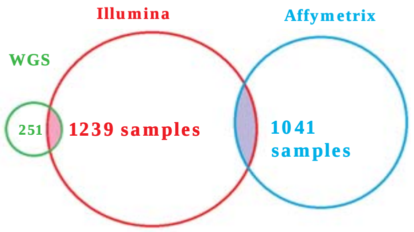
種豬性能育種用基因晶片

Gene Chip for Performance Evaluation of Breeding Pigs

2019/12/06
Genomics BioSci & Tech
Dr. Yu, Chairman

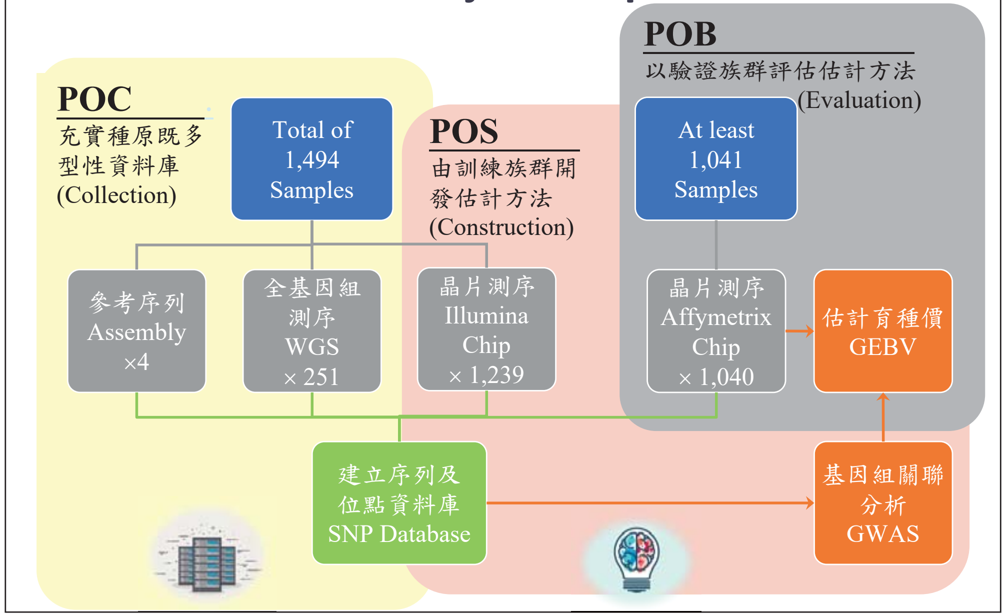
An Overview of Samples

平台 (Platform)		場域 (Donor)		品種 (Breed)		性別 (Gender)	
Illumina	1,239	ATRI	87	杜洛克 (D)	547	♂	382
		Central	342	藍瑞斯 (L)	417	♀	819
		Fuchang	734	約克夏 (Y)	237		
		Shueipo	76	未知 (?)	38	?	38
WGS	251	Central	132	杜洛克 (D)	94	♂	115
		Fuchang	119	藍瑞斯 (L)	111	♀	136
				約克夏 (Y)	46		
Affymetrix	1,041	ATRI	263	杜洛克 (D)	544	♂	721
				藍瑞斯 (L)	211		
		Central	426	約克夏 (Y)	134	♀	318
		Fuchang	109	黑豬 (P)	33		
				三元豬 (LYD)	6	?	2
		Shueipo	243	未知 (?)	2		
總樣本數 (total): 2,531							



○ : 251samples of WGS
○ : 1239samples of illumina
○ : 1041 samples of Affy
▨ : 32 samples of WGS + illumina
▨ : 92 samples of Affy + illumina

An Overview of Project Scope



POC
充實種原既多型性資料庫
(Collection)

Total of 1,494 Samples

POS
由訓練族群開發估計方法
(Construction)

POB
以驗證族群評估估計方法
(Evaluation)

At least 1,041 Samples

參考序列 Assembly ×4

全基因組測序 WGS × 251

晶片測序 Illumina Chip × 1,239

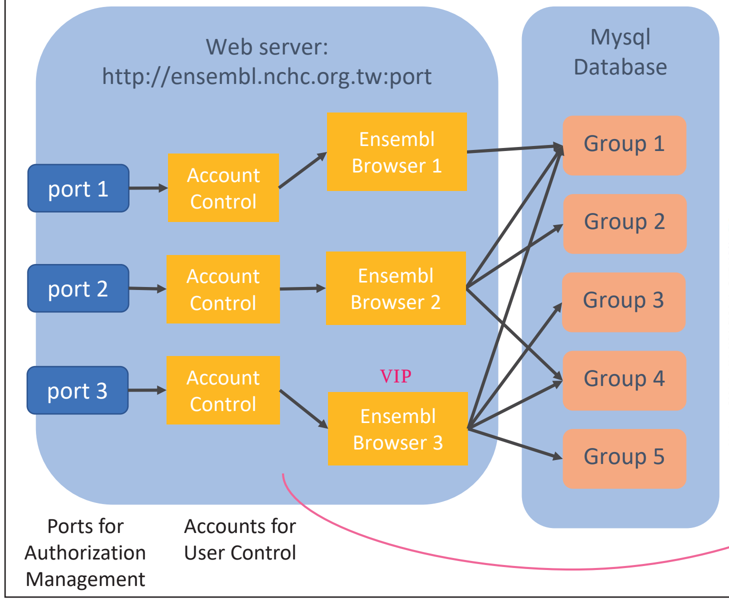
晶片測序 Affymetrix Chip × 1,040

估計育種價 GEBV

建立序列及位點資料庫 SNP Database

基因組關聯分析 GWAS

Integration and Utilization of Data



Web server:
http://ensembl.nchc.org.tw:port

Mysql Database

port 1
port 2
port 3

Account Control
Account Control
Account Control

Ensembl Browser 1
Ensembl Browser 2
VIP
Ensembl Browser 3

Group 1
Group 2
Group 3
Group 4
Group 5

Ports for Authorization Management

Accounts for User Control

rs332978863 as an example:

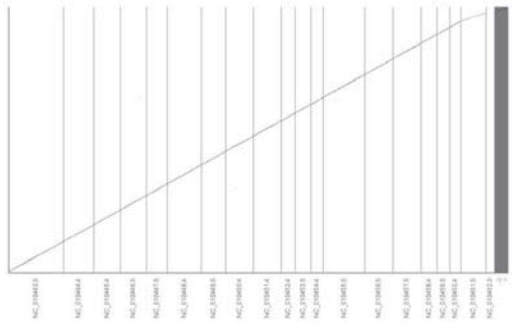
Database for Visualization & genotype management

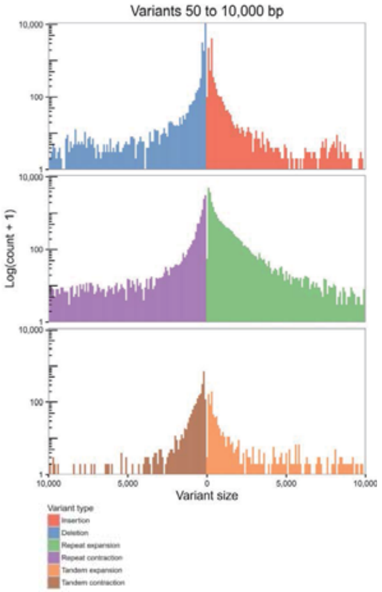
Reference-leveled De Novo Assembly

	Platform	Depth	Total length (Gb)	Total Gap (Mb)	Scaffold	
					Total number	N50 (Mb)
D208-3_v1	10X	82x	2.47	34.0	15,629	38.12
L1398-1_v1	10X	85x	2.47	33.9	16,069	35.57
Y1394-3_v1	10X	87x	2.48	36.6	17,833	35.11
高畜黑豬 (KHAPS black pig)	PacBio	69x	2.41	0	1,516	17.02
台灣水鹿 (Sambar Deer)	10X	90x	2.61	--	43,287	14.12
Sscrofa11.1 (Standard Ref)	PacBio	65x	2.55	29.9	706	88.23
USMARCv1.0	PacBio +Illumina	65x	2.76	132.3	14,157	131.46
Tibetan_Pig_v2	Illumina	131x	2.44	57.9	72,068	0.86
Bamei_pig_v1	Illumina	89x	2.46	27.1	129,335	1.53
Sscrofa10.2	Illumina	24x	2.81	289.4	9,906	0.58
Northern Sea Otter Genes (Basel). 2017 Dec; 8(12): 379	10X	60x	2.43	32.2	6,770	38.45

Deeper Understanding in Taiwan's Novel Breed

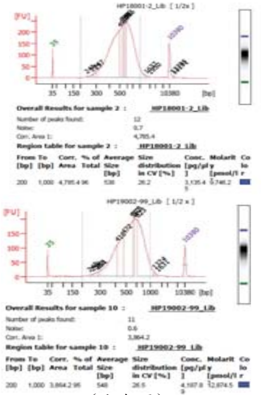
- 高畜黑豬 (K) vs. 參考序列 (*ref.* D)
  - 組裝的完整度及一致性都非常高
Highly completed and consist
  - 品種間的差異主要以小單位重複為主
Major variant is tri-nucleotide repeats
  - 有助了解黑豬品系特性，加速推廣擴散
Helps increase black pig popularity among farmers



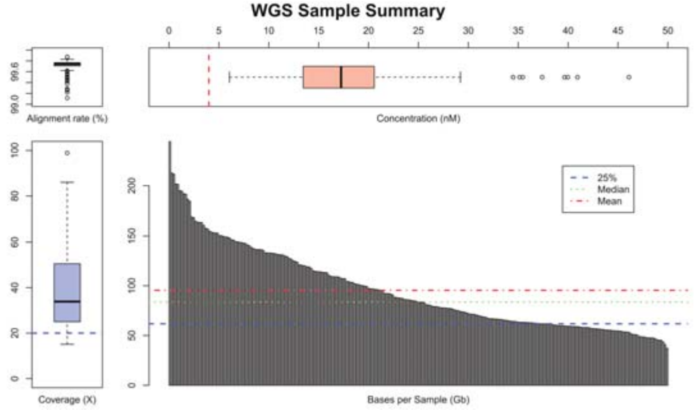


Extensive Whole-genome-shotgun Sequencing

Average	WGS (251)	KPI
Concentration (nM)	17.46	≧4nM
Read length / Library size (bp)	151 (library size = 470)	2 x 150bp / 470 ± 100bp
Coverage (X)	38.41 (total bases = 95.79Gb)	20
Q30 / Alignment rate (%)	92.85% / 99.72%	75% / 60%
Output data (Gb)	24,219.4	22,500



（代表圖）

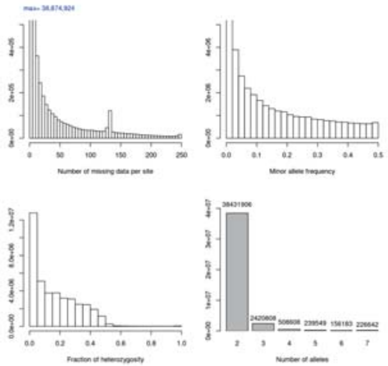


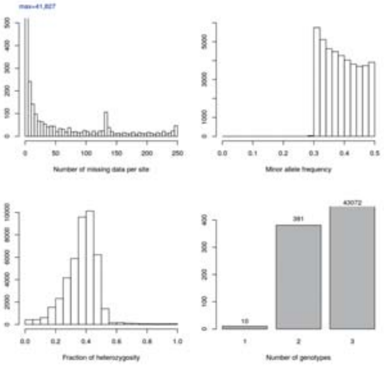
Enormous Variant Discovered

Remove non-chr SNP → Remove multiallelic SNP → Missing rate < 1% → MAF > 30% → Equal space 50k → Total 43,463 SNP

Initial SNPs: 41,983,696

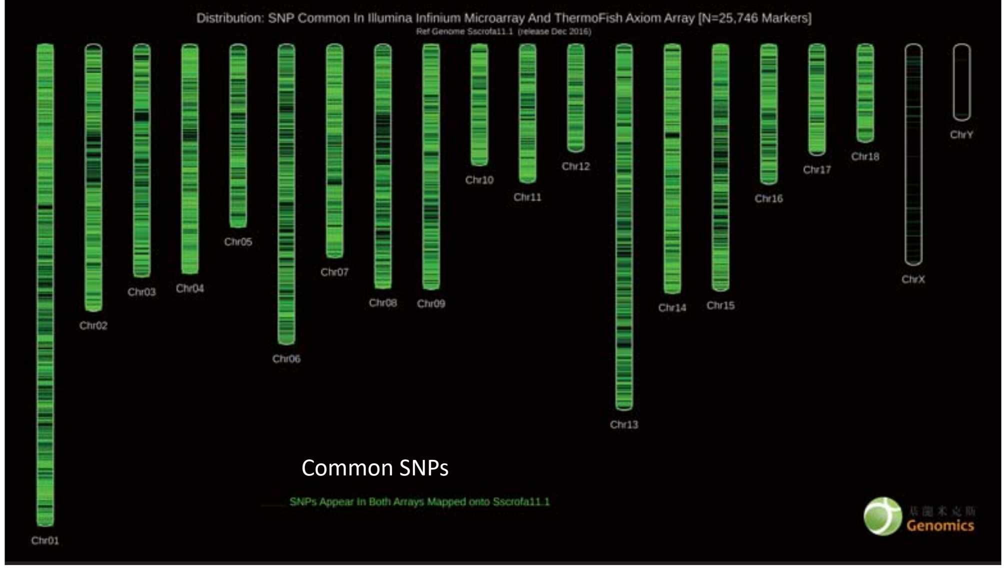
After Selection SNPs: 43,436





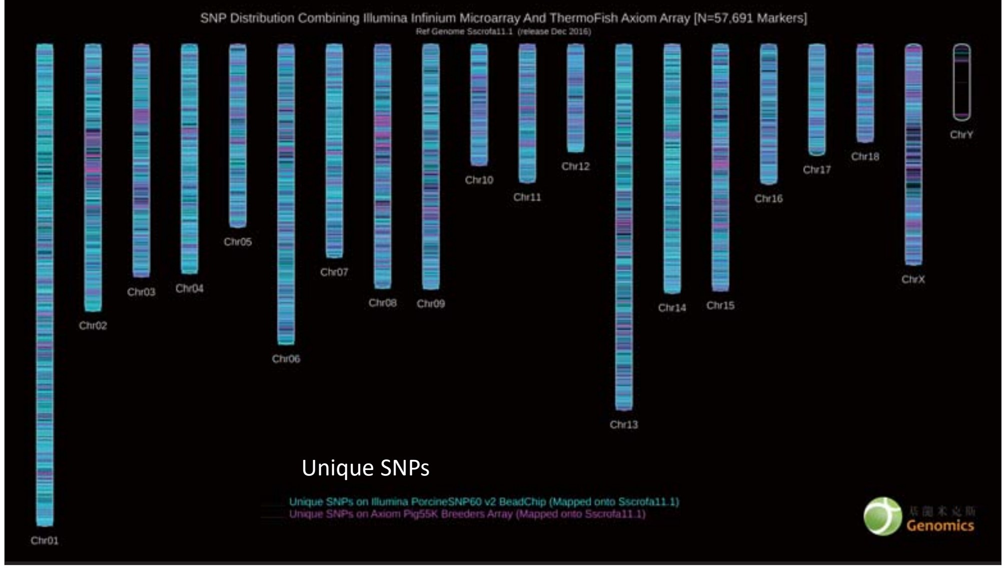
A/G:34.9%　A/C:　8.2%　A/T:　6.6%
C/T: 35.1%　G/T:　8.3%　C/G:　6.7%

Comparison of 2 Major Catalog Chips



Common SNPs

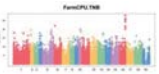
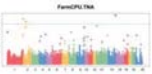
Comparison of 2 Major Catalog Chips



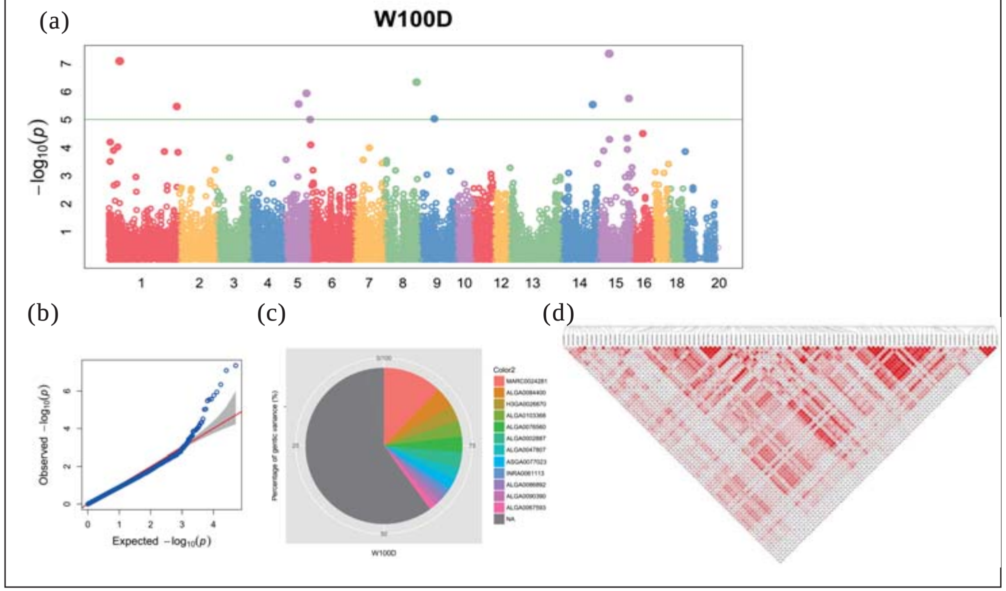
Unique SNPs

GWAS Result Summary

- Total number of processed traits: 41

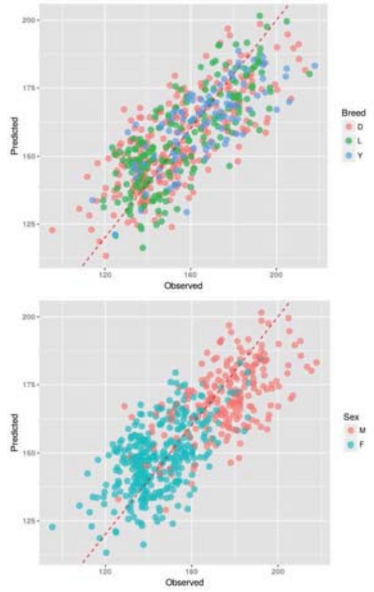
Item	# Traits	# Good traits	Example results	# Bad traits	Comments
Growth	7	4	 	3	結果符合前期文獻蒐尋與預期，除產品化及有實施場域外，將進行v.2改良 (Results as expected, working on v.2 model)
Re-productive	14	2	 	12	繁殖性狀具有挑戰性，已與種豬場進行其他延伸計畫。 (Difficult traits, extended project is under discussion with breeding farms)
Meat Quality	19	5	 	14	數個性狀有結果，部分顯示增加樣本並減少性狀測量誤差可有更佳結果。 (Low statistical power due to poor sample, working on setting collecting standard)
Genetic Defect	1	1		0	結果良好，已於其他場施行 (Result looks promising, field trail)

[Example] Days to 100kg



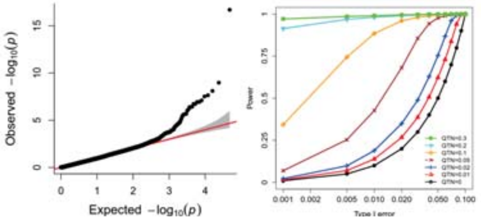
[Example] Marker Explanation

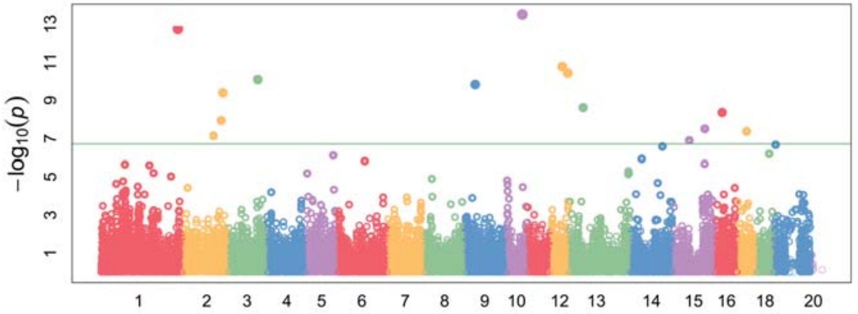
Chrom	Position (MB)	Weight (kg)	SNP	IND_1	IND_2
1	53.6	-4.6	A/G	AA(0)	AA(0)
3	49.6	4.5	G/A	GG(2)	GG(2)
8	42.7	-3.9	A/C	CC(2)	AA(0)
9	29.7	4.1	G/A	GG(2)	GG(2)
13	5.4	-4.8	C/T	CC(0)	CC(0)
14	27.2	-5.0	C/A	CC(2)	CC(2)
15	24.5	-4.1	T/C	CC(0)	CT(1)
15	112.1	1.6	T/C	TT(2)	CT(1)
15	134.3	2.3	G/A	AG(1)	AA(0)
15	136.7	3.3	T/C	TT(2)	TT(2)
16	35.3	-4.6	G/T	GG(0)	GG(0)
17	44.2	-3.7	T/G	GT(1)	TT(2)
other				198.4	153.3
Predicted				182.6	146.1
Observed				189.3	146.2



[Example] Genetic Defect

- Breakthrough
  - 公開研究中提出的數個位點在台灣族群上均無影響或表現
Public results do not fit in Taiwan's population
  - 關聯性分析成功找到約10個新的顯著 SNPs
Successfully identified 10 novel SNPs





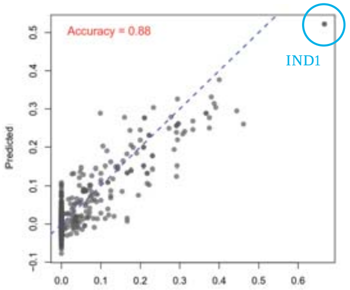
[Example] Marker Explanation

- The importance of top 10 SNPs:

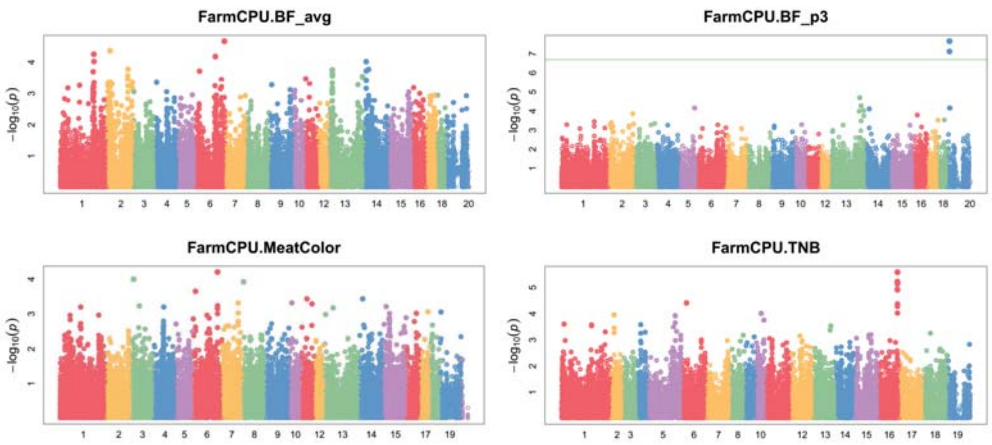
Chrom	Position (MB)	Weight (%)	SNP	IND1	IND2
1	297.2	-1.7	C/T	CT (1)	TT (0)
2	40.6	+1.5	C/T	TT (0)	CC (2)
2	57.5	-1.6	T/C	CC (0)	CC (0)
3	29.2	-5.7	G/T	GG (2)	TT (0)
4	40.5	+1.5	A/G	GG (0)	AA (2)
5	23.5	+2.5	C/T	CC (2)	CC (2)
6	156.1	-1.3	G/A	AA (0)	AA (0)
12	123.6	-2.3	C/A	CA (1)	AA (0)
17	34.0	-3.1	T/C	CC (0)	CC (0)
19	5.9	-5.7	G/A	GA (1)	AA (0)
other				68%	68%
			預測結果	52%	79%

IND2為將 IND1 Top 10 SNPs 中的6個位點改為更”嚴重”的基因型，最終預測每胎發生缺陷的後代為79%，較 IND1 增加 27%

Manually flipped 6 of 10 top SNPs of IND1 to “more severe” genotypes, and each litter of the resulting individual, IND2, is expected to have 79 % defects, 27% more than IND1.

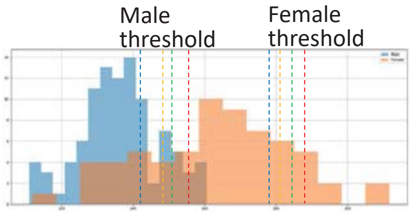


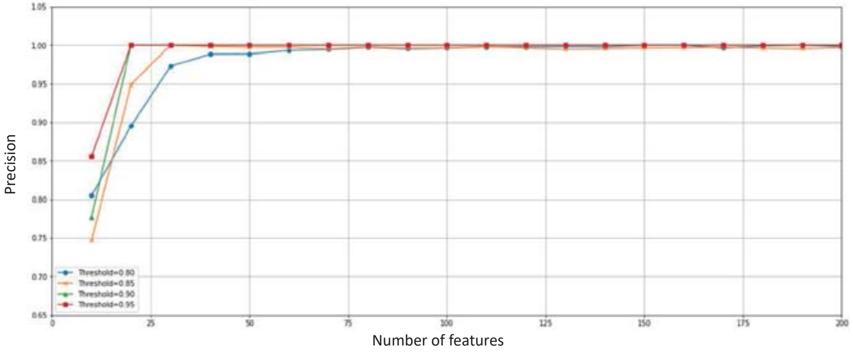
Sometimes the magic doesn't work



Another Success with Machine Learning

- 因為生長趨勢的差異，公、母豬分別有兩組閥值
  Two sets of thresholds for ♀ & ♂, respectively due to difference in growth rate
    - \> threshold → bad
    - ≤ threshold → good
- 機器學習預測為好/壞豬的準確度
  Precision on predicting good or bad pig.



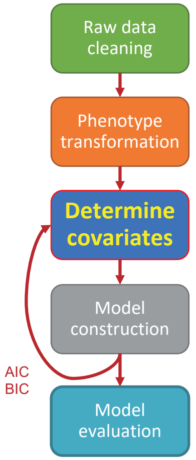


Some Experiences We Learned

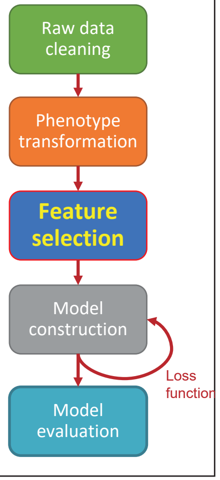
- 生長性狀 (Growth traits)
    - 性別效應於生長效應很強
      Growth traits clearly show gender-specific differences.
    - 背脂於第5肋(p1) 有較顯著關聯位點，然而最後肋(p2)、最後腰椎(p3)及三點平均則因潛在訊號受稀釋而無結果
      Back fat only mapped for p1 (5th rib) but not for p2 (the last rib), p3 (the last lumber), as well as 3-point average. Aggregating phenotypes may reduce statistical power in discovering associations.
- 繁殖性狀 (Reproductive)
    - 各母豬所有胎次平均無顯著結果 (同背脂三點平均)
      Averaging litter size did not work well (as indicated in back fat)
    - 低遺傳率複雜性狀可考慮加強變因控制、二階段建模(GBLUP=先取得性狀BLUP, 再進行GWAS)、貝氏模型或機器學習
      Consider following strategies for low heritability traits: more complex covariates, two-stage modeling (BLUP for phenotypes then GWAS), Bayesian model or machine learning.

Cope with The Shortage of Conventional GWAS

- GWAS 已有大量的研究結果，但基因組的異質性 (heterogeneity)、表型變異和線性模型前提等諸多因素造成探測多基因交互作用上有諸多困難。而機器學習可以用來解決這些局限性。
  Genomes heterogeneity, pleiotropy, penetrance, model assumptions and etc. make it hard to explore gene interactions; however, machine learning predicts using combinations of markers, which is a potential solution.



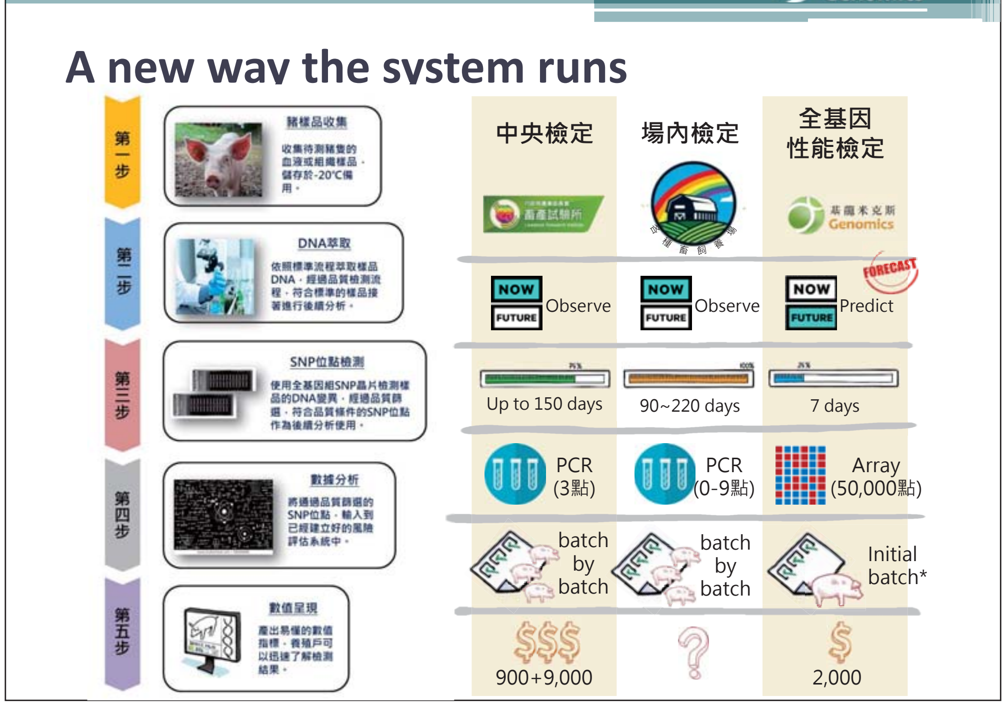
關聯性分析 (GWAS)		機器學習 (ML)
數萬個較佳 Tens of thousands	**位點** **SNPs**	可達數千萬個 Tens of millions
數十分鐘 Tens of minutes	**速度** **Speed**	數分鐘 Few minutes
一般電腦 Personal computer	**硬體** **Hardware**	規格較高 Often requires GPU
專注解釋單一位點效應 On single marker	**效應** **Effects**	位點間高維度的組合 Combinations of “features”
較多(個體,系譜,族群,環境...) More (indiv., popl, env., pedigree, and etcs)	**控制變因** **Factors**	較少(性別) Less (like only gender for growth trait)
可 Mostly yes.	**缺值** **Missing**	有些演算法不能 No for some methods
較仰賴人工視結果調整 Labor dependent	**調整** **Optimize**	演算法中有最佳化條件 Part of algorithm
模型意義直觀 Intuitive model meaning	**優勢** **Pros**	辨識非線性複雜關係 Identify Complex gene-gene interactions



The Wheel Starts Rotating

- 雖候選位點仍需進行註解及基因確認等後續工作，且分析的方式尚有調整進步的空間，然而全基因組分析上的成效，已吸引種畜場，將計畫結果應用於輔助該場核心族群留種，並計畫於 12月底前簽訂合作意向書。

  Although fine tuning of final models, annotating significant markers and other down-strand analyses are still undergoing, due to successful results demonstrated above, a breeding farm has shown interests in introducing this powerful tool to assist annual selection on its core population and accelerate breeding precision. MOU will be signed by the end of the year.



A new way the system runs

第一步 **豬樣品收集**
收集待測豬隻的血液或組織樣品，儲存於-20°C備用。

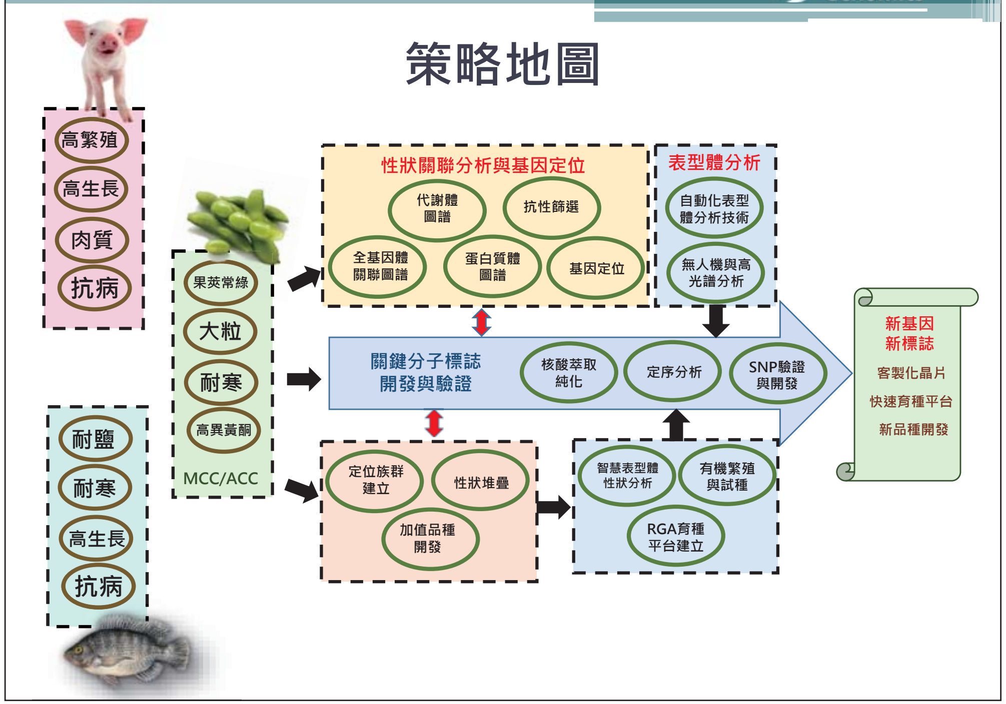
第二步 **DNA萃取**
依照標準流程萃取樣品DNA，經過品質檢測流程，符合標準的樣品接著進行後續分析。

第三步 **SNP位點檢測**
使用全基因組SNP晶片檢測樣品的DNA變異，經過品質篩選，符合品質條件的SNP位點作為後續分析使用。

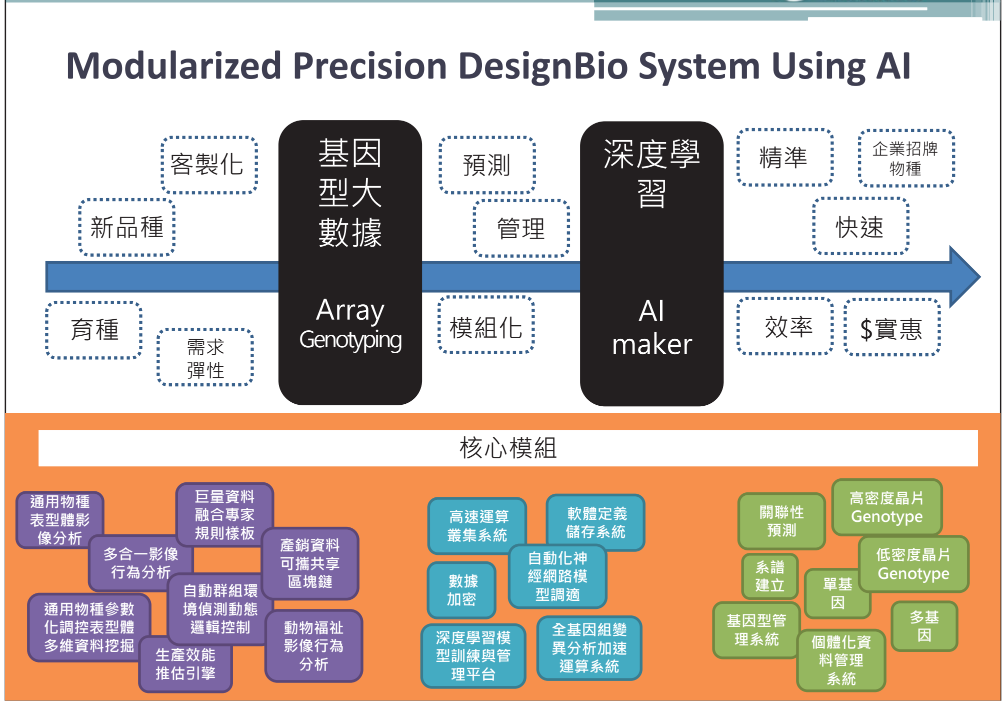
第四步 **數據分析**
將通過品質篩選的SNP位點，輸入到已經建立好的風險評估系統中。

第五步 **數值呈現**
產出易懂的數值指標，養殖戶可以迅速了解檢測結果。

中央檢定	場內檢定	全基因性能檢定
NOW FUTURE Observe	NOW FUTURE Observe	NOW FUTURE Predict
Up to 150 days	90~220 days	7 days
PCR (3點)	PCR (0-9點)	Array (50,000點)
batch by batch	batch by batch	Initial batch*
$$$ 900+9,000	?	$ 2,000



策略地圖



Modularized Precision DesignBio System Using AI

Thanks for your attention

Q&A