

生物信息学：导论与方法

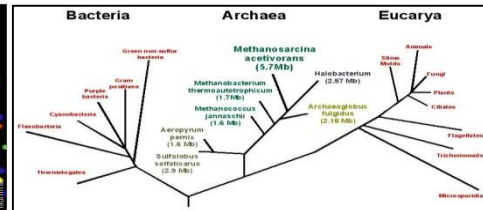
Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍

Center for Bioinformatics, Peking University



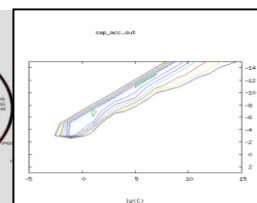
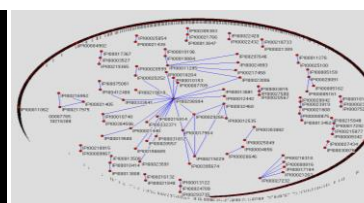
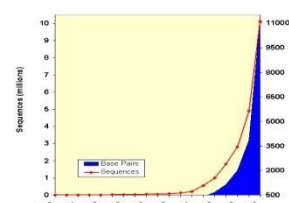
<https://www.coursera.org/course/pkubioinfo>



北京大学生物信息学中心 魏丽萍

Liping Wei, Ph.D.

Center for Bioinformatics, Peking University



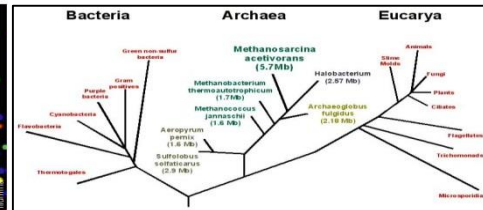
Lessons learnt beyond the methods themselves

Towards developing a new bioinformatic method:

- how to identify an important and unsolved biological question
- how to formulate it into a computational problem
- how to come up with the idea for an algorithm to solve the problem
- how to implement the algorithm
- how to evaluate them

Towards using a bioinformatic method:

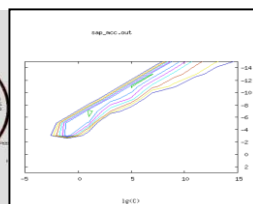
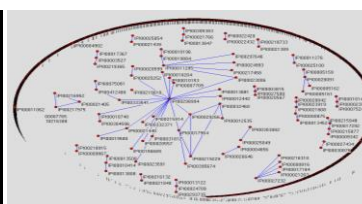
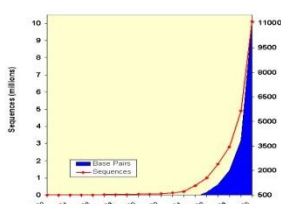
- what biological questions does it address?
- what are the underlying assumptions?
- what are the different parameters and what do they do?
- what are the accuracies of the method?
- what are its limitations?



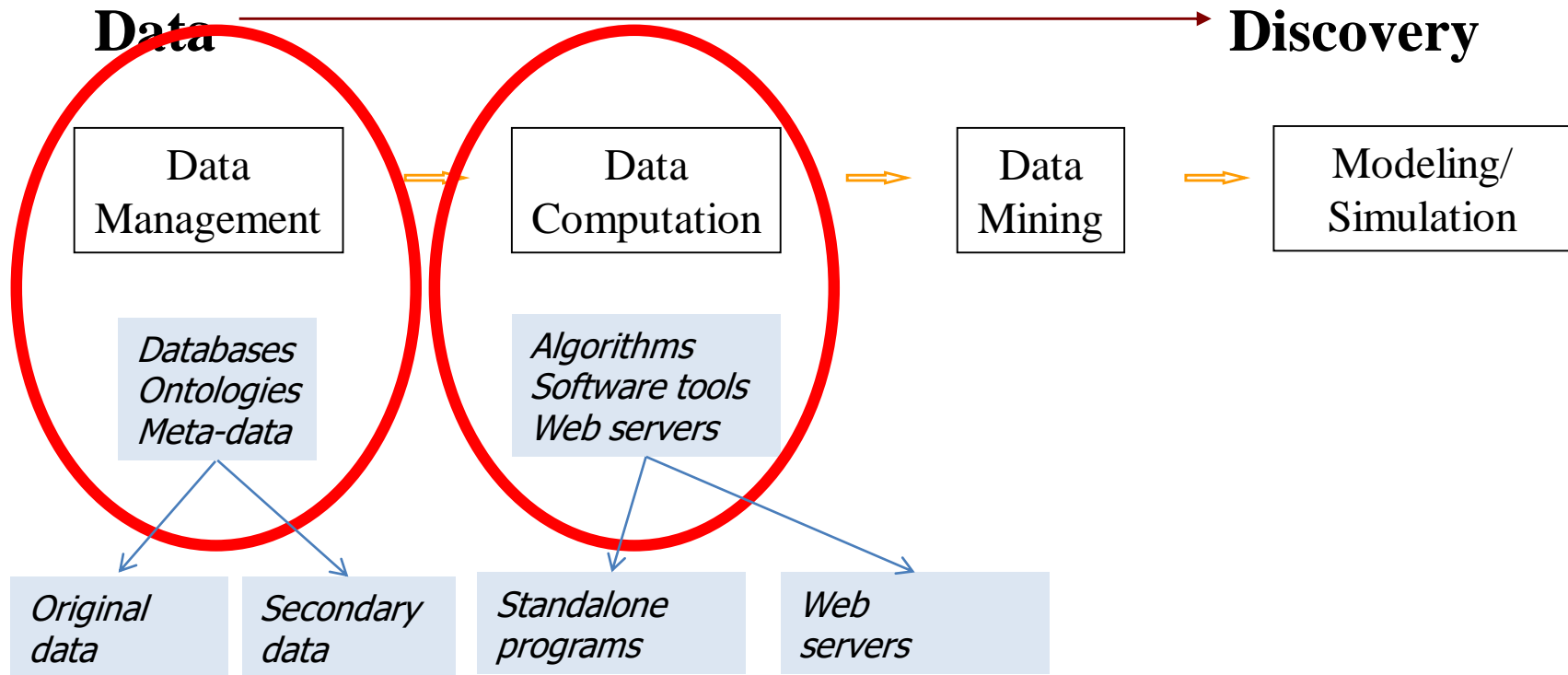
北京大学生物信息学中心 魏丽萍

Liping Wei, Ph.D.

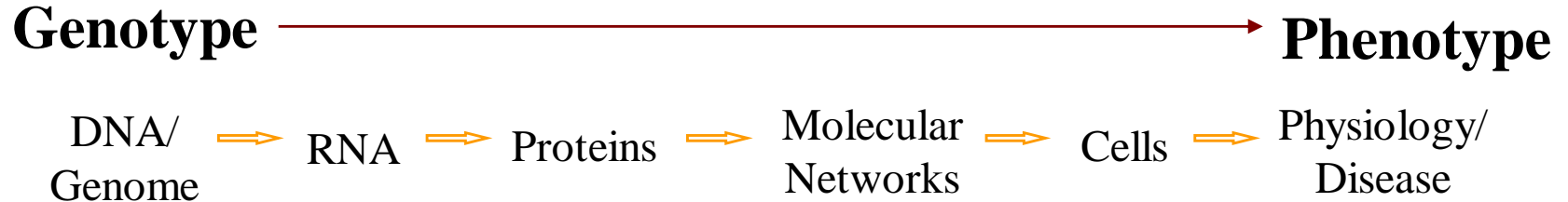
Center for Bioinformatics, Peking University



The –informatics in Bioinformatics



The Bio- in Bioinformatics



Centralized vs. individual resources

NCBI: <http://www.ncbi.nlm.nih.gov/>

The screenshot shows the NCBI homepage. A red rectangle highlights the left-hand navigation menu, which includes the following items: NCBI Home, Resource List (A-Z), All Resources, Chemicals & Bioassays, Data & Software, DNA & RNA, Domains & Structures, Genes & Expression, Genetics & Medicine, Genomes & Maps, Homology, Literature, Proteins, Sequence Analysis, Taxonomy, Training & Tutorials, and Variation. The 'Gene' dropdown menu is open, showing a list of databases including All Databases, Assembly, BioProject, BioSample, BioSystems, Books, ClinVar, Clone, Conserved Domains, dbGaP, dbVar, Epigenomics, EST, Gene (highlighted), Genome, GEO DataSets, GEO Profiles, GSS, HomoloGene, and MedGen. The main content area features the NCBI logo, a search bar, and a 'Sign in to NCBI' link. Below the search bar, there is a section for 'Popular Resources' with links to PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. A 'NCBI Announcements' section is also visible, mentioning a planned change in bacterial strain-level information management as of November 21, 2013. At the bottom, there is a 'Genetic Testing Registry' banner with a 'GO' button and a slide navigation bar.

NCBI Resources ▾ How To ▾ Sign in to NCBI

NCBI
National Center for Biotechnology Information

Gene ▾
All Databases
Assembly
BioProject
BioSample
BioSystems
Books
ClinVar
Clone
Conserved Domains
dbGaP
dbVar
Epigenomics
EST
Gene
Genome
GEO DataSets
GEO Profiles
GSS
HomoloGene
MedGen

NCBI
National Center for Biotechnology Information advances science and health by providing access to biomedical information.

Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Popular Resources
PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI Announcements
Planned change in bacterial strain-level information management
Nov 21, 2013
Please be aware that there is an upcoming change (January 2014) in how...
Exploring next-gen sequencing

Genetic Testing Registry
A portal to clinical genetics resources with detailed information about genetic tests and laboratories.
GO

1 2 3 4 5 6 7 8

EBI: <http://www.ebi.ac.uk/services>

EMBL-EBI

Services

Research

Training


About us


Services


OverviewA to ZService teamsSupport


Bioinformatics services


We maintain the world's most comprehensive range of **freely available** and up-to-date molecular databases. Developed in collaboration with our colleagues worldwide, our services let you share data, perform complex queries and analyse the results in different ways. You can work locally by downloading our data and software, or use our web services to access our resources programmatically.


 DNA & RNA
genes, genomes & variation


 Gene expression
RNA, protein & metabolite expression


 Proteins
sequences, families & motifs


 Structures
Molecular & cellular structures

 Systems
reactions, interactions & pathways


 Chemical biology
chemogenomics & metabolomics


 Ontologies


 Literature


 Other software


Popular


 Ensembl


 UniProt


 PDB


 ArrayExpress


 ChEMBL

 BLAST

 Europe PMC

 Reactome

 Train online

 Support

Bioinformatics training

Guide to resources

Service news

Copyright © Peking University

UCSC Genome Bioinformatics (<http://genome.ucsc.edu/>)

UCSC Genome Bioinformatics

[Genomes](#) - [Blat](#) - [Tables](#) - [Gene Sorter](#) - [PCR](#) - [VisiGene](#) - [Session](#) - [FAQ](#) - [Help](#)

[Genome
Browser](#)

[ENCODE](#)

[Neandertal](#)

[Blat](#)

[Table
Browser](#)

[Gene Sorter](#)

[In Silico PCR](#)

[Genome
Graphs](#)

[Galaxy](#)

[VisiGene](#)

[Utilities](#)

[Downloads](#)

About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the [ENCODE](#) and [Neandertal](#) projects.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work of annotators worldwide. The [Gene Sorter](#) shows expression, homology and other information on groups of genes that can be related in many ways. [Blat](#) quickly maps your sequence to the genome. The [Table Browser](#) provides convenient access to the underlying database. [VisiGene](#) lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. [Genome Graphs](#) allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering ([CBSE](#)) at the University of California Santa Cruz ([UCSC](#)). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our [public mailing list](#).

News



[News Archives](#) ►

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list.

27 November 2013 - 100 Species Conservation Track now available on hg19

After 15.4 years of CDL run time in 2,005,504 individual files and 90 clusters for last-minute alignment, we are excited to announce

UCSC Genome Browser

<http://genome.ucsc.edu/cgi-bin/hgTracks>

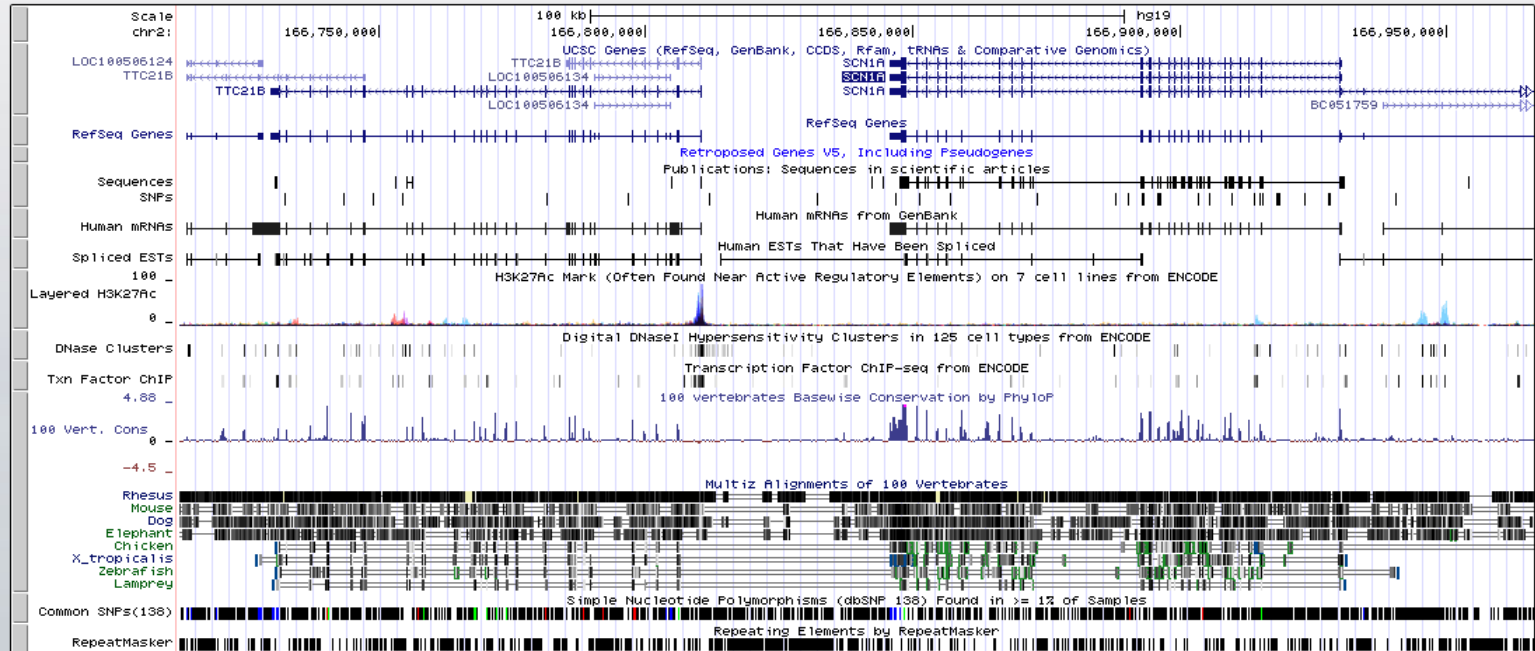
[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[About Us](#)[View](#)[Help](#)

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

chr2:166,712,755-166,966,287 253,533 bp.

go



move start

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track

move end

Center for Bioinformatics, Peking University (<http://cbi.pku.edu.cn>)



Center for Bioinformatics
Peking University

Tilsi Introduction

Chinese Version

HOME
PEOPLE
PUBLICATION
RESOURCE
COURSES
DOWNLOAD
FAQ
NEWS

Site Search

Advanced Site Search

LINKS 

Introduction

Welcome to CBI - the Center for Bioinformatics is the first bioinformatics center in China, founded in 1996. It is currently home to seven faculty and staff, one adjunct faculty, and 20 doctorate students. Located in the new Life Science Building on Peking University campus, it houses several computing labs with strong computer hardware and software facilities and a molecular biology lab. While continuing to maintain the first and largest online bioinformatics resource in China as China's official national node of EMBnet, CBI has published in the areas of gene expression regulation, genome analysis and evolution, and pathway networks. CBI is supported by the Ministry of Science and Technology, Ministry of Education, and Natural Science Foundation of China.

Bioinformatics Resource at CBI

PlantTFDB	Plant transcription factor database
Rice-Map	A Map Like Rice Genome Browser and Related Resource
FTP	FTP server for public data (e.g. GenBank, EMBL) downloading.
WebLab	Your lab on the web
Resource	CBI research, development or maintaining resource (hESlincRNABrowser , CPC , PlantTFDB , WebLab , FTP , ...)
Download	Public ftp/rsync server of database and software

NCBI/EBI/UCSC Summary Table

	NCBI	EBI	UCSC
Tools	BLAST	BLAST Exonerate ClustalW2	BLAT In-Silico PCR
Data Repository	GenBank GEO SRA	ArrayExpress ENA PDBe	ENCODE
DNA/Genome	Genome	Ensembl	Ideogram Recombination Rate GC Content

NCBI/EBI/UCSC Summary Table (Cont'd)

	NCBI	EBI	UCSC
DNA/Gene	Gene	Ensembl	UCSC Genes GENCODE RefSeq Genes
RNA	RefSeq	Ensembl	mRNAs ESTs UniGene
Proteins	Protein, RefSeq Conserved Domain	UniProt, InterPro, PRIDE (mass spec)	

NCBI/EBI/UCSC Summary Table (Cont'd)

	NCBI	EBI	UCSC
Expression	UniGene	Expression Atlas	Affy Exon Array Caltech RNA-seq Allen Brain
Regulation			Transcription TFBS Epigenetics DNaseI HS
Literature	PubMed		
Ontology		Gene Ontology	

NCBI/EBI/UCSC Summary Table (Cont'd)

	NCBI	EBI	UCSC
Comparative Genomics	Taxonomy HomoloGene	*Ensembl	Conservation Neandertal
Variation	dbSNP dbVar	*Ensembl	SNPs DGV RepeatMasker
Disease	OMIM MeSH dbGaP ClinVar		GAD COSMIC ClinVar GWAS Catalog QTLs

NCBI access

- All Resources

- Database:

- <http://www.ncbi.nlm.nih.gov/guide/all/#databases>

- Tools:

- <http://www.ncbi.nlm.nih.gov/guide/all/#tools>

- Resource List (A-Z)

- <http://www.ncbi.nlm.nih.gov/guide/sitemap/>

EBI access

- Bioinformatics services
 - by topic:
<http://www.ebi.ac.uk/services>
 - DNA&RNA, Gene expression, Proteins, Structures, Systems, Chemical biology, Ontologies, Literature, ...
 - Popular: Ensembl, UniProt, PDBe, ArrayExpress, ChEMBL, BLAST, Europe PMC, Reactome, Train online, ...
 - by name(A-Z):
<http://www.ebi.ac.uk/services/all>

UCSC access

- Table Browser
 - <http://genome.ucsc.edu/cgi-bin/hgTables>
- Genome Browser
 - <http://genome.ucsc.edu/cgi-bin/hgTracks>
 - track hubs
 - configure

Examples of individual Resources

Analyses		Databases and tools
DNA/Genome	Gene prediction	GENSCAN, Glimmer
	Genetic and Somatic Variations	HGMD, COSMIC, SIFT, PolyPhen, SAPRED
Expression regulation	Transcription factors & binding	TRANSFAC, PlantTFDB
	RNA annotation	Rfam
	microRNA annotation	miRBase
Epigenetics	DNA methylation	MethylomeDB
Pathway & network	Pathway	KEGG, PANTHER, BioCyc, REACTOME, KOBAS, DAVID
	Interaction network	PID, STRING
Evolution	Conservation	GERP++, PHYML

Examples of individual resources for mass spectrometry proteomics analyses

	Databases and tools
Mass spec data repository	The Global Proteome Machine
Peptide identification by database search	Sequest, Mascot, ProteinProspector, pFind, PEAKS, Byonic, Proteome Discoverer, SpectrumMill, Masslynx, X!Tandem
<i>de novo</i> peptide identification	pNovo, PEAKS, PepNovo
Quantitation	MaxQuant, Census

Examples of Individual Resources for Protein 3D Structures

Analyses	Databases and Tools
Protein 3D structures	PDB
Homology Modeling	Modeller, Swiss-Model, I-TASSER
Fold Recognition	3D-PSSM, Phyre2, PROSPECT
<i>ab initio</i> folding	QUARK, Rosetta
Protein 3D models	Swiss-Model Repository
Nucleotide structure	Mfold, PDB
Nucleotide interaction	RNAhybrid

Examples of Individual Resources for NGS analysis

Analyses		Databases and tools
Read mapping	Reads mapping (DNA)	BWA, Bowtie/Bowtie2
	Reads mapping (RNA)	TopHat/TopHat2
	Reads mapping related utilities	GATK, FastQC, RNA-SeQC, SAMtools, Picard
Assembly	de novo genomic assembly	Velvet, SOAP de novo
	de novo transcriptome assembly	Trinity, Velvet + Oases
	reference-based transcriptome assembly	TopHat + Cufflinks
Visualization	Genome Browser	GBrowse, JBrowse, IGV
Variant calling	SNP	GATK, SOAPsnp, SAMtools
	Indel	Pindel
	CNV and structural variation	CNVnator, BIC-seq, SVMerge, mrCaNaVaR, ExomeCNV, CoNIFER, HMMcopy, Control-FREEC
Expression	Differential expression	Cuffdiff, DESeq/DESeq2
ChIP-Seq	Peak calling	MACS

Other examples of individual resources & utilities

Purpose		Databases and tools
Model Organisms	Genome and gene annotations	Flybase, Wormbase, ZFIN, TAIR
Large-scale studies	Cancer	TCGA, CGP
	Epigenetics	Roadmap Epigenomics Project
	Brain	Allen Brain Atlas, Human Connectome project
Tools to assist wet-lab experiments	Primer design	Primer3/Primer3Plus, Electronic PCR
Software programming utilities	Sequence analysis tool package	EMBOSS
	R package	Bioconductor
	Perl package	BioPerl
	Python package	BioPython
Workflow	Workflow platform	Galaxy
	Workflow construction	Taverna

2013 NAR database issue

<http://nar.oxfordjournals.org/content/41/D1.toc>

OXFORD JOURNALS

[CONTACT US](#)

[MY BASKET](#)

[MY ACCOUNT](#)

Nucleic Acids Research

[ABOUT THIS JOURNAL](#) [CONTACT THIS JOURNAL](#) [SUBSCRIPTIONS](#)

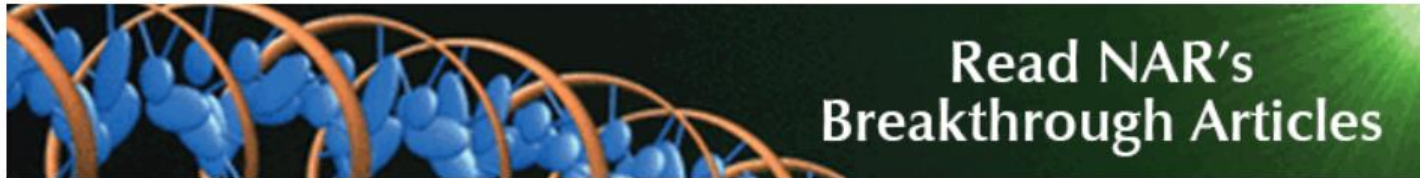
[CURRENT ISSUE](#)

[ARCHIVE](#)

[SEARCH](#)

Institution: Peking University [Sign In as Personal Subscriber](#)

[Oxford Journals](#) › [Life Sciences](#) › [Nucleic Acids Research](#) › Volume 41, Issue D1



Database issue

Volume 41 Issue D1 January 2013

For checked items

☐ view abstracts ☐ download to citation manager

[« Previous](#) | [Next Issue »](#)

This Issue

January 2013 41 (D1)

Search this journal:

[Advanced »](#)

Current Issue

[November 2013 41 \(21\)](#)

2013 NAR web server issue

<http://nar.oxfordjournals.org/content/41/W1.toc>

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Institution: **Peking University** Sign In as Personal Subscriber

Oxford Journals > Life Sciences > Nucleic Acids Research > Volume 41, Issue W1

Web Server issue


Volume 41 Issue W1 1 July 2013

For checked items
☐ view abstracts ☐ download to citation manager

Editorial

☐ Gary Benson
Editorial
Nucleic Acids Res. (1 July 2013) 41 (W1): W1-W2 doi:10.1093/nar/gkt559

» Extract » Full Text (HTML) » Full Text (PDF) » Screen PDF

OPEN ACCESS 

« Previous | Next Issue »

This Issue

1 July 2013 41 (W1)



Current Issue

November 2013 41 (21)



Copyright © Peking University

生物信息学：导论与方法

Bioinformatics: Introduction and Methods

Ge Gao 高歌 & Liping Wei 魏丽萍

Center for Bioinformatics, Peking University



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