discussion generated a number of questions, How do researchers know what data is real and what is noise? Which statistical model best represents reality? What are the next steps?

Coordinator James Olson, of Fred Hutchinson Cancer Research Center (FHCRC), presented a schedule for the two days, then introduced Tom Bird, of the Department of Neurology at the University of Washington, to present a Huntington's Disease patient. Rebecca, 47, lives by herself north of Seattle. She pointed out a number of the day-to-day effects of Huntington's Disease, including her inability to continue her job, difficulty with normal tasks such as brushing her teeth, the need to eat a high-fat diet to compensate for her body's constant activity and her inability to sleep. Rebecca's father died at 61, after 10 years of Huntington's Disease symptoms and her brother died early in January, 2001 also after about 10 years of Huntington's Disease symptoms.

### **The Origins of HDAG, The Questions that Remain**

Olson outlined the history of the HDAG, beginning with a discussion between himself and Ruth Luthi-Carter. Luthi-Carter was attempting to identify gene transcription changes occurring in Huntington's Disease, using a mouse model called R6/2. Her work arose from that of a colleague, Jang-Ho Cha, of the Massachusetts General Hospital, who, along with researchers in several other labs, had shown that the expression of certain individual genes, particularly those associated with neurotransmitter receptors, was reduced in Huntington's Disease patients and animal models.

Because Olson was using microarray technology at FHCRC in cancer studies, he and Luthi-Carter decided to apply microarray technology to a study of R6/2 mouse brains compared to wild-type mouse brains.

Gene chips reveal whether gene expression increases or decreases, and of the 140 to 150 genes changed significantly in R6/2 striatum compared to wild type, Luthi-Carter and Olson found that the number of genes whose expression decreased outnumbered the genes whose expression increased by two to one. Surprisingly, many of the genes whose expression changed fell into three broad categories, those associated with afferent neurotransmission signaling, those associated with calcium homeostasis and those associated with inflammation or stress. Only the latter group showed increased expression. [1]

The research yielded valuable data but left two important general question unanswered: Which of these gene expression changes are pathogenic and which are simply markers of cells that are damaged during the course of Huntington's Disease? And, if some of those gene changes can be linked to disease progression in Huntington's Disease, which ones may be viable targets for therapy in humans? A number of more specific questions also remain to be answered.

- Is there a sequential pattern of gene expression changes?
- Are gene expression changes restricted to a specific brain region?
- R6/2 mice are only one model of Huntington's Disease. What other models should be investigated with microarray technology?
- What is the role of normal huntingtin, and could expanded huntingtin block that role?
- What can yeast genetics tell us about Huntington's Disease?

Finally, Olson presented a question that echoed throughout the two-day meeting. How can we separate real gene expression changes from noise when using microarray technology?

### **The Models**

This report does not include the data presented by Olson at the meeting. However, during the discussion of the data and at other times throughout the meeting, participants discussed the essential features of each mouse model from which microarray data has been gathered, as well as yeast and cultured cell lines.

#### R6/2 mouse

Ruth Luthi-Carter explained the R6/2 mouse model, which contains a human exon 1 transgene driven by a human promoter. The resulting expansion is a very high CAG with about 135 to 150 repeats. The animals show spatial learning deficits beginning at 3.5 weeks, progress to overt neurological symptoms at 5.5 to 6 weeks and die at 12 weeks. The age of onset may vary up to ten percent more than a single standard deviation of a two-week period. The pathology is a general shrinking of the brain, including the cortex and striatum, but without massive cell loss in either area. These mice also have diabetes.

#### Borchelt mouse

The Borchelt mouse expresses a 171-amino acid huntingtin protein encoded by a cDNA fragment driven by the prion promoter. This is the same promoter used in the DRPLA mouse model. The first symptoms appear at six to eight weeks and the mice live four to five months.

#### Aronin mouse

Neil Aronin, of the University of Massachusetts Medical School, detailed the Aronin mouse model. Driven by an NSE promoter, the Aronin huntingtin cDNA is about 3,000 bases long, making a protein of about 980 amino acids. The model includes two lineages of each of two strains, one with a 46 repeat (46Q) and one with a 100 repeat (100Q), with an 18Q control. A spectrum of early phenotypes arises from both repeats. The 100-repeat mice show behavioral changes earlier, but both the 100Q and 46Q have very similar behavioral phenotypes as they age. Pathology mimics human disease in both the striatum and cortex.

Aronin mice may be mildly affected or severely affected. Mildly affected mice show a mild change in rotorod performance and some clasping, while severely affected mice generally show a severe rotorod deficiency and clasping, and/or gait activity change. Severely affected animals show a 20-40% cell loss, and begin to exhibit behavioral changes as early as 2 months. Most show clear behavioral changes by 8-10 months. Many can live more than a year.

Jim Olson asked how Aronin would go about determining which gene expression changes are pathogenic in severely affected mice. Aronin suggested experiments to find out if the affected genes are in neurons or glia. He would also look at early time points to check for gene expression changes that have not caused a phenotype. Once an animal is sick, many cellular processes may go wrong, causing changes in mRNA that are not relevant to pathogenesis of the disease, but are part of the processes of a sick cell.

#### YAC mice

Edmond Chan, of the Center for Molecular Medicine and Therapeutics at University of British Columbia, explained the YAC mouse model of Huntington's Disease. This model has two lines, a low-expressor (also known as 2511) [2] and a high-expressor (also known as line 44). Both lines carry the same YAC, which contains the full-length human Huntington's Disease gene with 72 CAG repeats. The YAC sequences also contain the endogenous Huntington's Disease promoter, which drives expression of mutant huntingtin. The YAC may also contain the alpha-adducin1 gene and while it's expected to be included in both lines, Chan has not yet specifically tested for its presence.

Low-expressors begin to show behavioral differences, beginning with increased activity, at six months, although electrophysiological differences can be detected at birth. Cell death is detected in medium spiny neurons of the striatum by 12 months, in addition to nuclear localization of huntingtin fragments. Microarray experiments performed so far with striata from low-expressors at 6 months of age have identified a number of genes with altered expression. A high stringency method of data mining has revealed 10 genes that show altered expression while the "Seattle-method" (see discussion below) has identified approximately 200 genes with altered expression. Efforts are currently aimed at confirmation of expression changes in addition to biological model testing experiments.

High-expressors YAC mice express four to five times the amount of mutant huntingtin as do the low-expressors—a result based on Southern and Western blotting—and appear to show much more rapid disease progression. High-expressor mice at 6 months of age show massive neuronal death in the striatum, primarily in medium spiny neurons of the lateral striatum. In contrast with the increased levels of striatal pathology, there were

no significant gene expression changes detected in microarray experiments. At least two ideas may explain these surprising data. First, the current method of dissection may miss the striatal cells most affected. Secondly, excessive cell death may have already taken place by six months in these mice, confounding the detection of gene expression changes. Chan plans to study the same mice at an earlier age, since many gene expression changes might take place before the cells die.

This brings up the important question of whether the YAC mice (both low- and high-expressors) correctly model Huntington's Disease. They should, Chan said, as they all carry the transgene and express full-length mutant huntingtin.

#### Yamamoto/Hen mouse

Ai Yamamoto discussed the mouse model she developed in Rene Hen's lab. The model expresses both exon 1—with a polyglutamine expansion of about 94 repeats—and lac z, used as a marker. Lac z shows up first in the cortex two weeks after doxycycline treatment is stopped. The mice are maintained on doxycycline throughout their prenatal life through to 4 weeks of age. By 8 weeks of age, high levels of transgene expression become evident.

#### Yeast

Robert Hughes, of the University of Washington, discussed work in yeast. One of the fortuitous things about trying to interpret yeast chip data is the large amount of work that has been done on RNA responses to many mutations. Comparing his gene lists to the literature, Hughes found work by Rick Young at MIT, who studies mutants in the histone-acetylase complex SAGA. Young has made databases of that work publicly available.

A search through that database turned up a reasonable match between Hughes's data and data for various knock-outs in the yeast SAGA complex. SAGA is homologous to the complex in humans that contains CBP and p300. Because the direction of the gene expression changes suggested a lack of acetylation, Hughes subsequently looked at the effects of trichostatin A (TSA), a histone deacetylase inhibitor. To Hughes's knowledge, there are no chemical agents that can directly augment the activities of acetylase, but compounds, such as TSA, that inhibit deacetylases are well known. So the idea was to ask if compounds such as TSA could, in a sense, correct at least some of these transcript perturbations.

Inhibiting histone deacetylases, is not likely to be a safe therapeutic approach, but the idea provides a way of thinking about a particular kind of pharmacologic intervention that could address what may be a primary mechanism of HD.

#### PC-12 cell-line model

Leslie Thompson, of the University of California Irvine, described the PC-12 cell line experiments she has undertaken with Barbara Apostol in her lab, in collaboration with Erik Schweitzer of Allan Tobin's lab at UCLA. PC-12 cells are derived from a rat pheochromocytoma. They have stably transfected these cells with an ecdysone-inducible exon 1 construct containing a C-terminal GFP tag. The construct includes alternating CAGCAA, with 25 repeats or 103 repeats made by Alex Kazantsev in David Housman's lab. Inducing the transgene in Thompson's cells has no obvious effect on viability. In parallel, Erik Schweitzer has used a bombyx-ecdysone inducible cell culture system with the same construct. Most of Schweitzer's cells die at 48 hours after induction.

Thompson compared lists of genes whose expression was called up or down with two analysis applications, CyberT, developed at UCI, and the "Seattle Analysis," using the Affymetrix software and difference calls. The most striking, even bothersome, result of this comparison, as Erik Schweitzer pointed out, was how little overlap the two analysis methods showed.

Olson replied that this discordance underlines again the need for the HDAG to complete 100 or 200 Northern analyses to help distinguish between the various analysis applications available, which presently include the Affymetrix software, CyberT, GeneSpring and a method developed by Wing Wong, of Harvard (see Data Analysis Statistics and Informatics, below).

#### Androgen receptors

Andrew Lieberman, of NINDS, uses a cell culture system consisting of a clone of motorneuron/neuroblastoma hybrid cells. These cells constitutively express either the full-length androgen receptor, which has 24 repeats (Q24), or a mutant receptor, which has 65 repeats (Q65).

Treated for 24 hours with androgen, the Q24 cells show no dramatic reprogramming. The mutation causes a partial loss of androgen receptor function. Some genes respond to the wild-type receptor but not to the mutant

receptor. Some genes respond to the mutant receptor, but not the wild-type receptor, suggesting that there may be an alteration in function caused by the mutation, not just a loss of function.

Both forms of the protein bind the ligand identically, but the mutant receptor is expressed at lower levels, according to Western blot analysis. RNA levels do not differ between the mutant and wild-type receptor, but pulse-chase studies show that the mutant receptor has a shorter half life, so the partial loss of function may be explained in part by lower levels of protein.

Both the wild-type and mutant androgen receptors behave normally in that they appear in the cytoplasm in the absence of ligand and translocate to the nucleus in the presence of ligand. Neither immunofluorescence nor Western blot reveal aggregates in these cells, even after the continuous presence of ligand for up to a week. Both forms of the androgen receptor promote the survival of these cells in the presence of low serum, so they exert some sort of trophic effect, similar to what's seen in vivo, though the wild-type receptor does it significantly better than does the mutant receptor.

#### Cattaneo cell-line model

Elena Cattaneo, of the University of Milano, has convincing data showing that normal huntingtin has some important gene transcriptional activity in cells, and that the loss of this protein function occurs in HD. Microarray data include changes in gene expression that result from the action of mutant huntingtin, but also gene expression changes arising from the loss of huntingtin function.

Cattaneo works with a cell line derived from the striatum. The cells express antigens specific to striatal medium spiny neurons. These antigens reveal, for example, DARPP-32 and D2 receptors on the cells [3] as well as functional A2A receptors whose activity is however altered in the presence of mutant huntingtin [4].

The cells appear to summarize quite nicely some of the properties of the striatal neurons. This doesn't mean that they are striatal neurons.

#### Discussion

After hearing about the mouse models, Lawrence Steinman, of Stanford University, commented that the mouse models appear just as complex as human tissue. For the most part, researchers cannot look at the early stages of HD in humans as they can with mouse models. Researchers in human tissue are looking not only at HD but asking about the effect of neurodegeneration in general. Many of the genes researchers thought would be specific to HD turn out to be altered in Alzheimer's and Multiple Sclerosis.

Allan Tobin, of HDF and UCLA, emphasized that changes in mRNA levels seen in microarray experiments could arise from a combination of inflammatory response, gliotic response and a response to the accumulation of aggregates. The accumulation of denatured proteins in cells triggers a number of responses, including the increased expression of heat-shock proteins. One way to control for these possibilities is to include earlier time points in experimental series. Another control would be to use cell lines to find a set of genes that are known to respond to the accumulation of aggregates.

Olson and Luthi-Carter have experiments in progress that will include very early time points. Analyzing time courses, however, will require new software. Current software can only handle comparisons of a single brain part at a single time versus wild type.

Donald Bergstrom, of FHCRC, spoke to the issue of time-course microarray experiments in a cell line. He studies myoD muscle transcription factor, looking at the transcriptional program and the message levels of various genes following myoD induction. They have done time courses and have found groups of genes that travel together throughout a time course.

#### Transcription Factors

Ruth Luthi-Carter discussed research from several groups showing that many transcription factors can interact directly with huntingtin and are included in huntingtin-containing aggregates, both in mouse models and in human tissue. Perhaps the transcription factor gets sequestered in the huntingtin-containing aggregate, blocking its normal role in the cell. The transcriptional activator CBP serves as an example. Any genes whose expression depended on CBP should be turned off if huntingtin or other polyglutamine-disease-causing proteins put CBP out of commission by aggregating it away from its target site, the nucleus.

Among genes commonly decreased in various model systems, the HDAG should look for target sites for CBP and cyclic AMP-driven transcriptional regulation, Luthi-Carter suggested. Cell culture experiments with reporter

constructs driven by the promoters from several of these genes may yield important information. We also may identify new target promoter sites that are non-CPB-dependent, based on considering our gene list further.

Leslie Thompson, of UC Irvine, reminded participants that CBP is an almost ubiquitous coactivator. Gene lists based on CBP activation would be huge.

Bergstrom added one more complication. More recent research on transcription factors has revealed cases where a factor considered a transcriptional activator can also function as a repressor, depending on the cellular environment. If HD or other expanded polyglutamine diseases sequester transcription factors, the result may be activation or repression. In one model gene expression may increase; in another model it may decrease. And the fact that different results appear in different experiments doesn't rule out the hypothesis that the mechanism is sequestration by polyglutamines. It may just show that the role of transcription factors is complex.

Do striatal cells die of murderers or suicide?

Tobin asked if the striatum is the right target. Is it possible these cells sustain damage as a result of earlier HD processes occurring in the cortex? That is to say, are striatal cells murder victims?

Jang-Ho Cha responded that nailing down the initiating event is important. But the microarray profiling experiments have revealed that a lot of things go wrong. The clue to treating HD is not going to be just one set of things. The effects on transcript levels revealed by microarray experiments may be secondary or tertiary, but that still has an impact on how the organism is going to respond, how the striatum or other brain region is predisposed to subsequent injury. Experiments perturbing striatal cells in culture should shed light on the relevance of the striatal microarray profiling results. Concentrating on the striatum might miss the initiating murderous event, but should help researchers sort out, whether we have a phenotype in cell culture of what an apoptotic stressor looks like in terms of these profiles, and whether it looks very similar or very different to what we see in mouse models.

Anne Young, of Massachusetts General Hospital, added that the earliest event in a disease progression may not always be the best therapeutic target. Wilson's disease, a hereditary inability to metabolize copper, is a good example. The best therapy is not aimed at the genetic cause of the disease, but at getting rid of the copper. Even though we know about the genetics of Wilson's now, and know something about what it does, the initial event is not the best target.

What role is chromatin playing in polyglutamine diseases?

Steve Tapscott, of FHCRC, explained that work in yeast and other systems has shown two basic mechanisms of regulating in an epigenetic fashion the availability of a promoter for transcription-factor binding.

One is the notion that histone acetylases and deacetylases that have global effects in gene expression and regulate the overall level of acetylation of the genome. In yeast, knocking out certain genes causes global changes in gene expression.

Superimposed on top of that is the targeting of specific chromatin remodeling factors or histone acetylases or deacetylases to specific promoters.

As Don Bergstrom pointed out, it's probably inappropriate to think of a transcription factor as just doing one thing, there are many instances—hormone binding proteins being well characterized—where depending on whether the transcription factor is targeting a deacetylase or an acetylase, it can, through signal transduction, either activate or inactivate a gene.

The global studies are very important, especially when put together with specific studies looking at histone modifications at a specific promoter. The other sort of chip assay asks "Is the transcription factor binding, if it's binding, is it associated with a change in histone?" A comparison of induced to uninduced or a comparison of mutant to wild type may show if there is a block at any one of those points of activity.

Finally, there are the more classical types of approaches, endonuclease access assays asking how relaxed the chromatin is at a particular spot.

## **Quality Control, Data Analysis, Statistics and Informatics**

Quality Control (Friday)

Olson reported three major quality control issues that arose in initial research.

Affymetrix recommends that the 5':3' ratio of actin and GAPDH genes be kept within 2.5-fold to gauge the quality of the RNA sample. The rationale is that since transcript amplification occurs at the 3' end of the gene, degraded RNA will have less detectable message from the 5' end of the gene. Olson and Luthi-Carter found that when the 5':3'

fell out of the recommended range, cDNA synthesis or amplification were usually at fault, rather than degraded RNA. When they reprepared samples, the ratio usually returned to normal. When RNA degradation was in fact a problem, a 2.5-fold margin of error was too sloppy within the framework of the HDAG consortium. They now require that the 5'message be at least 75 percent of the 3'message.

Using a test chip to determine 5':3' turned out not to be cost effective. Replacing primary chips that didn't meet the criteria proved less costly. Finally, the number of genes called present from chip to chip turned out to be a sensitive indicator of the quality of the sample. Regardless of whether the experiment looks for a global or a focused transcription dysregulation, the number of genes called present from chip to chip should remain within 10 percent of each other.

A second quality control question asks how to compare one sample to another. The results for each gene must be multiplied by a scaling factor unique to each chip. To normalize a chip, a particular amount of control genes, from *E. coli*, are added to each chip. Affymetrix recommends that scaling factors from chip to chip should not vary by more than 4-fold. Exceeding this threshold will produce false positive calls and false negative calls. The HDAG has been trying to keep the scaling factor within 2-fold because data quality is clearly better than 4-fold.

Finally, Olson suggested that experiments should be done in batches, with all enzyme reagents prepared at one time, and the whole experiment conducted over a period of a week or less. He suggested that over time, analysis software may compensate for batch-to-batch inconsistency, but for now recommends conducting all time courses on the same day or whole experiments within a week. Lawrence Steinman, of Stanford University, wondered if the relative abundance of transcripts affected this aspect of quality control, noting that the lower the abundance, the greater the variability. Olson agreed, adding that many of the most interesting genes lie in the low abundance end of the spectrum.

## The "Seattle Method"

Microarray experiments produce a continuous data set, though criteria are imposed to separate reliable differences from the noise. The development of the informatics technology has not kept up with the development of microarray technology, but progress in that realm is occurring. Several groups have developed analysis algorithms for microarray data. The HDAG has so far relied on Affymetrix analysis software, which has produced good qualitative results.

If Affymetrix software indicates that the expression of a gene is increased, there is a better than 90% chance that a Northern blot will confirm those results. The two analyses may not yield the same fold difference, however. For example, Affymetrix may show a gene's expression increased by 2.5-fold while the Northern shows a 4-fold or 1.3-fold increase. But the number of confirmatory Northern blots remains limited, and one of Olson's goals for the meeting was to convince participants to run 100 to 200 confirmatory Northern blots to improve our ability to assess analysis programs.

The Affymetrix algorithm combines measurements of several parameters to come up with a call: increased, decreased or no change. Each gene is represented by 13-20 probe pairs. Each probe pair actually consists of two sets of oligonucleotides, a perfect-match set of a million to ten million copies complementary to a portion of the gene, and a mismatch set containing a single base-pair mismatch. The analysis software determines if the change in experimental sample is sufficiently different from the control sample both in magnitude and by ratio to determine whether the probe pair can confidently be called increased or decreased. The software can produce a call for each probe pair.

Comparing each component of the Affymetrix algorithm to the overall gene call showed a good correlation between the fraction of probe pairs that go up or down and the respective call. Other parameters carry less weight.

Because the individual probe pair calls can be analyzed independent of the overall call, Olson found that using probe pair cutoffs correlate very well with the Affymetrix Increase or Decrease calls. A cutoff is defined as the difference between the percentage of probe pairs that increase and the percentage that decrease. For example, if 80 percent of probe pairs decrease and 10 percent increase, the net change would be a 70 percent decrease. A cutoff of 25 percent requires that in a gene called increased, the percentage of probe pairs that increased must exceed the percentage that decreased by at least 25 percent.

Probe pair analysis yields some advantages. The data set becomes continuous, allowing one to adjust the stringency of analysis, asking, for example, for 30 percent of probe pairs to be increased or decreased, and compare those results with a stringency of 20 percent. Probe pair analysis also makes it easier to compare experiments. Using the Affymetrix analysis, it becomes complicated to compare experiments with two samples to those with four samples. Probe pair analysis is independent of the number of replicates. More replicates will reduce noise, but the experiments remain comparable.

## Is Northern Analysis the Gold Standard?

Before presenting the preliminary data, Olson opened the floor to a discussion of confirmation analyses. Some researchers asked whether real-time PCR (rtPCR) a better standard for confirmation of microarray calls, since rtPCR could allow analysis of up to ten times as many genes per mouse compared to Northern analysis. rtPCR follows the amplification of product very closely, providing a result that is more quantitative than normal PCR. Despite the obvious advantages, Jang-Ho Cha noted, the introduction of potential artifacts by amplifying samples will keep rtPCR from supplanting Northern blots as a gold standard. Ruth Luthi-Carter raised the point that in-situ hybridization histochemistry might also be considered as an alternative to northern hybridization, because it requires less material and offers the opportunity to assess the spatial and cellular resolution of a change in mRNA expression. Participants agreed on the need to be able to separate signal from noise in microarray analysis.

## Wong Analysis

Wing Wong explained that his aim is to analyze several arrays at the same time instead of analyzing a single array and then to compare the expression measurement for each gene. The idea is to measure as 20-30 arrays, all of the same chip type and the same tissue type and assess the results in each probe set to reduce spurious data.

When you have two probe pairs interrogating the same gene, one probe pair may behave very well, the other may not respond at all. Probe pair-to-probe pair differences may be as large as or larger than the differences between arrays. In the current method of analysis, you take the differences between the perfect and the mismatch and average across the probe pairs in the set and you hope that by averaging, you average away the discrepancies among the probe pairs.

Although the probe pairs show very different behavior, comparing the behavior of a probe pair across many arrays yields very predictable characteristics. Some probe pairs are consistently high, showing a higher response than adjacent probe pairs. The reliability varies with the overall expression level of the gene in question. This analysis can be automated in a software application which Wong can share with other investigators.

## GeneSpring (Saturday)

Barrett Eynon and Ilya Kupersmidt, of Silicon Genetics, demonstrated the exploratory capabilities of GeneSpring, the company's software for microarray data mining and visualization. The Silicon Genetics web site, <http://www.sigenetics.com/> includes an on-line demonstration. They had loaded DRPLA cerebellum, R6/2 cerebellum, cortex and striatum, Aronin striatum and YAC striatum data for the demonstration. GeneSpring can represent data in a number of ways, including Venn diagrams that can show overlaps of gene behavior between experiments and tree diagrams, or clustering analysis, that reveal relationships between genes on a criterion chosen by the researcher. Clustering analysis might reveal families of genes whose expression goes up or down together within a single experiment. That might give an indication that a particular group of genes is similarly regulated, allowing questions such as "Are they modulated by the same transcription factor," or "Are they on nearby parts of the same chromosome and corrugated."

Eynon pointed out that a researcher must try several statistical analyses. There is no single statistical method that's agreed upon by the research community. Ultimately, the qualification of these results has to come from the biology. Statistics cannot solve the problem.

Donald Bergstrom further explained how he has used clustering analysis in studying myoD. Bergstrom and others working with myoD have 30 years of traditional analysis data to work from, so they know what they should be seeing from clustering analysis of microarray data. He predicted that the HD data will become easier to work with after a few important genes, such as enkephalin, are studied. In myogenesis, the genes that researchers expect to increase in expression are called increased by Affymetrix software. In general, the genes that should be changing in skeletal muscle—skeletal muscle structural genes, troponins, myosin heavy chains—behave on the arrays as one would expect them to behave. Steve Tapscott asked Bergstrom if clustering analysis would be better done on a small set of genes. Bergstrom agreed, pointing out that the GeneSpring demonstration grouped data from 15 experiments studying 11,000 genes. In a real experiment, some statistical and biological filtering would go on before feeding the data to GeneSpring.

Steve Tapscott noted earlier in the meeting that cluster analysis programs can group genes and allow researchers to ask about particular perturbations in a pathway. In particular, Tapscott's group blocked an arm of the MAP kinase pathway. They had a good molecular reason to expect that this would affect a particular aspect of a developmental program, and indeed saw that expression of a particular cluster of genes was abolished. With *in vitro* systems that can be tightly controlled and induced, at least, it is possible to develop informative clusters that could then be used

to test hypotheses.

#### Other Data Analysis and informatics efforts

Chun Cheng explained the statistical work a group at FHCRC is examining. The goal is to be able to predict, among all the positive calls, what percentage are truly increased gene expression that can be verified by other methods? On the other hand, of the negative calls, how many are true negatives that can be verified? In order to determine how many time points and how many replicates are needed for an experiment, Cheng and her mentor Lue Ping Zhao, plan to use a comparable data set to mimic the variance in the data anticipated in the proposed experiment. Using that comparable data set as a starting point, they will generate a simulation of the data set, then mimic experiments, varying the number of replicates, the time points and time intervals. Out of the simulation they hope to calculate a predicted true and false discovery rate.

Nathan Goodman, of 3<sup>rd</sup> Millennium, is working with the HDAG to identify data handling needs. He suggested that one way to approach the verification of data would be to run the entire HDAG data set through each analysis method. That can be done automatically, allowing the group to then look for concordance between analysis methods. This could help the group reach a consensus or informed nonconsensus about the various analytical methods.

Goodman's aim is to develop a system allowing each researcher to submit data electronically and have it automatically run through all of the chosen analysis methods.

#### Laser Capture

Sarah Augood, of Massachusetts General Hospital, explained a new method of selecting particular cells from tissue samples. The process involves using a laser on an inverted microscope to melt a plastic cap onto particular cells of a thin tissue section. When the cap is pulled off of the section, the cells come with it. Augood uses a Pixel II made by Arcturus. The company web site <<http://www.arctur.com/>> has complete information on the technology.

The technology has both advantages and disadvantages, Augood explained. It allows the harvesting of a homogeneous cell population from tissue sections that include a heterogeneous cell population. For example, a researcher can pick out particular interneurons of interest from sections of the striatum.

The major limitation is in the extremely small samples that result. Because of that, Augood cannot apply tissues collected this way to microarrays without amplifying the genes of interest. Instead, Augood uses radioisotope labels and cDNA marker arrays. The isotope labeling appears to be much more sensitive than fluorescence-based labeling.

Because the polymer, ethylene vinyl acetate, must melt onto the cells, tissue sections must be thin and perfectly flat on the glass slide so the polymer cap can contact the cells directly.

Jim Olson asked if it might be possible to use laser capture to dissect out so-called "sick" neurons, which stain with toluidine blue. Augood said that as long as a stain is compatible with ethanol fixative, this should be possible.

The minimum number of cells required for cDNA microarrays appears to be 500 to 1,000 for the smaller commercially available cDNA microarrays. To characterize the system, Augood's lab laser dissected 40,000 neurons for a completely unamplified cDNA microarray.

To compare laser capture to homogenization, Augood laser dissected dopamine neurons from substantia nigra and profiled those using Research Genetics cDNA microarrays. They then hand dissected the pigmented area, which should contain primarily just the dopamine neurons, from the same tissue block to the same tissue block and profiled those. Comparing a thousand mRNAs in terms of their relative abundance, they got a correlation coefficient in the range of .99.

However, it was pointed out that the section thickness is dramatically different for laser capture and toluidine blue staining.

#### Therapeutics

Robert Hughes and Carl Johnson, of HDF, discussed some of the issues of taking the HDAG data from experiments to therapeutics. Hughes asked how the information from arrays might lead us to drug targets or to target pathways. He admitted to having more questions than answers.

Before microarrays came along, a lot of effort went into fleshing out the basic biology of cellular dysfunction associated with the expression of expanded polyglutamine. The subtext was the discovery of potential drug targets. The consensus pathways that came out of those pre-chip experiments included chaperones, caspases, the proteosomal ubiquitin pathway and transcription factors. Given that those were implicated by previous experiments,

how much confirmation of dysregulation in those pathways have we been able to see in the chip experiments?

A number of transcription factors that have been shown to physically interact with expanded huntingtin. Are the genes that are known to be regulated by those factors are showing up in these chip experiments? Perhaps there's a way to figure that out in a systematic fashion.

Carl Johnson mentioned to Hughes that a p53 knockout mouse is available, and some experiments suggest that p53 dysfunction is associated with expanded huntingtin expression. So it might be possible, Hughes said, to find similarities between HD mouse models and the p53 knockout mouse.

Hughes also mentioned Olson's work with retinoic acid response elements. Among the genes decreased in the R6/2 mice, Olson explained, several have retinoic acid response elements in their promoters. The question arose whether treatment of the mice with retinoic acid could restore the expression of that particular subset of genes. Whether or not it provided benefit to HD mice would depend on whether or not the retinoic acid-responsive changes were pathogenic.

In terms of thinking about drug targets, Hughes continued, perhaps the most crucial question to ask about array data—even after noise is filtered out—is which of the real gene expression changes are compensatory and which are pathogenic in nature. Answering this question is going to require straightforward classical experiments.

Carl Johnson brings to HDF a background in using animal models to validate and preselect targets for development of human therapeutics. The goal is to find targets using chemical screens.

HDF is pushing a comparable approach. For HD, in fact, the availability of screens is impressive compared to other human diseases.

The only molecular target available is the expanded huntingtin protein. At a mechanistic level, one can look at aggregation or at formation of toxic protein conformations. In the next step up, a whole series of mammalian cell models exists, again looking potentially at anatomical aspects of the aggregation of huntingtin protein or possibly the toxicity of the protein to the cell. At the level of model organisms, researchers have yeast, in which they can ask about toxicity and aggregation, and animal models include *C. elegans* and *Drosophila*, within which cell death and cell dysfunction can be studied.

This collection of models can act as a high-throughput (approximately 50,000 compounds a month) or at least moderate-throughput filter to decide which compounds should be tested in mouse models. A compound found effective in all of these models stands a good chance of being effective in the mouse. Toxicity, side effects, pharmacokinetic and pharmacodynamic issues will have to be dealt with separately.

One of the hopes for the arrays is that looking at gene expression changes could contribute to identifying potential molecular targets other than huntingtin. A good molecular target should appear in a number of different analyses in addition to the arrays, interaction studies, yeast two-hybrid studies, and suppressor genetics in the model systems. Arrays have the potential in the near term to identify some candidates for validation and deselection studies. But the arrays are really quite a way away from identifying a target.

Where the arrays might make an equally important contribution is in generating better, faster and cheaper readouts for the mouse models. The readouts in the mouse models are all relatively slow and behavior-based. It would be nice, Johnson said, to have a more quantitative readout, and perhaps gene chips or some sort of condensed gene chips using selected genes could provide that.

## Biomarkers

Allan Tobin pointed out that it may be possible to develop biomarkers based on the gene chips. Enkephalin might be a good biomarker, and it has now been validated by the gene chips. Perhaps enkephalin or another protein, could be tracked noninvasively using PET.

Johnson suggested that a biomarker in peripheral tissues would help track the progression of the disease. That would certainly give a better readout in terms of whether a chemical intervention was disrupting the disease process.

Participants discussed the possibility of finding a biomarker that could be tracked in blood samples, including human samples. However, uniform handling of blood samples can be crucial, and proper collection and handling in the field may prove difficult. Another possible peripheral target for a biomarker would be the retina.

Johnson added that biomarkers will come up in a future HDF workshop.

## Data Confirmation

Data confirmation came up as a point of discussion several times during the meeting. Being able to confirm the data coming from the microarray experiments is crucial to believing the raw data, crucial to choosing a data analysis software package, or several, crucial to publication of the data and crucial to putting all of the research to practical use in searching for therapies.

Allan Tobin suggested that instead of starting with 11,000 genes from a dozen experiments, the HDAG should test a “constructed” data set, one where researchers know what to expect.

Jim Olson began the meeting emphasizing the need for each group to do a number of Northern blots, and he began the closing discussion with the same suggestion. Cell line experiments produce the largest amount of material, Olson pointed out, and are therefore a logical place to start with Northern blots. He suggested that each cell-line research group run 25-30 Northern blots and each mouse-model research group run 5-10 Northern blots. Phosphor imaging will be important so the data can be quantified. In addition, each group working on a mouse model should attempt ten Northern blots. Steve Tapscott recommended that confirmatory Northern blots be run on the same RNA samples as were used in the microarray experiment. With Northern blots in hand, each analysis algorithm can be tested.

The HDAG will be able to place confirmatory data on a web site. With confirmatory data public, the HDAG data set will become an important tool for any researchers working on microarray analysis algorithms.

## References

1. Luthi-Carter R, Strand A, Peters NL, Solano SM, Hollingsworth ZR, Menon AS, Frey AS, Spektor BS, Penney EB, Schilling G, Ross CA, Borchelt DR, Tapscott SJ, Young AB, Cha JH, Olson JM  
Decreased expression of striatal signaling genes in a mouse model of Huntington's disease.  
*Hum Mol Genet.* 2000 May 22;9(9):1259-71.
2. Hodgson *et al.*  
A YAC mouse model for Huntington's Disease with full-length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration.  
*Neuron* 23:181-192. 1999
3. Ehrlich *et al.*  
*Experimental Neurology* 2001
4. Varani *et al.*  
*FASEB Journal* 2001