动态网络标志物

生物医学大数据的“疾病前兆”

-- “未病”的定量诊断 --

陈洛南

中国科学院上海生命科学研究院

一旦发病，很难治愈

《黄帝内经》公元前221年

上医治未病，中医治欲病，下医治已病

↓

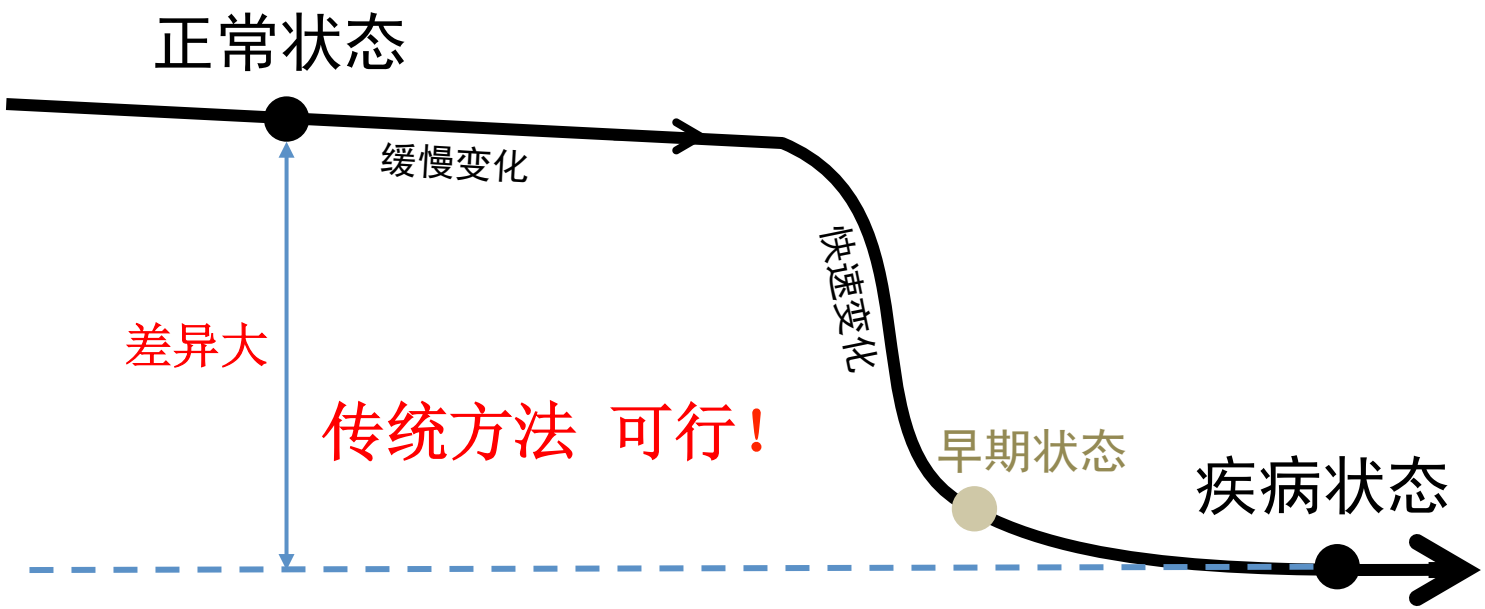
临界点和疾病前兆
动态网络标志物

复杂疾病的临界理论

－ 系统生物学的大数据理论 －

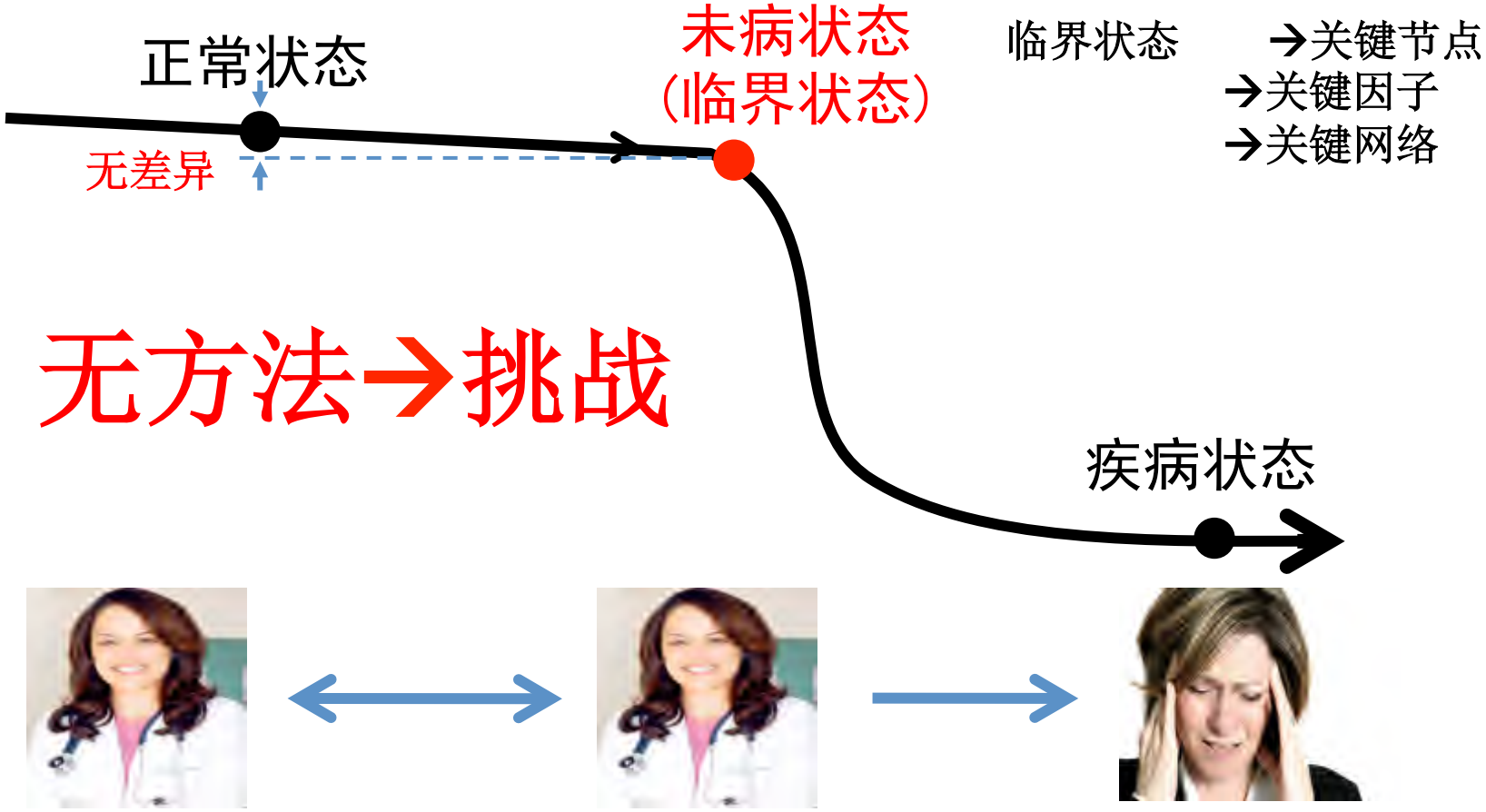
生物过程不是线性和静态，而是非线性和动态

（如，疾病过程，细胞分化）



疾病发生发展过程

科学问题： 临界状态、定量评估、疾病预警？

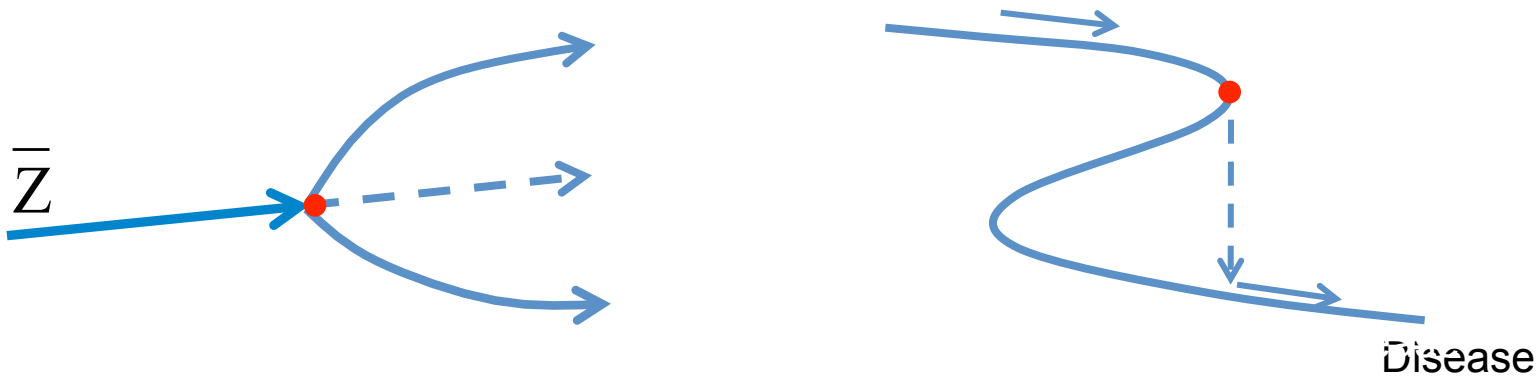


从Case/Control静态比较到动态过程研究

Bifurcation

$$\dot{Z}(t) = f(Z(t), P)$$

- $Z(t)=(z_1(t),\ldots,z_n(t))$ are state, P are parameters.
- The stability of one equilibrium will change from stable to unstable at bifurcation point P_c.



Chen, et al., Scientific Reports, 2, 342, 2012

Catastrophic bifurcation

System

$$\dot{Z}(t) = f(Z(t);P)$$

Driving Factors (unknown)

Parameters *P*

Slowly changing factors, e.g. genetic (SNP,CNV), epigenetic (methylation, acetylation) factors

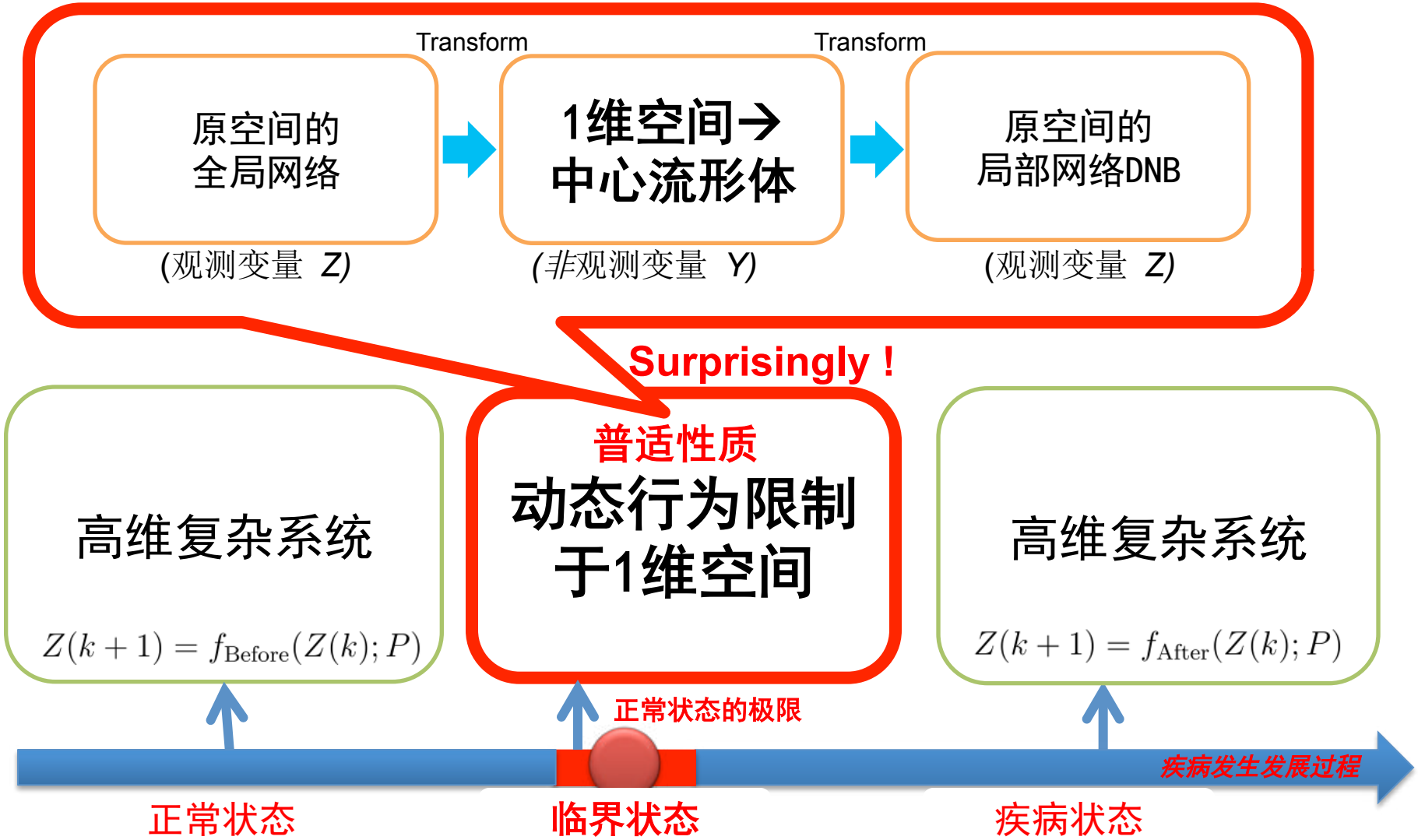
Responses (observable)

State Variables *Z*

e.g. gene expression, protein expression

The progression of a disease is considered as the evolution of a nonlinear dynamical system.

临界期的特殊动力学行为

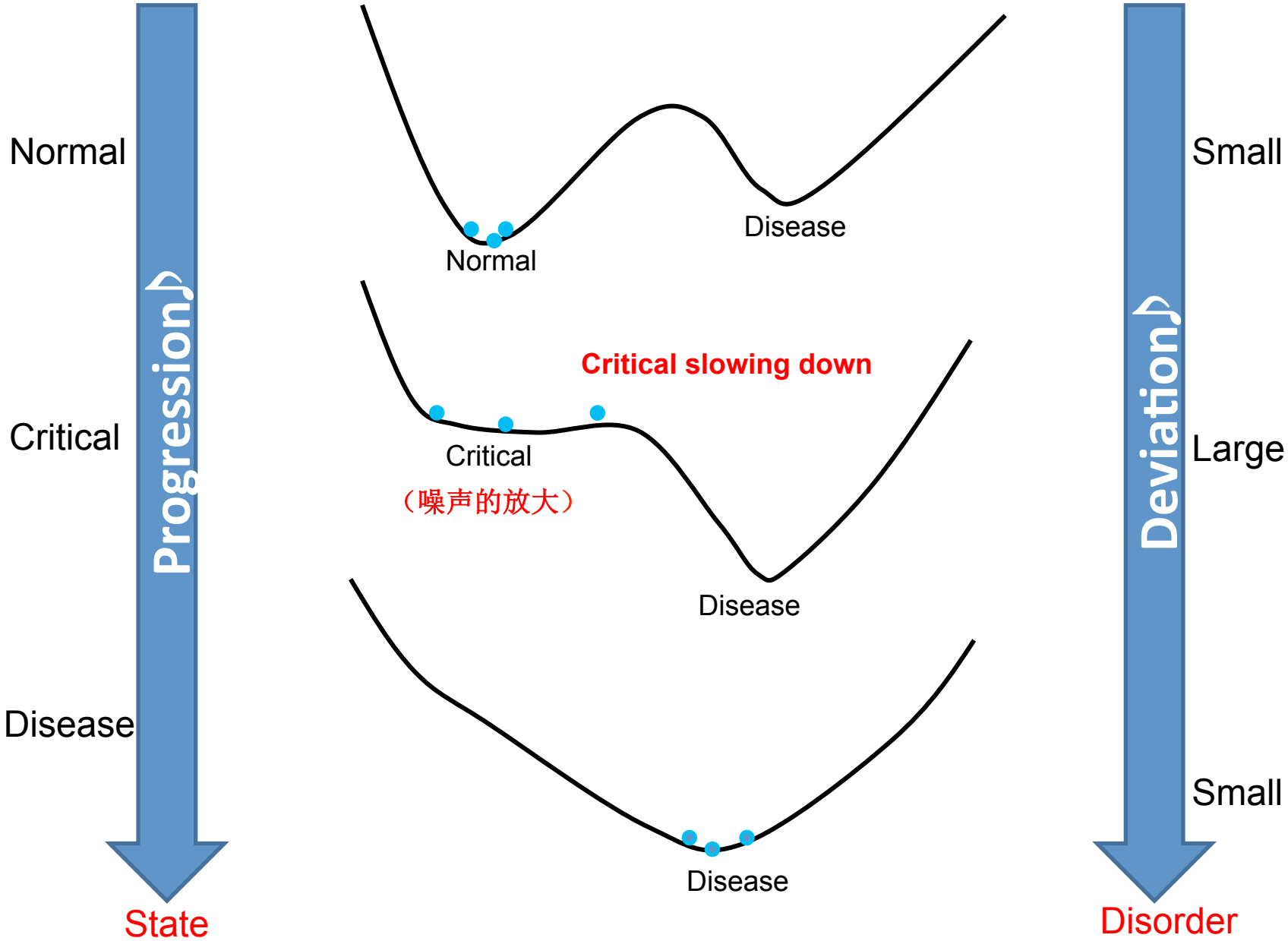


不依赖于个体差异的普适性质

Conceptual illustration from physics viewpoint

Landscape

Potential Energy Function

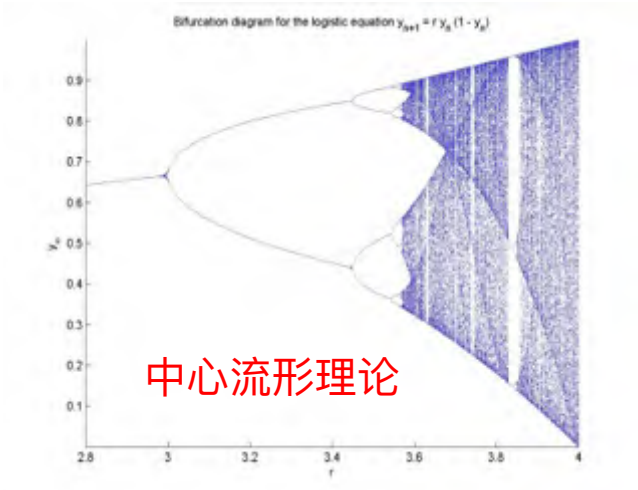


动力学理论模型的建立及分析

普适分叉分析(co-dim one : $\dot{x} = f(x\ ;p)$ at $\bar{x}$)

- Eigenvalue λ1 →0
  （鞍点分叉）
- Eigenvalue || λ1 ||→0
  (Hopf分叉)

非线性动力系统的普适分叉



不同的动力系统，细节上可能会有很大迥异，
到达临界点之前千差万别，
但是它们在临界点附近却有相同的普适分叉

小样本提取临界特征的理论依据

个体差异

L.Chen et al., Scientific Reports, 2012

疾病临界检测的理论(DNB)

由生物医学大数据

数学证明

1. 在子网络分子中

相关（网络上） $\mathrm{pcc}(x_i, x_j) \rightarrow \pm 1$

方差（动态上） $\mathrm{sd}(x_i)\ \mathrm{and}\ \mathrm{sd}(x_j) \rightarrow \infty$

2. 在子网络与其它分子之间

相关（网络上） $\mathrm{pcc}(x_i, x_j) \rightarrow zero$

方差（动态上） $\mathrm{sd}(x_i) \rightarrow \infty\ \mathrm{but}\ \mathrm{sd}(x_j) = bounded$

3. 在其它分子之间

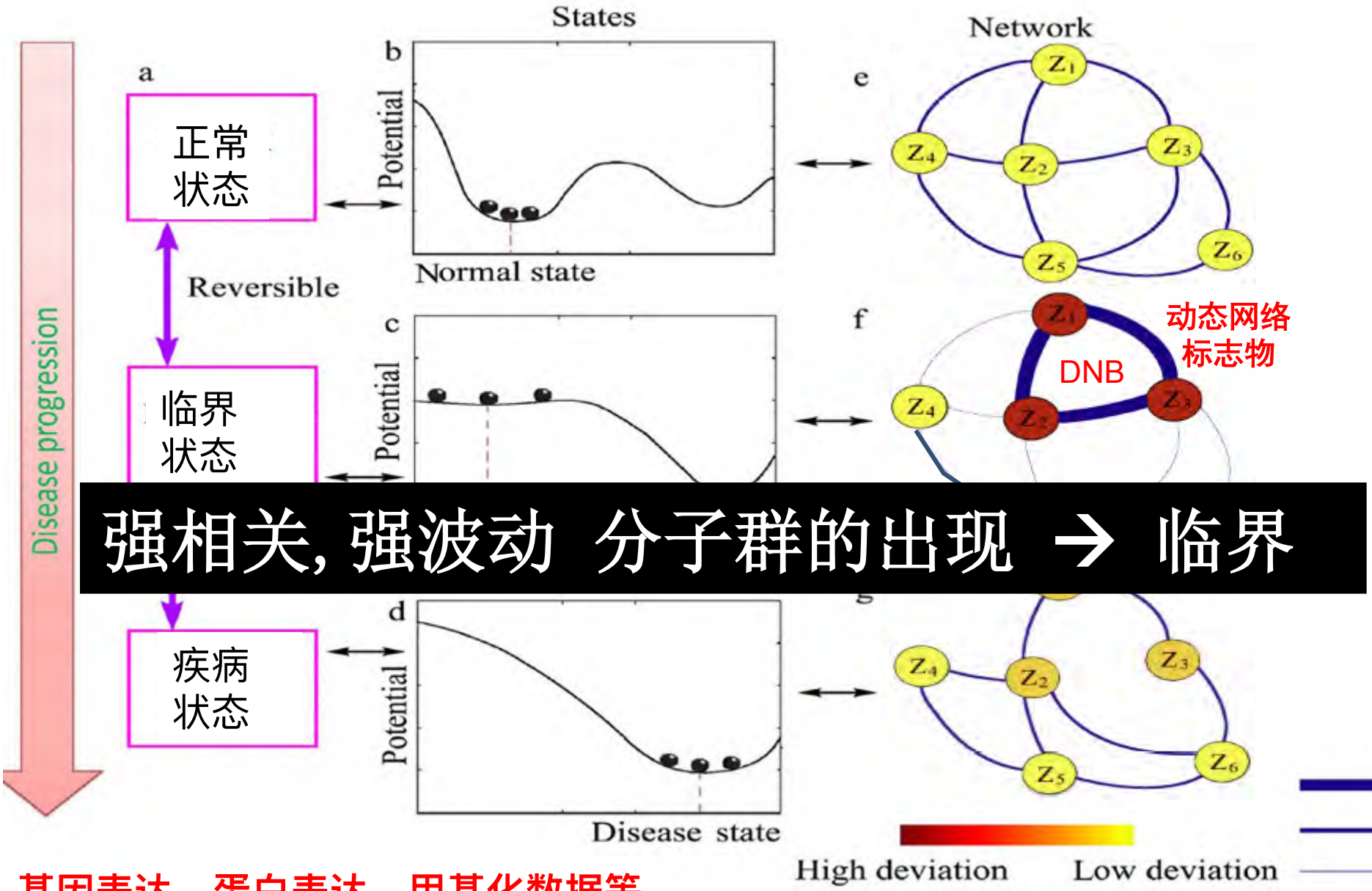
相关（网络上） $|\mathrm{pcc}(x_i, x_j)| \rightarrow a \quad (0 < a < 1)$

方差（动态上） $\mathrm{sd}(x_i)\ \mathrm{and}\ \mathrm{sd}(x_j) = bounded$

非线性动力学理论 → 大数据统计学方法 → 动态网络标志物

动态网络标志物及疾病前兆预测

Dynamical Network Biomarker (DNB)



强相关, 强波动 分子群的出现 → 临界

基因表达，蛋白表达，甲基化数据等

疾病诊断和预测要求不同的方法和信息

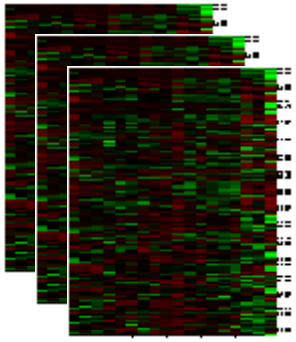
- “疾病”诊断 → 分子表达　静态差异
- “疾病”预测 → 分子表达　动态差异

临床应用

一个人采取三个样本以上的组学数据

二阶统计量

临界状态预测



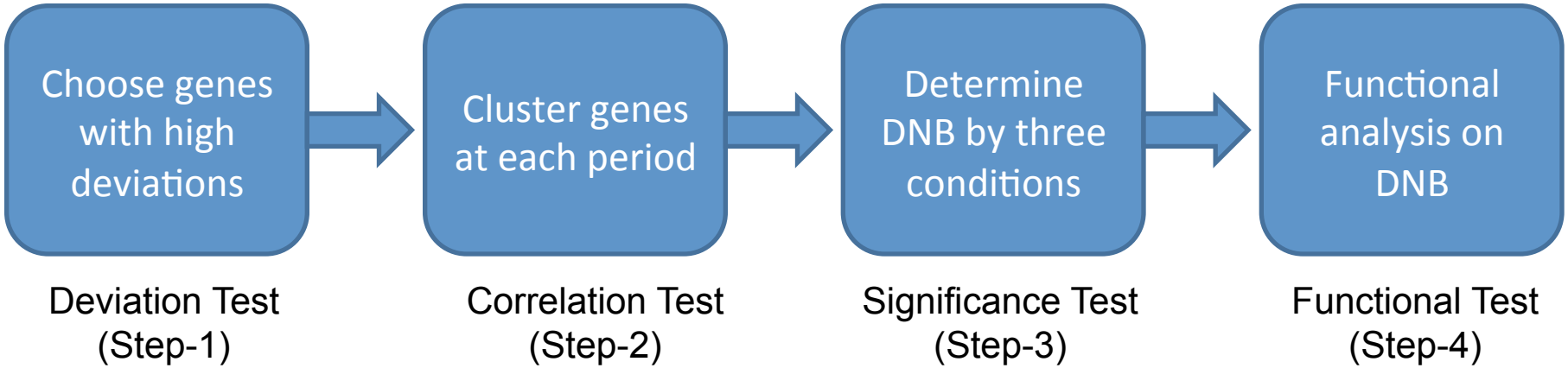
(如，组学数据)

(Big Data)

五个样本以上的组学数据→统计显著性结果

动态网络标志物和关键节点的方法

每个期间至少5个样本，或单样本高维大数据



满足3个条件的优化问题

$$I =: \frac{SD_d \cdot PCC_d}{PCC_o}$$

临界理论验证

- 肺损伤 Mouse
- 肝癌 Human
- 糖尿病 Rat
- 流感 Human

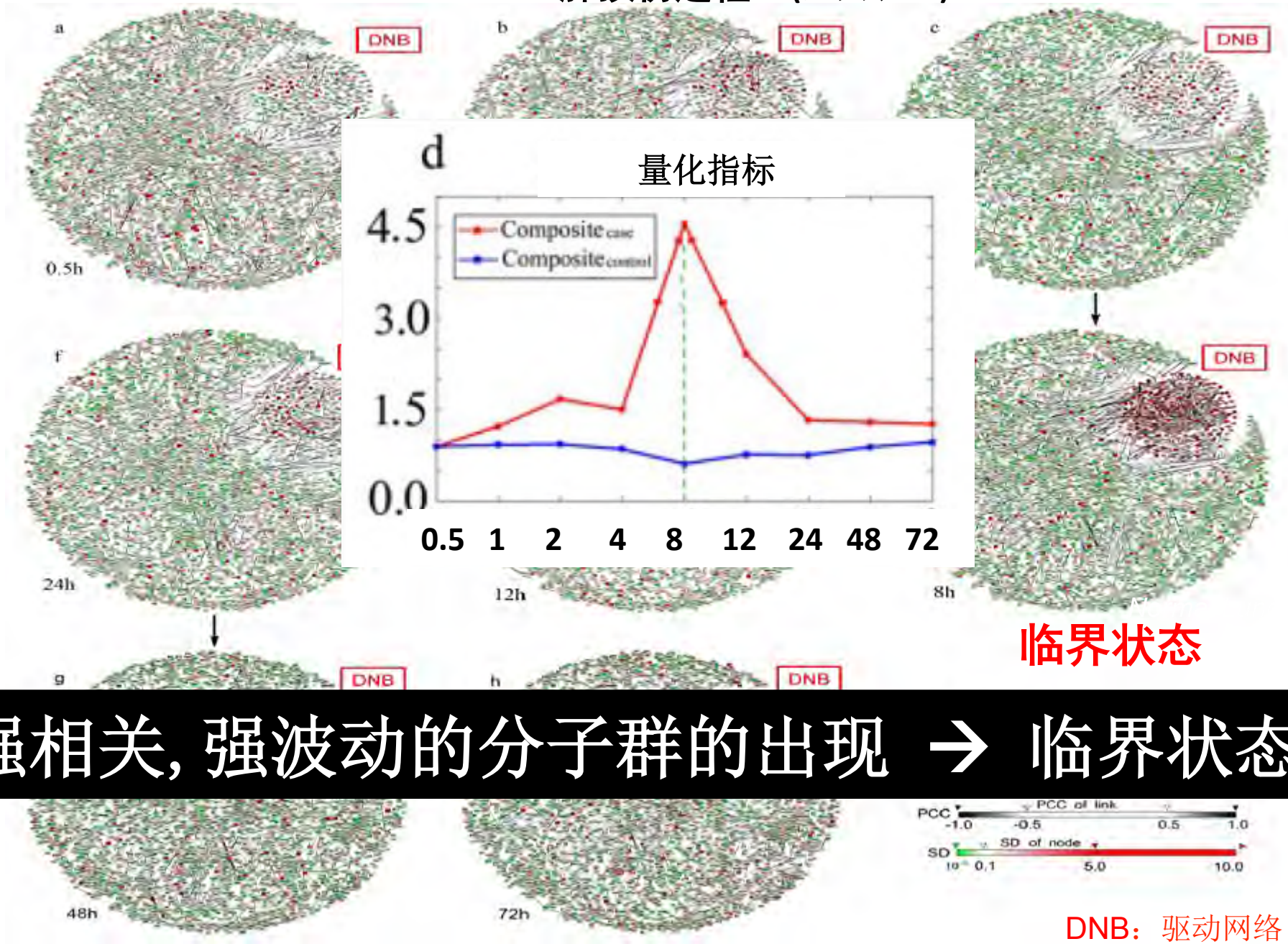
Chen, et al., Scientific Reports, 2, 342, 2012
Liu, et al., Bioinformatics, 2014; Salina, et al., JMCB, 2014
Liu, et al., Medicinal Research Review, 2013; Yu, et al., Bioinformatics, 2013

Chen, et al., Scientific Reports, 2, 342, 2012

The experimental data	Descriptions
Genomic data about the lung injury with Carbonyl chloride inhalation exposure[12]	
Sampling points	9 sampling points
	0, 0.5, 1, 4, 8, 12, 24, 48, 72 (hours)
The Number of observing objects	22691 RNAs
Groups	control group and case group (5 mice)

Dynamical changes of whole mouse network (3452 genes and 9238 links)

肺损伤过程 (基因表达)



临界状态

强相关, 强波动的分子群的出现 → 临界状态

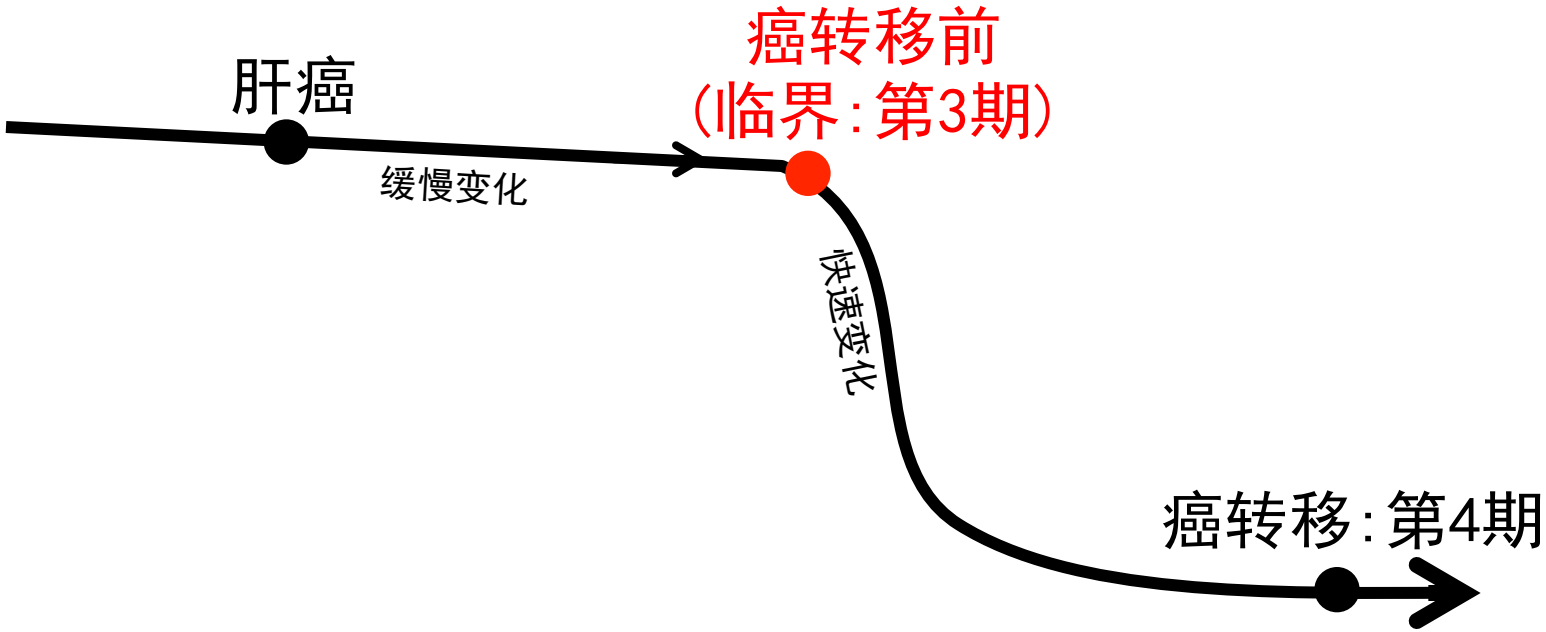
DNB：驱动网络

临界理论验证

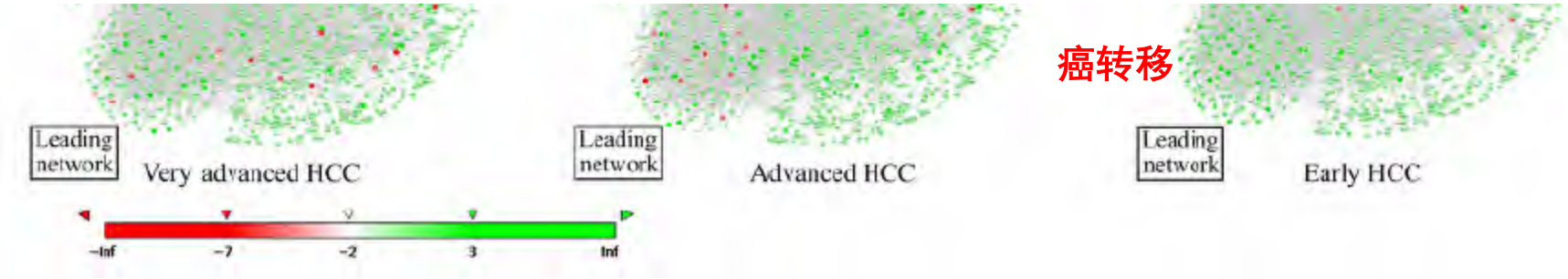
- 肺损伤 Mouse
- **肝癌** Human
- 糖尿病 Rat

Chen, et al., Scientific Reports, 2, 342, 2012
Liu, et al., Bioinformatics, 2014
Liu, et al., Medicinal Research Review, 2013

Dynamical changes in network (2291 genes) during disease progression of HCC（肝癌）

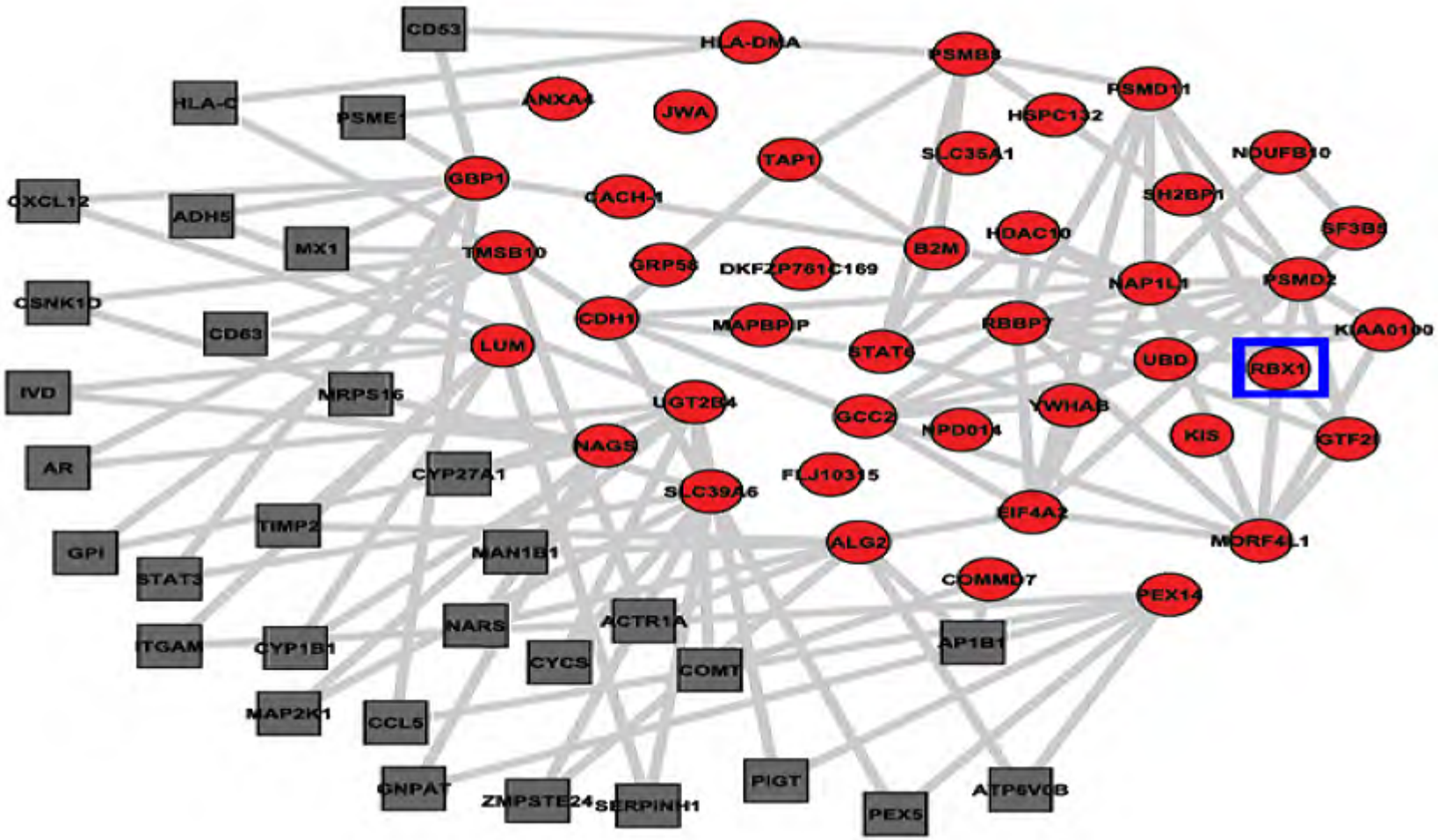


不仅用于疾病的发生，而且疾病的发展



Liu, et al., Medicinal Research Review, 2013

Human Liver Cancer
(DNB: 动态网络标志物)



9个癌症致病基因，5个肝癌致病基因

临界理论验证

- Lung injury — Mouse
- Liver Cancer — Human
- 糖尿病 — Rat

Chen, et al., Scientific Reports, 2, 342, 2012
Liu, et al., Bioinformatics, 2014
Liu, et al., Medicinal Research Review, 2013

Diabetes Rats (microarray data)

4wk 8wk 12wk 16wk 20wk

胰岛素抵抗
Insulin Resistance

Beta 细胞损伤
Beta Cell Failure

5 GK 5 GK 5 GK 5 GK 5 GK

糖尿病大鼠

5 Wistar 5 Wistar 5 Wistar 5 Wistar 5 Wistar

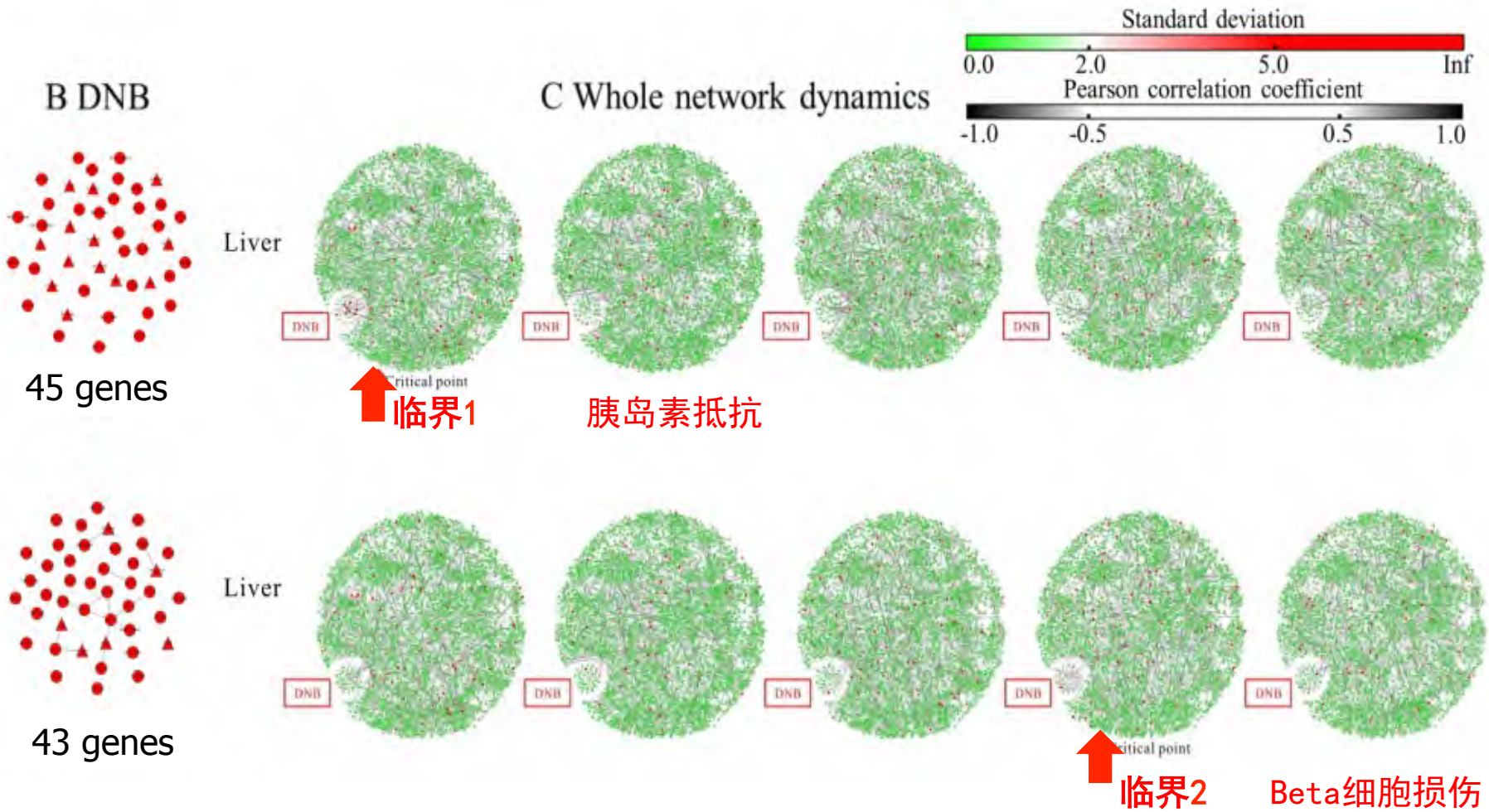
正常大鼠

Tissues		Genes
Liver		10729
Muscle		10729
Adipose		10729

3 tissues，5 stages ， 25 GK，25 Wister
（150 samples）

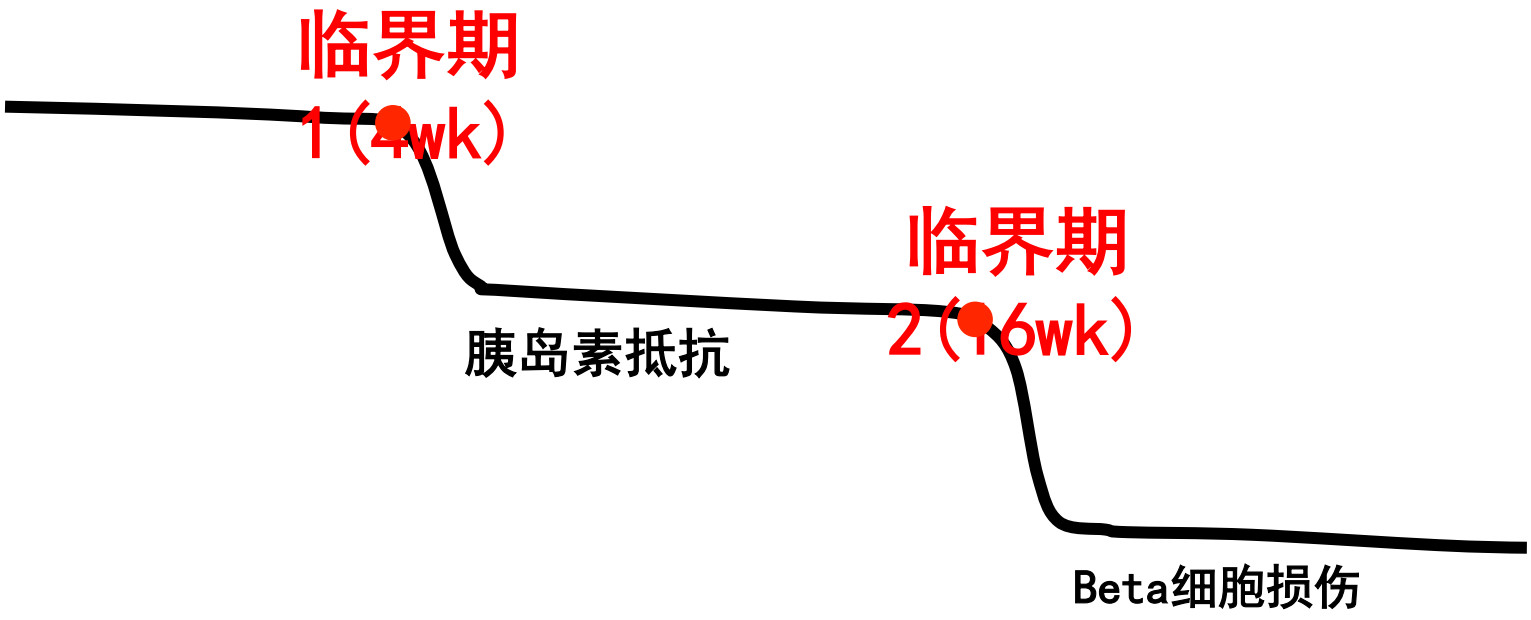
Liu, et al., Bioinformatics, 2014

DNB of GK rats （两个临界状态）



疾病的发生发展

糖尿病的发生发展的2个临界期



DNB和DEG的关联

(DEG: 差异表达基因)

动态网络标志物　　差异表达基因

Tissue	DNB	DNB		DEG		Overlap	P-value	Overlap
		gene	pathway	gene	pathway	gene		pathway
Liver	L1	45	81	2807	254	3	9.99E-01	81
	L4	43	35	1930	250	7	5.20E-01	34
Muscle	M1	72	92	3651	252	3	1.00E+00	92
	M4	58	75	1399	245	5	7.85E-01	75
Adipose	A2	146	147	1444	248	8	9.98E-01	147

DNB和DEG的 基因交集少，但通路交集大 ！

DEG is the result of the disease, but DNB is related to its cause

DNB和DEG的关联

同一通路中

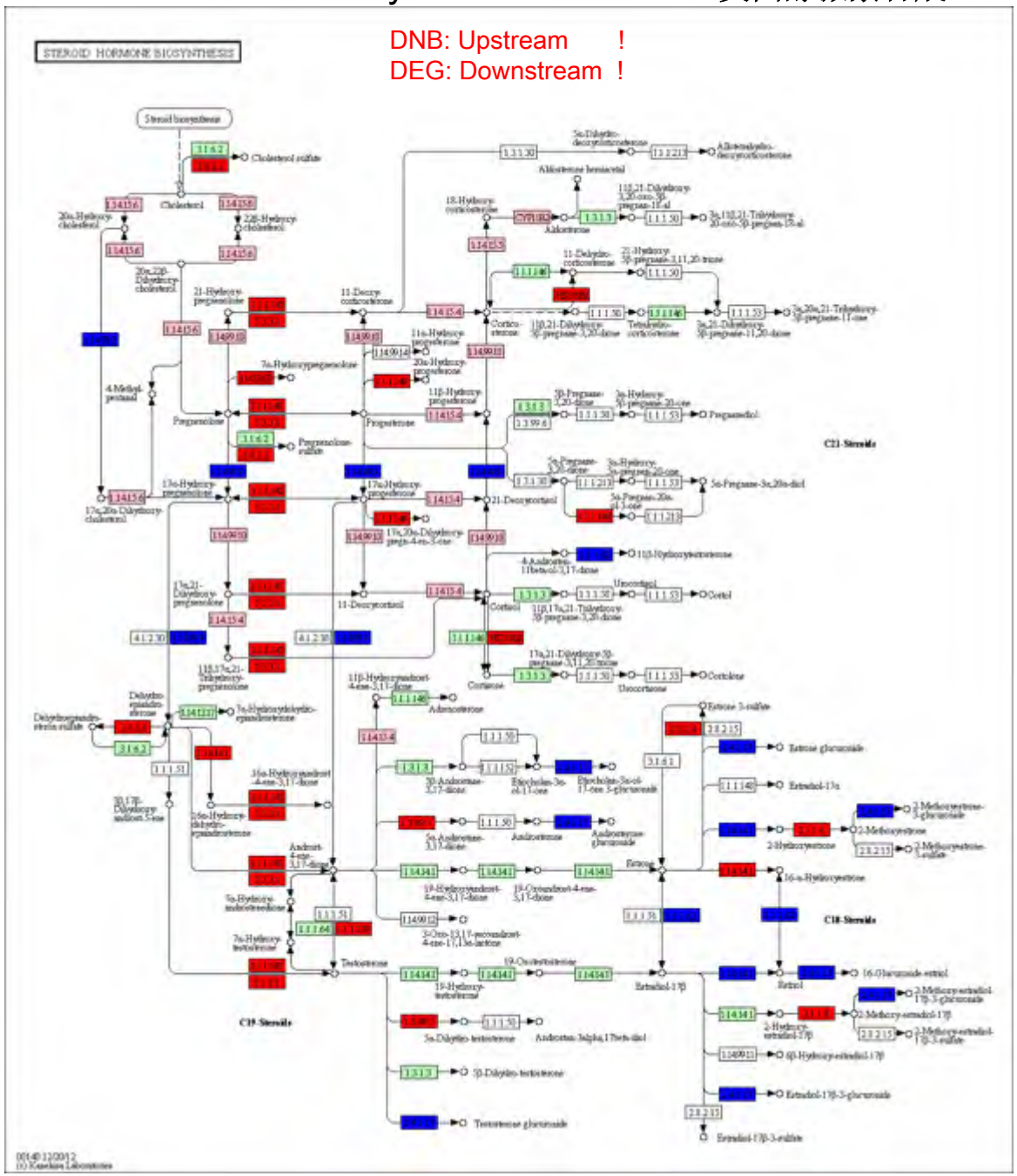
DNB在上游→关键基因
DEG在下游→靶点基因

1. Upstream
2. As messengers or receptors
3. As a part of the combined complex

DNB
over-expressed genes
under-expressed genes

White and Green: background

Steroid hormone biosynthesis – *L1* 类固醇激素合成



17人的流感预测 (H3N2/Wisconsin)

N Non-symptom (normal state or pre-disease state)

S Symptom (disease state)

17 subjects in live influenza experiment

S1	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S5	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S6	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S7	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S8	N	N	N	N	N	N	N	N	N	N	S	S	S	S	S	S
S10	N	N	N	N	N	N	N	N	N	N	N	S	S	S	S	S
S12	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S	S
S13	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S
S15	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S
S2	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S3	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S4	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S9	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S11	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S14	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S16	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S17	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Hour	-24	0	5	12	21	29	36	45	53	60	69	77	84	93	101	108

Symptomatic subjects

Asymptomatic subjects

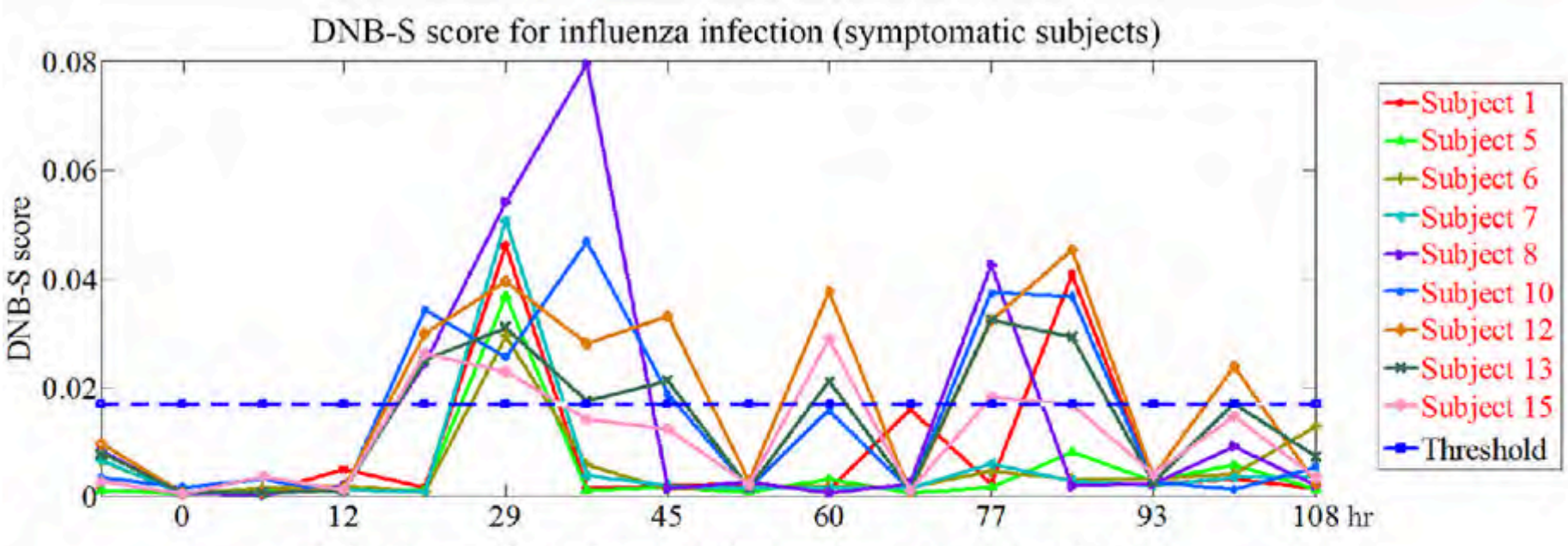
Time

Yu, et al., Bioinformatics, 2013

Clinical data

Class	Unique ID	Gender	Age (years)
Asx	flu011	Female	25
	flu002	Male	28
	flu003	Male	24
	flu004	Female	23
	flu009	Male	24
	flu014	Female	22
	flu016	Male	41
	flu017	Male	33
Sx	flu013	Female	29
	flu015	Male	26
	flu001	Female	29
	flu005	Female	25
	flu006	Female	28
	flu007	Male	35
	flu008	Male	25
	flu010	Female	22
	flu012	Male	27

DNB的预测结果


DNB-S score for influenza infection (symptomatic subjects)
DNB-S score
0.08
0.06
0.04
0.02
0
0
12
29
45
60
77
93
108 hr
Subject 1
Subject 5
Subject 6
Subject 7
Subject 8
Subject 10
Subject 12
Subject 13
Subject 15
Threshold

17人的流感预测结果 (H3N2/Wisconsin)

N — Early diagnosis before disease (pre-disease state)

N — Non-symptom (normal state or pre-disease state)

S — Symptom (disease state)

17 subjects in live influenza experiment

S1	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S5	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S6	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S7	N	N	N	N	N	N	N	S	S	S	S	S	S	S	S	S
S8	N	N	N	N	N	N	N	N	N	N	S	S	S	S	S	S
S10	N	N	N	N	N	N	N	N	N	N	N	S	S	S	S	S
S12	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S	S
S13	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S
S15	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	S
S2	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S3	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S4	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S9	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S11	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S14	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S16	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
S17	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Hour	-24	0	5	12	21	29	36	45	53	60	69	77	84	93	101	108

Symptomatic subjects

Asymptomatic subjects

Time

总结

DNB理论 → 各种临界现象预测

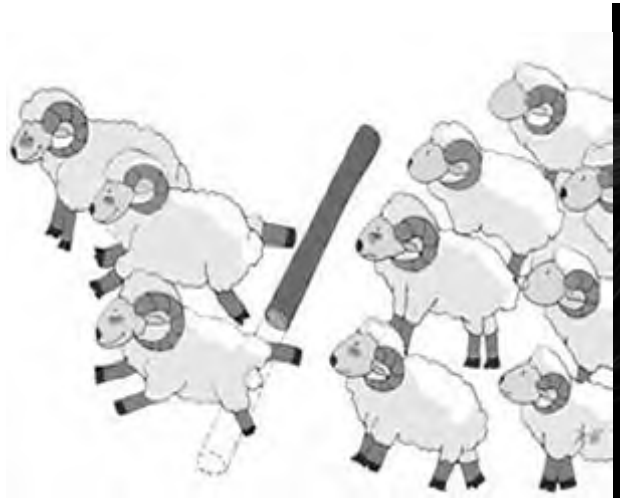
不仅应用于生物，医学，而且生态，金融等领域
Collapse of fish stocks, stock market, ice melting to water, earthquake, revolution

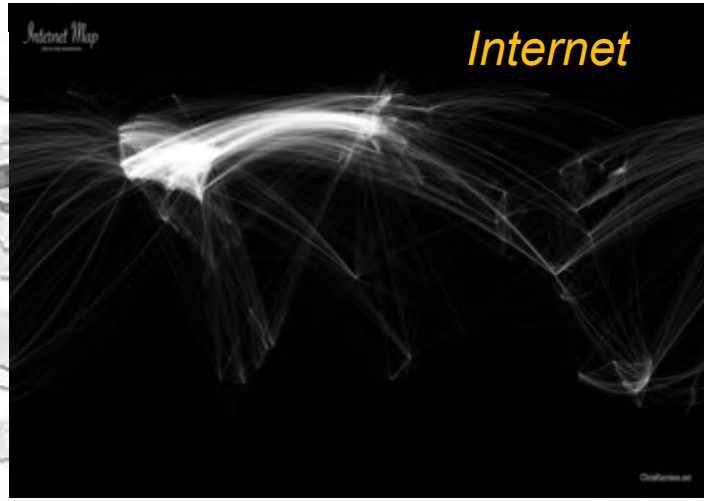
羊群效应
(社会，经济学)

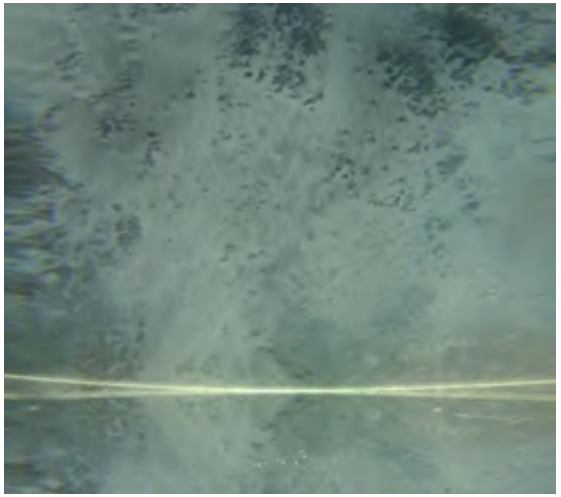
连接的雪崩
(网络科学)

临界松弛
(物理学)

DNB理论的一部分

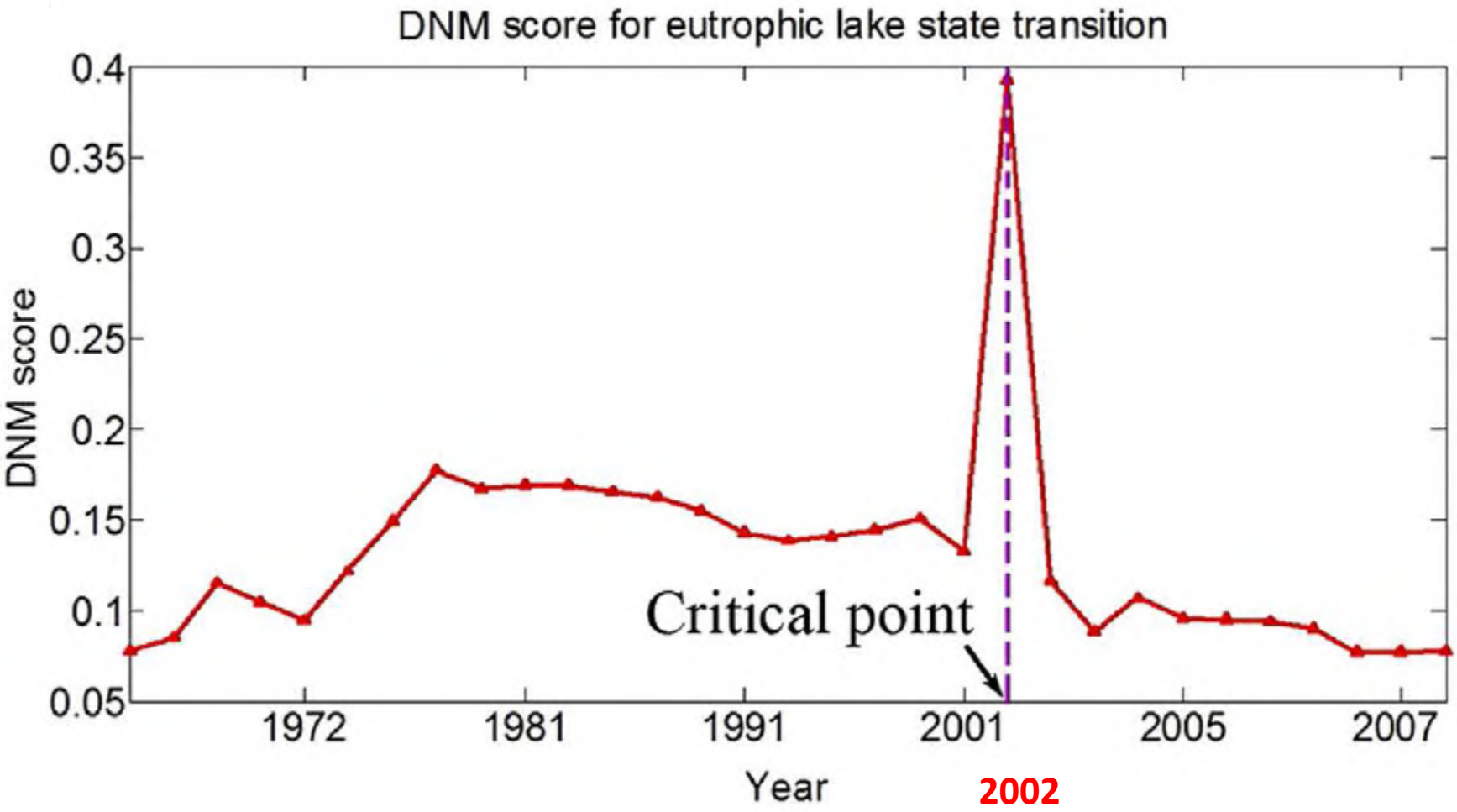






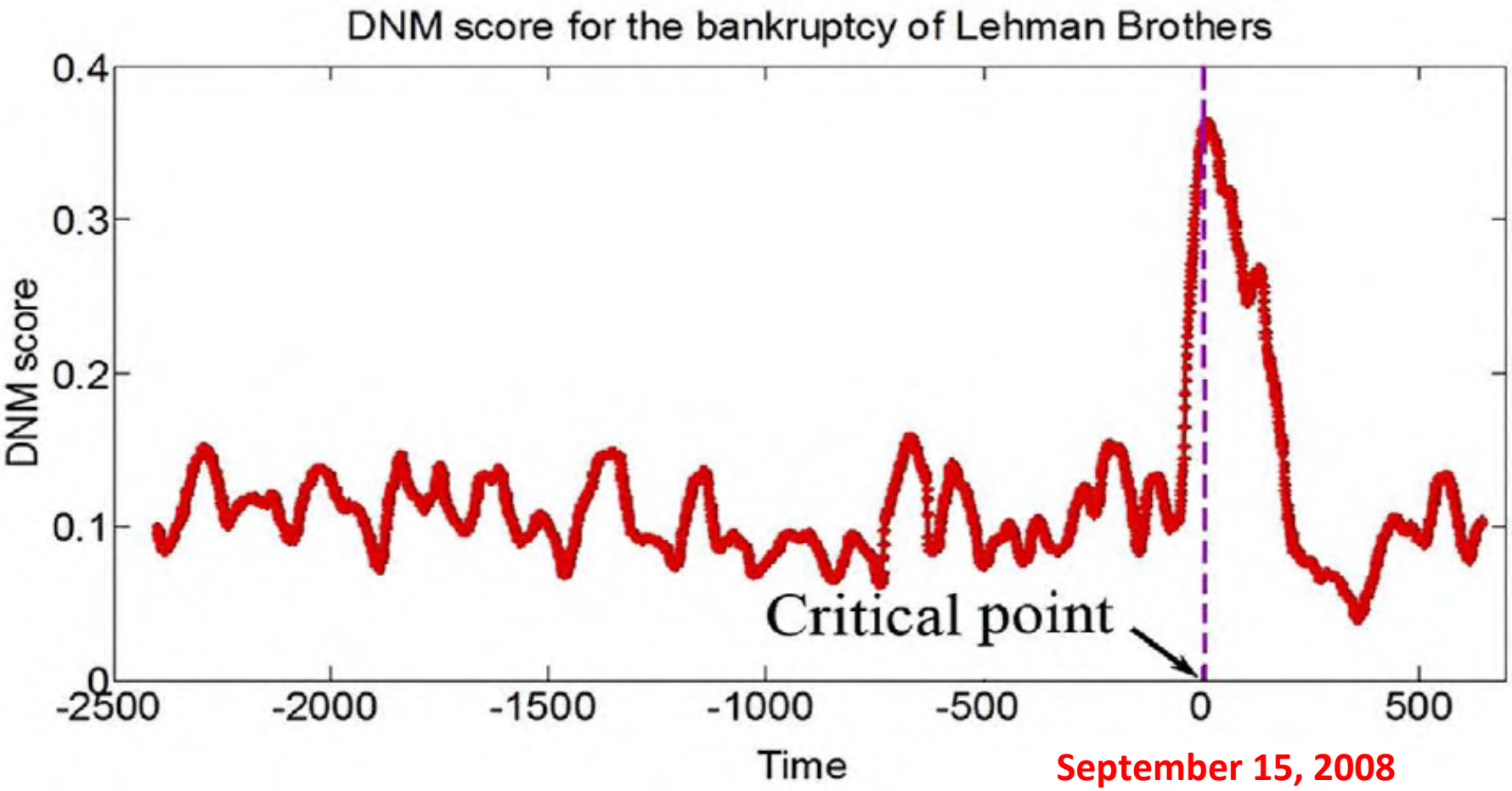
Oya, et al., New Journal of Physics, 2014; Quax, et al., Scientific Reports, 2013; Hefley, et al., Theoretical Ecology, 2013.

云南洱海 (生态环境)

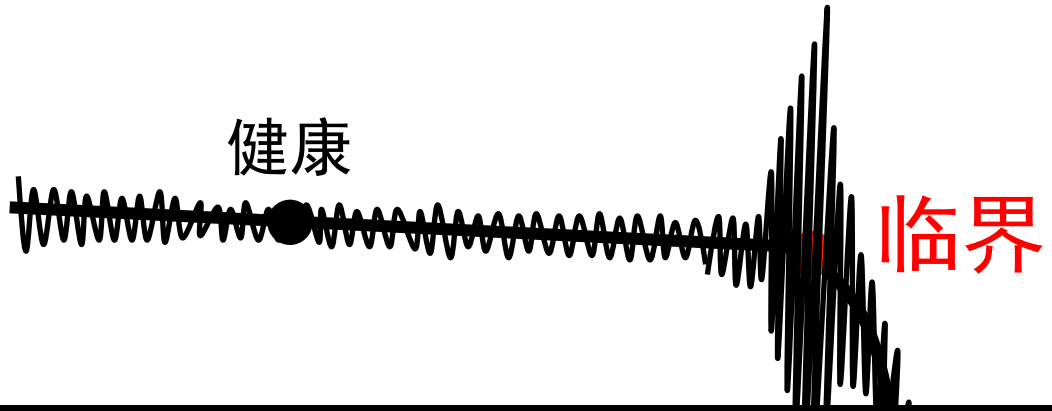


algal states from the oligotrophic state to the eutrophic state

Financial bubbles（金融市场）



临界信号不是来自（静态）统计而是来自动态性质



多变量系统 → DNB的检测

强相关及强波动的分子群

（疾病预测不是诊断的延长，需要全新的系统生物学方法）

From DNA to DNB

疾病临界现象及DNB理论

提出 临界状态 新概念
解决 疾病前兆 预测的缺口

大数据疾病前兆预测原创性理论和方法

↓

个体化医疗，精准医疗

↓

经济，网络，社会科学等的临界现象

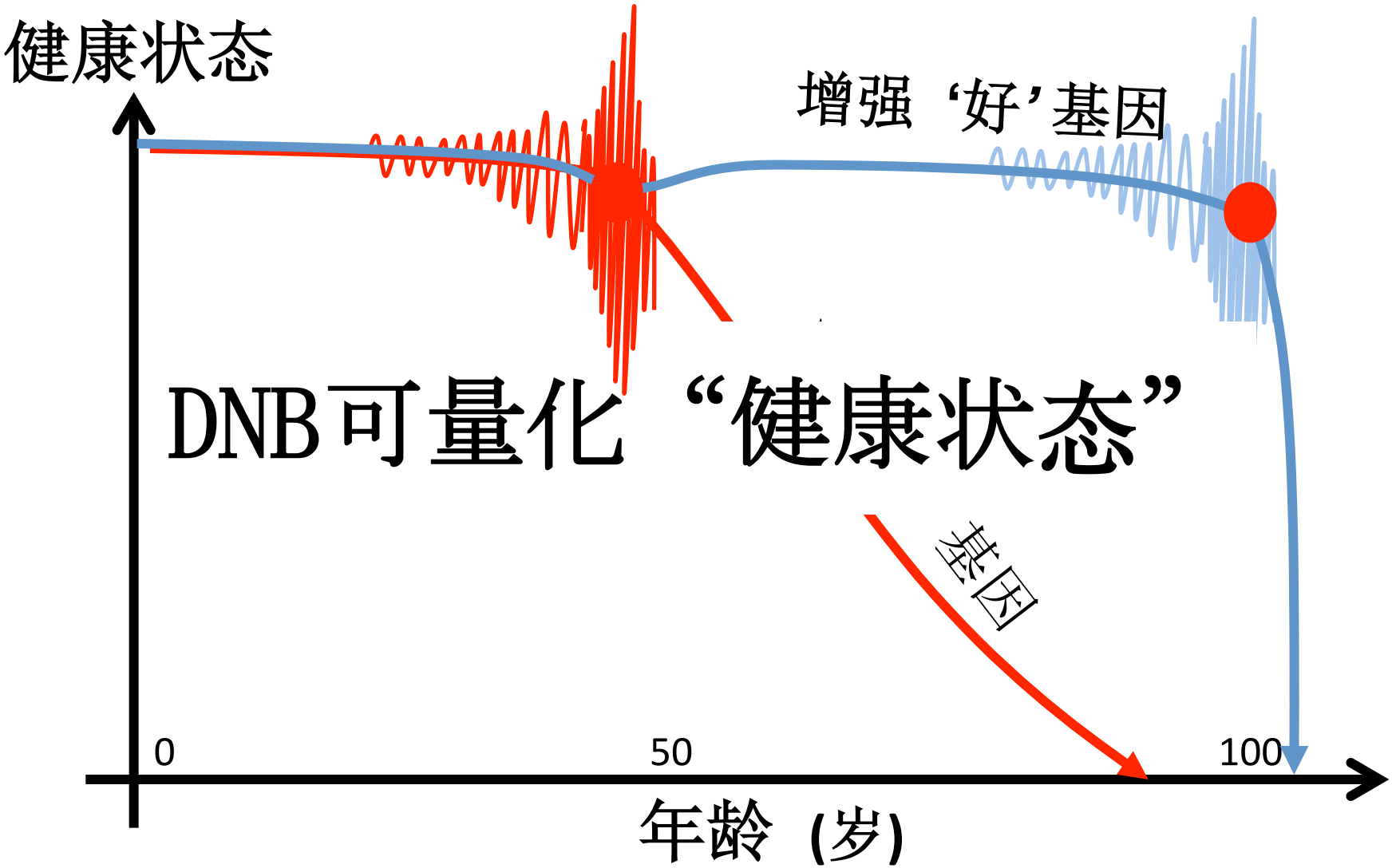
General Big Data → 大样本低维数据

large samples with low dimensions

Biomedical Big Data →小样本高维数据

Small samples with high dimensions

量化“健康状态” → *不是治疗‘坏’基因，而是增强‘好’基因*



“幸福状态” = “Happiness”

“健康状态” = “Wellne

感谢

- 东京大学　　Kazuyuki Aihara等
- 华南理工大学 **刘锐** 等
- 中科院上海生科院　**曾涛，刘小平, 李美仪**等

谢谢！