



IT Innovations for Better Life



LAB OF System Biology & Network Biology

中央研究院資訊科學研究所 @iis, Academia Sinica, TAIWAN

系統生物學與網路生物學實驗室

PI: Chung-Yen Lin
2018 Dec

Bioinformatics in



• 資訊科學研究所生物資訊實驗研究群

Principal Investigators:

宋定懿 Ting-Yi Sung(Chair)

何建明 Jan-Ming Ho

林仲彥 Chung-Yen Lin

施純傑 Arthur Chun-Chieh Shih

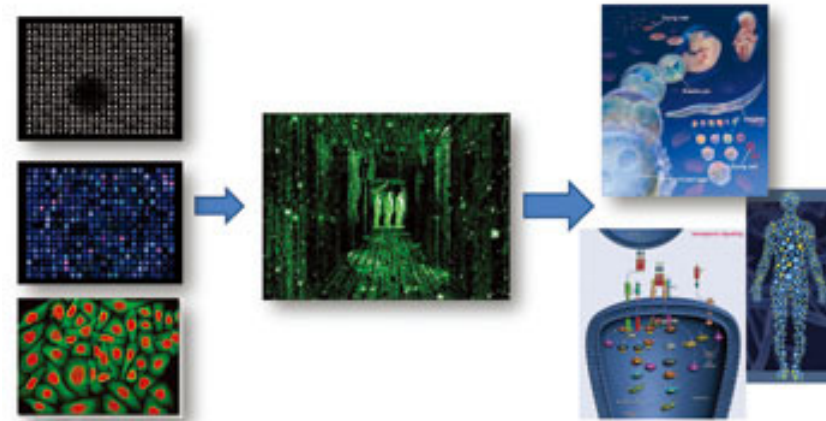
許聞廉 Wen-Lian Hsu

蔡懷寬 Huai-Kuang Tsai

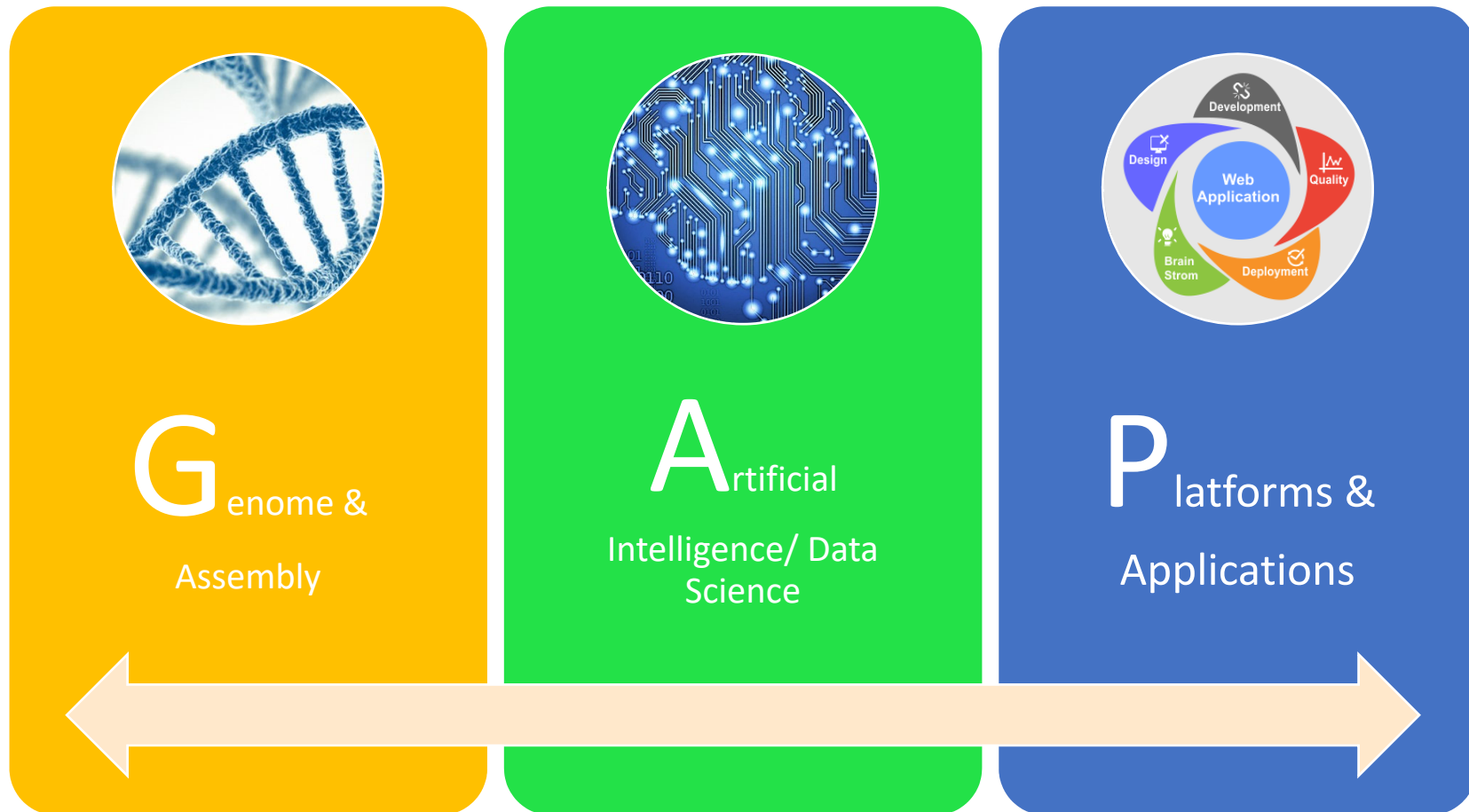


- 我們的研究是以資訊技術為主，針對不同生物體學(omics) 上的生物醫農問題進行解析，大致分為：

- 利用新世代序列定序儀進行基因體和轉錄體註解與分析
- 基因體/ 轉錄體重組演算法與程式開發
- 調控機制與網路
- 質譜儀分析的蛋白體學及代謝體學
- 蛋白質結構與功能預測
- 蛋白質交互作用網路分析
- 膜蛋白知識庫平台
- 人工智慧etc....



2016-2018





Genome Assembly

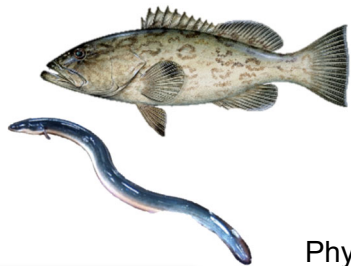




Initiate Genome Projects for Economic Important Organisms in Taiwan

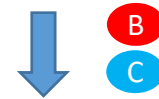
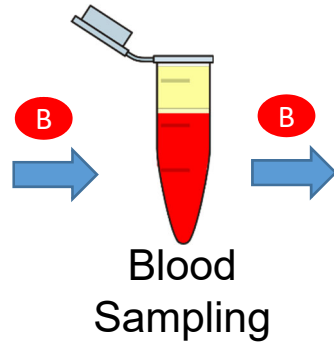
B : Knowledge of Biology
C : Bioinformatics/ IT

Supported by  創新轉譯農學研究計畫
Innovative Translational Agricultural Research Program



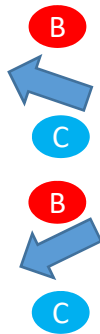
Physiological Regulation under Stress and infection

疾病防治



Genetic Breeding for Fries with disease resistance

分子育種









Construct Web database



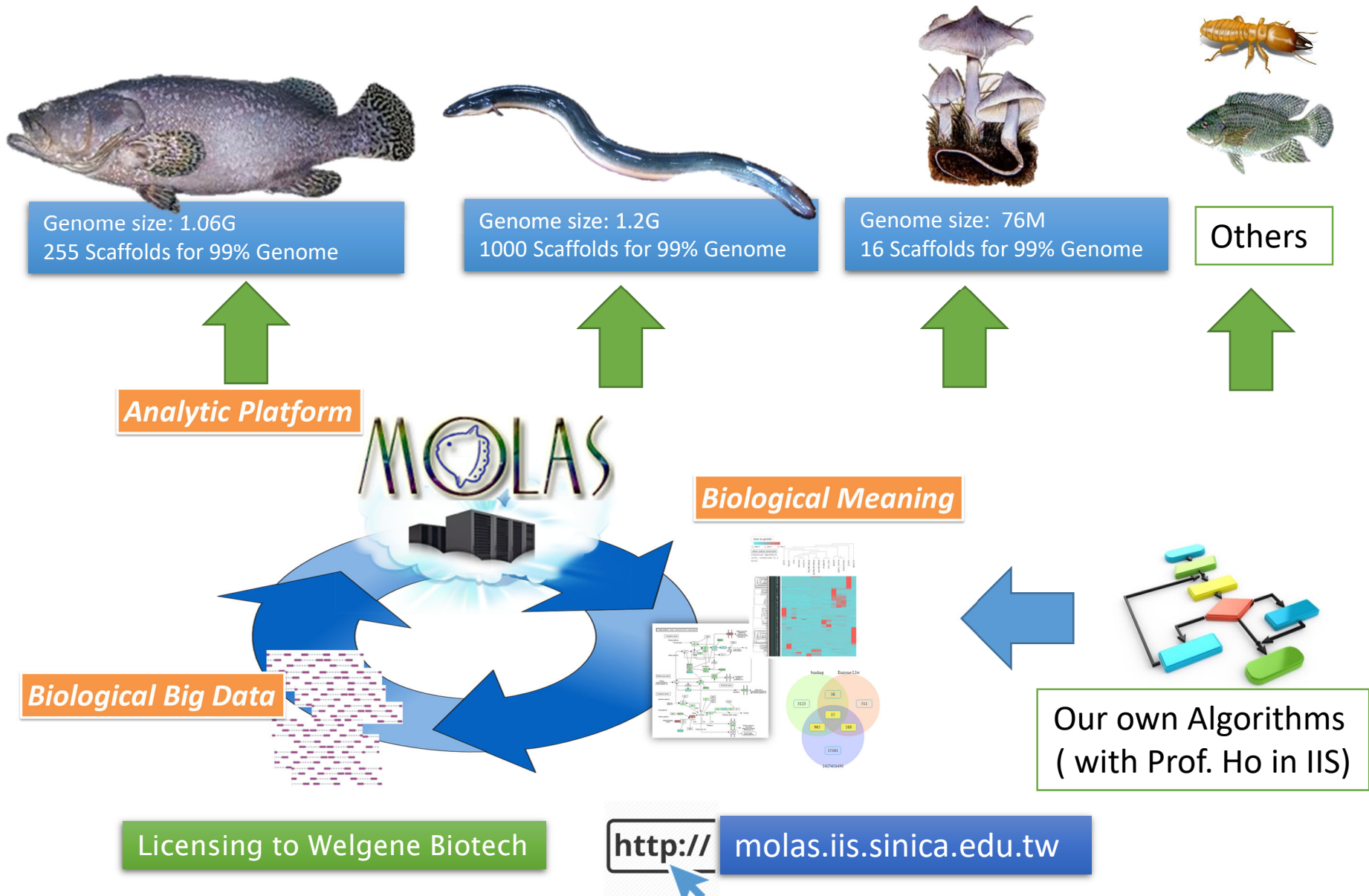
Data integration and Annotation



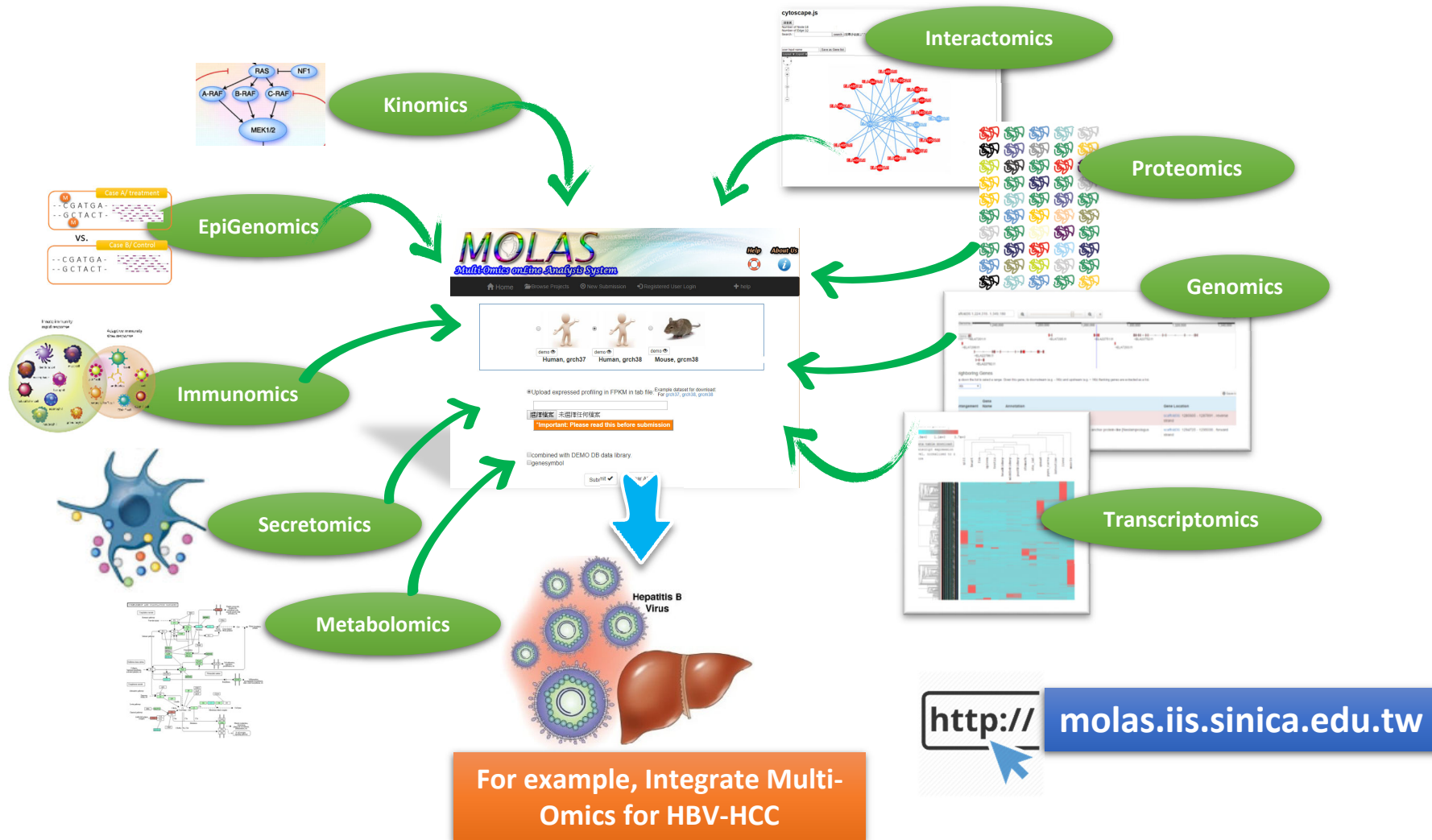
Introduce 3rd Sequencing Technology to Improve the Quality of Genome Assembly

Genome (Size) (# of Chromosome)	2017, Based on 2 nd Sequencing Tech, Illumina	2018, Introducing 3 rd Sequencing Tech, Pacbio
 Japanese Eel (1G)(19 pairs)	1,710 	948 scaffolds for 95% of Genome
 Giant Grouper (1G) (44 pairs)	308 	154 scaffolds for 99% of Genome
 Mushroom (Unknown) (76M)	34 	11 scaffolds for 99% of Genome

Assembly and Annotations for Several Genome



Implementation of Multi-Omics Platform for Biomedical Research





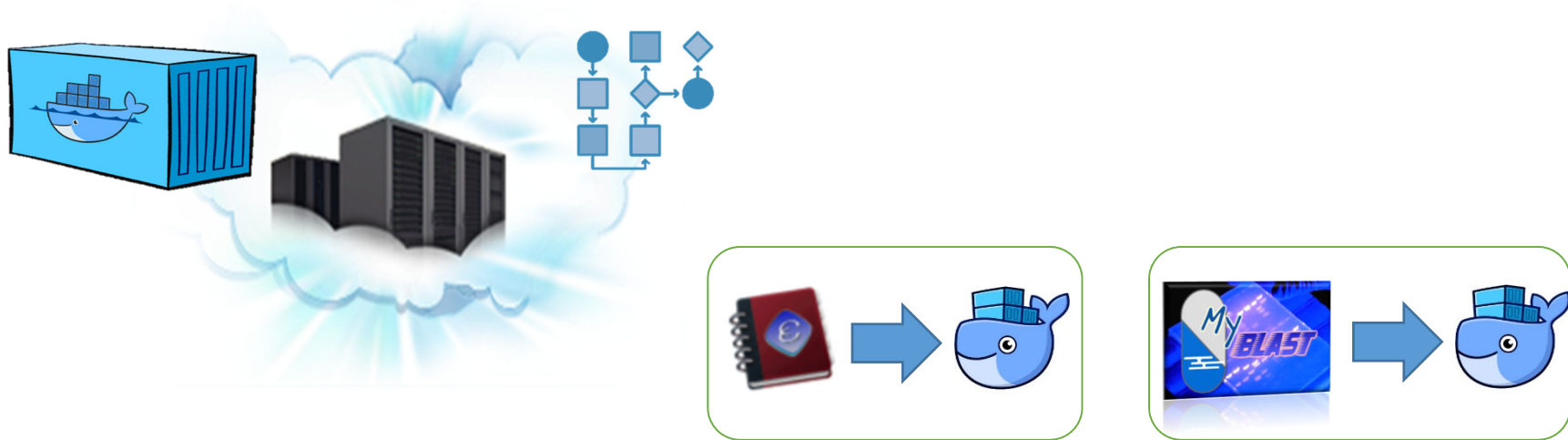
AI & Data Science



The Applications of AI/ Data Science on Biomedical Research

Toolbox on Cloud for High throughput Biology

2017 Seed Grant Proposal, Grand Challenges in Data Science
Research by Academia Sinica (資料科學種子研究計畫)

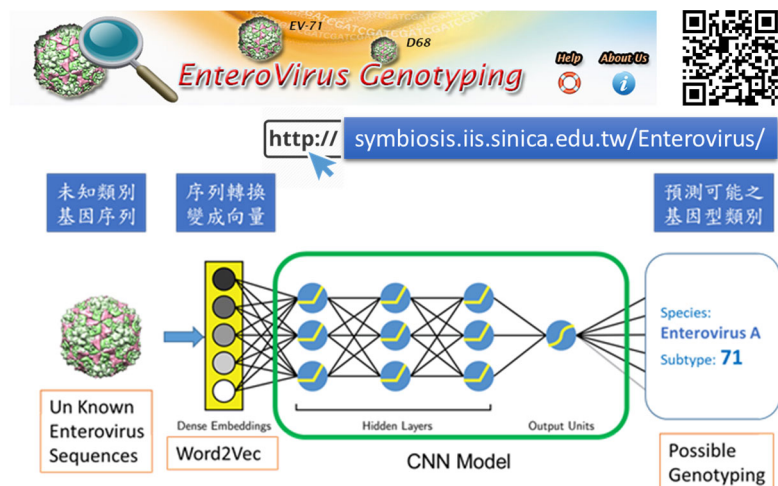


[http:// hub.docker.com/u/lbnnb/](http://hub.docker.com/u/lbnnb/)

以深度學習機制解析巨量腸病毒序列資料並建立基因型鑑別機制與網路即時分析平台

Analyze big data of Enterovirus Genome in AI and construct the web platform for real-time genotyping

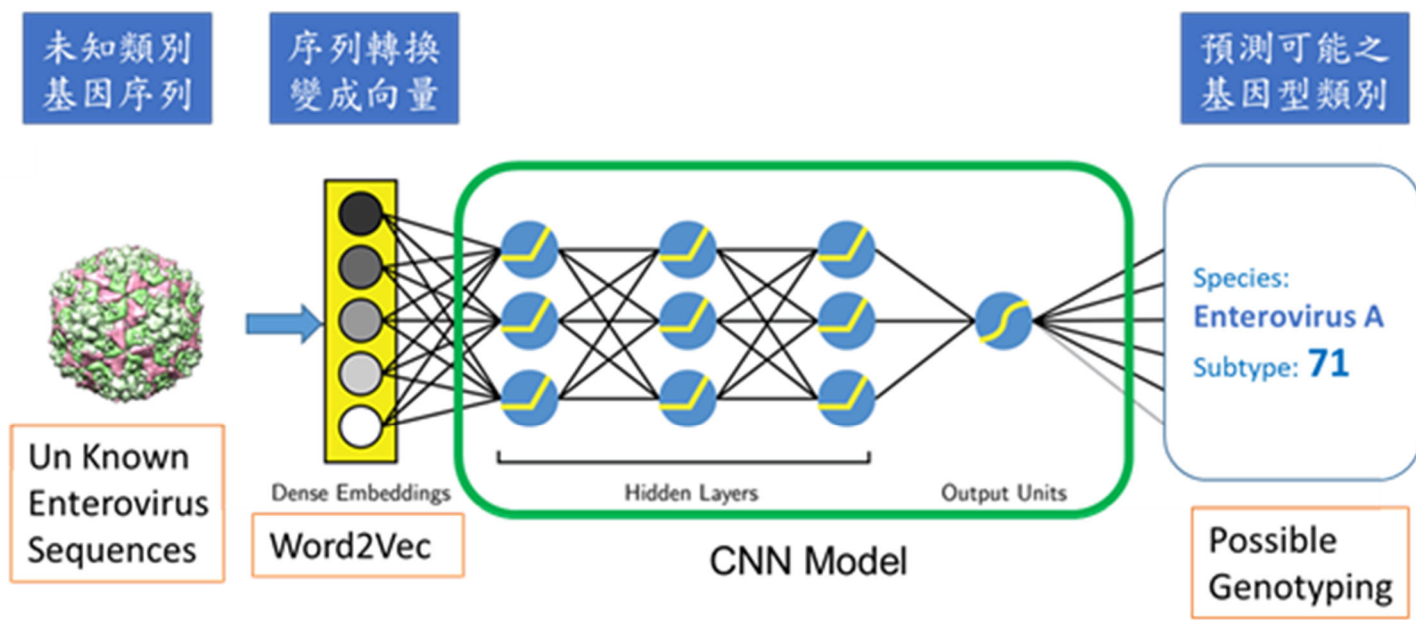
研發團隊整合深度學習模型及相關自有程式，建構了第一個以腸病毒基因體大數據為基礎的線上分類網站。此一線上腸病毒基因型分型平台，將能結合第三代小型低價定序儀(如Nanopore)，應用於在腸病毒感染初期(約三到五天內)的檢測上，應能及早發現所感染的病毒類型，如屬高致病品系，便能提早準備，防止重症的發生，減少寶貴生命受損的機率，同時也降低可能的醫療費用，並有效防禦及控制疫情，且減低醫療和社會經濟成本，進而實現預防醫學的目標。本成果可應用的領域業別，將包含第一線診所與醫院，生醫檢驗產業，及病毒學研究團隊與流病監控團隊。本研究為中研院資訊所林仲彥博士與國衛院熊昭特聘研究員團隊之合作成果，並榮獲本年度科技部未來科技獎，同時應邀於2018未來科技展展出。



Analyze Big Data Of Enterovirus Genome In AI And Construct The Web Platform For Real-time Genotyping



<http://symbiosis.iis.sinica.edu.tw/Enterovirus/>



Enterovirus Genotyping Online



<http://symbiosis.iis.sinica.edu.tw/Enterovirus/>

Portal for EV sequences

Input FASTA: FASTA Sequence

or Fileupload(*.txt): 選擇檔案 未選擇任何檔案 download example

Your Email (optional): Email

I have read and accept the Terms of Use.

Submit Reset Demo

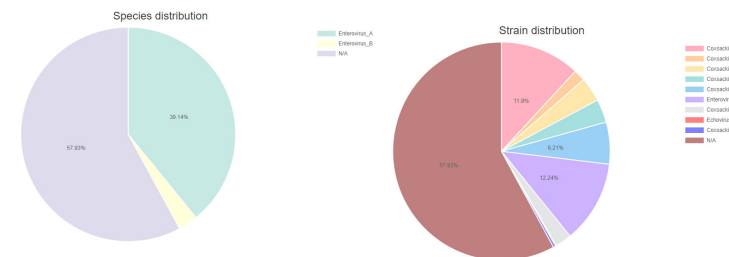
Total upload: 580 ,Enterovirus: 244 ,N/A: 336

Show 10 entries

ID	Species	Strain	
EV 16 00001_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	
EV 16 00002_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	
EV 16 00003_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00004_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00005_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00006_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00007_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00008_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00009_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00010_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99

Showing 1 to 10 of 580 entries

Previous 1 2 3 4 5 ... 58 Next



Prediction Results with Distribution of Strains

- Although the macro-average of prediction accuracy by five folds cross-validation (CV) is around 80%, the precision rates for severe strains like EV-71 and D-68 are up to 96.5% and 91%, respectively.
- Moreover, we have constructed the first web application in deep learning that is fully automatic to provide precise and rapid prediagnosis on EV genotype.



Platforms & Applications



Provide DOCKER Images for Big Biological Data

The screenshot shows the GitHub profile for 'lsbnb' (Institute of Information Science, Academia Sinica). The profile includes a bio, location (Taipei, Taiwan), website (http://eln.iis.sinica.edu.tw), and join date (June 2017). A table of repositories is displayed, with two specific repositories circled in orange and annotated with text boxes:

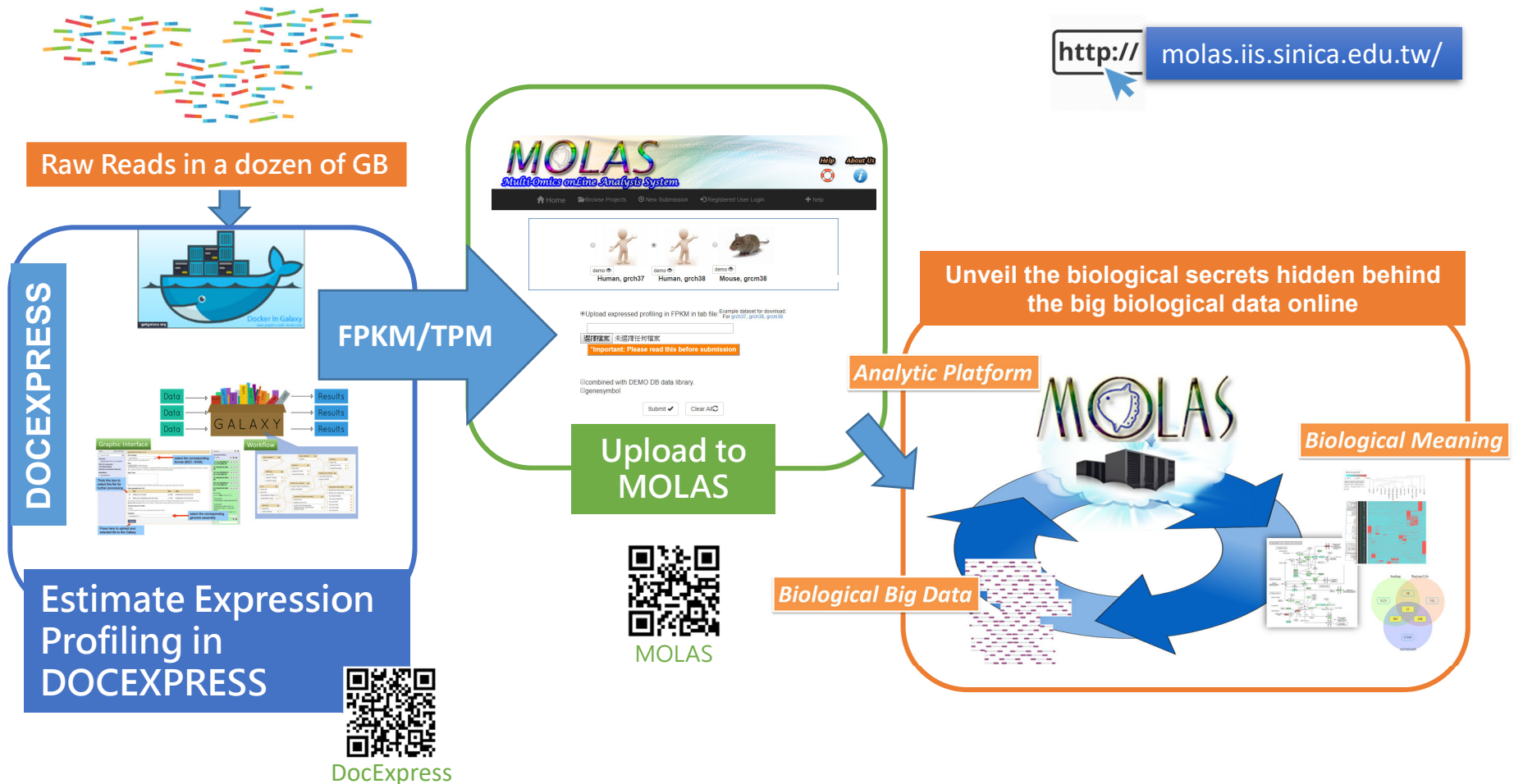
- lsbnb/galaxy-rnaanalysis**: Annotated with "For RNA-seq Data Pre-Process/轉錄體概況估計圖像化分析流程".
- lsbnb/docmethyl**: Annotated with "For WGBS Methylation Data Pre-Process/全基因體甲基化概況估計分析流程".

The repository table data is as follows:

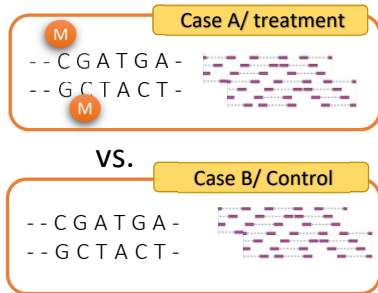
Repository	Stars	Pulls	Details
lsbnb/eln	1	95	>
lsbnb/myblast	0	77	>
lsbnb/galaxy-rnaanalysis	0	15	>
lsbnb/galaxy-epimolas-pe	0	13	>
lsbnb/ghostz_gpu	0	10	>
lsbnb/docmethyl	0	10	>

A blue button at the bottom right contains the URL: [http:// hub.docker.com/u/lsbnb/](http://hub.docker.com/u/lsbnb/)

Workflow and Platforms for Transcriptome Analysis



Workflow and Platforms for Whole Genome Methylation



Genome + GTF

Available Genome/ Gene Transfer Format (GTF) in epi-MOLAS



TAIR10



GRCh37,38



GRCm38



epiMOLAS

Raw Reads in fastq from WGS

DocMethyl (Web GUI)

Trim Galore

FastQC

Bismark

EpiMOLA
S
Extractor

Galaxy



EpiMOLAS (Web GUI)

mtable A

mtable B

mtable C

Upload mtable(s) via
Web

epi-MOLAS
Epi-genomics OnLine Analysis System

Show
Gene List

GO term
Enrichment

Gene	Chromosome	Start	End	Strand
BRCA1	17	41,276,112	41,490,517	-
BRCA2	12	24,320,000	24,440,000	+

GO term	Count	Adjusted P
GO:0003682	10	0.0001
GO:0003681	8	0.0005

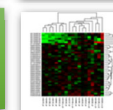
Methylated Genes
on Chromosome
map



KEGG Pathway
Enrichment



Venn
Diagram



HeatMap



DocMethyl

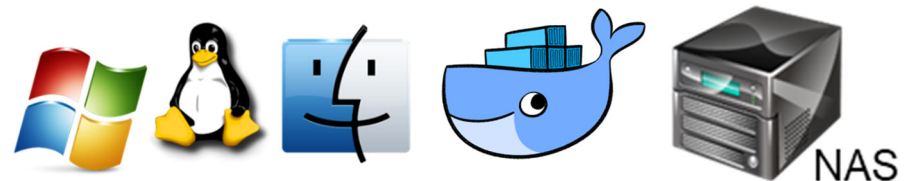
雲端共享架構的實驗室電子記事本

Electronic Laboratory Notebook on Cloud (Elegancy, ELN)

全球第一個提供多種平台的電子實驗紀錄平台 (ELN)，並完成 DOCKER 虛擬平台雲端版本的建置，可協助各型實驗室，以簡易的安裝方式，快速建置具高親和圖像介面、隱私性、資料互動分享、安全回復、防止竄改、會議時段預約與實驗室資源管理等模組之電子實驗室記錄協作平台，讓實驗室得以快速地將研究智慧與相關資源數位化。本研究為中研院資訊所林仲彥博士團隊之成果，並榮獲本年度國家新創獎，同時應邀於台灣醫療科技展展出。



<http://eln.iis.sinica.edu.tw/eln>



2018 Healthcare+ EXPO · TAIWAN 11.29 - 12.02 台北南港展覽館一樓 台灣醫療科技展

實驗室所產生的各類資料

儲存到雲端或個人電腦中

可在各種隨身設備上讀取修改

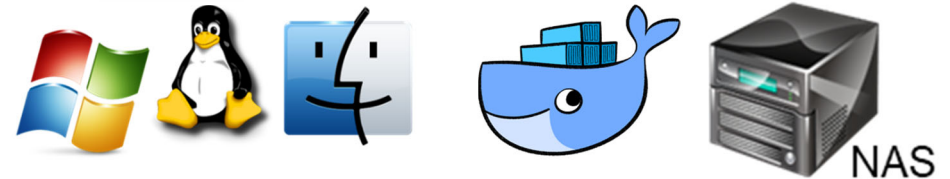


Electronic Laboratory Notebook (Elegancy, ELN)



[http:// eln.iis.sinica.edu.tw/eln](http://eln.iis.sinica.edu.tw/eln)

Digitalize Your Wisdoms



實驗室所產生的各類資料

儲存到雲端或個人電腦中

可在各種隨身設備上讀取修改



All information generated by Lab members



Migrate ELN to Cloud as multi-lab platform



Re-editing and Mining by Various Devices

Publications and Other Services (Nov, 2015- Nov, 2018)

- ✓ 12 Publications (BMC series, PNAS, Molecular Ecology..)
- ✓ 75,000 submissions/ usages for All web Applications & Databases (more than 20) world -wide
- ✓ Over 6,500 download for Cytohubba (New version, Starting from Jan , 2017) Collaborated with Prof. Kao@ IIS, AS



cytoHubba

Predicts and explores important nodes and subnetworks in a given network by several topological algorithms.

★★★★★ (123) 6475 downloads | posts | citations 3.0+ 

- ✓ Several Collaborations supported by

中華民國科技
Ministry of Science and Technology, R.O.C.



行政院
農業委員會

im.r

創新轉譯農學研究計畫
Innovative Translational Agricultural Research Program



線上資料庫 Web Databases



阿拉伯芥基因甲基化線上分析資料庫 (*BMC Genomics*, 2016) · <http://tea.iis.sinica.edu.tw>



細枝鹿角珊瑚與共生藻基因資料庫 (*Molecular Ecology*, 2016). <http://ips.iis.sinica.edu.tw/coral/>



半索動物轉錄體資料庫 (*Marine Genomics*, 2014). <http://molas.iis.sinica.edu.tw/hemichordate/>



龍膽石斑基因體與轉錄體資料庫, <http://molas.iis.sinica.edu.tw/grouper2015/>



日本鰻基因資料庫 (*PLoS ONE*, 2015) · <http://molas.iis.sinica.edu.tw/ipeel2016>



台灣鯛基因體資料庫 <http://molas.iis.sinica.edu.tw/tilapia2017/>



沉香基因表現資料庫 (*BMC Plant Science*, 2015) · <http://molas.iis.sinica.edu.tw/agarwood>



腸病毒基因分型與重組分析 (*BMC Genomics*, 2015) <http://symbiont.iis.sinica.edu.tw/evidence>



人類幽門桿菌蛋白質交互作用資料庫 (*Bioinformatics*, 2005). <http://dpi.nhri.org.tw/hp>



線上分析工具平台 Web Applications

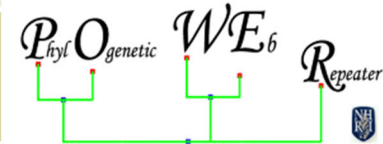
PDA: Primer design Assistant for PCR primer design, NAR 2003, <http://dbb.nhri.org.tw/primer>,
聚合酶反應核酸引子線上設計平台



myBLAST : customized databases and results management on your Desktop
在您的電腦上的高速序列比對平行化平台
web version: <http://mybioweb.nhri.org.tw/myblast>,
Windows/ MAC/ Linux version download
: <http://eln.iis.sinica.edu.tw>



POWER : Phylogenetic web repeater (NAR, 2005), <http://power.nhri.org.tw>
線上生物巨分子親緣關係分析平台



MySort : Estimate Immune Cell Composition (BMC Bioinfo, 2018)
<http://symbiosis.iis.sinica.edu.tw/mySORT/>
解析浸潤於癌組織中的免疫細胞組合



UPS 2.0: Unique Probe Selector (BMC Genomic 2010)
<http://array.iis.sinica.edu.tw/ups>
快速病原檢測開發平台/檢測晶片與基因晶片線上設計系統



PALM: Phylogenetic reconstruction by Automatic likelihood Model Selector. PLoS ONE 2009, <http://palm.iis.sinica.edu.tw>
最佳化親源分析高速分析平台



Enterovirus Genotyping in AI/
腸病毒基因型線上AI智慧定型
<http://symbiosis.iis.sinica.edu.tw/Enterovirus>

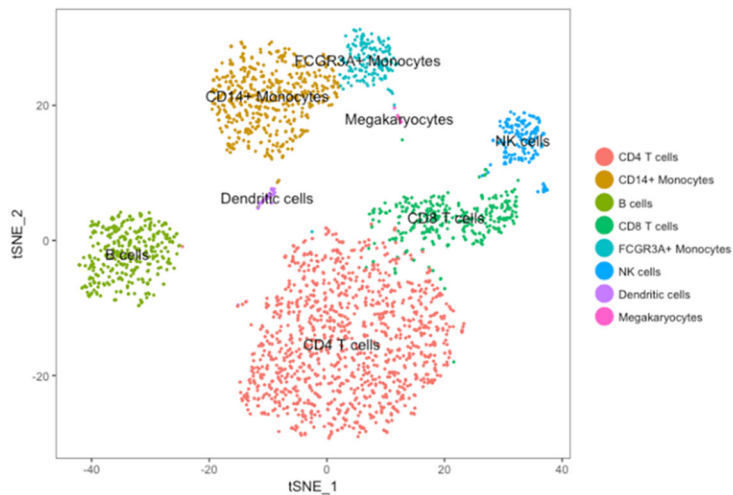


New Directions

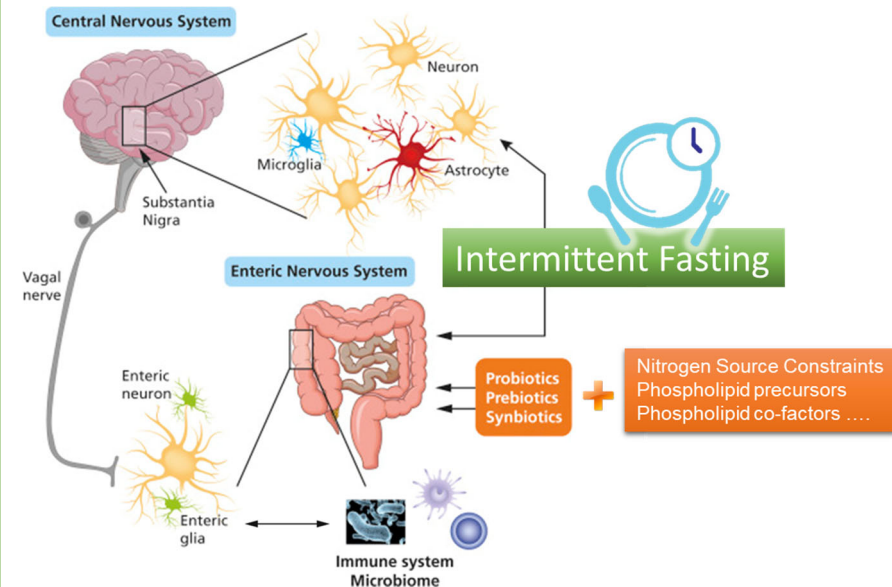


New Directions (2018-)

Single Cell analysis for cell-based therapy/ 用於細胞治療用之單細胞分析



Brain-Gut-Microbiome/ 腸道微生物腦軸腸線之研究



新世代長定序重組演算法之研發 與其在大型基因體定序策略之應用



計畫主持人：何建明
林仲彥

重大挑戰

定序技術的突破、大量數據(數百GB)的產生與價格的下降，已大幅改變原有基因體定序的策略，並進而提昇生物醫農領域基因體定序組裝的需求，以基因體更為全面的角度來解析生命的奧秘。

現有的限制

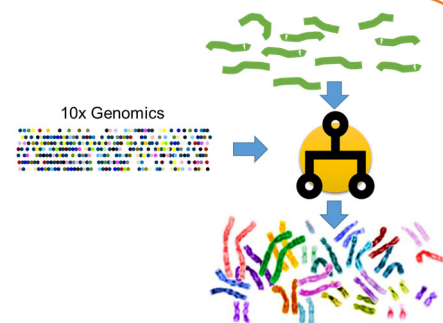
新的定序技術所帶來長度增長與產量遽增，使得原有的組裝演算法之基本假設大都不適用，也造成程式處理大量的長序列組裝時的效率低落，組裝品質也不佳。同時巨量序列與基因體模版之對應耗時，大量序列標誌(Barcodes)的同步處理不易，而且也沒有善用序列資料中所隱含的豐富資訊。

關鍵問題

如何在有限的定序經費與計算資源下，在定序策略(定序技術)與組裝品質之間取得平衡，是基因體研究領域當今所遇到的最大困境，這是我們的機會。

突破性的研究方案

將結合百萬個barcodes所對應的組裝序列條數(以10x Genomics所產出的短序列標誌連鎖組)，透過圖學拓樸分析，來重新建立原有二代組裝或第三代定序組裝基因體片斷，其長序列間的排列順序，並進行序列的遺缺修補，一舉提昇到組裝品質與條數到接近染色體數目的層次。



國際競爭力

在考量定序成本與基因體覆蓋倍率的平衡下，將能提供全新高品質、染色體層級之基因體組裝的新策略，建構第一個真正結合第三代與10x Genomics技術的新穎演算法，並提供國際研究社群使用。

Collaborations



國家衛生研究院
National Health Research Institutes





LAB OF System Biology & Network Biology

中央研究院資訊科學研究所

@iis, Academia Sinica, TAIWAN

系統生物學與網路生物學實驗室



Thanks for your Attention

