

Portfolio Analysis Workshop

Characterizing Overlap

February 6, 2012

Panel Members

Catherine Lewis, PhD, NIGMS

Carl Baker, MD, PhD, NIAMS

Richard Conroy, PhD, NIBIB

Israel Lederhendler, PhD, OD

Carl Roth, PhD, LLM, NHLBI

Areas of Overlap

- Overlap of grants/applications for individual PIs
- Overlapping projects within defined scientific portfolios
- Overlap between portfolios within an IC
- Overlap between ICs in areas of common interest
- Overlap issues resulting from unclear referral guidelines within an IC or between ICs; referral of emerging or gray areas
- Overlap of topics clustered by biological question or by technical approach
- Overlap of new RFAs or PAs with existing investigator-initiated projects
- Overlap of trans-NIH activities with existing IC portfolios
- Overlapping grants/portfolios among different funding agencies or foundations

Overlap Questions to Consider

- How does one assess overlap? What tools can we use? How well do these work? How objective are they?
- How much overlap is acceptable? When is duplication of research appropriate? When is it wasteful?
- How might overlap analysis contribute to program arguments for funding applications beyond a payline (reaches, HRHR, NI/ESI, skips or special emphasis)?
- Does overlap analysis within a portfolio or across portfolios provide a basis for decisions about program balance?

Special Topics from Panelists

- Assessing Overlap with Document Analysis Tools
Carl Baker, NIAMS
- First, Get the Data – Then be Open to Surprises
Israel Lederhendler, OD
- Overlap from the small Institute Perspective
Richard Conroy, NIBIB
- Assessing Overlap – the Known Unknowns
Carl Roth, NHLBI

Assessing Overlap with Document Analysis Tools

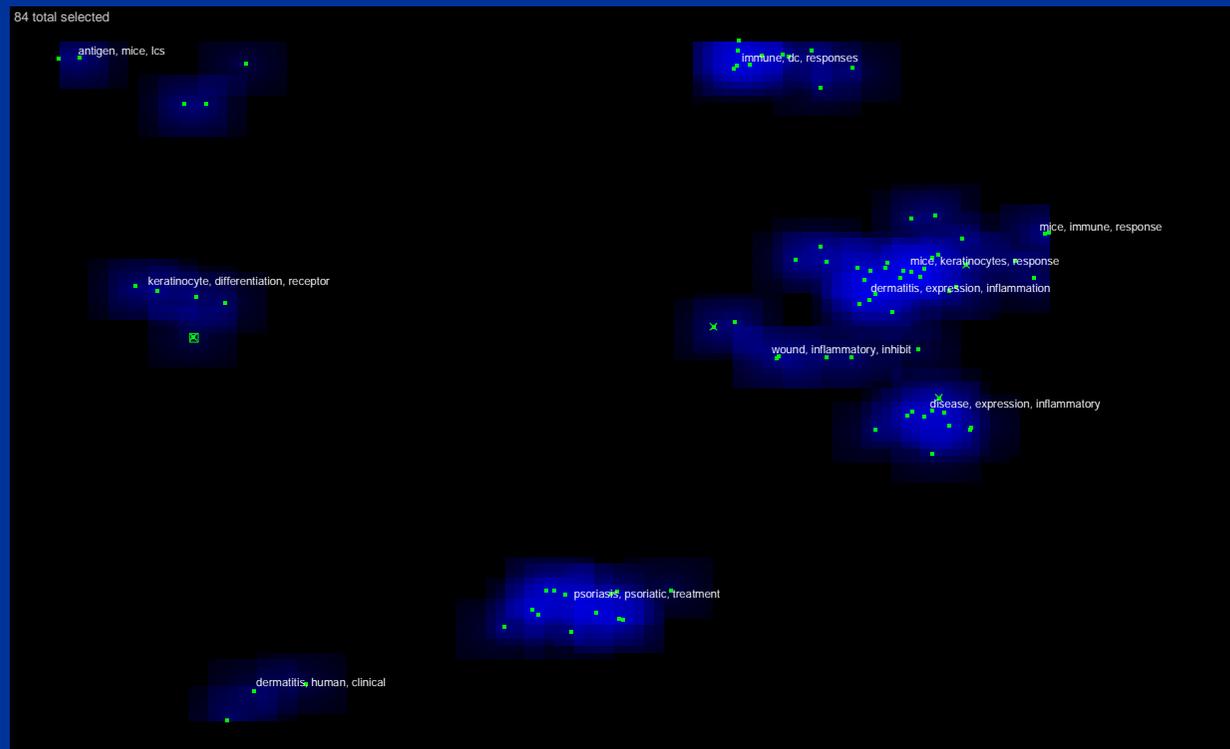
Carl C. Baker, NIAMS

Q: Can we use analysis tools to help make funding decisions?

Q: Do we capture new science by going beyond the payline?

IN-SPIRE: Visual Document Analysis Tool

- Analyzes multiple text documents and determines key topics or themes in each to create a signature for each document in the collection.
- Analysis is based purely on word usage (frequency, distribution) and not on scientific concepts.
- In Galaxy view, each dot represents a document and the distance between two dots indicates the degree of relatedness of the two documents.
- Similar documents are clustered together in a 2-dimensional space.



Portfolio analysis using IN-SPIRE:

- Grant title, abstract, and specific aims used for analysis.
- Fiscal year of competing award can be used for time analysis.
- Customization is required for optimal clustering: exclusion of common words (e.g. specific aims, hypothesis, data, etc.) and the inclusion of relevant scientific phrases (T cell, B cell, antimicrobial peptide, stem cell, etc.).
- Can add additional “Domain Knowledge”, but clustering will be more “biased”.

FY2010 New R01 in DSRD Programs

- 2008, 2009, 2010 Active R01
- 0-8%
- 8.1-12%
- 12.1-20%
- >750k

Hair
Follicle
Stem

Skin, Differentiation, Epidermal

Hair
Skin
Stem

Skin, Activation, Mice

Mice, Autoimmune, Activation

Genetic, Genes, Gene

RA, Arthritis, Inflammatory

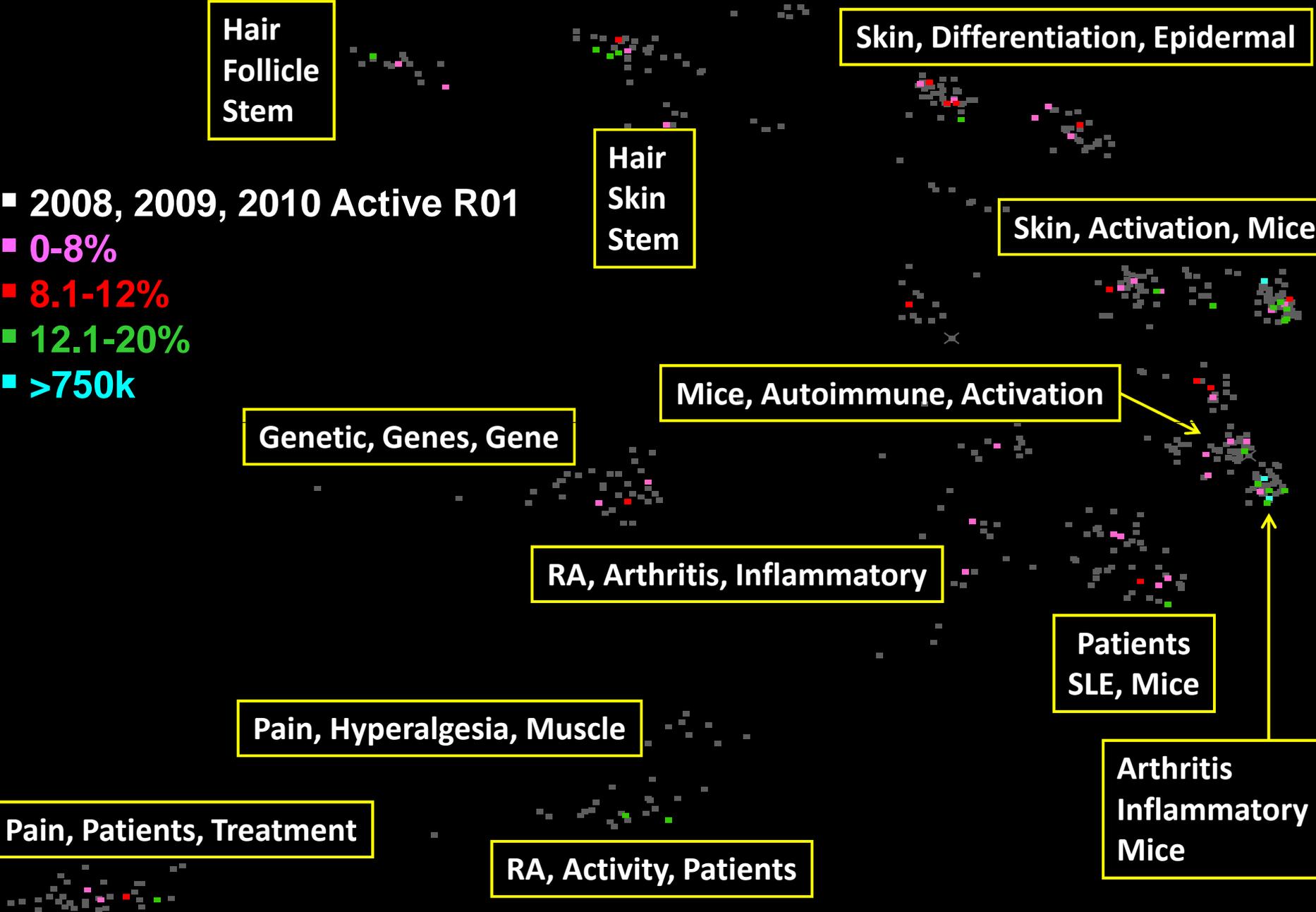
Patients
SLE, Mice

Pain, Hyperalgesia, Muscle

Arthritis
Inflammatory
Mice

Pain, Patients, Treatment

RA, Activity, Patients



Ideal Analysis Tool

- Ease in selecting and importing grants
- Scientific lexicon (terms, phrases, synonyms)
- Scientific concepts
- Kept current
- One-to-one or one-to-many relationship?
- Map grants to scientific literature
- Flexible

First, Get the Data ...

Israel Lederhendler, ORIS, OER, OD

- Is my program shrinking because of competition from another program?
- Can I marshal the evidence for the uniqueness of the science in my program so that even my director will understand it?
- Can I do so easily without needing to learn sophisticated software?

Use Scientific Concepts and the Carts Feature in QVR to Examine Overlap

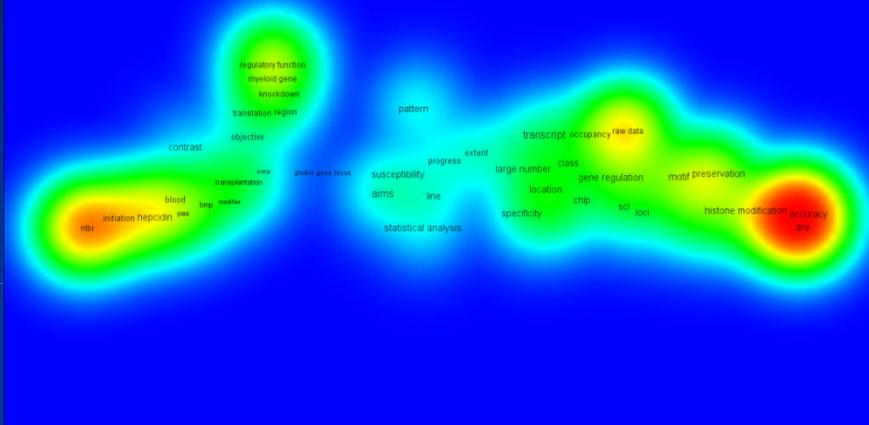
1. In QVR create Carts from relevant hitlists
2. Go to the Carts section
3. In the “Intersect” column select the carts being compared
4. Hit “GO”
5. The result is a list of projects whose fingerprints achieve a significant degree of overlap among the portfolios
6. In the comparison (to follow) of two PCCs, there were “zero” projects that intersected
7. However, using a second comparison based on “persons” in the two programs of about 200 projects, 12 PIs had projects in both programs

Not much overlap here, and what there is can be addressed strategically

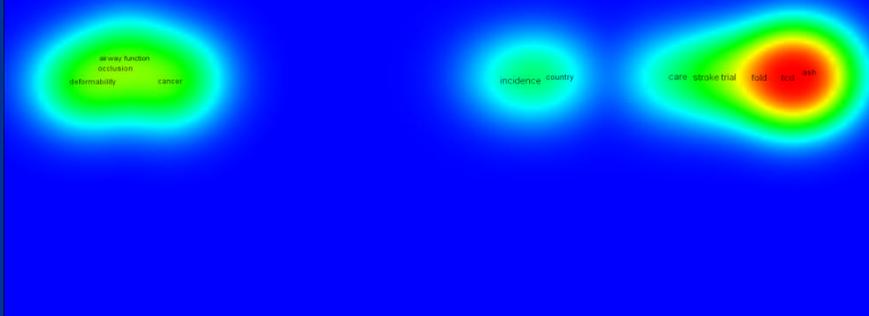
| PCC X | | PCC Y | |
|-------|-------------|-------|-------------|
| Terms | Base Wgt | Terms | Base Wgt |
| 1 | 100 | 1 | 100 |
| 4 | 100 | 3 | 100 |
| 6 | 100 | 4 | 100 |
| 2 | 96 | 6 | 100 |
| 3 | 87 | 13 | 89 |
| 7 | 73 | 2 | 86 |
| 5 | 72 | 10 | 83 |
| 8 | 64 | 5 | 79 |
| 15 | 62 | 16 | 74 |
| 11 | 58 | 12 | 71 |
| 17 | 52 | 15 | 68 |
| 12 | 50 | 7 | 63 |
| 16 | 44 | 9 | 59 |
| 9 | 43 | 14 | 54 |
| 32 | 42 | 8 | 53 |
| | exclusive X | | exclusive Y |

1. Create Carts in QVR for each PCC
2. Create Fingerprints and save them
3. Download Fingerprints into Excel
4. Sort each list of terms by their weights
5. Connect the terms that appear in each list
6. In the lists at left, the top 15 terms were examined

PCC Y



PCC X



VOS Viewer

(www.vosviewer.com) “Cluster Analysis for Dummies”

- A simple to use natural language clustering tool
- Great performance
- Useful for a quick overview
- Allows for terms refinement
- Open source
- Doesn't do much thinking for you

Overlap from the small Institute Perspective

Richard Conroy, NIBIB

- NIBIB Portfolio:
 - Early-stage technology development at interface of life & physical sciences
 - No disease focus – parse by technology
 - Non-hypothesis driven research
 - Small enough to consider non-automated approaches
- Is overlap measurable/important?
 - RCDC/'like' does not capture technology granularity
 - Funding determined by payline, few RFAs