

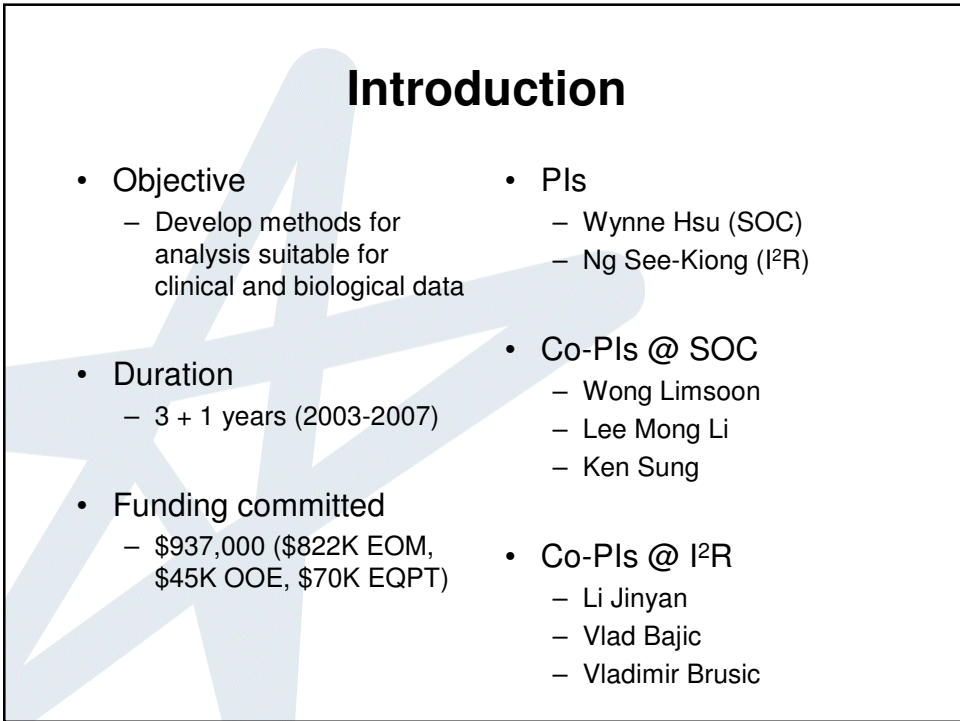


**I<sup>2</sup>R-SOC Joint Lab**

**Knowledge Discovery from  
Biological & Clinical Data**

A/Prof Wynne Hsu (SOC)  
Dr Ng See-Kiong (I<sup>2</sup>R)

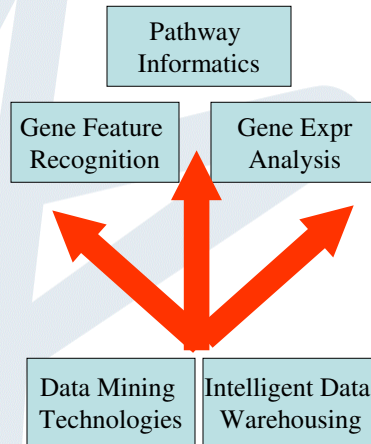
 



**Introduction**

- **Objective**
  - Develop methods for analysis suitable for clinical and biological data
- **Duration**
  - 3 + 1 years (2003-2007)
- **Funding committed**
  - \$937,000 (\$822K EOM, \$45K OOE, \$70K EQPT)
- **PIs**
  - Wynne Hsu (SOC)
  - Ng See-Kiong (I<sup>2</sup>R)
- **Co-PIs @ SOC**
  - Wong Limsoon
  - Lee Mong Li
  - Ken Sung
- **Co-PIs @ I<sup>2</sup>R**
  - Li Jinyan
  - Vlad Bajic
  - Vladimir Brusic

## Planned Deliverables



## Progress & Achievements

### *Funding/Resource Utilization*

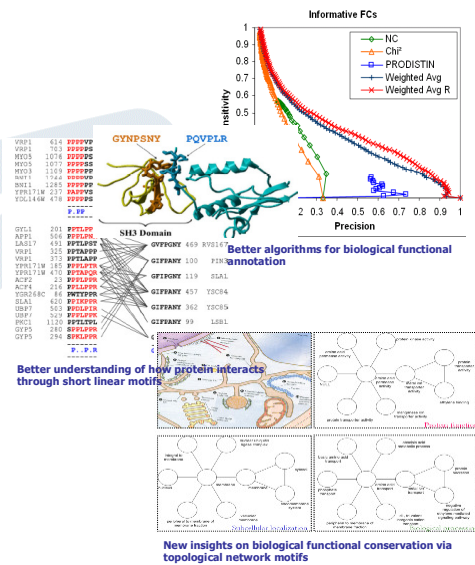
<i>Vote</i>	<i>Original Grant</i>	<i>Received Grant</i>	<i>Actual expenditure</i>	<i>% Utilization</i>
<b>EOM</b>	\$822,000	\$770,366.30	\$686,967.12	89%
<b>OOE</b>	\$45,000	67,994.50	\$53,228.83	78%
<b>EQPT</b>	\$70,000	57,124.54	\$54,189.92	95%
<b>Total</b>	\$937,000	895,485.34	\$794,385.87	89%

*Wynne: May need to say a few things on why 89% were used...*

# Progress & Achievements 1:

## Deepened understanding of molecular biology

- 1st to discover function-sharing of proteins in 2nd level PPI neighbours
- 1st to mine linear motifs from PPI data
- 1st to discover meso-scale network motifs in PPI networks
- Suite of bioinformatics algo's: IRAP, DeCAFF, NeMoFinder, D-STAR, FSWeight



# Progress & Achievements 2:

## Deepened understanding data mining methods

- Fundamental research on EP and ensemble-based classification methods led to new insights on freq pattern space and equiv classes
- Amongst the first to make the link betw max complete bipartite subgraphs and closed pattern pairs
- New mining methods developed: PCL & CS4
- New concepts developed: max quasi-bicliques for co-clustering, labeled network motifs for graph mining
- Applicable in other domains such as social network mining & financial data mining

## Progress & Achievements 3:

### Publications & Patents

- Published in leading conferences & journals in both computational biology and data mining
  - Computational Biology: Bioinformatics, Nucleic Acids Research, RECOMB, GIW, ...
  - Data Mining: Algorithmica, Machine Learning, KDD, ICDE, AAAI, ...
- 130 publications including
  - 10 books
  - 1 Best Paper
  - 1 Keynote Paper
- 1 patent filed

## Progress & Achievements 4:

### Manpower Training & Development

- 12 PhD students

PhD Student	Status	First job post-PhD
<b>3 recruited and directly funded by the lab</b>		
1 Chen Jin	PhD awarded	Postdoc at Stanford University
2 Wong Swee Seong	Thesis submitted	Lilly-Singapore Centre for Drug Discovery
3 Hugo Willy	Qualifiers passed	
<b>9 affiliated but funded by other sources</b>		
1 Liu Huiqing	PhD awarded	Postdoc at University of Georgia
2 Judice Koh	PhD awarded	Postdoc at University of Toronto
3 Li Haiquan	PhD awarded	Staff member at Noble Foundation USA
4 Kenny Chua	Thesis submitted	Research fellow at I2R
5 V. S. Sundarajan	Thesis submitted	Research fellow at NUS Dentistry
6 Feng Meng Ling	Confirmation passed	
7 Lee Terk Shuen	Qualifiers passed	
8 Donny Soh	Viva to be taken	
9 Hou Yuna	Withdrawn	

- 3 postdocs and 3 research assistants recruited and funded by the lab

## Other Achievements

- National Awards
  - Ken Sung
    - National Science Award 2006, for contributions to paired-end ditag sequencing technology with GIS
  - Limsoon Wong
    - Singapore Youth Award Medal of Commendation 2006, for sustained contributions to science and technology
- International Conferences & Workshops
  - BioDM Workshop @ PAKDD 2006
  - APBC 2005
- Invited Talks & Tutorials
  - Total 83
  - Incl. 8 keynotes

## Conclusions

- Successful joint-research on data mining tech, gene feature recognition, gene expression analysis, pathway informatics, & intelligent warehousing
- Deepened understanding of both molecular biology and data mining through computational analysis of biological data
- 3 directly-funded, 8 alternatively-funded PhD students, 3 postdocs, and 2 undergrads trained
- Published extensively in both bioinformatics and data mining

## Future Plans

- Seek more opportunities to continue the bioinformatics research
- Nourish the collaborative relationships resulted from the joint-lab
- Some ongoing collabs
  - Information fusion on network data
  - Protein function prediction
  - Protein complex prediction