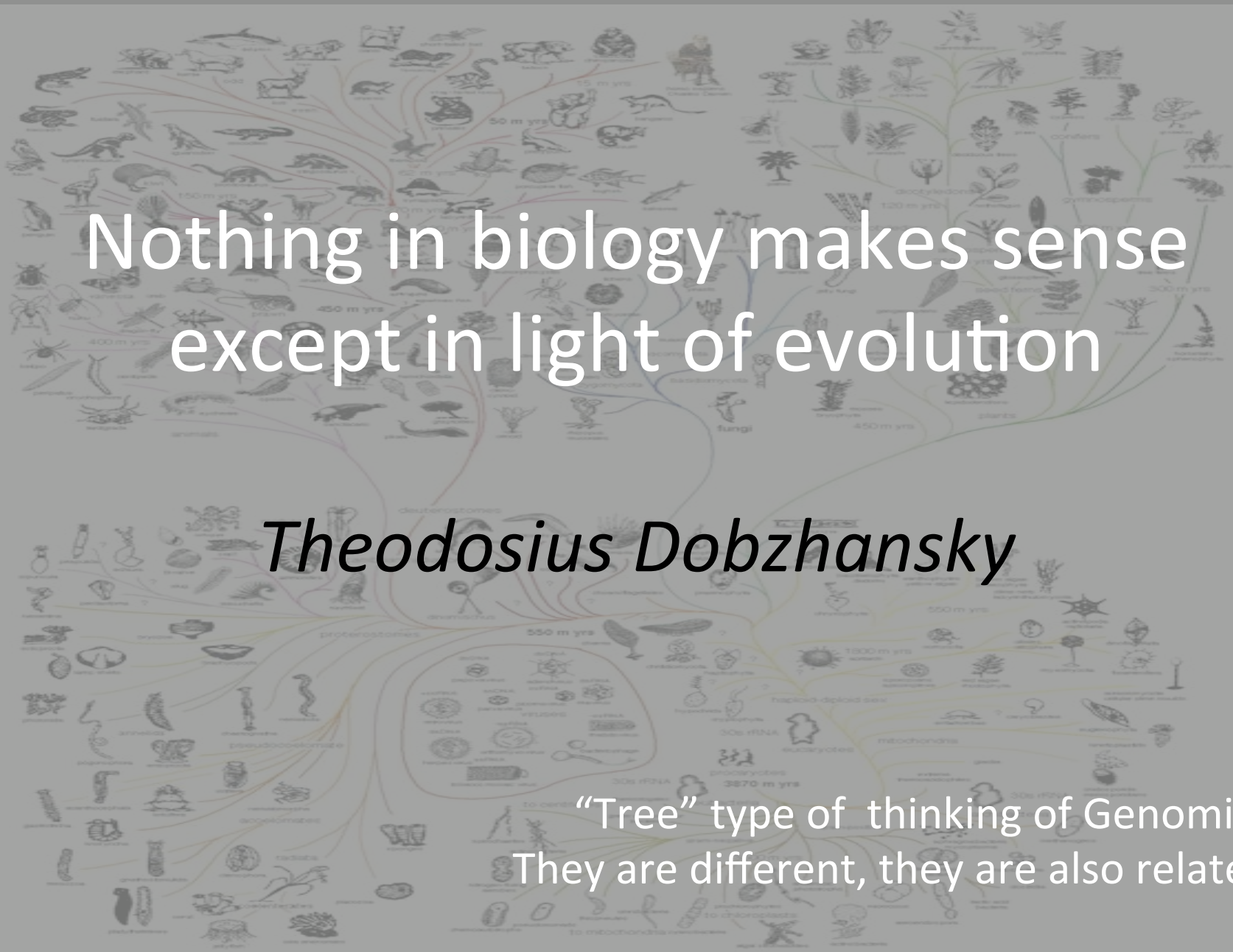
Resolving cancer heterogeneity by single cell sequencing

Xu Xun

Deputy Director of BGI

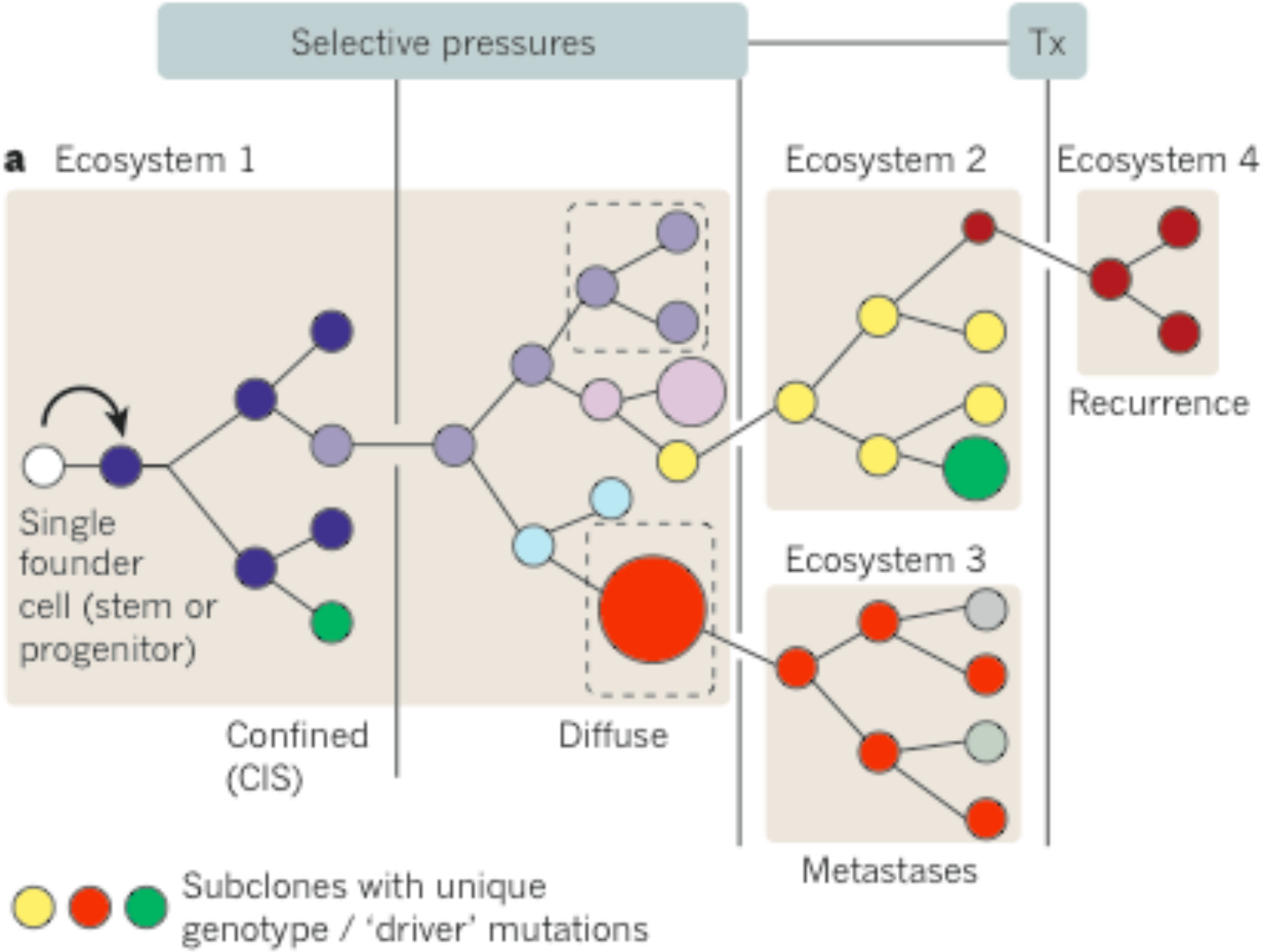


Nothing in biology makes sense except in light of evolution

Theodosius Dobzhansky

"Tree" type of thinking of Genomics
They are different, they are also related

Cancer is a game of cell evolution

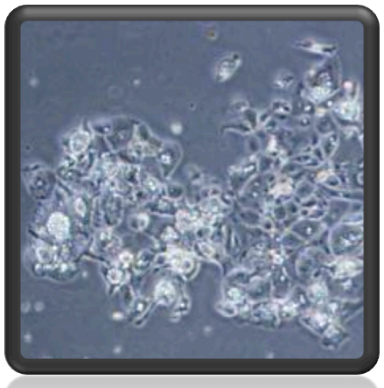


Mel Greaves & Carlo C. Maley Colonal evolution in cancer (2012) *Nature*

Single Cell Genomics Analysis

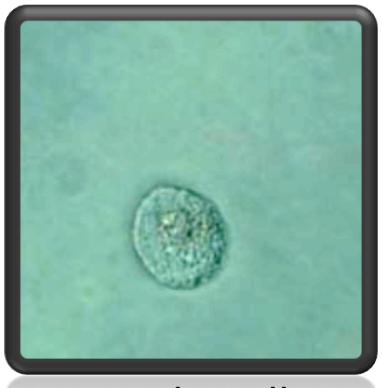
Single Cell Genomics

Sequencing the Single Cell Genome by Next Generation Sequencing (NGS)



Digested Tissues

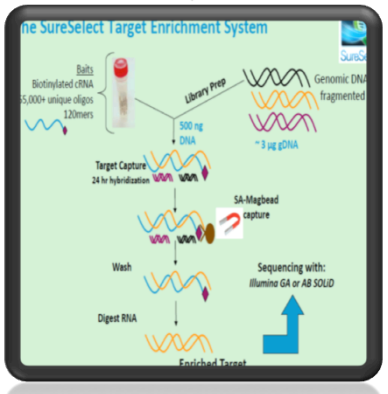
Cell isolation



Single Cell

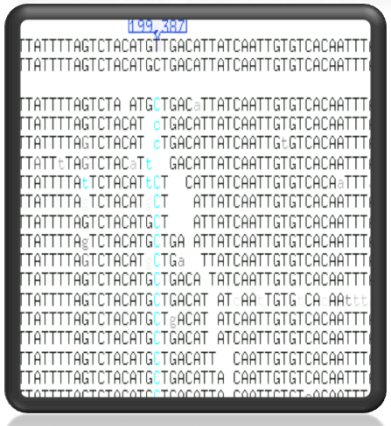
WGA (MDA)

Enough Amount of DNA



Library Preparation

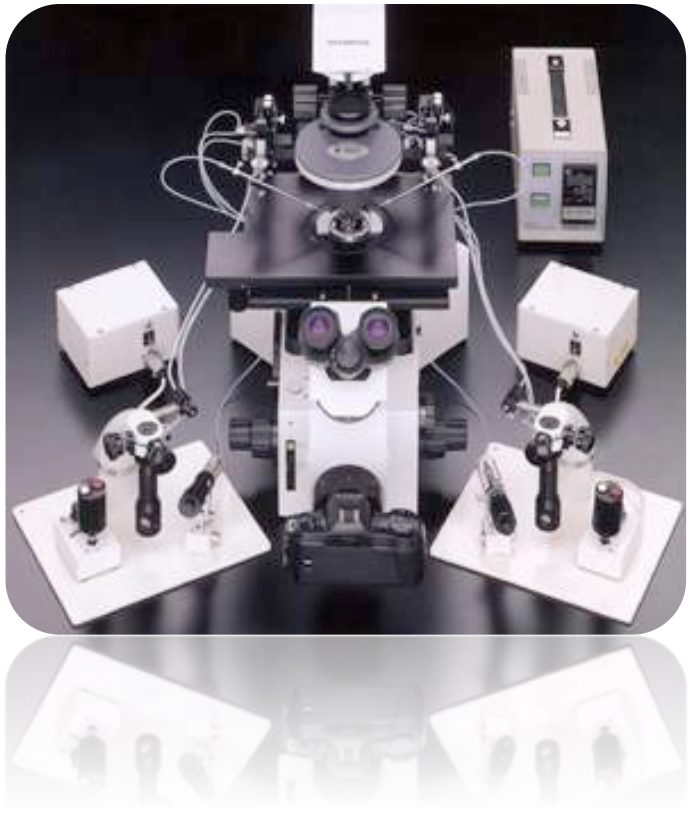
Hiseq 2000 Sequencing



Bioinformatics Analysis

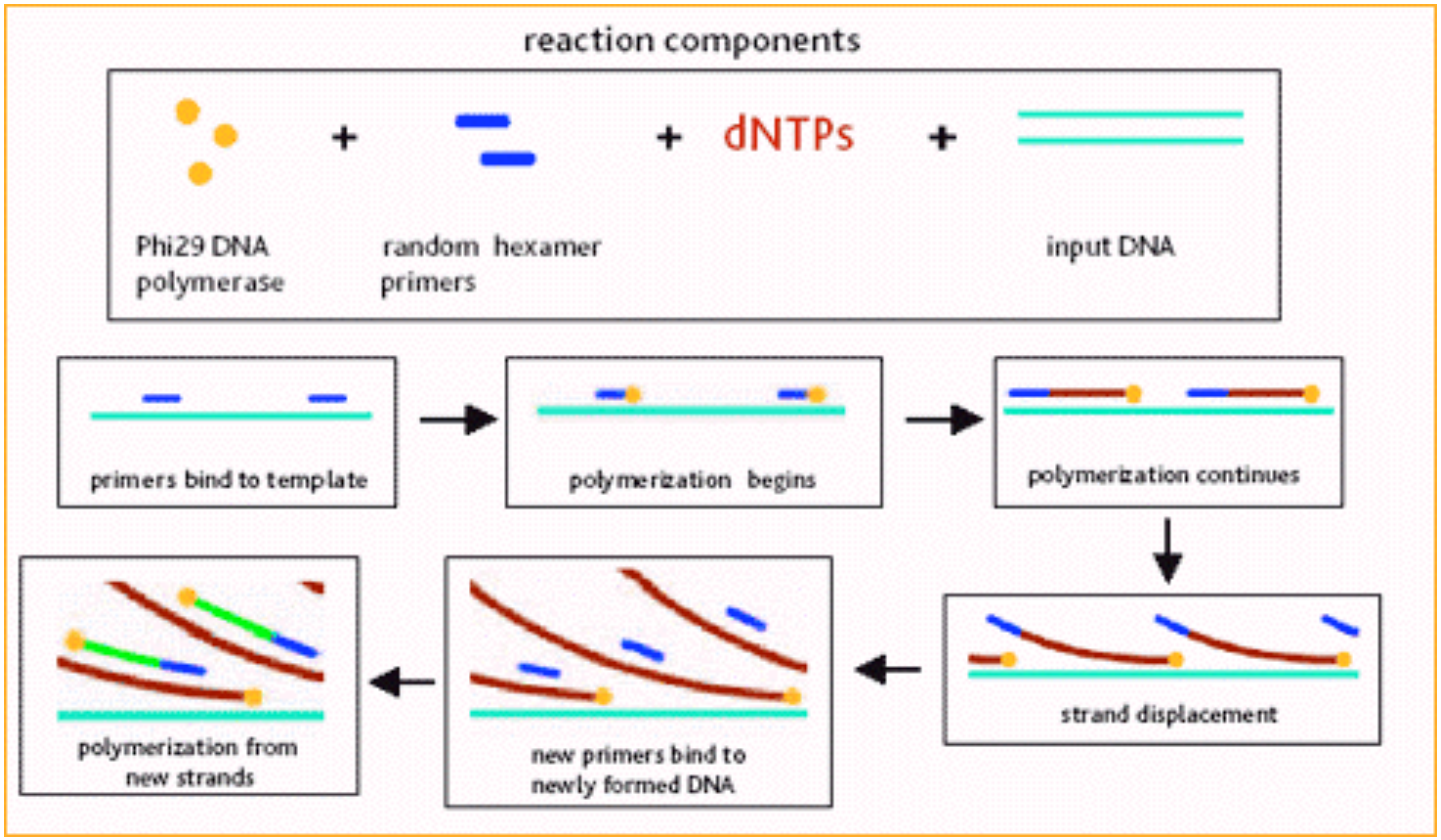
Single Cell Genomics

Single Cell Isolation



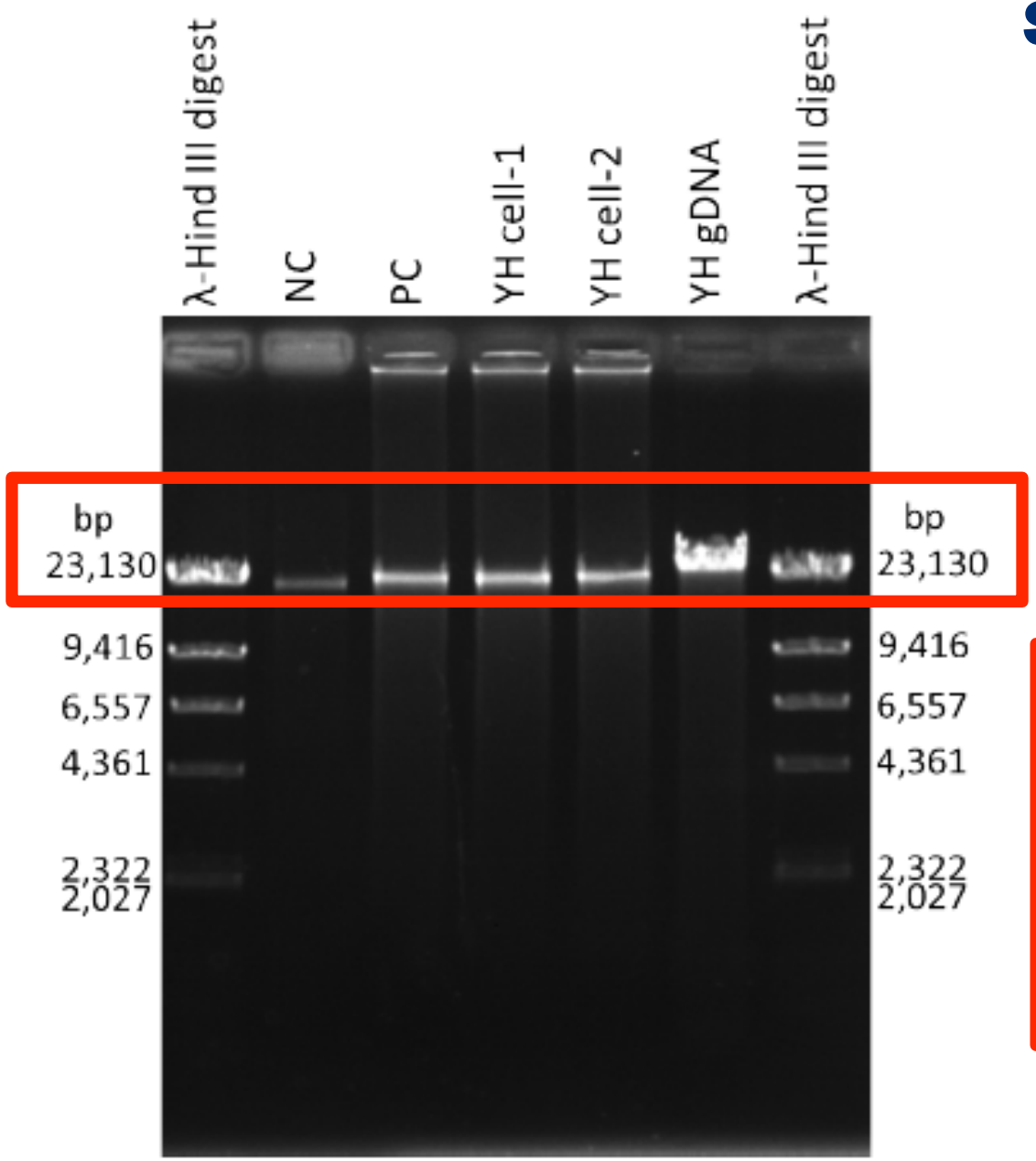
- Digest the tissue and randomly select the single cells by the inverted microscope and microcapillary pipetting.

Single Cell Genomics

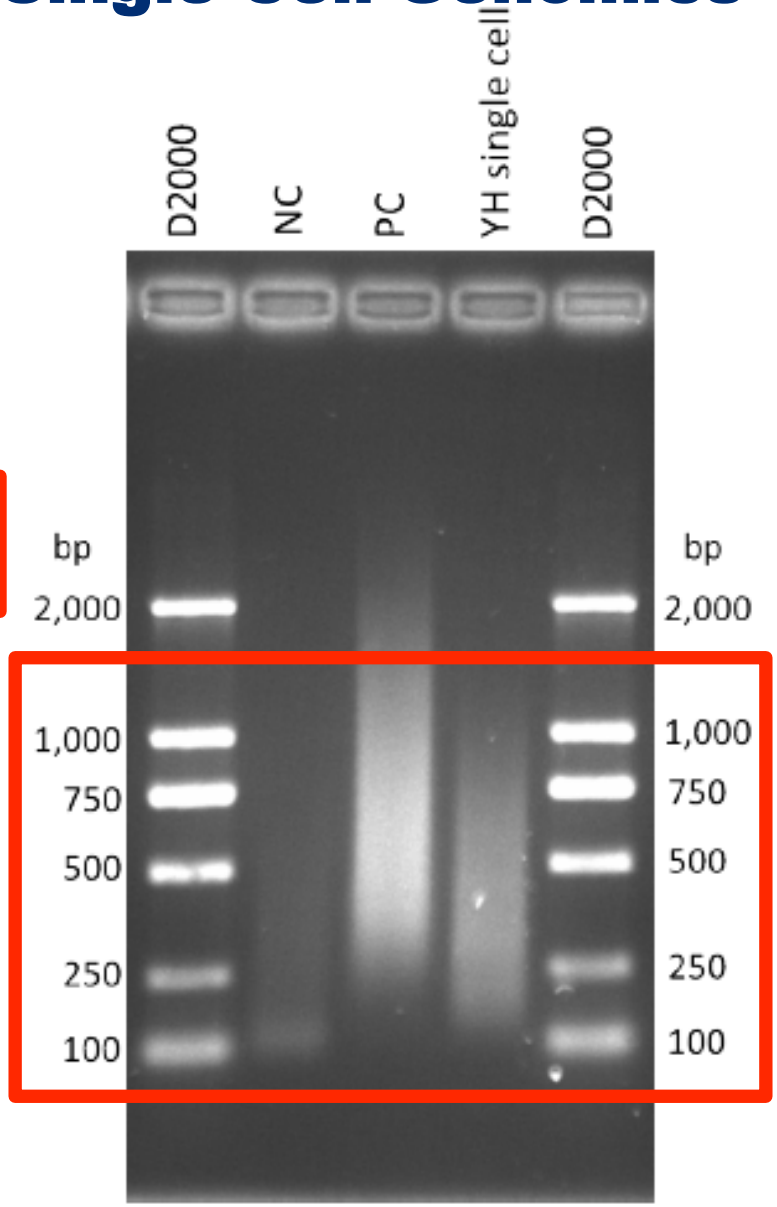


Whole-genome amplification (WGA) based on multiple-displacement amplification with the phi29 enzyme

Single Cell Genomics



Our method



PCR based method

Single Cell Genomics

Method Evaluation

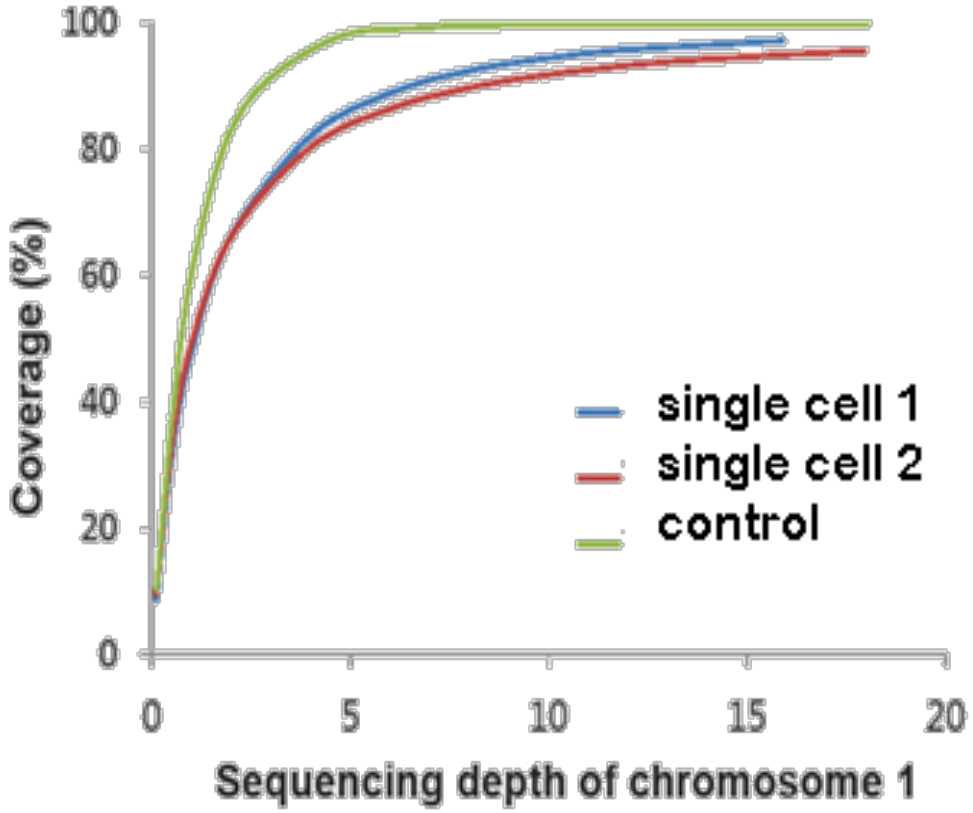
- Sample set: single cell from the first Asian genome donor (YH); and control form the same tissue.
- Data set : 13X and 18X for two replications

	Single cell 1	Single cell 2	Control
Raw data (Gb)	35.47	47.99	48.72
Average depth	13.32	17.82	18.03
Genome coverage (%)	**95.77**	**94.46**	99.91

Hou et al. *Cell* 2012

Single Cell Genomics

Method Evaluation

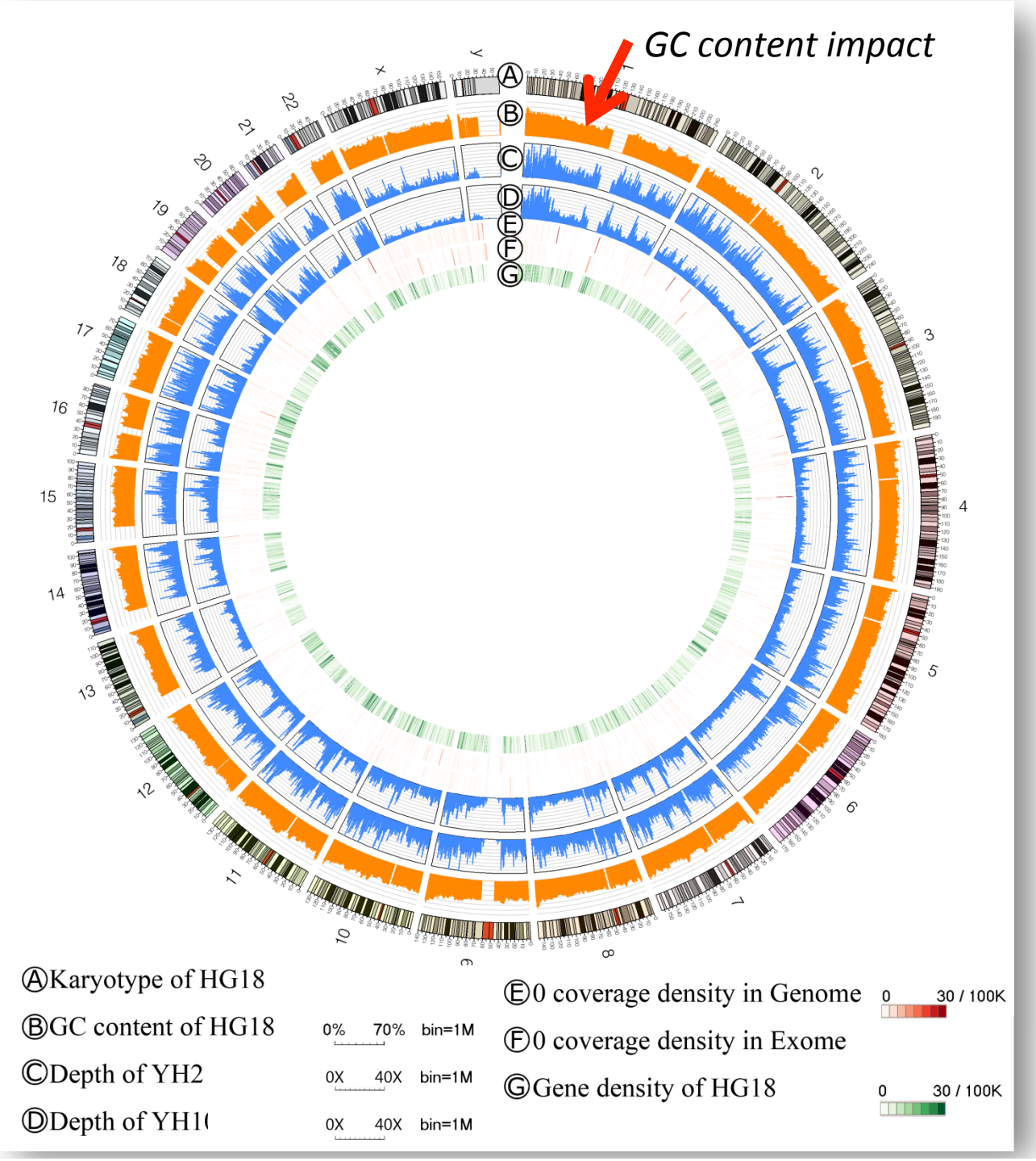


- *No obvious genome wide coverage limitation by single cell sequencing*

Single Cell Genomics

Method Evaluation

- *No obvious genome wide coverage limitation; GC content does impact the even distribution of WGA data.*



Single Cell Genomics

WGA Artifacts Rate Estimate

(Calculated by comparing consensus sequence between YH single cell and YH million cells data)

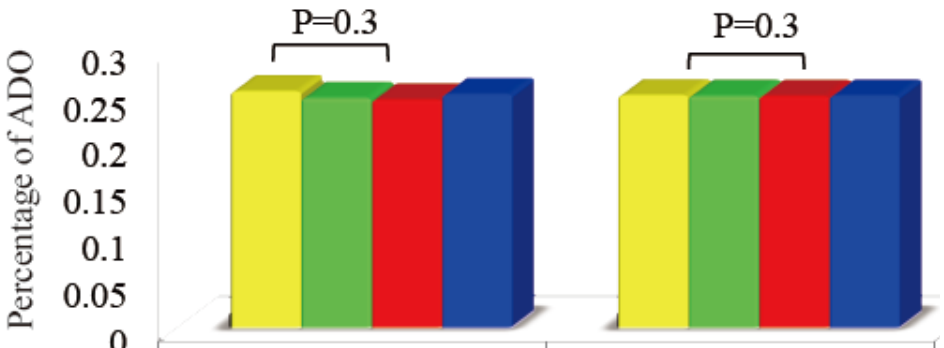
Sample ID	FDR (False positive mutation) rate
Single cell 1	7.2E-5
Single cell 2	8.9E-5

Note:

1. *FDR (False discovery rate) = Error SNP # in single cell/confident homo. SNP # in control;*
2. *Here FDR contains WGA error, sequencing error, and mapping error;*
3. *WGA error： E-5～E-6 (J. Guillermo Paez, et al. Nucleic Acids Research, 2004, Vol. 32, No. 9 e71）*

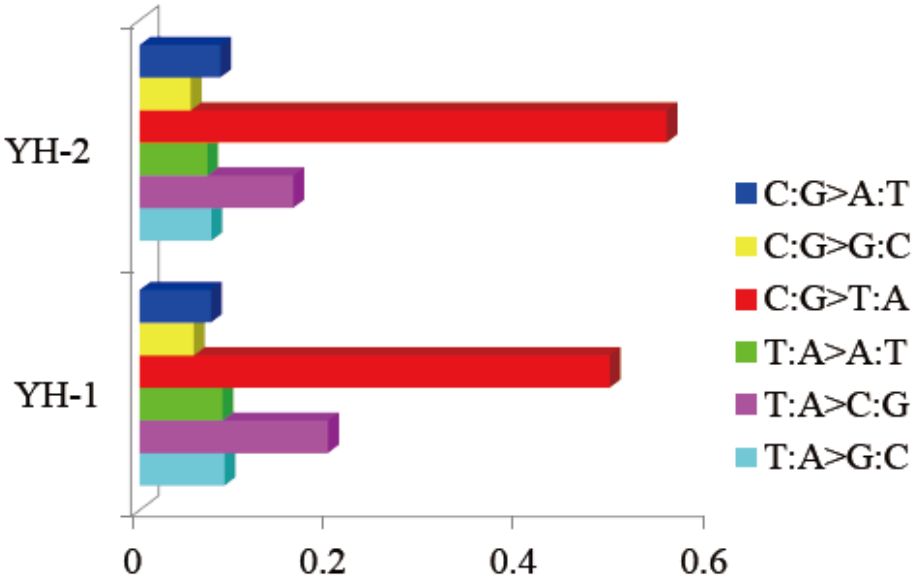
Single Cell Genomics

ADO (FN)



	YH-1	YH-2
A	0.254824616	0.250481154
C	0.247814255	0.249940295
G	0.245792036	0.249641768
T	0.251569092	0.249936783

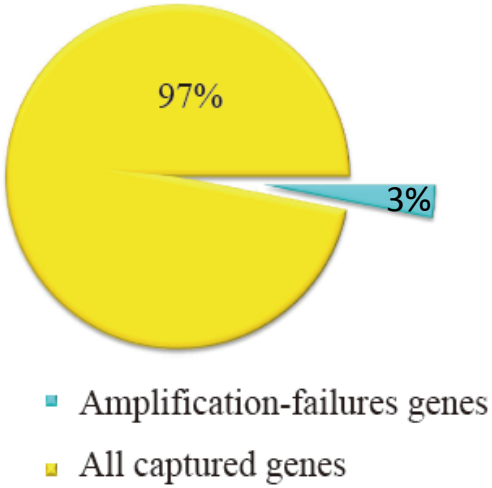
Amplification & Sequencing Error (FDR)

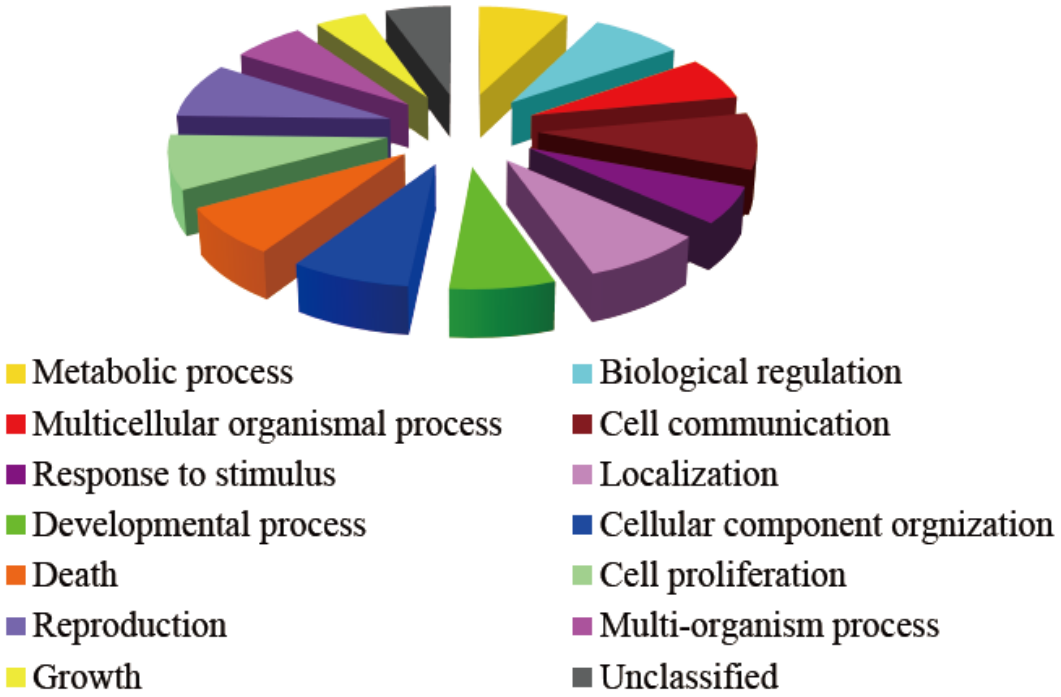


Fraction of false discovery sites

FN & FDR sites do not show specific base type bias beyond mutations

Single Cell Genomics





Pie Chart of Distribution of Biological Categories of Genes (GO) with Amplification Failure

Amplification failure genes do not show preference on different biological processes

Single Cell on Cancer Genomics

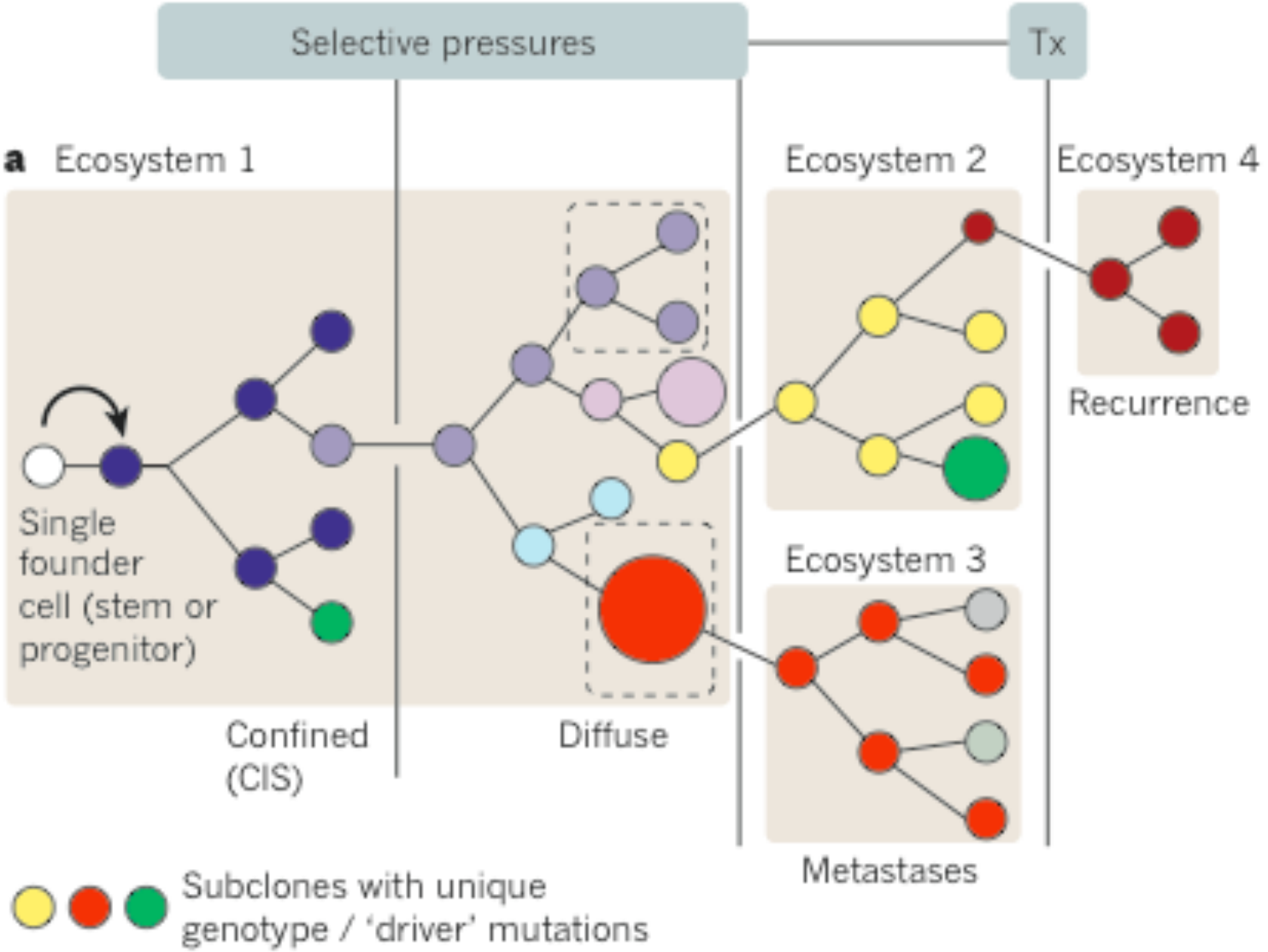
Blocks

- **Samples limitation**
- **Heterogeneity**
- **Cancer Progression**

Questions

- Where is the solutions for rare and rarity cancer samples?
- How can we differentiate such a mixed tumor tissues?
- What type of the genetic changes is relevant to cancer development?

Cancer is a game of cell evolution



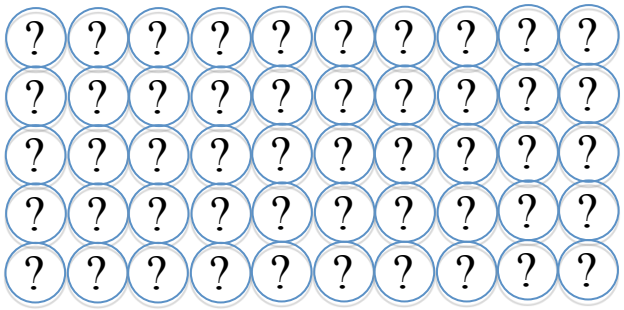
Mel Greaves & Carlo C. Maley Colonal evolution in cancer (2012) *Nature*

A Lesson From High Altitude Adaption

What is the genetic basis of difference in gifted ability to adapt for high altitude?

X. Yi, et.al. 2010. *Science* 329:75-8.

A Lesson From High Altitude Adaption



50 Tibetan Individuals

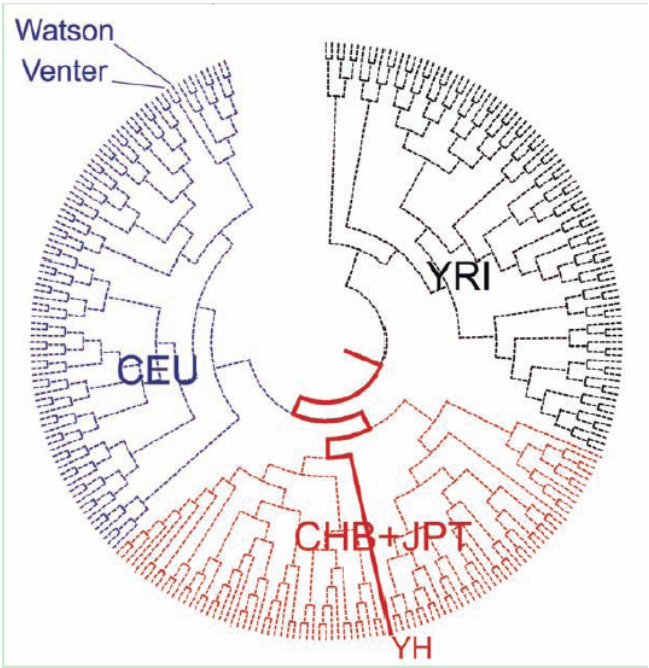
40 Han Chinese Individuals

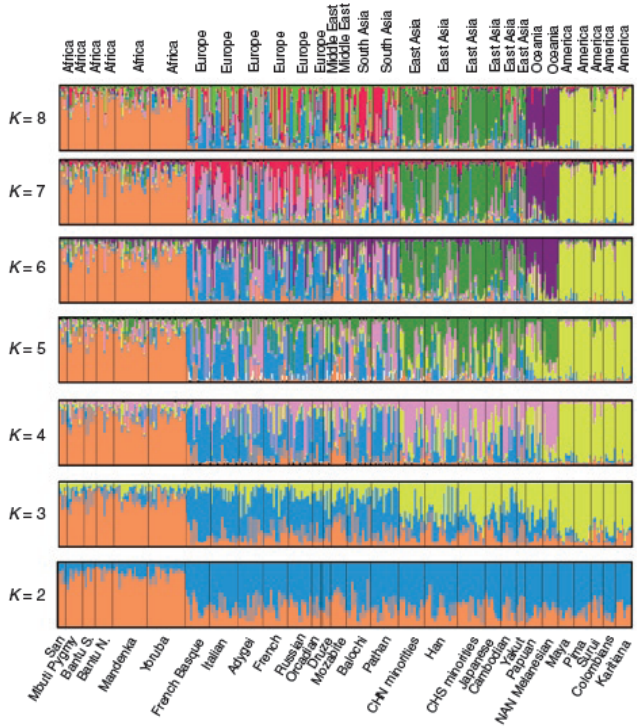
What is the genetic basis of difference in gifted ability to adapt for high altitude?

Xin Yi,et al Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude.
Science. **2010 July; 329(5987): 75-78**

A Lesson From High Altitude Adaption

- Building phylogenetic tree and structure analysis for different groups of people using SNP

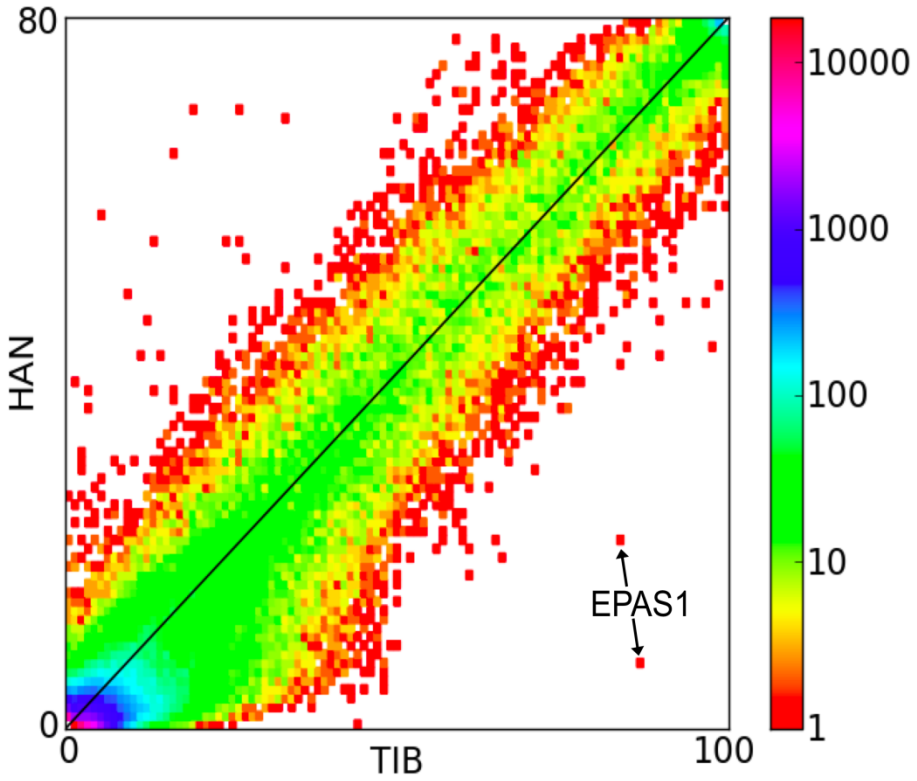


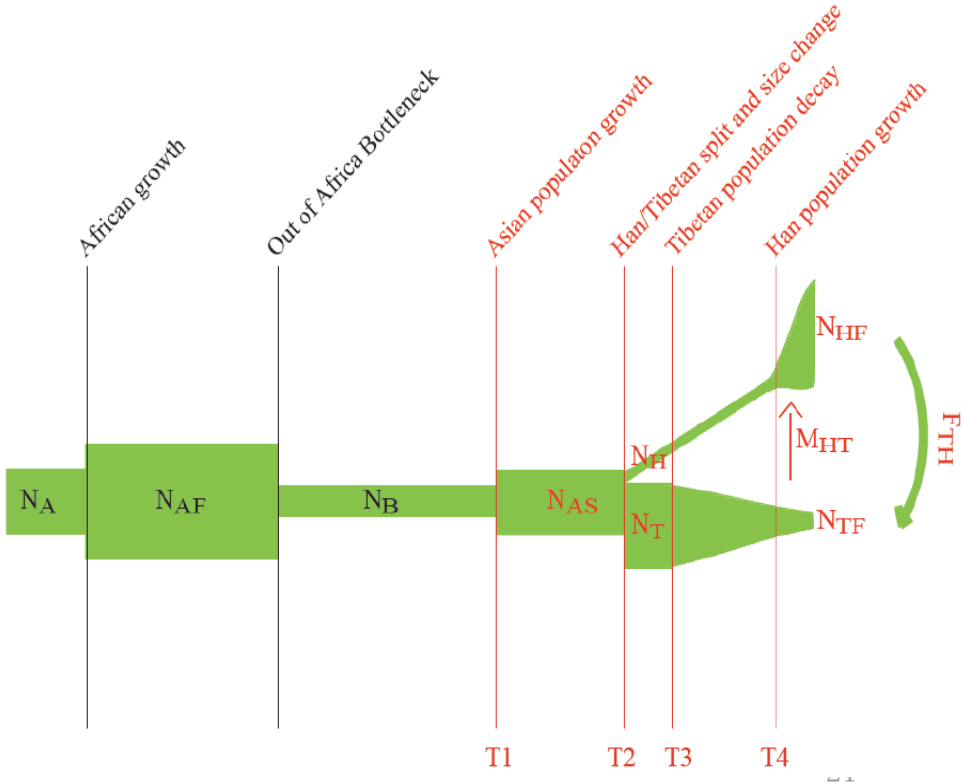


Jun Wang, *et al. Nature* 2008 Nov 6; 456(7218): 60-5. Li R, *et al. Nat Biotechnol.* 2009 Dec 7.

A Lesson From High Altitude Adaption

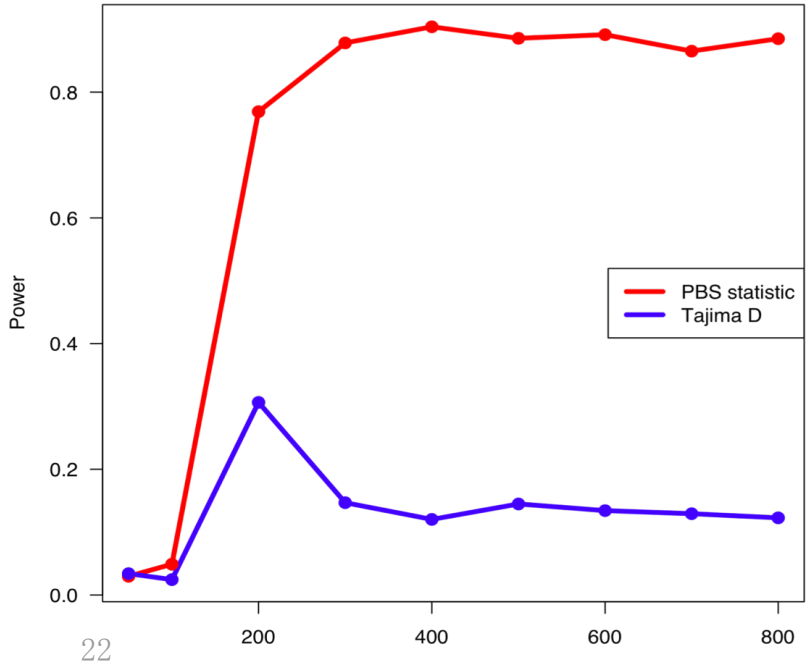
- SNP information shows that Tibetan and Han Chinese are genetically very similar and the evolution history was revealed as follow:

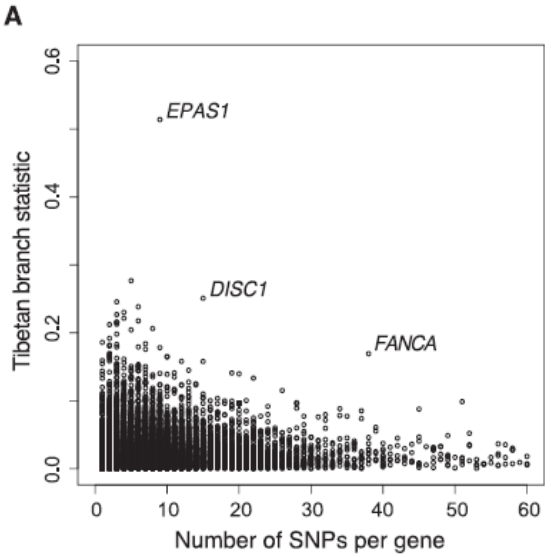


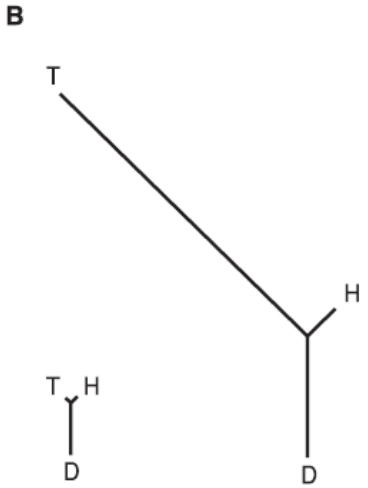


A Lesson From High Altitude Adaption

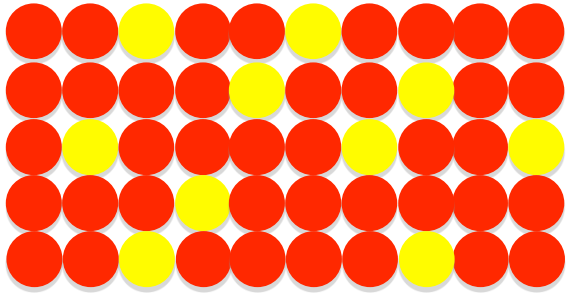
- Here are the power of Population Branch Statistic (PBS) and Genes with significant PBS selection signals



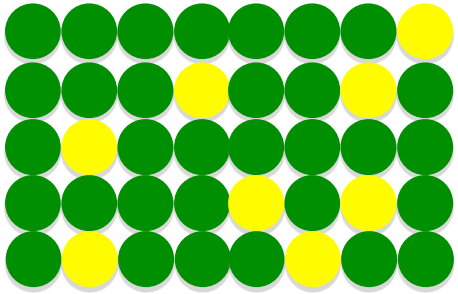




A Lesson From High Altitude Adaption



50 Tibetan Individuals



40 Han Chinese Individuals

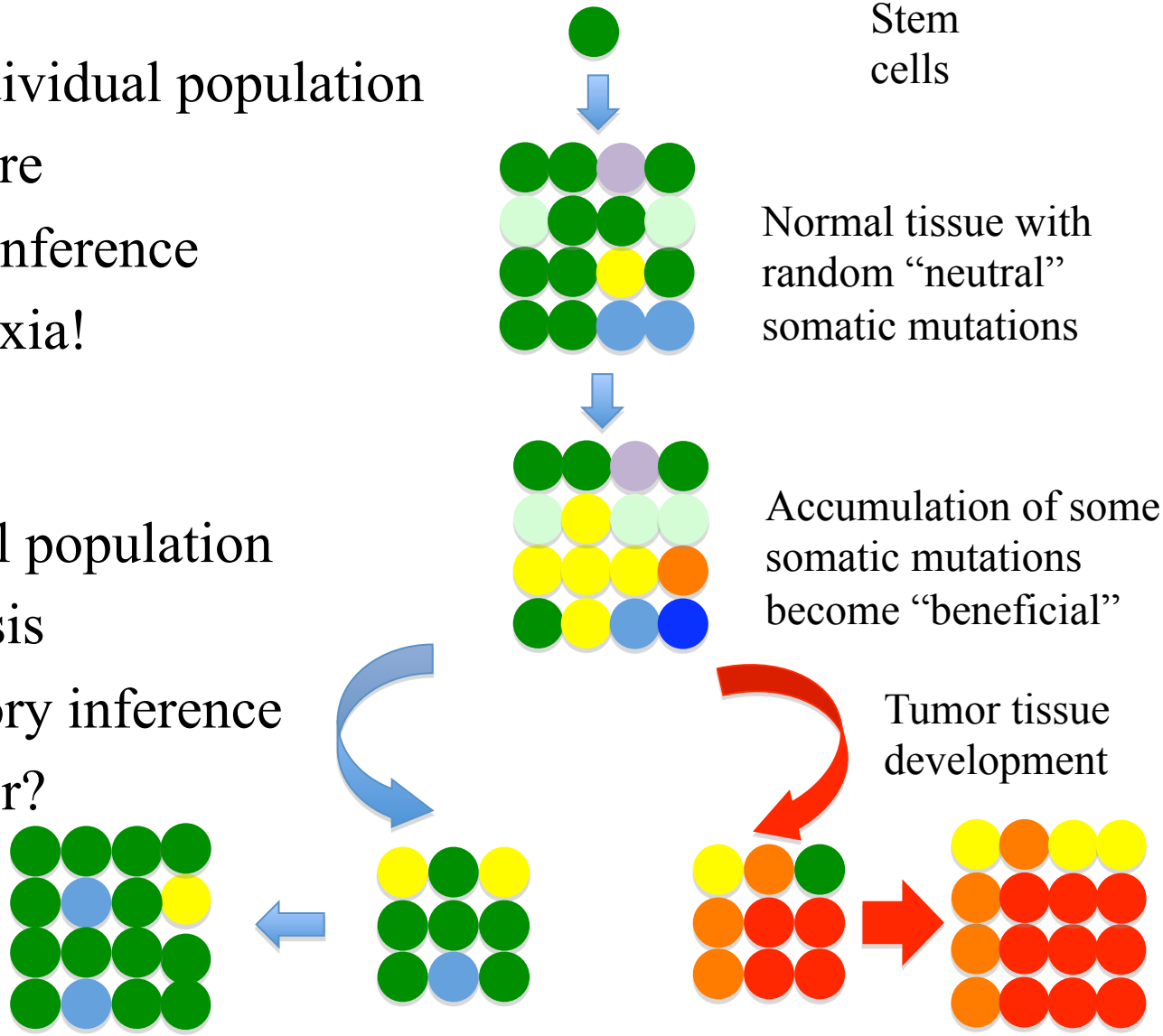
The gene (*EPAS1*) showing strongest selection signal (up to 80% frequency change in allele distribution)

Function further validated in
-Association with blood hemoglobin level
-Expression level difference in placenta

Apply Population Analysis to Cancer

- Heterogeneous individual population
- Phylogenic structure
- Evolution history inference
- Key genes to hypoxia!

- Heterogeneous cell population
- Cell lineage analysis
- Development history inference
- Key genes to tumor?



Four Cases from 1,000 Single Primary Tumor Cells Sequencing

Cancer type	Sample ID	Single cell number Cancer cell #/control cell #(sequencing available number)	Gender	Description
Essential thrombocytosis (ET)	ET	100/31(53/8)	M	**a *JAK2*-negative patient; published on *Cell, 2012***
Clear cell renal cell cancer (ccRCC-1)	CCRCC-1	20/6 (20/6)	M	**a VHL-wild type patient; large patient cohort also analyzed (Guo et al., 2011); published on *Cell* , 2012**
Bladder transitional cell cancer (BTCC)	BTCC	59/16 (47/11)	M	**a muscle invasive type patient; large patient cohort also analyzed (Gui et al., 2011); submitted on *Giga-Science***
Colon cancer	Colorectal	106/30 (64/6)	M	**large patient cohort also analyzed (unpublished); manuscript *in preparation***

Four Cases from 1,000 Single Primary Tumor Cells Sequencing

Data Sets

Sample name	Coverage(± SEM)	Depth(± SEM)	Cancer cell #/ normal cells #
ET	73.86%±5.08%	24.57±2.73	53/8
CCRCC-1	90.07%±1.93%	32.00±7.06	20/6
BTCC	85.17%±1.41%	40.23±2.21	47/11
Colorectal	78.27%±3.39%	15.65±1.05	64/6

Notes: All refer to target region of Agilent. We also sequenced the normal tissue (100X exome) or peripheral blood cells (30X exome), and cancer tissues (100X exome) to make quality control.

Four Cases from 1,000 Single Cancer Cells Sequencing

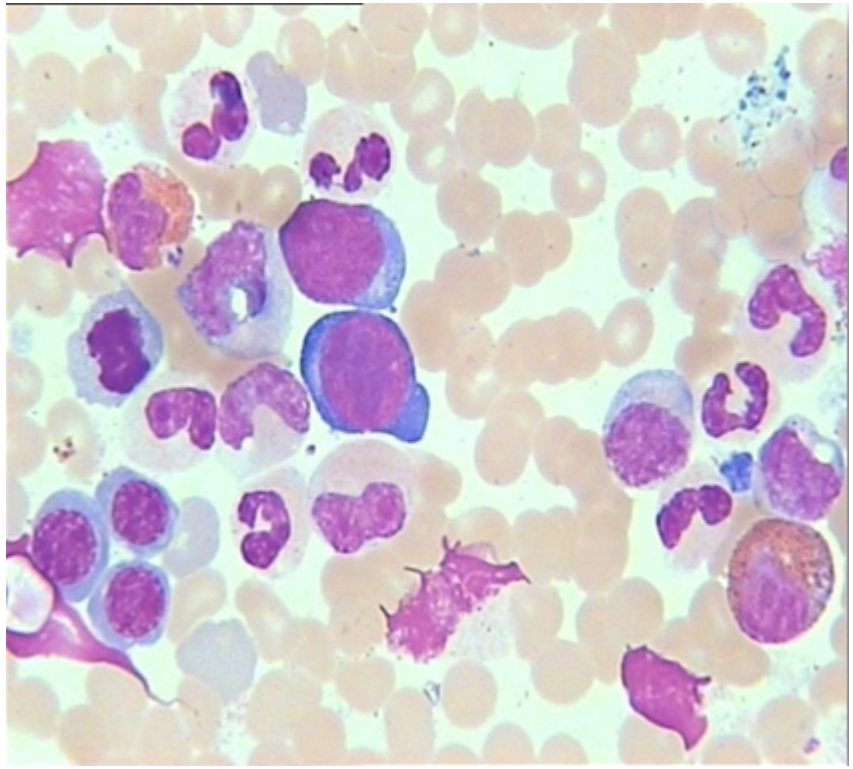
Analysis pipeline

Observe the somatic mutation pattern in single cell level

- Somatic mutation statistics
- Derived allelic frequency spectrum of somatic mutations
- Mutation prevalence of single cell level in different cancer
- Somatic mutation types of single cell level in different cancer
- Cancer-mutated genes
- Functional validation

Four Cases from 1,000 Single Primary Tumor Cells Sequencing

Essential Thrombocythemia



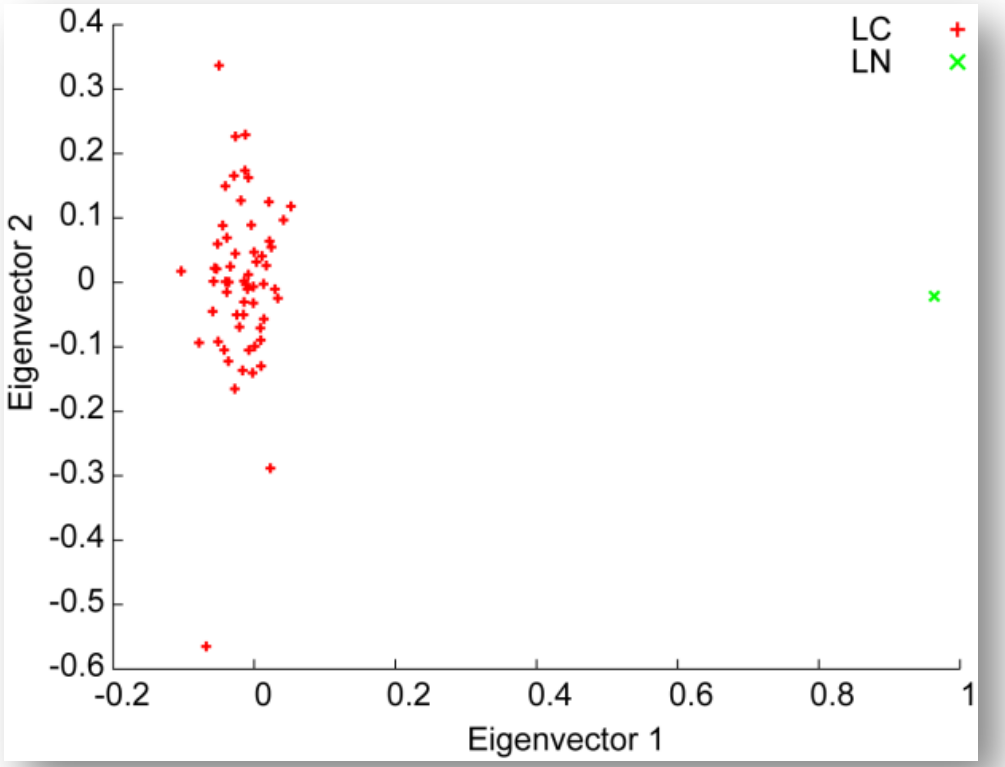
Myeloproliferative neoplasm is a kind of hematopoietic tumours that originate from the genetically variations contained hematopoietic stem cells or progenitors and lead to abnormal differentiation and myelopoiesis

A Wright's stained bone marrow aspirate smear of a ***JAK2*-negative** ET patient

Four Cases from 1,000 Single Tumor Cells Sequencing

Heterogeneity:

Essential Thrombocythemia



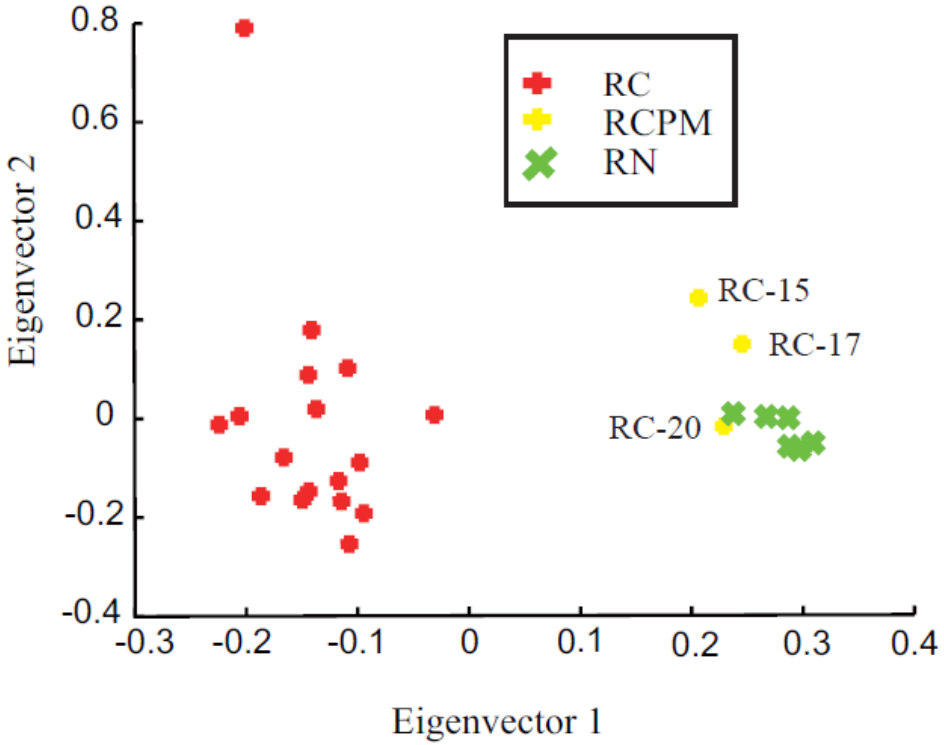
Somatic allele frequency between single cell sequencing and millions of cells shows consistency;

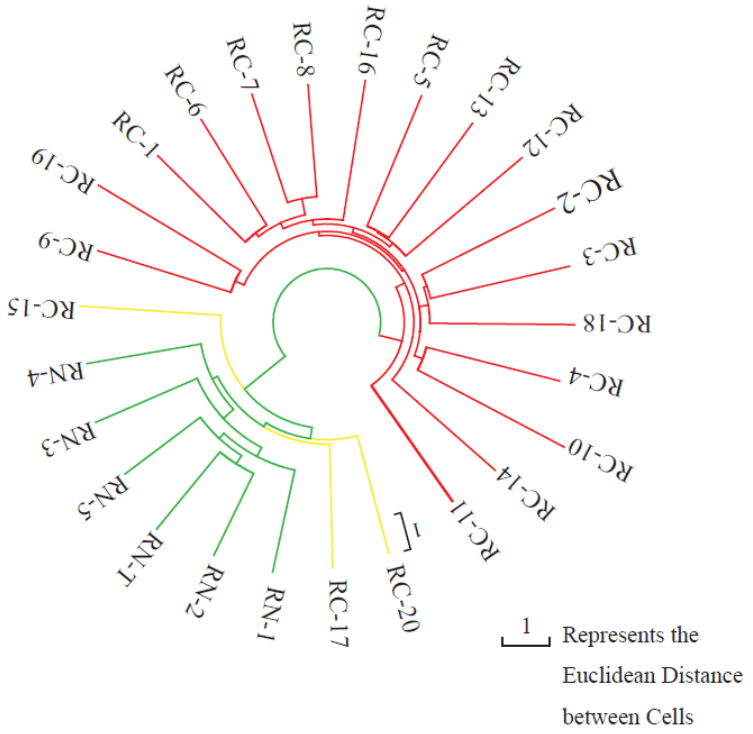
PCA analysis distinguish cancer and normal cells apparently

Four Cases from 1,000 Single Tumor Cells Sequencing

Heterogeneity:

Renal Cancer (ccRCC-1)

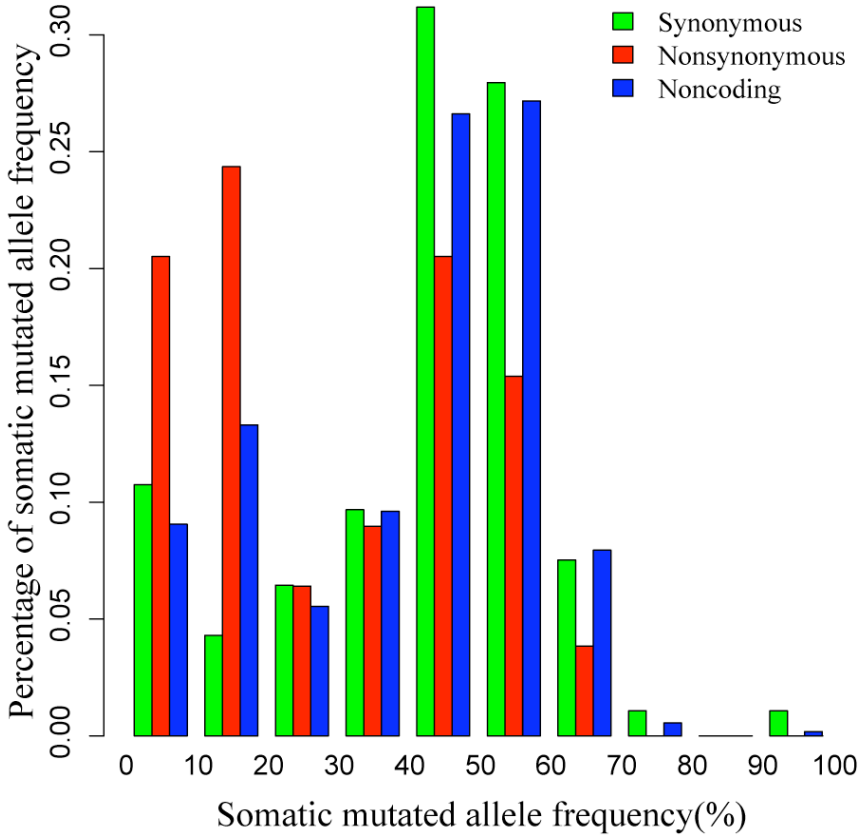


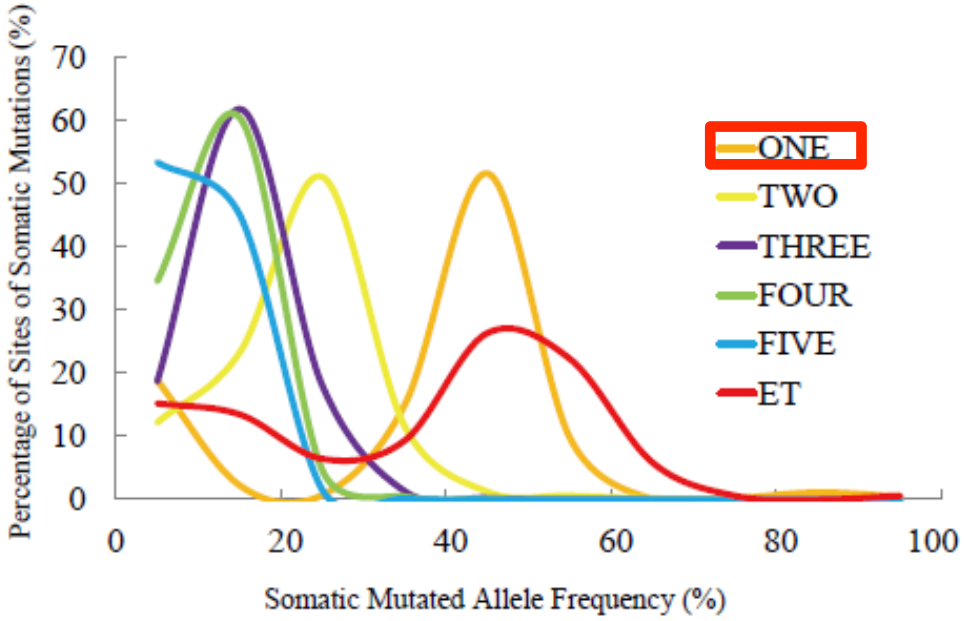


Phylogenetic analysis shows three "cancer" cells present among normal cells, and also showed the homogeneity of renal cancer (no obvious subpopulations were observed)

Four Cases from 1,000 Single Tumor Cells Sequencing

Progression: Essential Thrombocythemia



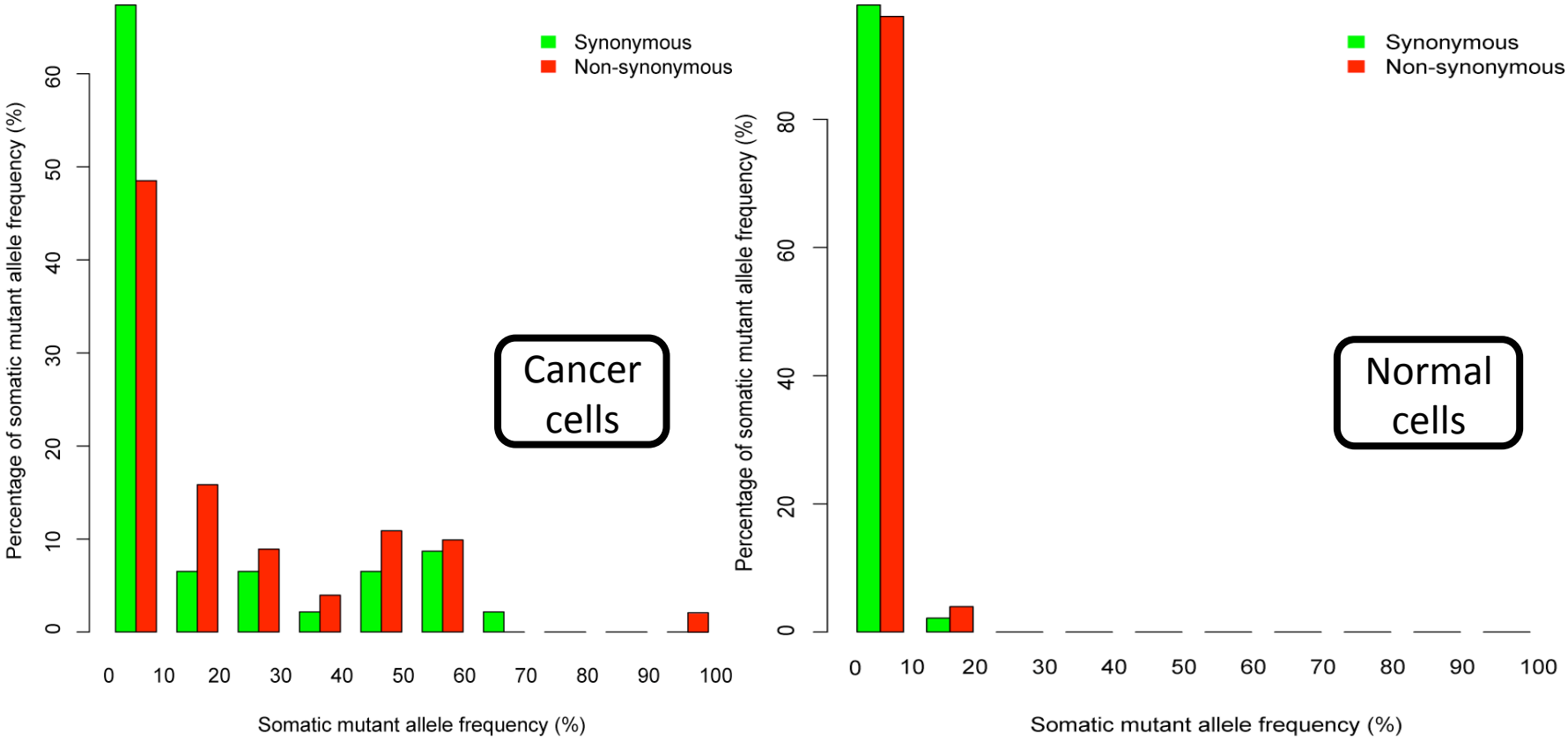


Comparing the simulation of somatic mutated allele frequency and our data shows the potential monoclonal origin of this kind of disease.

Four Cases from 1,000 Single Tumor Cells Sequencing

Progression:

Bladder Cancer (BTCC)

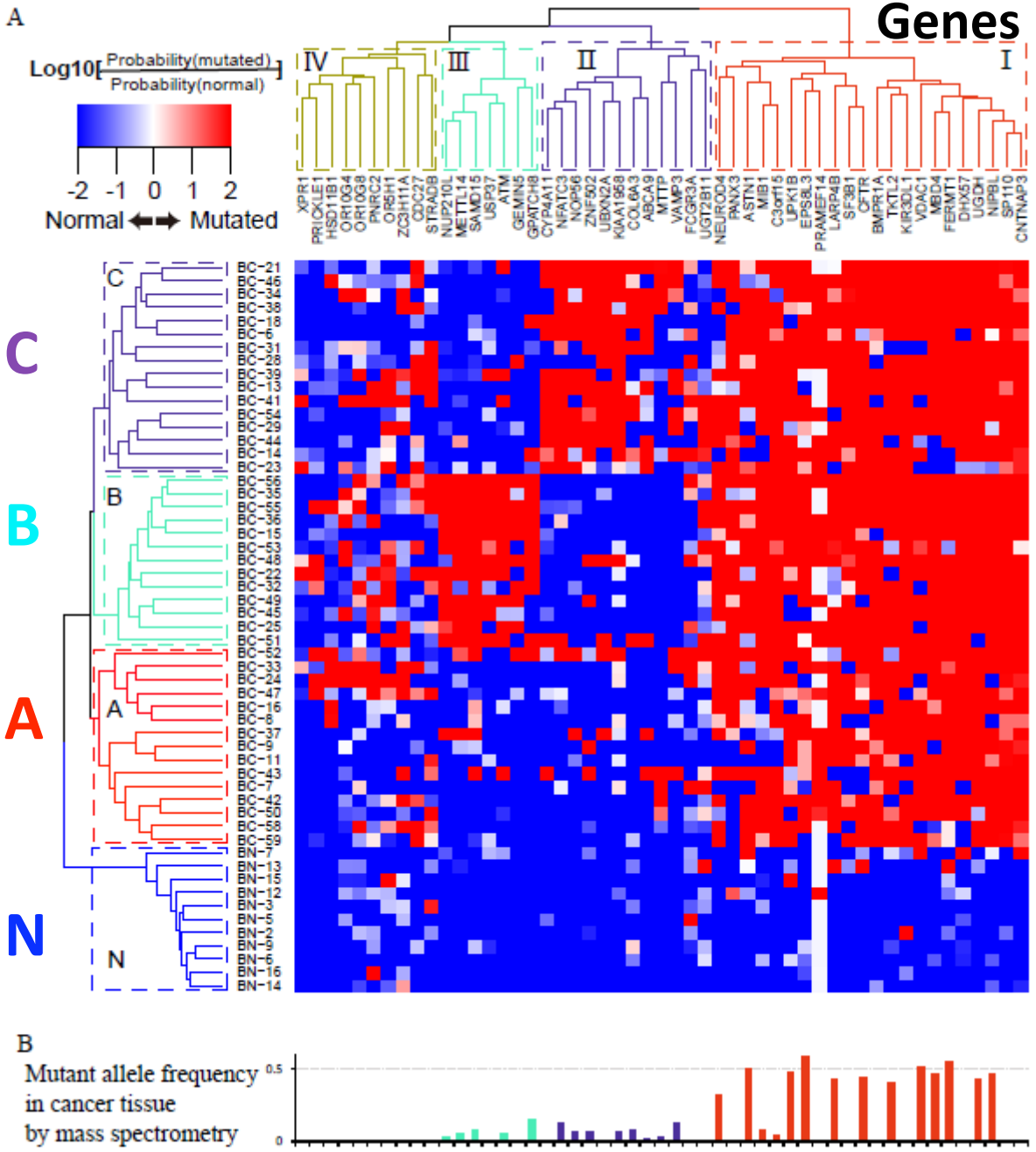


This indicates that this TCC is very likely to originate from only one ancestral tumour cell with heterozygous mutations

Progression:

Bladder Cancer (BTCC)

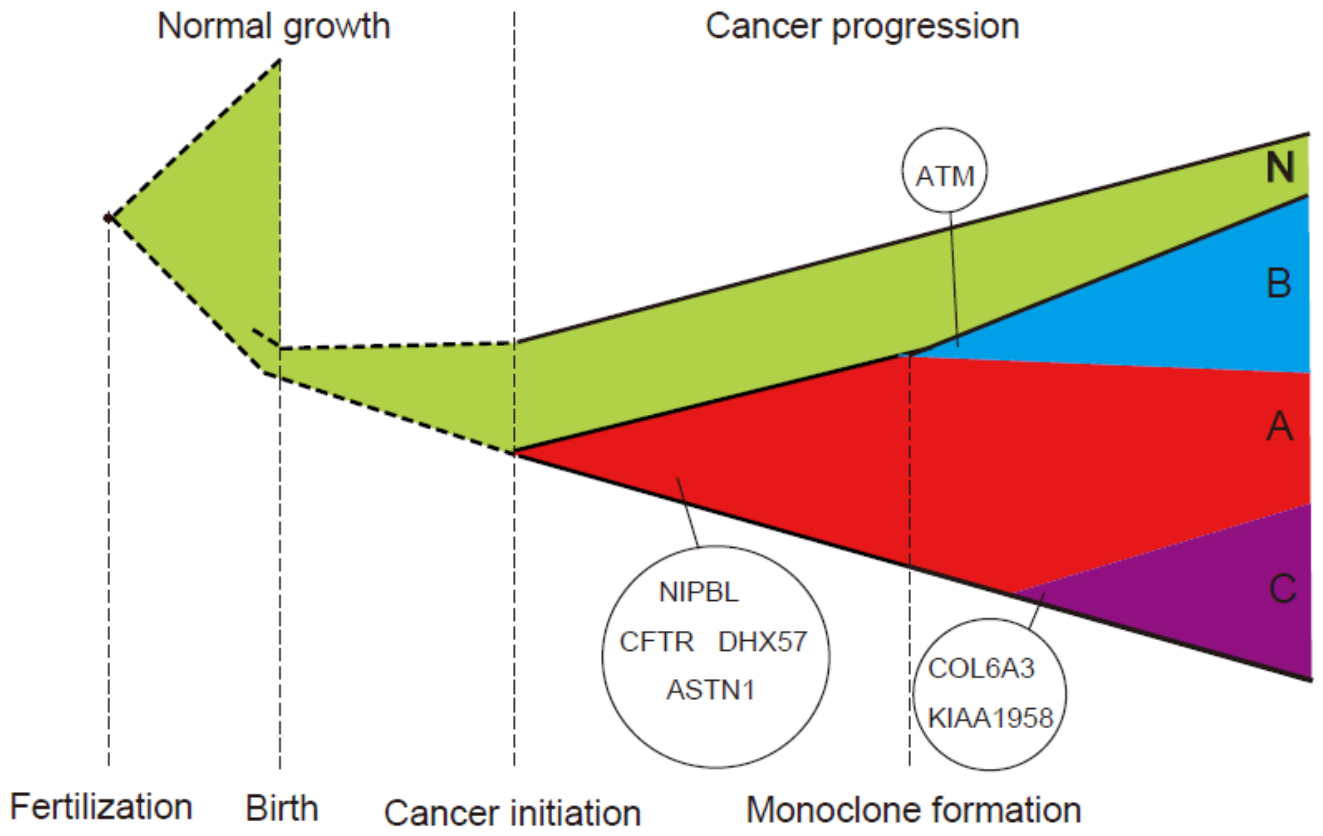
The tumour cells could be classified into 3 identifiable subclones with different genetic mutational signatures with 3 different groups of genes (A, B, C); N represents normal cells here.



Four Cases from 1,000 Single Tumor Cells Sequencing

Progression:

Bladder Cancer (BTCC)



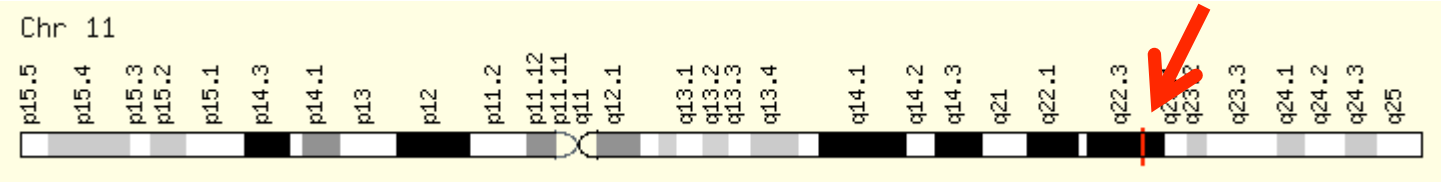
Muscle invasive?

The tumour evolution inferred by the heatmap

Four Cases from 1,000 Single Tumor Cells Sequencing

Progression:

Bladder Cancer (BTCC)



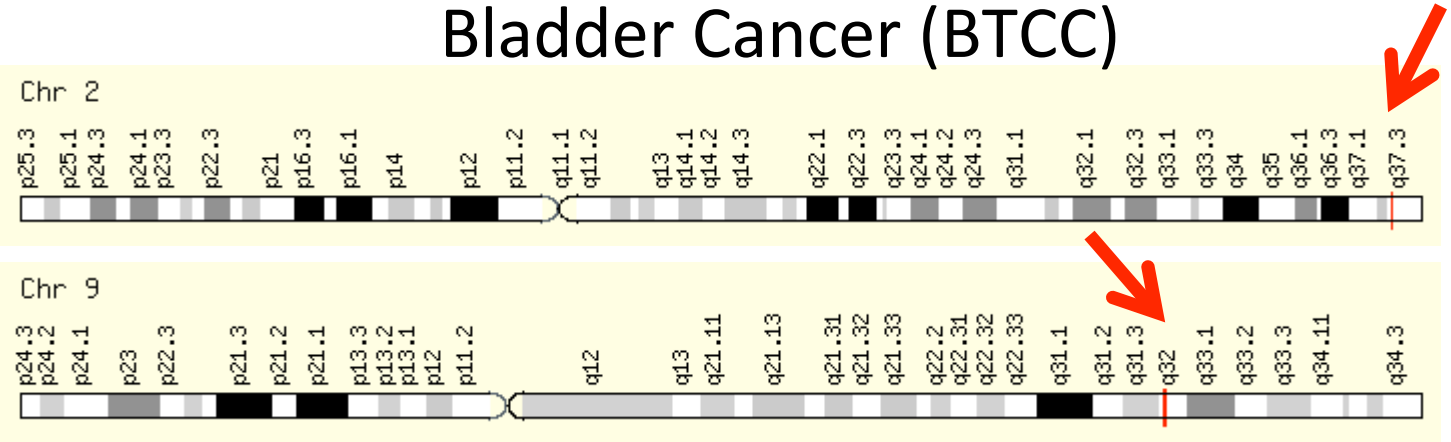
Gene
ATM

- In Clone B, *ATM* is specifically mutated and found recurring in 5 other TCC patients in the patient cohort. It is a known tumour suppressor that plays a key role as a cell cycle checkpoint kinase in response to DNA damage and is a regulator of a wide variety of downstream proteins (Rotman and Shiloh 1998; Branzei and Foiani 2008). Defects in this gene could increase mutation rate and genome instability and facilitate tumour progression

Four Cases from 1,000 Single Tumor Cells Sequencing

Progression:

Bladder Cancer (BTCC)

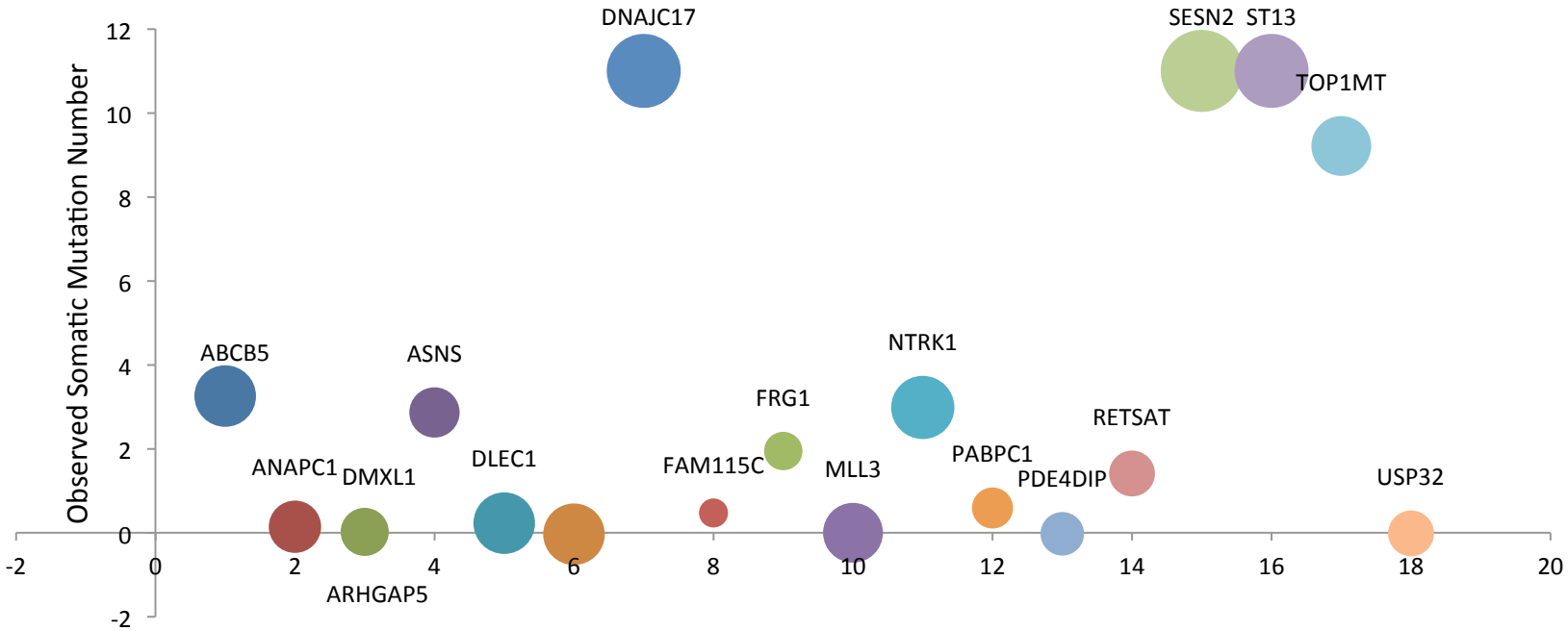


- Clone-C-specific mutated genes *COL6A3* and *KIAA1958* both recurred in 4 additional patients in the patient cohort. *COL6A3* is reported to have significant changes in expression level in tumour tissue(Smith, Culhane et al. 2009) and is a subunit of collagen IV, a cancer biomarker(Ohlund, Lundin et al. 2009). The *KIAA1958 gene* encode a unknown protein.

Four Cases from 1,000 Single Tumor Cells Sequencing

Key Mutations

Essential Thrombocythemia



Driver prediction of the non-synonymous somatic mutations: Q-score was calculated according to a modified method by (Youn and Simon 2011). Genes with Q-score more than 1 were identified as key genes.

Four Cases from 1,000 Single Tumor Cells Sequencing

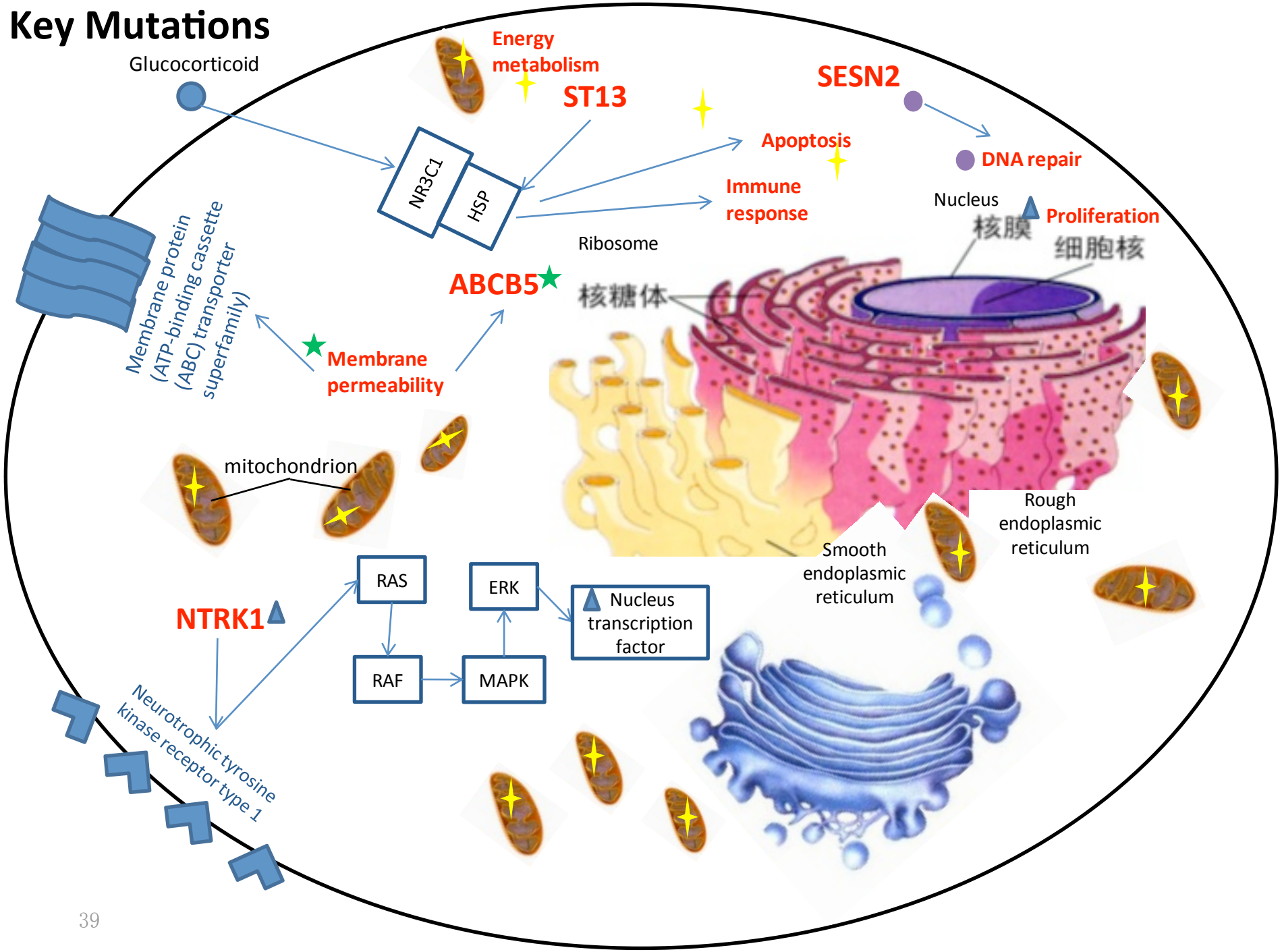
Key Mutations

Essential Thrombocythemia

Gene Name	Mutation Type	Monoclone-origin Gene	Functional Analyses
SESN2	Missense	Yes	SESN2 encoded a member of the sestrin family of SESN1-related proteins and was an antioxidant activated by p53. Mutation in SESN2 may lead to lack of DNA repair and damage prevention
ST13	Nonsense	No	ST13 encodes an Hsc70-interacting protein in controlling the activity of regulatory proteins such as steroid receptors and regulators of proliferation or apoptosis. Mutation in ST13 may contribute to loss the control of apoptosis and lead to abnormal proliferation.
NTRK1	Missense	No	A known oncogene, mutation in NTRK1 may contribute to sustained angiogenesis and cell proliferation
ABCB5	Missense	No	Up-regulation of ABCB5 was responsible for multidrug resistance in several cancers

Key genes with known function and correlation with cancer.

Key Mutations



Four Cases from 1,000 Single Tumor Cells Sequencing

Key Mutations

Renal Cancer (ccRCC-1)

Table 2. Key Genes Identified in This Patient

Gene Name	Mutations	Patient Prevalence (%)[a]	P Value[b] (Passenger Probability)	Mutant Allele Frequency in Cancer Tissue	Mutant Cell Number	Mountain/Hill[a]
AHNAK	g.chr11:62042132G > A; p.P5445 > S	5%	9.29×10^{-9}	20%	12	M
LRRK2	g.chr12:38985956A > G; p.I1294 > V	4%	4.28×10^{-4}	8%	8	H
SRGAP3	g.chr3:9041948T > A; p.R535[a]	2%	2.92×10^{-1}	34%	16	M
USP6	g.chr17:4976948C > G; p.T72 > R	2%	3.26×10^{-1}	1.99%	3	H

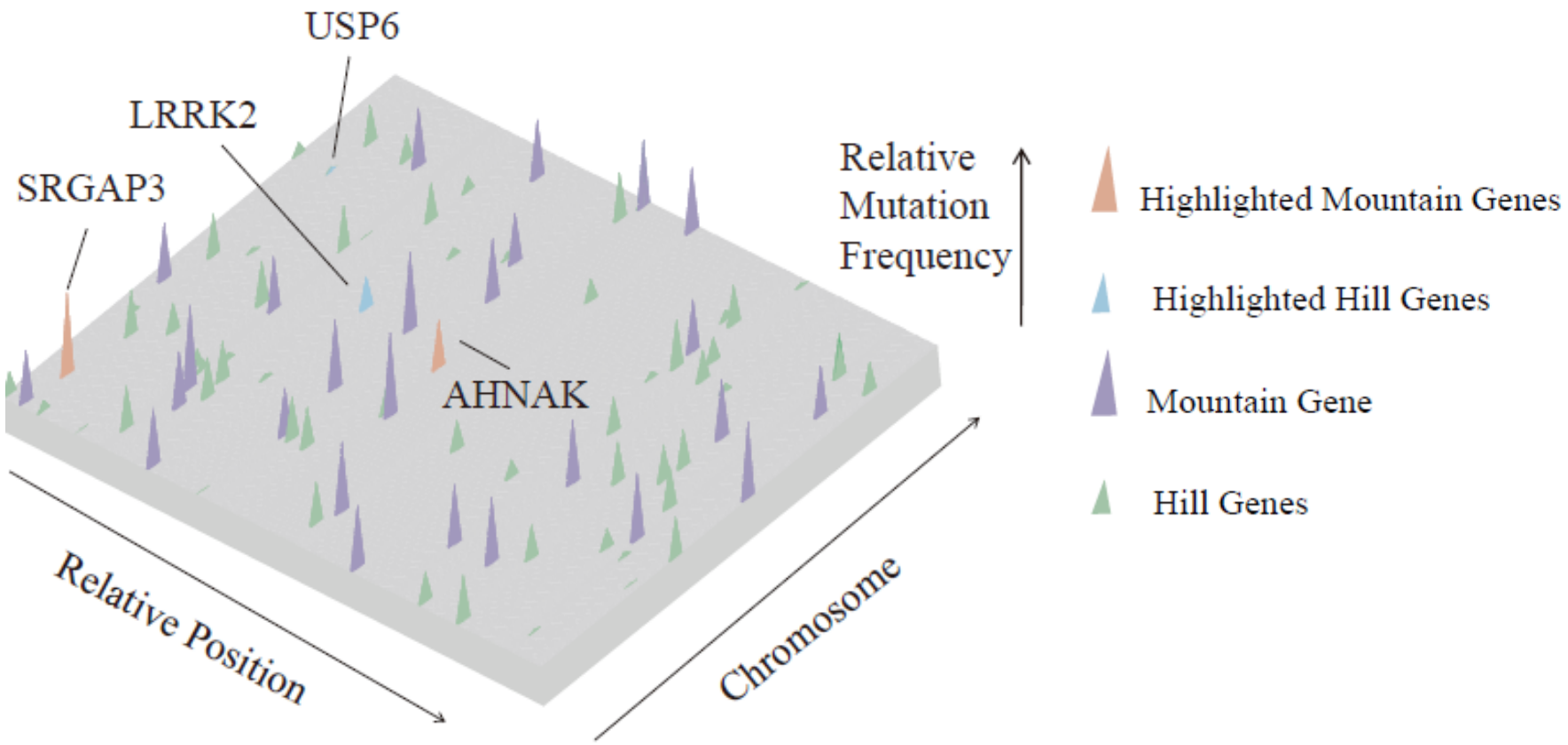
[a]Patient prevalence means the mutant genes recurred in the 99 ccRCC patients (including this patient); M/H represents mountain or hill gene.
[b]Significance of the observed mutation rate over the expected mutation rate in Guo et al. (2012).

Thanks to the large Chinese ccRCC patient cohort data, we compare the mutations in this patient and mutations in the large patient cohort, and found these recurrently genes.

Four Cases from 1,000 Single Tumor Cells Sequencing

Key Mutations

Renal Cancer (ccRCC-1)

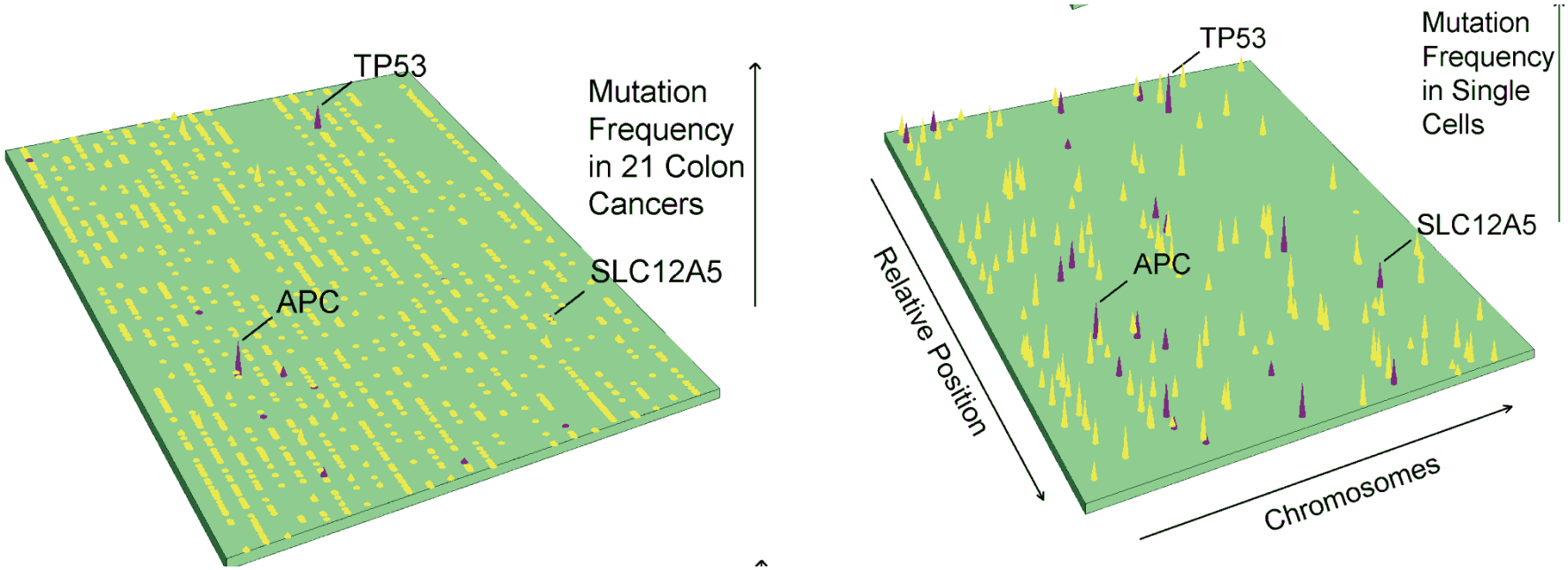


Mutated genes landscape: mountain (tissue common mutation) and hill(cell specific mutation) genes;

Four Cases from 1,000 Single Tumor Cells Sequencing

Key Mutations

Colorectal

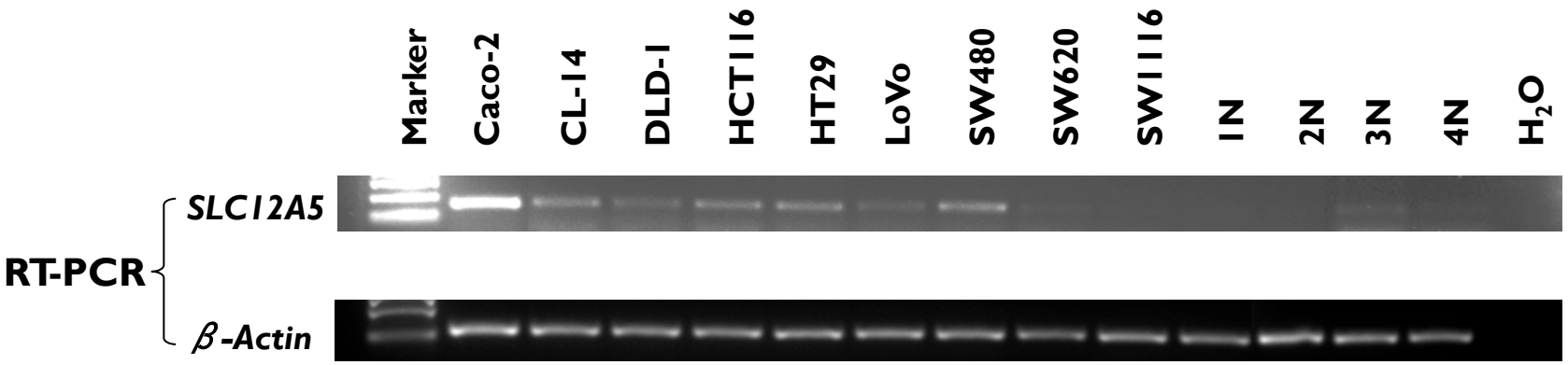


Thanks to the large colon cancer cohort data, we compare the mutations in this patient and mutations in the large patient cohort, and found recurrent genes which may play important roles in this individual.

Four Cases from 1,000 Single Tumor Cells Sequencing

Functional analysis

Colorectal



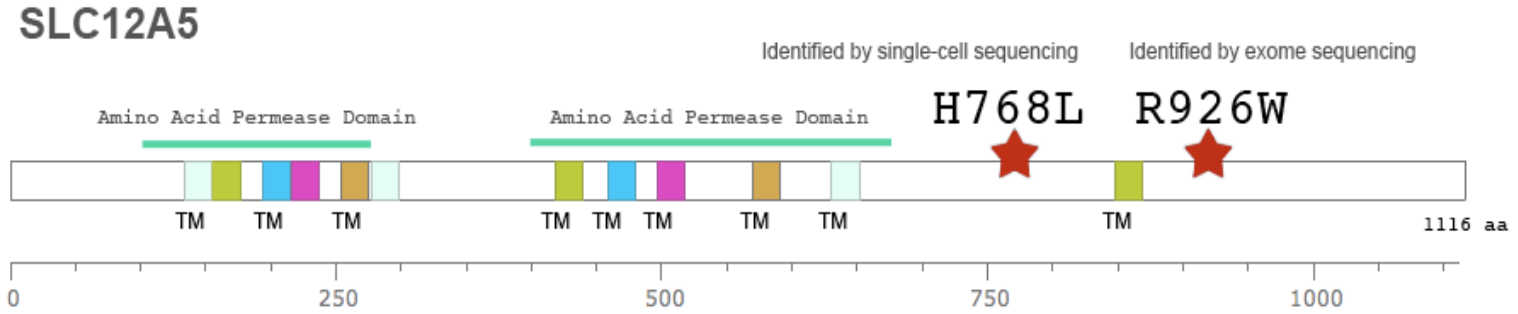
SLC12A5 was upregulated in colon cancer cell lines

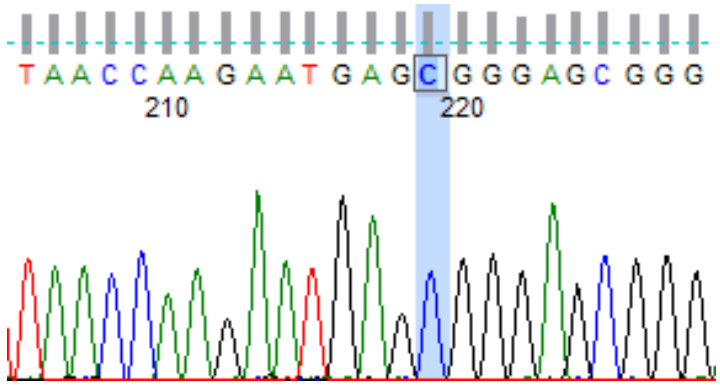
Four Cases from 1,000 Single Tumor Cells Sequencing

Colorectal

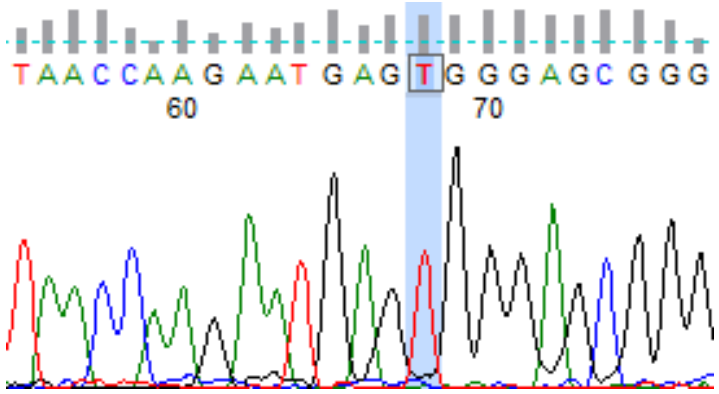
Construction of wild-type and mutant *SLC12A5* expression vector by site-directed mutagenesis

Functional analysis





SCL12A5-WT



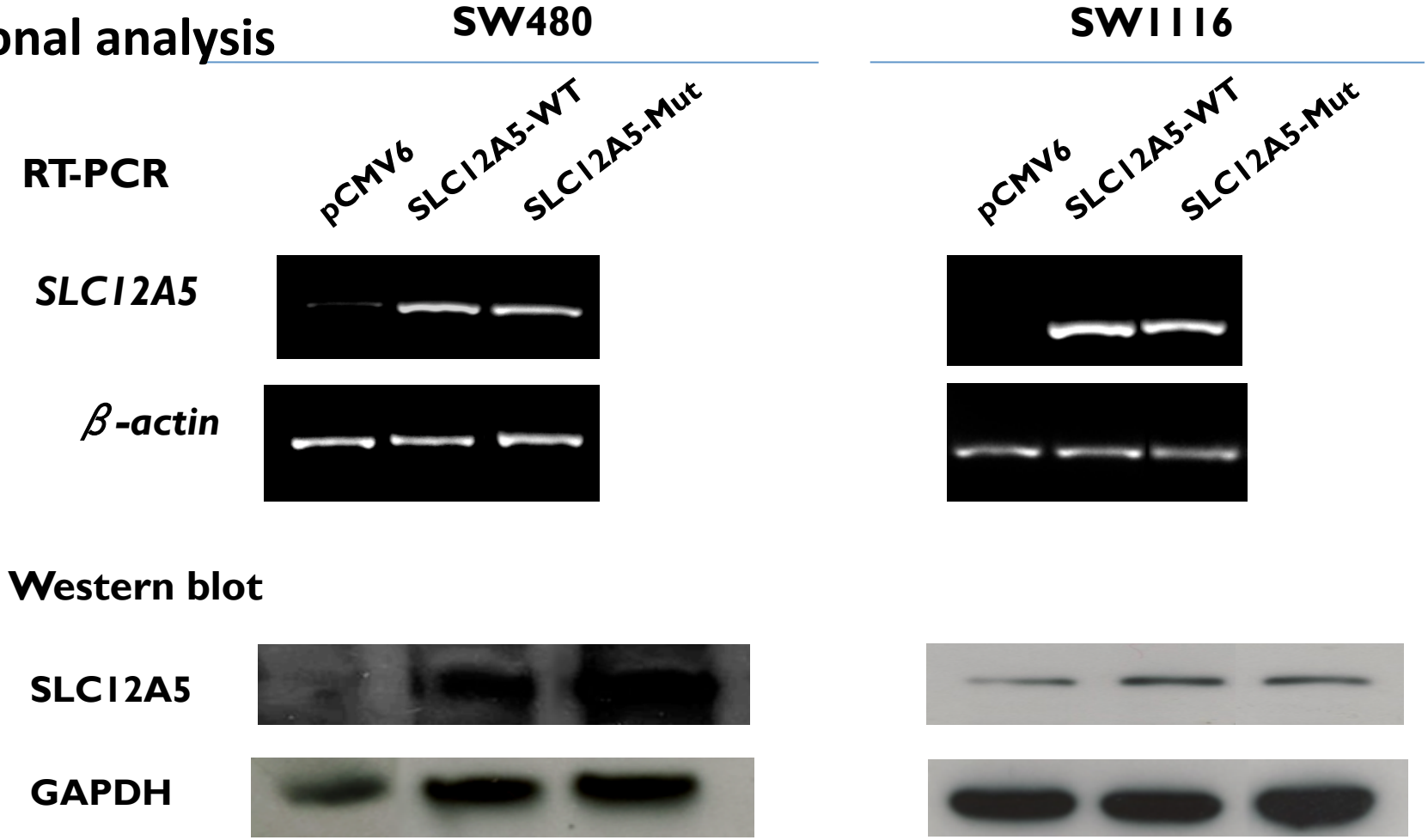
SCL12A5-Mut

Four Cases from 1,000 Single Tumor Cells Sequencing

Colorectal

mRNA & protein expression of *SLC12A5* in colon cancer cells transfected with WT and mutant *SLC12A5*-expressing plasmid

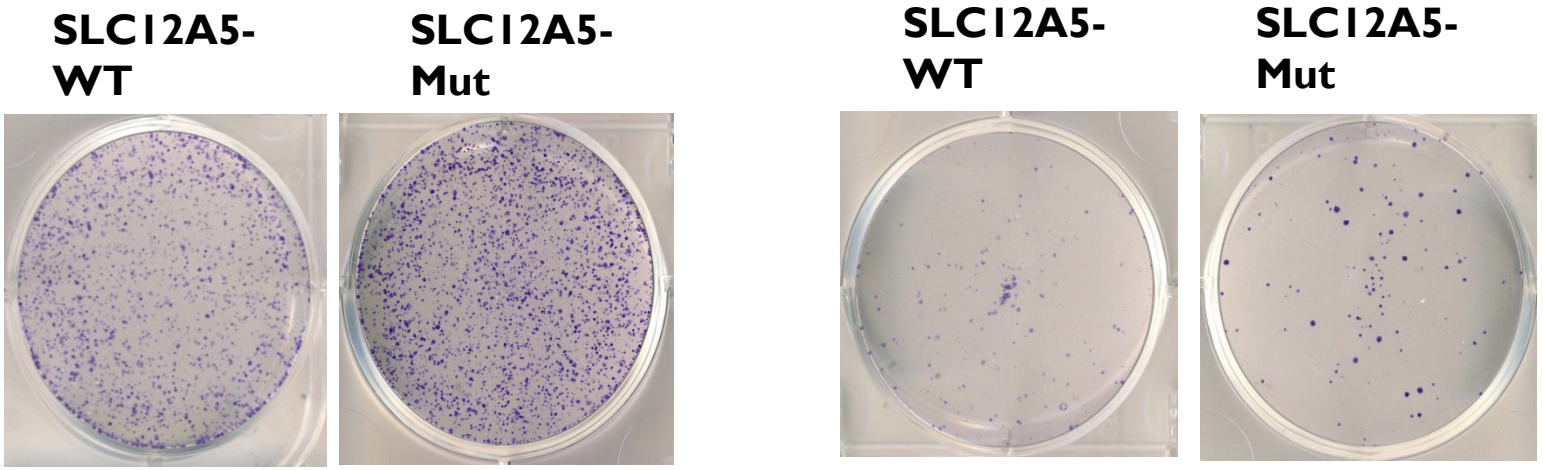
Functional analysis



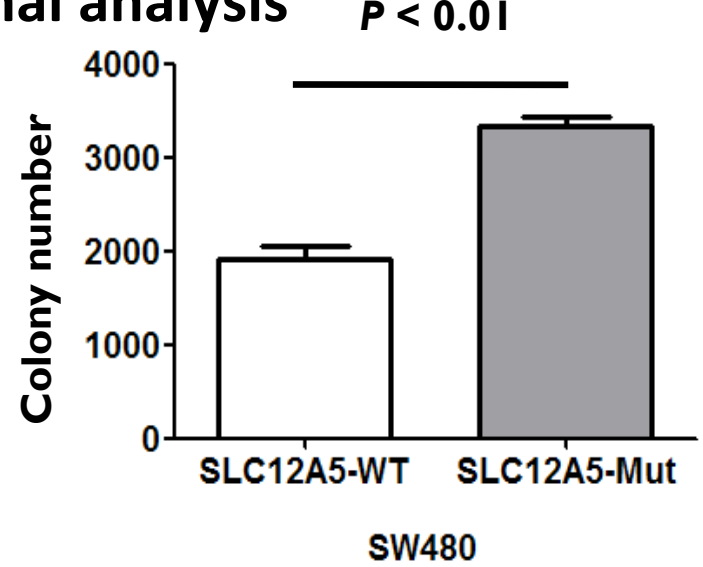
Four Cases from 1,000 Single Tumor Cells Sequencing

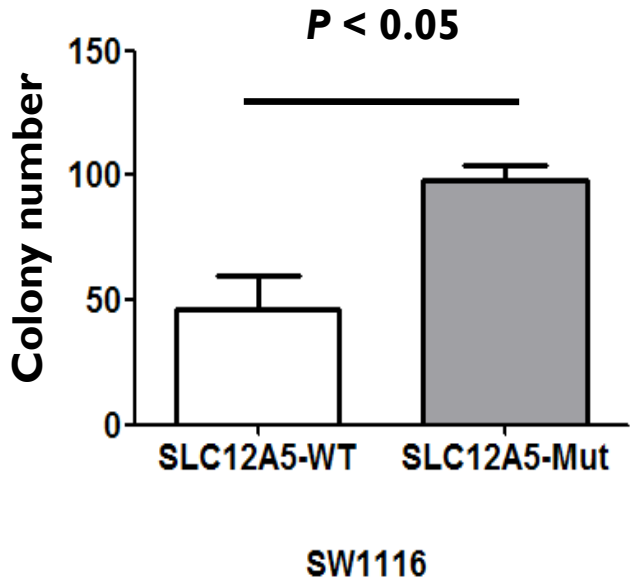
Colorectal

Mutant *SLC12A5* promoted colony formation



Functional analysis



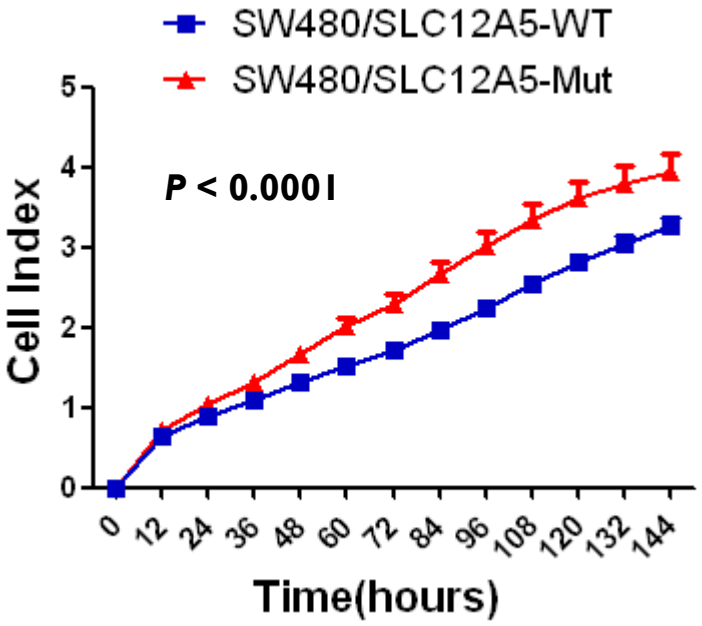


Colorectal

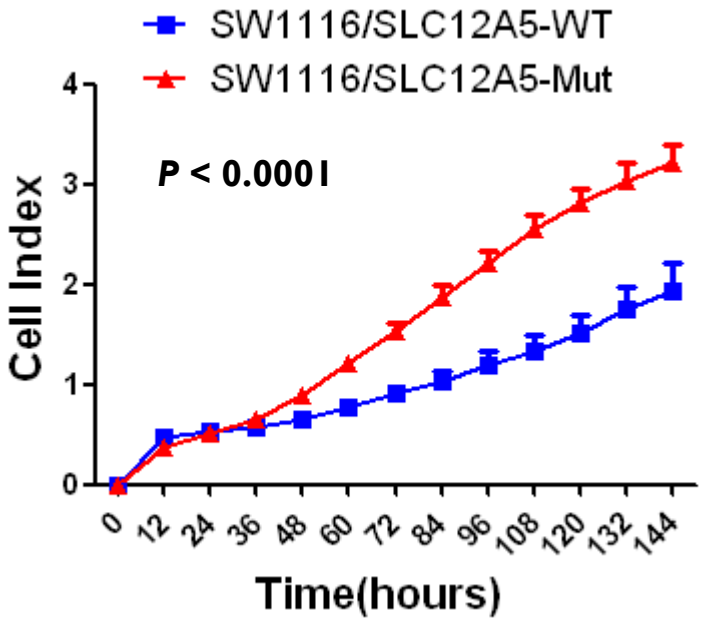
Functional analysis

Four Cases from 1,000 Single Tumor Cells Sequencing

Mutant *SLC12A5* promoted colon cancer cell proliferation



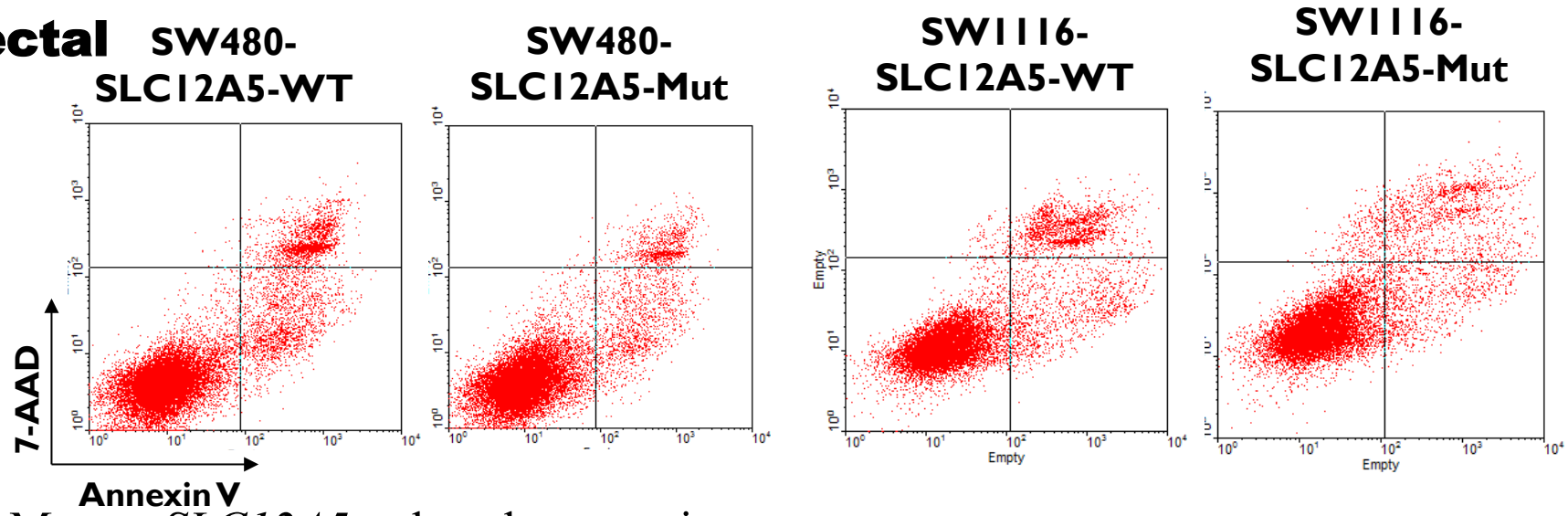
SW480 (Stable transfection)



SW1116 (Stable transfection)

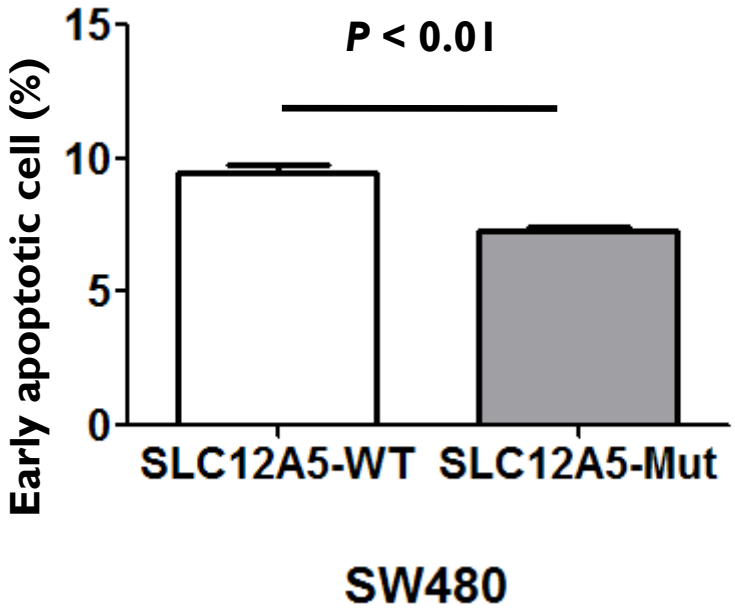
Four Cases from 1,000 Single Tumor Cells Sequencing

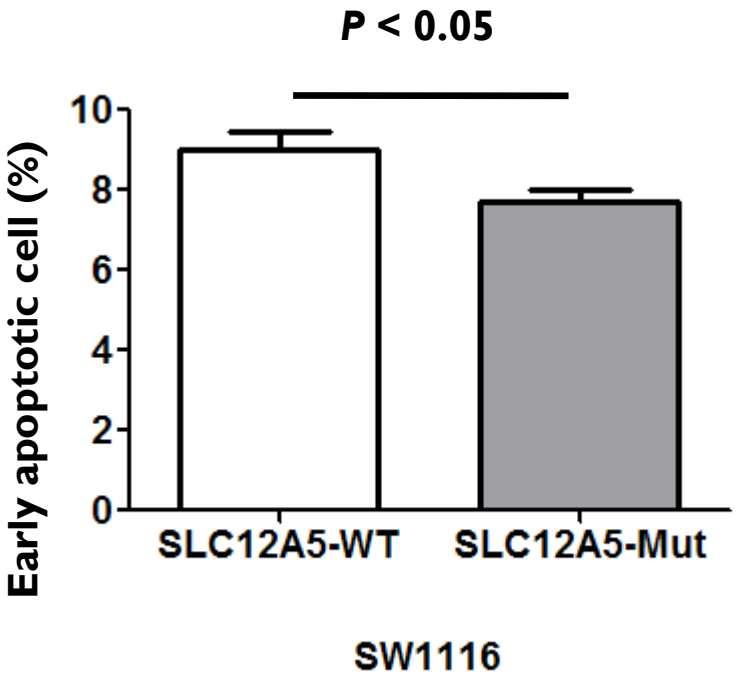
Colorectal



Mutant *SLC12A5* reduced apoptosis

Functional analysis





Four Cases from 1,000 Single Tumor Cells Sequencing

Functional analysis

Colorectal

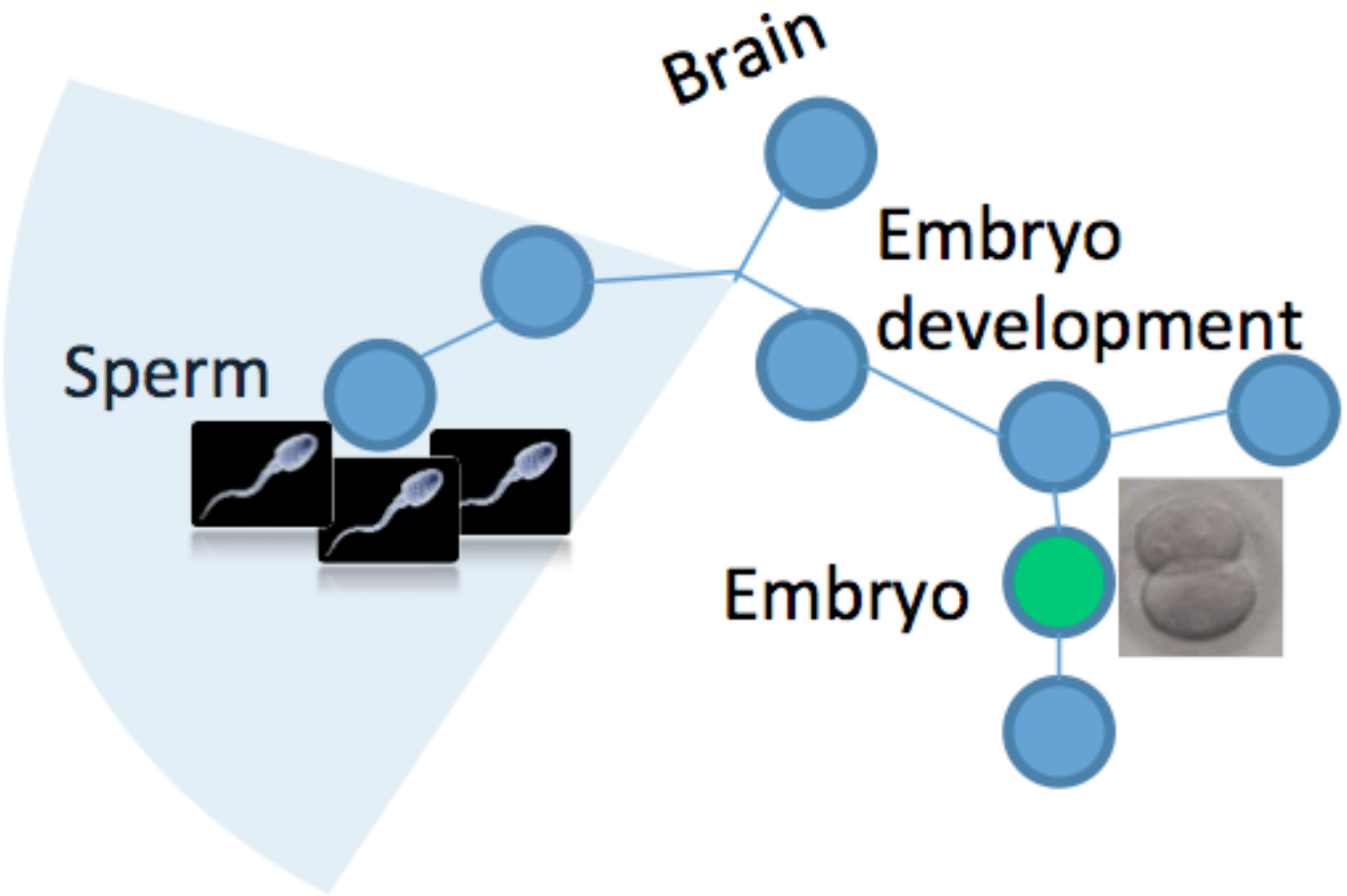
- A novel oncogenic mutation in *SLC12A5* with growth-promoting and anti-apoptotic function was identified

Four Cases from 1,000 Single Tumor Cells Sequencing

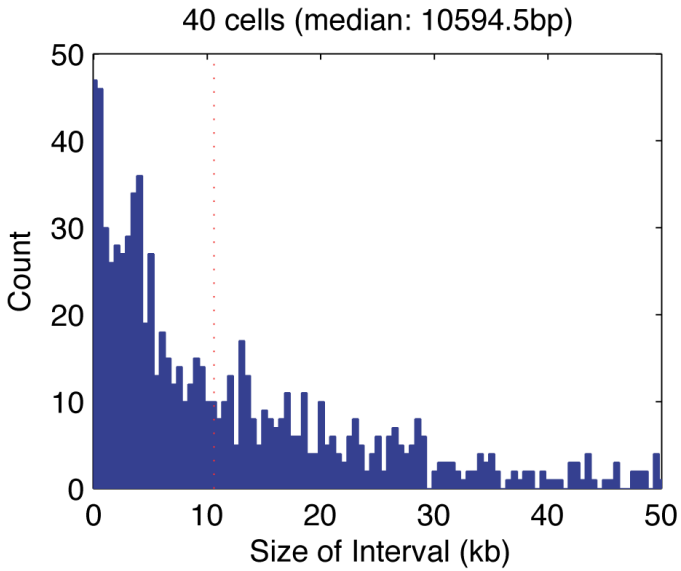
Summary

- We successfully infer the intratumoral heterogeneity and progression pattern from both blood tumor and solid tumor by single-cell exome sequencing;
- We identified key mutations and genes using independent methods in an individual tumor;
- Our results indicate the further application of single cell sequencing on cancer personalized medicine and target therapy.

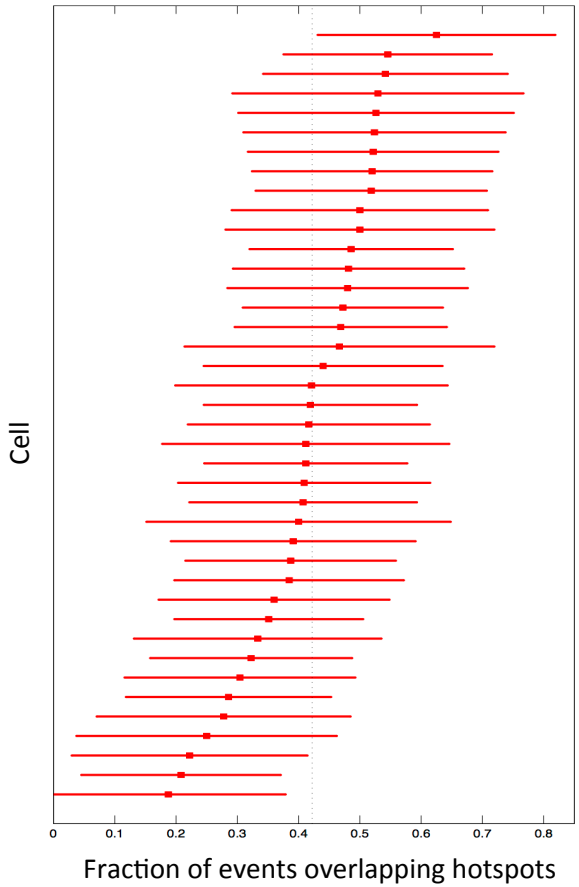
Life is a Game of Cell Evolution



Detection Resolution



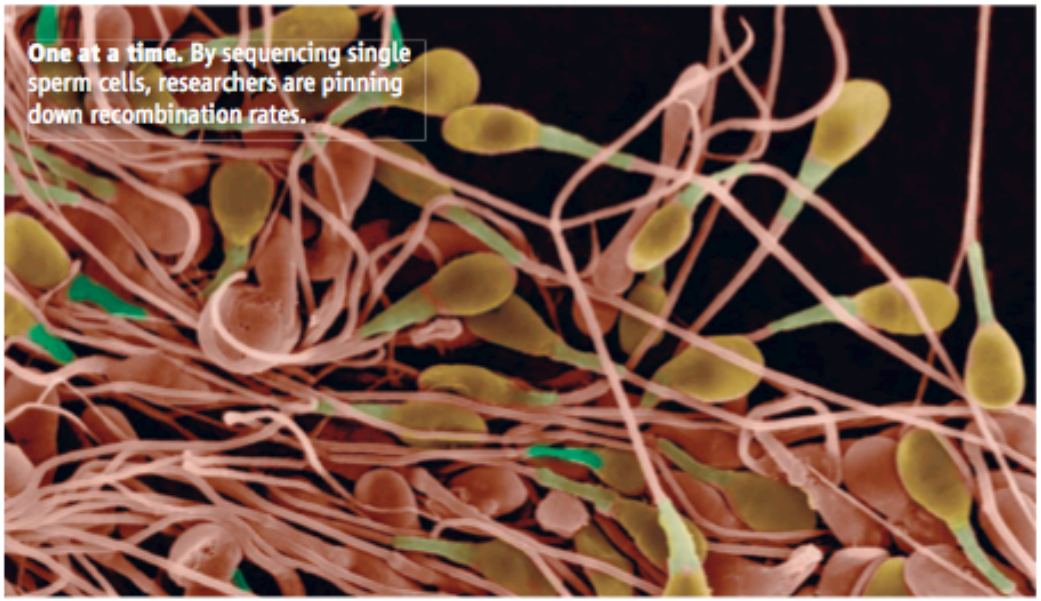
- 50% of events resolved to within 10.5kb
- 80% of events resolved to within 40kb



- **42.2% of events overlap known hotspots**, vs 23.4% overlapping 'coldspots' ($p<0.001$).
- Cannot reject null hypothesis of uniform hotspot usage across cells.

MEETINGBRIEFS>>

THE BIOLOGY OF GENOMES | 8–12 MAY | COLD SPRING HARBOR, NEW YORK



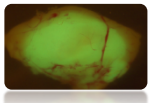
One at a time. By sequencing single sperm cells, researchers are pinning down recombination rates.

Single-Cell Sequencing Tackles Basic and Biomedical Questions

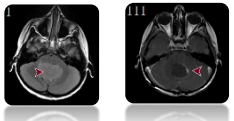
eager tc
lighted
cell anc
ing. Th
sequen
cells; o
each sp
a blooc
detecte
or abou
agrees
Recom
chromo
places
the mos
tified c
seen th
was als
some 1:
Step
in Palo
ing sin
lab on
of tiny

- Researchers have long sought a way to determine the amount of recombination that occurs in humans, and they have come up with several indirect ways to measure it in families or in populations.
- single-cell sequencing provides a window on recombination, the process by which matching chromosomes exchange pieces of their DNA during cell division. Recombination helps generate genetic diversity by putting various versions of genes together in new combinations.

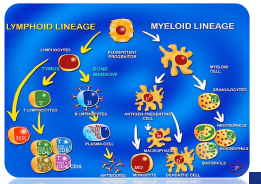
Life is a Game of Cell Evolution



Expression markers of cell lineage



Retro-transposon heterogeneity



TCR/BCR/MHC polymorphism

Brain

Immune

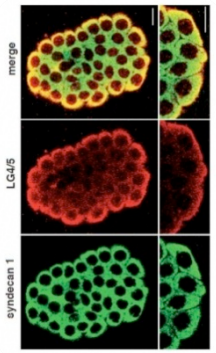
Expression markers of cell lineage

(GIGA)n SCIENCE

Embryo development

Primary tumor

Monoclonal cancer evolution

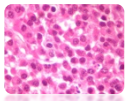


Sperm

Trans-omics marker

Evolution

Embryo

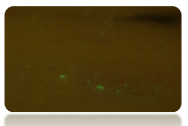


Element

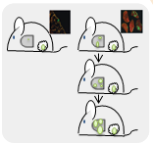
Trans-omics marker

Primary tumor

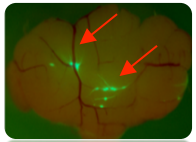
Metastasis/recurrent tumor



Gene and regulation network



Pharmacogenomics



Carcinogen (physical/ chemical/environment)

Pathogen (bacteria/virus)

polyclonal cancer evolution

Acknowledgement

- R&D center of BGI-Shenzhen: cell isolation, amplification, and analysis;
- BGI-Shenzhen Bioinformatics center: perform the analysis;
- BGI-Shenzhen NGS Platform: sequencing;
- Peking University Shenzhen Hospital and Shenzhen Second People's Hospital for providing the renal cancer and bladder cancer samples;
- Peking University First Hospital for providing the leukemia samples;
- Peking University Cancer Hospital for providing the gastric cancer and colon cancer samples;
- Stanford University School of Medicine Prof. Matthew Scott and his lab: collaborate on Medulloblastoma project

International Conference on Genomics

September 27 – 28, 2012 · Children's Hospital of Philadelphia

Come meet CEOs, directors, and senior executives from BGI, Merck, Pfizer, Sanofi, GSK, BMS, Illumina, Life Technologies, NHGRI, FDA, USDA, Harvard, UC Davis and many more!

Visit **www.icgamericas.org**

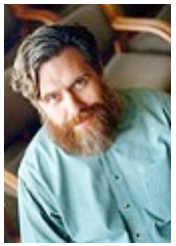
With over 40 renowned speakers, including:

Dr. Huanming Yang
Chairman & Professor
BGI

Dr. Jun Wang
Executive Director
BGI

Dr. Hakon Hakonarson
CAG Director
CHOP

Dr. Eric D. Green
Director
NHGRI



Dr. George M. Church
Professor
Harvard Medical School

Dr. Rupert Vessey
Senior VP
Merck

Dr. Stephanie Chissoe
Acting Head of Genetics
GlaxoSmithKline

Dr. James Watters
Head of Applied Genomics
Sanofi Oncology

Dr. Howard-Yana Shapiro
Global Director of
Plant Science
Mars Incorporated

Gregory T. Lucier
CEO
Life Technologies

Frank Witney
President & CEO
Affymetrix

Dr. Howard J. Jacob
HMGC Director
Medical College of
Wisconsin

Dr. Gabriel G. Haddad
Pediatrics Dept. Chair
UC San Diego

Dr. Richard W. Michelmore
Director
UC Davis Genome Center

Dr. Joe Gray
Director and Chair of
Biomedical Engineering
OHSU

Dr. Andy Shih
VP of Scientific Affairs
Autism Speaks

Dr. Eric Brown
Director
Division of Microbiology
FDA

Mr. Yingrui Li
Head of Bioinformatics
BGI

Visit www.icgamericas.org