

cceHUB

[Login](#)[Register](#)

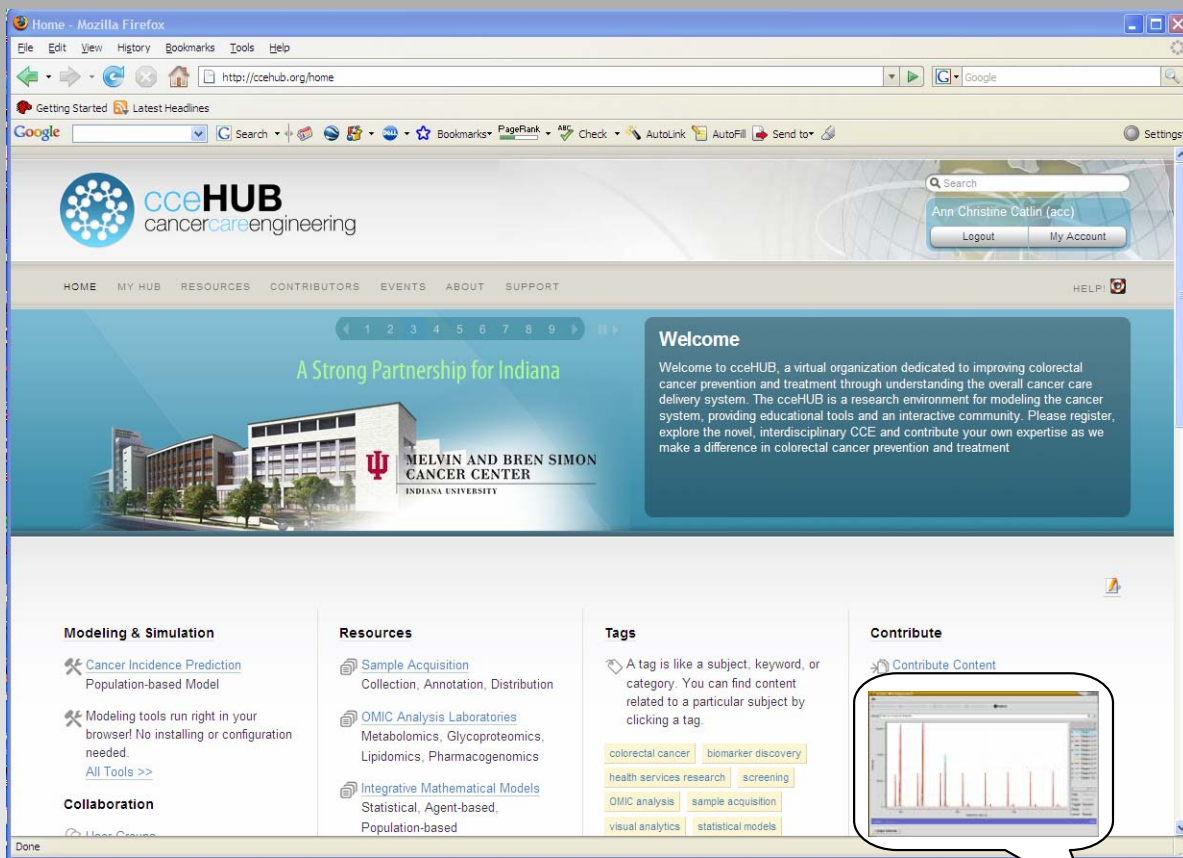
[HOME](#) [MY HUB](#) [RESOURCES](#) [CONTRIBUTORS](#) [EVENTS](#) [ABOUT](#) [SUPPORT](#)

A Knowledge Discovery Environment for Cancer Care Engineering Research

Ann Christine Catlin
HUBzero Workshop
November 7, 2008



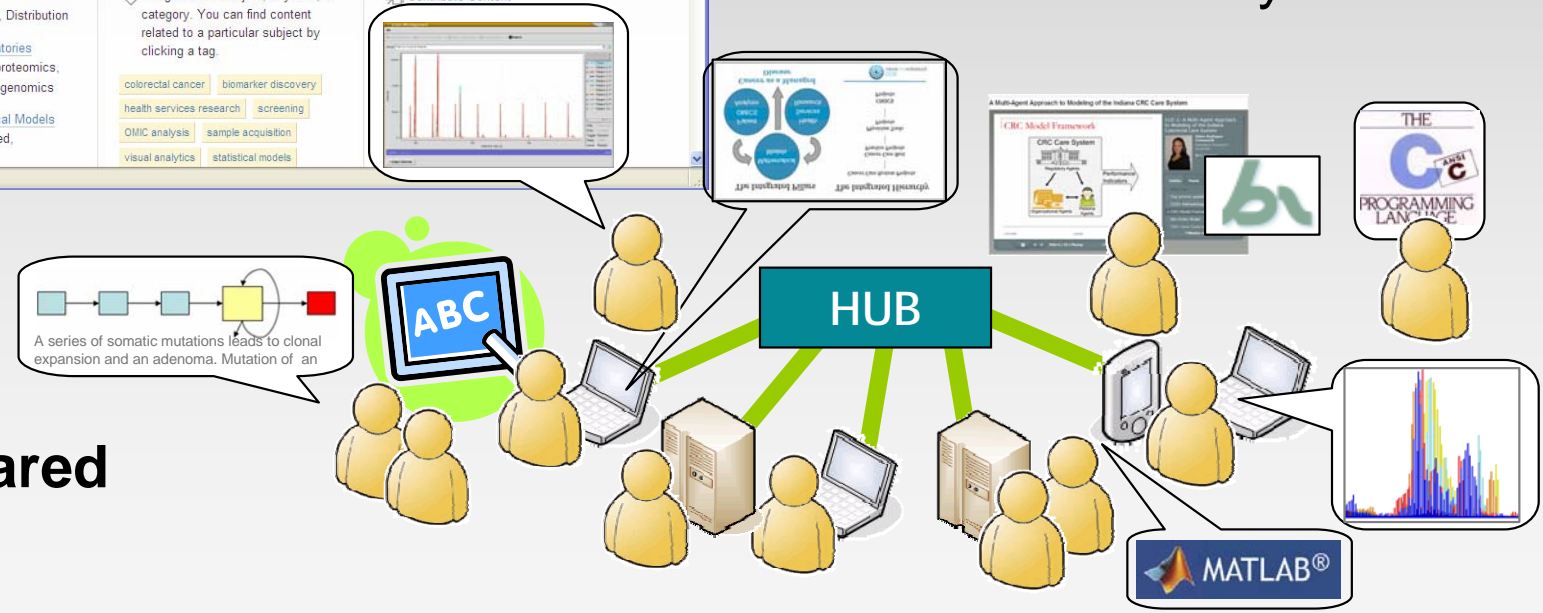
a HUB for Cancer Care Engineering



integrate/synthesize
biological OMIC data
→ biomarker knowledge

- OMIC Analysis Labs
- Statistical Modelers
- Visual Analytics

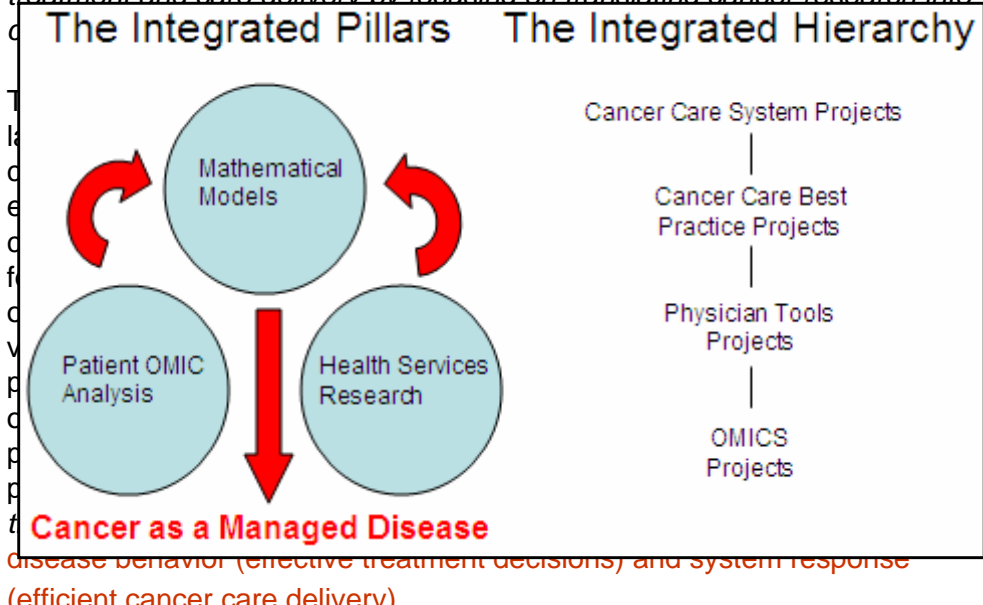
**Support for
Community-shared
Resources**





CCE integrated project hierarchy

The Cancer Care Engineering (CCE) project is a highly innovative, interdisciplinary, multi-institutional endeavor that holds promise for revolutionizing the current paradigms of cancer prevention, detection, treatment and care delivery by focusing on translating cancer research into



Cancer Care Engineering Projects funded by the Regenstrief Foundation

CCE-1: A Multi-Agent Approach to Modeling of the Indiana CRC Care System

CCE-2: An Indianapolis CRC Quality Improvement Initiative

...

CCE-5: A Fusion Center for Cancer Care System Information – the Cancer Care

... highly innovative, interdisciplinary, multi-institutional endeavor ...

CCE TEAM: more than 70 scientists, clinicians, statisticians, physicians, nurses, engineers, computer scientists, health service researchers, university and hospital staff

- Roudebush VA Center for Implementing Evidence-Based Practice
- IU Center for Health Services and Outcomes Research
- Department of Medicine, IU School of Medicine
- Department of Medicinal Chemistry and Pharmacology, Purdue University
- School of Chemical Engineering, Purdue University
- School of Electrical and Computer Engineering, Purdue University
- Department of Statistics, Purdue University

Laboratory Analysis, Conversion to Digital Data

...



CCE team: collaboration !

Cancer Care Engineering Retreat 2008
CCE Retreat 2008 Part V

Contributor(s) [Julie S. Nagel](#)

Abstract: Cancer Care Engineering Retreat Video Part V. (one hour)

The following presentations are included in the video:

- Reha Uzsoy, PhD, Clifton A. Anderson Distinguished Professor, Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University. Systems Approach to Colorectal Cancer Care. (00:00-58:56)
- Joe Pekny, Director, e-Enterprise Center, Purdue University. Wrap Up. (59:02-1:00:42)

Normal **Adenoma** **Advanced Adenoma** **Cancer**

Q&A session topics are discussed over lunch. CCE retreat participants wrap it up.

Credits: Julie Nagel, Managing Director of the

ects
to
are
ality
cer
ancer
-
sis
mple
n,
ersion



Using HUB Technology I

- Content Management System for Scientists
- Collaboration and Social Networking

RESOURCES

Contrib

Start a co

Shared in

The cceHUE
their work. I
the cceHUE

Contribute

There are si
submit materials for each type:

[Animations](#) | [Downloads](#) | [Public](#)

Present your work

Your contributions will become
locate there.

Events

Add an event

Year Month Week Day

17 Oct, 2008 [Improving CRC No-Show Rates: A Quality Improvement Initiative at IUMG and VAMC](#)

11:30 AM **Category:** Seminar
to 12:30 PM
Jamie Workman Germann, Associate Professor or Engineering Technology, IUPUI The Rengenstrief Center for Healthcare Engineering at Purdue University features Cancer Care Engineering health services research focused projects in its Brown Bag Seminar Series.

24 Oct, 2008 [cceHUB: Bringing Data into the Cyber Infrastructure](#)

11:30 AM **Category:** Seminar
to 12:30 PM
Ann Catlin, Rosen Center for Advanced Computing The cceHUB will be the first HUB community with a data layer incorporated. Learn what impact this will have on the CCE project.

29 Oct, 2008 [CCE Health Services Meeting: Visualization and Performance Dashboards](#)

02:00 PM **Category:** Seminar
to 03:30 PM
We welcome Dr. David Ebert, Director of the Purdue University Regional Visualization and Analytics Center (PURVAC) and a faculty partner of the Purdue Homeland Security Institute] and other Purdue researchers and partners to our October 29th CCE Health Services Monthly Meeting Agenda: ...

All Categories

2008

October

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

Teaching Materials

0.0 Ranking Sherer: [Predicting Patient-specific CRC ...](#)

0.0 Ranking Catlin: [A Hierarchical System Based ...](#)

0.0 Ranking Germann: [CCE-2: CRC Quality Improvement ...](#)



Integrative Mathematical Models

Colorectal Cancer Incidence Prediction Model - Mozilla Firefox

http://ccebhub.org/tools/incidence/

cceHUB
cancer.careengineering

HOME MY HUB RESOURCES CONTRIBUTORS EVENTS ABOUT SUPPORT HELP

Health Services

Colorectal Cancer Incidence Prediction Model

Contributors: Eric Sherer, Mohd Haziq Rahmad

At a glance: Stochastic simulation of polyp and colorectal cancer (CRC) incidence with patient age. Patient information such as demographics and colon history can be used to individualize incidence predictions.

Screenshots:

Description: This model describes the accumulation of somatic mutations within a single cell for genes commonly altered in colorectal adenomas and carcinomas. Acquisition of such mutations through time can lead to the formation of an adenoma (polyp) or a carcinoma.

Several default scenarios are available, where the mutation rates and genetics states have been fit to Indiana colorectal cancer incidence rates for various demographic groups (MF and B/W). Options are available for altering the model structure by changing any of the following:

- the genes considered
- the CRC genetic states
- the polyp genetic states
- the mutation rates between genetic states

The model can be compared with the appropriate actual data set.

Supporting Documents:

- Decision Tree Analysis: Adaptation to Colon Dynamics
- Mechanistic Modeling of Polyp and CRC Development
- Predicting CRC Incidence from Polyp Prevalence
- Introduction to the Theory

Launch Tool →

Preliminary investigation of a mutation network model for predicting CRC prevalence based on information from colonoscopies

Eric Sherer^{1,2,3,4}, Seza Orcun¹, Ann Rundell⁵, and Doraiswami Ramkrishna⁶

INDIANA UNIVERSITY
SCHOOL OF MEDICINE

PURDUE UNIVERSITY

1. e-Enterprise Center, Purdue University; 2. VA CoE on Implementing Evidence-Based Practice; 3. Indiana University Center for Health Services and Outcomes Research; 4. Regenstrief Institute, Inc.; 5. Biomedical Engineering, Purdue University; 6. Chemical Engineering, Purdue University

Abstract: A sequential somatic mutation network model is used to describe colorectal cancer (CRC) prevalence data. A hypothetical example is shown which demonstrates how model could be used to predict the likelihood and genetic characteristics of an individual patient's colorectal cancer based on results from colonoscopies. The model is currently being tuned to such information with the collection of Lagrangian polyp and CRC prevalence throughout patients' lifetimes. Future research will refine the genetic network by examining polyp and CRC biopsy characteristics collected throughout patients' lifetimes.

Sequential network => CRC prevalence

Asymmetric division

Objective: Develop a predictive model for the likely age incidence and genetic characteristics of CRC.

Methods: A multi-state model is developed where single features are altered sequentially in a stochastic manner (Michor et al. 2005). These features could potentially include CNV associated genes, MFI genes, or hypermethylation states. As an example, a 3-state model is applied to the Indiana CRC incidence data from 2000-2004. The 12 rate parameters are varied in a 2-factorial manner. Each parameter set provides the initial guess for a local parameter optimization with the resulting fit for a 'good' parameter set shown.

Future Work: The state model will be coupled with models predicting growth, invasion, and response to adjuvant chemotherapy. The presence of polyps and their characteristics could be fitted with the likelihood of CRC and its characteristics. These differences could be used to differentiate between the likely behaviors and responses of patients to various treatments.

$$x_{i+1} = x_i + \Delta t \cdot f(x_i, \theta)$$

$$x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$$

$$f(x_i, \theta) = [f_1(x_i, \theta), f_2(x_i, \theta), \dots, f_n(x_i, \theta)]^T$$

$$x_{i+1} = x_i + \Delta t \cdot f(x_i, \theta)$$

Advantages:

- analytic formulation allows for rapid solution
- Good fit to incidence data
- Finite number of discrete patient classes
- Preferred pathways and states unknown

Challenges:

- Multiple parameter sets for the mutation network can be identified that fit the CRC

Patient Differentiation

Patient Differentiation based on +/- "polyp" types

Polyp prevalence => likelihood of behavior

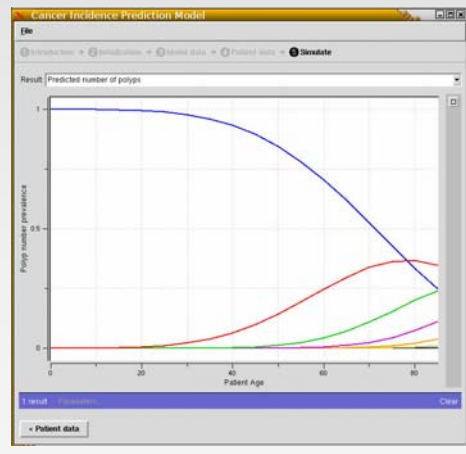
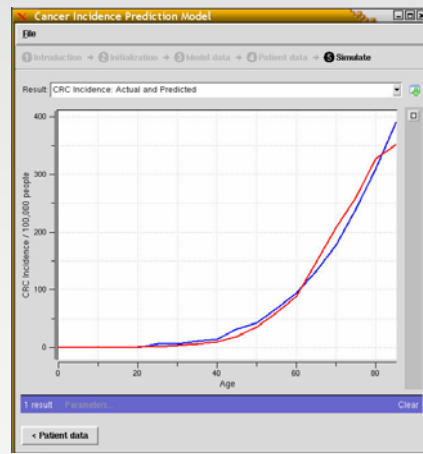
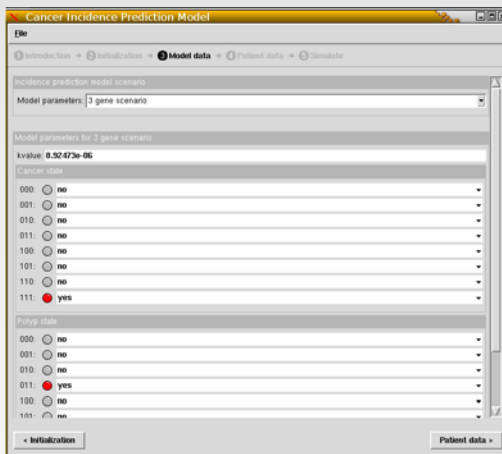
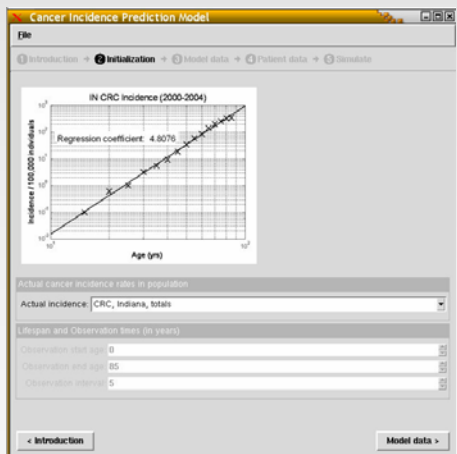
slight differences in behavior of mutation based on demographics implies that "neurological" conditions vary between the classifications. The variable conditions will produce different occurrence behaviors with time.

Parameter sets for the model were fit to the IN incidence data. Besides CRC incidence, these parameter sets also give the rate of formation for three hypothetical types of polyps. Since the parameter sets vary between the demographic groups, the likelihood of various polyps will also be different. So, the likelihood that an unknown patient - that is, of unknown race and gender - belongs to a given demographic group can be found by observing the polyp classification. The figures below are the probability that at least 1 polyp of a given type is present, but the resolution can be increased by including the odds of combinations of multiple polyps.

Question of interest: Can the demographic of an unknown patient be identified from observations?

Question of interest: Can the demographic probabilities - identified from observations - be used to predict the likelihood of developing CRC in the future?

Probabilities of patient classifications => likely CRC incidence





Shared models need shared data ...

Monte-Carlo Simulator

File

Getting Started → Model Setup → Model Conditions → Simulate

Nordling, 1953
CRC is the result of a series of somatic mutations in a stem cell. Each stem cell mutates independently.

Luebeck & Moolgavkar, 2002
A series of somatic mutations leads to clonal expansion and an adenoma. Mutation of an adenomatous cell begins a CRC lineage.

This tool performs Monte Carlo simulations on four colorectal cancer (CRC) incidence models which describe the likelihood that an individual will develop CRC at a certain age.

The sequential mutation model of **Nordling (1953)** showed that a series of roughly six somatic mutations captures the observed linear log (CRC incidence) versus log(age) relationship.

The model of **Luebeck and Moolgavkar (2002)** included the abnormal proliferation of an adenoma (which develops after a series of somatic mutations) as an intermediate step to carcinoma. Then, any of the proliferating cells can undergo the final transformation to a cancerous cell.

GCxGC MS Alignment

File

Getting Started → GC-GC-MS Datasets → Model Parameters → Output Selection → Analyze

xMass: the PDP Deconvolution Tool

1 OMICS Dataset → 2 Parameters → 3 Analyze

Result: Peak intensity vs Retention time

GCxGC MS Alignment

File

Getting Started → GC-GC-MS Datasets → Model Parameters → Output Selection → Analyze

Statistical Modeling of GCxGC MS Data

The model uses curve matching, including centering and rescaling, to align GCxGC MS data. In terms of retention time and mass charge (m/z), the two-dimensional Correlation Optimized Warping algorithm (2D-COW) tries to identify common patterns by automatic alignment.

INPUT: CDF datasets range of values for retention time 1, retention time 2, noise observed, m/z

OUTPUT: Total Ion Count, Selected mass charge values, Selected patient data, Comparison graphs

ALGORITHM:
Step 1: Calculate shift/alignment coefficients
Step 2: Apply alignment coefficients to GCxGC chromatograph generated at each m/z value in the m/z classification

GCxGC MS Alignment

File

Getting Started → GC-GC-MS Datasets → Model Parameters → Output Selection → Analyze

Choose input datasets

Choose input datasets from: cceHUB data repository

Search repository for GC-GC-MS collections

Search by collection name: Colon Cancer collection (public)

Choose sample datasets from: Search the repository by dataset collection name - some collections are marked as protected and can be accessed only by identifying cceHUB groups

Choose Colon Cancer sample: Colon Cancer collection (public): Collection of 49 colon cancer and control sample datasets

xMass: the PDP Deconvolution Tool

File

1 OMICS Dataset → 2 Parameters → 3 Analyze

Result: Peak intensity vs Retention time

xMass: the PDP Deconvolution Tool

File

1 OMICS Dataset → 2 Parameters → 3 Analyze

Choose an OMICS input dataset

Choose input dataset from: cceHUB data repository

Search cceHUB data repository

Choose dataset by: instrument

Search repository by instrument

Which instrument?: XCTPlus Ion Trap

Which XCTPlus Ion Trap instrument?: XCTPlus Ion Trap

Select LC parameters and instrument

Parameters >

xMass: the PDP Deconvolution Tool

File

1 OMICS Dataset → 2 Parameters → 3 Analyze

Analyte type: peptides

Instrument: Ion Trap

Data acquisition mode: Positive

Mass weight diff: 0

Max modification per peptide: 0

Centralization necessity: Profile (centralization is necessary)

LC peak width: 5

min effective retention time (in mins): 5.0

max effective retention time (in mins): 65.0

m/z variation between 2 isotopes: 0.12

< OMICS Dataset

xMass: the PDP Deconvolution Tool

File

1 OMICS Dataset → 2 Parameters → 3 Analyze

Result: View/download the output DLT file

Peak intensity vs Retention time

Scan: Mass/charge ratio of peptide peak vs Retention time

53.0 Molecular weight vs Retention time

54.0 Noise level vs Retention time

54.0 View/download the output PAR file

54.0 View/download the output DLT file

54.0 ---

54.0 Download

m/z	Retention Time (min)	Intensity	Isotope Ratio
555.0	445.329	446.337	20.4224, 22.0634, 1.9
55.0	555.0	671.456	672.464, 37.8179, 12.8296, 1.9
56.0	565.333	30.284	309.285, 405.21, 537.37, 0.373, 1.9
63.0	6345.0	766.844	767.851, 13.718, 11.2247, 1.9
70.0	7045.0	223.122	224.13, 5.59553, 8.66283, 1.9
72.0	7245.0	445.386	446.394, 12.2152, 23.6071, 1.9
75.0	7546.67	698.535	699.543, 18.3411, 20.0516, 1.9
76.0	7645.0	389.456	390.464, 41.0623, 75.1466, 1.9
77.0	7743.33	309.318	310.326, 8.28319, 13.531, 1.9
79.0	794.0	365.43	366.438, 12.3953, 31.0502, 1.9
88.0	8825.0	445.403	446.411, 10.2247, 25.198, 1.9
89.0	8925.0	581.547	582.555, 7.04737, 10.7763, 1.9
91.0	9125.0	335.373	336.381, 9.31713, 11.0533, 1.9
94.0	9425.0	307.314	308.322, 5.71286, 16.2451, 1.9

Find:

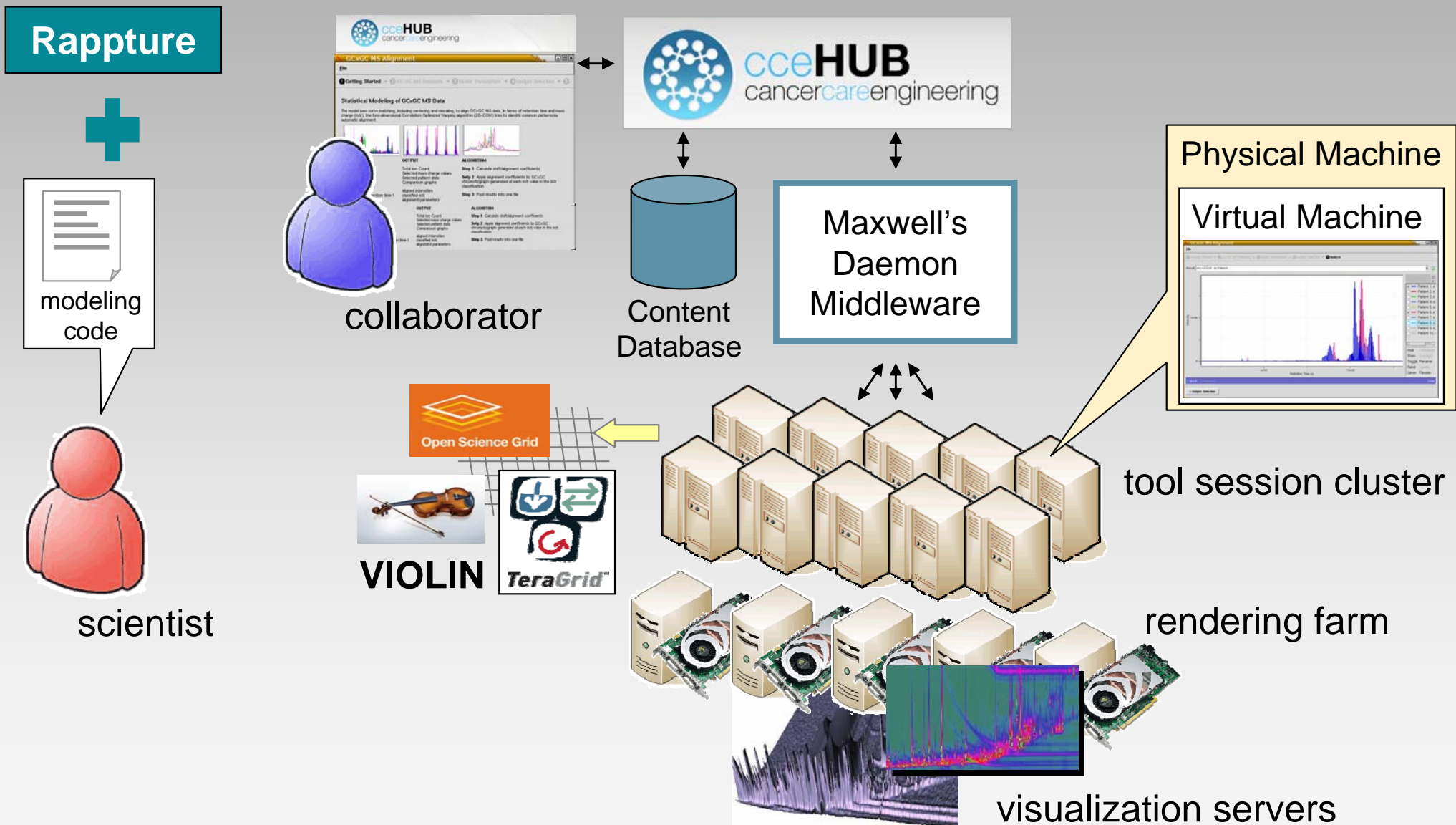
1 result Parameters Clear

< Parameters Analyze >



Using HUB Technology II

- Unique Middleware for Modeling and Simulation





What we need to support that's new ...

■ OMIC workflow

- biosamples → biological data → biomarker knowledge ...
“data lifecycle” support
- end-to-end user support

■ Data ... the new shared resource

- data repository & data support infrastructure
- metadata ... annotate, track, characterize content



OMIC experiment workflows

IU Simon Cancer Center

Sample acquisition
processing
transfer

Bindley Biosciences Laboratories

Sample preparation
Instrument analysis
Data generation
raw instrument data
pre-processed data

BLOOD SAMPLE
CRC PATIENT/CONTROL

clinical database,
demographics,
diet, diagnosis,
treatment

transfer to Purdue

sample



LECO
GC GC MS

instrument generated

PEG

350MB
spectrum, list of peaks,

data
converters

CSV

35MB
peak list

CDF

1.2GB
retention times,
m/z, intensities

data
converters

TXT

DAT/RAW

more ...

datasets

SAMPLE DATA
converted, reduced,
selected, filtered,
transformed

RAW
DATA

PRE-
PROCESSED
DATA

annotation

Interactive GUI

protocols

Document library
& document tagging

tracking

Metadata processing

external DB
linkage

Communication &
information exchange

annotation

Interactive GUI

methods

Document library
& document tagging

upload

Data transfer

tracking

Metadata processing

storage

File server & backup

metadata
linkage

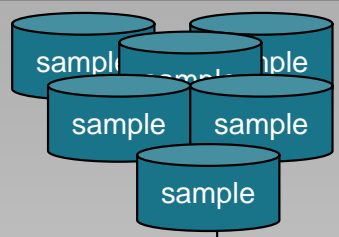
Metadata processing



OMIC data analysis workflows

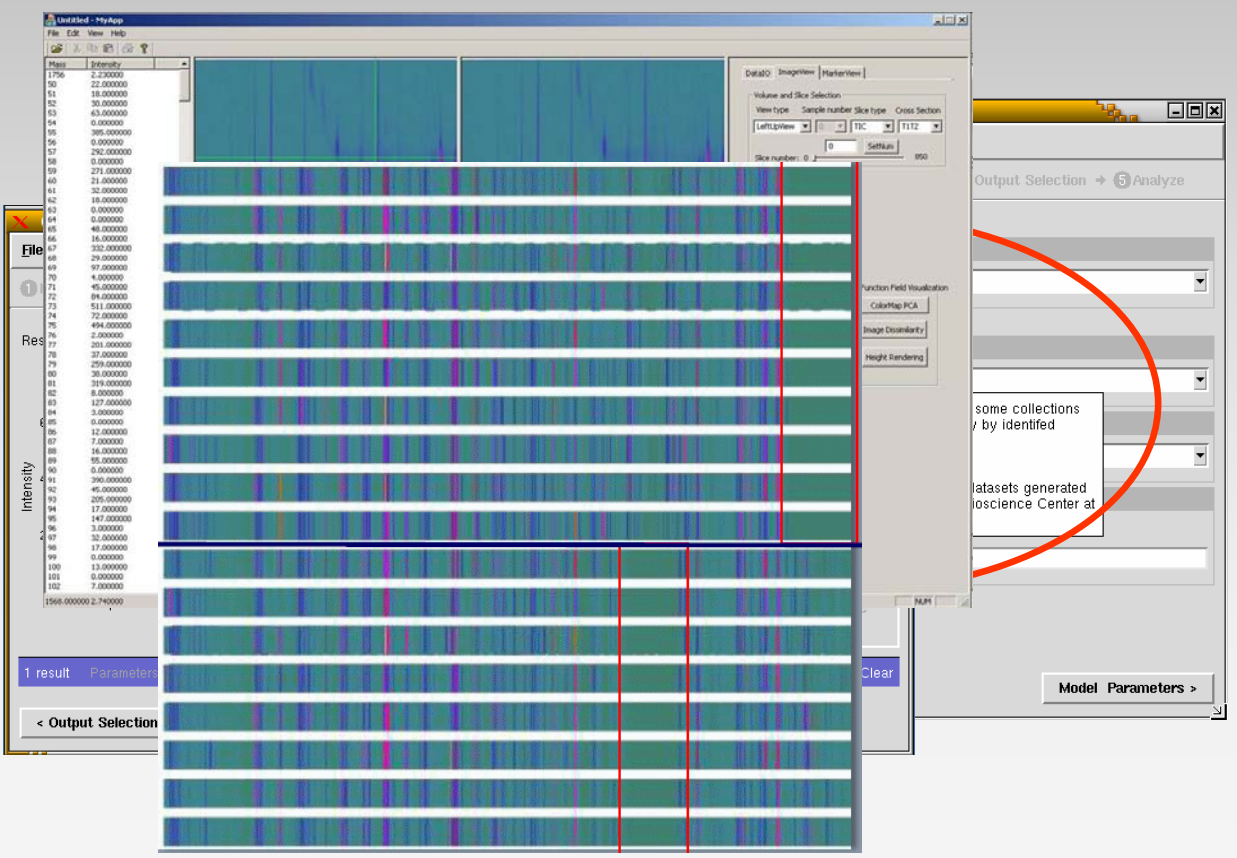
Laboratory Analysis
Statistical Modeling
Visual Analytics

Processing multiple samples
Comparisons across samples
Modeling across samples
Visualization across samples



ANALYZED DATA

datasets
→



- data exploration
- ontology support
- visualization
- modeling
- data tracking
- tool tracking
- data capture
- annotation
- metadata linkage
- storage



a HUB data support framework

