Abt Associates Inc.

NDPA Program Outcome Evaluation Design

Final

August 13, 2007

Prepared for
G. Stephane Philogene, PhD

Prepared by
Luba Katz, PhD
Junius Gonzales, MD MBA
Margaret Gwaltney, MBA

Contents

1. Introduction

The NIH Director's Pioneer Award (NDPA) program was launched in FY 2004 as part of the Roadmap Initiative *Research Teams of the Future*.[1] The stated goal of the program is to provide support "for scientists of exceptional creativity who propose pioneering approaches to major challenges in biomedical and behavioral research." [2] Since the launch of the program, three rounds of awards have been completed that resulted in the selection of 35 individuals (called Pioneers in this report). Each award is $500,000 in direct costs per year for the duration of five years.[3] The announcement of the fourth cohort of Pioneers is expected in September of 2007.

The award process is highly competitive, with a small number of scientists selected each year from hundreds of applicants (Table 1).[4] Each application undergoes a review by the NIH liaisons and external evaluators; approximately 20 finalists are invited for an interview, of which a dozen or so Pioneers are selected.

Fiscal year	Number of nominations	Number of applications	Number of awards
2004	1331	245	9
2005	833	283	13
2006	569	406	13

Table 1. Selection of Pioneers by Year

The NIH plans to conduct an Outcome Evaluation (OE) to assess the impact of the program and to ascertain if it is accomplishing its goal of supporting exceptionally creative scientists. We are conducting an exploratory phase for the evaluation—the Feasibility Study (FS)—to identify the appropriate study questions for the evaluation, to explore potential research methods and approaches, and to ultimately develop the optimal OE design. In this document, we propose an OE design.

The OE design is based on extensive background research. We conducted in-depth interviews and focus groups with a total of 27 individuals that included staff at the NIH, external evaluators, and Pioneers. From these discussions, we learned about key questions and data needs and garnered some general suggestions on the methods, the timeline, and the comparison groups. In parallel, we reviewed academic and evaluation literature, and spoke with several authors and program managers. These sources were used to identify suitable methodologies that have been validated by other researchers and to explore how these methodologies can be adapted for the NDPA program evaluation.

In addition, the FS was supported by three external advisors, selected to represent distinct areas of expertise that are key to the study. These advisors included a senior professor who studies creativity in science; a former director of a Federal program for innovation in science; and, a scientist who has received several awards for innovation. The evaluation framework has been presented to the NDPA Advisory Committee and changes/comments by the Committee are incorporated in this document.

[1] http://nihroadmap.nih.gov/pioneer/

[2] *ibid*

[3] *ibid*

[4] Data from the process evaluation provided by Bhavya Lal, STPI

Five study domains for the OE are identified. The first four domains represent traditional dimensions of a research program evaluation:

(1) Adequacy of selection process;
(2) Impact on the NIH;
(3) Outcomes for Pioneers, their students, and institutions; and
(4) Broad scientific outcomes and implications for human and public health.

In addition, the NIH is interested in exploring a more general question related to the career trajectories of creative people in biomedical research and outcomes of their creative ideas. The large number of talented but unsuccessful NDPA applicants (Table 1) represents a unique population of researchers that can contribute to exploring these questions in a systematic manner. Thus, a fifth domain is included in the OE framework:

(5) Trajectories of creative researchers and their ideas.

Please note the ongoing Process Evaluation[5] has explored many aspects of the selection process (domain 1). To avoid duplication with the Process Evaluation, the proposed OE focuses exclusively on the scientific aspects of the NDPA proposal selection, which are not being addressed by the Process Evaluation.[6]

Our view of the NDPA program is presented in the Logic Model (Figure 1). The logic model provides a framework for the program, highlighting how it is expected to work, what activities need to be implemented and in what order, and what results are desired; once the logic model is established, it guides the selection of performance indicators. Because research activities of unfunded applicants are not an explicit goal of the program, but rather its unanticipated consequence, we present this element of the Model in the dotted format. We expect that the types of outputs/outcomes of research efforts by unfunded applicants will be similar to those of the Pioneers.

The following sections present the OE study questions, methodologies to address them, some preliminary data, comparison groups, and implementation timeline. Where appropriate, this report discusses methods limitations and suggested ways to mitigate those limitations. Several background documents formed the foundation of this evaluation design and these are referenced in the text.

[5] Bhavya Lal, STPI

[6] The process evaluation is focused on the procedural aspects of the NDPA application process

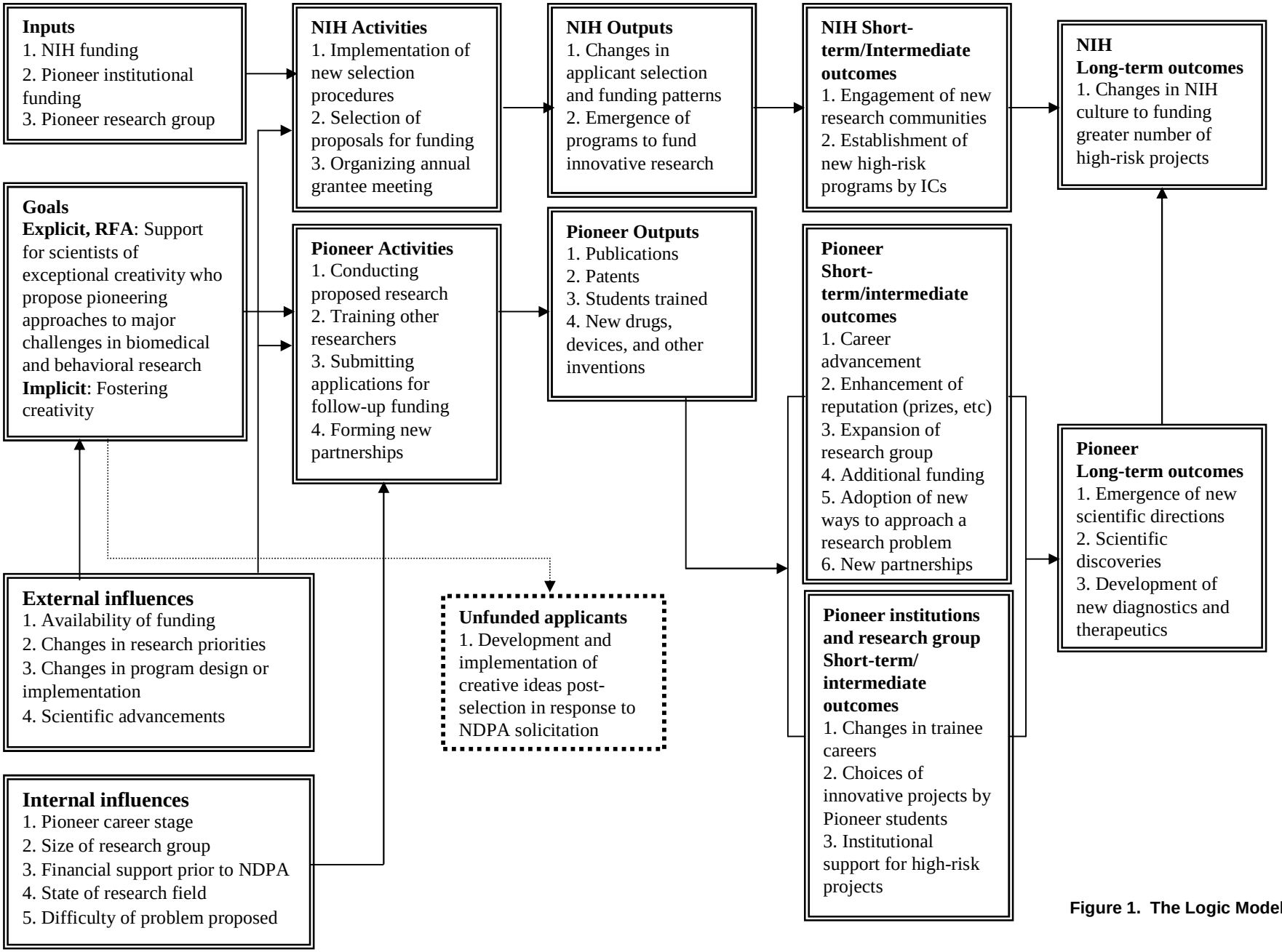

Figure 1. The Logic Model

2. Approach

2.1. Study questions

One major challenge in developing a suitable OE design is that considerable time—10 years or more—needs to pass before meaningful data can be collected on some program accomplishments, in particular on scientific outcomes. On the other hand, some valuable observations about the program can and should be documented early on, as these data can be key in determining whether and how the NDPA program (or its successor) should be modified or continued in future years. For parsimony, program outcomes are categorized into short-term (ST), intermediate (IM), and long-term (LT)—in 1-3, 4-7, and 8 or more years after the award was made. Each year, a new class of Pioneers joins the program (called cohort in this report); therefore, the timing when these various outcomes can begin to be captured will vary by cohort (Table 2).

Fiscal Year	Short-term outcomes	Intermediate outcomes	Long-term outcomes
2004	Cohort 1 (year 1)		
2005	Cohort 2 (year 1)		
2006	Cohort 3 (year 1)		
2007	Cohort 4 (year 1)		
2008	Cohort 5 (year 1)	Cohort 1 (year 4)	
2009		Cohort 2 (year 4)	
2010		Cohort 3 (year 4)	
2011		Cohort 4 (year 4)	
2012		Cohort 5 (year 4)	Cohort 1 (year 8)
2013			Cohort 2 (year 8)
2014			Cohort 3 (year 8)
2015			Cohort 4 (year 8)
2016			Cohort 5 (year 8)

Table 2. Pioneer Cohorts

The OE design will ascertain the program outcomes in five domains. The next few paragraphs present a series of study questions to capture outcomes in each of these domains. Also presented in brackets is the ideal timing for when these data should be collected.

Domain 1: Adequacy of selection process and impact on the NIH

Questions to assess short-term outcomes (can be measured at any time)

- Has the NDPA selection process resulted in the identification of exceptionally creative individuals and new research directions?
- In what way are the projects selected for funding different from other Pioneer research?
- How are the funded projects different from what is typically funded by the NIH?

Domain 2: Impact on the NIH

Questions to assess intermediate and long-term outcomes (4-7 years and 8 or more years after the NDPA program launch)

- As a result of the NDPA program, has there been a change at the NIH towards funding high-risk research?

Domain 3: Outcomes for Pioneers and their students

Questions to assess short-term and intermediate outcomes (1-3 and 4-7 years)

- What is the effect of the award on Pioneers and members of their research groups?
- Are Pioneers conducting research they would not have been able to do without the award?
- Are they able to obtain funding for work begun under the NDPA when the award ends?

Domain 4: Scientific outcomes and applications to human health

Questions to assess intermediate and long-term outcomes and impacts (4-7 years and 8 or more years)

- Have they formulated new ideas, developed new methodologies and instruments, asked new questions? (Ideally, this should be tracked over time)
- What are career tracks and research fates of exceptionally creative individuals? (8 or more years)
- How is Pioneer work viewed by the research community? (8 or more years)
- What are applications of the Pioneer work to the diagnosis and treatment of disease? (8 or more years)

Domain 5: Career trajectories of creative individuals and the outcomes of their ideas

Questions to explore trajectories of unfunded applicants and their proposed ideas (every 2-3 years)

- Did unfunded applicants choose to pursue the idea proposed in their NDPA application?
- If they pursued the idea, how was the work funded? What were the outcomes and outputs of the work and its impacts? Did the NDPA ideas result in paradigm-shifting discoveries?
- If the applicants decided to abandon the idea, why did they do it?
- How did the applicants' careers develop over the course of the years?

2.2. Comparison groups

Use of a comparison group strengthens the ability to assess the relative impact of the NDPA program. Three main criteria emerge in searching for possible comparison groups:

- Funding should be provided for approximately the same number of years as NDPA (5 years)
- Funding should be similar in monetary scope to NDPA ($500,000 a year)
- The program should have a similar goal of funding creative individuals rather than projects

After reviewing a large number of grant programs, the conclusion reached was that none completely satisfies all these criteria, but a few may satisfy one of the criteria.[7] For example, R01 funding is provided for five years.[8] Using this group has several advantages: the sheer number of grantees should allow matching by research topic and by year, and data on these individuals should be available from the NIH databases. Accomplishments of R01 grantees can be used for comparative analyses of scientific impact. In an interesting idea suggested by a Pioneer, outcomes of research funded by an R01 grant might be compared to these of NDPA *within the same laboratory*. The expectation of this comparison is that NDPA work will lag behind the R01 work in terms of publications and other tangible research products, but that progress when made will be a paradigm shift rather than an incremental advance.

[7] Environmental Scan, Abt Associates, 2007

[8] 78 percent of unfunded applicants have NIH funding (Process Evaluation of 2004 cohort), presumably many of them R01 grants

The Howard Hughes Medical Institute (HHMI) has a stated goal of "funding people rather than projects." Thus this group is a suitable comparison when evaluating career and research outcomes. Access to data may be a potential limitation for this group, however, as it will require HHMI cooperation.

Instead of using a single comparison group for the entire OE, a mixed comparison group approach appears best—turning to the most suitable population depending on the nature of the study question. Furthermore, many important outcomes can be documented without comparison—for example, the impact of the NDPA program on the NIH.

2.3. Timeline and summary of approaches

Table 3 is an illustration of the proposed timeline and methods. In-depth discussion of each method is included in the next chapter.

We propose a 3-phase evaluation plan: Phase I, in 2008-2009; Phase II, in 2011-2012; and Phase III, in 2014-2015. The timing of the phases corresponds to the achievement of short-term, intermediate, and long-term outcomes, respectively, for the entire population of awardees (Table 3). Career paths of creative individuals and outcomes of their ideas proposed in the NDPA application can be studied in parallel by means of a web-based survey. Please note that the survey of the first two cohorts of unfunded applicants, conducted for this Feasibility Study, indicated that many of them might be unwilling to respond to multiple requests for information related to their NDPA application.[9] Thus, for the remaining cohorts of unfunded applicants, we recommend that the survey be combined with the data collection activities of the Process Evaluation, where a survey of unfunded applicants is already part of the design.

2007	2008	2009	2010	2011	2012	2013	2014	2015
	PHASE I (ST)			**PHASE II (IM)**			**PHASE III (LT)**	
	Pioneer interviews/visits			Pioneer interviews/visits			Pioneer interviews/visits	
	Review of annual reports			Review of annual reports			Review of annual reports	
	Expert Panel 1			Bibliometric analyses			Expert Panel 2	
	Analysis of UAS data (c3-5)			Interviews with NIH staff			Pioneer trainee survey	
				Analysis of UAS data (c1-5)			Bibliometric analyses	
							Interviews with NIH staff	
							Analysis of UAS data (c1-5)	

Table 3. Timeline

ST = short-term; IM = intermediate; LT = long-term outcomes
C = cohort (e.g., c1= cohort 1)
UAS = Unfunded Applicant Survey

[9] Unfunded Applicant Survey Report. Abt Associates, 2007

3. Methodologies

Creativity in science is difficult to study, because of the complexity of the subject and the lack of rigorous empirical data, measures, or methods. This OE is, in itself, a pioneering opportunity to contribute to a scant literature. To enhance the validity and interpretive power of data collected, we propose a combination of approaches, some of which are widely used in evaluation practice and others that are more novel and innovative, but less tested.

3.1. Pioneer interviews and site visits

In-depth interviews with Pioneers represent a unique source of information. Only Pioneers can describe what the award has allowed them to do, whether it has changed their scientific thinking and research approaches, how they balance the risks of taking their work in new directions with competing academic career demands for productivity, and how the NDPA funded work is different from other work that they do. In-person interviews are typically best, but as the number of Pioneers is becoming relatively large (35 as of the time of writing with an additional 20-25 awards projected), telephone interviews may be more practical and cost-efficient.

Pioneer interviews should be augmented with site visits; tours of the laboratory space, conversations with students, postdoctoral fellows, and department chairs are valuable additional sources of data that would be difficult to gather from individual interviews. During our literature review, we found abundant evidence that creativity flourishes in a communal setting[10] and the OE of the NDPA program is a unique opportunity to test this observation on an academic group. Professor Amabile of Harvard University and her colleagues have developed and validated an instrument called KEYS designed to measure organizational creativity.[11] KEYS is a questionnaire that probes all the dimensions of the work environment that are thought to be related to creativity.[12] This instrument was not designed for assessing organizational creativity of a scientific group, but can be adapted for this purpose. The site visit should be timed with a research group meeting, as these meetings are often a telling manifestation of the laboratory environment and organizational style.

We anticipate that visiting all labs may be too costly and instead recommend selecting a subset of 5-6 labs during each evaluation phase; selection criteria will be developed in order to achieve heterogeneity across key parameters of interest (examples of parameters might include research area, individual Pioneer characteristics, NIH funding history).

3.2. Expert panels

An independent expert panel is a powerful method to evaluate both the NDPA selection process and Pioneer research accomplishments. This methodology is routinely used in program evaluation. Grant and Allen, for

[10] For example, Kevin Dunbar. How scientists think: On-line creativity and conceptual change in science. Chapter in Conceptual structures and processes: Emergence, discovery, and change. American Psychological Association Press.

[11] Amabile T, Conti R, Coon H, Lazenby J, Herron M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*. 39(5), p1154-1184.

[12] Details about the instrument and about organizational creativity can be found in the Environmental Scan (Abt Associates).

example, have convened expert panels for the evaluation of the Wellcome Trust Showcase awards (awards for innovative research).[13] The authors assessed the innovativeness of the funded proposals by engaging an expert panel that blindly scored 20 Showcase awards and 20 standard grants previously funded by the Trust. The panel was asked to assess each research summary using criteria such as: "adventurous, innovative, novel, risky, and speculative." The summaries were numerically rated for each category. Grant and Allen collected 156 Showcase responses and 154 standard grant evaluations by 24 expert panelists and found that the Showcase awards were ranked significantly higher in all categories.

Expert Panel 1. We recommend that the OE use the approach tested by Grant and Allen to assess the adequacy of the selection process. The panel composed of 5-10 individuals will be requested to blindly score funded NDPA and R01 application summaries (opinions from more than one expert should be obtained on each application, to enhance the validity of scoring). Because application formats for these grants differ, the summaries will need to be revised, so that the type of application cannot be discerned from its style. We tested the feasibility of this approach on a randomly selected R01 and NDPA application and found that the two can be made indistinguishable though minor changes (Chapter 4 and Appendix). Ideally, the applicants should review the revised summaries to ensure that their content has not been inadvertently altered. The panelists would be asked to rate the applications on innovation and creativity and to indicate whether the application could be funded through the standard NIH programs. Expert Panel 1 can be convened any time after the award is made.

Expert Panel 2. Pioneer research work can also be evaluated by an expert panel 10 years or so after the award to examine long-term outcomes. In this case, Pioneers would be requested to identify papers that represent their most creative work and these would be subjected to the panel review. We propose using the typology of creativity developed by Heinze and Shapira to guide the panelists.[14] The authors suggested five types of scientific creativity, each accompanied by a famous historical example: (1) formulation of new ideas that open up a new cognitive frame or brings theoretical claims to a new level of sophistication (Einstein's Theory of specific relativity); (2) discovery of new empirical phenomena that stimulated new theories (Darwin's Theory of evolution); (3) development of a new methodology, by means of which theoretical problems can be empirically tested (Spearman's Theory of mental abilities); (4) invention of novel instruments that open up new search perspectives and research domains (invention of tunneling microscopy); and (5) new synthesis of formerly dispersed existing ideas into general theoretical laws enabling analyses of diverse phenomena within a common cognitive frame (Bertalanffy's General systems theory). Heinze and Shapira successfully used this framework to solicit community nominations for creative scientists of the past decade. Based on the input from the Advisory Committee, we will replace the historical examples of Heinze and Shapira with more recent famous scientific accomplishments drawn from the biological sciences.[15]

While Pioneers agree to participate in an evaluation as a condition to receiving the award, we were concerned that they would object to this level of scrutiny. This topic was discussed with them during the focus group. Pioneers told us that they would be reluctant to share unpublished work, but otherwise they foresaw no problem having their work or proposals evaluated.

---

[13] Grant J and Allen L (1999). Evaluating high Risk Research: an assessment of the Wellcome Trust's Sir Henry Wellcome Commemorative Awards for Innovative Research. *Research Evaluation* 8(3):201-204.

[14] Heinze-Shapira framework is only suitable to retrospective assessment and this is not suitable for Panel 1.

[15] For example: formulation of new ideas (apoptosis); discovery of new empirical phenomena (splicing); development of new methodology (DNA sequencing); invention of novel instruments (gene chips); new synthesis of dispersed ideas (systems biology).

3.2.a. Method limitations

Recruiting an expert panel may be difficult if the proposed workload is significant. On the other hand, external evaluators reported that reviewing the NDPA proposals was a great experience and they would continue to participate in the process; we hope that this level of enthusiasm is shared by other experts. If each panelist is asked to evaluate a small number of NDPA applications/R01 control summaries (5 each, for example) or one or two research papers, is given sufficient time to do it, and is not required to travel—participation will probably not be difficult to secure. Monetary incentives could also be offered to boost participation. Advance work with the relevant offices would be carried out to determine eligibility for the panel in order to avoid including experts on the panel who have previously been exposed to Pioneers' work. If a sufficient number of such experts in the United States cannot be recruited, international scientists can be approached.

3.3. Content analysis of Pioneer annual progress reports

Request for NDPA applications stipulates that "awardees will be required to submit the Non-Competing Grant Progress Report (PHS 2590) annually."[16] The form PHS 2590[17] includes request for information in all areas of progress including on key personnel, publications and research progress, and expenses and if completed diligently can be a rich source of data. In the course of the Feasibility Study, we analyzed two annual reports—for Pioneers from cohorts 1 and 2. The results of our analysis are presented in Chapter 5. We also spoke with the Program Officer to inquire whether Pioneers as a group typically file reports and how comprehensive they are. The Program Officer reported that he looked at about 20 progress reports and he considers them all adequate.[18]

3.4. Unfunded applicant survey

Given that the number of awards is so small, it is inevitable that many excellent applicants will not be selected. We propose administering a short web-based survey to all unsuccessful applicants who have submitted a full proposal. We have tested this approach on 500 individuals from the first two cohorts. Results of the survey and our recommendations for future implementation are discussed in Chapter 4. We are also in the process of designing the database that will include all information available on the applicants. The survey and the development of the tracking database were launched in advance of the Outcome Evaluation to avoid further delays in contacting unfunded applicants.

Survey results could be supplemented with extant data that are available on unfunded applicants, such as the content of their web pages or the application/review materials contained in the IMPAC II database. We have explored what type of data could be gathered using these sources (Chapter 4).[19] We found that this activity was very labor intensive and that some of the key questions could be difficult to answer. Thus, we do not recommend this approach for the OE.

[16] http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-07-005.html

[17] http://grants1.nih.gov/grants/funding/2590/phs2590.pdf

[18] Personal communication, April 27, 2007

[19] Case Studies Report. Abt Associates.

3.4.a. Method limitations

Achieving a high response rate is a concern for the unfunded applicant survey.[20] Unfunded applicants have put significant effort into competing for the award, and were probably disappointed not to be selected. They have been recently approached to complete a survey for the process evaluation, which may have exhausted their enthusiasm for providing feedback on the program. Therefore, we were concerned that the applicants may have little interest in participating in the evaluation. We took several steps to circumvent this potential limitation: the survey was brief and easy to use; in our letter of invitation, we emphasized the importance of applicant feedback not only for the NDPA program, but also for the NIH as a whole because it comes at a time when the Institutes are considering duplicating the NDPA-like funding mechanism. We sent two follow-up emails to non-respondents and tried to contact them by telephone. Finally, we offered a small ($20) contribution to one of three charities for each completed survey. Despite these efforts, our response rate was 34 percent.

3.5. Bibliometric analyses

Publication counts, ranking of journals, and citation analyses (how frequently a given paper is cited by others), are all standard techniques to measure scientific productivity and influence. From our literature review to prepare the Outcome Evaluation design, we found several less known, but promising measurements. One, called "thematic breadth index," is an indicator that combines the number of journals where the papers were published with their frequency in each journal and can be used as a measure of multidisciplinarity. If a scientist published 10 articles in two journals, for example, the index would be lower than if s/he published 10 articles in five journals. Heinze and Shapira observed that creative scientists showed broader thematic spectrum in their work than the comparison group of peers.[21]

Heinze and Shapira also demonstrated that creative scientists showed greater propensity to build collaborative networks than less creative (but equally productive) controls[22] using Social Network Analysis. Specifically, parameters such as brokerage index,[23] network size and density that describe the network can be calculated (our test of the approach is described in Chapter 4). We propose using these more novel measures in addition to standard bibliometric indicators (publication and citation counts, and journal rankings) as they were explicitly tested on a population of creative researchers.

3.5.a. Method limitations

While unquestionably a useful tool, results of bibliometric analyses may be difficult to interpret in the context of the program. In fact, several NIH staff interviewed suggested that Pioneers who start publishing high profile papers soon after the award was made might have been improperly selected. Publication delays and citation lags are patterns generally expected from the NDPA grantees. Thus, the results of bibliometric

[20] As part of the on-going Process Evaluation, a web survey has been administrated to unfunded applicants and resulted in 60-70 percent response rates depending on the cohort.

[21] Heinze, T., & Shapira, P. (2006). Research creativity. Analyses of unconventional, path-opening solutions in science. *Proceedings of Science & Technology Policy Research, 40th Anniversary Conference, Brighton (UK)*: pp. 11-13.

[22] Heinze T & Bauer G. (2007). Characterizing creative scientists in nano-S&T: Productivity, multidisciplinarity, and network brokerage in a longitudal perspective. *Scientometrics* 70(3), p811-830.

[23] Brokerage index measures the percentage of those peer scientists who are connected to the network only through Pioneer

analyses should not be over-emphasized. In addition, while Social Network Analysis is a measure of interconnectedness and collaboration, it does not necessarily measure research quality or impact.[24]

3.6. Interviews with NIH staff knowledgeable about research programs

Because of its unique funding mechanism, visibility, and prestige, the NDPA program will have ramifications that extend beyond its direct grantee beneficiaries. In fact, in the course of this Feasibility Study, NIH staff and others reported that the program has attracted much interest within NIH, and that several Institutes and Centers are establishing similar programs of their own. Furthermore, Dr. Zerhouni has recently announced the launch of the NIH Director's New Innovator Award program. [25]

To get a more systematic view of what is happening at the NIH, interviews with staff at each Institute and Center will be conducted and those who are in a position to know about high-risk research programs that are being planned or implemented in their part of the agency will be sought out. We understand that these individuals may be somewhat difficult to identify, as there are typically several staff, with different titles, that are involved in these activities within each IC.[26] Therefore, a first step will be interviewing members of the Extramural Policy and Management Committee (EPMC). Each Institute is represented in the Committee, and the representatives are well informed about different funding activities. References to other knowledgeable parties could be obtained from EPMC members if needed.

3.7. Pioneer trainee survey

If the NDPA selection process is successful—and highly creative proposals to initiate new research directions are chosen— doctoral and postdoctoral trainees of Pioneer PIs may find themselves treading in uncharted scientific territory. Does the experience of working on a high risk and potentially high pay-off project produce a lasting, career-changing effect on these junior researchers? Are their subsequent choices of institutions, advisors, and research directions influenced by the experience of working on the NDPA project? Have the experiences changed the way they think about the problem? These and other questions can be addressed by implementing a short survey a few years after Pioneer students began working on the NDPA-funded project. The survey can be administered to Pioneer trainees at any stage of their career, although the results will probably be most telling for postdoctoral fellows who have since become independent researchers because they have the greatest freedom in the choice of research program. Pioneers are likely to have contact information for their former trainees.

[24] National Academy of Sciences, *Facilitating Interdisciplinary Research*, Washington, DC, 2005

[25] http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-07-009.html

[26] Mary C. Dufour, MD, MPH, personal communication

4. Preliminary Data

Some methods and instruments recommended in Chapter 3 have been pre-tested. Results of these exploratory studies are presented below.

4.1. Bibliometric analyses

As discussed in Chapter 3, several measures of multidisciplinarity and influence, such as brokerage index, thematic breadth index, network size and density, appear to be higher in creative researchers than in their less creative colleagues.[27] Some of these indicators could be derived using Social Network Analysis software UCINET.[28] We constructed a social network for a 2004 Pioneer using all publications that he has co-authored from 2005 to present (Figure 2). Publication data were obtained from PubMed; each node (a circle) represents an author.

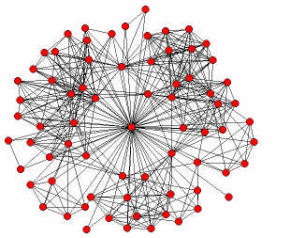

Figure 2. Pioneer network, 2005-2007

The network has the following characteristics: Its size, the total number of authors, is 71; the number of ties, all connections between the authors, is 724; and density, percent of all possible links present, is 14 percent.

4.2. Unfunded applicant survey

A web-based survey instrument was developed using the commercial software package QuestionPro.[29] The web instrument was pretested on four randomly chosen applicants from both cohorts and their comments were incorporated. The survey was administered to 456 individuals from the first two NDPA cohorts. The response rate was 34 percent. Applicants who did elect to complete the survey appear to have taken the task seriously. As a group, they have submitted nearly 1,000 open-ended responses. Most said that they would be willing to complete another survey in 2-3 years. Thus, while the group is small, it may remain stable as a data source over the years.

Results of the survey suggest that most applicants remain committed to the project proposed in the NDPA application and many have already made some progress (or even significant progress) in its implementation. Since many of the applicants are very accomplished researchers who have stable funding, they were able to support the research proposed in the NDPA application by culling the moneys from other grants, institutional funds, or by contributing their time *pro bono.*

Most applicants appear to be making incremental career progress that is typical of a biomedical researcher. In the past 2-3 years, they have published papers, received funding support, expanded their groups, and formed new collaborations. In a few cases, more dramatic changes have taken place—such as loss of the position or a change in research direction.

---

[27] Heinze et al, 2007.

[28] Borgati S, Everett M, Freeman L (2002). *Ucinet 6 for Windows. Software for Social Network Analysis*. Analytic Technologies, Natick.

[29] www.questionpro.com

4.3. Analysis of annual reports

Pioneers are required to submit annual reports documenting progress on the grant. Two of these reports were reviewed—from 2004 and 2005 Pioneers (cohort 1 and cohort 2). Most of the information included is devoted to the description of research accomplishments (and is written in highly technical language). Both Pioneers explained what they proposed to do in their NDPA application and what progress has been made towards each aim. Publications resulting from the project are also included in the report. One of the Pioneers has highlighted his partnerships for each element of the project. He also listed all his current funding (the other Pioneer did not, but she may not have any). Beyond research progress, annual reports contain little information. For example, virtually nothing is included on the personnel involved (one of the Pioneers mentioned that one new key personnel was added). If other annual reports are similar in content, their main use to the evaluator is for reviewing Pioneers' scientific progress.

4.4. Expert Panel 1

In the methodology section, we propose revising the abstracts in the NDPA and R01 proposals for blind evaluation by members of the expert panel. This method has been successfully applied for the evaluation of the Wellcome Trust Showcase awards.[30] In order to ascertain the feasibility of this approach, we reviewed an abstract from an R01[31] and NDPA application.[32] We found that the two abstracts are written in a similar manner and can be made indistinguishable from each other with minor changes. First, the sentence in the Pioneer abstract "In this essay I outline the following proposal for testing this novel idea" was modified to read "Our approach will involve the following." This change was made because the word "essay" is not used to describe an R01 application. The second change involved removing references from the NDPA abstract, which we noticed to be absent from the R01 abstract. The final, truly trivial and probably unnecessary change was in reformatting the list of three items in the NDPA abstract to make it look more similar to the R01 abstract. Even without these changes, the two abstracts are virtually identical in the layout, language, and length, as the reader can judge by reviewing both documents, which we include in the Appendix (the changes that we made are marked in bold red font).

In our opinion, the abstracts contain sufficient information for an expert panel to make an informed judgment as to the novelty/creativity of an idea. In fact, these abstracts are close to 400 words, and are likely to be much longer than the abstracts used in the Wellcome Trust evaluation. The article that described the method used by the Wellcome Trust staff stated that "summaries of proposed research [were] similar to an abstract of a scientific paper;" these summaries typically cannot exceed 200 words.

4.5. Case studies

The purpose of the case studies was to explore the nature of the data that can be collected on the NDPA applicants without contacting them directly.[33] Case studies were completed for two Pioneers and two applicants not selected by the program.

---

[30] Grant J and Allen L (1999). Evaluating high Risk Research: an assessment of the Wellcome Trust's Sir Henry Wellcome Commemorative Awards for Innovative Research. *Research Evaluation* 8(3):201-204.

[31] Data from CRISP

[32] Data from IMPAC II

[33] Case Studies Report, Abt Associates.

We found that many aspects of an applicant's professional situation can be ascertained: their title and affiliation, research interests, publication record, funding received from the NIH, speaking engagements and other public appearances (e.g., mention in a newspaper), awards and prizes, and in some cases composition of their research laboratory or group. It is much more difficult, however, to ascertain whether a specific idea proposed in an unfunded application was pursued and to what end (Pioneer annual reports are a good source for this type of data). In principle, published papers do contain this information, but we found that in-depth technical understanding of the applicant proposal and a general knowledge of the field of inquiry are necessary to be able to make a connection between the proposed scientific idea and the results reported in a publication.

Another disadvantage of case studies is that they are time-consuming. We estimate that, on average, it took a senior scientist with doctoral and postdoctoral training about two days to compile, analyze, and synthesize the information on each applicant.

From this exercise, we concluded that given the size of the unfunded applicant population, this data collection strategy is significantly inferior to other approaches, such as interview or survey, as it is time consuming and may not yield all the information sought.

5. Design Summary

5.1. Ranking of approaches

We understand that time and cost constraints may make it difficult to implement a large array of methods described in this proposal and the decision would have to be made what approaches to eliminate from the OE. To assist the NIH in making this decision, we rank all proposed methods by priority; priorities were assigned as "high" and "low" in accordance with the expected value of data that they would generate. In the next section, we present a matrix that links our proposed methodologies with study questions and indicators.

High priority:

- Pioneer interviews
- Site visits to select Pioneer labs
- Expert panels (both to evaluate Pioneer proposals and their research accomplishments)
- Interviews with EPMC staff, NIH
- Unfunded applicant survey (Process Evaluation) and analysis of survey data (Outcome Evaluation)

Low priority:

- Trainee surveys
- Analysis of annual reports
- Bibliometric analysis

5.2. Summary table

EVALUATION QUESTION	EVALUATION INDICATOR	EVALUATION APPROACH/ DATA SOURCE
Has the NDPA selection process resulted in the identification of exceptionally creative individuals and new research directions?	1. Creativity ranking	1. Blind evaluation of NDPA and R01 application summaries by Expert Panel 1
In what way are the projects selected for funding different from other Pioneer research?	1. Pioneers hired new staff with different background; 2. Pioneers attended conferences in new areas; 3. Pioneers plan to publish in different journals	1. Pioneer interviews 2. NIH staff
How are the funded projects different from what is typically funded by the NIH?	1. Ranking by expert panel	1. Blind assessment of NDPA and R01 application summaries by Expert Panel
As a result of the NDPA program, has there been a change at NIH towards funding high-risk research?	1. Increase in the number of RFAs/PAs in new/emerging disciplines; 2. Changes in review criteria and/or review process	1. Interviews with members of EPMC and other individuals at NIH
What is the effect of the NDPA on Pioneers?	1. Time to promotion/tenure; 2. Number of awards; 3. Other signs of recognition (requests to serve on panels and editorial boards); 3. Additional funding; 4. Thematic breadth index and journal rankings;	1. Pioneer interviews 2. Bibliometric analyses 3. Review of annual reports

EVALUATION QUESTION	EVALUATION INDICATOR	EVALUATION APPROACH/ DATA SOURCE
What is the effect of the NDPA on members of the Pioneer research group?	1. Trainees select more risky research projects 2. Signs of recognition/success (publications, prizes, funding, etc)	1. Survey of Pioneer trainees
Are Pioneers conducting research that they would not be able to do otherwise?	1. Number of unfunded proposals on the NDPA topic 2. Number of rejected manuscripts on the NDPA topic	1. Pioneer interviews
Are they able to obtain funding for work begun under the NDPA when the award ends?	1. Number of grants	1. Pioneer interviews
Have Pioneers formulated new ideas, developed new methodologies and instruments, asked new questions, initiated new research areas?	1. Patents 2. Rating by Expert Panel 3. Bibliometric indicators: thematic breadth index, ranking of journals	1. Pioneer interviews 2. Assessment of Pioneer research portfolio by Expert Panel 2 3. Bibliometric analyses
What are applications of the Pioneer work to the diagnosis and treatment of disease?	1. Patents 2. Publications in medical journals 3. New diagnostic equipment, drugs, etc.	1. Pioneer interviews 2. Bibliometric analyses
How is Pioneer work viewed by the research community?	1. Bibliometric indicators: thematic breadth index, ranking of journals 2. Evidence of recognition (awards, requests to serve on panels, etc). 3. Ranking by Expert Panel	1. Pioneer interviews 2. Bibliometric analyses 3. Expert Panel 2 4. NIH staff
Did unfunded applicants choose to pursue the idea proposed in their NDPA application? If they pursued the idea, how was the work funded? What were the outcomes and outputs of the work and its impacts? Did the NDPA ideas result in paradigm-shifting discoveries? If the applicants decided to abandon the idea, why did they do it? How did the applicants' careers develop over the course of the years?	1. Funding sources 2. Publications 3. Career progression (promotion, tenure, etc) 4. Signs of recognition (requests to serve on boards and panels) 5. Bibliometric indicators	1. Web-based survey administered every 3 years

Table 4. Summary table

Appendix A: Applicant survey

Demographic information:

1A. Name

1B. Institution

1C. Gender

- ❑ Male
- ❑ Female

2. Would you have chosen to submit an application containing your NDPA idea to the NIH if the NDPA program did not exist?

- ❑ Yes
- ❑ No
- ❑ I do not know
- ❑ I cannot recall

3. How would you rate your NDPA idea on risk?

- ❑ Not risky
- ❑ Of medium risk
- ❑ Very risky
- ❑ I do not know
- ❑ I cannot recall

4. How different was the NDPA idea from your other work at the time of application?

- ❑ Not at all
- ❑ Somewhat different
- ❑ Very different
- ❑ I cannot recall

5. In what way was your NDPA idea different from what is typically funded by the NIH?

6. Did you choose to pursue the idea(s) proposed in your NDPA application?

- ❑ Yes
- ❑ No

7. Was the project funded?

- ❑ Yes
- ❑ No

7A. What was the funding source?

- ❑ NIH
- ❑ NSF
- ❑ Private Foundation
- ❑ Other

7B. Do you plan to apply for funding in the next 12 months?

- ❑ Yes
- ❑ No

7C. How were you able to work on the NDPA idea? Please check all that apply:

- ❑ By using institutional funds
- ❑ By using other grants that I have
- ❑ I collaborate with a colleague who provides the funding
- ❑ My work is theoretical and requires minimal funding for supplies and I donate my time
- ❑ Other _____

8. If the project got funded, to what extent did you modify the idea proposed in your NDPA application?

1 - Not at all
2
3
4
5
6
7 - Substantially

9. Has the work proposed in the NDPA application resulted in a publication(s)?

- ❑ No
- ❑ Manuscript in preparation
- ❑ Yes. Please specify the journal(s): _____

10. Has the work proposed in the NDPA application resulted in a presentation at scientific conference(s)?

- ❑ No
- ❑ Yes

10A. How did you get selected to present at the conference?

- ❑ I was invited by the organizers to give a presentation at a conference
- ❑ My presentation was selected through competitive submission

11. To date, how would you evaluate the impact of the project? Please select all that apply

- ❑ The idea has already made some contribution to my research field
- ❑ The idea has already made significant contribution to my research field
- ❑ The idea has already made some contribution to another research field
- ❑ The idea has already made significant contribution to another research field
- ❑ No impact so far
- ❑ I do not know

12. How would you evaluate the potential impact of the project on your research field?

- ❑ Low
- ❑ Medium
- ❑ High
- ❑ I do not know

13. How would you evaluate the potential impact of the project on another research field?

- ❑ Low
- ❑ Medium
- ❑ High
- ❑ I do not know

14. Have other scientists given you any feedback on the impact of your project? Please select all that apply:

- ❑ Other scientists think that the idea has made some contribution to my research field
- ❑ Other scientists think that the idea has made some contribution to another research field
- ❑ Other scientists think that the idea has made significant contribution to another research field
- ❑ I did not get any feedback
- ❑ I cannot recall

15. Why did you choose not to pursue the idea proposed in your NDPA application? Please select all that apply:

- ❑ I applied for grants to fund the project, but was unable to get the funding
- ❑ I did not apply for grants to fund the project
- ❑ I am no longer interested in the idea
- ❑ I do not have time to pursue the idea
- ❑ The work proposed has been done or is being done by someone else
- ❑ I made a decision that the idea was not practical/feasible
- ❑ Other

16. Please indicate which of the following important developments have taken place since your NDPA application? Please check all that apply:

- ❑ I received an award(s).
- ❑ I was promoted

- ❑ I published a key paper(s)
- ❑ I filed a patent application or have been granted a patent
- ❑ I received additional funding
- ❑ I expanded my research group
- ❑ I formed new partnerships/collaborations
- ❑ I changed my research direction
- ❑ I changed institutions
- ❑ Other

16A. What awards did you receive?

17. Prior to your NDPA applications, has you research been funded by the NIH?

- ❑ Yes
- ❑ No

18. Did you approach a scientific problem differently in your NDPA application compared to other funding applications?

- ❑ No
- ❑ I cannot recall
- ❑ Yes. In what way?

19. In our effort to continue following the paths of exceptional researchers with creative ideas, we would like to administer this survey again in 2-3 years. It would be a tremendous help to have your input again. Would you be willing to complete this survey at that time?

- ❑ Yes
- ❑ No
- ❑ Maybe

20. As our thanks to you for completing this survey, we will donate $20 to a charity of your choice. Please select one of the charities below for your contribution.

- ❑ UNICEF
- ❑ American Cancer Society
- ❑ Doctors Without Borders

21. Please feel free to share any other thoughts with us related to the NDPA program or tell us more about yourself.

Appendix B: Abstracts from an NDPA and R01 applications

[NDPA application]

The fundamental scientific problem we propose to address is to determine the basic neural and molecular requirements for vocal learning, the behavioral substrate for spoken language. Language is one of the essential behaviors that make us human. With it, we are able to communicate complex concepts, pass on knowledge culturally, and advance human civilization. Without it – due to brain damage, trauma, or developmental diseases - we live a life of impoverished social communication and life dependency on others. Studying this fundamental problem requires that we compare the vocal behavior and associated brain pathways of the few rare groups that have vocal learning - four groups of distantly related mammals (humans, cetaceans, elephants, and bats) and three groups of distantly related birds (parrots, hummingbirds, and songbirds) – with the vast majority of species that do not have it - non-human primates, rodents, suboscine songbirds, pigeons, chickens, etc **[Two references deleted]**. Remarkably, although vocal learners are distantly related to each other, of those whose brains that have been studied (humans, parrots, hummingbirds, and songbirds), evidence suggests that they share a similar vocal pathway forebrain organization: a premotor or anterior vocal pathway (AVP) necessary for vocal learning, including syntax learning, and a motor or posterior vocal pathway (PVP) necessary for production of learned vocalizations **[One reference deleted]**. These forebrain pathways are not found in vocal non-learners. Yet, vocal non-learners appear to possess similar brain pathways for learning and production of non-vocal motor behaviors. Given these findings, we have proposed that the fundamental difference between vocal learners and non-learners is a genetic difference or several genetic differences that control the connection of forebrain motor learning pathways onto brainstem motor neurons that normally control the production of innate vocalizations **[One reference deleted]**. Our approach will involve the following: **[The sentence used to read: In this essay, I outline the following proposal for testing this novel idea:]**: 1) Discover molecular differences in the motor learning pathways between vocal learners and non-learners; 2) Manipulate their network connectivity by developing novel gene manipulation tools; and 3) Use these tools to modify vocal nuclei connectivity and thus vocal behavior of a vocal non-learner, potentially allowing other species to modify and imitate sounds and allowing correction of damaged vocal learning brain pathways in vocal learners. **[This list was formatted in a different way: each number started from a new line.]** Inducing such connectivity and behavioral changes in vocal non-learners would have profound impact towards understanding molecular mechanisms of vocal learning and evolution of language. Repairing the pathway in vocal learners, when damaged, would have profound impact for correcting neurological disorders of speech.

[R01 application]

Most vertebrate genes contain multiple introns which must be precisely removed from the primary transcript prior to its export from the nucleus to create the proper mRNA to direct translation. The process of RNA splicing which is responsible for removal of introns and ligation of exons is therefore an essential step in the expression of most genes. However, the basis for the specificity of this process is not well understood. The goal of this proposal is to understand the rules which are used by the vertebrate RNA splicing machinery to identify exons, introns and splice sites in primary transcripts and to encode these rules in computer programs which predict the splicing pattern of an arbitrary input primary transcript sequence. This will be accomplished by in-depth computational and statistical analysis of available primary transcript and mRNA sequences of vertebrate genes, taking advantage of the recent progress of large-scale genome sequencing and cDNA sequencing efforts. The approach will involve: 1) analysis of the detailed compositional properties of 5' and 3' splice signals and branch signals of vertebrate introns; 2) identification of exonic and intronic splicing enhancers and repressors; and 3) integrated computer models of slicing specificity enhancers and repressors; and 3) integrated computer models of splicing specificity. A variation of the Gibbs sampling algorithm will be used to characterize the branch signal and other signals which occur at a characteristic but variable distance from splice junctions. Clustering algorithms will be used to identify natural subgroups of 5' and 3' splice signals composition and to assign scores to potential splice signals. A statistical approach will be applied for identifying short sequence motifs which are likely to function as exonic or intronic splicing enhancers or repressors based on differences in oligonucleotide composition between exons and introns with weak versus strong splice signals. Conservation of putative splicing enhancers and repressors between homologous exons and introns from different vertebrates will be explored. As knowledge accumulates about splicing specificity, it will be integrated into computer models which predict the splicing patterns of primary transcripts. These models will be adapted to the problems of gene identification in genomic sequences and prediction of the splicing phenotypes of human mutations and polymorphisms. Deciphering the 'splicing code' will be essential to understanding the basis of alternative splicing, an important regulatory mechanism involved in development, differentiation and apoptosis. Computational methods for predicting splicing patterns will also aid in identification of genes including human disease genes and for understanding the effects of disease gene mutations, approximately 15% of which affect splicing.