WONG, Limsoon

- Kwan-Im-Thong-Hood-Cho-Temple Professor
- Dept of Computer Science, National University of Singapore (NUS)
- ResearchID: E-5033-2010; ORCID: 0000-0003-1241-5441
- Homepage: www.comp.nus.edu.sg/~wongls

Professional preparation

- BSc(Eng)(Computing), first-class honours, 1998, Imperial College, London, UK
- PhD, in computing and information science, 1994, University of Pennsylvania, Philadelphia, USA

Appointments

- Current: **Kwan-Im-Thong-Hood-Cho-Temple Professor, Dept of Computer Science, National University of Singapore (NUS); and Deputy Dean, NUS Graduate School.** I joined NUS in 2005 and held the several positions (viz. professor of computer science; professor of pathology; Provost's chair professor of computer science; head of computer science dept; acting executive director, NUS Graduate School for Integrative Sciences and Engineering) before being appointed to the two current ones.
- Past: **Deputy Executive Director (Research), Institute for Infocomm Research, A*STAR, Singapore**. I held various positions at this institute and its predecessor organizations during 1988 to 2005. Last position held was Deputy Executive Director (Research).

Current PhD students

1. CHAN Weixin
2. Mohammad Neamul KABIR
3. Lakshmi ALAGAPPAN
4. Henry KASSIM

Past PhD students

1. LIU Huiqing, "Effective use of data mining technologies on biological and clinical data", PhD 2005, NUS. Current position: Principal Investigator, Incyte, Philadelphia, PA, USA.
2. Rajesh CHOWDHARY (main supervisor: Vladimir Bajic), "Modeling and recognition of histone promoters on genome-wide scale by Bayesian networks", PhD 2007, NUS. Current position: Chief Scientific Officer, Insilicom, Tallahassee, FL, USA.
3. Vijayaraghava Seshadri SUNDARARAJAN, "Progressive data mining: An exploration of using whole-dataset feature selection in building classifiers on three biological problems", PhD 2008, NUS. Last known position: Asst Director at GSTF, Singapore
4. CHUA Hon Nian (co-supervisor: Ken Sung), "Graph-based methods for protein function prediction", PhD 2008, NUS. Current position: Head of R&D, DataRobot, Singapore
5. WONG Swee Seong (main supervisor: Ken Sung), "String matching and indexing with suffix data structures", PhD 2008, NUS. Current position: NGS Bioinformatics Director, LifeOmic, Indianapolis, IN, USA.

6. Stanley NG Kwang Loong (main supervisor: Santosh Mishra), "Computational identification of novel microRNAs using intrinsic RNA folding measures", 2008, NUS. Last known position: Senior Manager, Exploit Technologies, Singapore.
7. FENG Mengling (co-supervisors: Yap Peng Tan, Jinyan Li), "Frequent pattern space maintenance: Theories and algorithms", PhD 2010, Nanyang Technological University. Current position: Assistant Professor, Saw Swee Hock School of Public Health, NUS, Singapore.
8. Donny SOH (co-supervisor: Yike Guo), "Understanding pathways", PhD 2010, Imperial College London. Current position: Senior Technical Advisor, Future-Moves Group, Singapore.
9. DONG Difeng, "Relapse prediction in childhood acute lymphoblastic leukemia by time-series gene expression profiling", PhD 2011, NUS. Current position: Portfolio Manager, Dymon Asia Capital, Singapore.
10. KOH Chuan Hock, "Embracing noise in bioinformatics", PhD 2013, NUS. Current position: Data Science Manager, Indeed.com, Tokyo, Japan.
11. WANG Yue, "Efficient computational techniques for tag SNP selection, epistasis analysis, and genome-wide association study", PhD 2013, NUS. Current position: Founder and CEO, Populstay, Tokyo, Japan.
12. ZHOU Hufeng (William CHEW), "Computational studies of host-pathogen protein-protein interactions---a case study of the H. sapiens-M. tuberculosis H37Rv system", PhD 2013, NUS. Current position: Instructor, Harvard Medical School and Research Scientist, Dept of Biostatistics, Harvard School of Public Health, Boston, MA, USA.
13. Wilson GOH Wen Bin (co-supervisor: Marek Sergot), "Computational proteomics using network-based strategies", PhD 2013, Imperial College London. Current position: Assistant Professor, LKC School of Medicine, Nanyang Technological University, Singapore.
14. JIN Jingjing (co-supervisor: Nam-Hai Chua), "Investigating lipid and secondary metabolisms in plants by next-generation sequencing", PhD 2014, NUS. Current position: Principal Investigator, Zhengzhou Tobacco Research Institute, Zhengzhou, China.
15. YONG Chern Han, "Discovering dynamic protein complexes from static interactomes: Three challenges", PhD 2015, NUS. Current position: Research Fellow, School of Computing, NUS.
16. LIM Junliang Kevin, "Using biological networks and gene-expression profiles for the analysis of diseases", PhD 2015, NUS. Current position: Research Fellow, Wilmar Research Labs, Singapore.
17. Michal WOZNIAK (main supervisor: Jerzy Tiuryn), "Analysis of bacterial interactome for fighting drug resistance", PhD 2015, Warsaw University. Current position: With a hedge fund in Warsaw.
18. Abha BELORKAR, "Gene expression analysis in the presence of heterogeneity", PhD 2018, NUS. Current position: Assistant Professor, Department of Computer & Information Sciences, Temple University, Philadelphia, PA, USA.
19. Chayaporn Suphavilai (main supervisor: Niranjan Nagarajan), “Predicting cancer drug response using a recommender system”, PhD 2019, NUS. Current position: Research Fellow, Genome Institute of Singapore.
20. LI Chenhao (main supervisor: Niranjan Nagarajan), “Predicting microbial interactions with modelling approaches”, PhD 2019, NUS. Current position: Research Fellow, Broad Institute at MIT, Boston, MA, USA.
21. XU Weinan, “Evolutionary computation for solving constrained an large-scale optimization problems”, PhD 2020, NUS. Current position: Data Scientist, Lazada, Singapore.
22. XIE Luyu, “Improvement and evaluation of genome assembly”, PhD 2020, NUS. Current position: Research Fellow, Duke-NUS.

Selected professional activities and recent memberships of commissions of trusts

- Member, External review committee, Institute of Medical Sciences University of Tokyo (IMSUT).
  IMSUT is one of Japan’s largest national institutes in the life sciences with its own hospital. Limsoon served on the external committee which reviewed the performance of IMSUT over the April 2016 – June 2020 period and the institute’s proposed new structure and direction for the next period.

- Member, Awards Committee, International Society for Computational Biology (ISCB) (2018- present)
- Member, Publication Committee, International Society for Computational Biology (ISCB) (2020- present)
- Managing Editor, Journal of Bioinformatics and Computational Biology (2002- present)
- Advisory Editor / Associate Editor, Bioinformatics (2005- 2010)
- Advisory Editor, Drug Discovery Today (2004- present)
- Area Editor, Information Systems (2008- present)
- Associate Editor, IEEE/ACM Transactions on Computational Biology and Bioinformatics (2010- present)
- Served on the program committees of about 70 conference series, many multiple times. Some notable ones include ISMB, RECOMB, ECCB, GIW, APBC, InCoB, PODS, ICDT, SIGMOD, VLDB, ICDE, and WWW.
- Conference and/or PC chair of APBC2005/2006, BIBM2010, InCoB2005/2009, RECOMB2008, and GIW2001/2009/2013.
- Gave ~240 invited keynotes, tutorials, and talks worldwide. Some notable ones include ICFP2000, ICDT2003, AAAI2005, GIW2006, APBC2007, ICDE2007, SIGIR2008, GIW2014, BIBE2016, InCoB2017.

Selected research awards

- Fellow of the ACM, inducted 2013, for contributions to database theory and computational biology
- International Conference on Database Theory (ICDT) Test of Time Award, 2014, for the paper "Naturally embedded query languages" published in ICDT1992.
- Asian Innovation Award (Gold), Far Eastern Economic Review, 2003, for treatment optimization of childhood acute lymphoblastic leukemias.

Publications

- 11 books authored or edited; 20 book chapters; ~300 research articles, mostly in top-tier journals and conferences; 3 patents granted; ~240 keynotes and invited talks in international conferences, workshops, and other events.
- ~20% of Limsoon's research articles are heavily cited (~16,600 times), and many of them are among the best-cited in their respective fields. Limsoon's H-index is 62.
- A comprehensive list of Limsoon's paper can be found on Google Scholar, https://scholar.google.com/citations?user=Inqu8SkAAAAJ

Representative accomplishments in database theory

(i) Enabling database query language theory to catch up with SQL:

[BNTW95] P. Buneman, S. Naqvi, V. Tannen, L. Wong. **Principles of programming with complex objects and collection types**. *Theoretical Computer Science*, 149(1):3--48, 1995.

[Wong96] L. Wong. **Normal forms and conservative extension properties for query languages over collection types**. *Journal of Computer and System Sciences*, 52(3):495--505, 1996.

[LW93] L. Libkin, L. Wong. **Aggregate functions, conservative extension, and linear orders**. *Proceedings of 4th International Workshop on Database Programming Languages*, pages 282--294, 1993.

[LW97] L. Libkin, L. Wong. **Query languages for bags and aggregate functions**. *Journal of Computer and System Sciences*, 55(2):241--272, 1997.

[DLW00] G. Dong, L. Libkin, L. Wong. **Local properties of query languages**. *Theoretical Computer Science*, 239:277--308, 2000.

Much work on understanding the precise boundary of the (extensional) expressive power of database query languages (i.e., what functions can be or cannot be expressed in a database query language) has been based on the relational algebra and its equivalence to first-order logic. However, a lot of results on the limit of the expressive power of database query languages could not be applied to SQL, the de facto query language of the database industry. SQL has significantly higher expressive power than the relational algebra. For example, while a test of whether two departments have the same number of employees given a department-employee table is inexpressible in relational algebra, this test is easily expressed in SQL. The extra expressive power of SQL is due to its group-by, aggregate function, and arithmetic operators. The presence of such operators is difficult to deal with using earlier approaches for analyzing expressive power. Thus, theory lagged far behind practice in the case of SQL.

Limsoon and his collaborators presented a new paradigm for the design of database query languages where the primitive operations are organized around types [BNTW95]. Viewing a relational database as consisting of sets of records, this paradigm dictates that we should investigate separately operations for records and sets. There are two immediate advantages of this paradigm. First, it provides a language for structures in which record and set types may be freely combined: nested relations or complex objects. Second, the fundamental operations for sets are closely related to those for other "collection types" such as bags and lists, and this suggests how database languages may be uniformly extended to these new types.

Limsoon investigated various questions on the expressive power of query languages resulting from this paradigm. In particular, he contributed a series of fundamental theorems on the conservative extension property [Wong96, LW93], finite-cofiniteness property [LW97], locality property [DLW00], and bounded degree property [DLW00] of various nested relational calculi augmented with aggregate functions. These properties characterize functions (on generic unordered atomic data types) that are expressible in various nested relational calculi augmented with aggregate functions. They define a more precise limit on the expressive power of these query languages: If a function does not satisfy them, then the function is not expressible in these query languages. These theorems are major breakthroughs which settled important long-standing issues in understanding the expressive power of realistic query languages such as SQL. I.e., these theorems have enabled query language theory to catch up with practice.

(ii) Settling the Kanellakis Conjecture:

[BDLW98] M. Benedikt, G. Dong, L. Libkin, L. Wong. **Relational expressive power of constraint query languages**. *Journal of the ACM*, 45(1):1--34, 1998.

A relational database is traditionally formalized as a collection of finite relations. But this has limitation in some applications such as representing and querying spatial data. Thus, Kanellakis, Kuper, and Revesz in their famous PODS'90 paper introduced the idea of constraint databases and query languages for dealing with this type of data and queries. The original idea of constraint query languages is that quantifiers in a query can range over the entire universe of real numbers, as opposed to only over those real numbers that occur in the underlying database. That is, constraint query languages permit the use of quantifiers under the "natural domain semantics", while traditional database query languages interpret quantifiers strictly under the "active domain semantics". The presence of such quantifiers and arithmetic operators on real numbers would give constraint query languages more expressive power on spatial data. However, it was not known whether these quantifiers would give them more expressive power on queries that were not spatial related. The late Kanellakis conjectured that these constraint query languages could not express the transitive closure query.

Limsoon's second important contribution to database theory was the solution of the Kanellakis Conjecture on the inexpressibility of transitive closure in a large class of constraint query languages. He made an insightful and

fundamental generalization of this conjecture that arithmetic operations occurring in this class of query languages could be eliminated for all queries satisfying a simple genericity condition under both the natural domain semantics interpretation (i.e., all quantifiers are interpreted to range over all the entire universe of the real closed field) and the active domain semantics interpretation. He then proceeded to confirm this conjecture in collaboration with M. Benedikt, G. Dong, and L. Libkin in a much celebrated paper [BDLW98]. The long-standing Kanellakis Conjecture followed as a direct corollary.

(iii) Clarifying the intensional expressive power of nested relational calculi:

> [Wong13] L. Wong. **A dichotomy in the intensional expressive power of nested relational calculi augmented with aggregate functions and a powerset operator**. *Proceedings of 32nd ACM Symposium on Principles of Database Systems*, New York, 22-27 June 2013.

Most existing studies on the expressive power of query languages have focused on what queries can be expressed and what queries cannot be expressed in a query language. They do not tell us much about whether a query can be implemented efficiently in a query language. Yet, paradoxically, efficiency is of primary concern in computer science. Thus an important and insufficiently explored frontier in query language theory is the development of powerful general methodology for studying the intensional expressive power of query languages, especially those that support nested relations, aggregate functions, and powerset or recursion operations.

The study of intensional expressive power is much more fine-grained and challenging than the study of (extensional) expressive power. It is not sufficient to show that a function cannot be expressed. To get a more precise limit on the intensional expressive power of a query language, one has to consider functions that are expressible in the query language (i.e., some algorithms for realizing these functions can be implemented using the query language) and then show whether other algorithms for realizing these functions are inexpressible in the query language. Thus there had been little progress on this topic over the last decade and a half after some early results.

Limsoon's third important contribution to database theory was made quite recently, when he returned to the field after a break of more than a decade (he has switched his attention to computational biology in between) and looked at this difficult topic of intensional expressive power. Almost instantly, he made an ingenious technical breakthrough and proved a dichotomy theorem [Wong13] which significantly generalizes several old results, e.g., Suciu and Paredaens's (that all implementations of transitive closure in the relational algebra augmented with a powerset operation must take exponential space) and Van den Bussche's (that all implementations of parity query in the same algebra must take exponential space) in three significant ways: 1/ the result is generalized to SQL augmented with a powerset operation; 2/ the result is generalized to all queries on severely dichotomous classes of structures (which are very general classes of structures); and 3/ the proof technique factors through the conservative extension property and locality property (thus, if language L having a powerset operation and the conservative extension property, has the locality property after removing the powerset operation from it, then the dichotomy theorem holds on L too). This is probably the most general result and most general proof technique on the intensional expressive power of query languages.

Representative accomplishments in bioinformatics and computational biology

The successful application of computational approaches to solve problems in biology and medicine often involves four different dimensions, which are described below along with Limsoon's key contributions in them.

(i) Integration of multiple sources of data:

[BDHO95] P. Buneman, S. Davidson, K. Hart, C. Overton, L. Wong. **A data transformation system for biological data sources**. *Proceedings of 21st International Conference on Very Large Data Bases*, Zurich, Switzerland, 158-169, 1995.

[Wong00] L. Wong. **Kleisli, a functional query system.** *Journal of Functional Programming*, 10(1):19--56, 2000.

In order to analyze a problem in biology or other domains, data in multiple databases must often be compiled and integrated. These databases are usually complex, heterogeneous, and dispersed geographically. Limsoon's effort in this dimension of the problem is epitomized in the Kleisli Query System, which pioneered the use of high-level query languages with type inference and self-describing exchange formats for broad-scale data integration. The system was the first to solve a series of bioinformatics data integration problems considered "impossible" by the US Department of Energy at that time.

Prior to Limsoon's work on the Kleisli Query System in the early-mid 1990s, it was widely believed that an integrated global schema was necessary to a successful integration of multiple databases. Limsoon's pioneering concepts of wrappers, self-describing data exchange formats, and high-level query language with type inference in Kleisli completely changed that belief. His insight was that, in practice, in any query, only small parts of a small number of databases would be accessed. Thus there should be no need to construct a fully integrated global scheme beforehand. A high-level query language with type inference would make it possible to infer the structure or schema of those small parts of the small number of databases needed by the query. Light-weight wrappers and self-describing exchange formats would make it possible to encode, decode, and exchange the data needed by the query from these databases. This seminar work was also considered to be one of the six important practical achievements of the 1990's in functional programming (for Kleisli was implemented using a functional programming language, SML) as highlighted in P. Wadler's 1998 "Angry Half Dozen" article in *SIGPLAN Notices*.

The Kleisli Query System was a technology ahead of its time. It led to a spin-out company, geneticXchange, which marketed and deployed Kleisli in large pharmaceutical companies such as Schering Plough. It had also been deployed in large banks such as Lloyds.

(ii) Extraction of information from free text:

[Wong01] L. Wong. **A Protein Interaction Extraction System**. *Proceedings of Pacific Symposium on Biocomputing 2001*, pages 520--530, 2001.

[HPTW02] L. Hirschman, J. C. Park, J. Tsujii, L. Wong, C. H. Wu. **Accomplishments and challenges in literature data mining for biology**. *Bioinformatics*, 18:1553--1561, 2002.

The information that is needed for the analysis of a problem in biology or other domains may not be available in well-structured databases. Nevertheless, such information may be found in free-text form in the scientific literature. Prior to Limsoon's work on PIES (Protein Interaction Extraction System) and his effort to champion bioNLP research in the late 1990s, this was not a challenge that received much attention in the bioinformatics community. Limsoon was among the first to identify automatically extracting information from the published literature as an important approach in bioinformatics. Such information is necessary to augment data in structured bio databases that are used in computational analysis of biological and clinical data further downstream.

He collaborated with S.-K. Ng to develop PIES, one of the first NLP/IE systems for extracting protein interaction and regulation information from literature. Together with Ng, Limsoon put PIES and its associated data curation processes into a start-up company, Molecular Connection, in India. Two decades later, the profitable company has organically grown 400x to more than 2000 engineers and curators. The company is considered one of the best SMEs in India in several studies by, e.g., Deloitte and the Indian Institute of Economics Studies.

Not only did Limsoon co-developed one of the earliest "bioNLP" systems, PIES, he also went out of his way to actively bring researchers' attention to this area by organizing the first three bioinformatics conference tracks dedicated to this topic, as well as lead-authored a paper [HPTW02] that called the community into action. This is now a vibrant subfield in bioinformatics; e.g., there are now tens of thousands of Google search hits on the term "bioNLP"; a top journal of the field, viz. *Bioinformatics*, has a special section explicitly for papers on bioNLP; and international bioNLP competitions are regularly held.

(iii) Distillation of knowledge from the information and (iv) Application of that knowledge to make useful predictions:

[YRSW02] E.-J. Yeoh, M. E. Ross, S. A. Shurtleff, W. K. William, D. Patel, R. Mahfouz, F. G. Behm, S. C. Raimondi, M. V. Reilling, A. Patel, C. Cheng, D. Campana, D. Wilkins, X. Zhou, J. Li, H. Liu, C.-H. Pui, W. E. Evans, C. Naeve, L. Wong, J. R. Downing. **Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling**. *Cancer Cell*, 1:133--143, 2002.

[YLDL18] Allen E.-J. Yeoh, Zhenhua Li, Difeng Dong, Yi Lu, Nan Jiang, Jan Trka, An Moy Tan, Hai Peng Lin, Thuan Chong Quah, Hany Ariffin, Limsoon Wong. Effective Response Metric: A novel tool to predict relapse in childhood acute lymphoblastic leukaemia using time-series gene expression profiling. British Journal of Haematology, 181(5):653--663, 2018.

[CSW06] H. N. Chua, W.-K. Sung, L. Wong. **Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions**. *Bioinformatics*, 22:1623-1630, 2006.

[SDGW10] D. Soh, D. Dong, Y. Guo, L. Wong. **Consistency, comprehensiveness, and compatibility of pathway databases**. *BMC Bioinformatics*, 11:449, 2010.

[SDGW11] D. Soh, D. Dong, Y. Guo, L. Wong. **Finding consistent disease subnetworks across microarray datasets**. *BMC Genomics*, 12(Suppl. 13):S15, 2011.

[LW14] Kevin Lim, Limsoon Wong. **Finding consistent disease subnetworks using PFSNet**. *Bioinformatics*, 30(2):189--196, January 2014.

[GLCW12] W. Goh, Y. H. Lee, M. Chung, L. Wong. **How advancement in biological network analysis methods empowers proteomics**. *Proteomics*, 12(4--5):550--563, 2012.

[GLRS12] W. Goh, Y. H. Lee, Z. Ramdzan, M. Sergot, M. Chung, L. Wong. **Proteomics Signature Profiling (PSP): A novel contextualization approach for cancer proteomics**. *Journal of Proteome Research*, 11(3):1571--1581, 2012.

[GW16] Wilson Wen Bin Goh, Limsoon Wong. **Advancing clinical proteomics via analysis based on biological complexes: A tale of five paradigms**. *Journal of Proteome Research*, 15(9):3167--3179, 2016.

Traditionally, computational biologists tend to employ "one-dimensional" approaches. For example, for protein function prediction, they usually rely solely on sequence homology search. As another example, for biomarker selection from gene expression, they usually rely solely on statistical analysis of the correlation of gene expression profiles to patient phenotypes. Yet sequence homology search is not applicable to a significant portion of novel proteins that do not have informative homologs in sequence databases. Similarly, due to practical limitation on sample size, the expression profiles of many false-positive genes are as well correlated with patient phenotypes as those of true-positive genes and, thus, cannot be eliminated by pure statistical means. This has begun to change in the early 2000s. Limsoon is among the group of computational biologists to actively pioneer knowledge discovery approaches that deeply incorporate biological background---such as biological pathways and networks, often integrated from multiple sources and extracted from free text---to model and solve these problems.

Limsoon pioneered many new ideas in the integrated analysis of biomedical data and the discovery of insights into the invariants underlying the associated biological problems. A paper in CIKM2006 ranked him as a top-10 social

actor who powered data mining research into bioinformatics. His seminar results on gene expression profile analysis in childhood leukemias [YRSW02] and the cleansing and use of biological networks in, e.g., protein function prediction [CSW06] are well known. His team was winner of *Far Eastern Economic Review*'s Asian Innovation Gold Award in 2003 on the former topic and the DREAM2 Challenge in 2007 on the latter topic.

An early success was the extremely widely cited work of Limsoon and his team (Liu Huiqing, Li Jinyan, and Allen Yeoh) with St Jude's that pioneered knowledge discovery from patient gene expression profiles of childhood acute lymphoblastic leukemias (ALL) [YRSW02]. It was the first gene expression profile-based work to distinguish all major pediatric ALL subtypes and the first to demonstrate the use of gene expression profiling to identify treatment relapse in certain childhood ALL subtypes.

More recently, Limsoon and his student Dong Difeng proposed the genetic status shifting model that models the shift of gene expression patterns (from non-cell-sorted samples) as a child suffering from ALL gets his treatment [YLDL18]. In collaboration with Allen Yeoh, this model has been observed to be more accurate than the current clinical gold standard, MRD33, in several trials. Moreover, it could also be done after one week into treatment, as opposed to the one month into treatment required by MRD33. This was the first time ever that a gene expression profile-based approach consistently achieved significantly better relapse and treatment response prediction than MRD33. The same concept was also successfully applied to a small cohort of acute myeloid leukemia patients, suggesting the general applicability of the genetic status shifting model to other liquid tumours.

To achieve a better understanding of the causal factors underlying patient gene expression profiles, Limsoon also advocated a (sub)network-based approach to analyze gene expression profiles that integrated his technologies in all three dimensions (data integration, information extraction, and knowledge discovery). In this approach, data from multiple structured databases, as well as information extracted from literature, are integrated into a comprehensive set of biological networks. Novel knowledge discovery approaches were then developed by Limsoon and his students, Donny Soh and Kevin Lim, to identify subnetworks that would explain the differential gene expression in different patient phenotypes [SDGW10, SDGW11, LW14]. This approach was tested on multiple diseases and, each time, it was shown that selected subnetworks and associated genes were much more consistent than those selected by other state-of-the-art approaches.

The power of this end-to-end network-based approach is not limited to gene expression profile analysis. For example, Limsoon and his student, Chua Hon Nian, also applied this to protein function prediction, a key problem in computational biology. There, in a widely cited paper [CSW06], they discovered that proteins have clear functional association with their level-two neighbours in protein-protein interaction networks (PPIN). They successfully exploited this phenomenon to make reliable protein function prediction for many proteins even in the absence of sequence homology.

As another example, Limsoon and his student, Wilson Goh, insightfully pointed out that, for disease phenotypes caused by the non-functioning of a protein complex, a major change in any protein component of the complex would lead to the non-functioning of the complex and, thus, the disease phenotype. Therefore, while one often could not see a consistent proteomic profile at the level of individual proteins (since different ones could change in different samples), one would see a consistent proteomic profile at the level of protein complexes. This insight has led to successful novel network-based approaches to overcome consistency and recovery challenges that have so far plagued untargeted proteomic profiling [GLCW12, GLRS12, GW16].