centocor
research & development inc

tranSMART

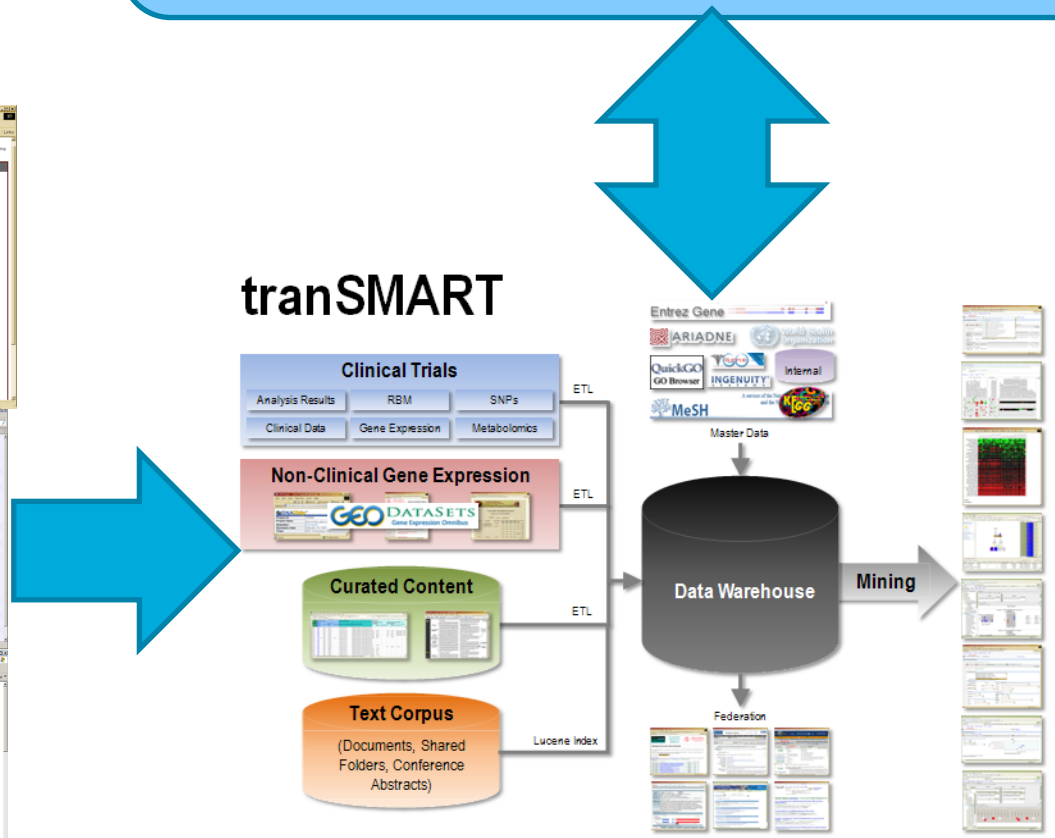
December 2010

Aim

- Focus - Translational Research
- Sampling of questions to be answered
  - what is the correlation between animal models and human data?
  - what is the best biomarker strategy for a given compound ?
  - what is the best indication for a given compound?
  - how can a disease stratified based on clinical data?
  - is there support for a target of interest based on clinical data?

centocor
research&development inc

Precompetitive Sharing



CIO100

Bio·IT World Best Practices AWARDS 2010

- Indication selection
- Target/pathway hypothesis
- Biomarker hypothesis
- Clinical study design

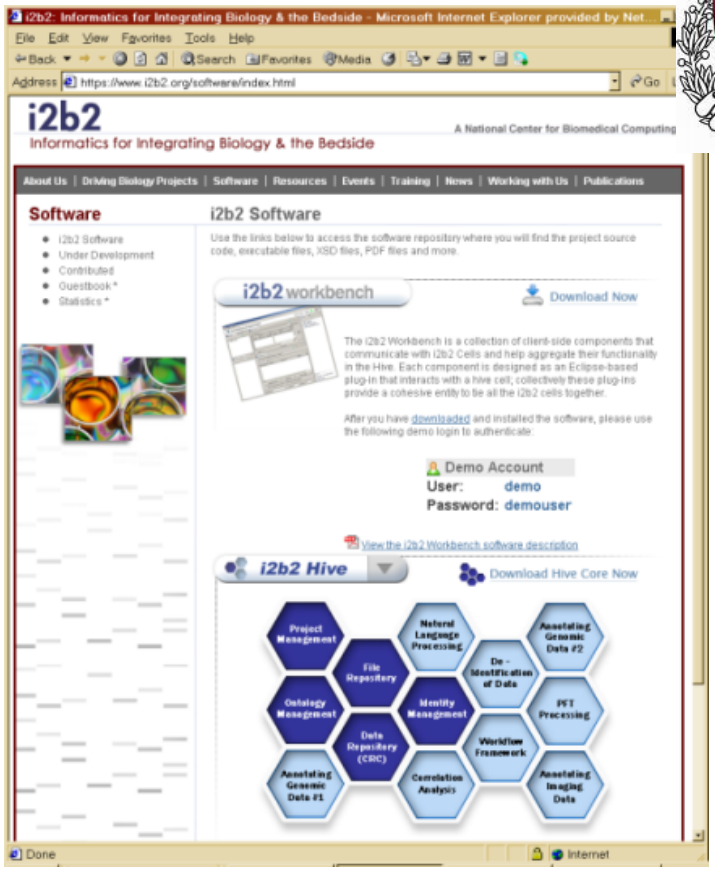
centocor research & development inc

Academic Partnerships

DANA-FARBER CANCER INSTITUTE

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

The Cancer Institute of New Jersey℠ Bringing research to life.

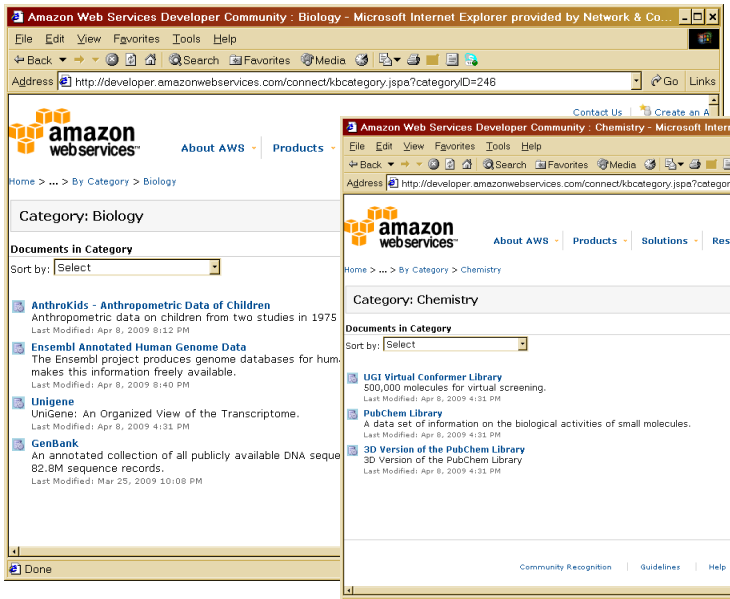


- Based on i2b2 release 1.3
- GDE Consortium is being formed

centocor research & development inc

Infrastructure

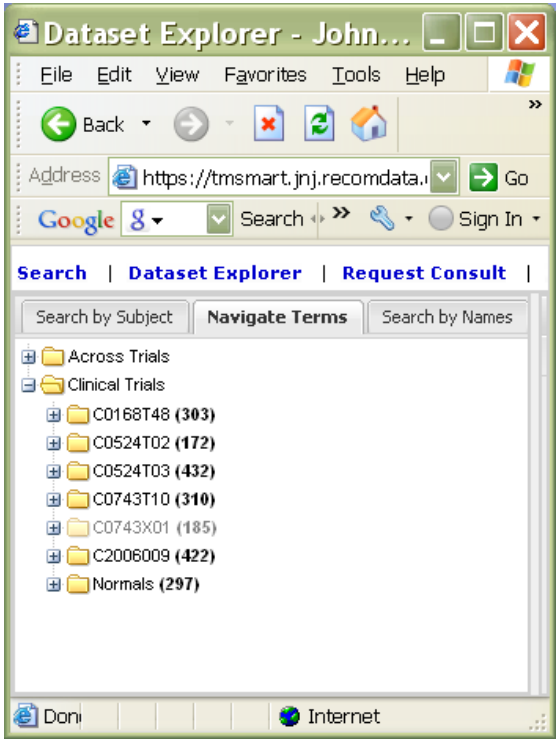
- Amazon Elastic Compute Cloud
  - Security
  - Cost
  - Scalability
  - HPC
  - Amazon data feeds

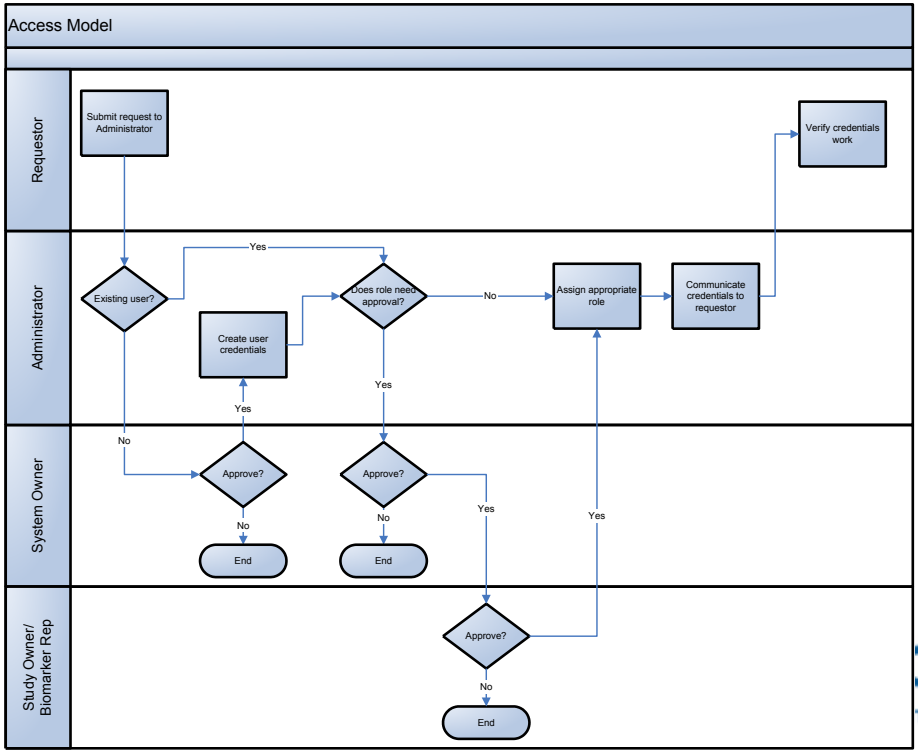


centocor
research & development inc

Fine-grained Security Model

- Strong authentication
- Strong governance – study owners, established process
  - Training
  - Access granting process is established and followed
  - Clinical trial nodes grayed-out if no access is granted



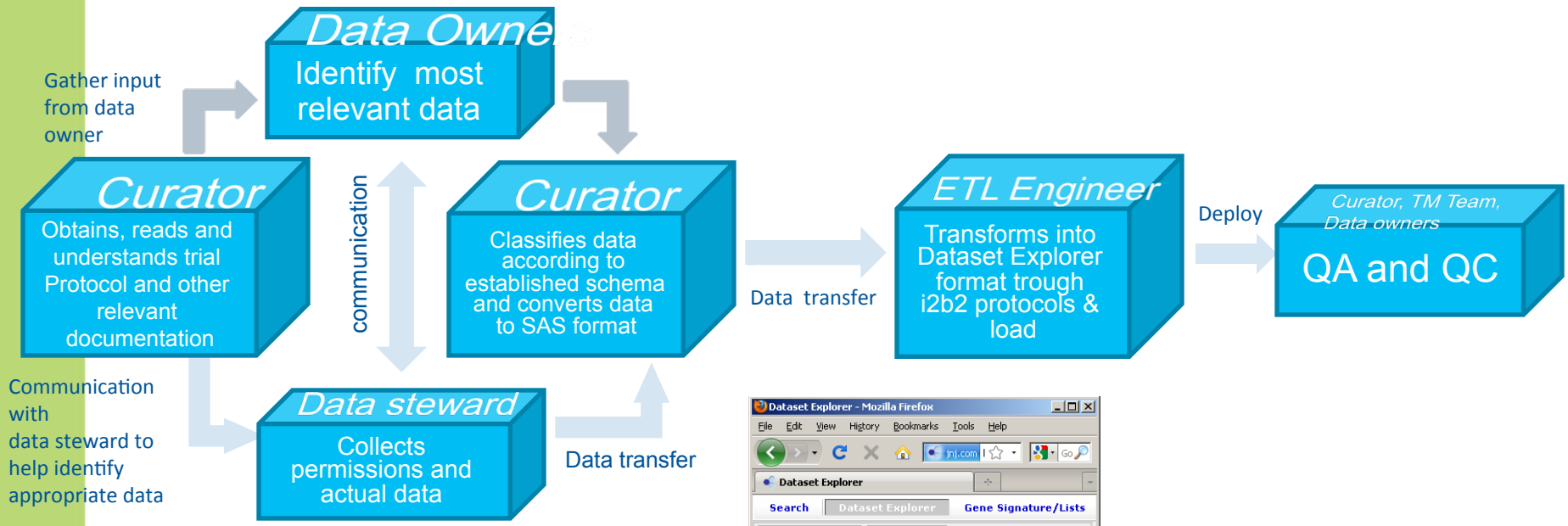


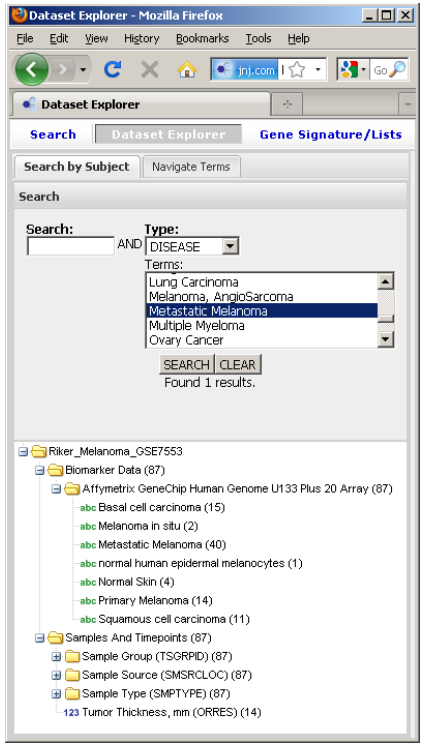
Dictionaries, Ontologies and Master Data

- Gene/protein names and alternatives (Entrez)
- Gene name mapping service (Affy ID, SNP, protein)
- Pathways (GO, KEGG, GeneGo, Ingenuity, Ariadne, MSigDB)
- Internal signatures
- Diseases (MeSH, ICD10)
- Clinical trial observations (MedDRA)
- J&J drugs dictionary
- Clinical trials metadata dictionary
- Cell line dictionary
- Curated inhibitors (CAS id)

centocor
research&development inc

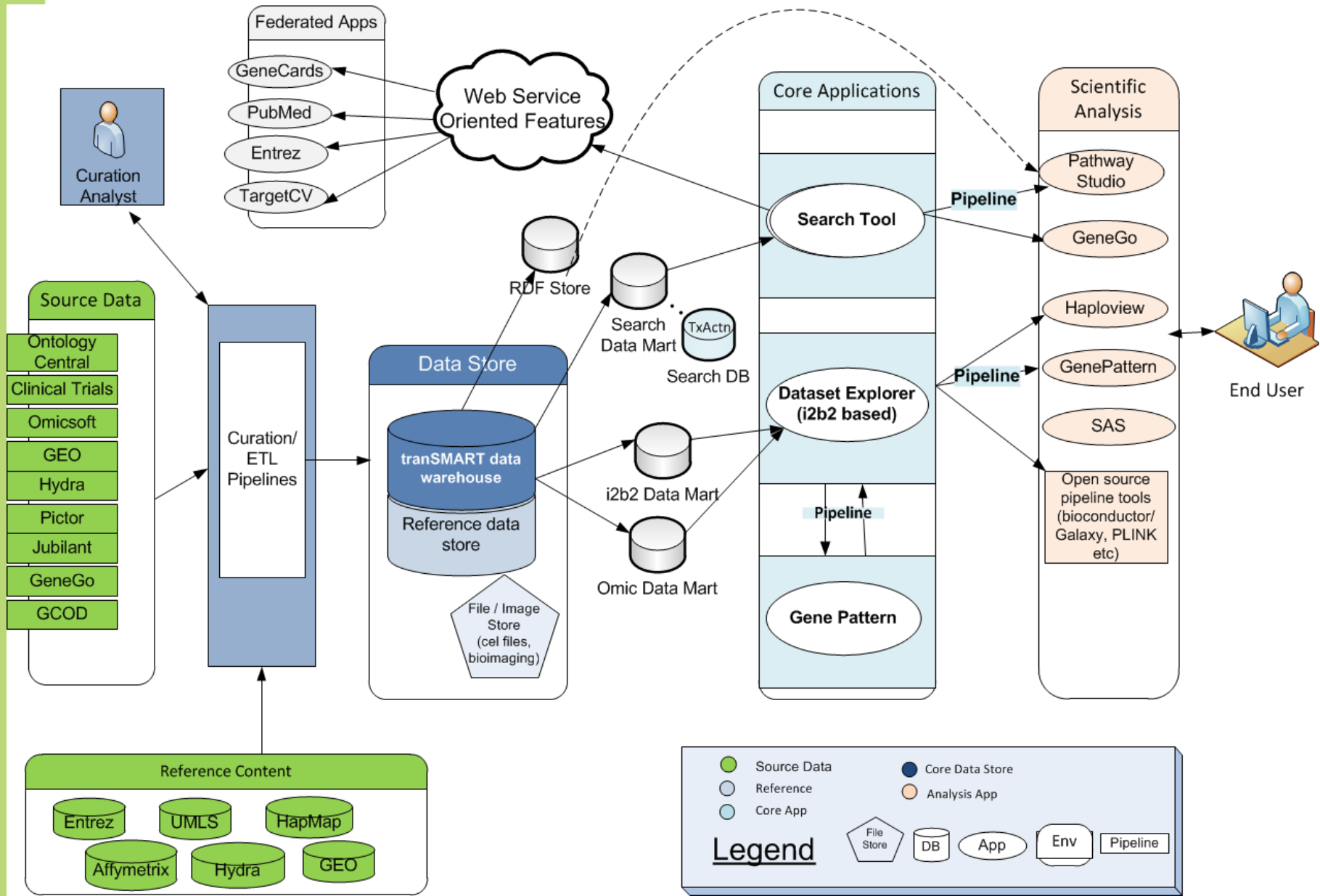
Curation Process for Internal Trials





centocor
research & development inc

Big Scary Diagram



tranSMART Data Warehouse

- Data
  - Structured Data
    - Clinical trials, clinical and pre-clinical gene expression, protein profiling (RBM), SNP, PD markers, metabolomics, proteomics
      - In-house – immunology (large and small mol), oncology, psychiatry
      - Public and commercial
  - Unstructured Data
    - Curated text
    - Text indexing
  - Master data, ontologies, vocabularies and metadata
  - Federated sources
- User Interfaces
  - Search
    - Gene, pathway, disease, compound, trial, and combinations
  - Hypothesis testing
    - Cohort selection and comparison/analysis
  - Hypothesis generation
    - Gene signatures
  - Analytics workflows

JUBILANT BIOSYS

OmicSoft Corporation

ARIADNE

gene

centocor research&development inc

Statistics

- Dataset Explorer
  - 42 internal studies (clinical trials, experimental medicine studies, in vivo and ex vivo experiments)
  - 43 public studies
  - Oncology, immunology, cardiovascular, CNS
- Search
  - 10 internal studies
  - 9,000 GEO/Array Express comparisons
  - 97 DFCI curated studies
  - 100,000 curated biomarker assertions

centocor research&development inc

Platforms and Sources Roadmap

Experimental Platform	Q2 2009	Q3 2009	Q4 2009	Q1 2010	Q2 2010	Q3 2010	Q4 2010	Q1 2011	Q2 2011	Q2 2011
Clinical Trial Data										
Rules-based Medicine										
Affymetrix gene expression										
Candidate SNP										
OmicSoft analyzed mRNA										
ELISA										
Flow cytometry										
Cell counts										
Metabolomics										
Proteomics										
MSD										
Luminex										
MS counts										
Illumina gene expression										
SNP Chip/CNV										
ex vivo										
Histology										
in vivo										
miRNA										
CTC										
Full genome sequencing										
RNA-seq										
PK/Tox										
Epigenetics										
Phosphorylation										
Lipidomics										

Experimental Data Analysis	Q2 2009	Q3 2009	Q4 2009	Q1 2010	Q2 2010	Q3 2010	Q4 2010	Q1 2011	Q2 2011	Q3 2011
Jubilant Oncology										
Pathways										
Beerse preclinical mRNA pipeline										
ClinicalTrials.gov (Link)										
Hydra (Link)										
Entrez (Link)										
GeneCards (Link)										
Google Scholar (Link)										
OmicSoft public mRNA (GEO, EBI)										
La Jolla preclinical mRNA										
Dana-Farber CI oncology mRNA										
Conference abstracts (indexing)										
DIP folder (indexing)										
Jubilant Asthma										
Biomarker folder (indexing)										
Resnet (text mining)										
GeneGo Prostate Cancer Content										
ClinicalTrials.gov (text mining)										
GeneGo Disease Pages										
dbSNP										
P53 mutation DB										
COSMIC Link										
J&J Biobank Link										
Histology (Aperio) Link										
GVK Bio biomarkers										
Broad Connectivity Map										
Cancer Gene Data Curation										
Prous Integrity										
GeneGo asthma maps										
AsthmaMap										
OASIS										
Cellminer										

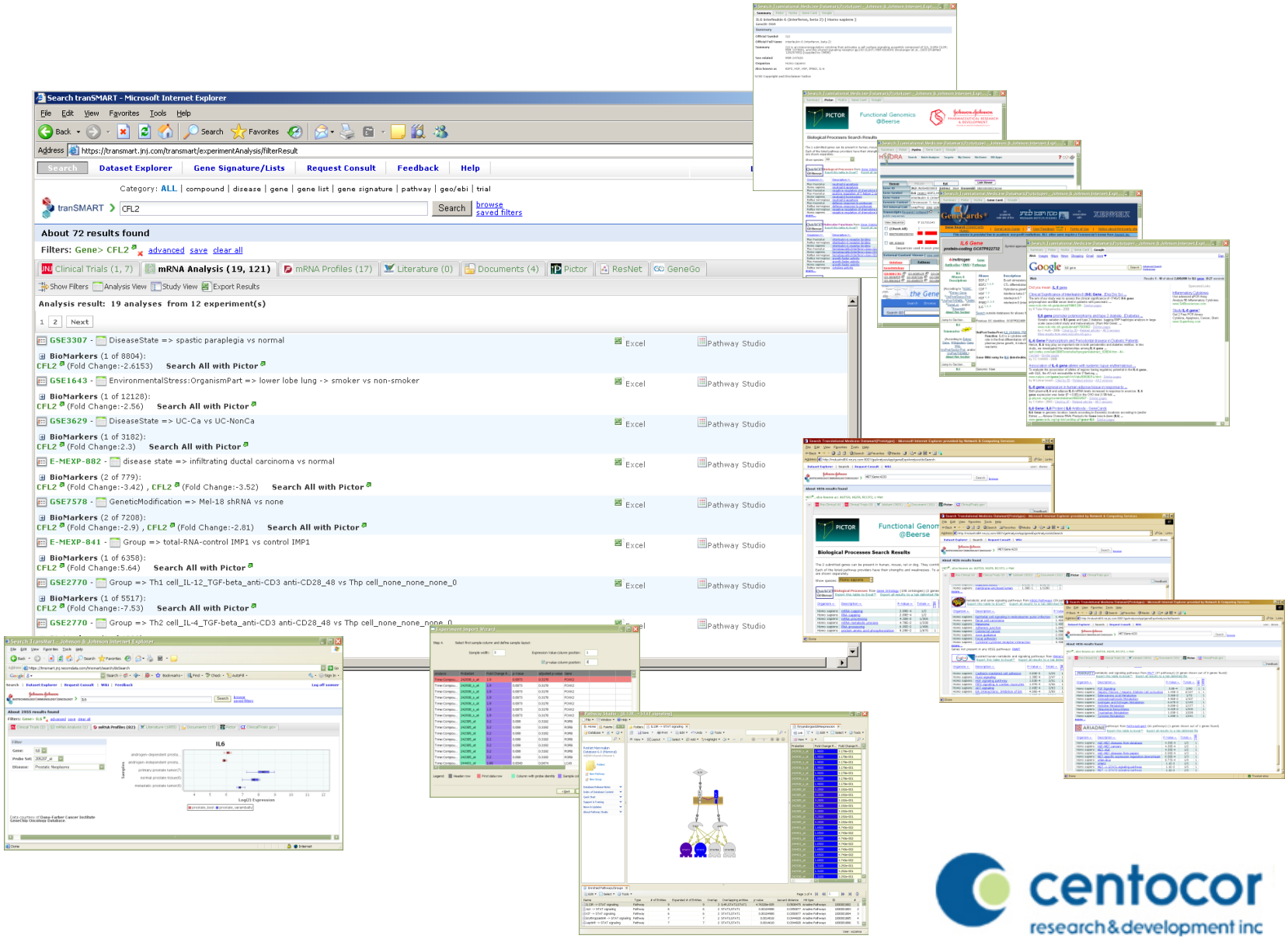
centocor
research&development inc

Use Cases

- Knowledge Management/Search
- Hypothesis Testing
- Hypothesis Generation

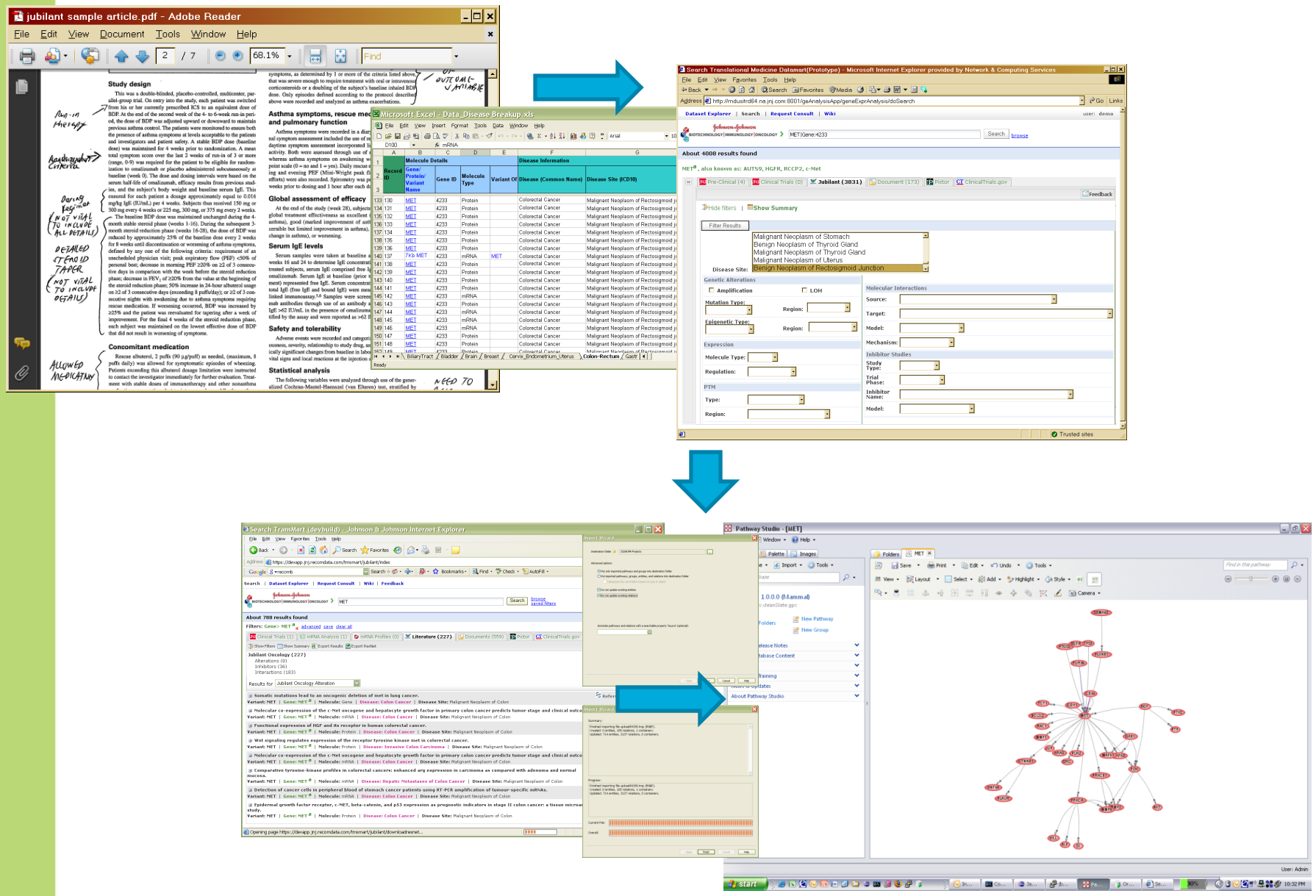
centocor
research & development inc

Search

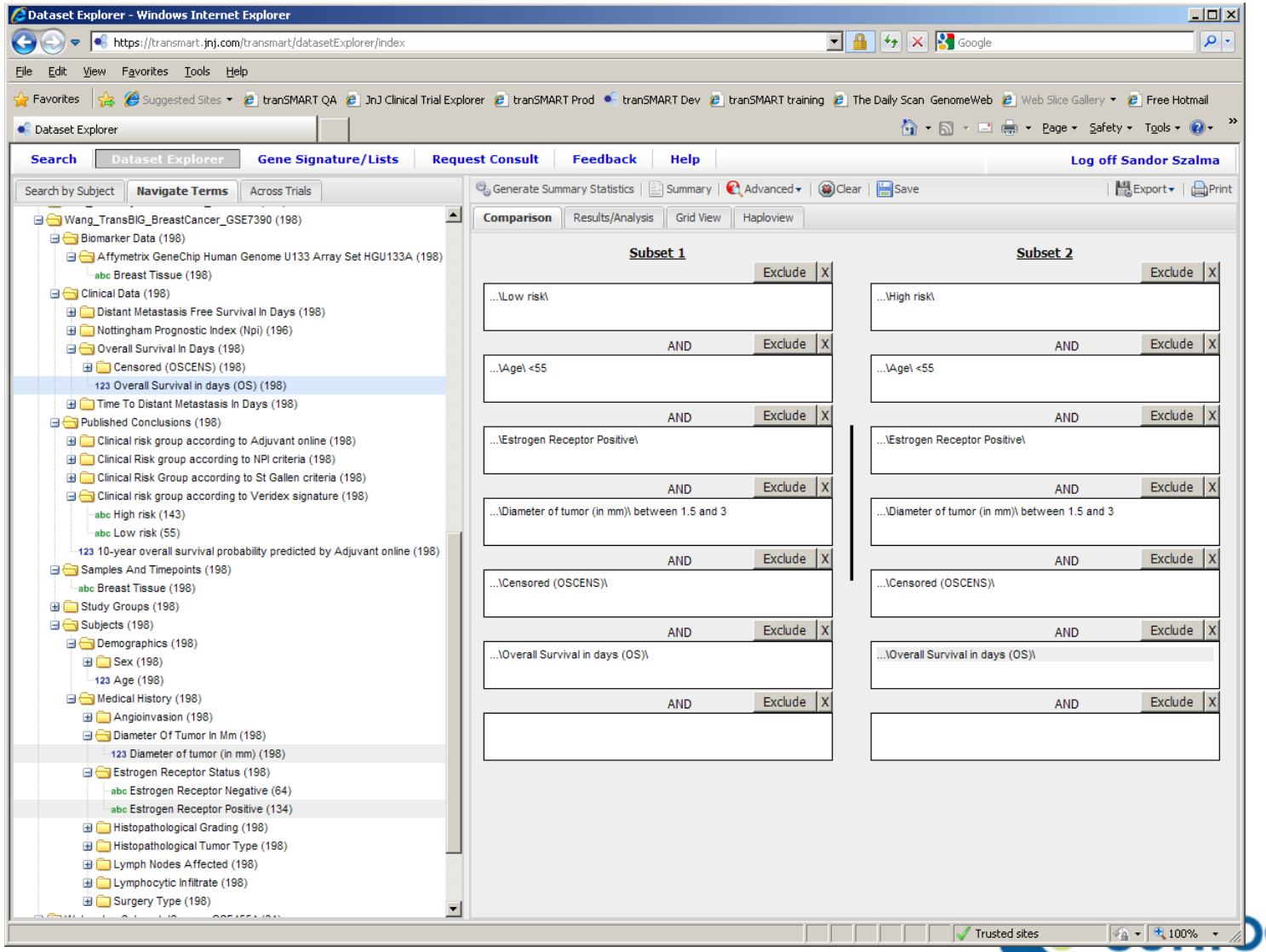


centocor
research&development inc

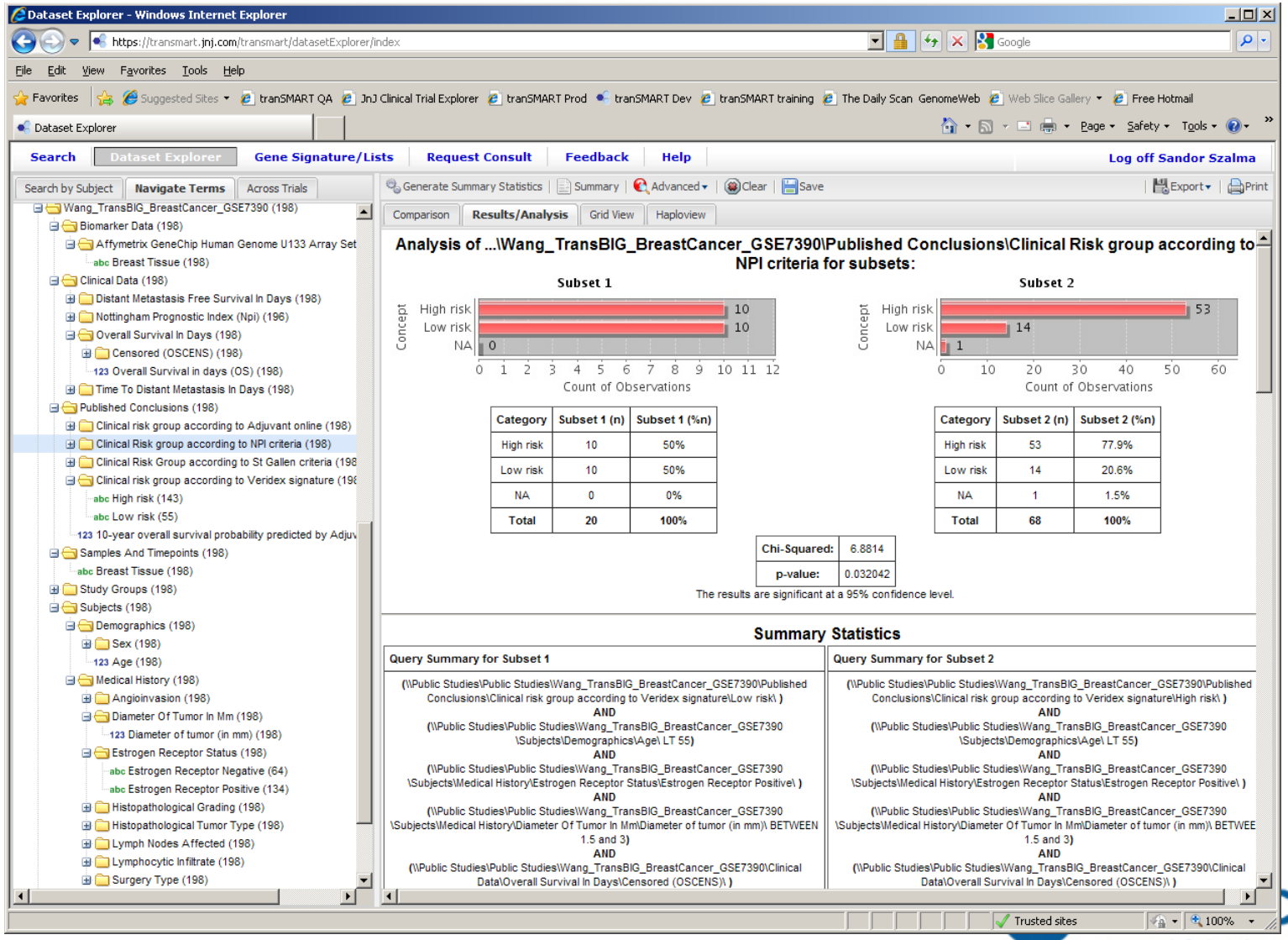
Curated Content in Semantic Networks



Hypothesis Testing – Flexible Queries



Hypothesis Testing – Simple Statistics



Hypothesis Testing – Advanced Statistics

Survival Analysis - Windows Internet Explorer

Survival Analysis

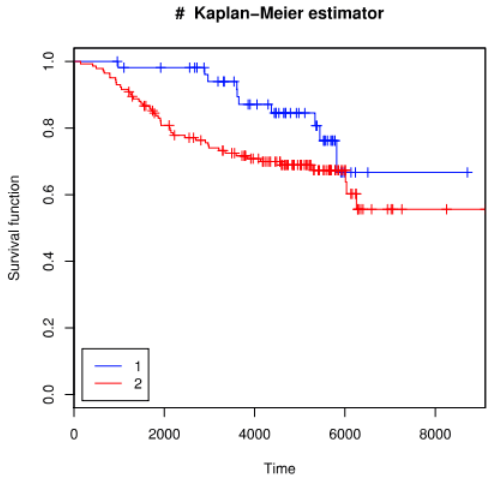
Subset 1 Query	Subset 2 Query
(\Public Studies\Wang_TransBIG_BreastCancer_GSE7390\Published Conclusions\Clinical risk group according to Veridex signature\Low risk\ ) AND (\Public Studies\Wang_TransBIG_BreastCancer_GSE7390\Clinical Data\Overall Survival In Days\Censored (OSCENS)\ ) AND (\Public Studies\Wang_TransBIG_BreastCancer_GSE7390\Clinical Data\Overall Survival In Days\Overall Survival in days (OS)\ )	(\Public Studies\Wang_TransBIG_BreastCancer_GSE7390\Published Conclusions\Clinical risk group according to Veridex signature\High risk\ ) AND (\Public Studies\Wang_TransBIG_BreastCancer_GSE7390\Clinical Data\Overall Survival In Days\Censored (OSCENS)\ ) AND (\Public Studies\Wang_TransBIG_BreastCancer_GSE7390\Clinical Data\Overall Survival In Days\Overall Survival in days (OS)\ )

Cox Regression Result:

Number of Subjects	198
Hazard Ratio (95% CI)	1.94 (0.98 - 3.86)
Relative Risk (p Value)	0.664 (0.057)

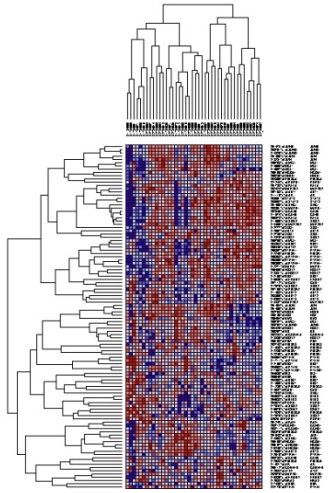
Survival Curve Fitting Summary:

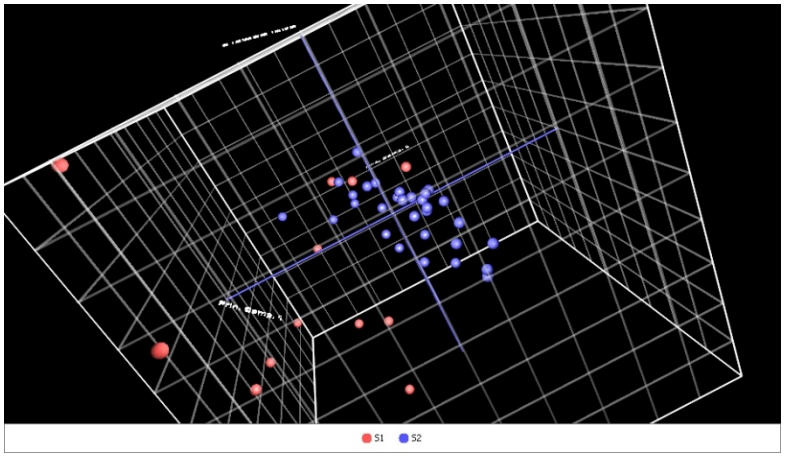
Subset	Number of Subjects	Number of Events	Median Value	Lower Range of 95% Confidence Level	Upper Range of 95% Confidence Level
Subset 1	55	10	infinity	5816	infinity
Subset 2	143	46	infinity	6036	infinity

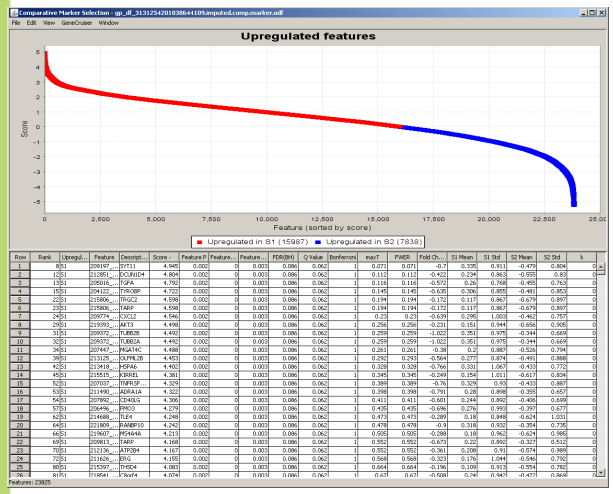


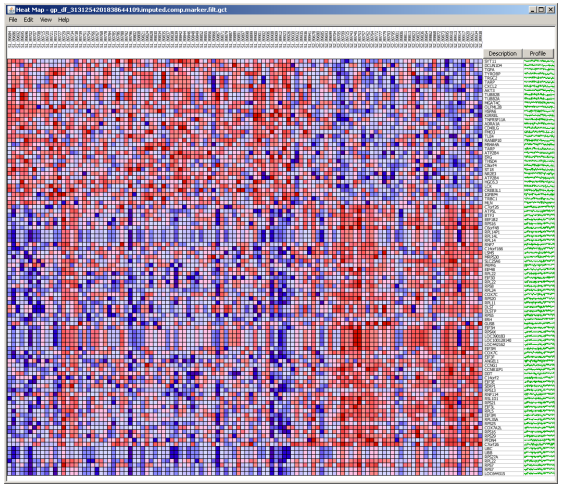
ntocor
h&development inc

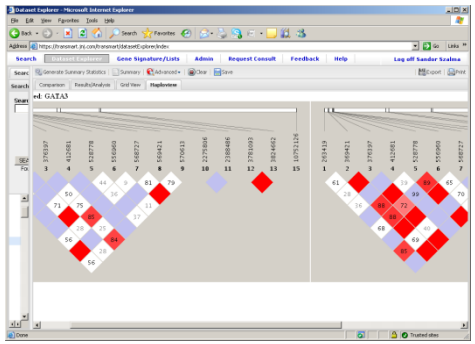
Advanced Biomarker Analysis





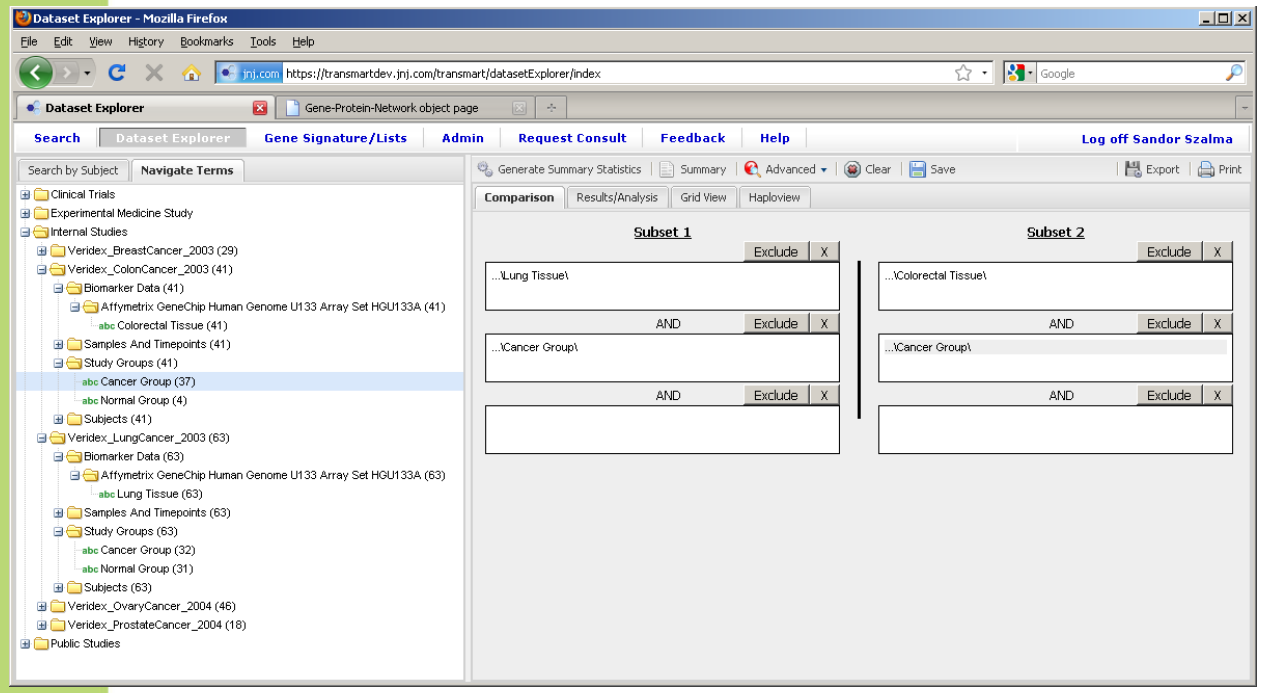


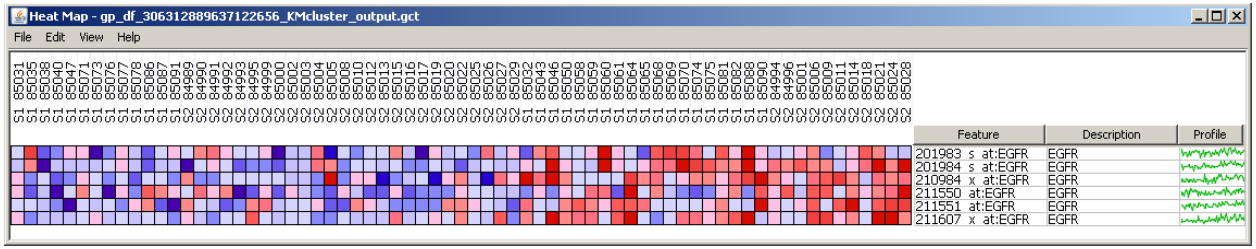




centocor
research & development inc

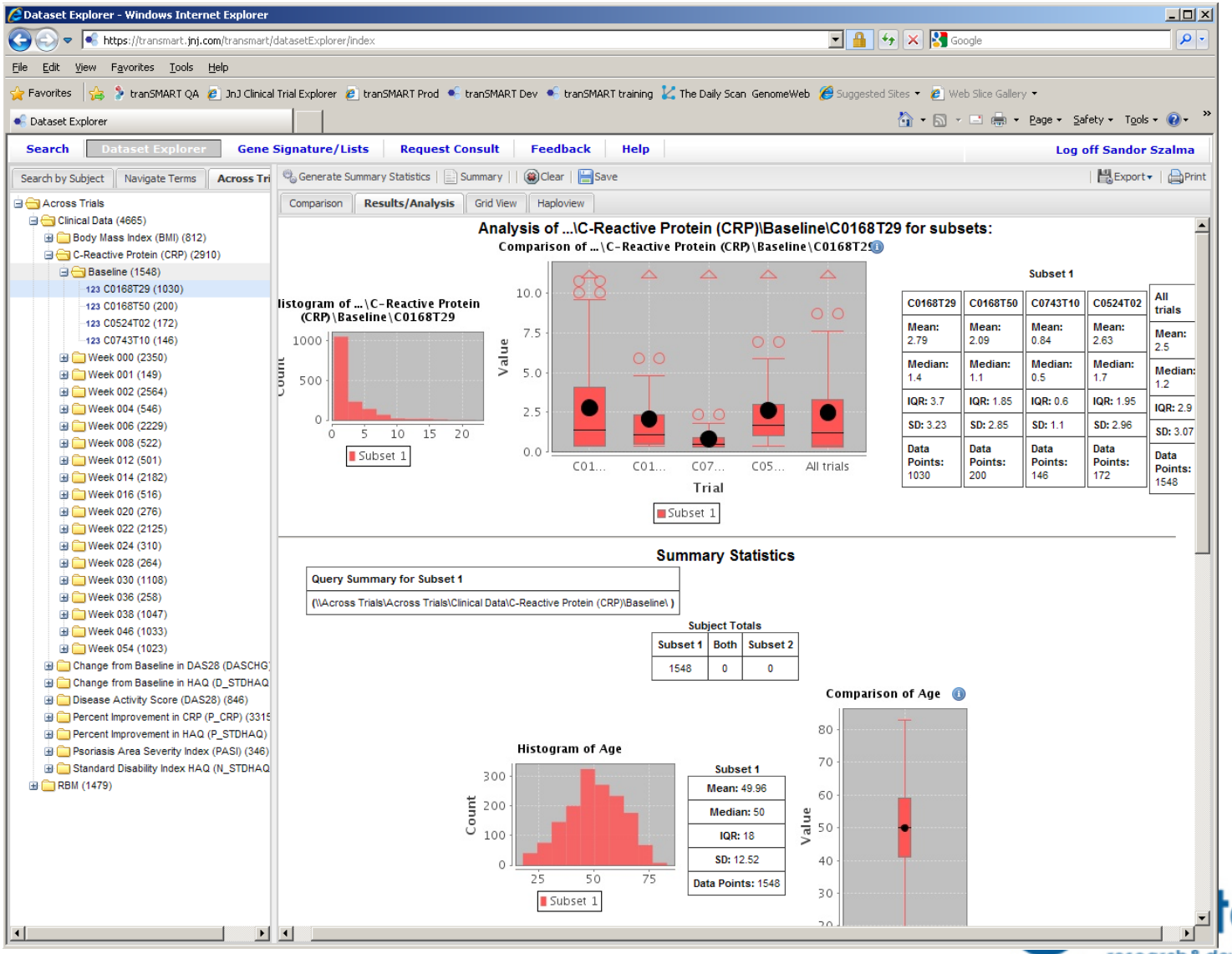
Meta-analysis





centocor
research&development inc

Cross-Study Queries

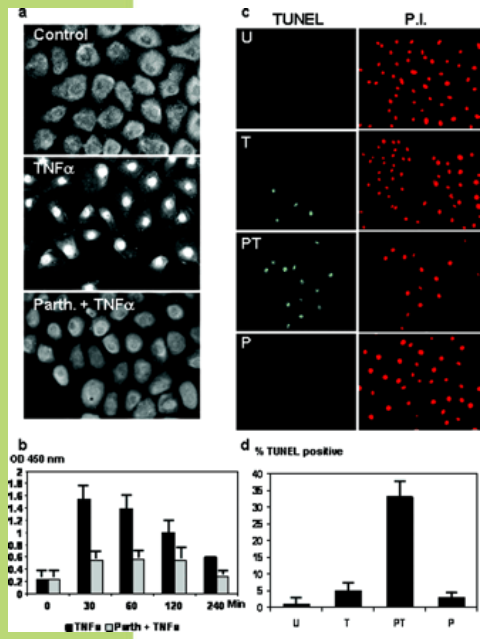


Indication Hypothesis Generation

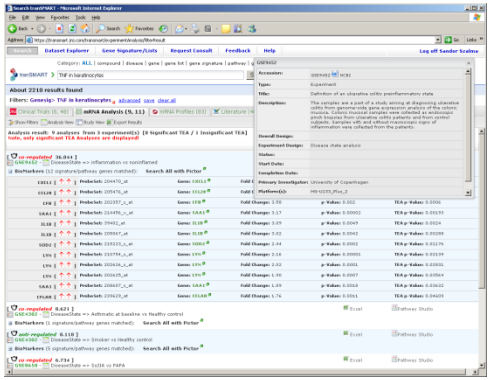
- Disease Profiles Database
  - Well curated
  - Well annotated
  - QCd
    - Database of 9,000 public comparisons from NCBI GEO and EBI Array Express
    - Subset is molecular profiles of disease vs. normal samples
  - Also, proprietary data!
- Drug or Pathway Modulation Signatures
  - In vitro, in vivo or clinical samples
  - Treatment vs. placebo or before vs. after or …
    - Drug, siRNA, Ab, KO, …
  - Measure gene expression, protein expression, …
  - Statistical analysis and derivation of statistically significant list of genes (proteins, …) with direction

centocor research&development inc

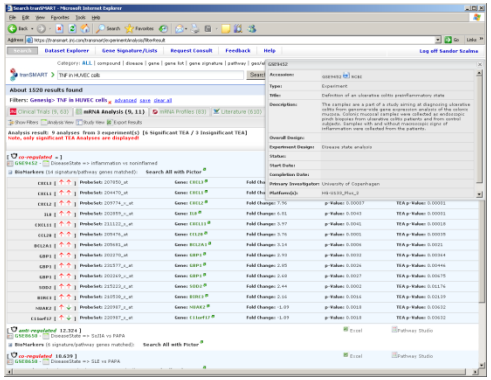
Indication Hypothesis Generation



in vitro
ex vivo
in vivo



TEA in tranSMART



Disease	Keratinocyte	HUVEC	Rat
TNF Stimulation	40	40	40
IBD	40	40	40
Crohn's disease	40	40	40
Ulcerative colitis	40	40	22.7
TNF Stimulation	40	40	40
Crohn's disease	40	40	40
TNF Stimulation	40	40	40
Inflammation	40	40	40
Mycobacterium tuberculosis	40		12.6
Influenza A	33.4	40	30.9
RA	32.7	40	27.5
TNF Stimulation	32.2	35.6	25.7
Mycobacterium tuberculosis	32.2	33.8	13.1
Ulcerative colitis	30.2	40	19.8
RA	27.9	40	20.2
TNF Stimulation	24.8	23.6	12.1
TNF Stimulation	23.4	22.4	11.4
Infectious Colitis	22.4	36.7	15.6
Dermatomyositis	18.4	20.6	6
Crohn's disease	17.3	34.4	15.3
Osteoarthritis	16.8	13.4	
RA	16.1	14	
Asthma	16.1		
Influenza A	13	24.5	15
Ulcerative colitis	12.4	8	6.2
RA		21.6	19.4
TNF Stimulation		21.3	

Manuscript in preparation.

centocor
research&development inc

Implemented Disruptive Technology

- Open source
- Cloud computing
- Pre-competitive sharing

- Informatics/IT
- Recombinant Data
- Business partners across multiple TAs (Biology, Biomarkers, Translational Medicine, Clinical)
- Academic partners