Network-based analysis of mouse testicular phosphoproteome

Yu Xue

Huazhong University of Science and Technology
Human Proteome Project

- Little known: $\frac{2}{3}$ of the 20,300 protein-coding human genes
- ~6000 (30%) genes lack the protein information
- HPP:
  - Mass spectrometry, antibody, and bioinformatics
  - Quantitative, sequencing, and PTMs
  - In heath and disease
- Chromosome-Centric Human Proteome Project:
  - Complete proteome
  - Genome annotation
- 2014, CNHPP: encyclopedia

Legrain et al., MCP, 2011, 10, M111.009993
Phosphoproteomics

- Large-scale identification of “in vivo” phosphorylation/PTM sites
- Data integration & resources
  - dbPTM 3.0: 208,521 PTM sites
  - SysPTM 2.0: 471,109 PTM sites, 53,235 proteins
- Prediction of regulatory kinases
  - PKIS: composition of monomer spectrum (CMS)
  - PSEA: Phosphorylation Set Enrichment Analysis

Li et al., Database, 2014, 2014:bau025
Lu et al., NAR, 2013, 41:D295-305
Zou et al., BMC Bioinformatics, 2013, 14:247
What can we learn?

- Poor protein-PTM correlation
- vertebrate-specific functional modules (VFM)s are more conserved than basic functional modules (BFMs)
- Phosphorylation Sites has strong subcellular specificity
- Non-functional p-sites: 65%

In Vivo SILAC-Based Proteomics Reveals Phosphoproteome Changes during Mouse Skin Carcinogenesis

Wang et al., MBE, 2011, 28:1131-40
Chen et al. Bioinformatics. 2014, pii: btu598
Landry et al., Trends Genet., 2009, 25:193-7
Functional PTM prediction

- How many PTM events are functional?
  - Molecular mechanisms
  - Biological forecasting
  - Drug targets

- A big problem:
  - The reproducibility is low

- A network approach: reverse engineering
  - Site-specific kinase-substrate network
  - NetworKIN & iGPS: motif + PPI
  - Tissue-specific: protein expression data

Bodenmiller et al., Nat Methods, 2007, 4:231-7
Linding et al. (2007) Cell, 129, 1415-1426
Song et al., MCP, 2012, 11: 1070-1083
Wang et al., ISB2013, 129-133
Phosphoproteomics-based network medicine

- Kinases: targets of ~75% of complex diseases
- Hypothesis: more sites, higher activity
  - iKING: integrative KINase Gauge
  - KSEA: Kinase-substrate enrichment analysis

Birds of a feather flock together

Known disease genes

Candidates

More sites, higher activity

Condition 1

Statistical comparison

Condition 2

Enriched

Kinase $X$

RxRx$\alpha$pS

Kinase $Y$

$pSP\alpha P$

Song et al., MCP, 2012, 11: 1070-1083
Casado, et al. Sci Signal, 2013, 6, rs6
Semen quality

- Decreasing quality of semen in western countries
- Semen quality is poor in China
  - Associated: region, season & abstinence duration
  - No effect: Age, smoking, alcohol use & BMI

Carlsen et al., BMJ, 1992, 305:609-13
Gao et al., Hum Reprod., 2007, 22:477-84
Li et al., Hum Reprod., 2009, 24:459-69
Spermatogenesis

- Sperm-generating process
  - Mitosis of spermatogonia (精原)
  - Meiosis of spermatocytes (精母)
  - Spermiogenesis of spermatids
  - ~1,000 sperms per heart beat

- Phosphorylation regulated
  - MAPKs, CDC2, POLO-like kinases (PLKs)

Li et al., Trends Mol Med., 2009, 15:159-68
Shi et al., Toxicol Lett., 2013, 221:91-101
Phosphoproteomics

- Swiss-Webster mice, Nine organs
  - brain, brown fat, heart, liver, lung, kidney, pancreas, spleen, & testis

- Phosphoproteomic identification
  - ~36,000 p-sites, 6296 proteins
  - Different network topologies

- Limitations
  - 3-week-old male
  - No sperms at all

Huttlin et al., Cell, 2010, 143:1174-89
Our strategy

- **Adult C57BL/6 mice**
  - 8-week-old male
  - six testes/replicate

- **Phosphopeptide enrichment**
  - IMAC: doubly replicated
  - TiO$_2$: triply replicated
Testicular phosphoproteome

- Total: 17,829 p-sites in 3,955 proteins
  - IMAC: 12,670 sites
  - TiO$_2$: 11,309 sites
- The residue distribution is similar

Qi et al., Mol Cell Proteomics, pii: mcp.M114.039073
IMAC & TiO$_2$

- The data overlapping is limited
  - IMAC: ~48% identified by TiO$_2$
  - 8.7% covered in all replicates

<table>
<thead>
<tr>
<th></th>
<th>IMAC</th>
<th>TiO$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>5,079</td>
<td>5,257</td>
</tr>
<tr>
<td></td>
<td>4,101</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>1,198</td>
<td>870</td>
</tr>
<tr>
<td></td>
<td>849</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>243</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>209</td>
<td></td>
</tr>
</tbody>
</table>
Comparison

- Gygi’s: 11,713 p-sites in 3,087 proteins
  - ~27% covered by our data set
GO analysis: similar!

Total

Spermatogenesis
RNA splicing
mRNA processing
Transcription, DNA-dependent
Negative regulation of transcription from RNA polymerase II promoter
DNA repair
Response to DNA damage stimulus
Ubiquitin-dependent protein catabolic process
Protein phosphorylation
Cell cycle
In utero embryonic development

IMAC

Spermatogenesis
mRNA processing
Response to DNA damage stimulus
RNA splicing
Protein phosphorylation
DNA repair
Cell division
Transcription, DNA-dependent
Intracellular signal transduction
Actin cytoskeleton organization
Protein autophosphorylation
Actin cytoskeleton organization
Negative regulation of transcription from RNA polymerase II promoter
Negative regulation of transcription, DNA-dependent
Positive regulation of translation
Cell cycle

Titanium dioxide

Spermatogenesis
Ubiqitin-dependent protein catabolic process
RNA splicing
mRNA processing
DNA repair
Negative regulation of transcription from RNA polymerase II promoter
Transcription, DNA-dependent
Response to DNA damage stimulus
Cell cycle
Regulation of Rho protein signal transduction
Regulation of phosphoprotein phosphatase activity
Protein deubiquitination
In utero embryonic development

Gygi’s

DNA repair
Response to DNA damage stimulus
Regulation of Rho protein signal transduction
Protein autophosphorylation
mRNA processing
Negative regulation of transcription from RNA polymerase II promoter
RNA splicing
Actin cytoskeleton organization
In utero embryonic development
Protein phosphorylation
Ubiquitin-dependent protein catabolic process
DNA replication
Intracellular signal transduction
mRNA polyadenylation
Protein localization to kinetochore (GO:0034501)
Network construction

- iGPS: ‘in vivo’ GPS
  - [http://igps.biocuckoo.org](http://igps.biocuckoo.org)
  - site-specific kinase-substrate relations (ssKSRs)

- kinase-substrate phosphorylation networks (KSPN)
  - 17,065 edges, 402 kinases, 1,066 substrates

---

Song et al., MCP, 2012, 11, 1070-1083
Spermatogenesis-related KSPN

- **Spermatogenesis-related proteins**
  - QuickGO: spermatogenesis (GO:0007283)
  - Spermatogenesis *Online* database

- **Sub-KSPN**: 106 proteins, 371 edges

Phosphorylates 47 (~44%) proteins
Kinases with higher activities

➢ Background: ~36,000 p-sites in 9 organs
➢ iKING: top15 kinases with highest activities
Kinase with higher activity

- **MAPKs**
  - ectoplasmic specialization dynamics
  - $p$-value < 0.05: JNK2, ERK2 and p38s

- **CDK2**: chromosome pairing in meiosis

- **CDC2**: sperm activity and testicular function

*Li et al., Trends Mol Med., 2009, 15:159-68*
*Viera et al., J Cell Sci., 2009, 122:2149-59*
*Shi et al., Toxicol Lett., 2013, 221:91-101*
Polo-like kinases (Plks)

Liu, et al., Brief Bioinform, 2013, 14:344-60
Plk1-PBD interact with Phos-Mis18B

A. PIK1

B. GPS-Polo 1.0

C. Cell lysates Pull down

D. Cell lysates IP

T14, S48, T221
Plk1 co-localizes with Mis18B
Plk1 phospho-binding regulates Mis18B stability
A potential role in testis?

- Plks: Not annotated in Spermatogenesis Online
- Plk sub-KSPN: 122 proteins, 499 edges
  - Substrates: many spermatogenesis-related proteins
  - Phenotypes: cell morphology related

<table>
<thead>
<tr>
<th>Phenotype ID</th>
<th>Description</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP:0003111</td>
<td>Abnormal cell nucleus morphology</td>
<td>2.26E-10</td>
</tr>
<tr>
<td>MP:0003077</td>
<td>Abnormal cell cycle</td>
<td>5.71E-08</td>
</tr>
<tr>
<td>MP:0004046</td>
<td>Abnormal mitosis</td>
<td>8.76E-08</td>
</tr>
<tr>
<td>MP:0000358</td>
<td>Abnormal cell morphology</td>
<td>3.10E-07</td>
</tr>
<tr>
<td>MP:0002022</td>
<td>Increased lymphoma incidence</td>
<td>3.97E-07</td>
</tr>
<tr>
<td>MP:0002020</td>
<td>Increased tumor incidence</td>
<td>4.03E-07</td>
</tr>
<tr>
<td>MP:0010274</td>
<td>Increased organ/body region tumor incidence</td>
<td>5.83E-07</td>
</tr>
<tr>
<td>MP:0002019</td>
<td>Abnormal tumor incidence</td>
<td>9.06E-07</td>
</tr>
</tbody>
</table>
Experiments

- Spermatocyte GC2 cell line
- BI2536: Plks inhibitor
- Okadaic acid (OA)
  - Phosphatase inhibitor
  - Reverses the phosphorylation changes after inhibition of Plks
- DMSO: control
- RT-PCR: Plk1-3 existence
Validation of Plk activation

- pT210: the major p-site and correlates with Plk1 activity
- Testes: 1, 2, 3, 4, 8 weeks
- Control: 8 tissue mixtures; 3 and 8 weeks
- Plk1 activity is significantly higher

PI3k inhibition: G2/M arrest

- After 6h
  - G1 cells: decreased
  - S cells: unchanged
  - G2/M cells: increased

- BI2536: induces arrest in the G2/M phase and inhibits cell proliferation
Discussion

- The differential activities of kinases can be readily and robustly predicted from poorly reproducible data
- Plk1: colon and lung cancers
- Plk inhibitors: potential anti-cancer drugs
  - BI2536: Phase I and II
  - BI6727/Volasertib: granted by the FDA in AML (with cytarabine), in older patients
  - GSK461364A: Phase I

*This compound is an investigational agent. Its safety and efficacy have not been established.

**VOLASERTIB**, AN INVESTIGATIONAL POLO-LIKE KINASE (PLK) INHIBITOR

Gjertsen et al., Leukemia. 2014, doi: 10.1038/leu.2014.222
Perspectives

- Biological forecasting: network-based prediction
- Network targets in Liver cancers and neurodegenerative diseases
- Key regulators in Autophagy
- Can we learn more things from the phosphoproteomic data?
Acknowledgements

- Zexian Liu (HUST)
- Jian Ren (SYSU)
- Jiahao Sha & Xuejiang Guo (NJMU)
- Luonan Chen (Sysbio)
- Hanfa Zou & Mingliang Ye (DICP)
- Ping Xu (BPRC)
- Min Li (HKBU)
- Longping Wen (USTC)
- Members of the CUCKOO Workgroup
Thanks!

Any questions?