

Portfolio Analysis— An Experimental Space

Dinah Singer, Ph.D.
Former Acting Head, Portfolio Analysis, DPCPSI
Director, Division of Cancer Biology and Head,
Molecular Regulation Section, Experimental
Immunology Branch, NCI

DPCPSI mission includes:

“Identify research that represents important areas of emerging scientific opportunities, rising public health challenges, or knowledge gaps that deserve special emphasis...”

NIH Reform Act of 2006

A Duality is Implicit in this Mission Statement:

Development of the capabilities to extract novel concepts and knowledge from the scientific achievements represented in the research portfolio to identify emerging areas of research and scientific opportunities.

Computational approaches that can probe large data sets do not exist yet

Dissemination of these capabilities to the broad NIH community.

Dissemination should include:

Providing information on available analytic tools

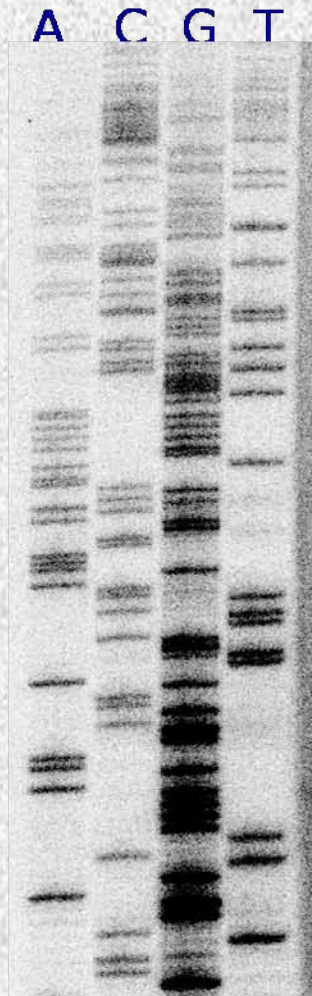
Providing training

Performing analyses

Definition of Portfolio Analysis

- In the context of DPCPSI, “portfolio analysis” is a knowledge discovery endeavor defined as:
the electronic integration and analysis of data derived from NIH research portfolios and other sources to identify emerging concepts/areas, opportunities, and gaps in research that will assist programmatic strategic planning of future areas of NIH support.

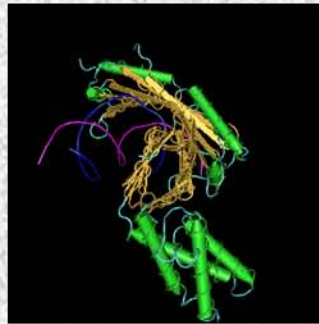
Steps in Biological Analyses



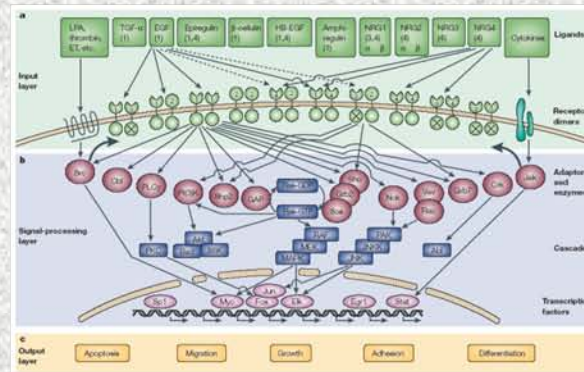
Sequencing

CTGGTGGGAAGACTTGAAC TTGAATCTCGAACCAC TGCATCTCCGACTCTGCCAGACTC
TTCAC TCCGCGGCACCCCTCAAACCC CAGCCAGGCCGGGGCGCACAAGCCAGCCAGCGCA
CCTGCAGTCTCTCGCCCGGACGCGCGCGCCCCCTCGGAACCAGGCTCTGCTCCGAGCAGC
CTTCGCCCTCAAGCCAGCCACAGTCCCCGCCAGGCCGGTGGGCGTCAAGATGAAGGCG
M K A

Derived DNA
sequence



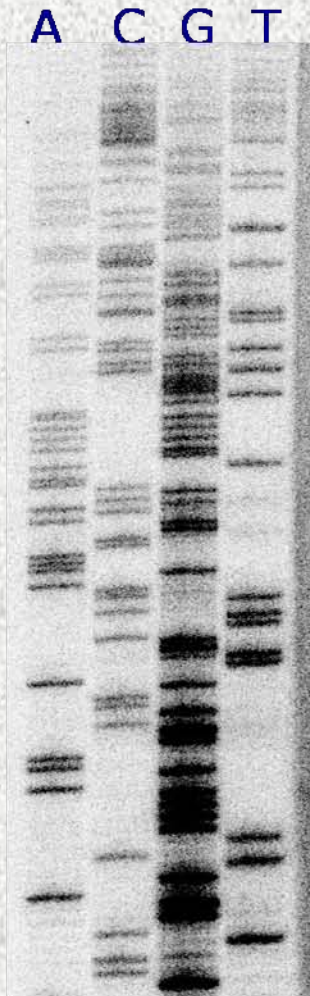
Amino Acid
Sequence and
Protein Structure



Position and
function in
biological
pathway

(Computational Algorithms)

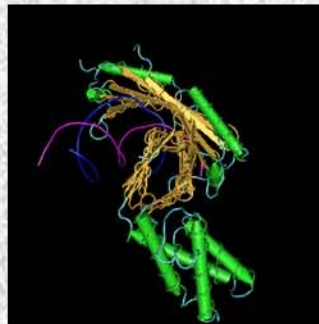
Steps in Knowledge Analyses



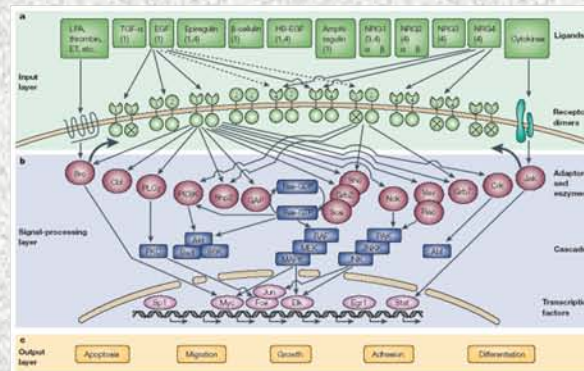
Inventory of
Applications
(Reporting)

CTGGTGGGAAGACTTGAACCTGAATCTCGAACCAGTGCATCTCCGACTCTGCCAGACTC
TTCACCTCCGCGGCACCCCTCAAACCCAGCCAGGCCGGGGCCGACAGCCAGCCAGCGCA
CCTGCAGTCTCTCGCCCGGACGCGCGCGCCCCCTCGGAACCAGGCTCTGCTCCGAGCAGC
CTTCGCCCCCTCAAGCCAGCCACAGTCCCCGCCAGGCCGGGTGGGCGTCAAGATGAAGGCG
M K A

Derived Concepts



Generated
Knowledge



Directions,
Opportunities
and Gaps
(OD, ICs)

(Computational Algorithms)

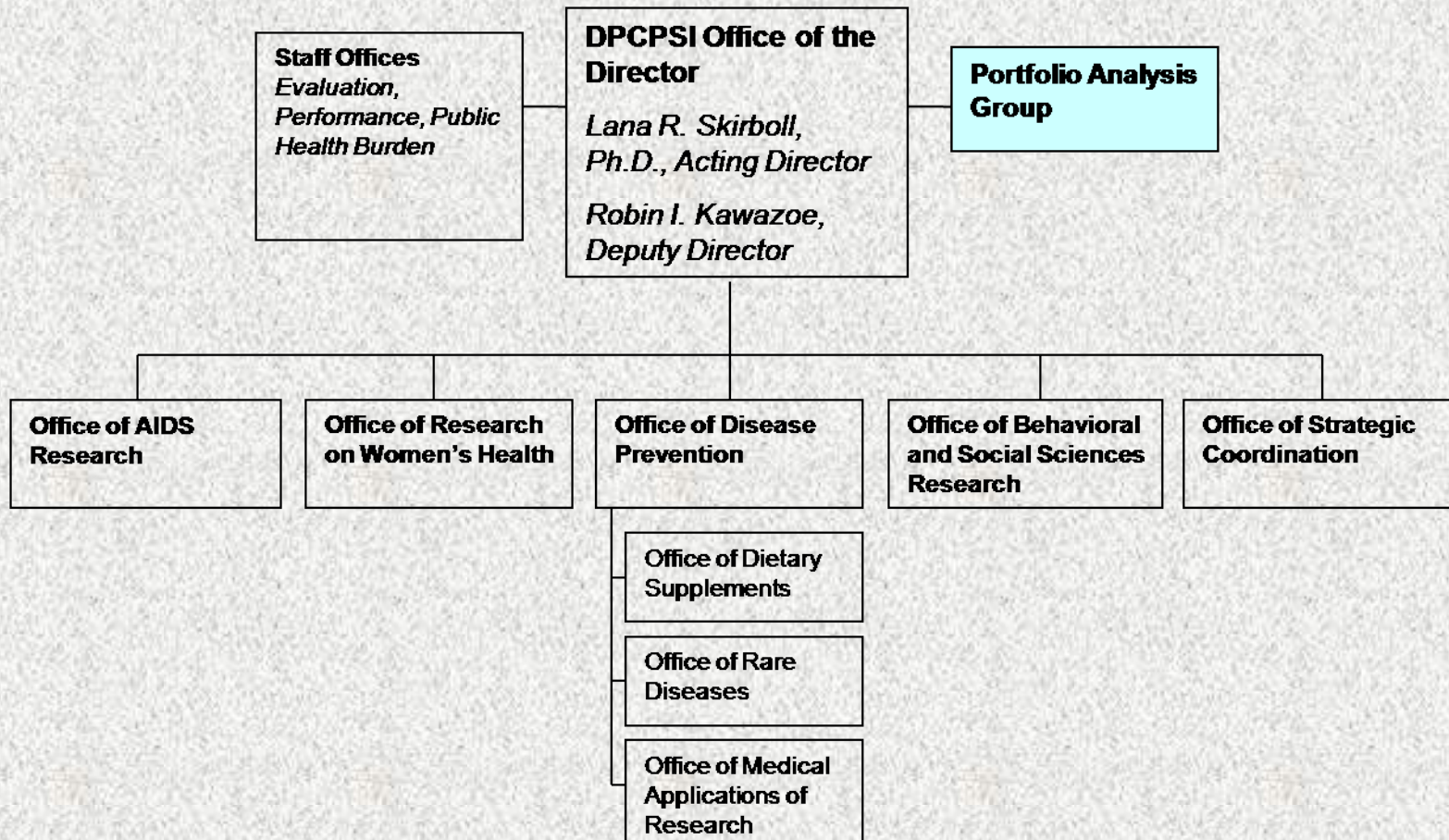
Analysis vs. Reporting vs. Evaluation

- Portfolio analysis is distinct from reporting and evaluation, in that:
 - It uses reporting tools (e.g. RCDC) to extract data which are analyzed with a variety of computational algorithms to derive new understanding about the state of science, which informs programmatic planning and/or evaluation.



Portfolio Analysis Group

PAG is responsible for enabling the analysis of the NIH portfolio of biomedical research



PAG's Functions

- The Portfolio Analysis Group serves dual roles:
 - Support research in and development of the emerging area of knowledge assessment/portfolio analysis
 - Serve as a resource on scientific, data analysis and IT tool capabilities for portfolio analysis at the NIH OD, IC, and programmatic level

Who Does What?

- DPCPSI Portfolio Analysis Group: Provide scientific analysis expertise in response to grant portfolio questions from the OD and ICs; research and develop analytic tools that extract novel concepts and knowledge from abstracts, publications, and other sources.
- OER Division of Information Services Reporting Branch: Ensure overall data quality and integrity of databases; develop analytical reports, on-line statistical models and visualization tools for extramural programs.
- CIT High Performance Computing and Informatics Office: Provide expertise and resources in high performance computing, computational science, biomedical informatics, and modern information technology to the NIH scientific community.
- NLM National Center for Biotechnology Information: Develop research-related data resources, text mining, predictive analytics.

PAG interacts with other NIH components to achieve its mission:

- OER Division of Information Services Reporting Branch:
 - Development of fingerprints for novel reporting capabilities using NIH automated enterprise reporting tools such as RCDC, QVR.
 - Semi-monthly meetings to discuss common interests
- CIT High Performance Computing and Informatics Office:
 - Established collaboration to develop new computational algorithms for portfolio analysis, using text mining, artificial intelligence, natural language processing, etc, approaches.
- NLM National Center for Biotechnology Information Research
 - Informal exchange of ideas, approaches, tools

Summary of Needs Assessments

2008-2009

- Hire staff with strong scientific credentials and interest in portfolio management to develop analytical applications
- Data integration and analysis
 - Include in analyses, additional data elements, such as unfunded applications, full text from grant applications, multi-component grants.
 - Fully integrate grants databases across NIH into a common structured format.
 - Develop novel analytic approaches, tools with improved search, analytic and visualization capabilities.
 - Conduct retrospective pilot projects to assess validity and completeness of analyses.
- Provide training on available software tools and increase overall awareness of the principles of portfolio management.
- Obtain on-going input from subject experts, analysis experts and users.
 - Work with Program Officers to apply new methods of portfolio management. Revive the Scientific Interest Group in Portfolio Analysis and/or convene focus groups.
 - Active outreach to Program Officers to ensure that NIH staff are aware of new developments.

Summary of Needs Assessments

2008-2009

- Hire staff with strong scientific credentials and interest in portfolio management to develop analytical applications
- Data integration and analysis
 - Include in analyses, additional data elements, such as unfunded applications, full text from grant applications, multi-component grants.
 - Fully integrate grants databases across NIH into a common structured format.
 - Develop novel analytic approaches, tools with improved search, analytic and visualization capabilities.
 - Conduct retrospective pilot projects to assess validity and completeness of analyses.
- Provide training on available software tools and increase overall awareness of the principles of portfolio management.
- Obtain on-going input from subject experts, analysis experts and users.
 - Work with Program Officers to apply new methods of portfolio management. Revive the Scientific Interest Group in Portfolio Analysis and/or convene focus groups.
 - Active outreach to Program Officers to ensure that NIH staff are aware of new developments.

Portfolio Analysis Group Current Staffing Level

3 Biomedical scientists
3 Computational scientists
1 Data analyst
1 Support staff

Summary of Needs Assessments

2008-2009

- Hire staff with strong scientific credentials and interest in portfolio management to develop analytical applications
- Data integration and analysis
 - Include in analyses, additional data elements, such as unfunded applications, full text from grant applications, multi-component grants.
 - Fully integrate grants databases across NIH into a common structured format.
 - Develop novel analytic approaches, tools with improved search, analytic and visualization capabilities.
 - Conduct retrospective pilot projects to assess validity and completeness of analyses.
- Provide training on available software tools and increase overall awareness of the principles of portfolio management.
- Obtain on-going input from subject experts, analysis experts and users.
 - Work with Program Officers to apply new methods of portfolio management. Revive the Scientific Interest Group in Portfolio Analysis and/or convene focus groups.
 - Active outreach to Program Officers to ensure that NIH staff are aware of new developments.

PAG Development of Analytic Capabilities

Research on existing analytic tools:

Text parsing and data cleansing:

Perl

Clustering/Topic Modeling:

In-spire

Topic Map

Lingo 3G/Carrot

Mallet

NINDS Visual Browser

Network/Visualization:

R

NodeXL

HeatMap

PAG Pilot Projects Designed to Develop Analytic Capabilities

- High risk/high reward research – validation of novel algorithm
- Lung cancer – identification of trends and emerging areas
- Translational research – tracking of broad areas of research
- RSV vaccine – mapping of scientific contributions to the development of a therapy

PAG Pilot Projects Designed to Develop Analytic Capabilities

- 
- High risk/high reward research – validation of novel algorithm
 - Lung cancer – identification of trends and emerging areas
 - Translational research – tracking of broad areas of research
 - RSV vaccine – mapping of scientific contributions to the development of a therapy

Retrospective Analysis for High Risk High Reward Applications

19

- Background:
 - NIH Reform Act 2006 required a demonstration project related to Roadmap for High Risk High Reward Research (HRHR)

Definition:

“Research with an inherent high degree of uncertainty and the capability to produce a major impact on important problems in biomedical/behavioral research”

Created by HRHR Demonstration Oversight Group for use in this Demonstration Project only

- CIT developed a text mining tool to categorize HRHR applications based on the language in individual grant summary statements.

Retrospective Analysis for High Risk High Reward Applications (Cont.)

- PAG has designed a pilot demonstration project
 - **Retrospective analysis** for HRHR applications.
 - Main purpose: assist in validating the methodology
- Goals:
 - Validate the categorization results
 - Conduct the portfolio analysis for both funded & unfunded HRHR research from 2003-2004
 - Assess the scientific success of the validated HRHR applications by examining publications, grant renewals, other sources of funding

Collaboration with OPA, CIT, and OER DIS

Retrospective Analysis for High Risk High Reward Applications (Cont.)

- **Approach:**

- Cluster the funded and unfunded research into three groups

Determined "Meritorious" by Peer Review
Determined "Good Enough" by Peer Review
Determined "Not Merited" by Peer Review


- Find distribution of HRHR+ and HRHR- applications in the above three groups
- Validate categorization results for a sample of HRHR+ and HRHR- from each group
- Assess scientific success
 - Conduct PubMed and Patent searches for the validated HRHR applications from three groups
 - Assess how many R01's identified as HR/HR were competitively renewed.
 - Attempt to determine how many unfunded HR/HR applications received support from other sources.

Retrospective Analysis for High Risk High Reward Applications (Cont.)

- Progress to Date:
 - Data selection criteria
 - NCI Type 1 R01 applications from FY2003 and FY2004
 - Data gathering
 - Query and Meta-data retrieval - DIS
 - 9000+ summary statements download
 - CIT ran HRHR application on the data set
 - Initial validation of random sampling of HRHR and not-HRHR by Program Staff.

Archana Bhandari, PAG

PAG Pilot Projects Designed to Develop Analytic Capabilities

- High risk/high reward research – validation of novel algorithm
-  • Lung cancer – identification of trends and emerging areas
- Translational research – tracking of broad areas of research
- RSV vaccine – mapping of scientific contributions to the development of a therapy

Lung Cancer Pilot Project

- Scientific Questions

- What were the major topic areas 5 years ago?
- What are the current scientific topic areas?
- How is the lung cancer portfolio evolving?
- Can we project the future direction of the lung cancer portfolio?
- How are these changes reflected in the broad areas of:
 - Basic biology of lung cancer?
 - Translational research?
 - Clinical research?

- Technical Question

- What resources do we have to help answer these questions?

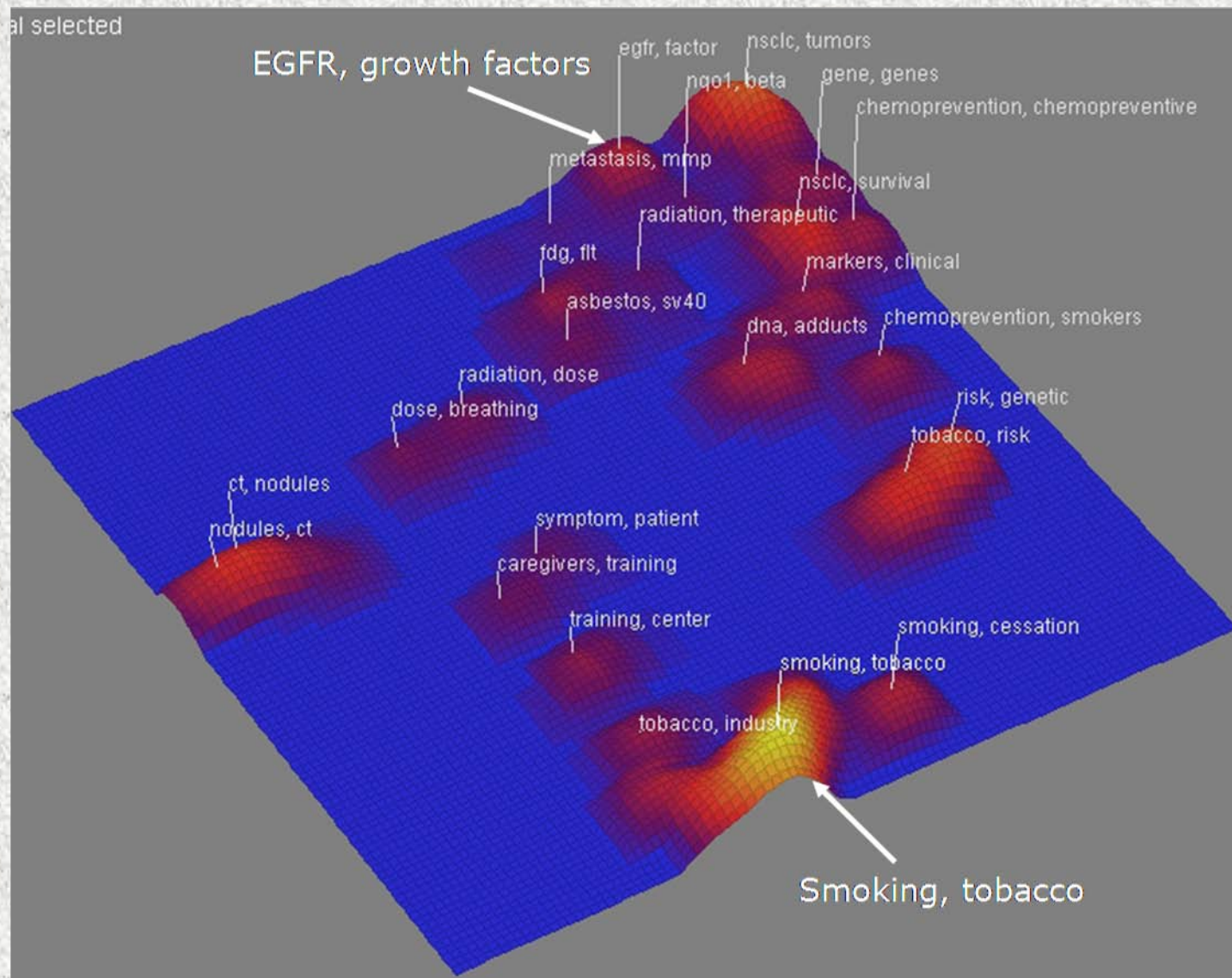
Lung Cancer Pilot Project

Technical Issues

- What resources do we have to help answer the scientific questions?
 - Currently available tools:
 - QVR
 - RCDC
 - e-SPA
 - Datasets
 - NCI Funding Database
 - Meeting Abstracts
- Do visualization tools such as IN-SPIRE capture information gleaned manually?

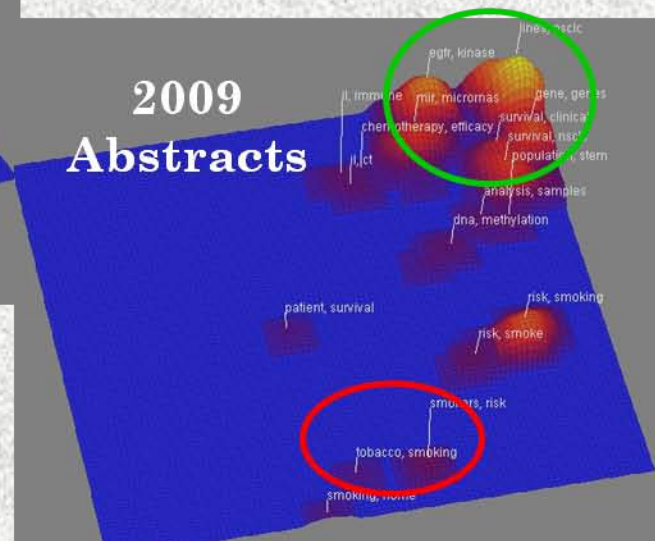
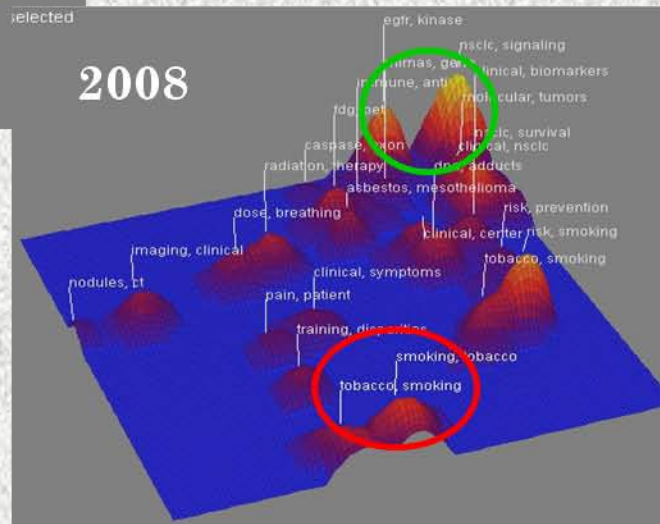
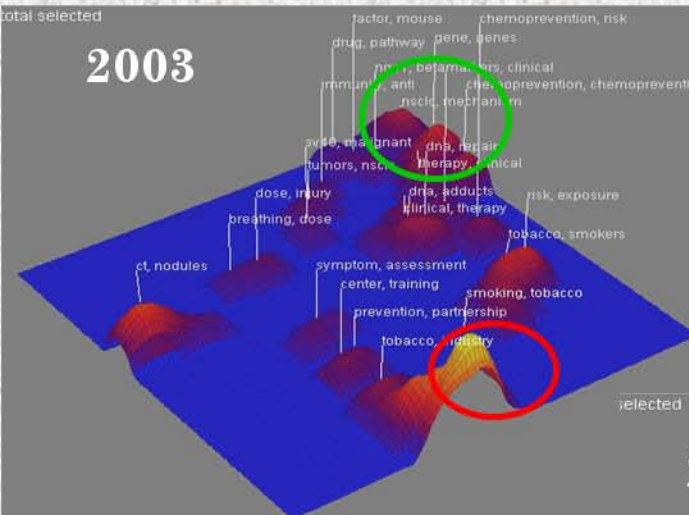
Carole Christian
Archana Bhandari

Visualizing the NIH portfolio 2003-2008 projects



Evolution of the Lung Cancer Portfolio

2003-2009



PAG Pilot Projects Designed to Develop Analytic Capabilities

- High risk/high reward research – validation of novel algorithm
- Lung cancer – identification of trends and emerging areas
-  • Translational research – tracking of broad areas of research
- RSV vaccine – mapping of scientific contributions to the development of a therapy

Summary of Needs Assessments

2008-2009

- Hire staff with strong scientific credentials and interest in portfolio management to develop analytical applications
- Data integration and analysis
 - Include in analyses, additional data elements, such as unfunded applications, full text from grant applications, multi-component grants.
 - Fully integrate grants databases across NIH into a common structured format.
 - Develop novel analytic approaches, tools with improved search, analytic and visualization capabilities.
 - Conduct retrospective pilot projects to assess validity and completeness of analyses.
- Provide training on available software tools and increase overall awareness of the principles of portfolio management.
- Obtain on-going input from subject experts, analysis experts and users.
 - Work with Program Officers to apply new methods of portfolio management. Revive the Scientific Interest Group in Portfolio Analysis and/or convene focus groups.
 - Active outreach to Program Officers to ensure that NIH staff are aware of new developments.

Training and Outreach

Within PAG:

- Sharepoint site for exchange of information, resources, etc
- Training manual for new staff

Within NIH:

- • Tools website as a central repository for access to available analytical tools, with descriptions and instructions
- Portfolio analysis for NIH OD and IC's
- Planned training capabilities, either through one-on-one sessions or courses

Preview of Tools Web Site

The screenshot shows the IN-SPIRE web site in a Microsoft Internet Explorer browser window. The address bar displays <http://www.od.nih.gov/opasi/OPA/tools/Pages/IN-SPIRE.aspx>. The page header includes the National Institutes of Health (NIH) logo and the text "DPCPSI Portfolio". A search bar is located in the top right corner.

On the left side, there is a vertical menu with the following links: e-SPA, IN-SPIRE, NIH Visual Browser, QVR, and RCDC. A purple arrow labeled "Index of Tools" points to this menu.

The main content area features the "IN-SPIRE" title and a "Link:" section. Below this, a "Description" section explains the tool's capabilities. A purple arrow labeled "Site available to NIH" points to the top of this section. Another purple arrow labeled "Tool Description" points to the "Description" section. A third purple arrow labeled "Sample output" points to two visualizations on the right side of the page.

The two visualizations are:

- Unfunded epigenetic research 2007 (858 projects)**: A 3D surface plot showing the distribution of research projects across various epigenetic markers.
- Epigenomes all years (Pilot, Grant and NIH 2008 funded Epigenetic projects (Red)**: A 3D surface plot showing the distribution of research projects across various epigenetic markers, with a red overlay indicating funded projects.

The bottom of the browser window shows the "Done" button and the "Local intranet" status.

Training and Outreach

Within PAG:

- Sharepoint site for exchange of information, resources, etc
- Training manual for new staff

Within NIH:

- Tools website as a central repository for access to available analytical tools, with descriptions and instructions
- • Portfolio analysis for NIH OD and IC's
- Planned training capabilities, either through one-on-one sessions or courses

Portfolio Analysis for NIH OD and ICs

Examples:

- Bioethics
- Biomarkers
- Climate Change
- • Comparative Effectiveness Research
- Behavioral and social sciences research
- Health economics
- Breast cancer

Comparative Effectiveness Research

Background:

- NIH has a long history of supporting comparative effectiveness research (CER) studies.
- NIH was awarded \$400 million of the \$1.1 billion allocated for CER under ARRA.
- PAG was asked to develop a CER “fingerprint” for annual CER reporting and for tracking the use of ARRA funds.

Strategy for Identifying CER Portfolio

- Refine and agree to a definition of CER (based on CBO definition)
- Generate a training set from examples provided by ICs
- Develop broad and narrow “fingerprints” of CER using terms for RCDC
- Query the entire NIH portfolio with the fingerprints
- Validate the fingerprints by expert review of a sample set of 200 projects into each of three categories: narrowly defined, broadly defined, not CER

Lessons Learned in Identifying CER Portfolio:

- Analysis is limited by the ability to extract valid datasets
- The ability to extract valid datasets is limited by:
 - A definition of the topic, agreed upon by experts, and a valid training set
 - The ability to include complex concepts in the query tool. For real analysis, more sophisticated natural language, text mining and AI tools need to be developed
- Even with an agreed upon definition, individual experts interpret the definition differently

Summary of Needs Assessments

2008-2009

- Hire staff with strong scientific credentials and interest in portfolio management to develop analytical applications
- Data integration and analysis
 - Include in analyses, additional data elements, such as unfunded applications, full text from grant applications, multi-component grants.
 - Fully integrate grants databases across NIH into a common structured format.
 - Develop novel analytic approaches, tools with improved search, analytic and visualization capabilities.
 - Conduct retrospective pilot projects to assess validity and completeness of analyses.
- Provide training on available software tools and increase overall awareness of the principles of portfolio management.
- Obtain on-going input from subject experts, analysis experts and users.
 - Work with Program Officers to apply new methods of portfolio management. Revive the Scientific Interest Group in Portfolio Analysis and/or convene focus groups.
 - Active outreach to Program Officers to ensure that NIH staff are aware of new developments.

Input from Subject Experts

Within NIH:

- Focus groups convened during 2009 provided recommendations

Across research community:

- “Think Tank” meeting of experts in knowledge analysis and knowledge management in Spring 2010
 - Provide input to PAG on the kinds of approaches that are being developed elsewhere and expertise that PAG will need going forward.
 - Major goal of Think Tank to allow NIH to frame a research program within PAG for knowledge analysis to improve our ability to:
 - Systematically analyze existing biomedical research
 - Project future opportunities for advancing important areas of research

Ongoing and Future Plans

- Expand collaborative research efforts with CIT, to include:
 - Developing analytic algorithms
 - Identifying specific business requirements of toolsets
 - Outlining specific methodologies and/or procedures to be followed in approaching the analysis
 - Building proof of concepts
 - Specifying the types/sources of data needed and cleaning the datasets.
- Expand collaborations with the OD and ICs
- Develop a plan to allow rotations through PAG to learn analysis techniques and tools

Questions to Consider

- What criteria should be used to prioritize projects requested from within NIH?
- How could portfolio analysis be helpful to Council members?
- What will be the role of the PAG in responding to ad hoc queries from outside of NIH?
 - Doing the analysis?
 - Referring it to the appropriate IC(s)?
 - Coordinating the analysis?

Portfolio Analysis & the Common Fund

- What information would be helpful to have about the portfolio of projects supported by the Common Fund?
- What information would be helpful to have for you to advise the IC Director where you are or were a member of the Advisory Council?
- What information are you receiving now that would be a useful resource for other Council members?

Acknowledgements

Lana Skirboll
Lora Kutkat
Robin Kawazoe

Chris Siemon, NCI

Tim Hays
Archana Bhandari
Carole Christian
Izja Lederhendler
Chuck Lynch
Geetha Senthil

Ehsan Haque
Ajay Shukla
Starla Collins
Alison Wohl

Questions & Discussion