ABRF Conference 2014

Learnings from Albuquerque...
ASSOCIATION OF BIOMOLECULAR RESOURCE FACILITIES

• The Association of Biomolecular Resource Facilities (ABRF) is dedicated to advancing core and research biotechnology laboratories through research, communication, and education.

• Initially started as an Association of protein researchers – but has now expanded into Genomics and Metabolomics Research.

• ABRF members include over 800 scientists representing 267 different core laboratories in 41 countries, including those in industry, government, academic and research institutions.

• A relatively closely-knit group (as compared to ASMS) with associations spread over decades.
ABRF 2014 HIGHLIGHTS

- Skyline Workshop.
- Next Generation Proteomics by Albert Heck.
- Proteomics Research Group
- Multi-omics from Snyder Lab
- ENCODE Project
- From Binder Clips to Lasers
- Digital Enterprise: Philip Bourne
- Conclusions
Quantitative Proteomics Analysis using Skyline and R Statistical Computing Software

Skyline is a freely-available Windows client application for building Selected Reaction Monitoring (SRM) / Multiple Reaction Monitoring (MRM), Parallel Reaction Monitoring (PRM - Targeted MS/MS and DIA/SWATH) and targeted DDA with MS1 quantitative methods and analyzing the resulting mass spectrometer data.

https://brendanx-uw1.gs.washington.edu/labkey/project/home/software/Skyline/begin.view
Quantitative Proteomics Analysis using Skyline and R Statistical Computing Software

https://brendanx-uw1.gs.washington.edu/labkey/project/home/software/Skyline/begin.view
Skyline v2.5 Released on 2/8/2014
Multi-peptide chromatogram graphs - use protein selection or multiple selection (shift- or ctrl-click) to see annotated chromatogram peaks for many peptides at once

Skyline v2.1 Released on 9/8/2013
Support for building chromatogram libraries on Panorama and using them in Skyline
Demultiplexing of overlapped DIA/SWATH methods
Automatic Python installation
Support for MSstats with new GUI form [download]
Improved integration with PanoramaWeb
MSGF+ pepXML/mzXML

Skyline v1.4 Release Updated on 3/18/2013
Support for building spectral libraries from MaxQuant Andromeda msms.txt

Skyline v1.4 Release Updated on 12/17/2012
Support for building spectral libraries from PRIDE XML
Improved retention time alignment features for MS1 filtering

Skyline v1.3 Released on 6/20/2012
Advanced support for data independent acquisition (DIA) across vendors:
AB SCIEX SWATH™

Skyline v1.1 Released on 6/11/2011
Import results from WIFF files much improved (50-fold faster for large scheduled runs - thanks to AB SCIEX)
Full-scan method export for Thermo LTQ
Integrated support for Unimod modification definitions
Modification auto-detect support for pasted/inserted annotated peptide sequences
Native method export for AB SCIEX QTRAP
Spectral library build support for Scaffold, Waters MS^e and OMSSA
Manuscript ready charts
SpectraST library support
Summary result statistics in reports

Skyline v0.5 Preview Released on 5/30/2009
Scheduled and unscheduled transition list support for instruments from: Agilent, Applied Biosystems, Thermo Fisher, Waters…

Skyline v0.2 Released on 2/17/2009
Retention time prediction
Spectral libraries (NIST, GPM, BiblioSpec)
Building spectral libraries from your results in TPP pepXML/mzXML
Skyline is getting popular and powerful in MS/MS and MRM analysis.

Plan for a Skyline Learning Session?
Plan for a Skyline Learning Session?
Next Generation Proteomics by Albert Heck.
Proteomics Research Group
Multi-omics from Snyder Lab
ENCODE Project
From Binder Clips to Lasers
Digital Enterprise : Philip Bourne
Conclusions
“Exposing the Proteome in Full Glory Through Advances in Enabling Technologies"
Albert Heck.

‘The rapid development of high-throughput technologies in the past decade, which is linked to a reduction in their costs, opens up new possibilities to interrogate a biological system at multiple regulatory levels and simultaneously offers us an unprecedented vision.’

New proteoforms.
Quantification of the entire set of proteins expressed in a complex biological system.
In combination with new approaches to isolate specific PTMs, MS-based studies are revealing a much higher order of proteome complexity in which most proteins are modified by several PTMs that crosstalk in intricate mechanisms to regulate protein function.

ABRF 2014 HIGHLIGHTS

• Plan for a Skyline Learning Session?
• Next Generation Proteomics: Heck of a Talk!
• **Proteomics Research Group**
• Multi-omics from Snyder Lab
• ENCODE Project
• From Binder Clips to Lasers
• Digital Enterprise: Philip Bourne
• Conclusions
Research Groups are established to
A) provide mechanisms for the self-evaluation and improvement of procedural and operational accuracy, precision and efficiency in resource facilities and research laboratories.
B) to contribute to the education of resource facility and research laboratory staff, users, administrators, and interested members of the scientific community.

Antibody Technology Research Group (ARG)
DNA Sequencing Research Group (DSRG)
Flow Cytometry Research Group (FCRG)
Genomics Research Group (GVRG)
Glycoprotein Research Group (gPRG)
Light Microscopy Research Group (LMRG)
Metabolomics Research Group (MRG)
Molecular Interactions Research Group (MIRG)
Nucleic Acids Research Group (NARG)
Protein Expression Research Group (PERG)
Protein Sequencing Research Group (PSRG)
Proteomics Research Group (PRG)
Proteomic Informatics Research Group (iPRG)
Proteomic Standards Research Group (sPRG)
### ABRF SPRG 2013 Study Results

#### Table of Results

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Manufacturer</th>
<th>Instrument</th>
<th>Software</th>
<th>Number Of Peptides</th>
<th>Log2Peak Area Ratio Correlation With Consensus Area Ratio</th>
<th>Number Of Heavy Peptides (Retention Time)</th>
<th>Retention Time Correlation With Mean Percentile Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS6842-01</td>
<td>Thermo</td>
<td>Orbitrap Elite</td>
<td>MaxQuant</td>
<td>764</td>
<td>0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MFIG13</td>
<td>Thermo</td>
<td>Orbitrap Velos</td>
<td>MaxQuant</td>
<td>718</td>
<td>0.78</td>
<td>718</td>
<td>0.98</td>
</tr>
<tr>
<td>56365A</td>
<td>Thermo</td>
<td>Orbitrap Velos</td>
<td>MaxQuant</td>
<td>639</td>
<td>0.27</td>
<td>726</td>
<td>1.00</td>
</tr>
<tr>
<td>BMS161</td>
<td>Thermo</td>
<td>Progenesis</td>
<td></td>
<td>395</td>
<td>0.95</td>
<td>395</td>
<td>1.00</td>
</tr>
<tr>
<td>AS6842-02</td>
<td>Thermo</td>
<td>Orbitrap Elite</td>
<td>ProteomeDiscoverer</td>
<td>518</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FR5411_1</td>
<td>Thermo</td>
<td>Orbitrap Classic</td>
<td>ProteomeDiscoverer</td>
<td>492</td>
<td>0.93</td>
<td>914</td>
<td>0.96</td>
</tr>
<tr>
<td>12358f</td>
<td>Bruker</td>
<td>Impact</td>
<td>Skyline</td>
<td>668</td>
<td>0.79</td>
<td>786</td>
<td>1.00</td>
</tr>
<tr>
<td>02940a</td>
<td>Thermo</td>
<td>LTQ-FT</td>
<td>Skyline</td>
<td>542</td>
<td>0.51</td>
<td>707</td>
<td>0.71</td>
</tr>
<tr>
<td>1014os</td>
<td>Thermo</td>
<td>LTQ-FT</td>
<td>Skyline</td>
<td>485</td>
<td>0.13</td>
<td>90</td>
<td>0.60</td>
</tr>
<tr>
<td>80232a</td>
<td>Thermo</td>
<td>Orbitrap Classic</td>
<td>Skyline</td>
<td>399</td>
<td>0.12</td>
<td>399</td>
<td>0.30</td>
</tr>
<tr>
<td>p7iacd</td>
<td>Thermo</td>
<td>Orbitrap Elite</td>
<td>Skyline</td>
<td>487</td>
<td>0.66</td>
<td>523</td>
<td>0.99</td>
</tr>
<tr>
<td>1e356x</td>
<td>Thermo</td>
<td>Orbitrap Elite</td>
<td>Skyline</td>
<td>756</td>
<td>0.85</td>
<td>932</td>
<td>0.99</td>
</tr>
<tr>
<td>291013</td>
<td>Thermo</td>
<td>Orbitrap Velos</td>
<td>Skyline</td>
<td>575</td>
<td>0.94</td>
<td>667</td>
<td>1.00</td>
</tr>
<tr>
<td>814277</td>
<td>Thermo</td>
<td>Orbitrap</td>
<td>Skyline</td>
<td>659</td>
<td>0.88</td>
<td>718</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Brian Searle:

- We are reasonably reproducible in RT, intensity and identification.
- We do not estimate true ratios well.
- And we are too willing to over report.
- Brian also mentioned interference by "chimeracy" (or mixture peptides) in true quantitation studies in discovery proteomics.

Questions were raised on how well can SILAC or MS1 results be trusted given that only 300 / 1000 peptides correlated well across labs and MS platforms.

http://tinyurl.com/ABRF-SPRG

- ProteoMonitor: ABRF sPRG Study Finds Field is Precise but Still Struggles with Accuracy in Quantitative Proteomics

- "Results of the sPRG 2013 Study for 1000 "Spiketides" available at http://tinyurl.com/ABRF-SPRG Consistent IDs (RT); Not Quant (Ratios). #ABRF2014" Tweet by @proteomics
Christopher Colangelo - Yale University.
Brian Searle – ProteomeSoftware.

“Calling all ABRF Proteomics/Mass spectrometry. Help reorganize the sPRG, PRG, and iPRG”
ABRF 2014 HIGHLIGHTS

• Plan for a Skyline Learning Session?
• Next Generation Proteomics: Heck of a Talk!
• Proteomics Research Group needs members
• Multi-omics from Snyder Lab
• ENCODE Project
• From Binder Clips to Lasers
• Digital Enterprise: Philip Bourne
• Conclusions
Multi-omics - Dynamic Omics Methods for Personalized Medicine.

iPOP: Integrated Personal Omics Profiling.

http://snyderome.stanford.edu/

Rui Chen
George Mias
Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes.

*Cell 148:1293-1307*

- Physiological states analyzed by integrative personal omics profiling
- Extensive molecular changes revealed during different health states
- Individual disease risk predicted from integrated omics data

Personalized medicine is expected to benefit from combining genomic information with regular monitoring of physiological states by multiple high-throughput methods. *Here, we present an integrative personal omics profile (iPOP), an analysis that combines genomic, transcriptomic, proteomic, metabolomic, and autoantibody profiles from a single individual over a 14 month period.* Our iPOP analysis revealed various medical risks, including type II diabetes. It also uncovered extensive, dynamic changes in diverse molecular components and biological pathways across healthy and diseased conditions. Extremely high-coverage genomic and transcriptomic data, which provide the basis of our iPOP, discovered extensive heteroallelic changes during healthy and diseased states and an unexpected RNA editing mechanism. This study demonstrates that longitudinal iPOP can be used to interpret healthy and disease states by connecting genomic information with additional dynamic omics activity.
Personal genomes, quantitative dynamic omics and personalized medicine

Quantitative Biology
March 2013, Volume 1, Issue 1, pp 71-90
PROTEINS

Quantitative Mass Spectrometry Sample Workflow

- Quantitation sample preparation
  SILAC, iTRAQ, TMT, label-free

- Data annotation/consolidation
  Uniprot/NCBI annotation

- Normalization
  Ratios (μ = 1); distinct runs

- Quality control, false discovery rate estimations

- MS/MS spectra identification
  X!Tandem, SEQUEST, OMMSA, Mascot

- Convert to .mzML for open source MSConvert, TPP

METABOLITES

Mass Spectrometry Sample Workflow

- Gas or liquid chromatography - mass spectrometry profiling
  (GC/LC-MS)

- Annotation
  PubChem, KEGG, Metlin, MetaCyc, Reactome

- Quality control
  Retention time filtering, average replicates, Id missing data

- Align mass and retention time data
  XCMS, SIEVE, Matlab, MassHunter Profiler, MzMine

- Converted to .mzML for open source MSConvert

The diagram illustrates the workflow for proteins and metabolites in quantitative mass spectrometry.
integrative Personal Omics Profiling

I. RISK EVALUATION

- Whole genome sequencing
- Disease risk evaluation
- Medical history & environment
- Pharmacogenomic evaluation

II. LONGITUDINAL OMICS PROFILING OF MULTIPLE PHYSIOLOGICAL STATES

- Transcriptomics
- Proteomics
- Metabolomics
- Clinical tests
- Autoantibodyomics
- Microbiomics
- New omics

Healthy → Infected → Recovery → Healthy

III. INTEGRATION OF MULTIPLE OMICS AND TEMPORAL RESPONSES MATCHED AGAINST iPOP DATABASES

- iPOP database
...Continues to collect data on during viral infections, normal days, diabetes and after exercise...
...Continues to collect data on microbiomes from skin, saliva, etc.
...Will be extended to 10 more individuals.
ABRF 2014 HIGHLIGHTS

• Plan for a Skyline Learning Session?
• Next Generation Proteomics: Heck of a Talk!
• Proteomics Research Group needs members
• Multi-omics on Snyder = Synderomics
• **ENCODE Project**
• From Binder Clips to Lasers
• Digital Enterprise: Philip Bourne
• Conclusions
Accessing and Using ENCODE Data

Peggy J. Farnham

Citation: Farnham, P. (2013), "Accessing and Using ENCODE Data", in Carey, M. and Smale, S. (eds), Eukaryotic Gene Regulation: From Chromatin to Transcription to mRNA Processing, The Biomedical & Life Sciences Collection, Henry Stewart Talks Ltd, London (online at http://hstalks.com/?t=BL0083641-Farnham)
The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

http://hstalks.com/?t=BL0083641-Farnham

http://www.genome.gov/10005107
How many human genes are encoded in our $3 \times 10^9$ bp?

- *C. elegans* (worm)
  - 959 cells and $1 \times 10^8$ bp
  - 20,000 genes

First guess for human genes
- 100,000-150,000

2001 genome draft
- up to 40,000 genes

2012: ~20,000 genes

http://hstalks.com/?t=BL0083641-Farnham
ENCODE 3 Consortium

Production Groups
1. Broad Institute
2. Cold Spring Harbor; Centre for Genomic Regulation (CRG);
3. University of Connecticut Health Center; UConn
4. HudsonAlpha; Pennsylvania State; UC Irvine; Duke; Caltech
5. UCSD; Salk Institute; Joint Genome Institute; Lawrence Berkeley National Laboratory; UCSD
6. Stanford; University of Chicago; Yale
7. University of Washington; Fred Hutchinson Cancer Research Center; University of Massachusetts Medical School

Technology Development Groups
1. MIT
2. Washington University; St. Louis
3. USC; Ohio State University; UC, Davis
4. University of Washington
5. Sloan-Kettering; Weill Cornell Medical College
6. Princeton; Weizmann
7. University of Michigan
8. Broad Institute
9. University of Washington; UCSC
10. Advanced RNA Technologies, LLC
11. Harvard

Computational Analysis Groups
1. Berkeley; Wayne State University
2. MIT
3. University of Wisconsin
4. Sloan-Kettering; Broad Institute
5. Stanford
6. UCLA

Affiliated Groups
1. Wellcome Trust Sanger Institute
2. Florida State University

http://hstalks.com/?t=BL0083641-Farnham
ENCODE ASSAYS

3D looping

Open chromatin

Gene expression

TF binding

http://hstalks.com/?t=BL0083641-Farnham

© 2014 Regents of the University of Minnesota. All rights reserved.
An integrated encyclopedia of DNA elements in the human genome.

Nature 489: 57-74, 2012

Jan 2011
data freeze

http://hstalks.com/?t=BL0083641-Farnham
Publications Using ENCODE Data

- Papers from Non-ENCODE Authors
- Papers from ENCODE 2 Production Groups

http://hstalks.com/?t=BL0083641-Farnham
The Encyclopedia of DNA Elements (ENCODEx) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cell and circumstances in which a gene is active.

ENCODEx data are now available for the entire human genome. All ENCODE data are free and available for immediate use via:

- Search for displayable tracks and downloadable files
- Visualization in the UCSC Genome Browser (ENCODEx data marked with the NHGRI logo)

To search for ENCODEx data related to your area of interest and set up a browser view, use the UCSC Experiment Matrix or Track Search tool (Advanced features). The Experiment List (Human) and Experiment List (Mouse) links provide comprehensive listings of ENCODEx data that is released or in preparation.

All ENCODEx data is freely available for download and analysis. However, before publishing research that uses ENCODEx data, please read the ENCODEx Data Release Policy, which places some restrictions on publication use of data for nine months following data release. Read more about ENCODEx data at UCSC.

http://www.genome.ucsc.edu/ENCODEx/

http://hstalks.com/?t=BL0083641-Farnham
Protein coding transcripts

- There are 20,687 coding genes
- Average of 6.3 transcripts/gene
- Average of 3.9 different proteins/gene
- Longest transcript: 100,272 nt
- Most spliced isoforms: 65

http://hstalks.com/?t=BL0083641-Farnham
Summary: Accessing and Using ENCODE Data

- What is ENCODE
  www.genome.gov/10005107

- What has been learned so far
  www.nature.com/encode/

- How to access ENCODE data
  www.genome.ucsc.edu/ENCODE/

- How to use ENCODE data
  www.genome.ucsc.edu/ENCODE/usageResources.html

http://hstalks.com/?t=BL0083641-Farnham
ABRF 2014 HIGHLIGHTS

- Plan for a Skyline Learning Session?
- Next Generation Proteomics: Heck of a Talk!
- Proteomics Research Group needs members
- Multi-omics on Snyder = Synderomics
- ENCODE Datasets: A hypothesis generating toolkit
- From Binder Clips to Lasers
- Digital Enterprise: Philip Bourne
- Conclusions
"From Binder Clips to Lasers: Creating Technology over Four Decades"

Patrick O'Farrell, Professor of Biochemistry & Biophysics, School of Medicine, University of California, San Francisco.
“There were others who had tried it – but I succeeded because I had a curious blend of stubbornness, confidence and optimism that led me to believe that I knew exactly what was needed to improve the separation no matter the failures and the ugly looking results along the way. In other words – I was young!”

- O'Farrell #ABRF2014

https://conf.abrf.org/sites/default/files/docs/08-12_proteomics.pdf
The Development of Two-dimensional Electrophoresis by Patrick H. O’Farrell

High Resolution Two-dimensional Electrophoresis of Proteins

Patrick H. O’Farrell received his B.Sc. in 1969 from McGill University in Montreal, Quebec. He then went on to graduate school at the University of Colorado, Boulder, where he worked with Jacques Pène. At Boulder, O’Farrell isolated a series of mutations affecting development in the colonial alga, Volvox. His goal was to define the normal pattern of gene expression during development and the effects that the mutations had on this regulatory program. Similar studies had been done in bacteriophage using sodium dodecyl sulfate gel electrophoresis to separate, identify, and quantify the proteins. However, Volvox has approximately 100 times the genetic complexity of bacteriophage and thus needed a separation system with a much higher resolution. Although an adequate system did not exist, two-dimensional methods for increased resolution in chromatography and electrophoresis had been well established. So, O’Farrell recalls, “I set out to combine the two most powerful electrophoresis methods available” (1). This is the subject of the Journal of Biological Chemistry (JBC) Classic reprinted here.

O’Farrell realized that to optimize resolution, he would need to separate the proteins according to independent parameters. He decided to use isoelectric focusing in the first dimension and sodium dodecyl sulfate electrophoresis in the second dimension. This permitted the simultaneous determination of molecular weight and isoelectric point for the proteins. Because the two parameters are unrelated, it was possible to obtain an almost uniform distribution of protein spots across the two-dimensional gel. Using his technique, O’Farrell was able to resolve 1100 different components from Escherichia coli and predicted his system should be capable of resolving up to 5000 proteins.

O’Farrell’s advisor, Pène, left Boulder around this time, but O’Farrell stayed on to complete his degree under the sponsorship of David Hirsh. After defending his thesis, O’Farrell submitted his two-dimensional electrophoresis manuscript to the JBC and moved to San Francisco to do postdoctoral studies with Gordon Tomkins in the Department of Biochemistry and Biophysics at the University of California, San Francisco (UCSF). Several months later, O’Farrell received a letter from the JBC rejecting his paper on the basis of two unfavorable reviews. The reviewers had concluded that the manuscript appeared to be “highly speculative in places and to be extrapolated in terms of usefulness far beyond what the author has any reason to expect.” Fortunately, with the help of several members of the journal’s editorial...
The O’Farrell Lab

http://biochemistry.ucsf.edu/labs/ofarrell/

Interview:
https://conf.abrf.org/sites/default/files/docs/10-02_my_qa_cb_.pdf

Personal Account of discovery of 2DE method:
https://conf.abrf.org/sites/default/files/docs/08-12_proteomics.pdf

University of Minnesota
ABRF 2014 HIGHLIGHTS

• Plan for a Skyline Learning Session?
• Next Generation Proteomics: Heck of a Talk!
• Proteomics Research Group needs members
• Multi-omics on Snyder = Synderomics
• ENCODE Datasets: A hypothesis generating toolkit
• From Binder Clips to Lasers: Moving along.
• Digital Enterprise: Philip Bourne
• Conclusions
"Biomedical Research as Part of the Digital Enterprise"

Philip Bourne, Associate Director for Data Science, National Institutes of Health

http://www.slideshare.net/pebourne/abrf032514
1. The Era of Open Has The Potential to Deinstitutionalize & Democratize
47/53 “landmark” publications could not be replicated

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

If a job is worth doing, it is worth doing twice
Researchers and funding agencies need to put a premium on ensuring that results are reproducible, argues Jonathan F. Russell.

University of Minnesota
Center for Mass Spectrometry and Proteomics | Phone | (612)625-2280 | (612)625-2279

http://www.slideshare.net/pebourne/abrf032514
Time to Reproduce the Method

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Time (hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Familiarization with workflow and running software</td>
<td>160</td>
</tr>
<tr>
<td>SMAP steps</td>
<td>32</td>
</tr>
<tr>
<td>SMAP result sorter steps</td>
<td>8</td>
</tr>
<tr>
<td>Merger steps</td>
<td>4</td>
</tr>
<tr>
<td>Get significant results</td>
<td>4</td>
</tr>
<tr>
<td>FATCAT URL checker</td>
<td>8</td>
</tr>
<tr>
<td>FATCAT step</td>
<td>4</td>
</tr>
<tr>
<td>Remove significant pairs</td>
<td>4</td>
</tr>
<tr>
<td>Create dip files</td>
<td>8</td>
</tr>
<tr>
<td>Create ideal ligands</td>
<td>8</td>
</tr>
<tr>
<td>Ideal ligand checker</td>
<td>8</td>
</tr>
<tr>
<td>Autodock Vina</td>
<td>16</td>
</tr>
<tr>
<td>Data visualization steps</td>
<td>16</td>
</tr>
<tr>
<td>TOTAL</td>
<td>280 hours</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0080278.t001

2. It's not that we could not reproduce the work, but the effort involved was substantial

http://www.slideshare.net/pebourne/abrf032514
3. Data are accumulating!

DATA KEEPS GROWING

The volume of digital data worldwide is growing rapidly, as the annual IDC Digital Universe study reveals. From 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child in 2020). From now until 2020, the digital universe will about double every two years.

The majority of information in the digital universe, 68% in 2012, is created and consumed by consumers watching digital TV, interacting with social media, sending camera phone images and videos between devices and around the Internet, and so on.

Explore the IDC report, watch videos, and more.

http://www.slideshare.net/pebourne/abrf032514
What Are Big Data?

- Large datasets from high throughput experiments
- Large numbers of small datasets
- Data which are “ill-formed”
- The why (causality) is replaced by the what
- *A signal that a fundamental change is taking place – a tipping point?*

http://jamia.bmj.com/content/21/2/194.full
One Possible End Point

1. User clicks on thumbnail
2. Metadata and a webservices call provide a renderable image that can be annotated
3. Selecting a features provides a database/literature mashup
4. That leads to new papers

PloS Comp. Biol. 2005 1(3) e34

http://www.slideshare.net/pebourne/abrf032514
New/Extended Support Structures Will Emerge

IDEAS – HYPOTHESES – EXPERIMENTS – DATA - ANALYSIS - COMPREHENSION - DISSEMINATION

Discipline-Based Metadata Standards

Git-like Resources By Discipline

Community Portals

Data Journals

New Reward Systems

Training

Institutional Repositories

Commercial Repositories

Commercial & Public Tools

Lab Notebooks

Data Capture

Software

Analysis Tools

Visualization

Scholarly Communication
http://z.umn.edu/10simplerules
ABRF 2014 HIGHLIGHTS

• Plan for a Skyline Learning Session?
• Next Generation Proteomics: Heck of a Talk!
• Proteomics Research Group needs members
• Multi-omics on Snyder = Synderomics
• ENCODE Datasets: A hypothesis generating toolkit
• From Binder Clips to Lasers: Moving along.
• Digital Bourne Identity: Big Data and 10 Simple Rules
• Conclusions
ABRF 2014 HIGHLIGHTS

• Plan for a Skyline Learning Session?
• Next Generation Proteomics: Heck of a Talk!
• Proteomics Research Group needs members
• Multi-omics on Snyder = Synderomics
• ENCODE Datasets: A hypothesis generating toolkit
• From Binder Clips to Lasers: Moving along.
• Digital Bourne Identity: Big Data and 10 Simple Rules
THANK YOU!