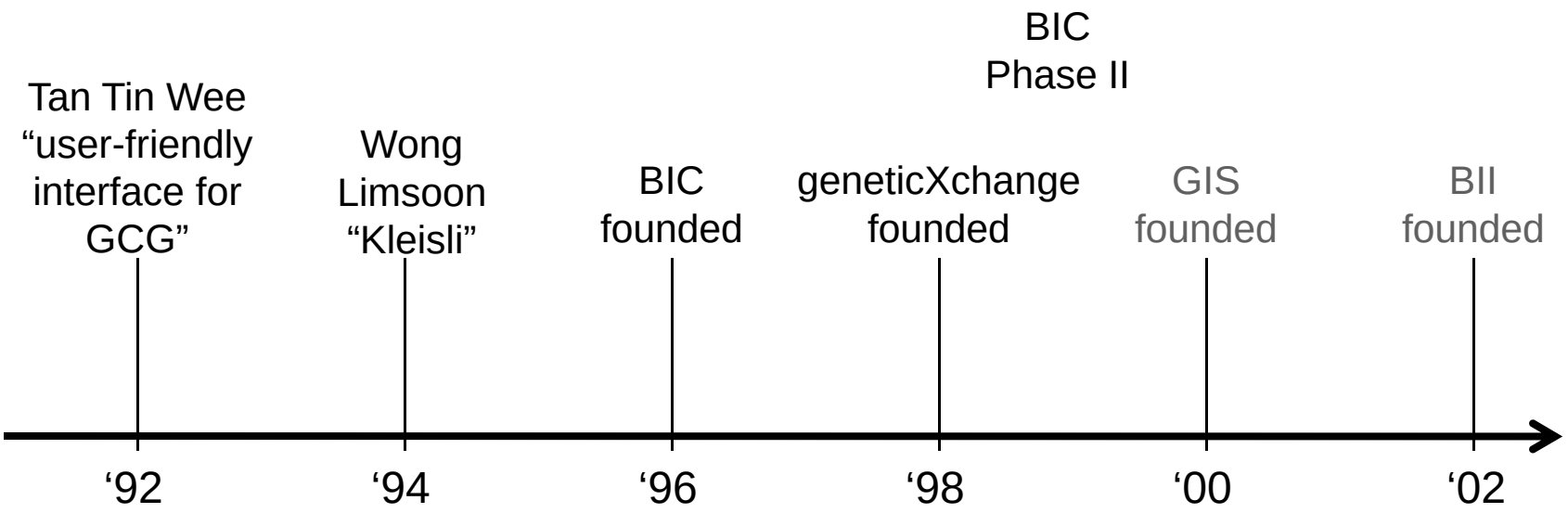
Singapore Bioinformatics in the Early Days

Limsoon Wong

17 March 2011

Significant Milestones



NB. There were a couple of earlier papers in CABIOS from Singapore. But those werent bioinformatics ones.

1st Bioinformatics Paper from S'pore

Tan Tin Wee

CABIOS Vol.9. no.5. 1993 Pages 581 – 586

A general UNIX interface for biocomputing and network information retrieval software

B.K.Kiong and T.W.Tan[1,2]

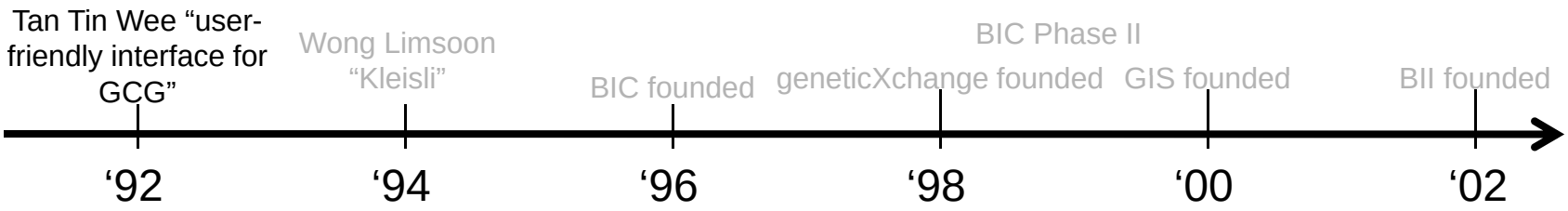
Abstract

We describe a UNIX program, HYBROW, which can integrate without modification a wide range of UNIX biocomputing and network information retrieval software. HYBROW works in conjunction with a separate set of ASCII files containing embedded hypertext-like links. The program operates like a

network access to remote databases. This has been hitherto difficult to achieve.

One strategy that is frequently adopted for proprietary packages in the design of unified interfaces is by writing new software or rewriting existing programs. However, the pace at which new software is being generated is faster than what a commercial package can reasonably incorporate and distribute.

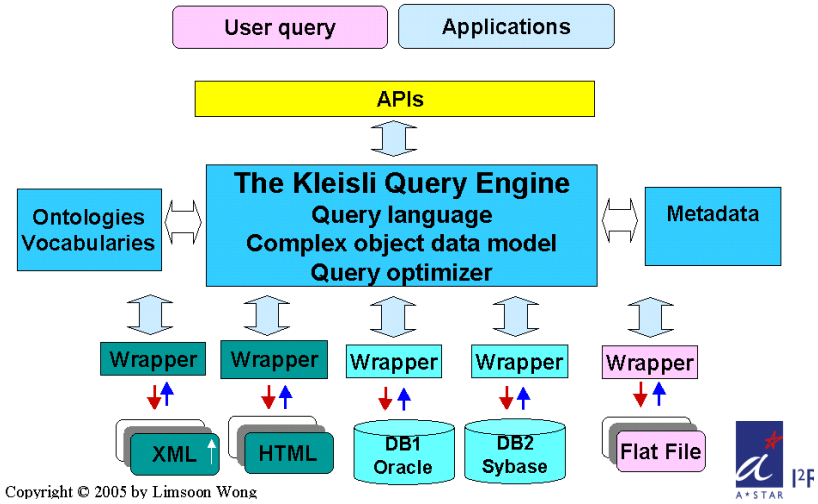
TTW is now Deputy Head of Biochem Dept in NUS



Kleisli --- the 1st Major Bioinformatics Solution from S'pore

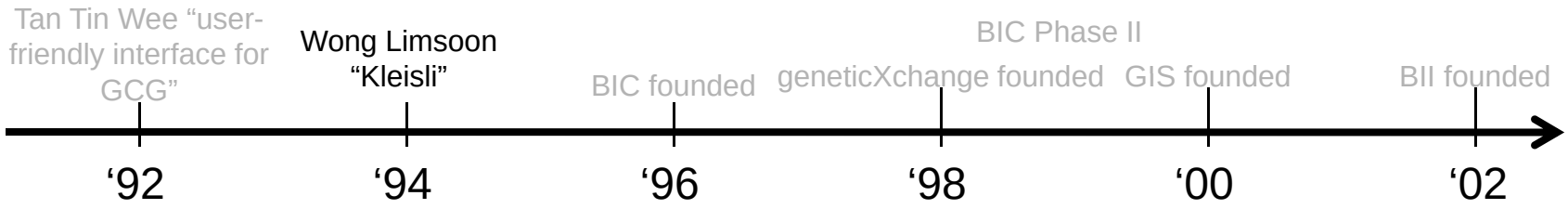
Wong Limsoon

G. Christian Overton



Buneman, Davidson, Hart, Overton, Wong, VLDB '95

WLS is now Head of CS Dept in NUS
GCO passed away in 2000



Integration: What are the problems?

A US DOE “impossible query”, circa 1993:

For each gene on a given cytogenetic band, find its non-human homologs.

source	type	location	remarks
GDB	Sybase	Baltimore	Flat tables SQL joins Location info
Entrez	ASN.1	Bethesda	Nested tables Keywords Homolog info

Data Integration Results

- **Using Kleisli:**
  - Clear
  - Succinct
  - Efficient
- **Handles**
  - heterogeneity
  - complexity

```
sybase-add (#name:"GDB", ...);
create view  L from locus_cyto_location using GDB;
create view E from object_genbank_eref using GDB;
select
    #accn: g.#genbank_ref,   #nonhuman-homologs: H
from
    L as c,  E as g,
    {select u
     from g.#genbank_ref.na-get-homolog-summary as u
     where not(u.#title string-islike "%Human%") &
           not(u.#title string-islike "%H.sapien%")} as H
where
    c.#chrom_num = "22" &
    g.#object_id = c.#locus_id &
    not (H = { });
```

NUS BioInformatics Centre

Desai N.

influ ence

Chu Swee Yeok

Sup- port

Philip Yeo

fund

Tan Tin Wee

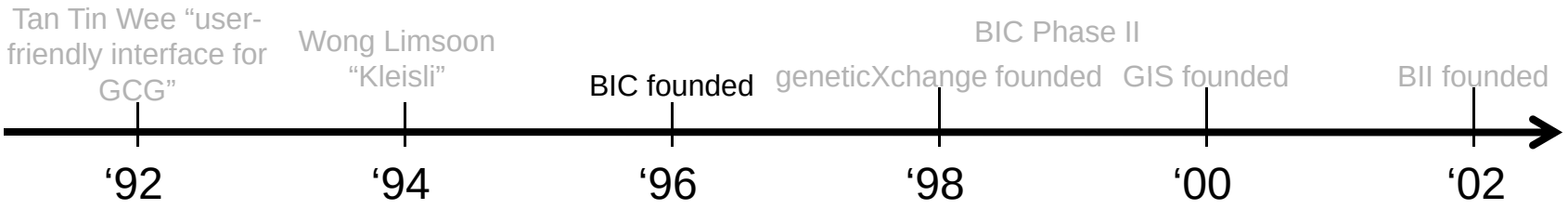
S. Subbiah

Baey Yam Keng

PY became chairman of A*STAR. He is now chairman of Spring

Wong Limsoon

BYK is now a Member of Parliament



Some BIC Stories

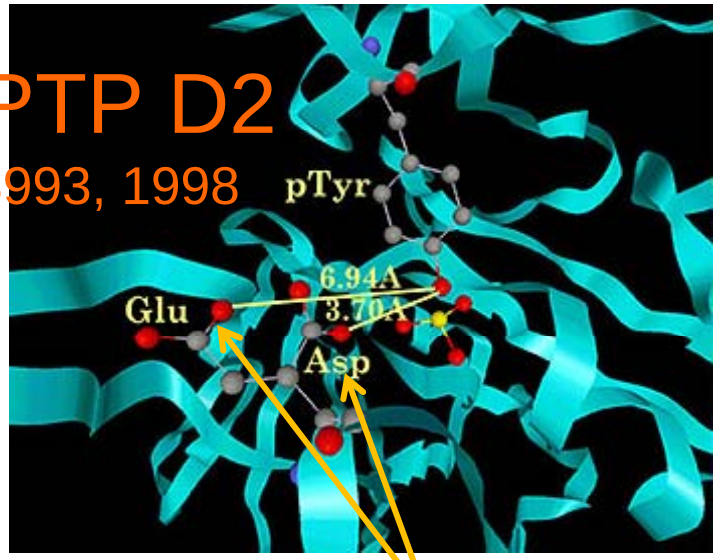
- **Provided bioinfo resources for NUS**
- **Developed Kleisli**
- **Made first steps into comp bio**
  - 1st structure solved in S'pore
  - WebPHYLIP
- **Recruited talent**
  - 1st "algorithmic" comp biologist
  - 1st "structural" comp biologist

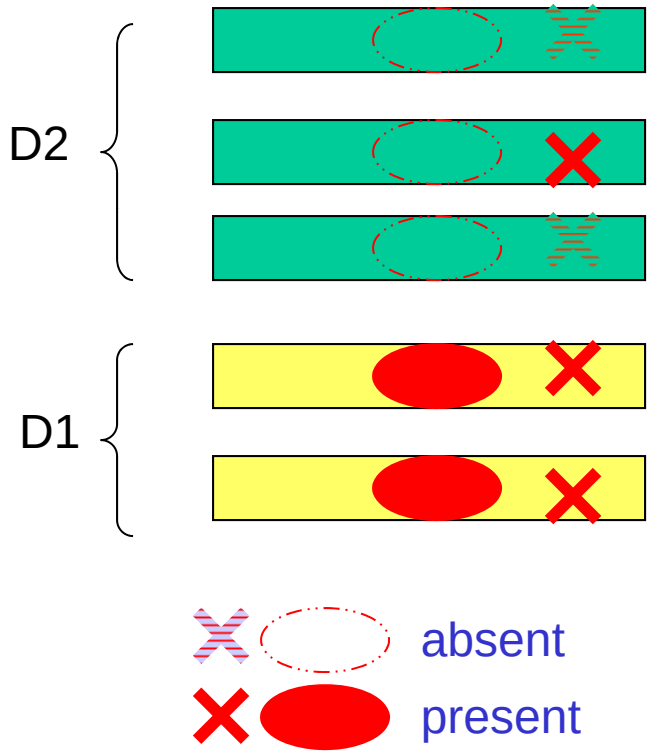
Prasanna Kolatkar

Zhang Louxin

Key Mutation Sites of PTP D2

K.L.Lim et al., *JBC*, 273:28986--28993, 1998






```
                                  ?    !   ?
gi|00000|P D2     QFHFHGWPEVGIPSDGK(
gi|126467|        QFHFTSWPDFGVPFTPI(
gi|2499753        QFHFTGWPDHGVPYHAT(
gi|462550|        QYHYTQWPDMGVPEYAL]
gi|2499751        QFHFTSWPDHGVPDTTD]
gi|1709906 D1     QFQFTAWPDHGVPEHPT]
gi|126471|        QLHFTSWPDFGVPFTPI(
gi|548626|        QFHFTGWPDHGVPYHAT(
gi|131570|        QFHFTGWPDHGVPYHAT(
gi|2144715        QFHFTSWPDHGVPDTTD]
                  * ..  **. *.*
```

Joe Chow

Beh Swan Gin

open doors

CEO

S. Subbiah

BSG is now a MD of EDB

$$$

technologies

Wong Limsoon

Tan Tin Wee

- **Lessons**
  - Biz plan
  - Mgmt
  - Tech renewal

Tan Tin Wee "user-friendly interface for GCG"

Wong Limsoon "Kleisli"

BIC founded

BIC Phase II

geneticXchange founded

GIS founded

BII founded

'92 '94 '96 '98 '00 '02

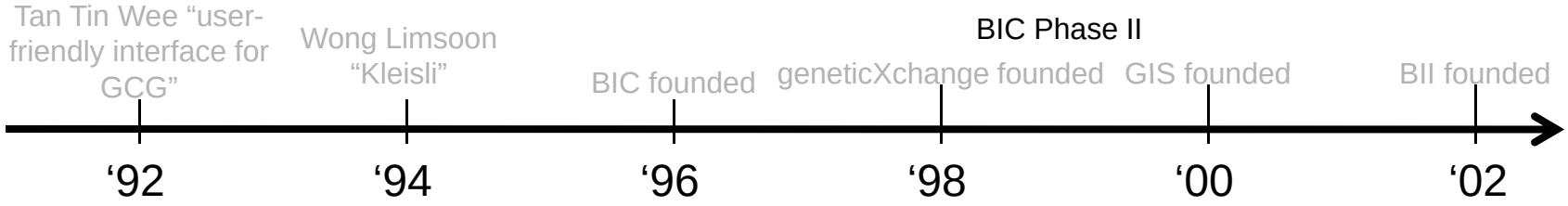
BIC Phase II

Wong Limsoon

Peter Saunders

PS was Deputy Director S'pore Health Services when he passed away in mid 2000s

- **Pioneer a new area in comp bio → immunoinformatics**
- **Pioneer a new area in bioinfo → bioNLP**
- **Leap frog in an established topic**
- **Compete in an emerging topic**

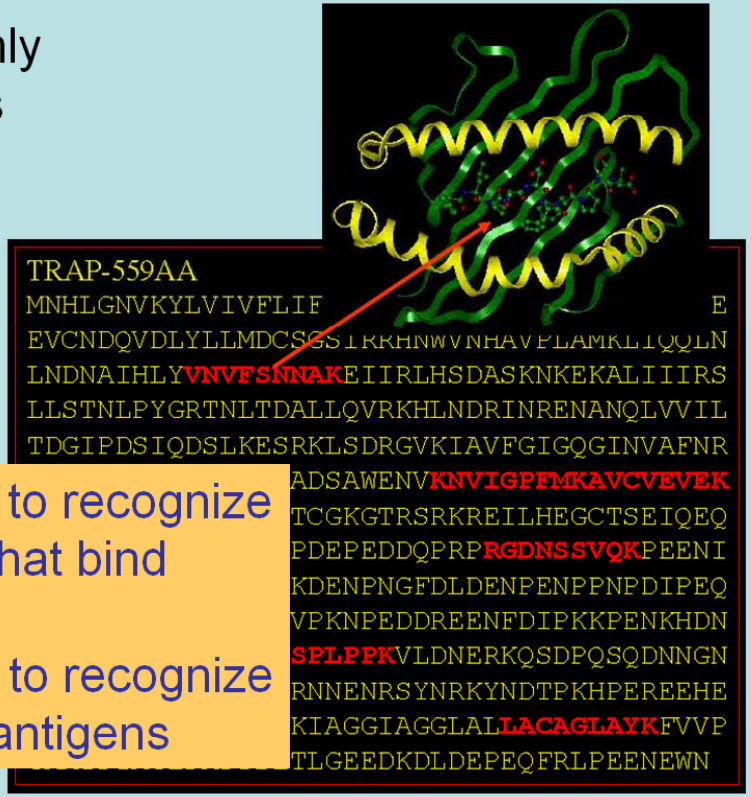


Pioneering Immunoinformatics

Predict epitopes, Find vaccine targets

Institute for Infocomm Research A*STAR

- Vaccines are often the only solution for viral diseases
- Finding & developing effective vaccine targets (epitopes) is slow and expensive process



- Develop systems to recognize protein peptides that bind MHC molecules
- Develop systems to recognize hot spots in viral antigens

Vladimir Brusic

VB is now Director of Bioinfo at Dana Farber Cancer Vaccine Center

Victor Tong who took over from VB won a 2009 SYA for immunoinfo

Pioneering BioNLP, Starting Molecular Connections

Jignesh Bhate

CEO

Ng See-Kiong

Wong Limsoon

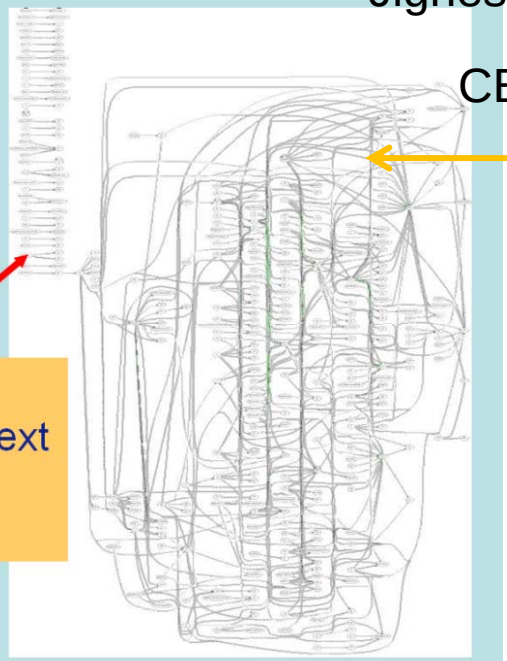
tech & proc

Understand proteins, Fight diseases

- Understanding function and role of protein needs organised info on interaction pathways
- Such info are often reported in scientific paper but are seldom found in structured databases

- Knowledge extraction system to process free text
- extract protein names
- extract interactions

Jak1

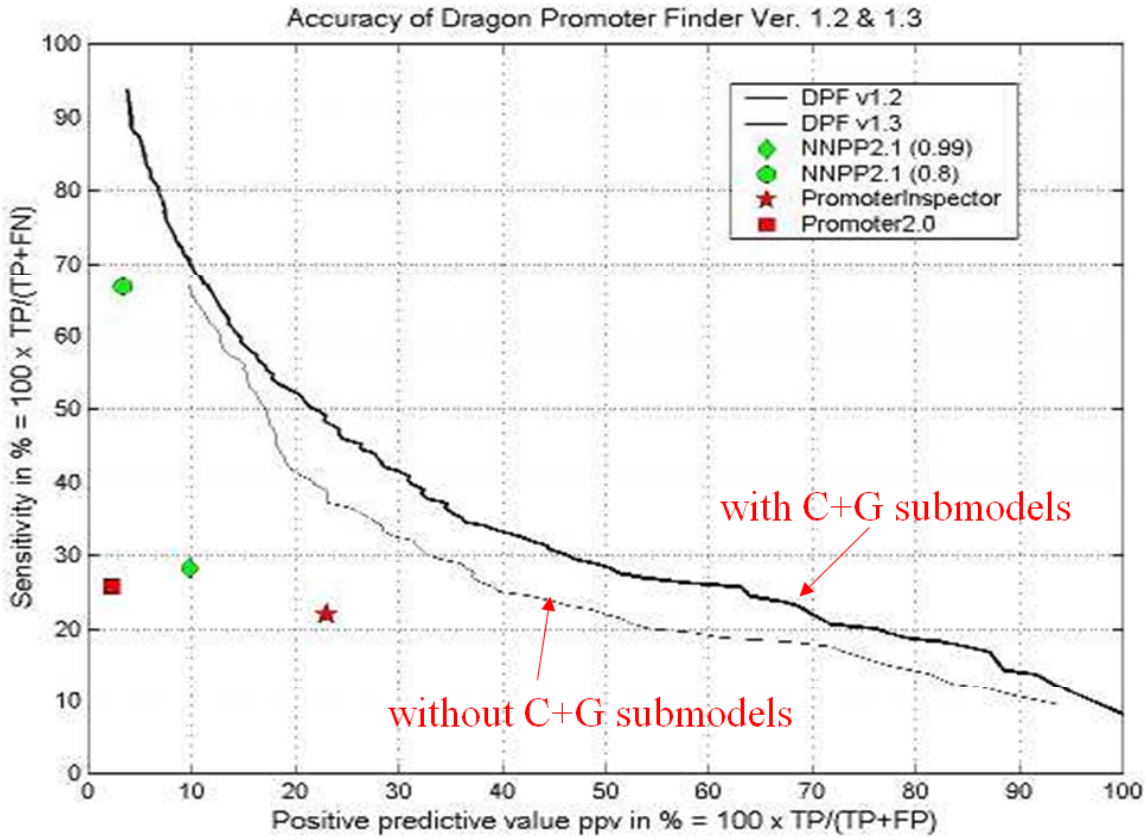


- **MC has grown profitably to ~700 curators and engineers**

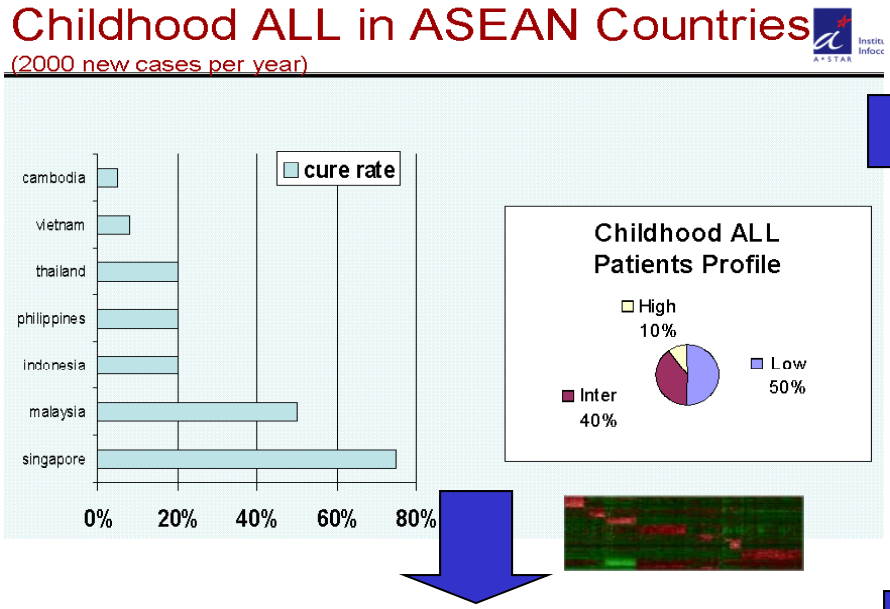
Leapfrog in Promoter Finding

Vladimir Bajic

VBB became Director of SANBI. He is now Director of Comput Bio Science Center at KAUST



Compete in an Emerging Topic – Gene Expression Analysis



Conventional Tx:
- **intermediate intensity to all**

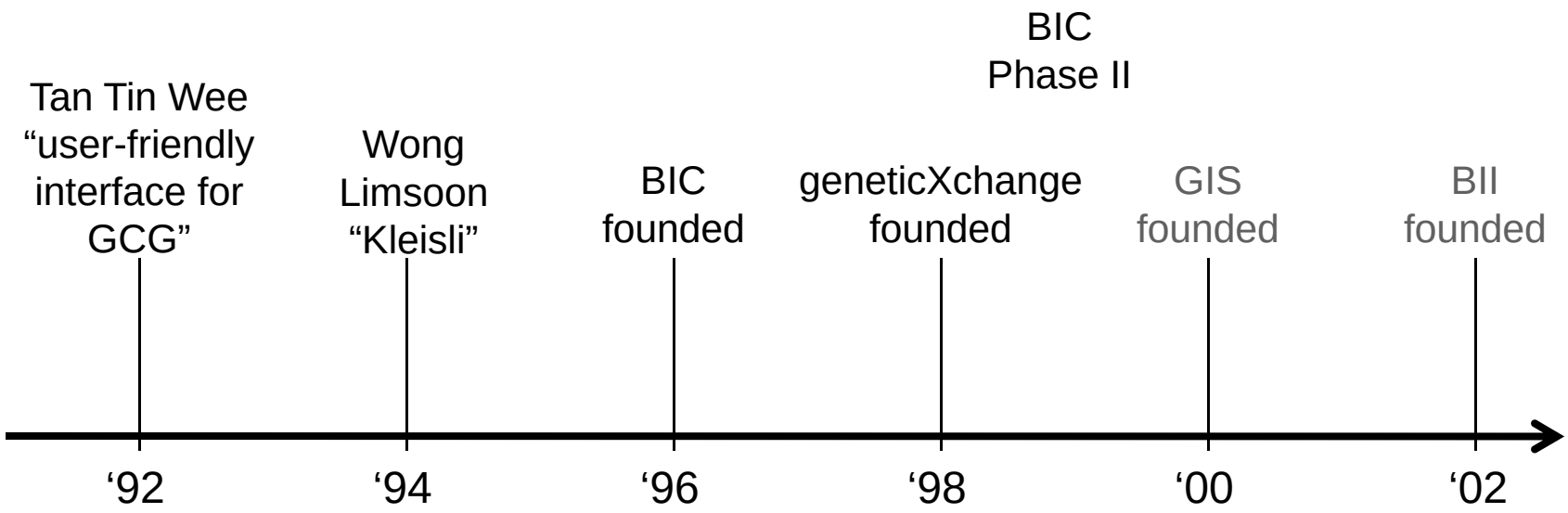
⇒ **10% suffers relapse**
⇒ **50% suffers side effects**
⇒ **costs US$150m/yr**

Bioinformatics-optimized Tx:
- **high intensity to 10%**
- **intermediate intensity to 40%**
- **low intensity to 50%**
- **costs US$100m/yr**

•High cure rate of >80%
• Less relapse
• Less side effects
• Save US$51.6m/yr

Yeoh et al, *CANCER CELL*, 2002

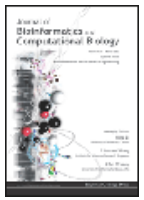
The first 10 years were an exciting fruitful experience…



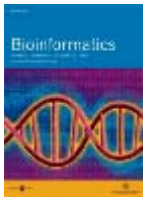
Today, there are ~100 bioinfo & comp bio RSEs in Singapore. Their activities in 2008/09…

- **Journals edited:**

TCBB



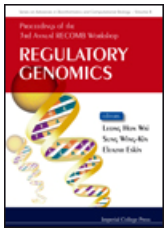
JBCB



Bioinformatics

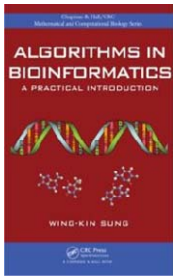
DDT

- **Books/Proceedings edited:**



3rd Regulatory Genomics

RECOMB'08



Algo in Bioinfo

- **Involved in ~20 bioinformatics conf prog & org committees**
  - RECOMB08, ISMB08/09, CSB08/09, GIW08/09, APBC08/09, …
- **Published ~70 papers**
  - Bioinformatics, JCB, BMC, JBCB, DDT, PLoS CB, EMBO, Cell Stem Cell, NAR, Cell, JBC, TKDE, VLDBJ, SIAM J Comput, Clin Cancer Res, …
- **~30 keynotes & invited talks in conferences**

NUS National University of Singapore

A photo from my first meeting with …

BGRS 1998, Siberia