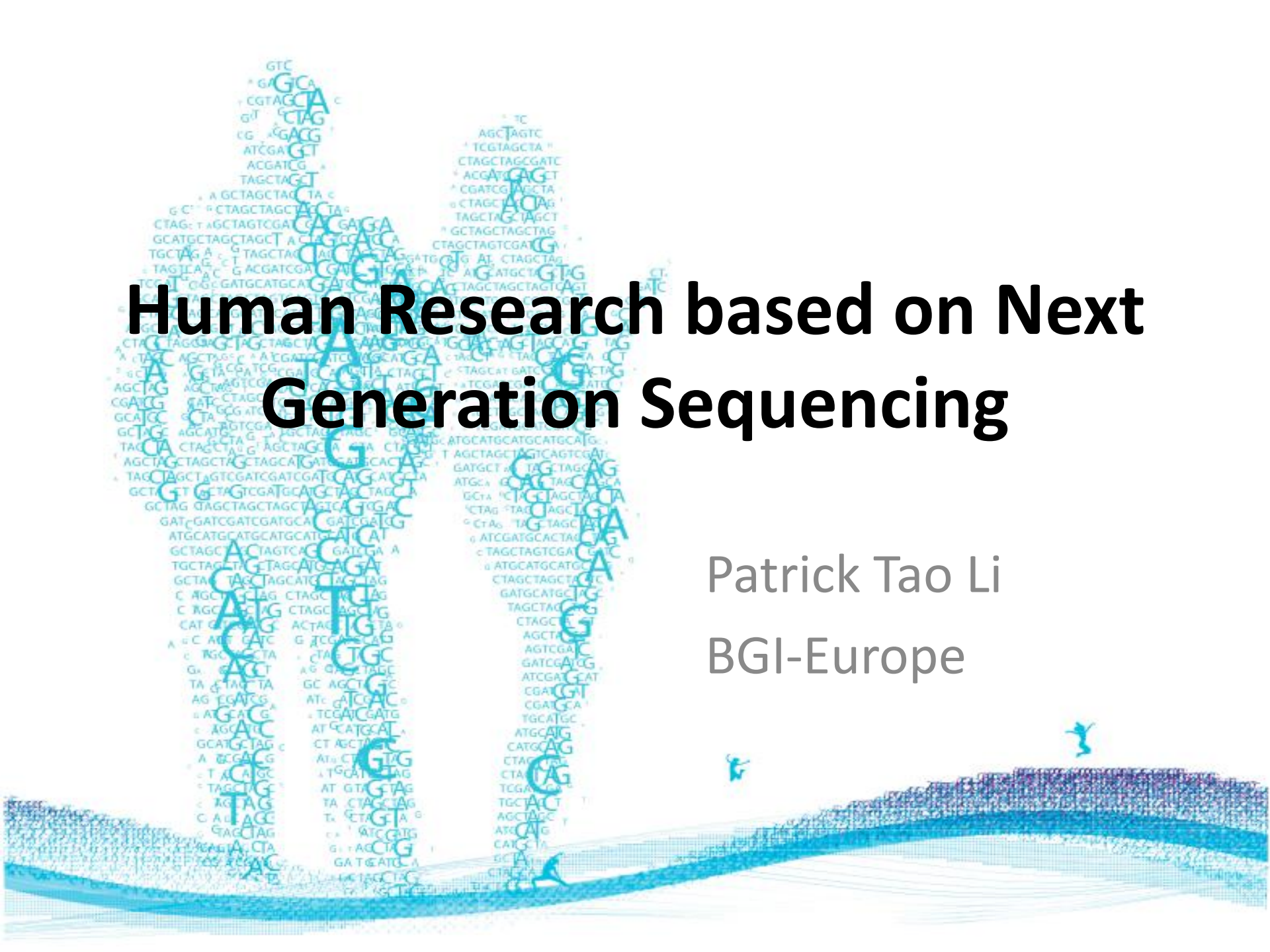


Human Research based on Next Generation Sequencing

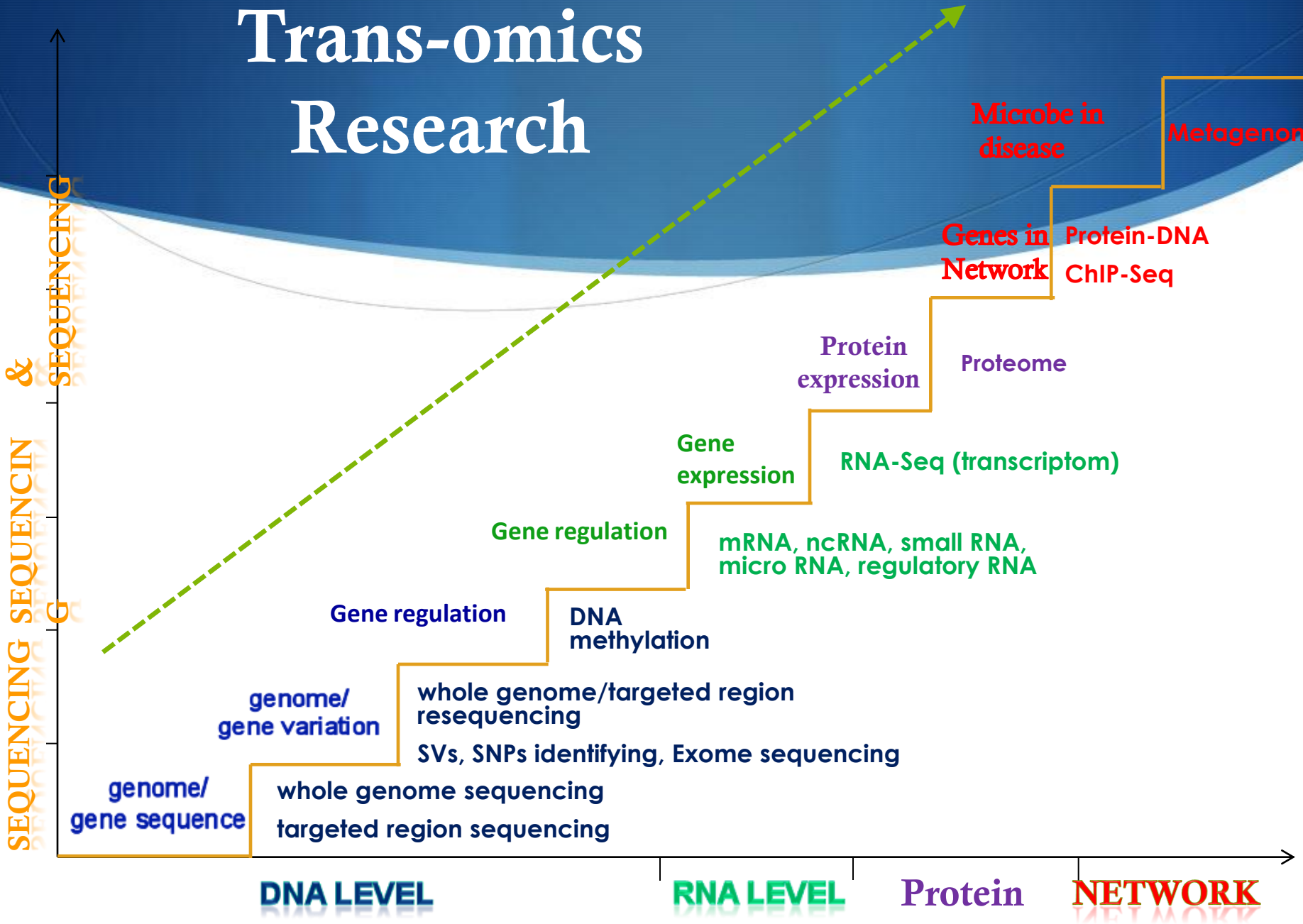
Patrick Tao Li

BGI-Europe

The candidate gene approach
hypothesis-driven

Do it in a different way!
data-driven, large scale analysis
-Slide from Francis Colins

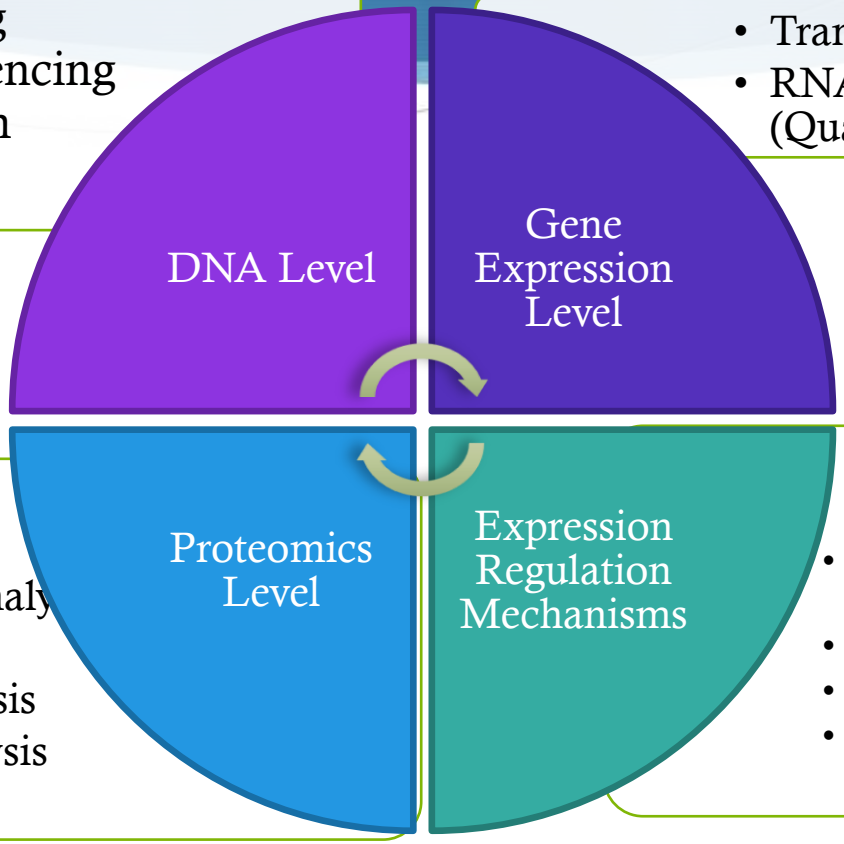
Trans-omics Research



Total Solution for Human Research

- Whole Genome Resequencing
- Exome Sequencing
- Target Region Sequencing

- Transcriptome Sequencing
- RNA Sequencing (Quantification)

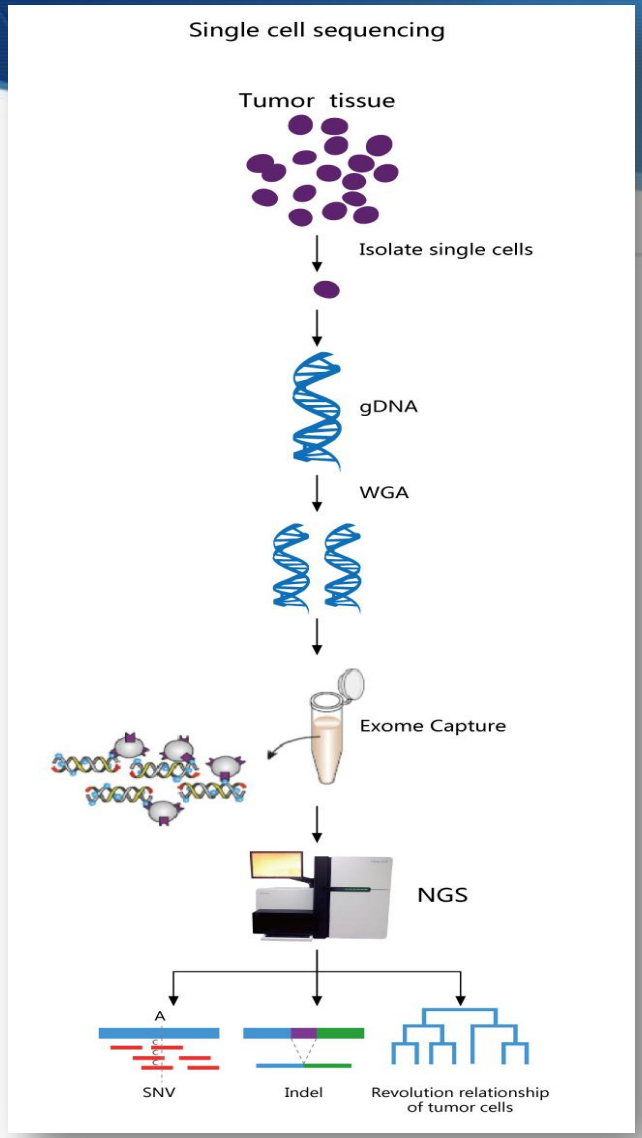


- Whole Genome Analy
- Target Analysis
- Quantitative Analysis
- Modification Analysis

- Whole Genome Bisulfite Sequencing
- RRBS
- ChIP Sequencing
- Small RNA Sequencing

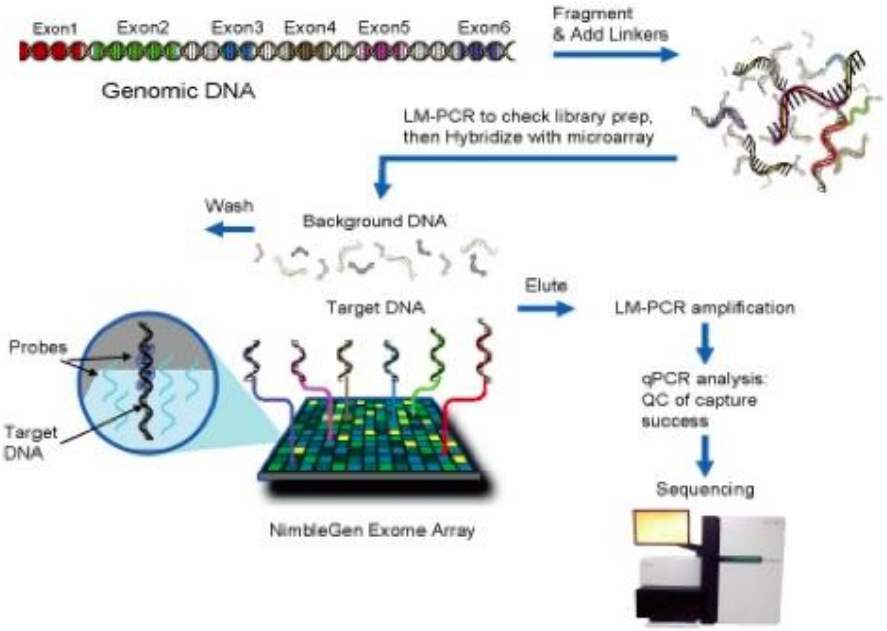
Whole Genome Profiling

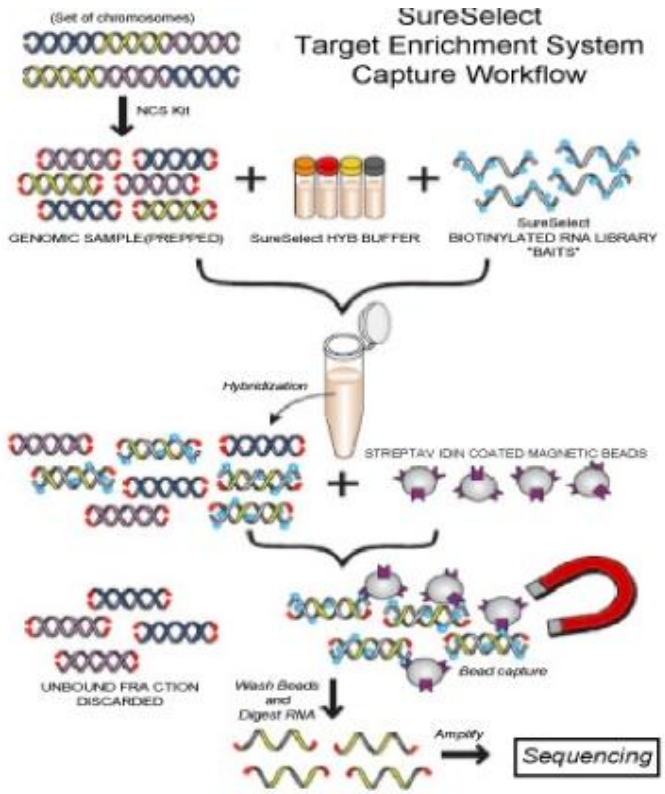
--Single Cell Resequencing



- Cancer cell evolution during tumor progression
- Large-scale epidemiological tumor research
- Early diagnosis and prognosis of cancer

Exome / Target Region Sequencing





Epigenomic Research based on NGS

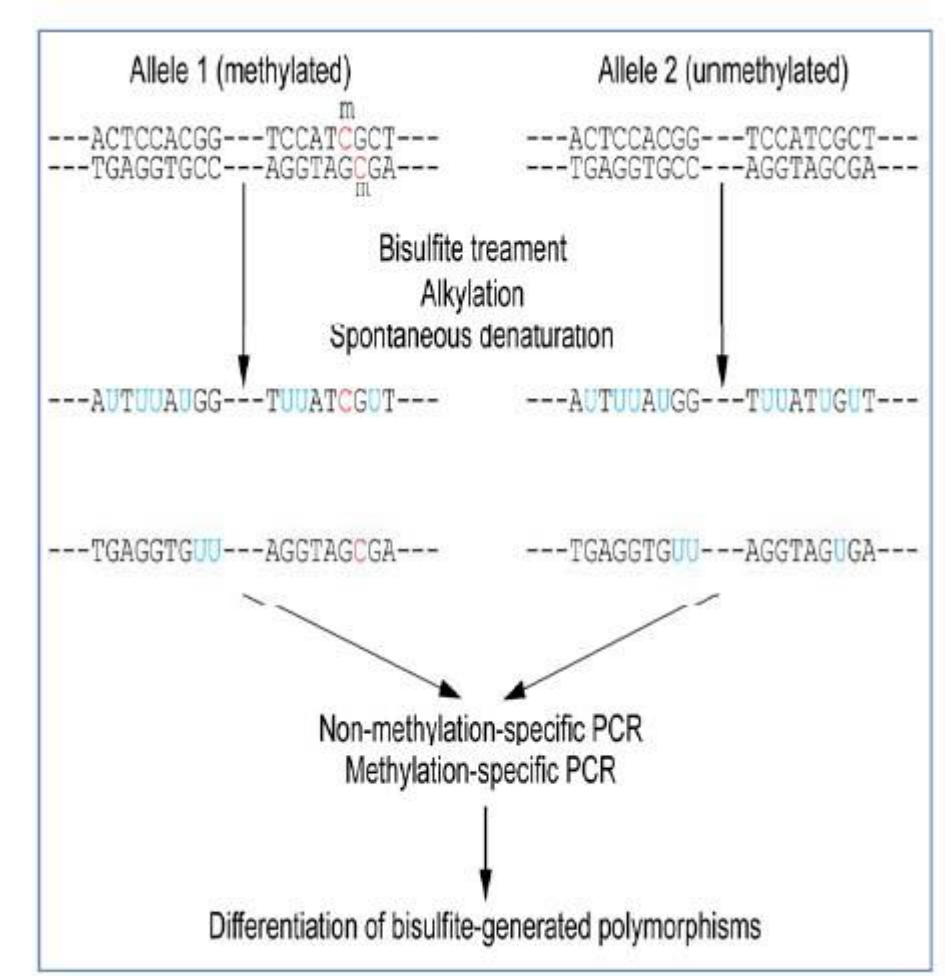
DNA Methylation

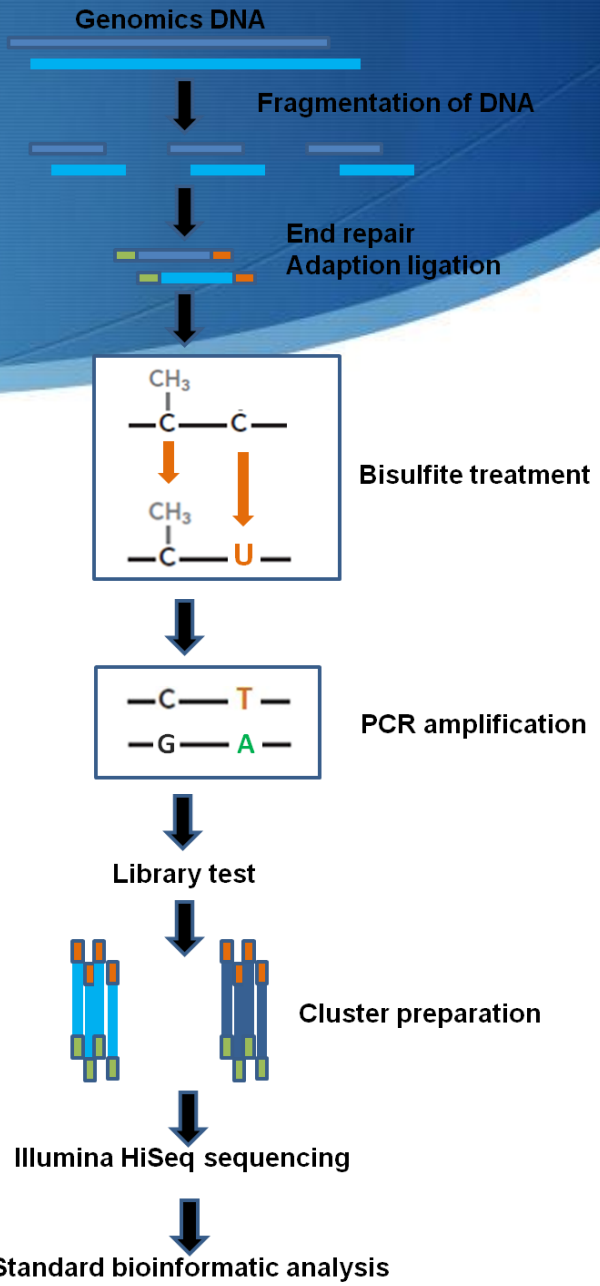
- Whole Genome Bisulfite Sequencing
- RRBS (Reduced Representative Bisulfite Sequencing)

DNA-Protein Interation

- ChIP Sequencing

Whole Genome Bisulfite Sequencing

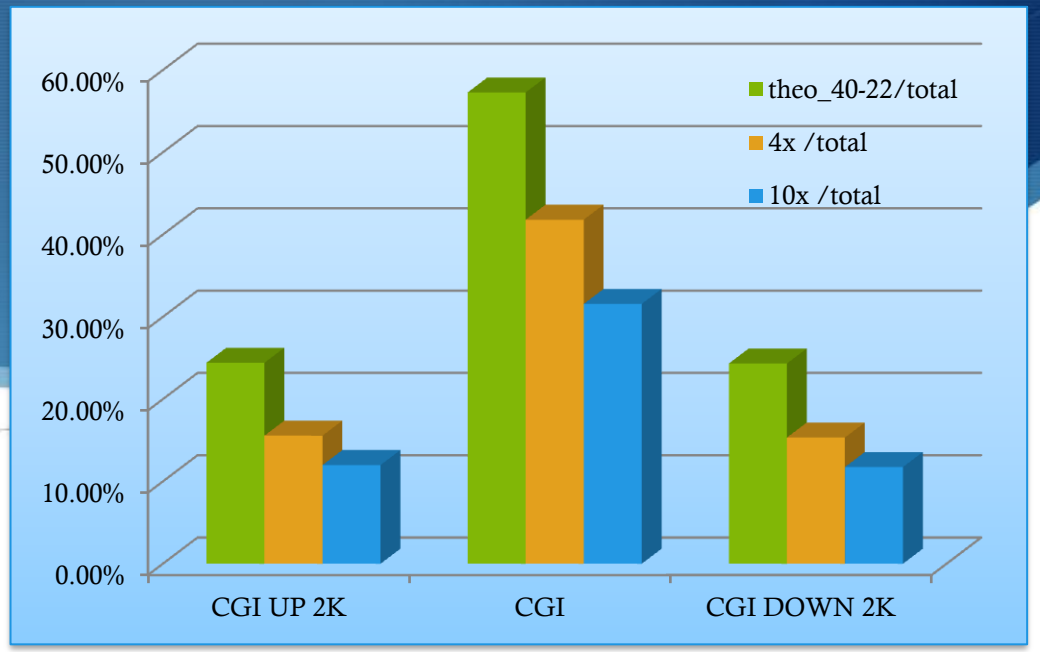


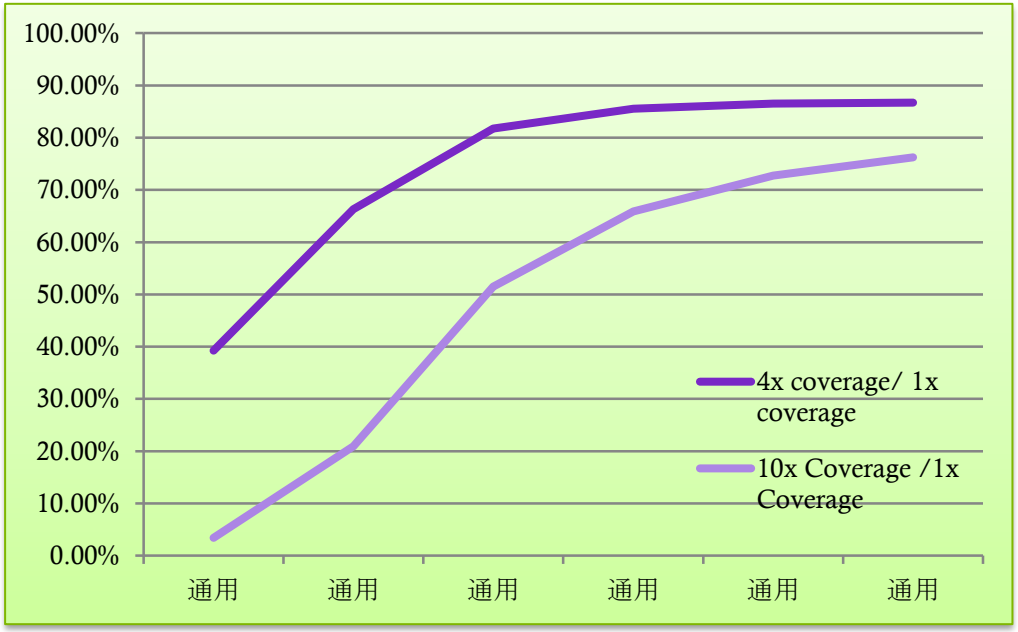


RRBS (Reduced Representative Bisulfite Sequencing)
Cost Effective Bisulfite Sequencing

- very limited genomic DNA, even if it is degraded (FFPE samples)
- Nearly 5M CpG sites have been covered, about 65-75% of them are covered by >=10x coverage depth.
- They are mainly distributed in CpG Island and gene upstream.
- The overlap between different human samples is 85-95%.

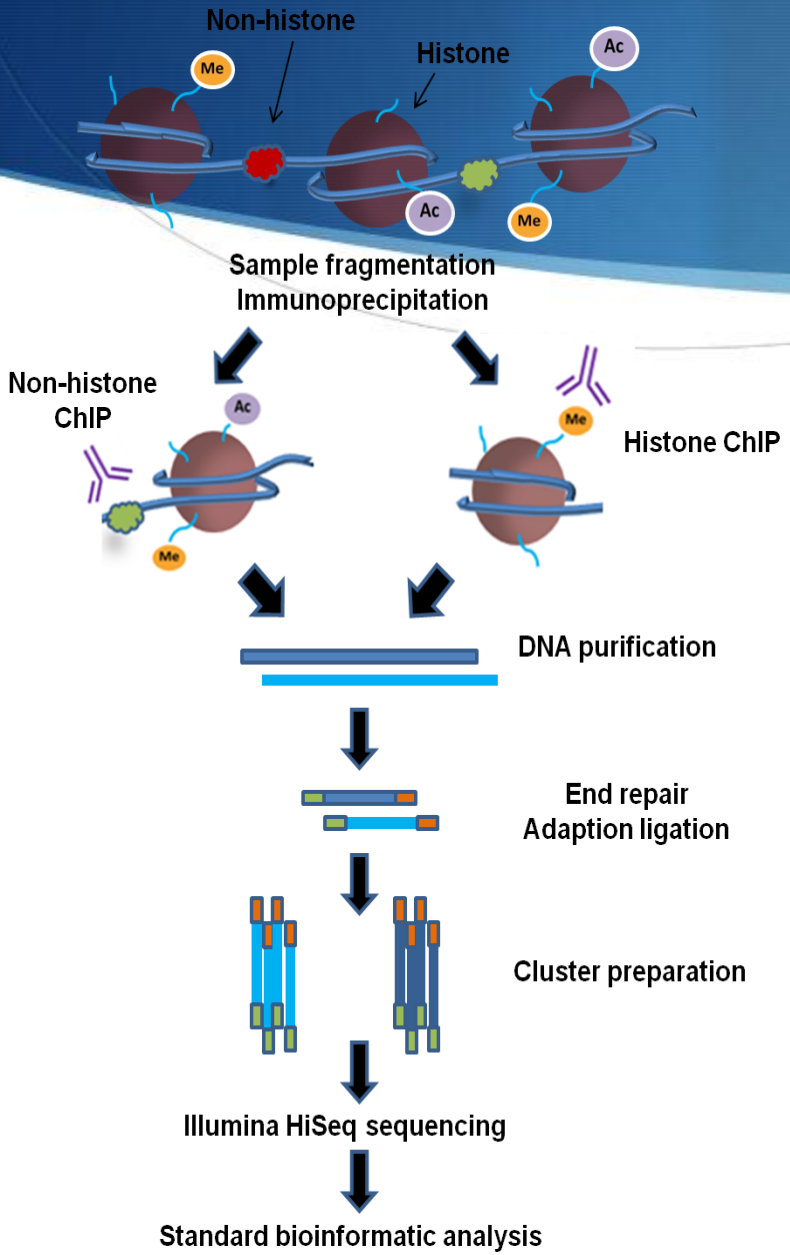
The distribution of CpG sites covered by RRBS data





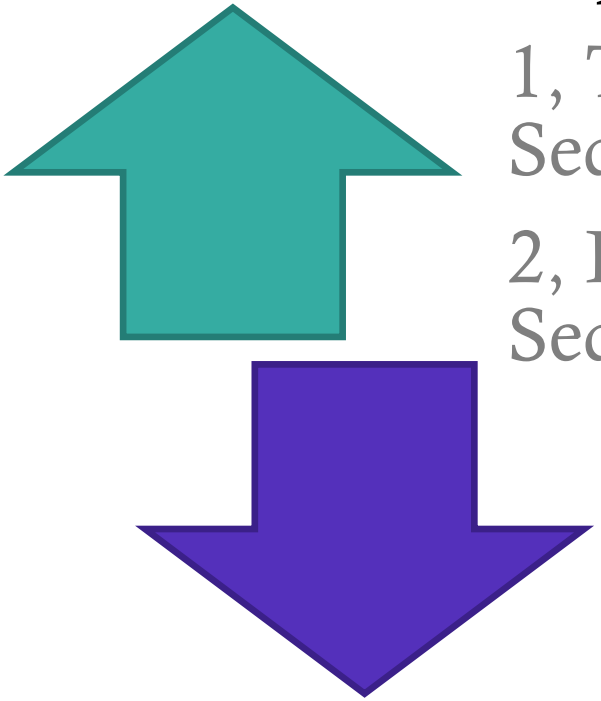
The percentage of CpG sites covered by <=4x depth and <=10x depth (in all the CpG sites covered by more than 1x depth)

DNA-Protein Interaction ChIP Sequencing



- whole genome wide
- Cost efficiency
- Start from limited amount of ChIPed DNA:
  5-10ug enrichment DNA product is available for library preparation

RNA Level



Expression level

1, Transcriptome Sequencing

2, RNA Quantitifcation Sequencing

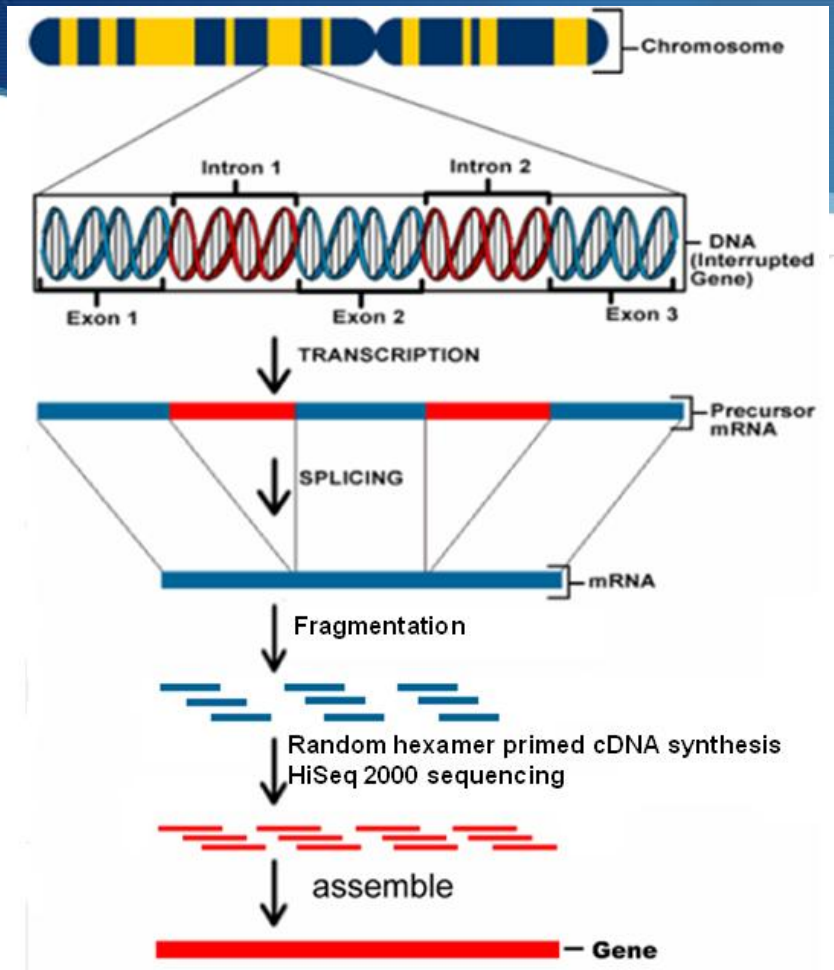
Non-coding RNA regulation:

Small RNA Sequencing

RNA Sequencing Technologies

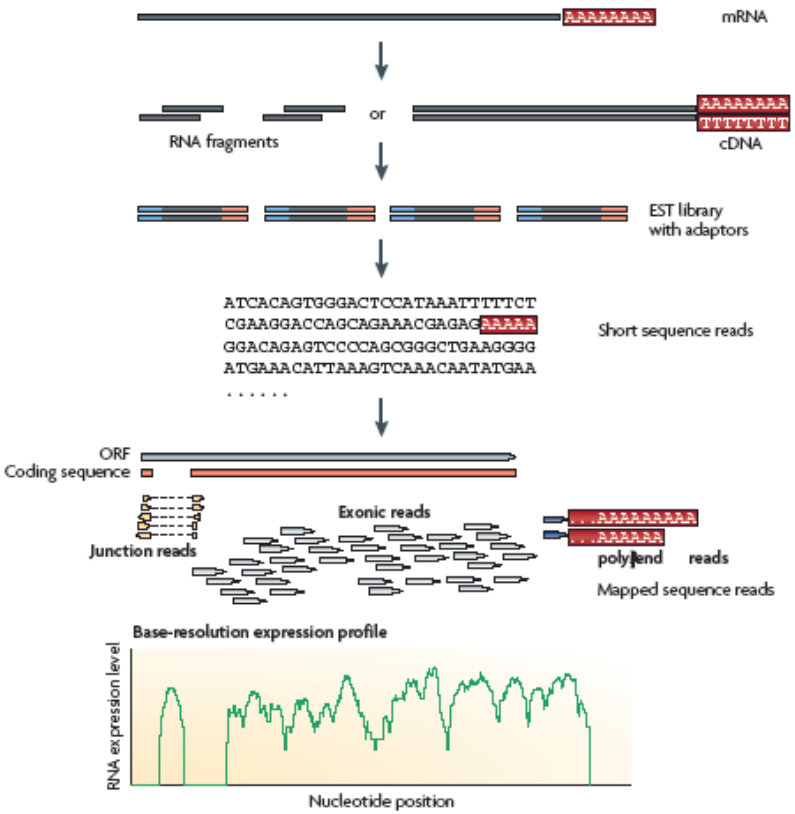
Sequencing Technologies	Transcriptome Sequencing	RNA-Seq	Small RNA Sequencing
Object	mRNA	mRNA	Small RNA
Identify new transcripts	O	X	O
Expression profiling research	O	O	O
Screen drug targets and biomarkers	O	O	O
Analyze gene structure	O	X	O
Genome-wide EST Sequencing	O	X	X
Detect gene fusion	O	X	X

Transcriptome Sequencing



De novo Transcriptome assembly

Transcriptome Resequencing

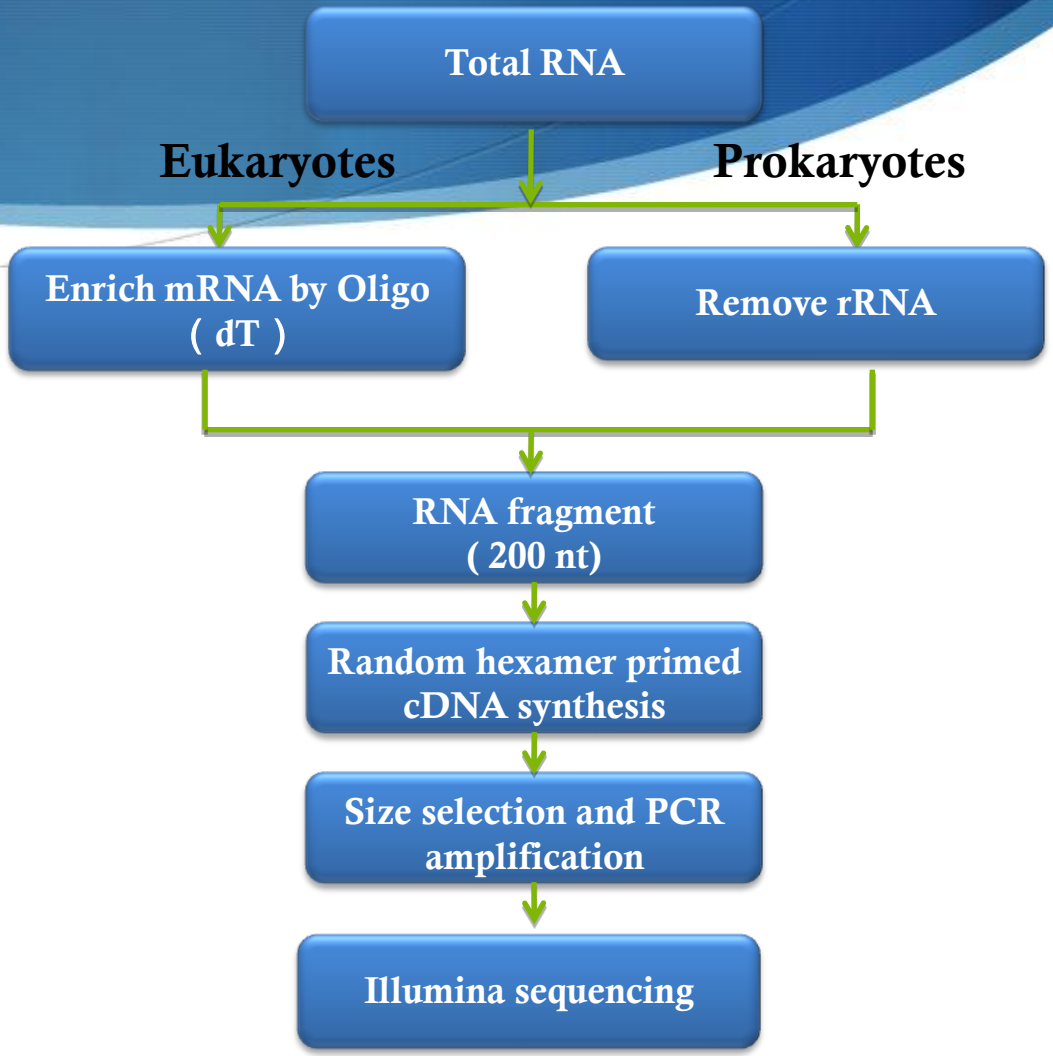


Transcriptome Sequencing

- Differentially expressed genes
- Alternative spliced transcripts
- Fusion genes

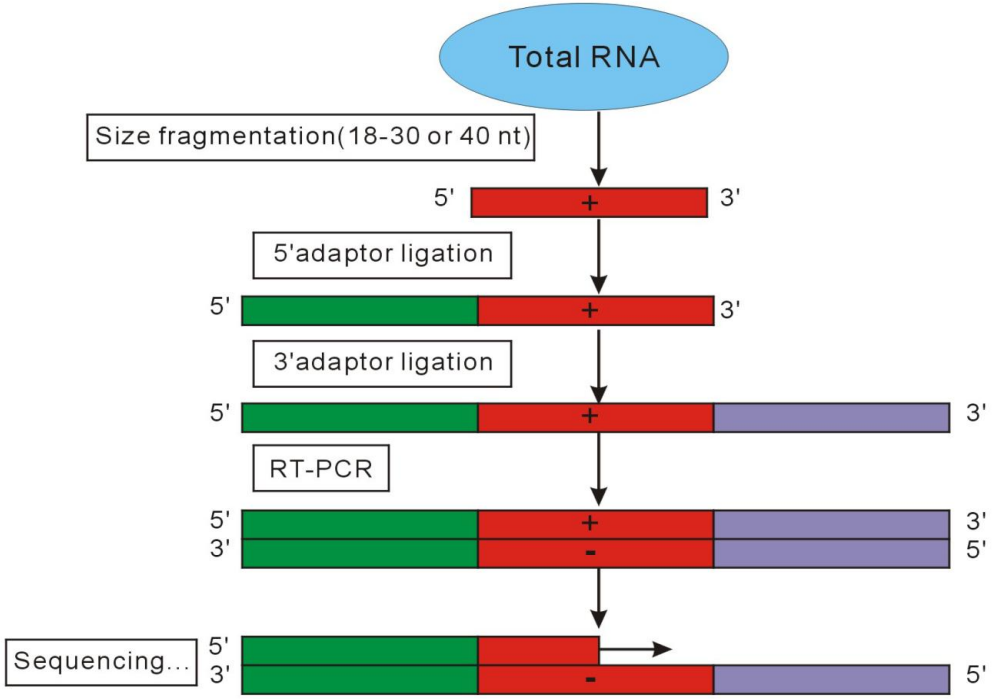
RNA Quantification Sequencing

- No noise, no cross-hybridization
- High repeatability and sensitivity
- 25% more genes than microarray
- Work well for rather high or low-abundance of genes
- A method for most of the species (modle/non-modle plants&animals)



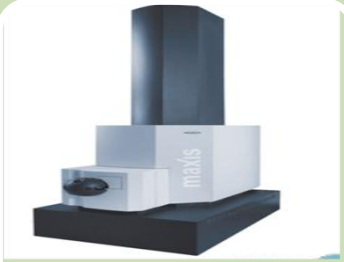
Small RNA Sequencing

- Screen differentially expressed small RNAs
- Identify small RNAs as target candidates



Proteome in BGI

Orbitrap velos, Thermo Scientific



maXis Q-TOF, Bruker

QTRAP 5500, AB SCIEX

UltrafleXtrem Bruker

Proteome profiling

modification proteomics

Quantitative proteomics

expression (different tissue/state/time)

Target proteomics

target protein expression

Protein ID

identification protein for low complexity sample

华大基因
BGI

Selected Cases

BGI: 1% - 10% - 100%

- **The International HapMap Consortium**
- Pro. Henry Huanming Yang: The Chinese HapMap Consortium

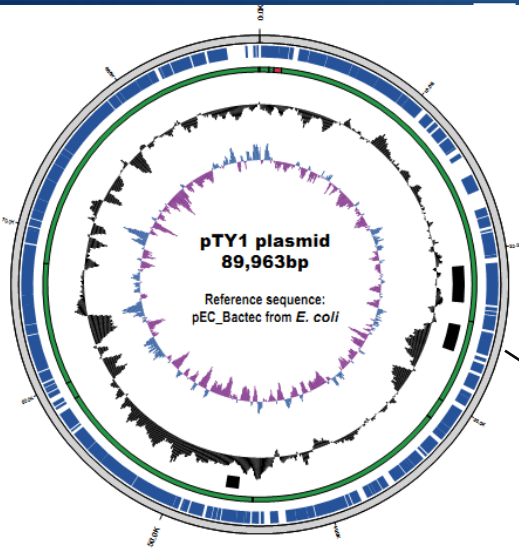
A Haplotype Map of the Human Genome. ***Nature***. 2005

A public database of common variation in the human genome: more than 1M SNPs in 269 samples from 4 populations.

Integrating common and rare genetic variation in diverse human populations. ***Nature***. 2010

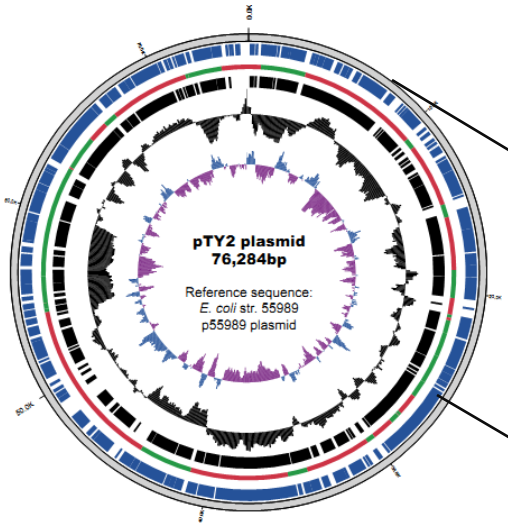
1.6M SNPs in 1,184 reference individuals from 11 global populations.

nished genome of this deadly E. coli



pTY1 plasmid (89,963bp)

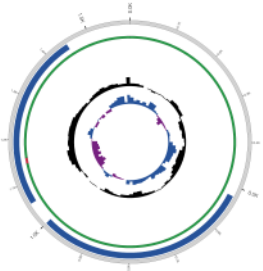
Two antibiotics resistant genes



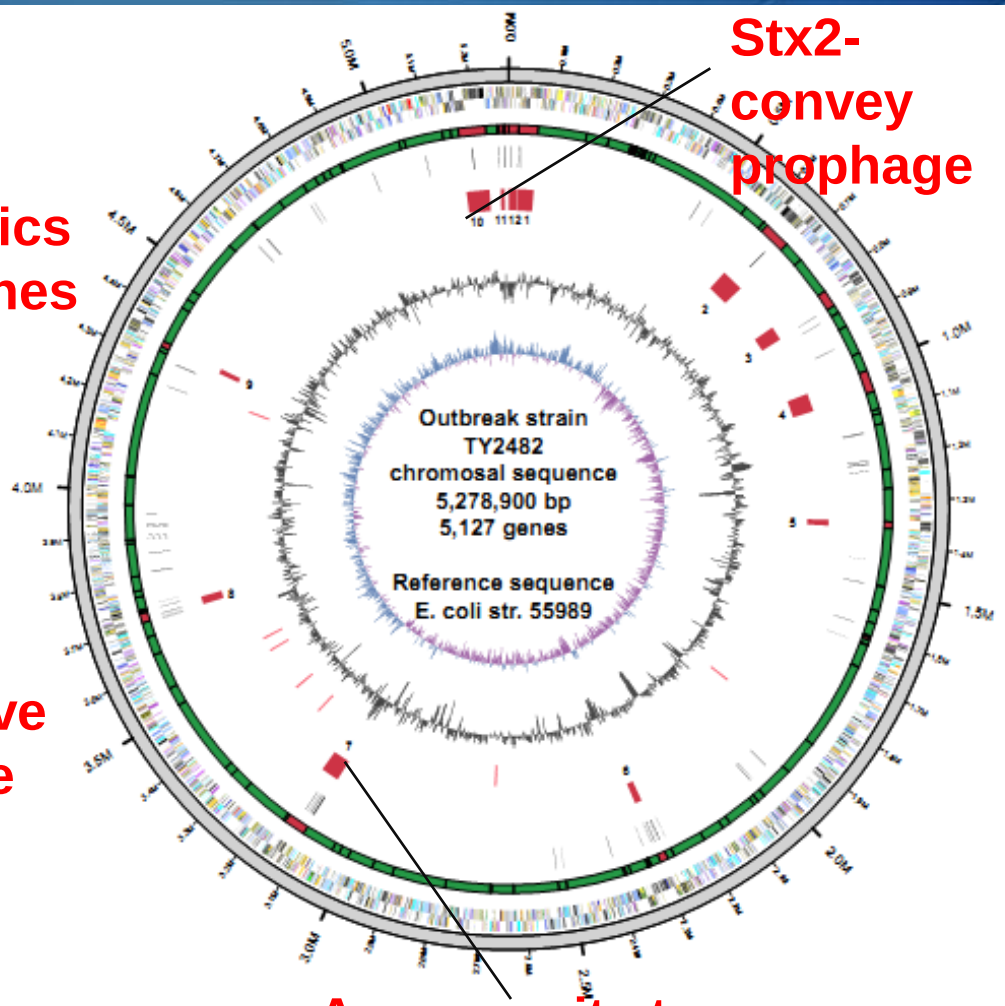
Sep A

pTY2 plasmid (76,284bp)

aggregative adherence fimbria I



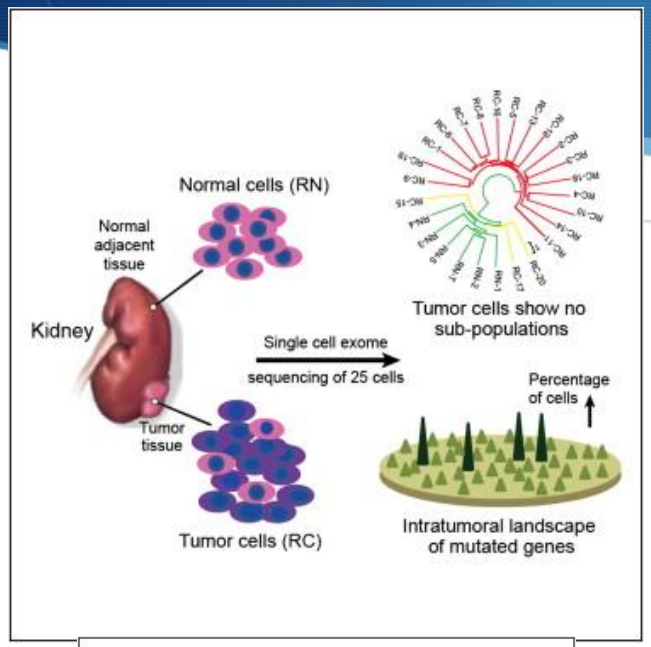
pTy3 plasmid

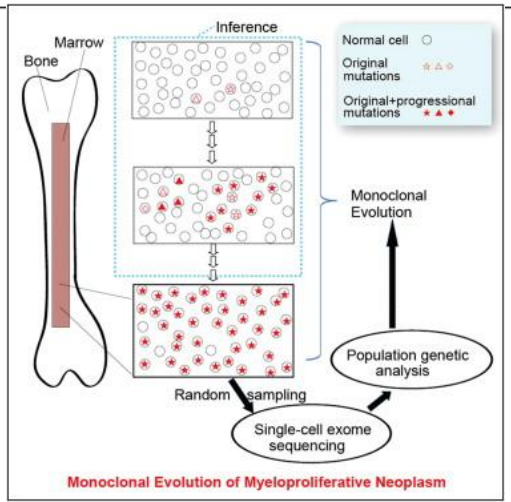


Stx2-convey prophage

A composite transposon harboring multi-resistant genes

Single cell sequencing on cancer research





Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor

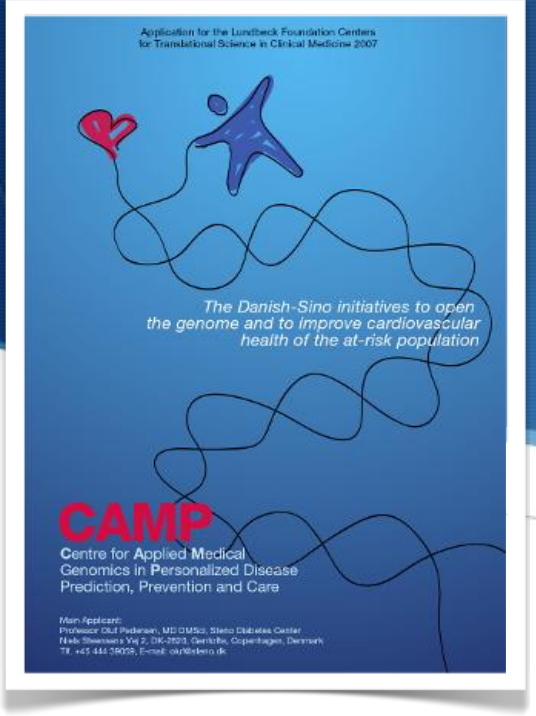
Cell 2012

- We present the genetic landscape of 25 single cells from a ccRCC patient
- No significant subpopulation of tumor cells could be observed within this tumor
- Different types of genetic lesion occurred depending on frequency of mutation
- Recurrent genes in patient population do not predict mutations in an individual tumor

Single-Cell Exome Sequencing and Monoclonal Evolution of a*JAK2*-Negative Myeloproliferative Neoplasm

Cell 2012

LuCamp Project



Patients with the combined at-risk metabolic phenotypes of visceral obesity, type II diabetes and hypertension

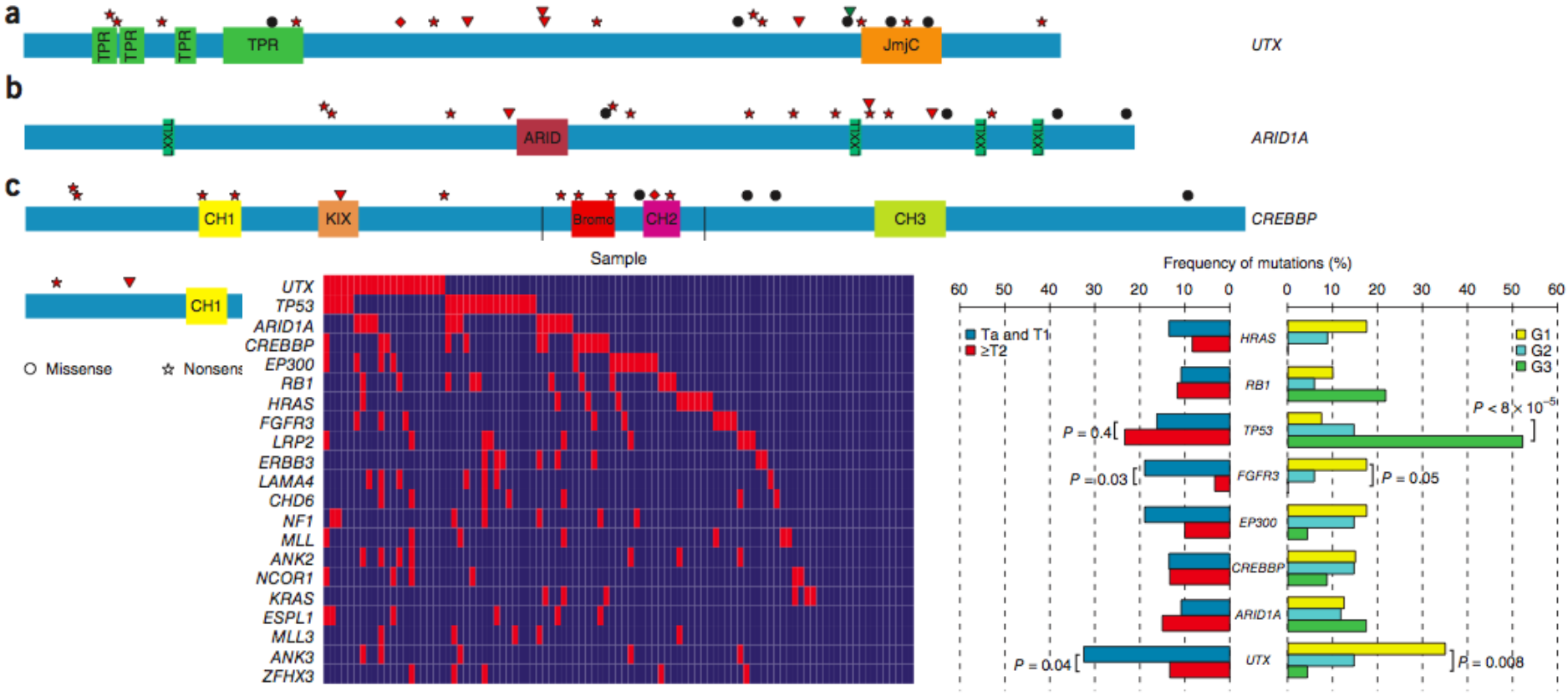
Exome Sequencing
1000 Case + 1000 Control
SNP Screening

Genome-wide associated genes and variations study

Re sequencing of 2000 human exomes identifies an excess of low-frequency non-synonymous coding variants. ***Nature Genetics***. 2011

Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. ***Nature Genetics***. 2011

Genetic aberrations of the chromatin remodeling genes (UTX, MLL-MLL3, CREBBP- EP300, NCOR1, ARID1A and CHD6) in 59% of our 97 subjects with TCC.



International Cancer Genome Consortium

BGI Involved in ICGC -- Gastric Cancer

Clinic & Pathology

Sequencing & Analysis

- Whole genome sequencing
- Whole exome sequencing

Complementary Studies

- RNA-Seq
- MeDIP-Seq
- Micro RNA

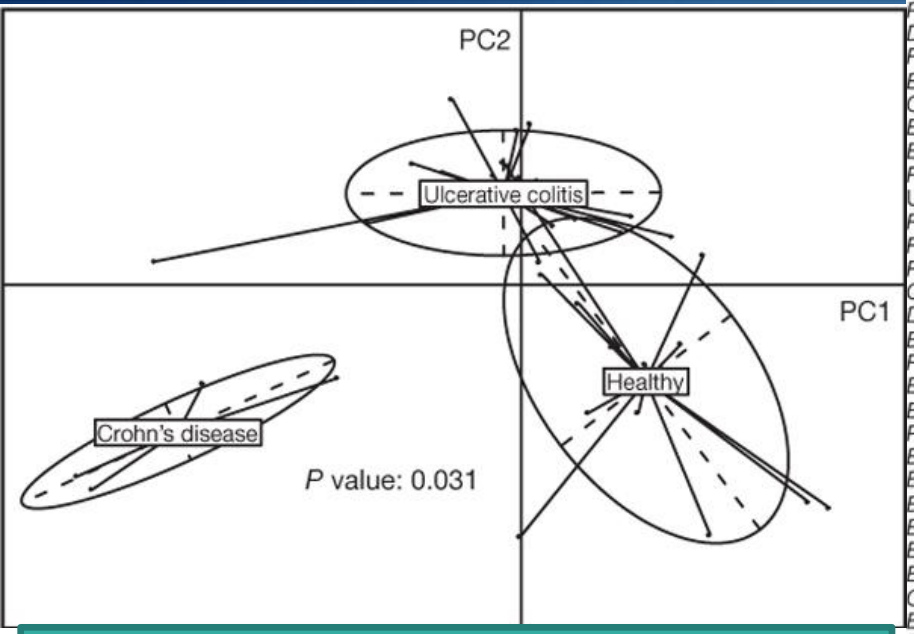
Data Storage, Analysis & Management

Human Metagenomics Research

A human gut microbial gene catalogue established by metagenomic sequencing. ***Nauture***. 2010

Define and describe the gut metagenome and the gut bacterial genome based on faecal samples of 124 European individuals

MetaHIT

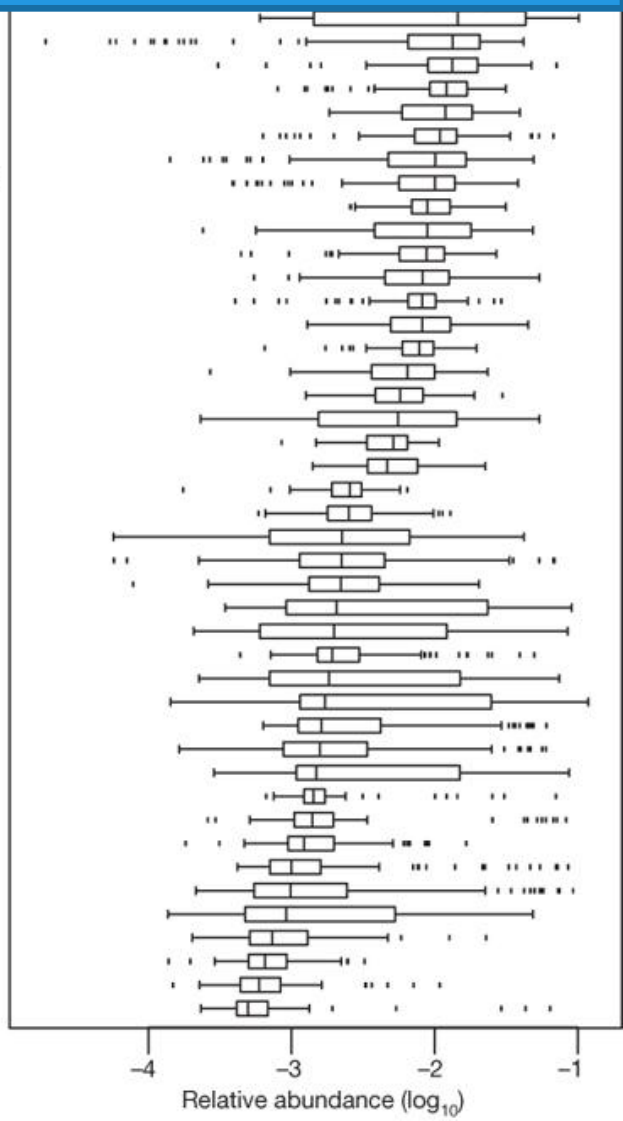


Bacterial species abundance differentiates IBD patients and healthy

- 576.7Gb data production
- 3.3Mb non-redundant microbial genes
- The gene set, ~150 times larger than the human gene complement

Bacteroides uniformis
Alistipes putredinis
Parabacteroides merda
Dorea longicatena
Ruminococcus bromii
Bacteroides caccae
Clostridium sp. SS2–1
Bacteroides thetaiotaom
Eubacterium hallii
Ruminococcus torques
Unknown sp. SS3 4
Ruminococcus sp. SR1
Faecalibacterium praus
Ruminococcus lactaris
Collinsella aerofaciens
Dorea formicigenerans
Bacteroides vulgatus ATCC 8482
Roseburia intestinalis M50 1
Bacteroides sp. 2_1_7
Eubacterium siraeum 70 3
Parabacteroides distasonis ATCC 8503
Bacteroides sp. 9_1_42FAA
Bacteroides ovatus
Bacteroides sp. 4_3_47FAA
Bacteroides sp. 2_2_4
Eubacterium rectale M104 1
Bacteriodes xylanisolvens XB1A
Coprococcus comes SL7 1
Bacteroides sp. D1
Bacteroides sp. D4
Eubacterium ventriosum
Bacteroides dorei
Ruminococcus obeum A2–162
Subdoligranulum variabile
Bacteroides capillosus
Streptococcus thermophilus LMD–9
Clostridium leptum
Holdemania filiformis
Bacteroides stercoris
Coprococcus eutactus
Clostridium sp. M62 1
Bacteroides eggerthii
Butyrivibrio crossotus
Bacteroides finegoldii
Parabacteroides johnsonii
Clostridium sp. L2–50
Clostridium nexile
Bacteroides pectinophilus
Anaerotruncus colihominis
Ruminococcus gnavus
Bacteroides intestinalis
Bacteroides fragilis 3_1_12
Clostridium asparagiforme
Enterococcus faecalis TX0104
Clostridium scindens
Blautia hansenii

Relative abundance of 57 frequent microbial genomes among individuals of the cohort.



Largest ever Epigenetics Project

KING'S College LONDON

华大基因 BGI

Purpose and Objectives：

5000 twins

↓

MeDIP-Seq

↓

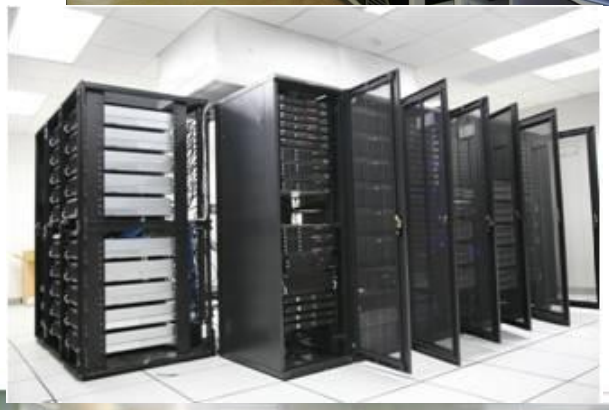
Finding genes related methylation differentially that cause of disease


华大基因
BGI

BGI (formerly known as Beijing Genomics Institute) was founded in Beijing on Sept 9th, 1999, and has undergone rapid development and expansion with the establishment of new branches in China, Europe and USA. Nowadays, BGI has become the biggest genomic research institute in Asia and one of the world's top three genomic research institutes.

New-gen sequencing

- 137 Illumina Hi-Seq 2000
- 27 Life Tech SOLiD 4
- Supercomputer: 102 T flops, 10 PB storage



- Applications:

1. Genome sequencing
2. RNA sequencing
3. Epigenome and ChIP-seq
4. Metagenome

Thank you!

Patrick Tao Li
litao@genomics.cn
Tel: +08 4406 7759
www.genomics.cn