生物信息学：导论与方法
Bioinformatics: Introduction and Methods

https://www.coursera.org/course/pkubioinfo

生物信息学：导论与方法
Bioinformatics: Introduction and Methods

北京大学生物信息学中心 高歌、魏丽萍
Ge Gao & Liping Wei
Center for Bioinformatics, Peking University

Transcriptome Analysis with noncoding RNAs

北京大学生物信息学中心 高歌

Ge Gao, Ph.D.

Center for Bioinformatics, Peking University
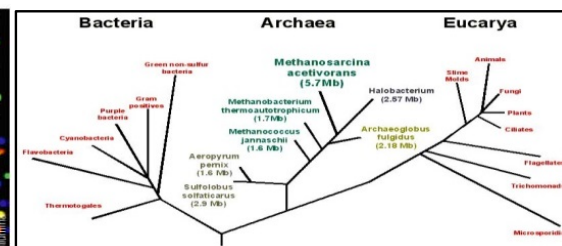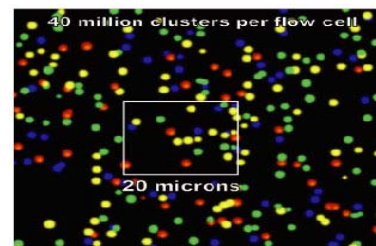
Unit 1:
From Information to Knowledge

北京大学生物信息学中心 高歌
Ge Gao, Ph.D.
Center for Bioinformatics, Peking University

A GENE REGULATORY NETWORK

INPUT signal A

INPUT signal B

receptor proteins

inactive transcription factor A

inhibitory

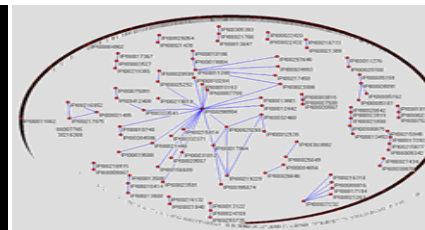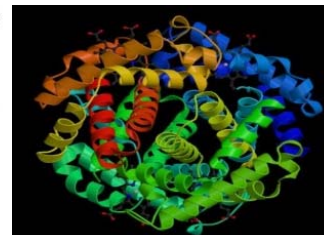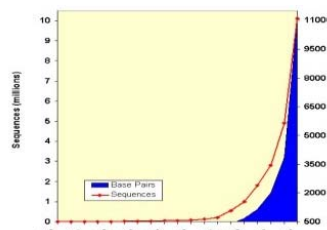active transcription

OUTPUT mRNA

Components

Network

System

cis-regulatory DNA sequence elements

feedbacks
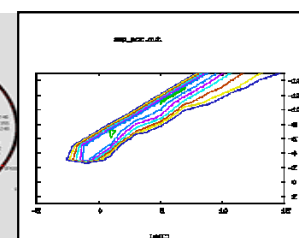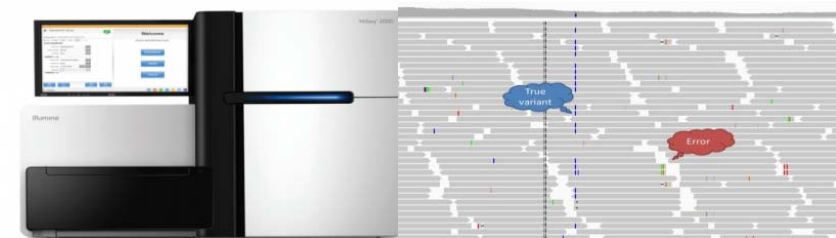
	B	C	D	E	F
1	gene	nsc1	nsc1 SE	nsc2	nsc2 SE
2	brain protein	18.9574	3.79952	21.5848	3.02241
3	Cluster Incl AW1	110.513	7.84625	114.894	7.95669
4	Cluster Incl AI8	235.873	35.6748	210.349	27.612
5	Cluster Incl AV3	47.4605	3.94976	29.6941	3.6586
6	Cluster Incl AV1	28.4527	3.74512	15.2986	3.62097
7	Cluster Incl AV1	80.302	6.45368	107.23	8.09591
8	Cluster Incl AV3	40.8113	5.13418	54.0835	3.18591
9	Cluster Incl AI1	53.1437	3.63392	58.635	5.50994

(Biological) Knowledge

(Expression) Information

(Sequencing) Data

(Expression) Information
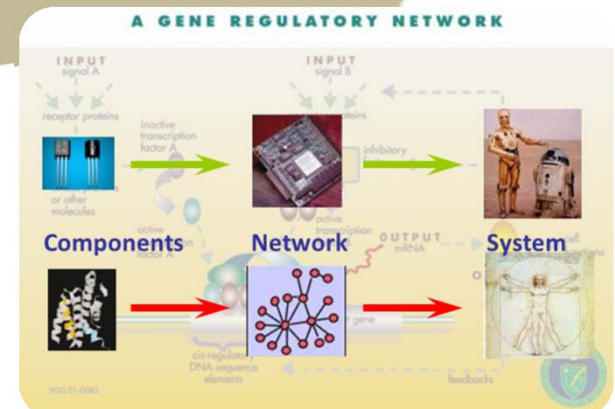
Differentially Expression Calling

Clustering/Classifying
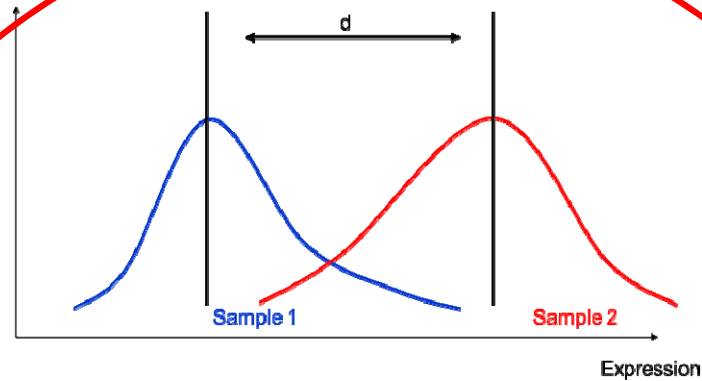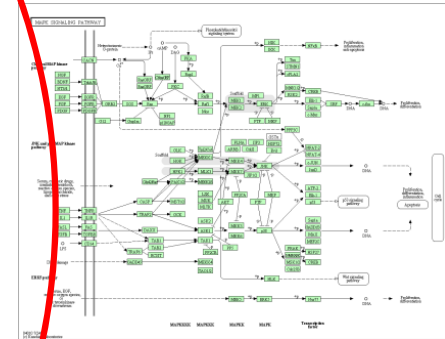
Pathway

- Functional assignment
- Pathway Enrichment
- Biological Processes
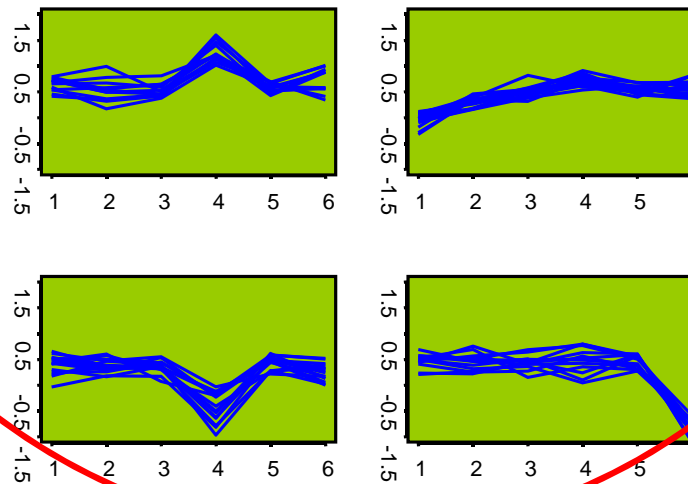
Copyright © Peking University

Probability: Given the information in the pail, what is in your hand?

Statistics: Given the information in your hand, what is in the pail?

(Figure Source: http://ocw.mit.edu/OcwWeb/Economics/14-30Spring-2006/CourseHome/index.htm)

Statistical Learning-guided Mining
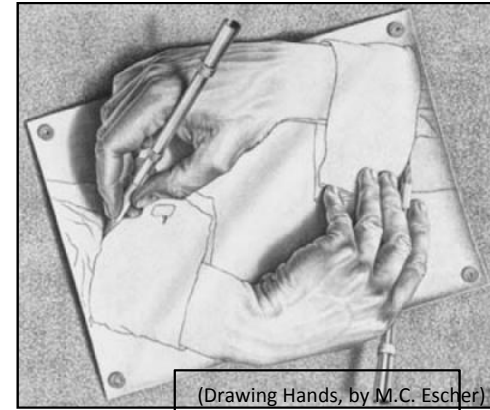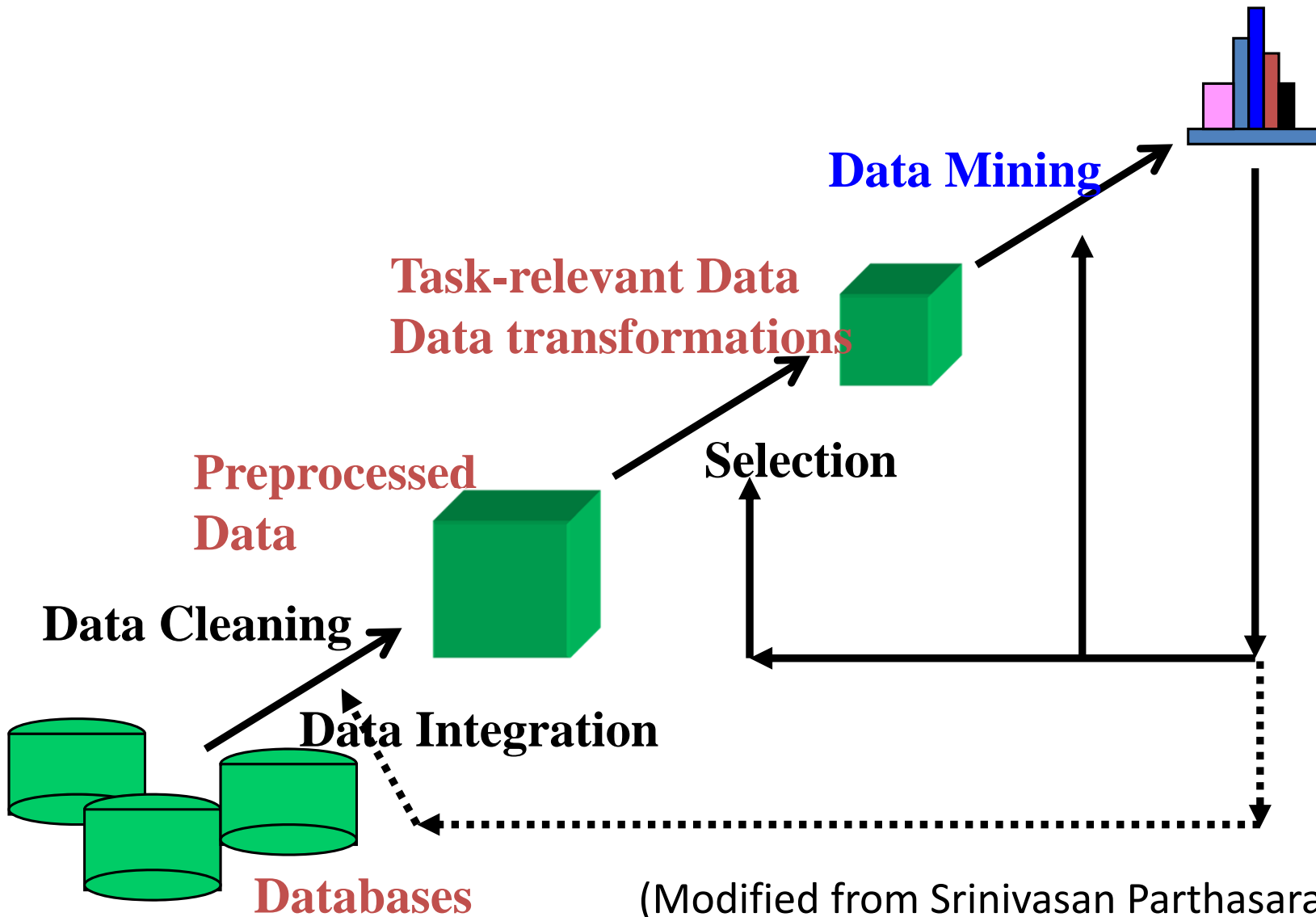
(Drawing Hands, by M.C. Escher)

Data Mining

Task-relevant Data
Data transformations

Preprocessed Data

Data Cleaning

Selection

Data Integration

Databases

(Modified from Srinivasan Parthasarathy, Ohio State Univ)

Data
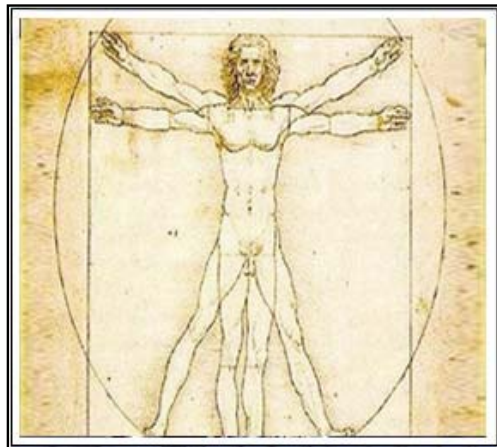
Model/Algorithm

(Prior) biological knowledge

(Domain Knowledge)

Parameters

- the transcriptome

cellular functions and processes

… – growth – differentiation – apoptosis – migration – cell cycle regulation – signal transduction – transcription - …

A **non-coding RNA (ncRNA)** is any RNA molecule that could function without being translated into a protein.

The DNA sequence from which a non-coding RNA is transcribed as the end product is often called an RNA gene or **non-coding RNA gene**.

Early discovered ncRNAs are mostly housekeeping

- "Assist" in translation in a necessary, but passive roles
- Constitutively expressed
- Include
  - rRNA
  - tRNA
  - snRNA
  - snoRNA
  - tmRNA
  - telomerase RNA
  - …

Recently discovered regulatory ncRNAs since 2000

- actively regulate gene transcription and translation
- are involved in various gene regulations through multiple mechanisms
- Many have specific expression patterns
- are widely encoded in the genome
  - The ENCODE (ENCyclopedia Of DNA Elements) pilot project suggested that over 90% of the human genome may be represented in primary transcripts.
  - Over 95% of all transcripts are noncoding. Some estimate the number of ncRNAs to be ~30,000.

INSIGHTS OF THE DECADE
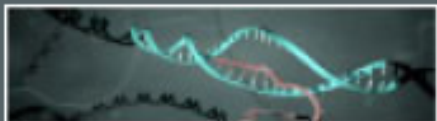
THE DARK GENOME
Since the publication of the human genome sequence in 2001, scientists have found that the so-called junk DNA that lies between genes actually carries out many important functions.

(http://www.sciencemag.org/site/special/insights2010/)

Representative Regulatory Mechanisms of ncRNAs

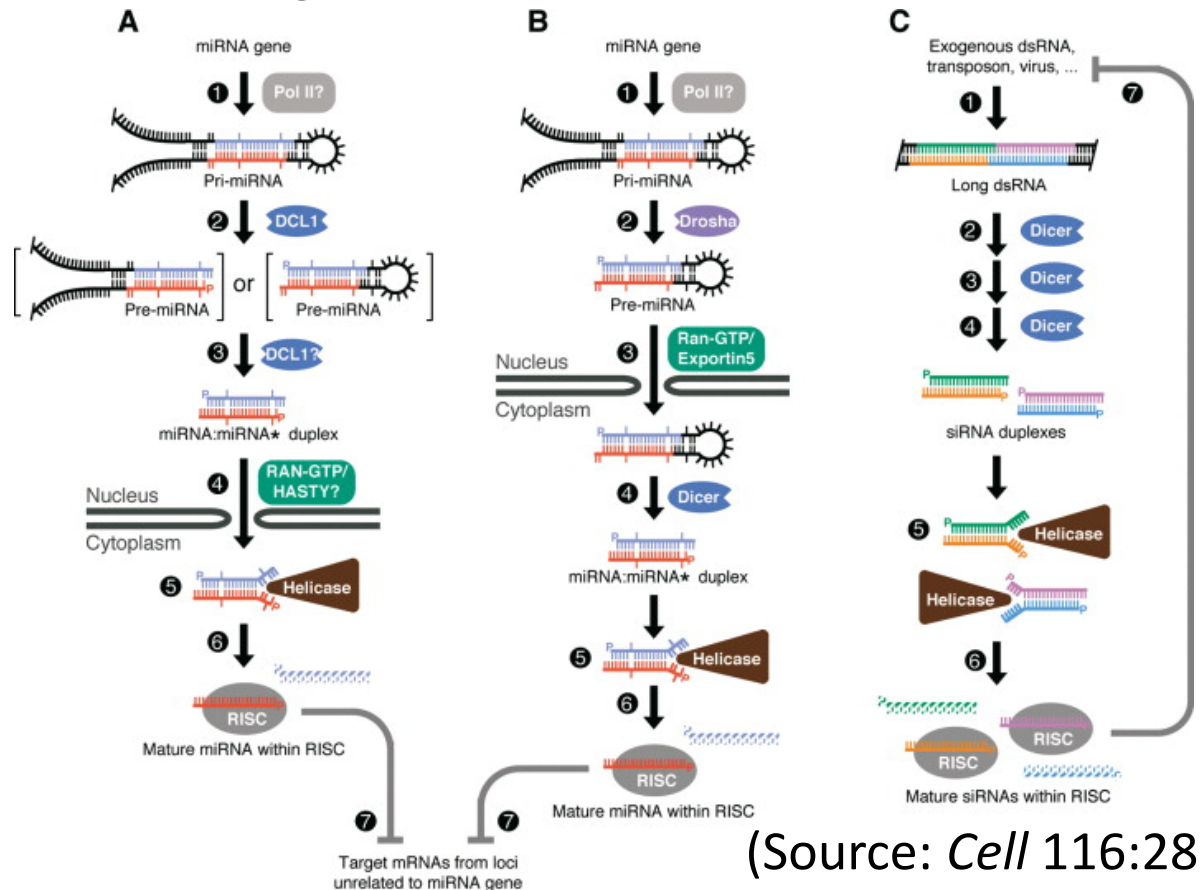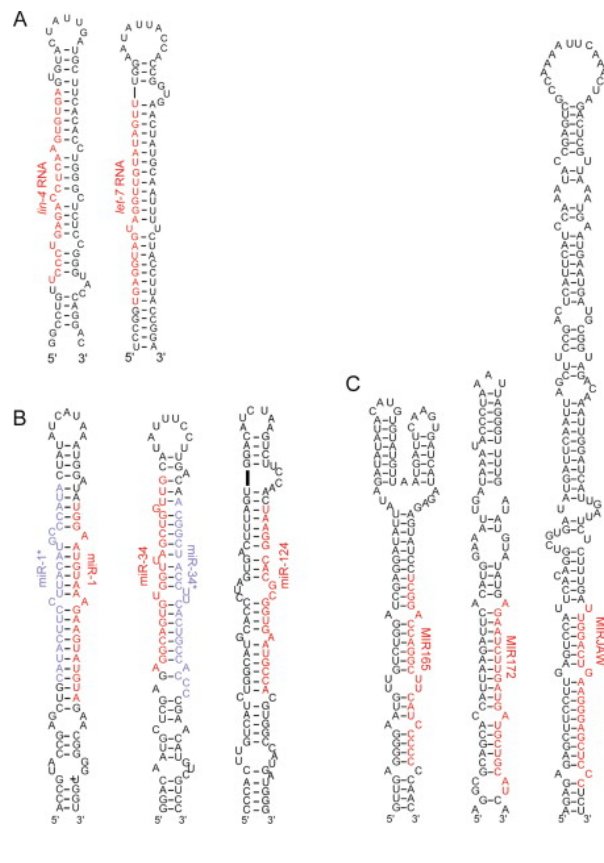| Mechanism	Orgnism	Example
Transcriptional repression	Several orgnisms	Riboswitches
Post-transcriptional regulation	Mouse	miR-196
Translational repression	E. coli	DicF
Translational activation	E. coli	RprA
DNA methylation	Arabidopsis	miRNA
DNA demethylation	Human	KHPS1a
Modification of the histone proteins	Arabidopsis	ncRNA
Regulation of chromatin structure	Yeast	ncRNA
Regulation of mRNA stability	Mouse	Makorin1-p1
Dosage compensation	Drosophila	roX1/roX2
Genomic imprinting	Human	AIR
X chromosome inactivation	Human	XIST
X chromosome activation	Human	TSIX

Qi, Sci China '06

Table 4   ncRNAs regulate various physiological and pathological events

Event	Organism	Example
Normal events		
Embryo development,	human	Let-7, miRNAs
Cell differentiation	human	NRSE, miR-143
Cell proliferation	Drosophila	Bantam
Regulation of apoptosis	human	ADAPT33
Fat metabolism	Drosophila	Mir-14
Modulation of behaviour	mouse	Bc1
Formation of photoreceptors	rat	TUG1
Regulation of insulin secretion	mouse	miR-375
Regulation of protein localization	Drosophila	hsr
Disease events		
Breast cancer	human	BC200
Colon cancer	human	miR-143, miR-145
Prostate cancer	human	PCGEM1
Lung cancer	human	Let-7
Liver cancer	rat	H19
Myeloid leukemia	mouse	HIS-1
B-CLL	human	miR-15a, miR-16a
B-cell neoplasia	human	BCMS
Angelman syndrome	human	UBE3A/SNURF-SNRPN
Beckwith-Wiedemann Syndrome	human	LIT1
Schizophrenia and bipolar	human	DISC2
Spinocerebellar ataxia	human	SCA8
Prader–Willi syndrome	human	ZNF127AS
Alzheimer's disease	human	BC200
Psoriasis	human	PRINS
Russel-Silver syndrome	human	MESTIT1

microRNA (miRNA)

- single-stranded RNAs of 21-23 (or some say 20-25) nt RNAs with regulatory functions when associated with a protein complex.

- In plants miRNAs can silence gene activity via destruction of homologous mRNA or blocking its translation. In animals, miRNAs inhibit translation by binding with imperfect homology to the 3' untranslated region of mRNA.

(Source: *Cell* 116:281)

Cancer type*	MiRNA profiling data	Significance	Refs
Chronic lymphocytic leukaemia	A unique signature of 13 genes associated with prognostic factors (ZAP70 and IgVH mutation status) and progression (time from diagnosis to therapy)	MiRNAs as diagnostic markers (the identification of two categories of patients)	49,35
Lung adenocarcinoma	Molecular signatures that differ with tumour histology; miRNA profiles correlated with survival (*miR-155* and *let-7*)	MiRNAs as prognostic and diagnostic markers	53
Breast carcinoma	MiRNA expression correlates with specific pathological features	MiRNAs as prognostic markers	50
Endocrine pancreatic tumours	A signature that distinguishes endocrine from acinar tumours; the overexpression of *miR-21* is strongly associated with both a high Ki67 proliferation index and the presence of liver metastases	MiRNAs as diagnostic and prognostic markers	54
Hepatocellular carcinoma	MiRNA expression correlated with differentiation	MiRNAs as prognostic markers	52
Papillary thyroid carcinoma	MiRNA upregulation (for example, *miR-221* and *miR-222*) in tumoral cells and normal cells adjacent to tumours, but not in normal thyroids without cancers	MiRNAs probably involved in cancer initiation	37 114
Glioblastoma	A specific signature compared with normal tissues	MiRNAs as diagnostic markers	51
Human cancers	MiRNA-expression profiles accurately classify cancers; an miRNA classifier classes poorly differentiated samples better than a messenger RNA classifier	MiRNAs as diagnostic markers	41
Human solid cancers	Common signature for distinct types of solid carcinomas	Specific miRNAs are involved in common molecular pathways	47

*Only data from microarray studies reporting results on human primary tumours were included in this table. IgV$_{H}$, immunoglobulin heavy-chain variable-region, MiRNA, microRNA. ZAP70, 70 kDa zeta-associated protein.

(Source: *Nat Rev Cancer* **6**, 857)

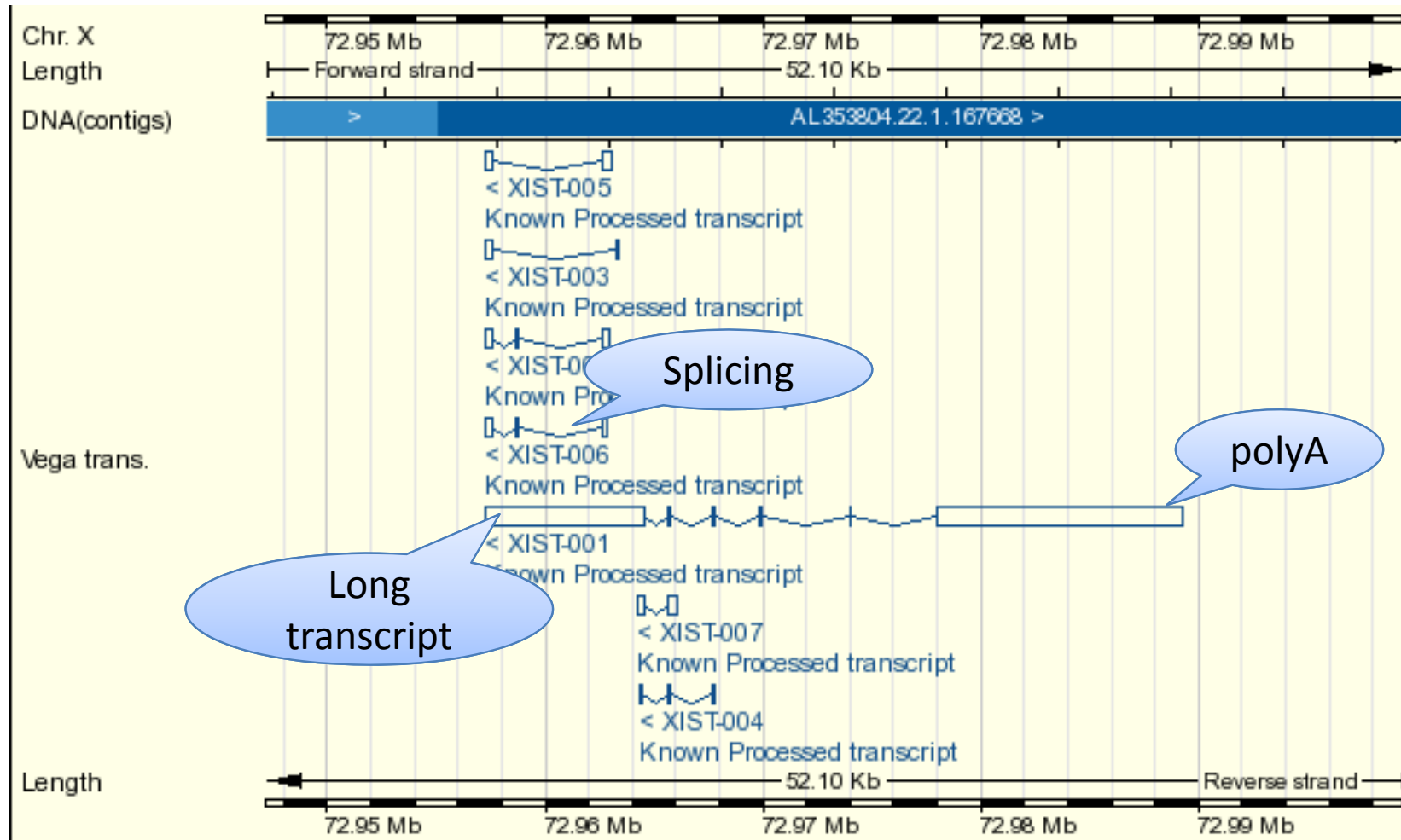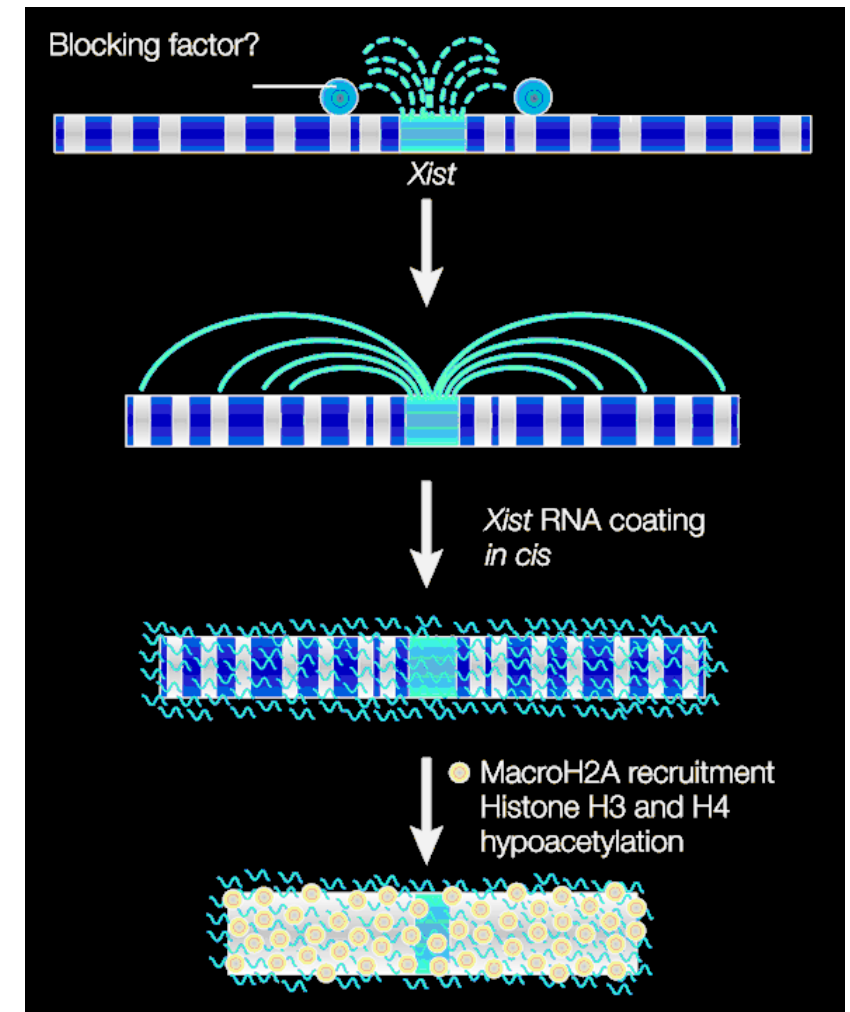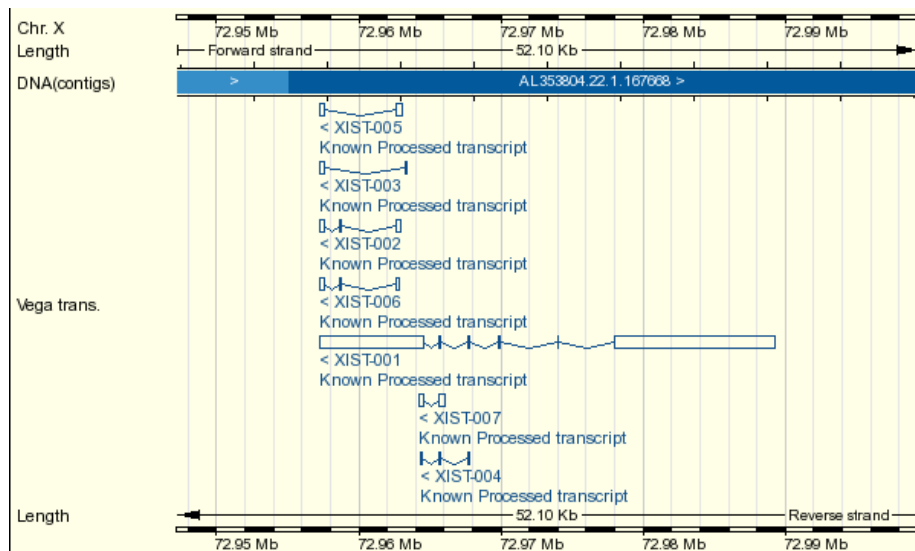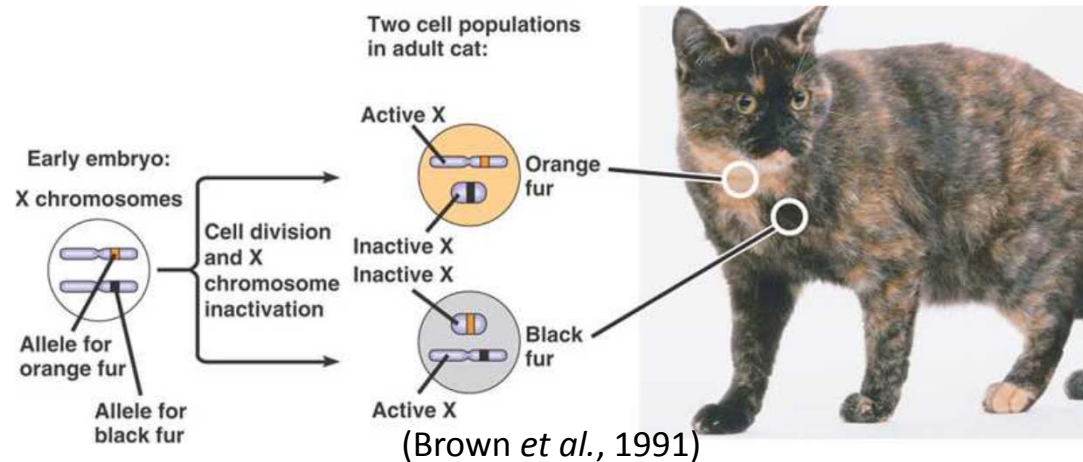| Generic Name	Originator	Status	Pharmacology	Target	Indication
SPC-3649	Santaris Pharma	Phase I	MicroRNA inhibitor	microRNA 122	Infection, hepatitis-C virus Hypercholesterolaemia
antagomirs, Alnylam	Alnylam	Preclinical	MicroRNA inhibitor	Unspecified	Unspecified
anti-inflammatory mi-croRNA,Reg	Alnylam*	Preclinical	MicroRNA inhibitor	Unspecified	Unspecified
anticancer microRNA, Regulus	Alnylam*	Preclinical	MicroRNA inhibitor	Unspecified	Unspecified
anti-miR-122 oligo, Regulus	Alnylam*	Preclinical	MicroRNA inhibitor	microRNA 122	Infection, hepatitis-C virus
miRNA inhibitors, Mi-ragen	Miragen Therapeutics	Preclinical	MicroRNA inhibitor	microRNA 208a	Heart failure
miRNA mimetics, Mi-ragen	Miragen Therapeutics	Preclinical	MicroRNA stimulant	Unspecified	Heart failure
prostate cancer miRNAs, Mirna	Mirna Therapeutics	Preclinical	MicroRNA stimulant	Unspecified	Cancer, prostate
AML miRNA therapy, Mirna	Mirna Therapeutics	Preclinical	MicroRNA stimulant	Unspecified	Cancer, leukaemia, acute mye-logenous
nsclc miRNA therapy, Mirna	Mirna Therapeutics	Preclinical	MicroRNA stimulant	microRNA let-7a-1	Cancer, lung, non-small cell
herpes virus therapy, Rosetta	Rosetta Genomics	Preclinical	MicroRNA inhibitor	Unspecified	Infection, Epstein-Barr virus Infection, herpes simplex virus
miR-34a mimetics, Rosetta	Rosetta Genomics	Preclinical	MicroRNA stimulant p53 stimulant Apoptosis agonist	microRNA 34a tumour protein p53	Cancer, liver
hepatitis-C therapy, Rosetta	Rosetta Genomics	Preclinical	MicroRNA inhibitor	Unspecified	Infection, hepatitis-C virus
HIV therapy, Rosetta	Rosetta Genomics	Preclinical	MicroRNA inhibitor	Unspecified	Infection, HIV/AIDS

*Alnylam/Isis Pharmaceuticals joint-venture   (http://www.pharmaprojects.com/therapy_analysis/microRNA-0808-therapeutictarget.html)

Xist : Beyond "small" ncRNA

Xist – X inactive-specific transcript

(Brown *et al.*, 1991)

Blocking factor?

Xist

Xist RNA coating
in cis

MacroH2A recruitment
Histone H3 and H4
hypoacetylation

(Avner *et al.*, 2001)
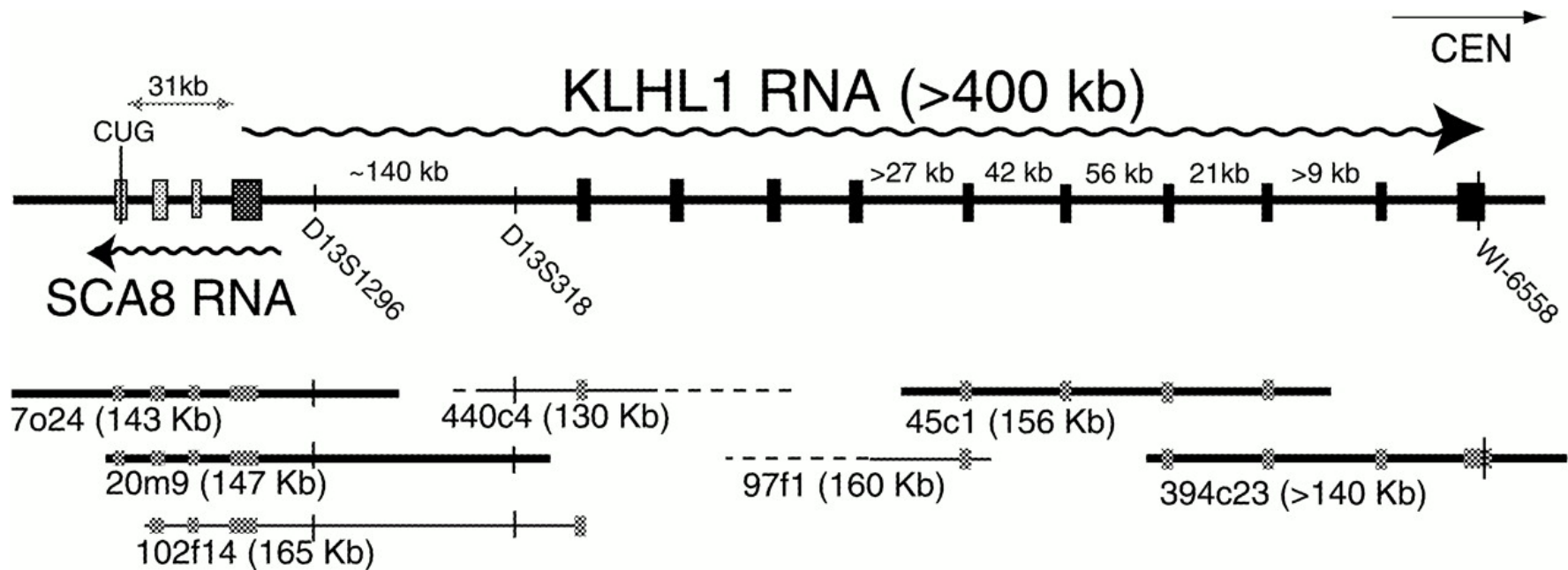
SCA8:
Long ncRNA in Human Disease

- SCA8 is mutated in one form of spinal cerebella ataxia

(Nemes, J. P. *et al.* 2000)

Long ncRNAs

- Estimated ~2000+ in human.
- Some, but not all, are mRNA-like, with Poly(A) tails.
- Most have unknown function. Many may function via *cis* or *trans* antisense pairing.
  - Dosage compensation (e.g. XIST)
  - Neuron development (e.g. SCA8)
  - Genetic imprinting (e.g. IGF/H19)
  - Post-transcriptional regulation
    - mRNA degradation or stabilization
  - Translational regulation
  - Modulate protein function by directly binding to the protein

How many non-coding transcripts?

What are the functional roles of those ncRNAs?

生物信息学：导论与方法
Bioinformatics: Introduction and Methods

https://www.coursera.org/course/pkubioinfo