2013.2.19

再生医療時代の
iPS細胞のゲノム解析

Center for iPS Cell Research and Application, Kyoto University
Genomics and Epigenomics Core Facility / Yamanaka Lab.
Akira Watanabe

Genomics and Epigenomics on iPS cell research

Safety

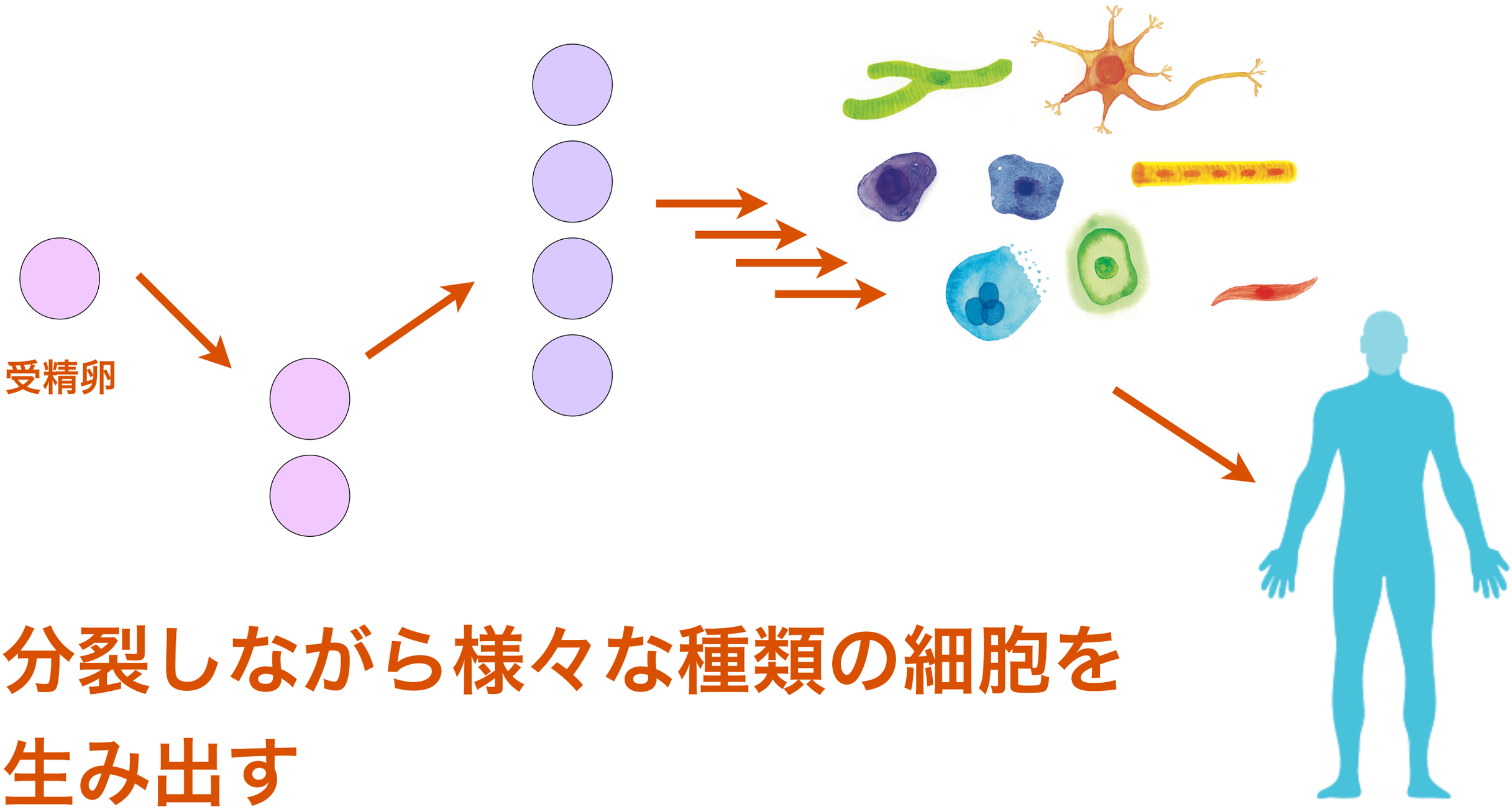
Quality

Genomics and Epigenomics on iPS cell research

Safety

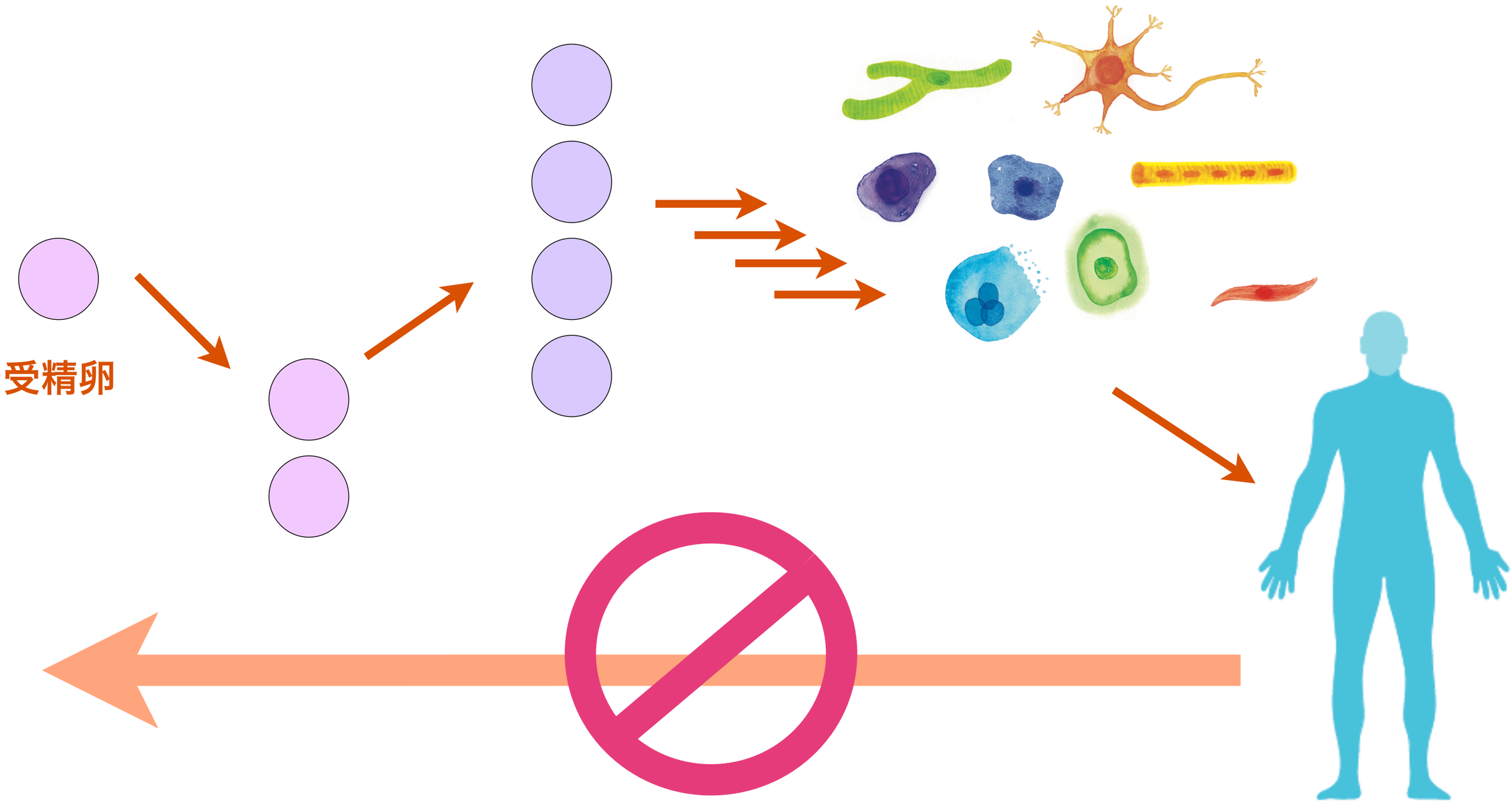
Quality

細胞は変化しながら増え続ける

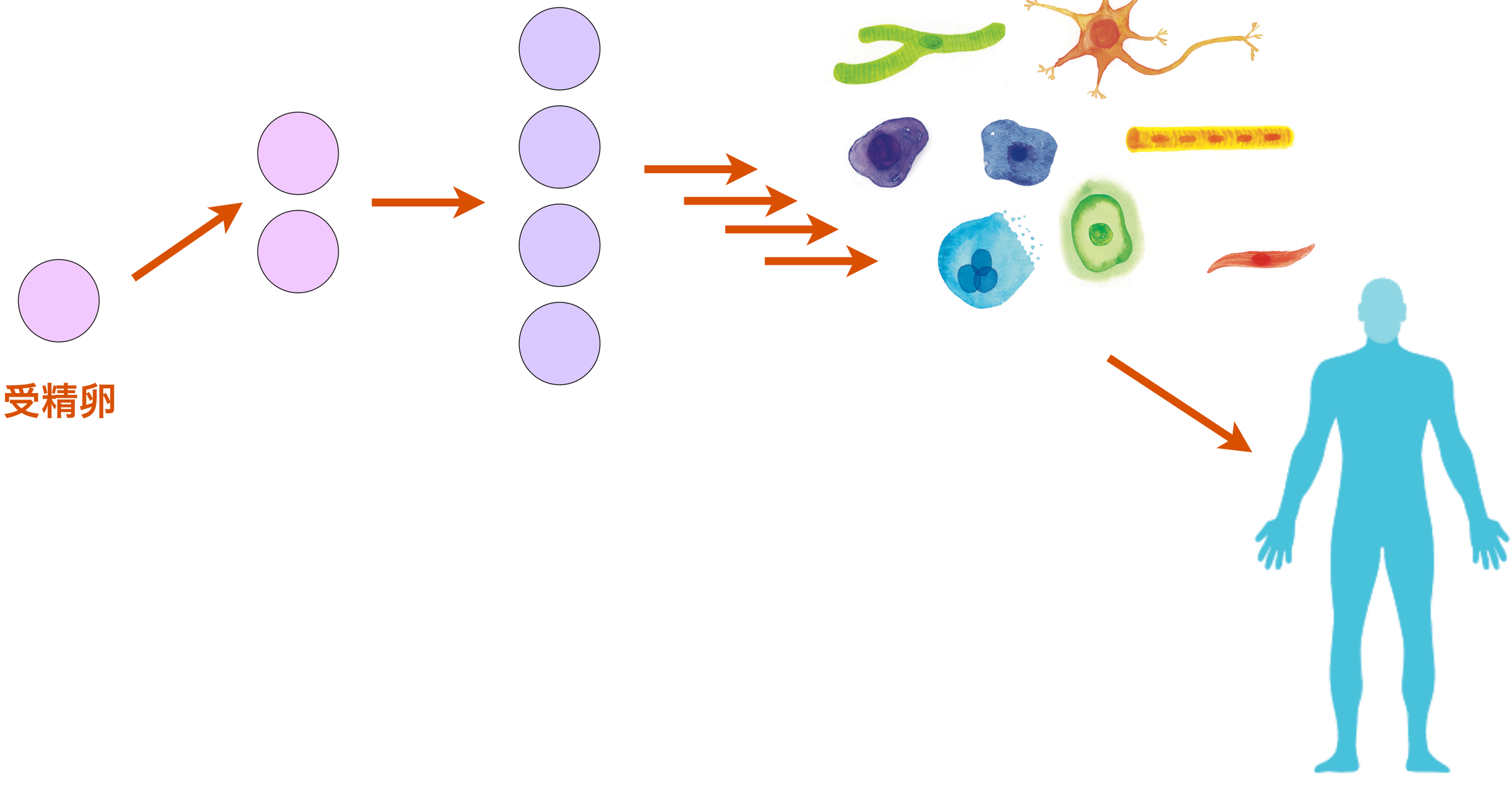


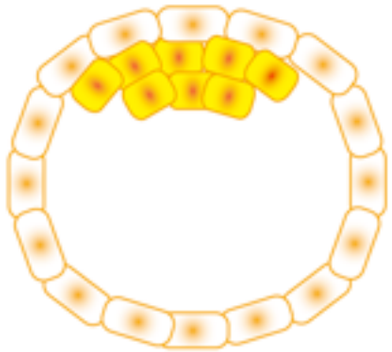
分裂しながら様々な種類の細胞を生み出す

細胞の変化は逆戻りしない

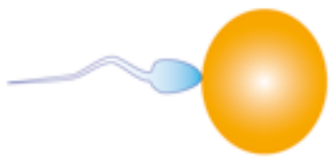


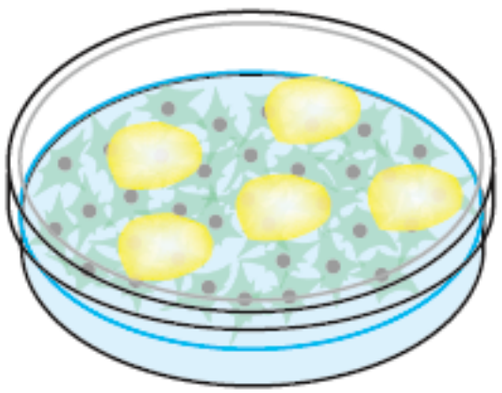
細胞の分裂回数は約50回





初期胚





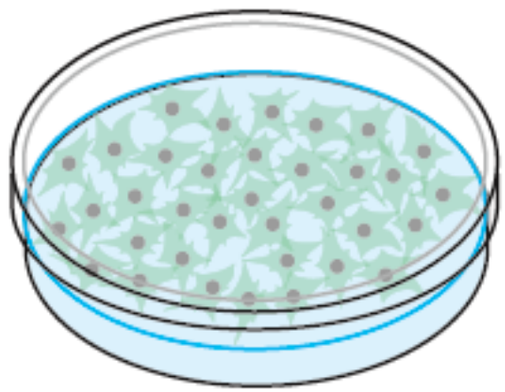
iPS細胞

Oct4

Sox2

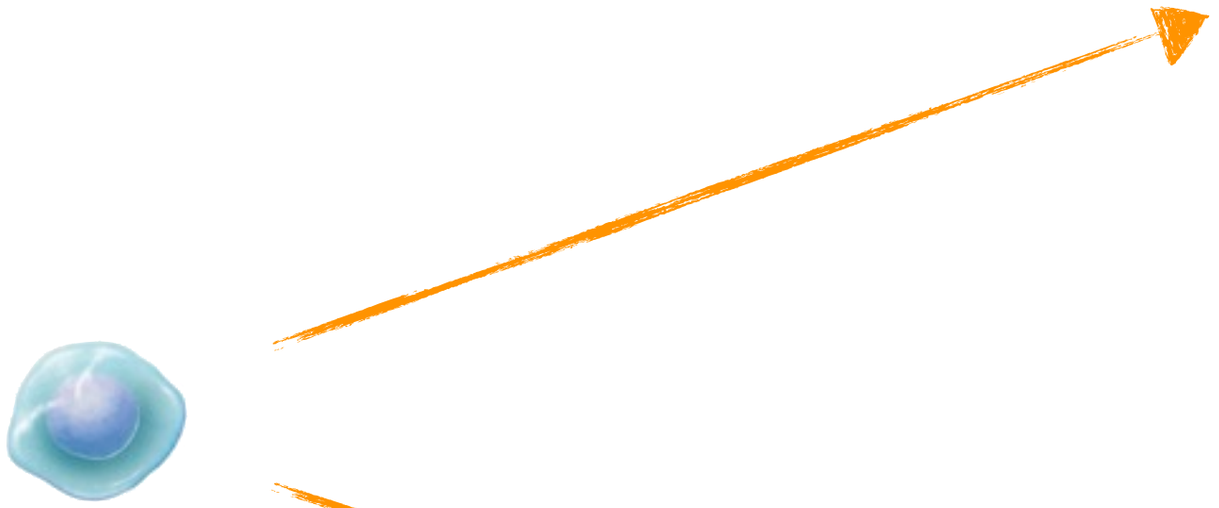
Klf4

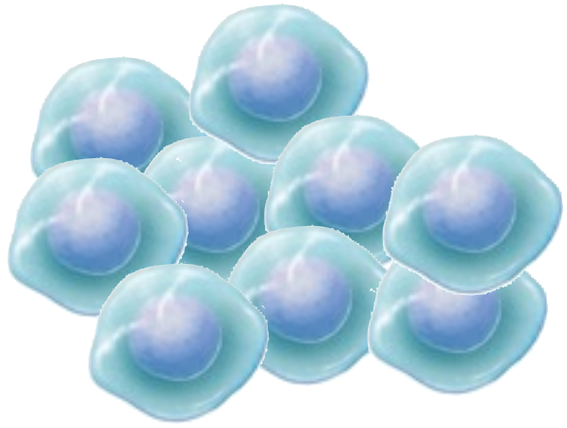
Myc



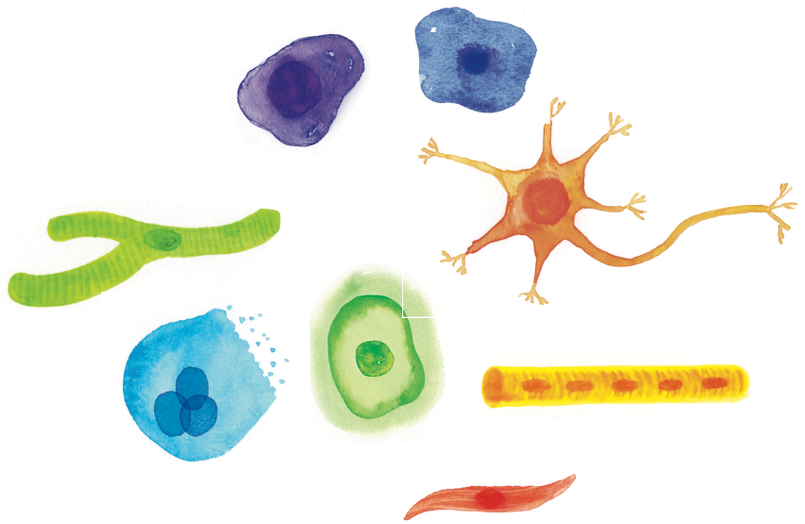
線維芽細胞

多能性幹細胞とは？





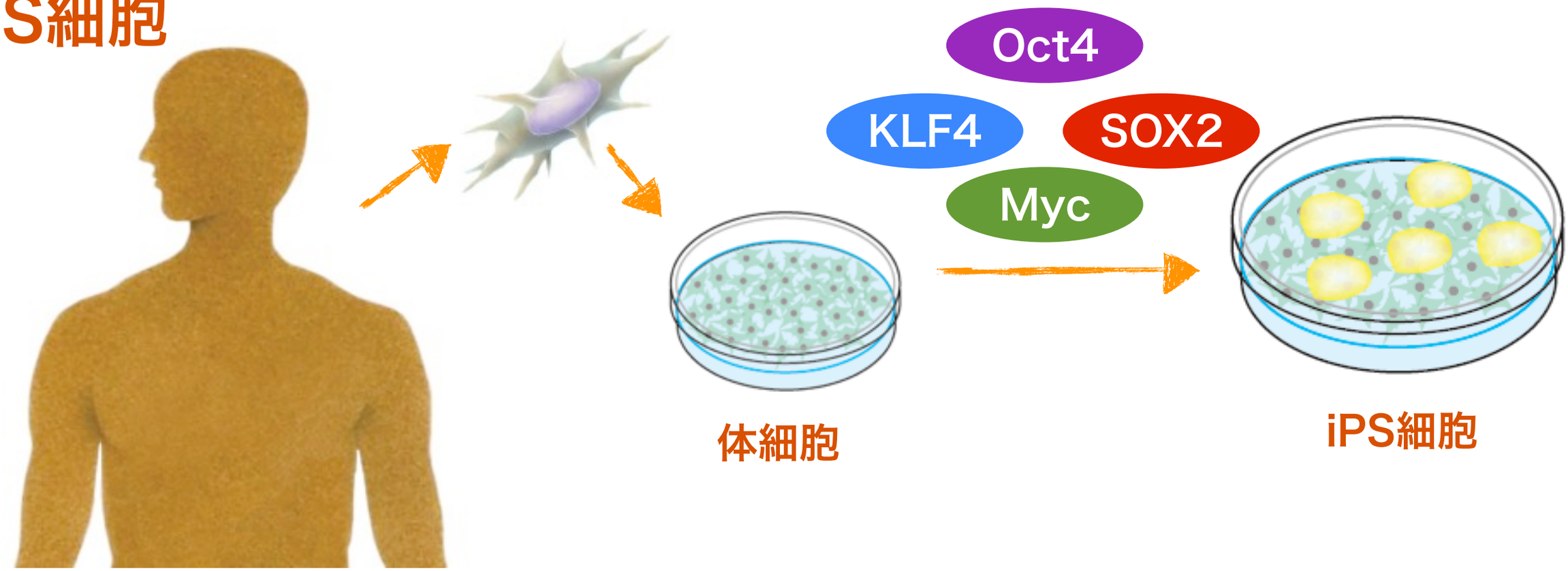
自己複製能



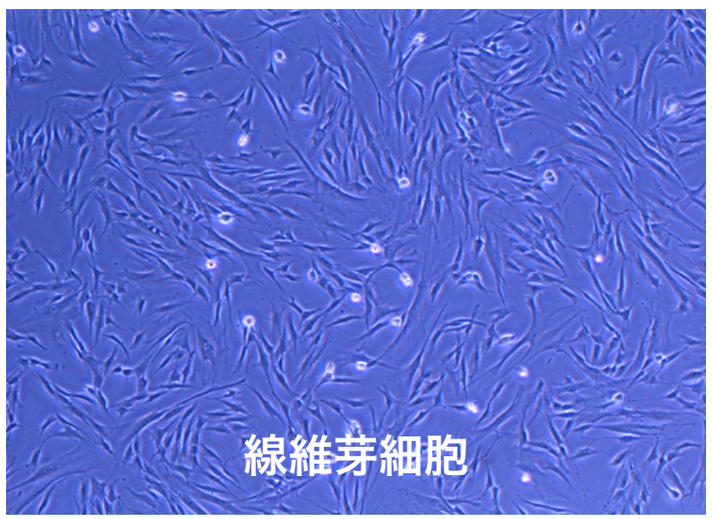
多分化能

ES細胞

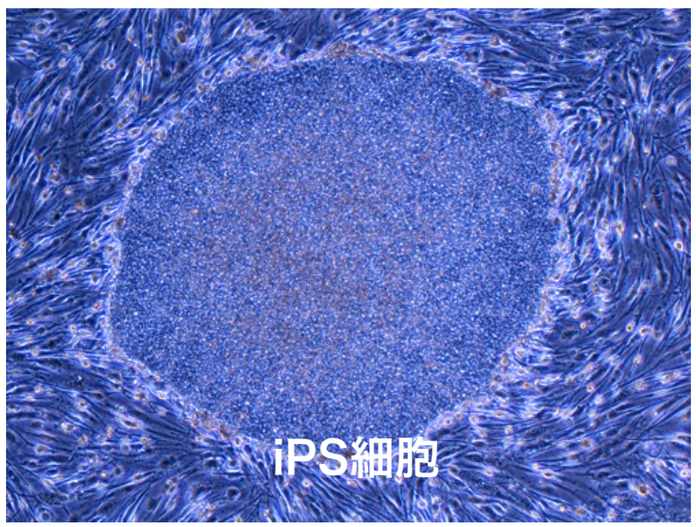
iPS細胞



iPS細胞の特徴



Oct4
KLF4
SOX2
Myc



コロニーの形態
三胚葉への分化（テラトーマ形成）
多能性マーカー(Nanog等)の発現
メチル化パターンの変化

従来の薬

化合物：合成・天然物

HPLC等で精製や純度決定

従来の薬

化合物：合成・天然物

HPLC等で精製や純度決定

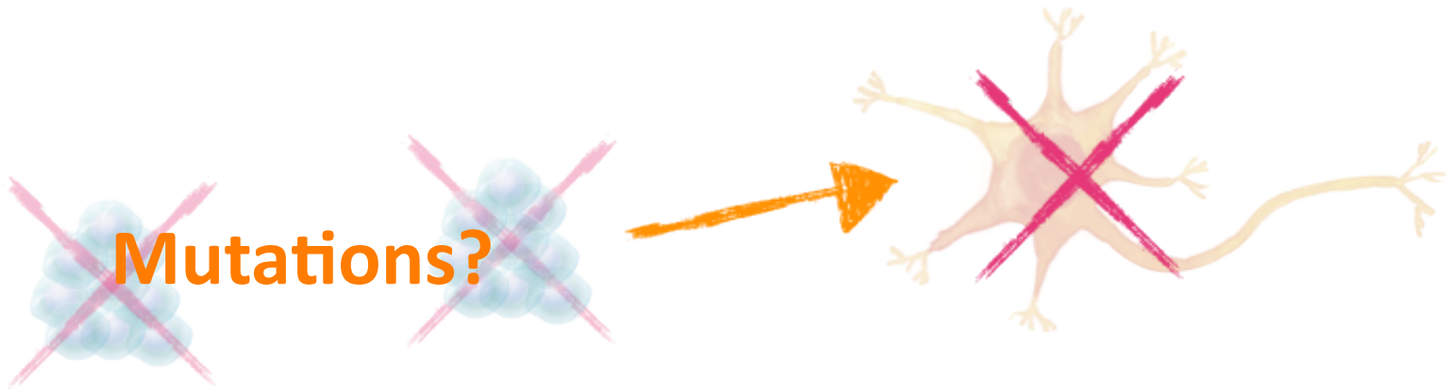
次世代の薬

生細胞

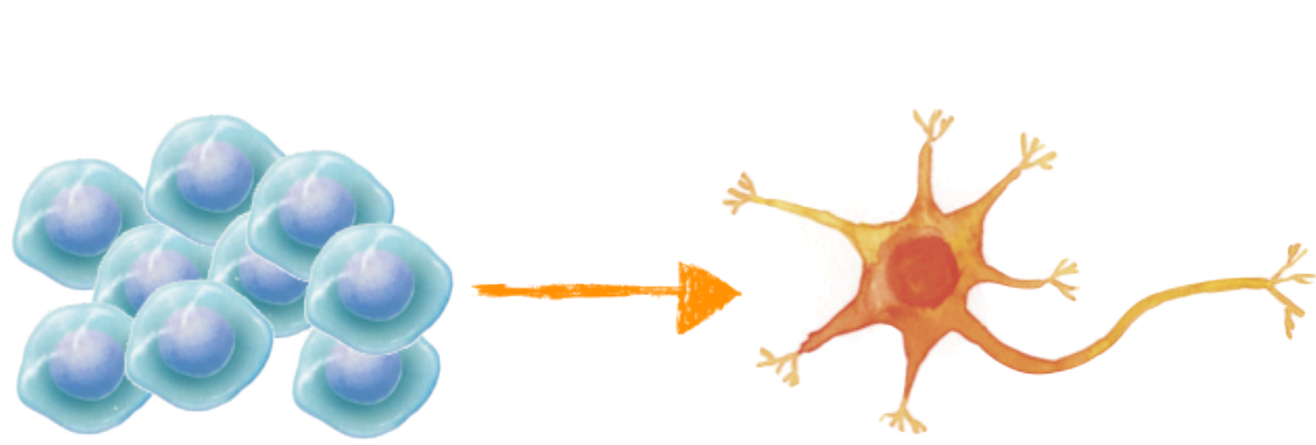
ゲノム・エピゲノム状態の安全性

Safety of iPS Cells

Mutations?



Abberant DNA copy number?



Abnormality in epigenetic profile?

Evaluation of safety of iPS cells? Coming soon!

iPS細胞のクオリティを評価するスタンダードがない！

安全なiPS細胞を選べば

問題ない！

細胞を培養している間にDNAに傷が入る？

iPS細胞に傷があるか

判定することが大切

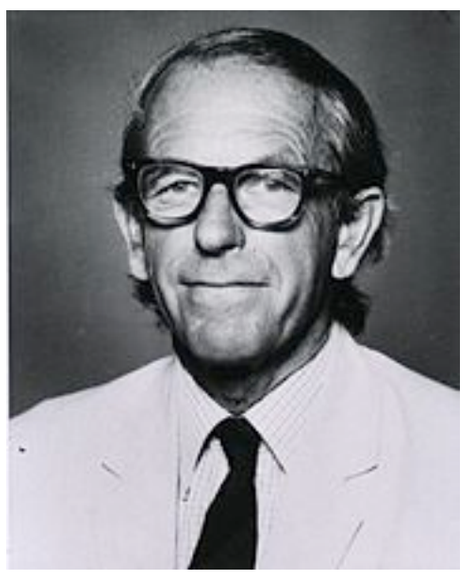
Genome		Epigenome	
sequence	CNV	Gene expression	DNA methylation

Sequence Variations

1975

Sanger Sequence

was developed



Dr. Frederick Sanger (1918-)

Novel Prize in Chemistry 1958 for Protein sequencing of insulin

Novel Prize in Chemistry 1980 for *Sanger sequencing of DNA*

1Q75

Microsoft®

HiSeq2000

illumina

600G塩基 / 2週間

Capillary sequencer：2.4M塩基 / ２週間

Genome		Epigenome	
sequence	CNV	Gene expression	DNA methylation

Sequence Variations

Methodology
SNVs in iPSCs
New Application

Genome		Epigenome	
sequence	CNV	Gene expression	DNA methylation

Sequence Variations

Methodology

SNVs in iPSCs

New Application

Heterogeniety and Genomic Seq

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

Xu et al., Cell 2012

Clear cell renal cell carcinoma is the most common kidney cancer
and has been believed the cancer with very few mutations.

The authors performed single-cell exome sequencing for
- **5** single-cell exome sequencing from adjacent **normal tissues**
- **20** single-cell from the **tumor**

12 mutations within the **normal** population
(average ~20.4 mutations per single normal cell).

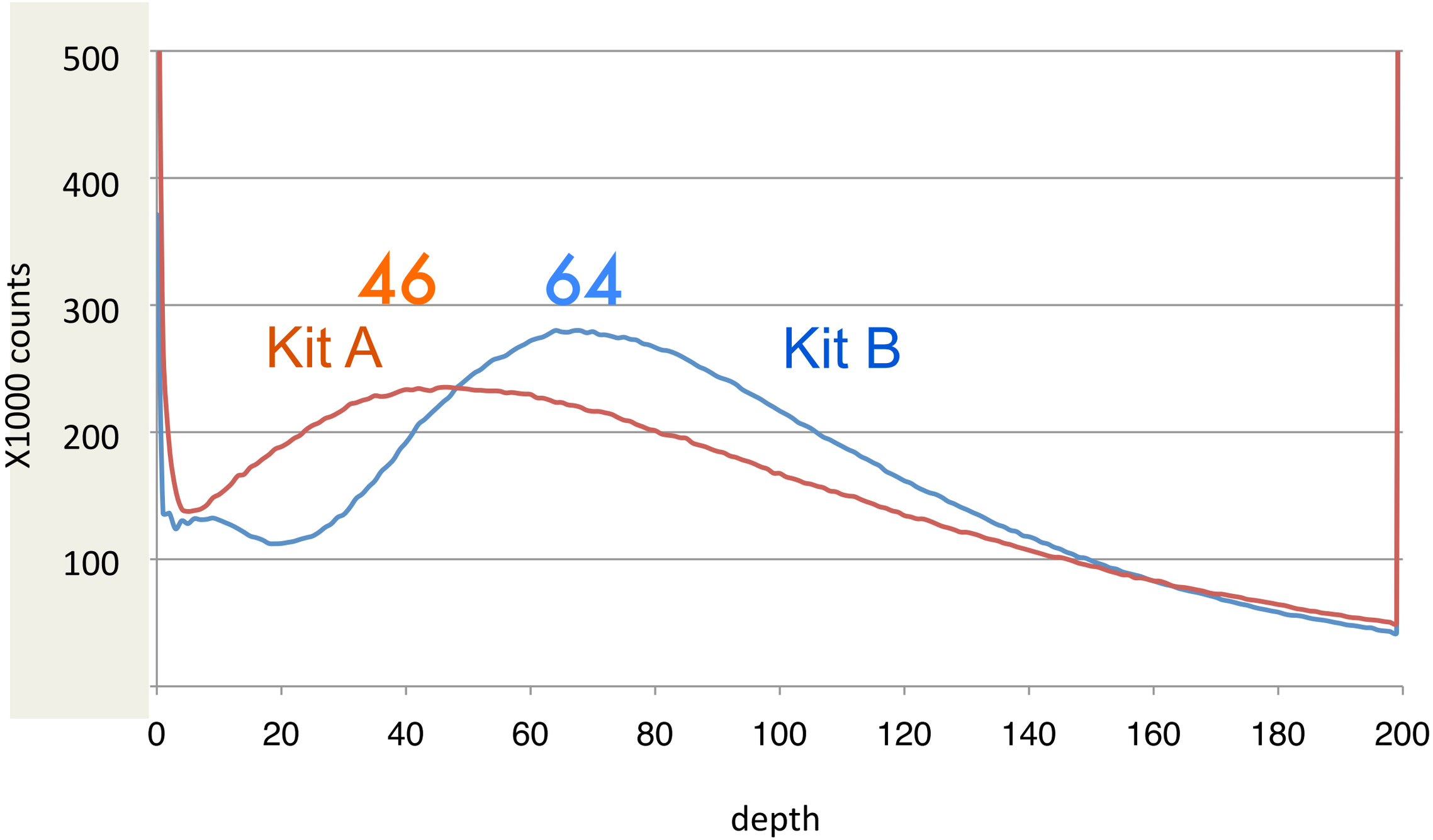
260 mutations between **cancer** and normal population
(average ~78.9 mutations per single cancer cell)

gDNA → Exon enrichment → Deep seq

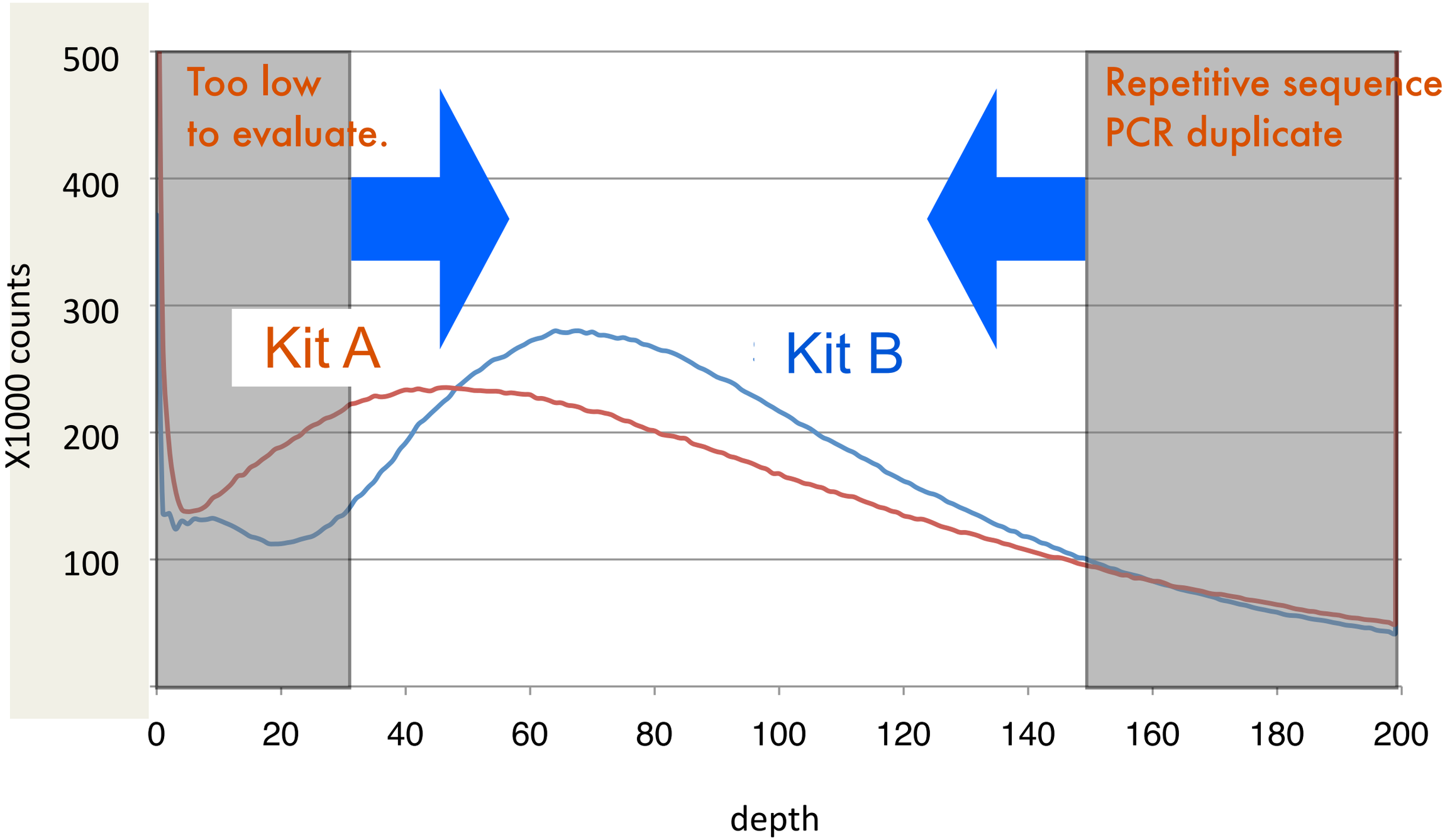
100bpXPE

Mapping → Variation call → SNP database

The Efficient Enrichment



Sharp distribution of the depth.



SNV, Single-nucleotide variation

Personal Variation?

~difference among individuals~

Acquired Mutation?

~difference between original and established cell lines~

Genome		Epigenome	
sequence	CNV	Gene expression	DNA methylation

Sequence Variations

Methodology

SNVs in iPSCs

New Application

Original fibroblast

vs

Established iPSCs

Difference in sequence among the same individual

Result 1

clone-4

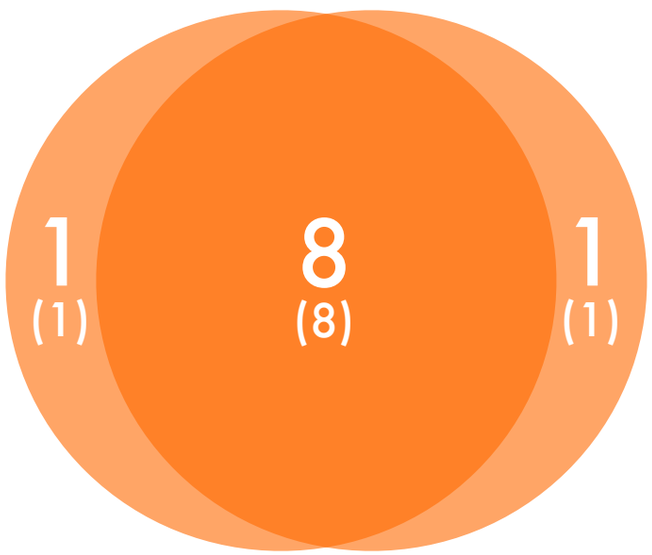
2
(2)

No synonymous mutations
are found in exonic region of clone-5

clone-5

1
(0)

clone-1
total, 9
(non-syn., 9)

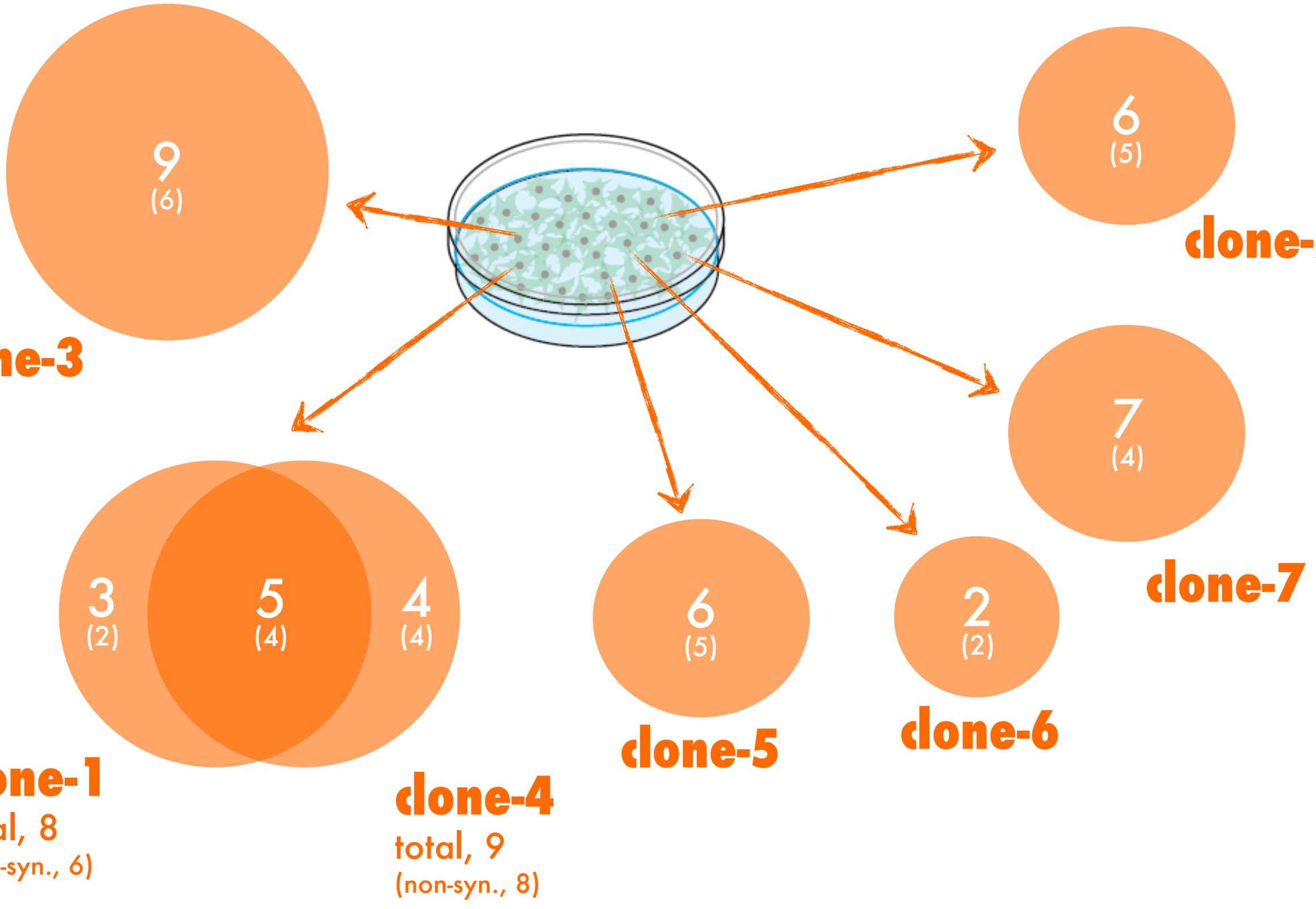


clone-2
total, 9
(non-syn., 9)

clone-6

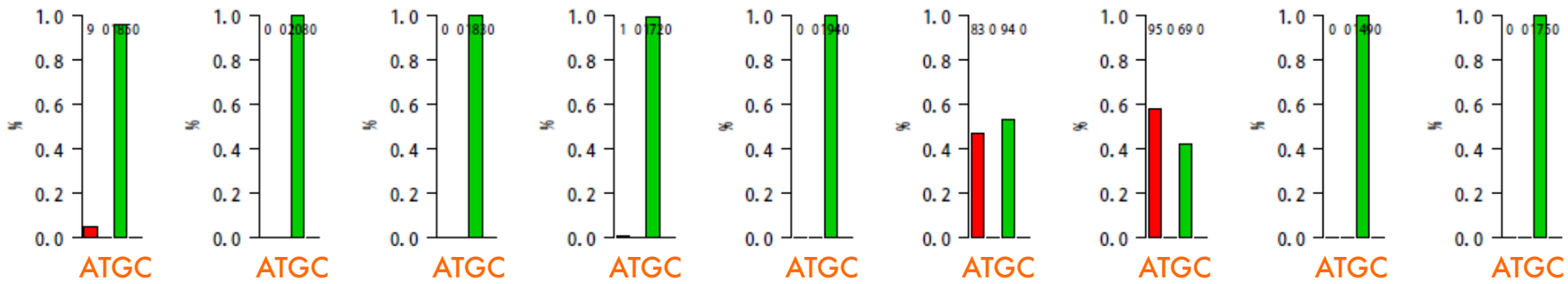
2
(2)

Result 2


9
(6)
clone-3
6
(5)
clone-2
7
(4)
clone-7
3
(2)
5
(4)
4
(4)
clone-1
total, 8
(non-syn., 6)
clone-4
total, 9
(non-syn., 8)
6
(5)
clone-5
2
(2)
clone-6

Somatic Mosaicism

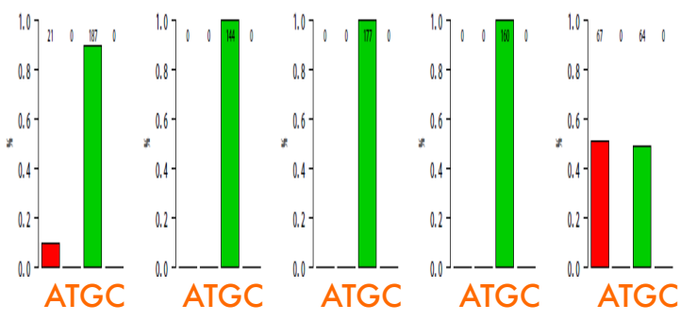
case 1



fibro

iPSC

case 2



fibro

iPSC

Is the sequencing by HiSeq **reliable**?

Is the **depth** enough to validate genotype with **heterogeniety**?

How is the frequency of **sequencing error**?

MiSeq,a personal deep sequencer

300bp/read (PE)

3 Gb seq. in 27 hrs

¥150,000/run

Multi-plex available

PCR

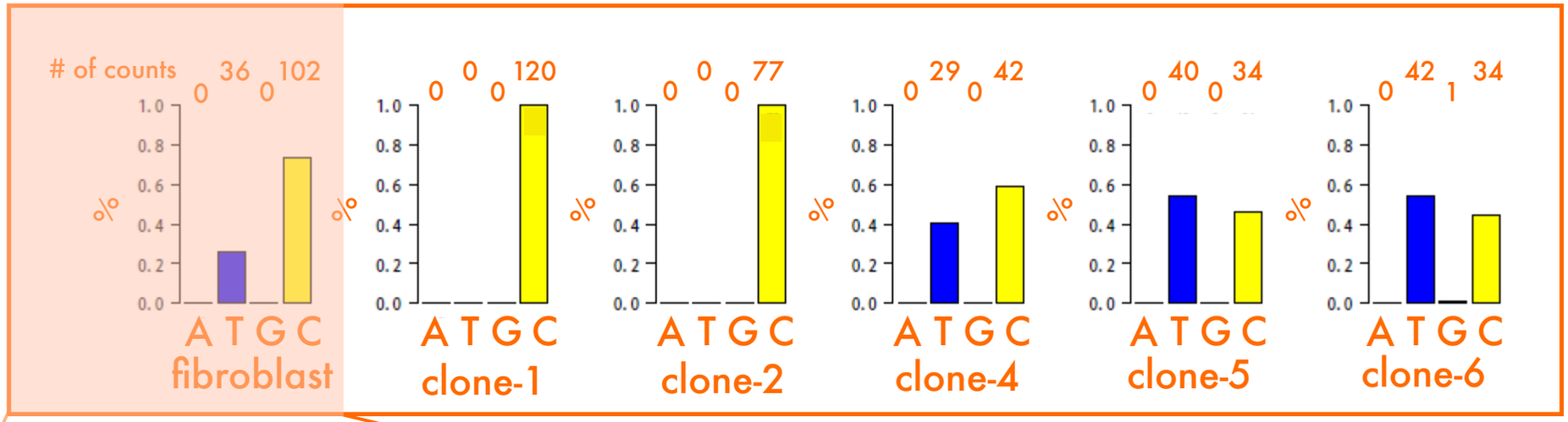
Adaptor-ligation
6 hrs

Multi-plex sequencing
by MiSeq
27 hrs

2 days for **50-100 samples** with **>10,000X depth**

Validation of HiSeq by MiSeq

Exome seq by HiSeq

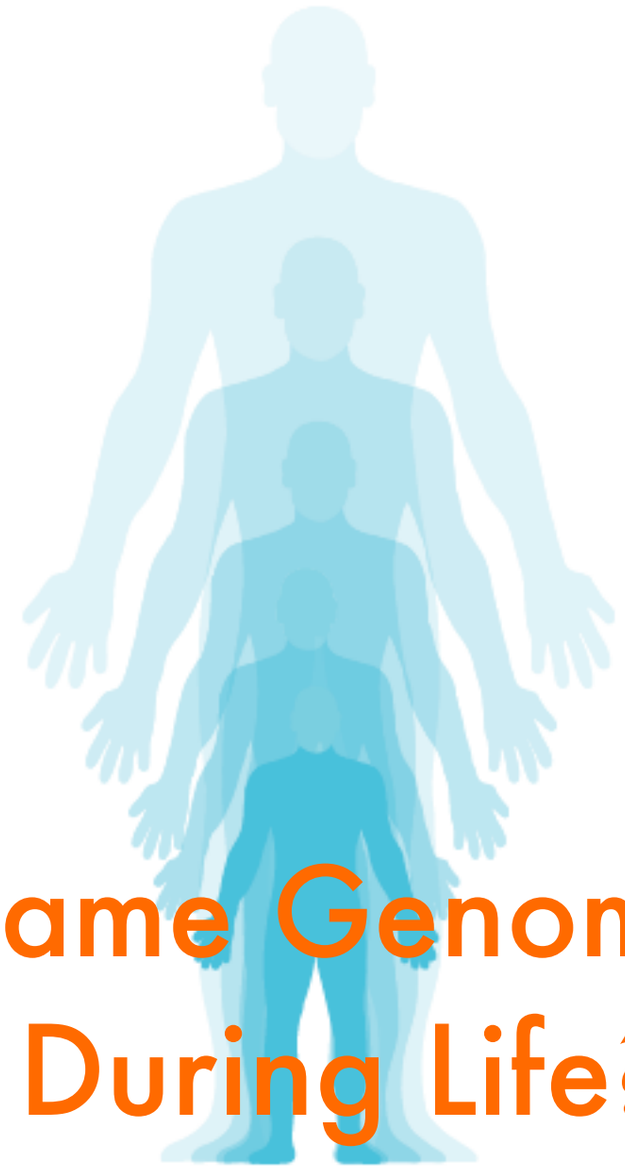


PCR-Seq by MiSeq

NCBI reference	T (next to the target)				C				G (next to the target)			
count (%)	A	T	G	C	A	T	G	C	A	T	G	C
	0 (0%)	88973 (~100%)	0 (0%)	10 (~0%)	8 (~0%)	26044 (29.1%)	0 (0%)	63361 (70.9%)	6 (~0%)	8 (~0%)	89138 (~100%)	0 (0%)

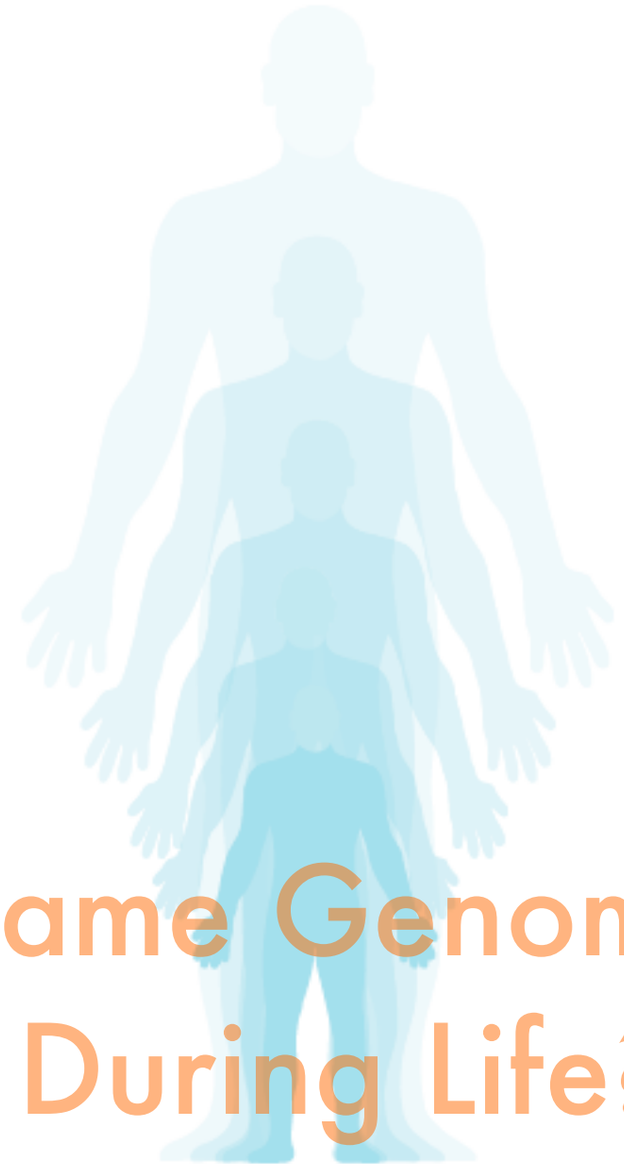
sequencing error or contamination?

We have



Same Genome
During Life?

We have



Same Genome
During Life?

No, Variable Genome

Personal Variation?

~difference among individuals~

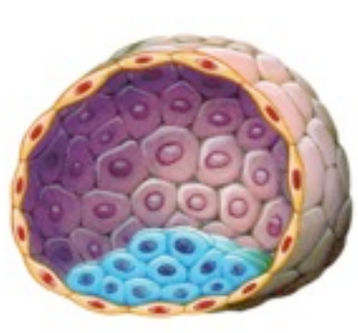
Acquired Mutation?

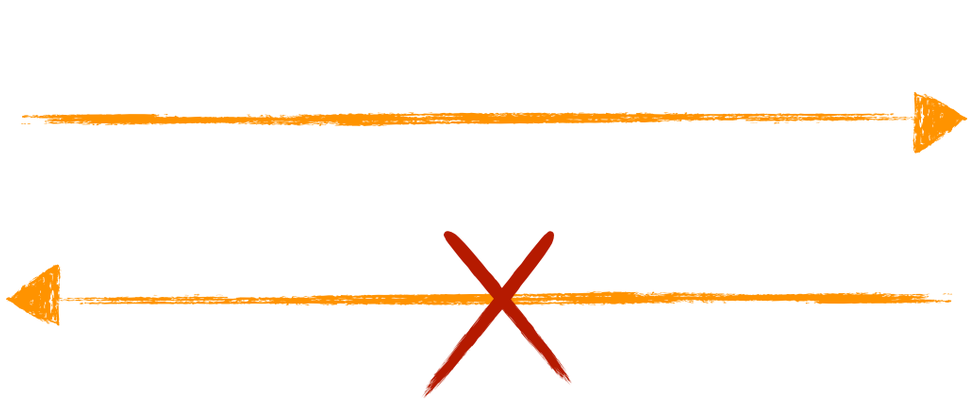
~difference between original and established cell lines~

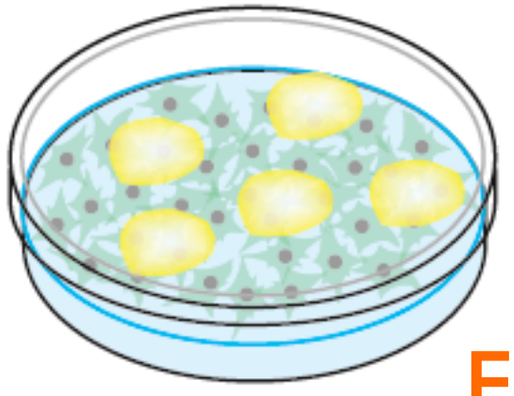
Somatic Mosaicism

~difference in one individual~

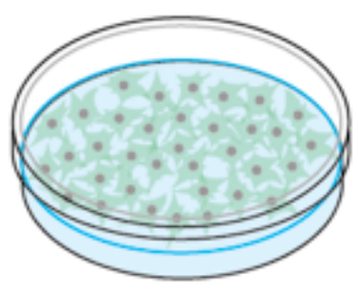
Origin of ES cells can not be accessible

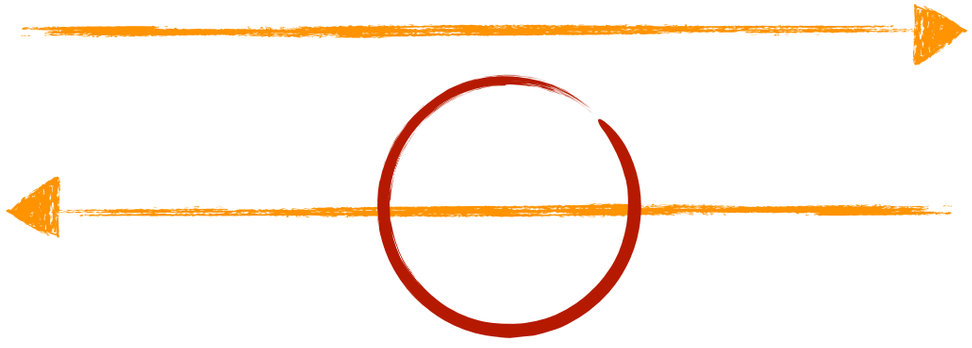


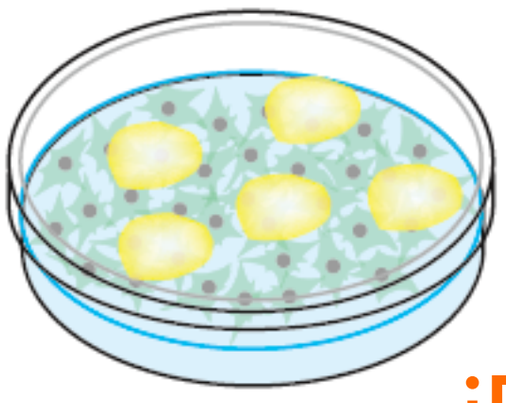




ESC

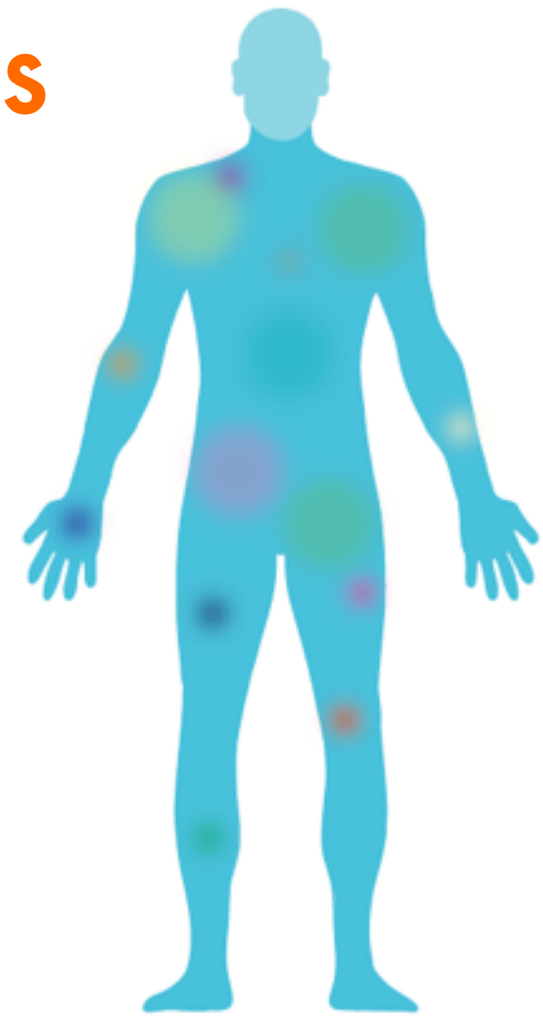






iPSC

Mosaicism should be considered
not only for iPSCs but also for ESCs



We have to compare the cell
model with paired original cells

Genome Diagnosis

Finding

50 sample/2weeks
depth >50

Validation

100 sample/2days
depth >10,000

Strategy

2 weeks

Exome by HiSeq

screening

1 weeks

Variation Call

3 days

Validation by Deep-seq with MiSeq

high-resolution analysis with depth >50000

Genome		Epigenome	
sequence	CNV	Gene expression	DNA methylation

Sequence Variations

Methodology

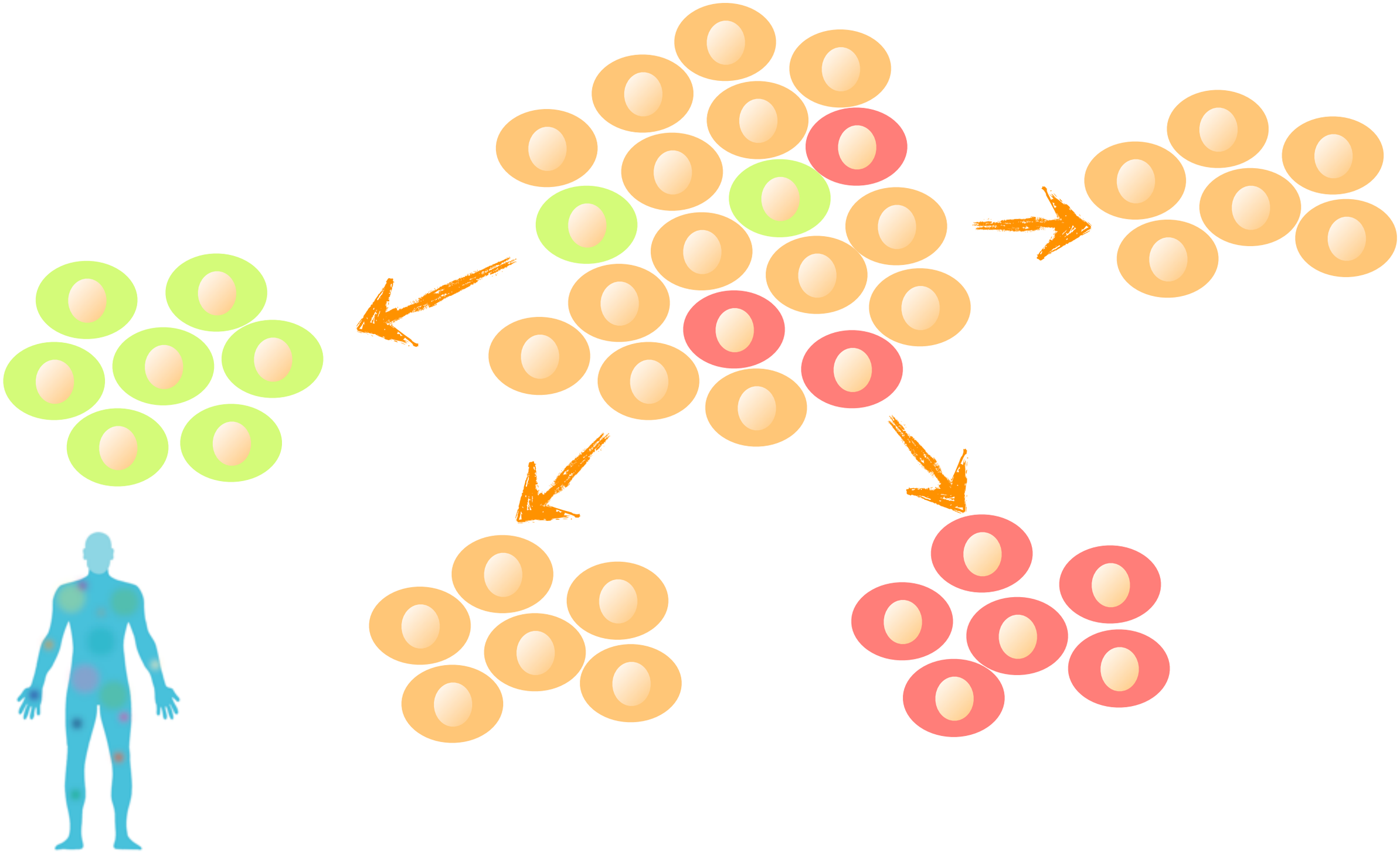
SNVs in iPSCs

New Application

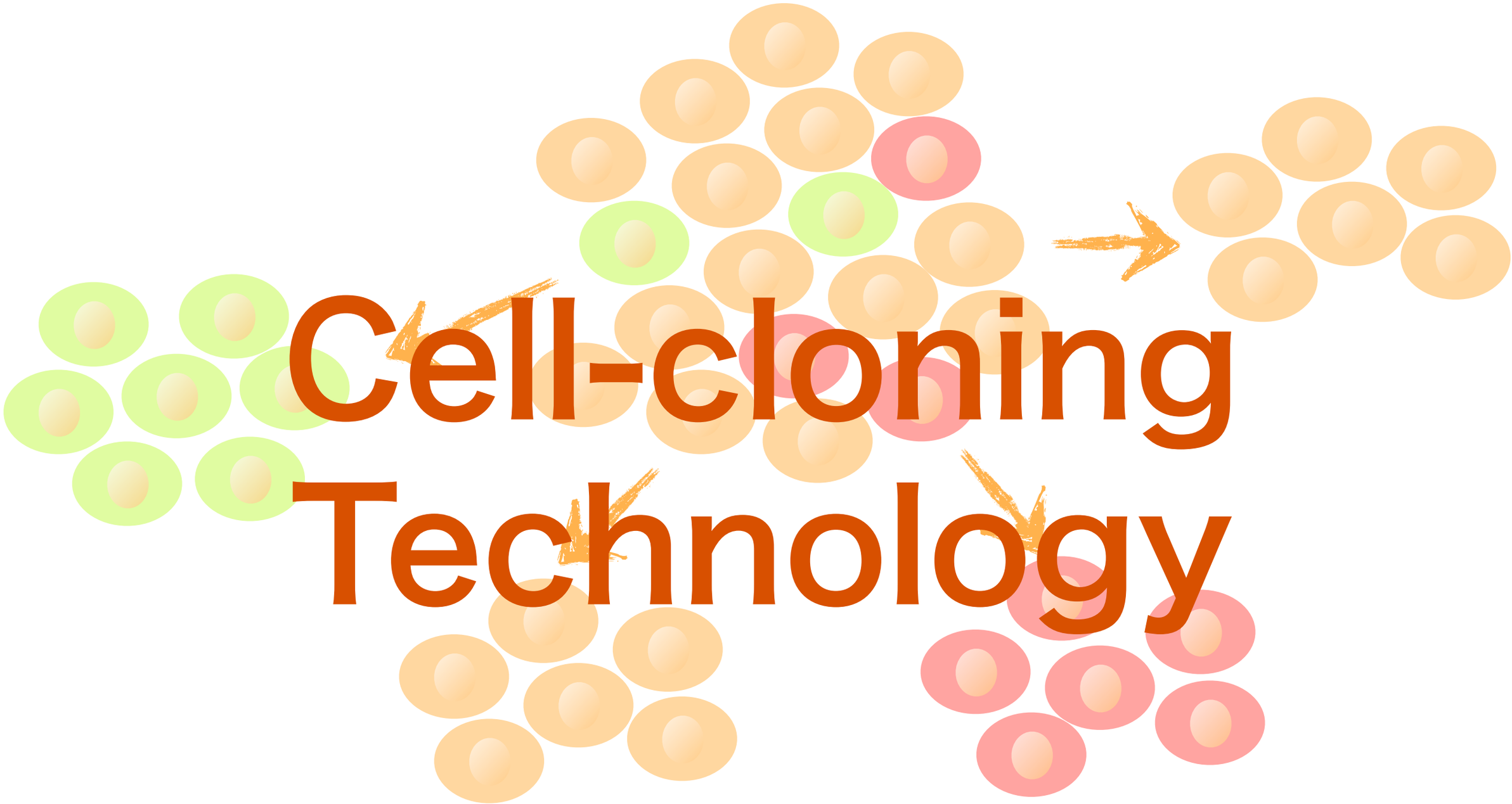
iPS cells,

A tool for cloning of the cell

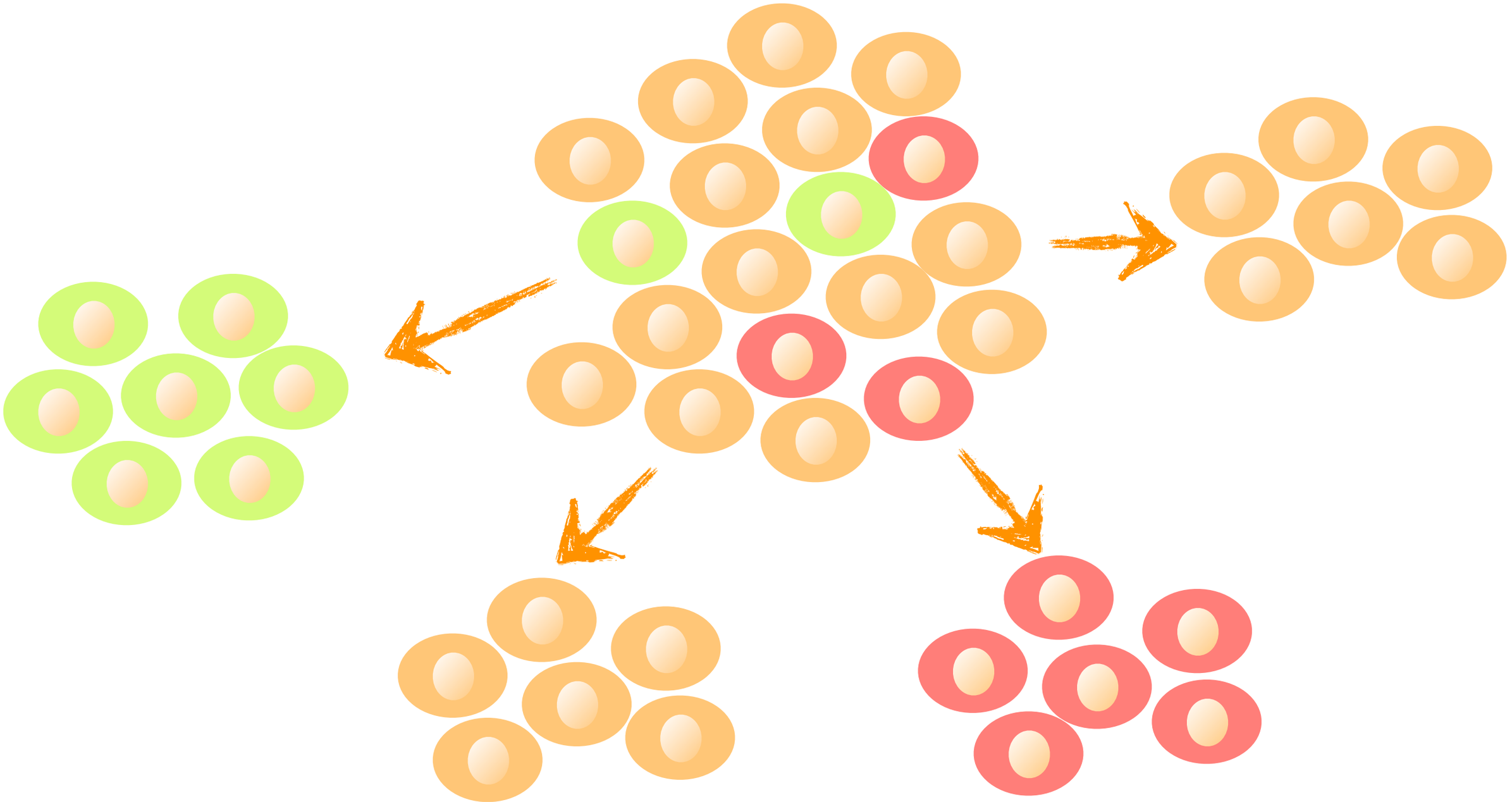
Heterogeniety in human Genome



Heterogeniety in human Genome



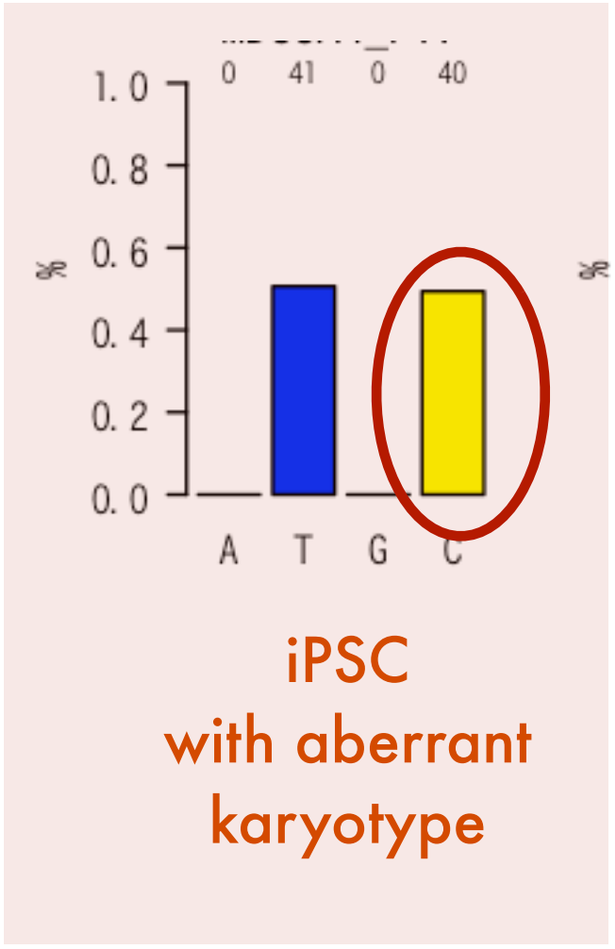
Cell-cloning Technology



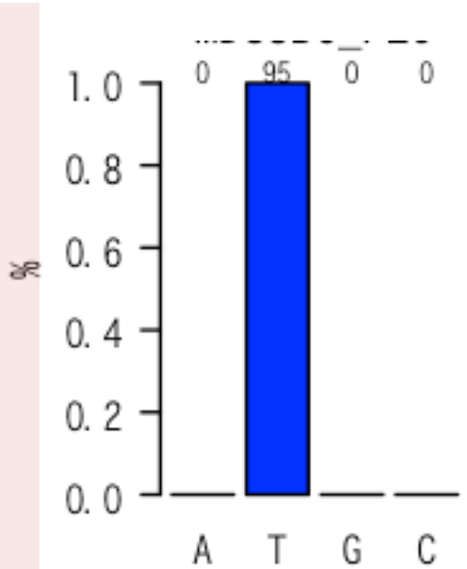
Normal cells with
normal phenotype

Pre-cancerous cells with
aberrant phenotype

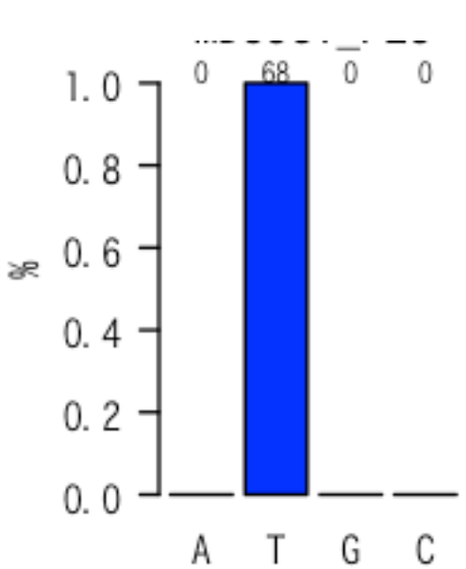
A Mutation in RTK



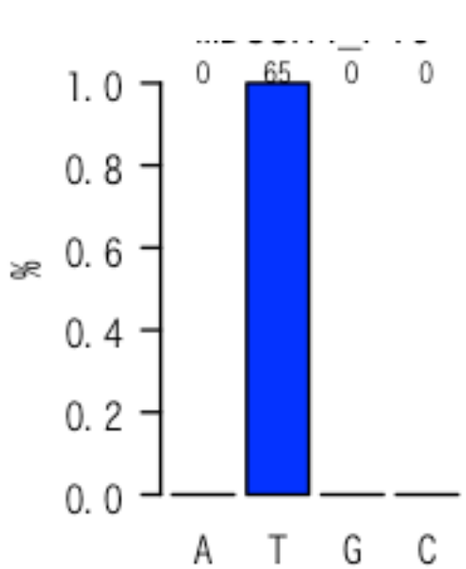
iPSC
with aberrant
karyotype



iPSC
with normal
karyotype



iPSC
with normal
karyotype



iPSC
with normal
karyotype

Y to H
at a phosphorylation target site

Sequence Analysis Opens New Drus Screening

US President Nixon Started *War on Cancer* from 1970's.

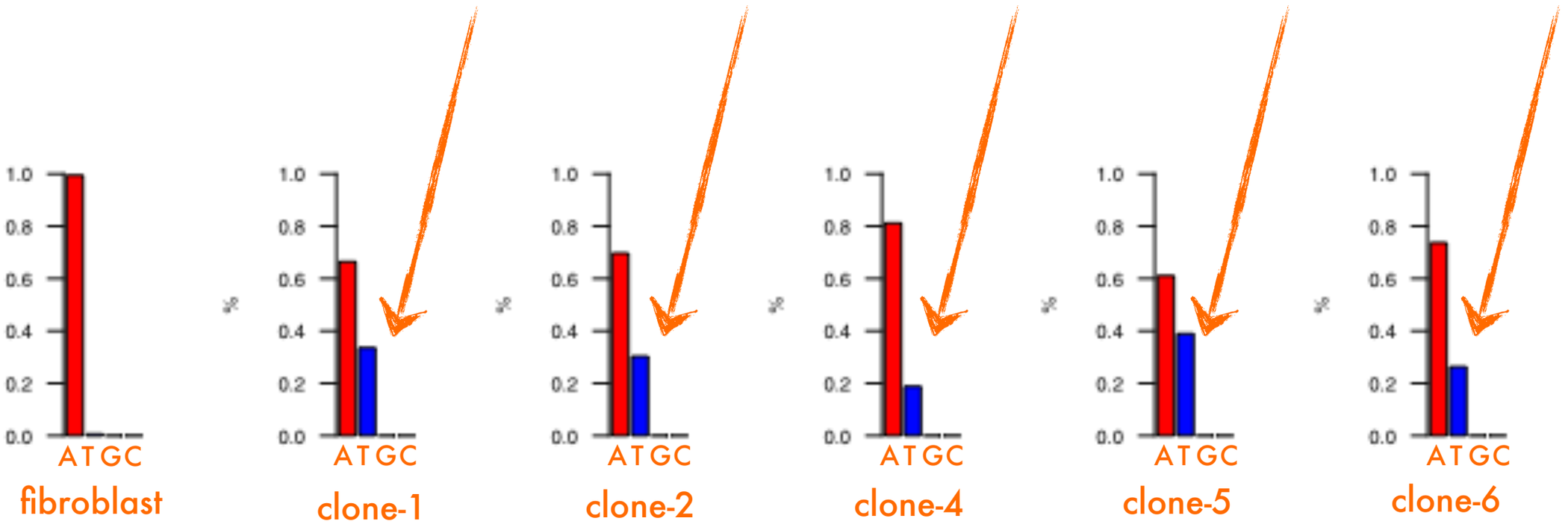
The molecular-based approach

realizes **Gleevec**,

with a dramatic impact on leukemia.

Contamination of mouse feeder genome

Contamination of mouse feeder cells



Bioinfotmatics

No Consensus Algorithms to Detect Variations.

Each Program for variation call provides different results.

Mapping to Reference Genome (Human)

Variant calling by Samtools

Variant calling by GATK

Variant calling by CLC-bio

Quality Filter

Quality Filter

Quality Filter

Samtools

GATK

GATK-unfiltered

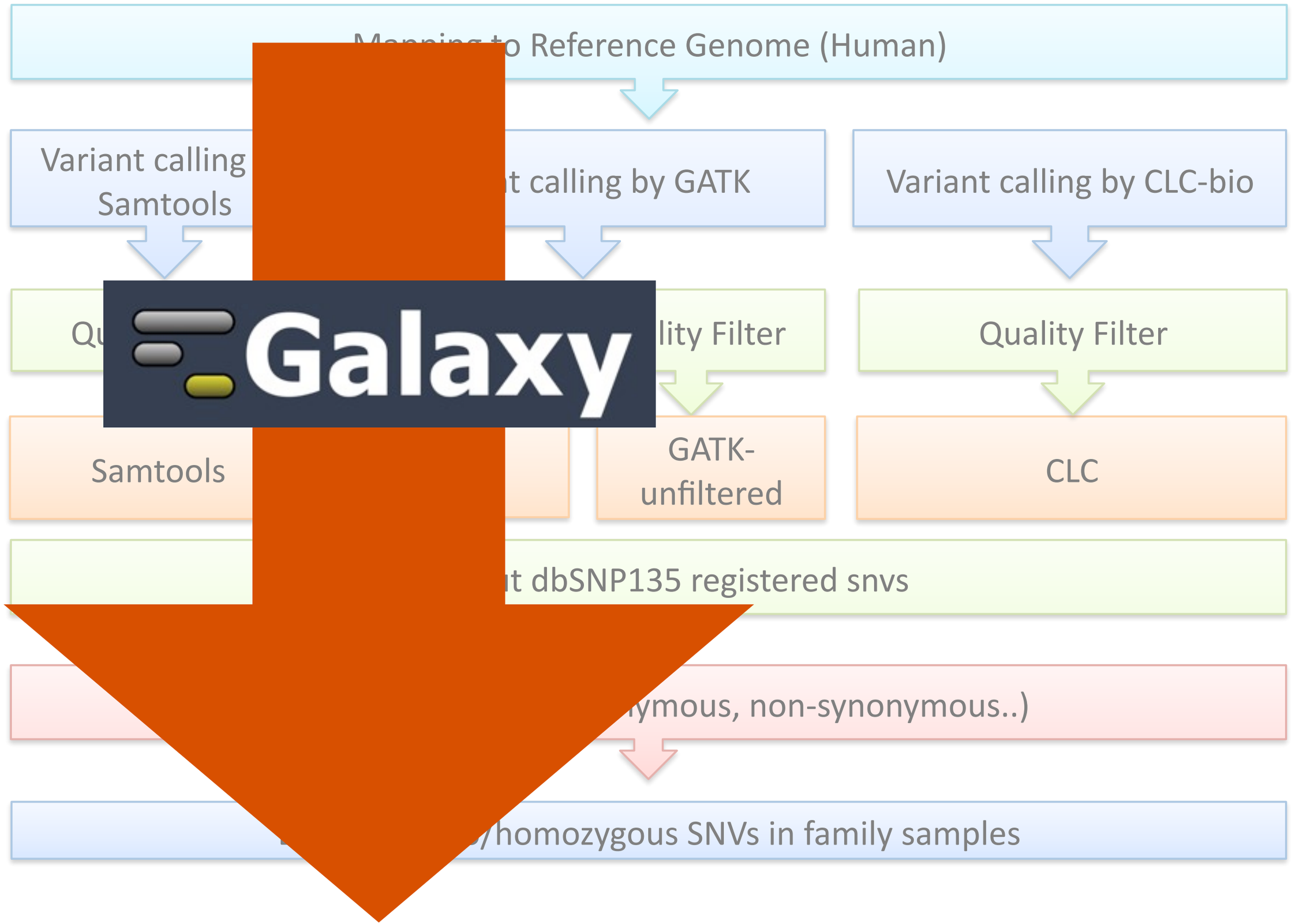
CLC

Filter out dbSNP135 registered snvs

Annotate variants (synonymous, non-synonymous..)

Detect hetero/homozygous SNVs in family samples

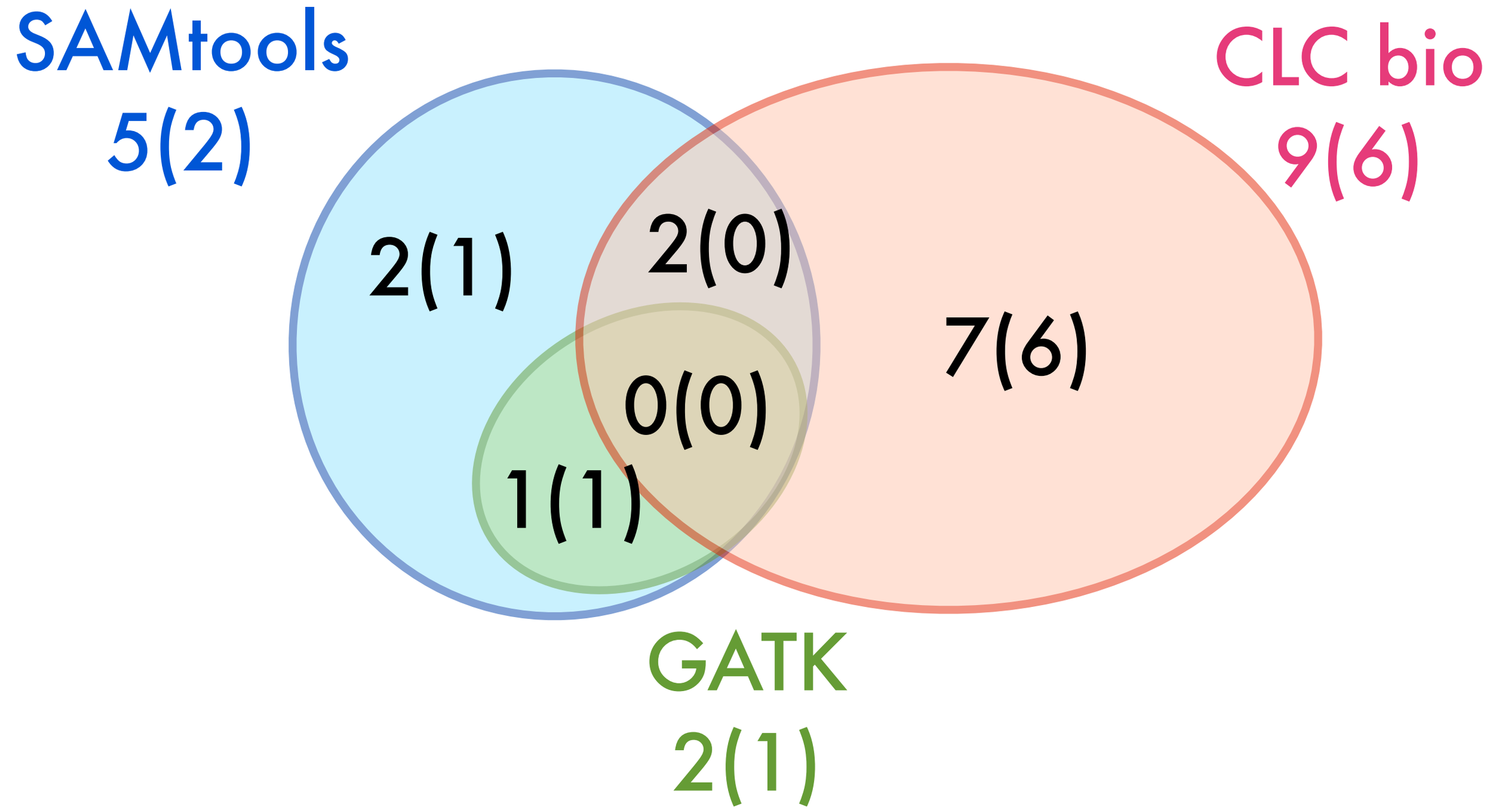
Automatic Analysis



Different Results in Variation Call with Different Algorithm

66 loci (42 target sites)	SAMtools	GATK	CLC
Validated/detected	5/6	2/4	9/14
accuracy rate	**83%**	**50%**	**64%**

Different Results in Variation Call with Different Algorithm



Genome		Epigenome	
sequence	CNV	Gene expression	DNA methylation

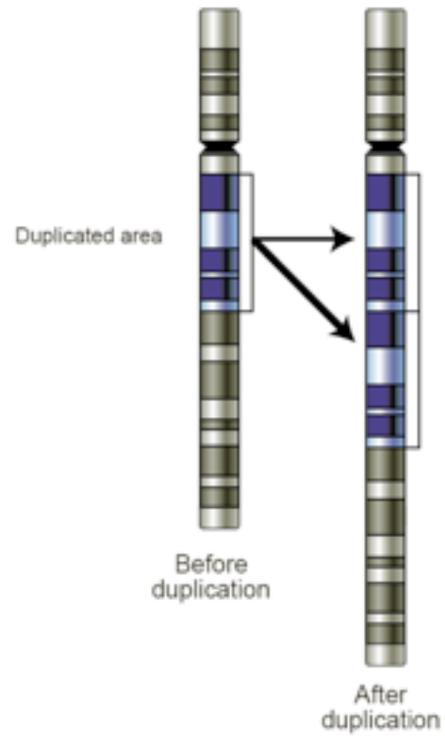
Structural Variations

Copy Number Alterations

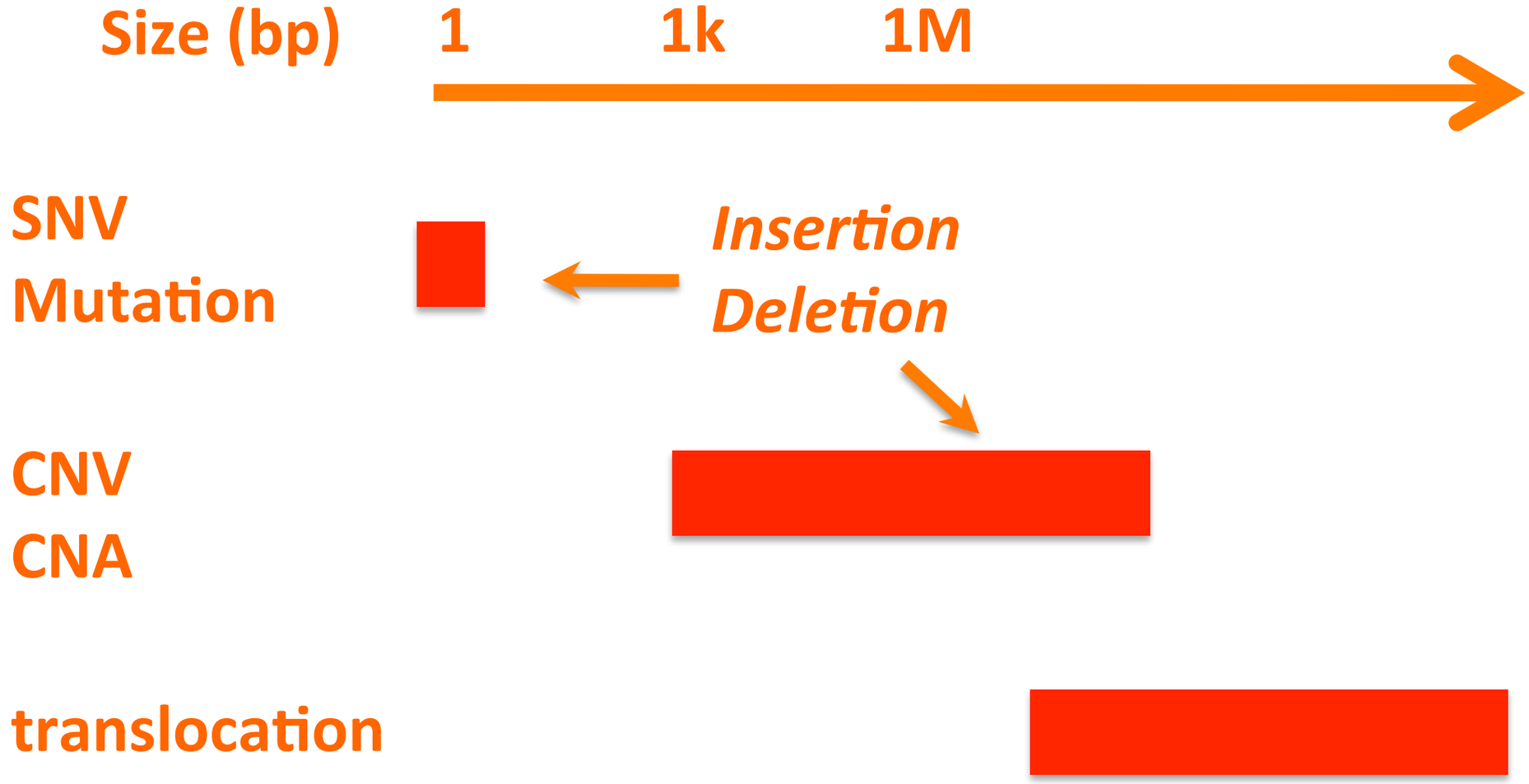
Possible mechanism for differential gene expression

Copy number variations (CNVs), a form of structural variation, are alterations of the DNA of a genome that results in the cell having an abnormal number of copies of one or more sections of the DNA.

CNVs may leads dosage imbalances in gene expression.

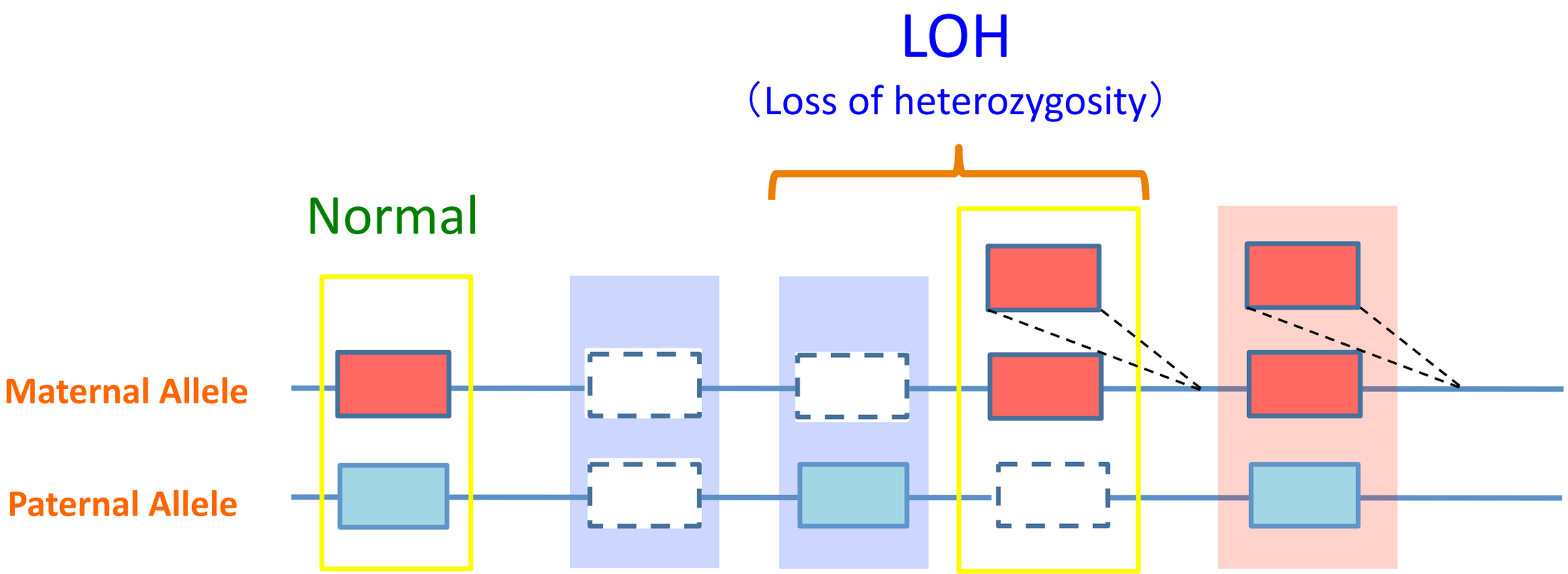


Genomic Alterations



SNV: single nucleotide variation
CNV: copy number variation
CNA: copy number aberration

Allele-specific CNV detection



Allele-specific	1	0	0	2	2
Copy Number	1	0	1	0	1
Copy Number	2	0	1	2	3

....

Deletion Deletion Duplication

Copy-neutral LOH

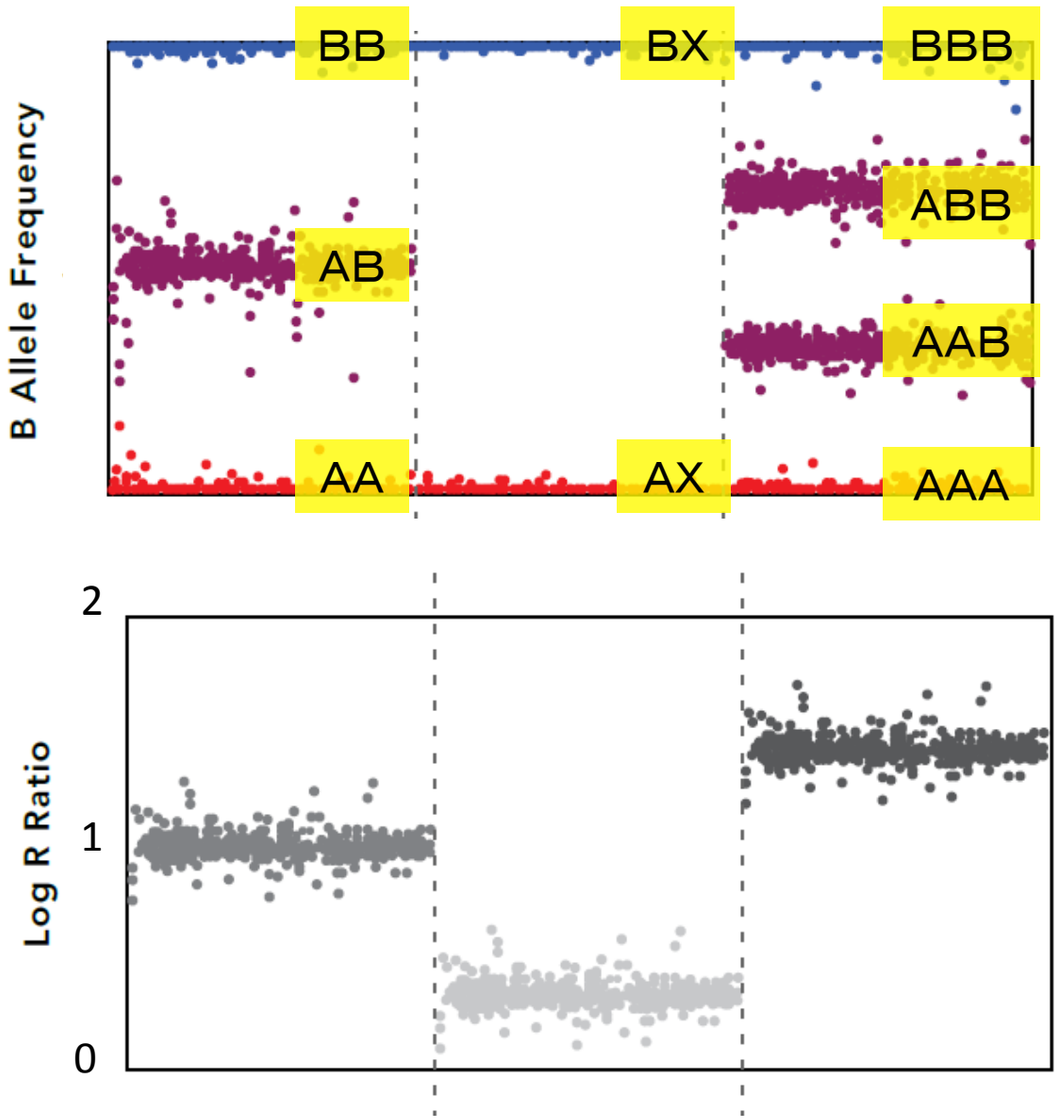
CNV

by SNP genotyping array

CNV Call by SNP Genotyping

Affymetrix SNP6.0

illumina Omni1

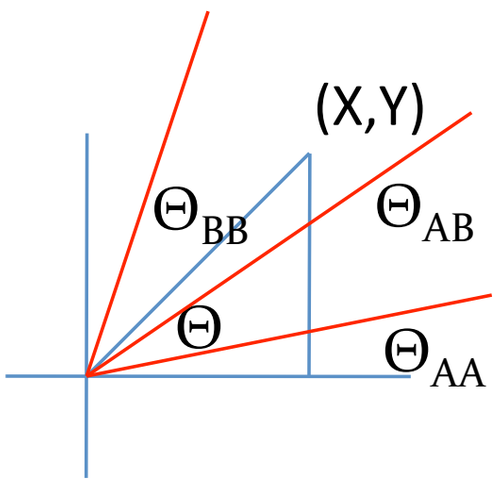


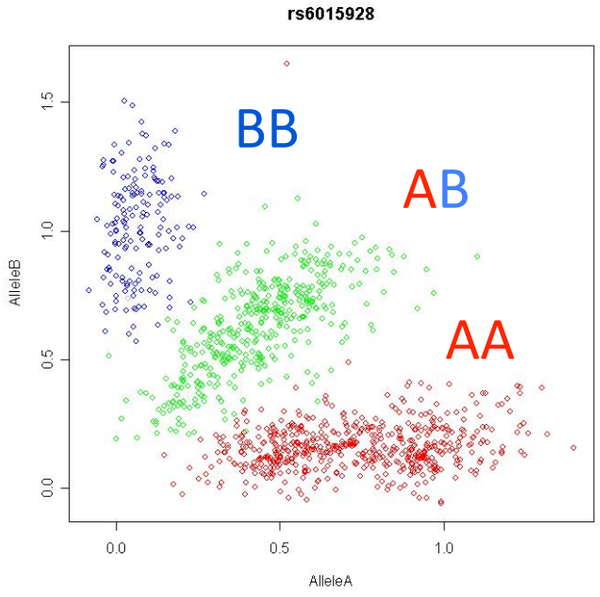
Accurate Model
Using log R ratio and B Allele Freq.

Hidden Markov Model designed for high resolution CNV detection in whole genome SNP genotyping data.

Log R ratio (LRR): total fluorescent intensity signals from both sets of probe/allele at each SNP.

B Allelle Frequency (BAF) : relative ratio of the intensity signals between two probes/allele at each SNP.





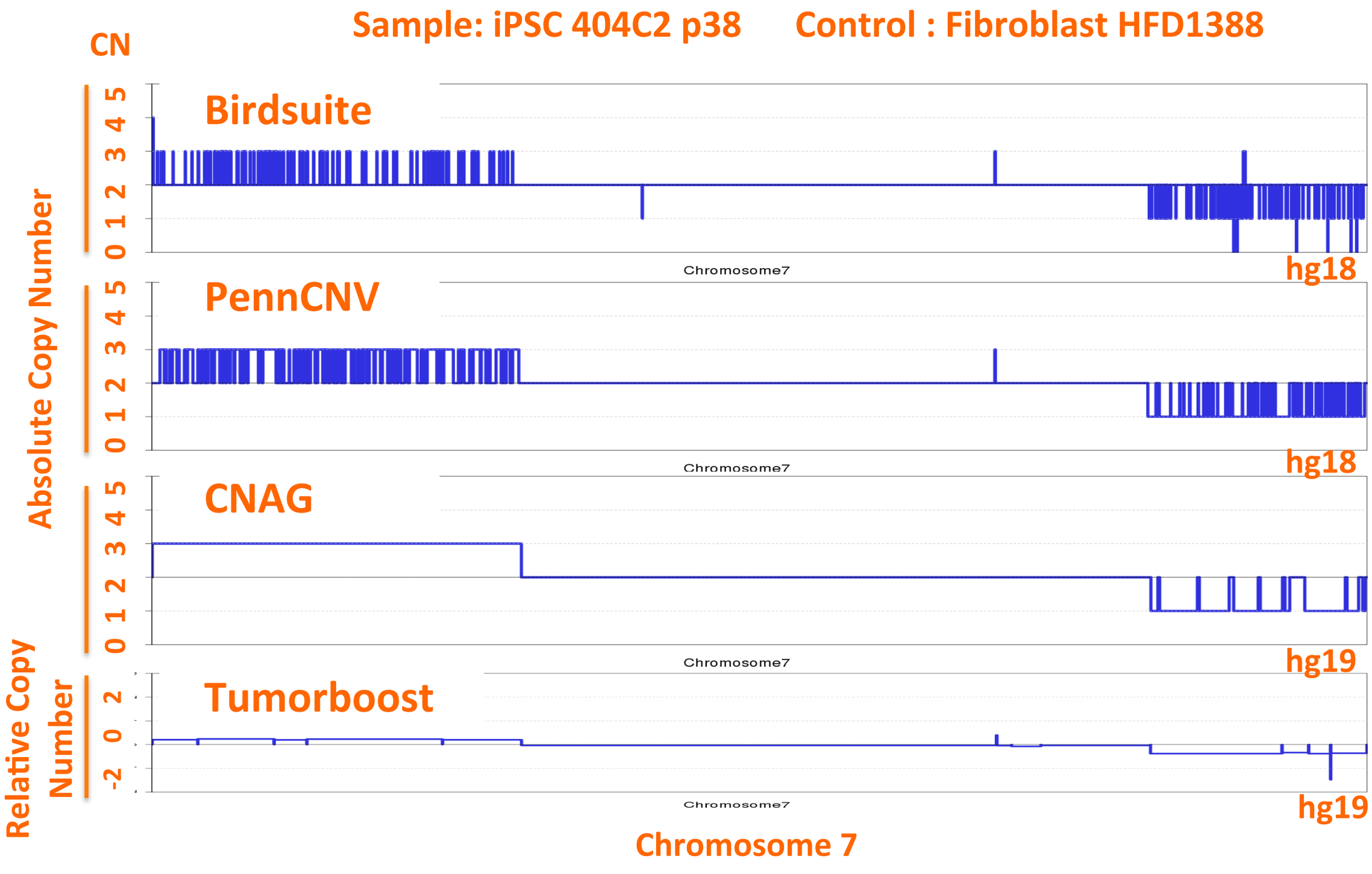
X, Y : normalized signal intensity
R = X+Y : total signal intensity
Θ = arctan(Y/X)/(π/2)

Wang et al., Genome Res., 2007

Combination of CNV Analysis

Platforms		Algorithm
SNP6.0 Affymetrix	X	PennCNV
Omni1M illumina		CNAG
Omni2.5M illumina		TumorBoost

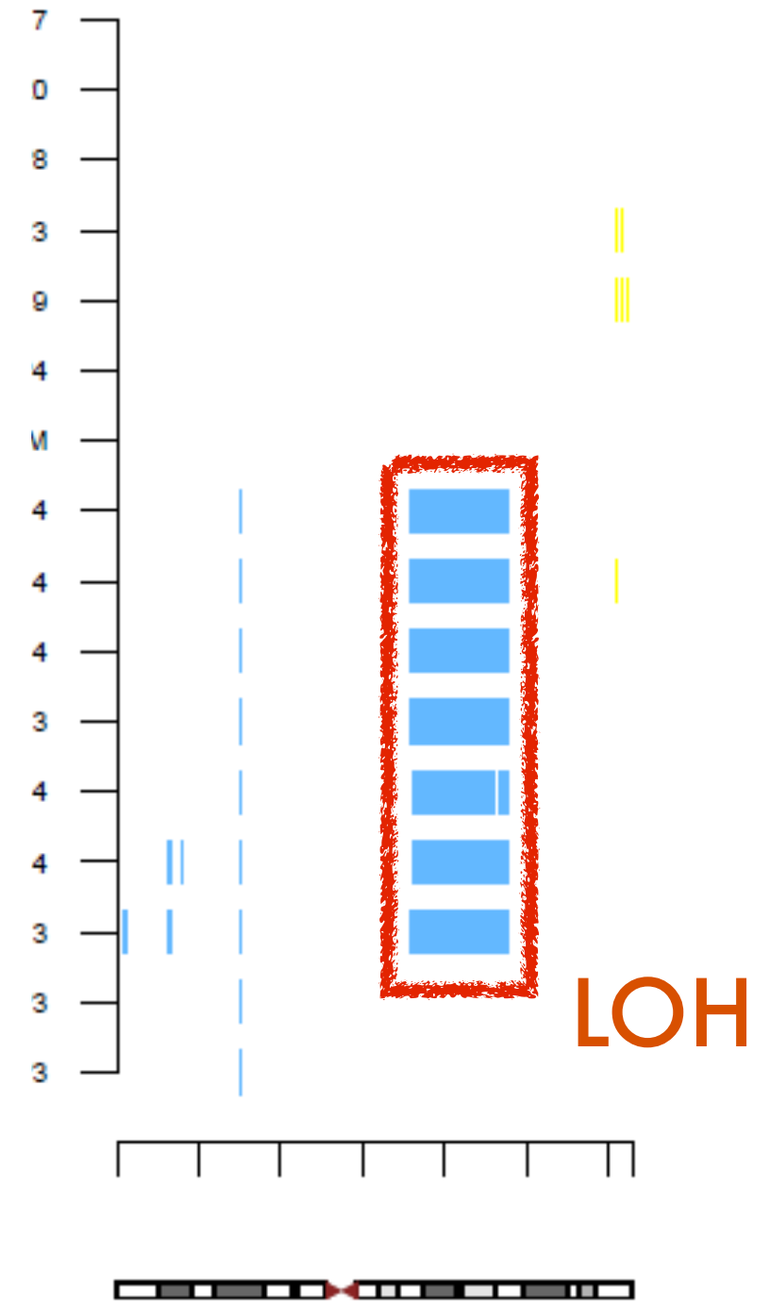
Different CNVs by Algorithms



CNV Affects SNVs

We detected many SNVs (hetero to homo) on this region.

But they may be due to loss o heterozygosity.



Available but Platform-dependent Algorithms

Program	analysis type	allele specific	Affymetrix	illumina
PennCNV	single	NA	✓	✓
BirdSuite	single	✓	✓	NA
CNVPartition	single	NA	NA	✓
Tumorboost	Pair*	NA	✓	✓
CNAG	Pair*	✓	✓	NA
Our Program	Pair*	✓	✓	✓

*CNV is calculated by subtraction of signal of control sample (fibroblast).

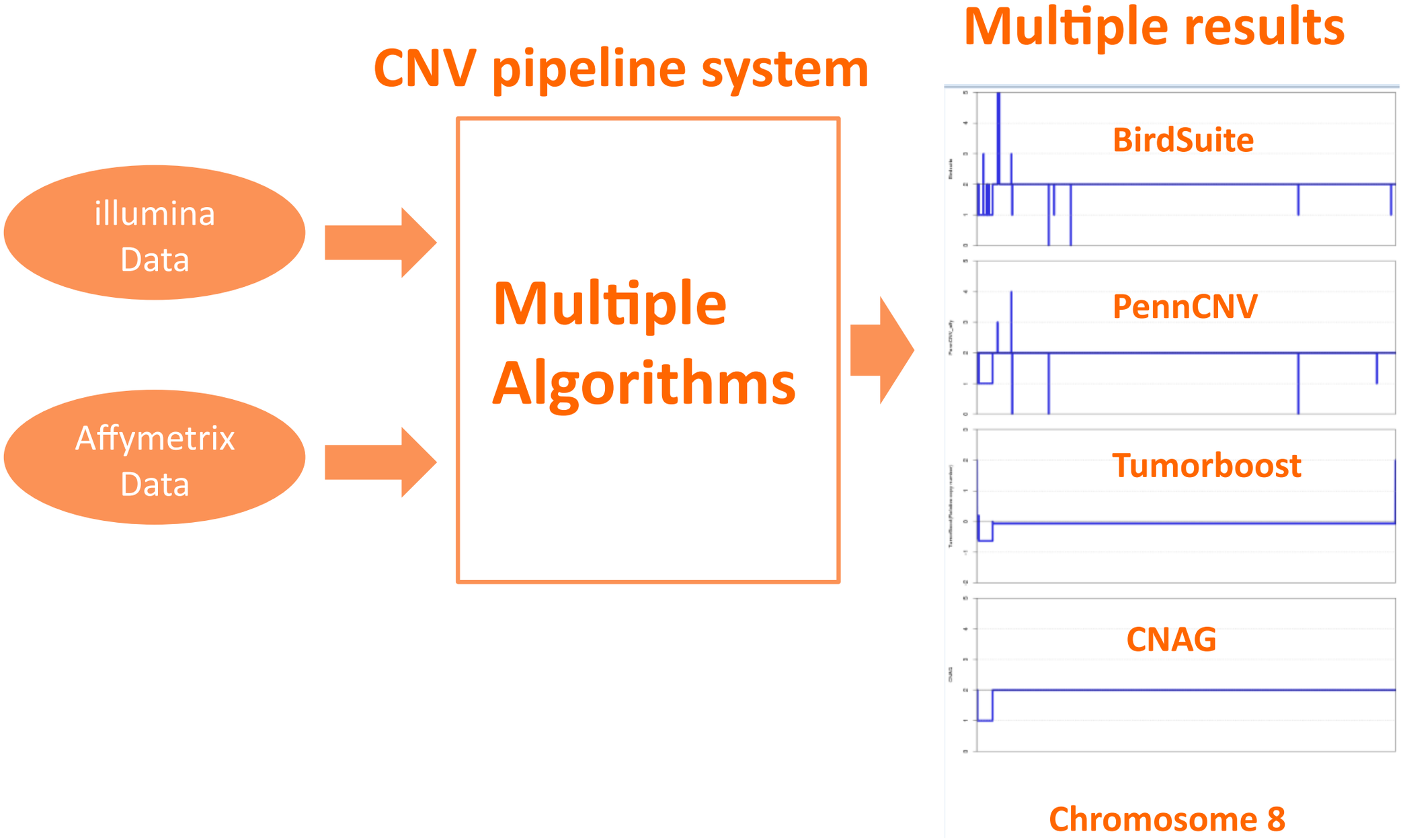
Comparison of Algorithms

Number of Overlapping Regions

Algorithm (detected region)	BirdSuite	PennCNV_affy	Tumorboost	CNAG
Birdsuite (7Mb)	-	81.5%	99.9%	88.3%
PennCNV_affy (50Mb)	10.8%	-	100%	95.1%
Tumorboost (77Mb)	8.6%	65.1%	-	93.7%
CNAG (72Mb)	8.1%	66.0%	100.0%	-

Sample: iPSC Control : human fibroblast

Multiple CNV Detection System

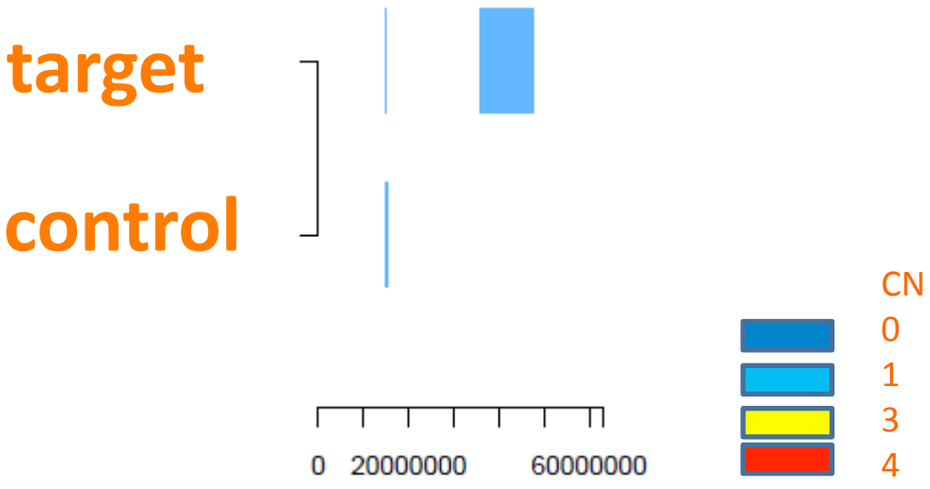


Next-gen. CNV call

SNV and CNV call at once

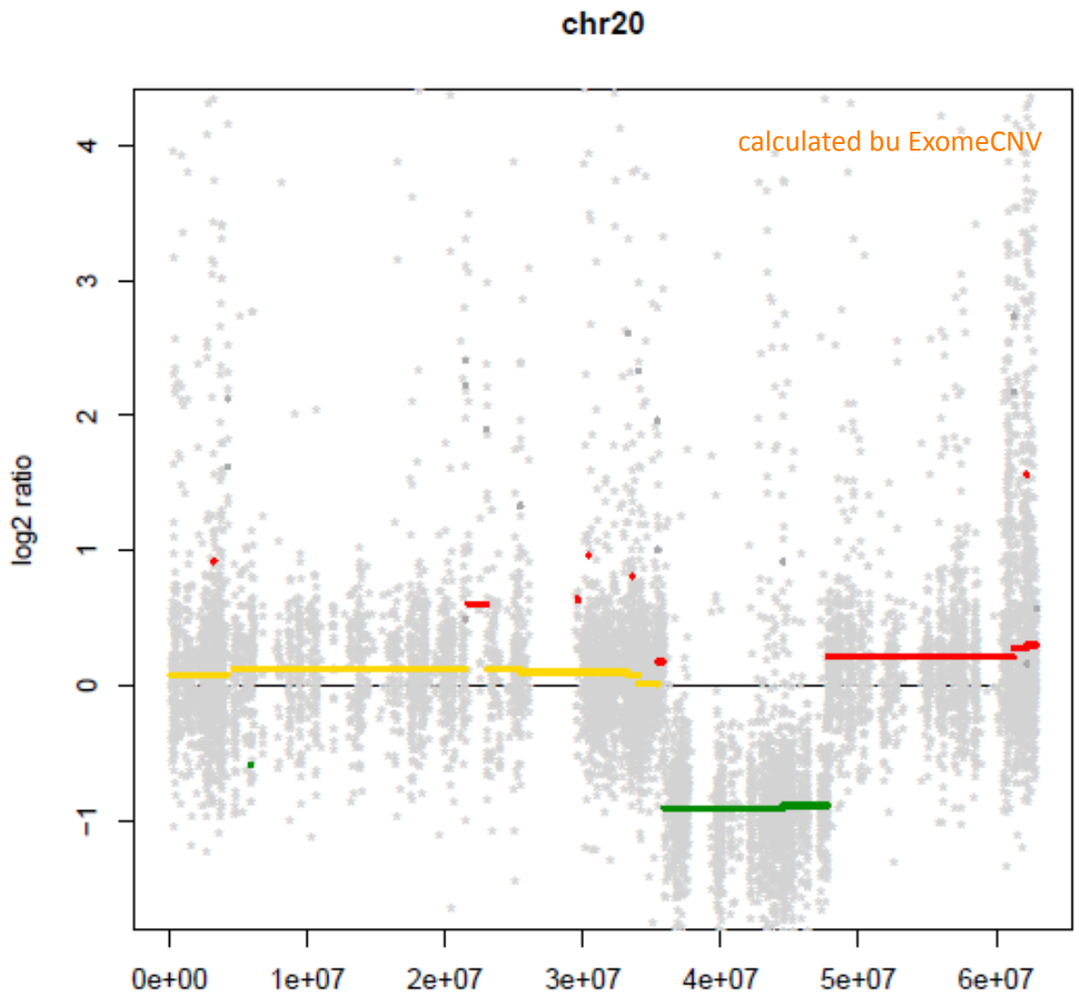
CNV Detection Using Exome Data

SNP Array



Exome Sequencing

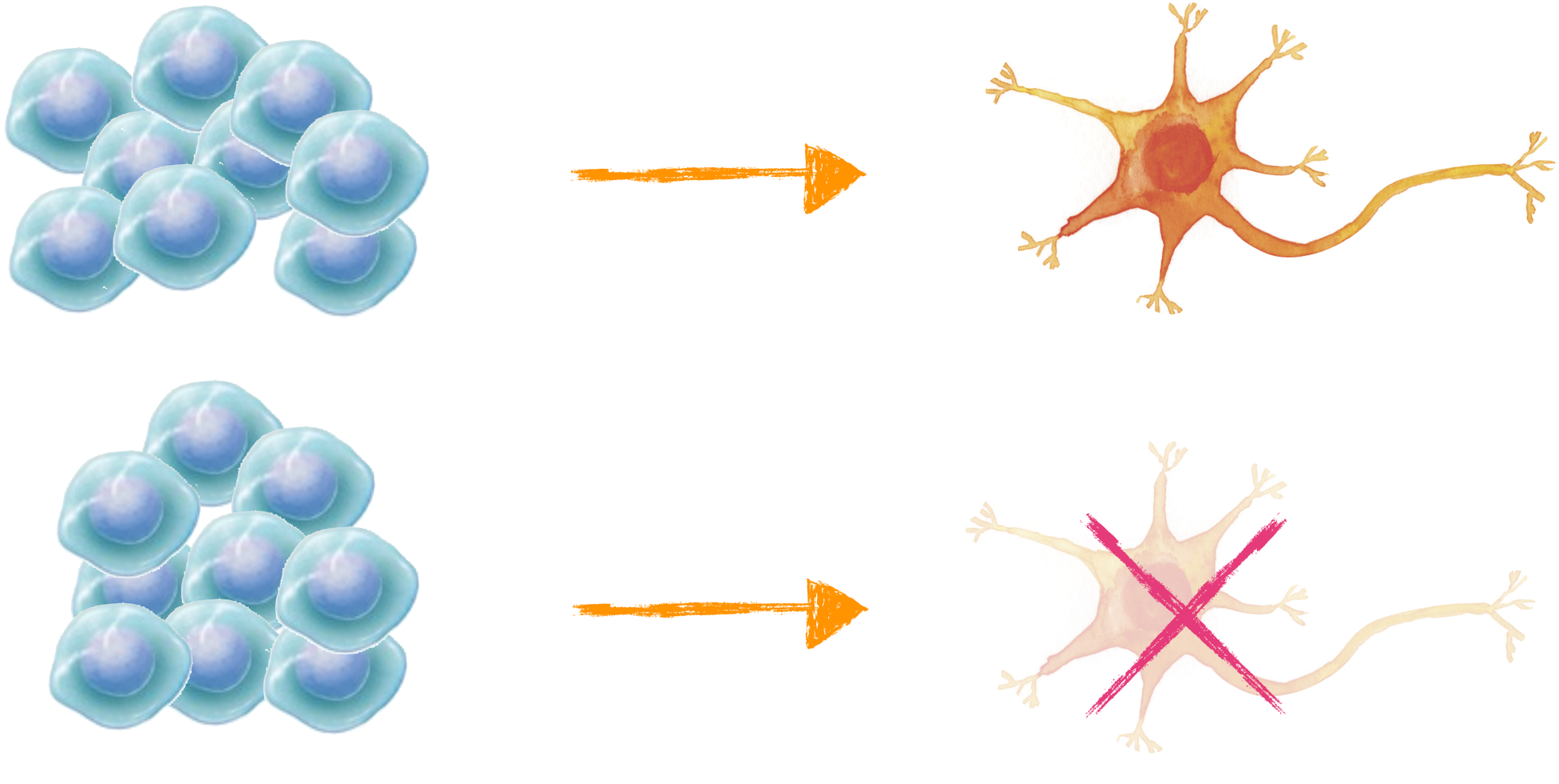
$$\frac{\text{\# of Tag [target A]}}{\text{\# of Tag [Control]}}$$



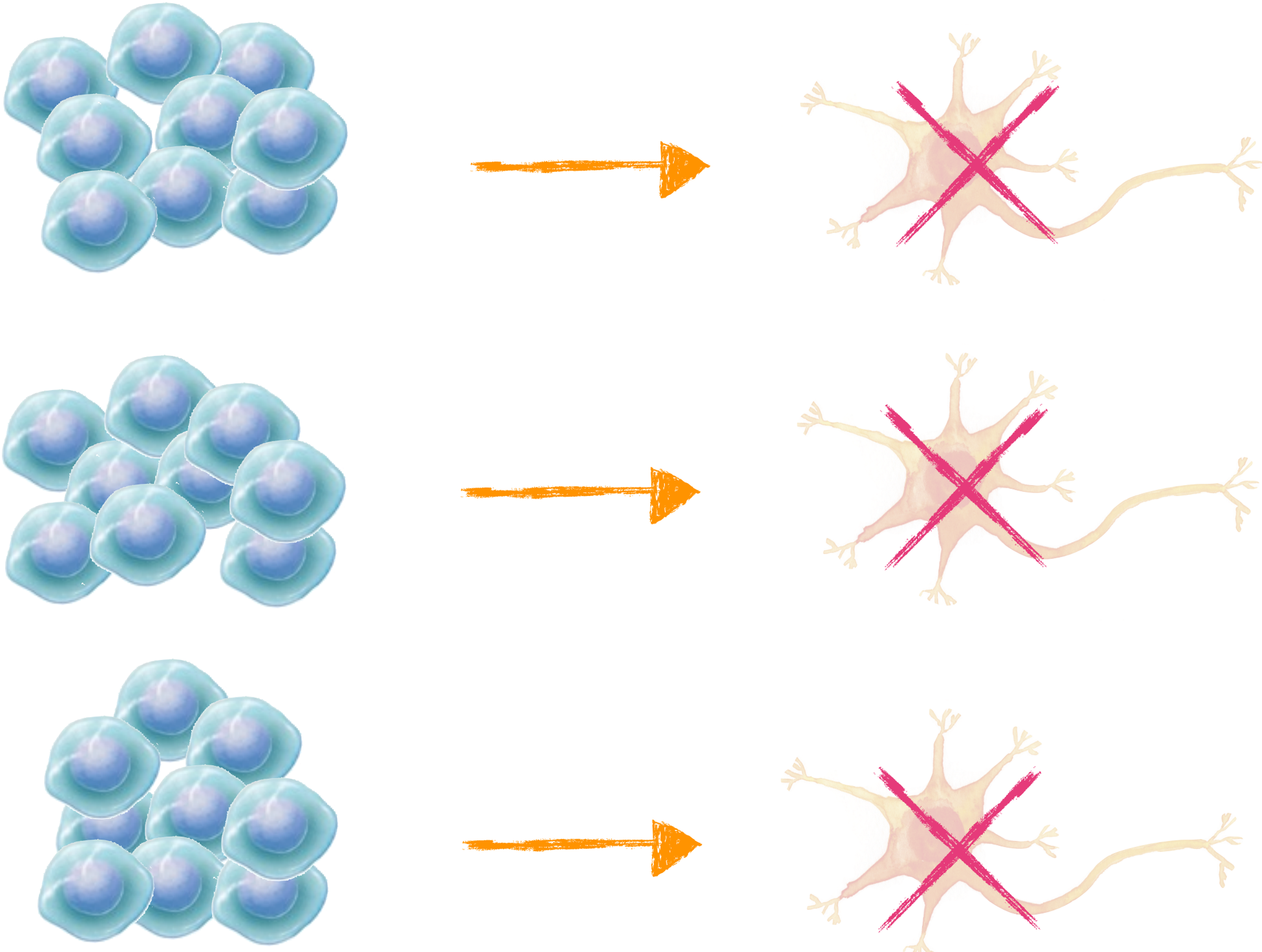
Sequence-based CNV detection such as exome and whole genome re-seq. can detect not only small CNV but also break point of DNA copy number.

Call for Members

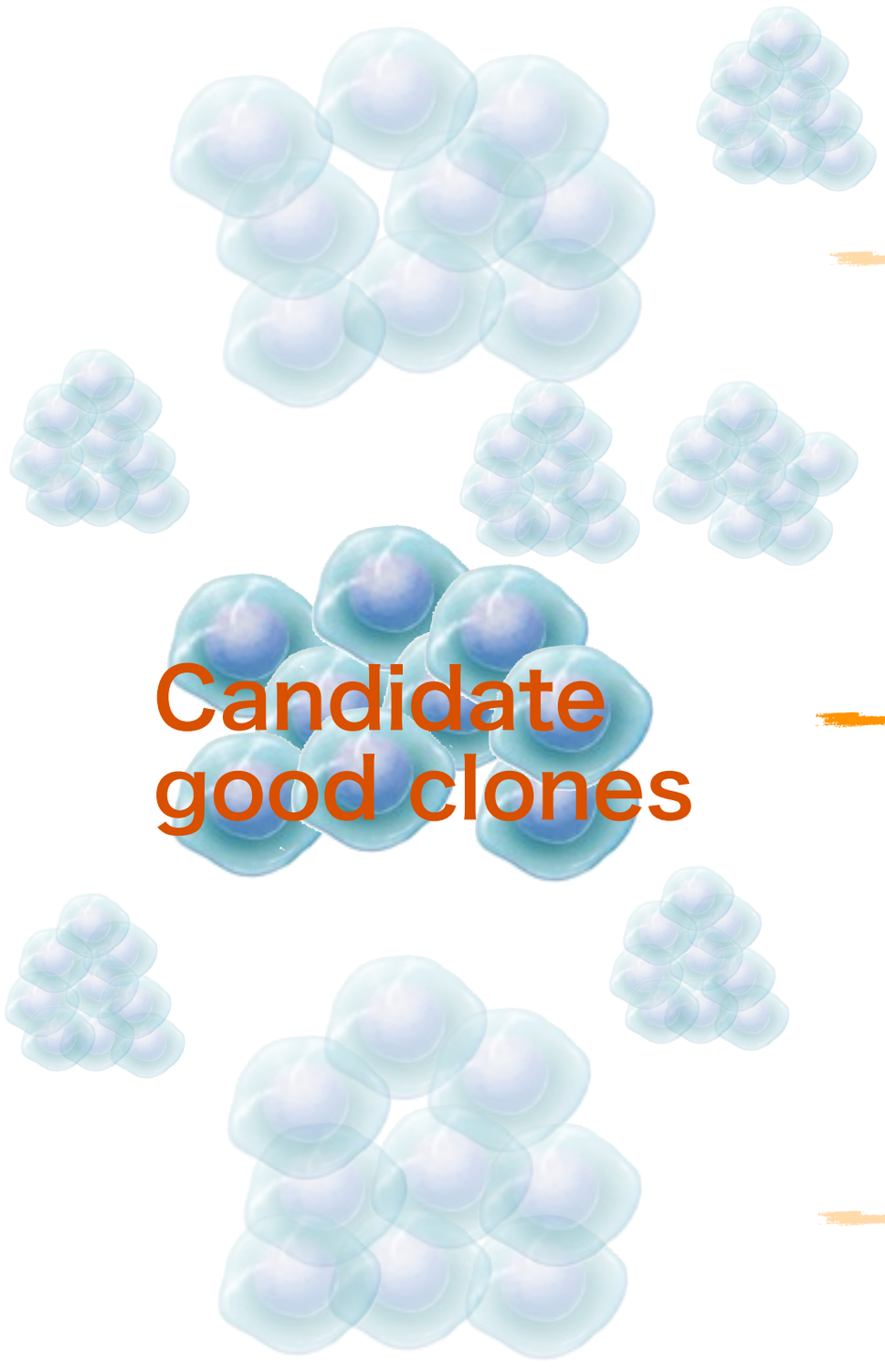
30 days!



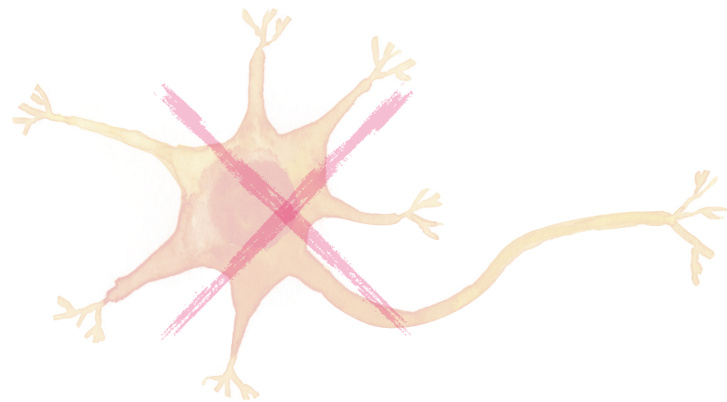
Need to Re-try?

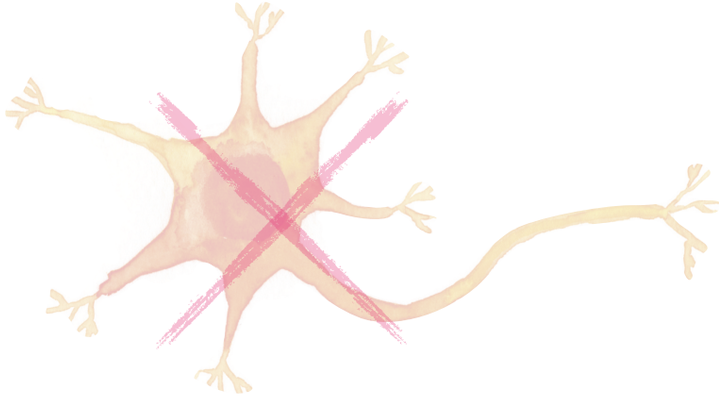


Only Good Clones!

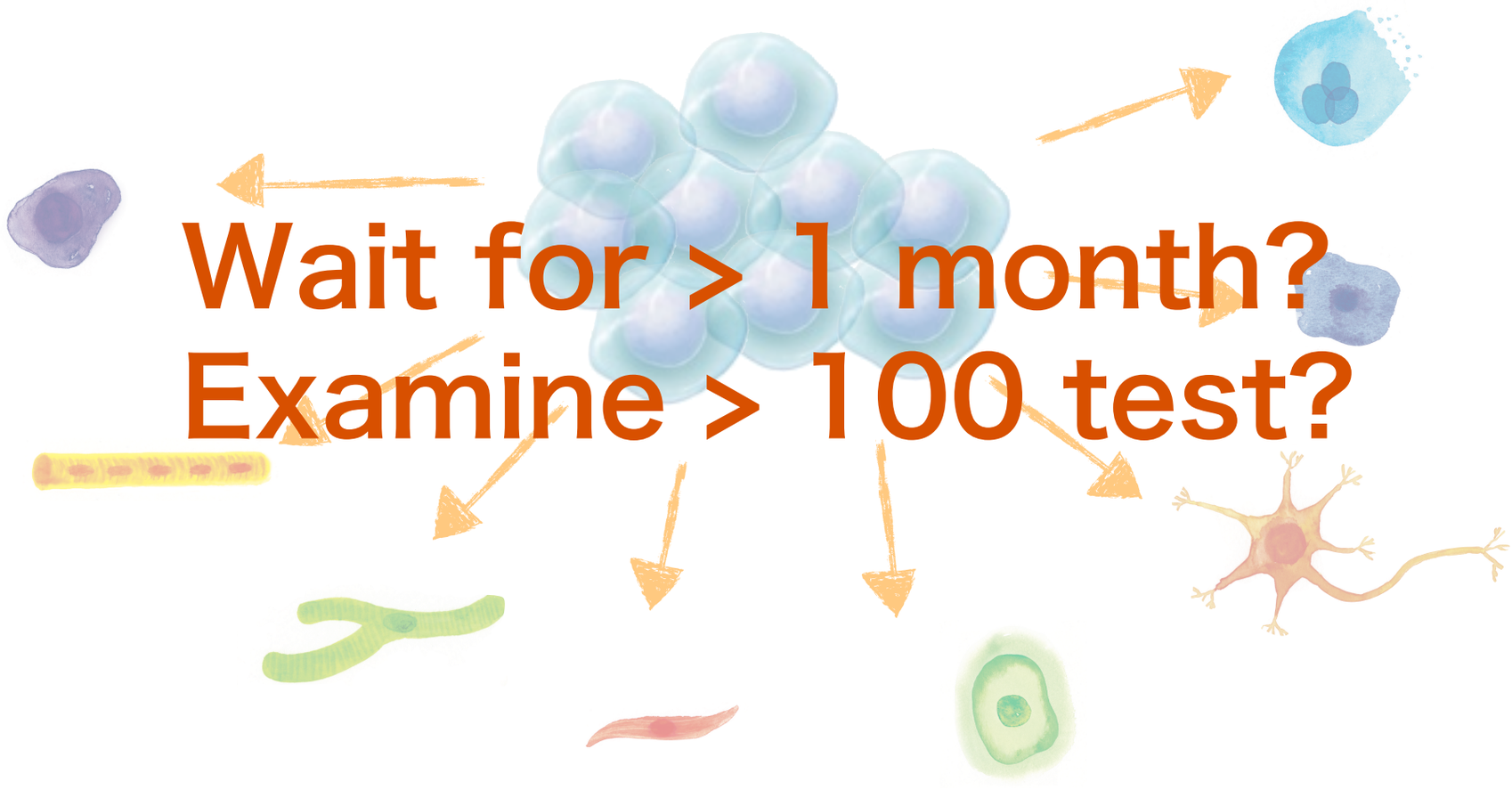


Candidate good clones





- Taking a long time to test differentiation
- Big labor to test of iPSC clones for multiple cell type
- Heterogeniety in cell populations of iPSC



Wait for > 1 month?
Examine > 100 test?

2013.2.26

エピゲノム解析による iPS細胞の特性解析

Center for iPS Cell Research and Application, Kyoto University
Genomics and Epigenomics Core Facility / Yamanaka Lab.
Akira Watanabe