Next Generation Sequencing For Medical Sciences

Department of Computational Biology and Medical Sciences,
University of Tokyo

Yutaka Suzuki

Development and Applications of Genome Technologies

Previous contributions as a sequence center


ゲノム支援

「ゲノム支援」はゲノム科学の
すそ野拡大とピーク作りをめざします

SGMJ

Genome-Shien, MEXT


CREST 戦略的創造研究推進事業

HOME

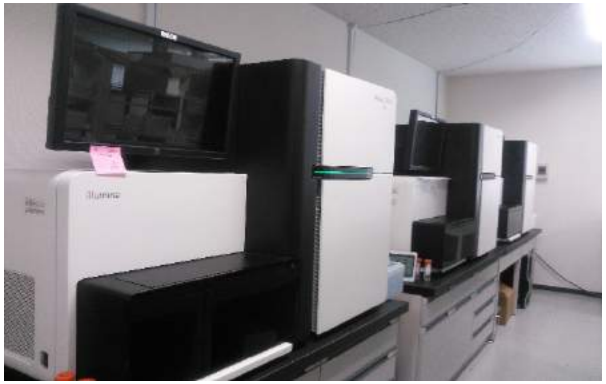
サイトマップ

IHEC, CREST

Hiseq2500 x 3+ Hiseq4000 x 1 & Related instruments

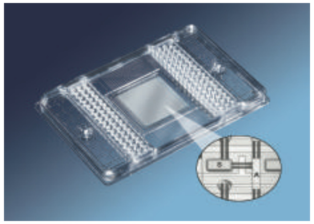
Approximate amount of data produced so far

	Genome	Transcriptome	Epigenome
#samples	2000	3000	5000



Single cell analytical platform

Single cell analyzer;
C1 (Fluidigm)



“Chip” of C1

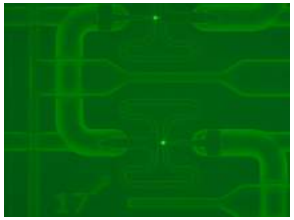
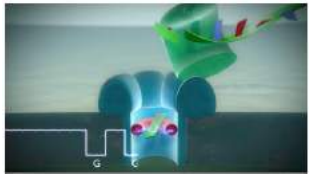


Image of single-cell
capture on C1

And more…
(Incubation center for new genome technologies)

PI: Yutaka Suzuki.
Univ. Tokyo

Operators:
Technicians 5
Programmers 4



Beta-test of nanopore sequencer

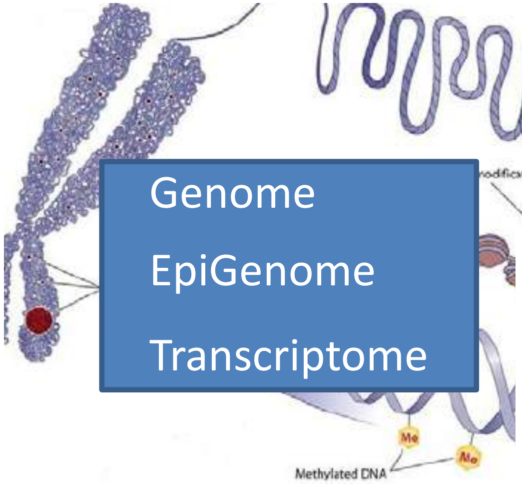
ysuzuki@hgc.jp

Providing latest analytical platforms for every layer of omics studies

Genome

EpiGenome

Transcriptome

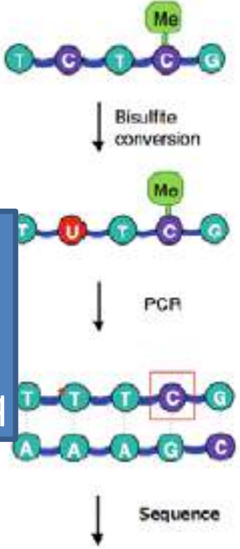


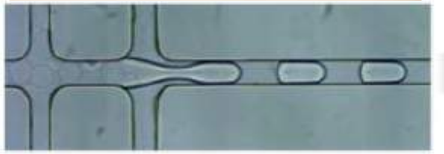
Sequencers
-HiSeq/MiSeq
-PacBio

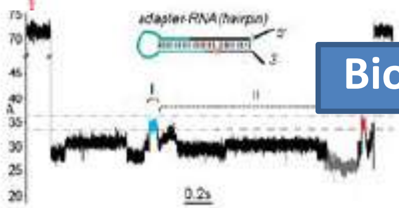
WES/WGS

ChIP Seq/BS Seq

RNA Seq/CLIP Seq

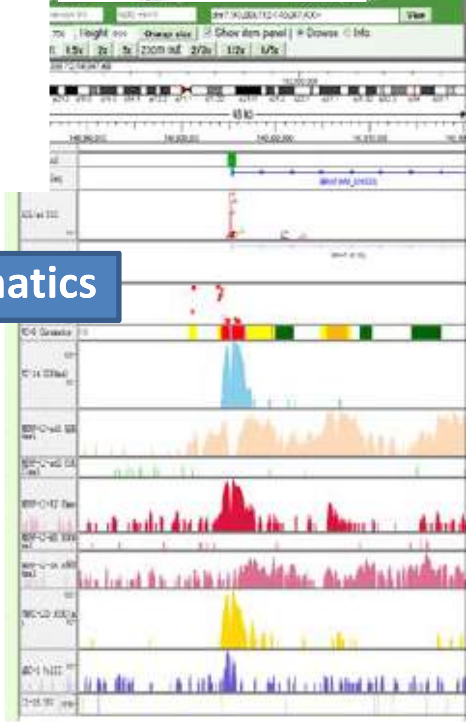






Analytics Platforms
-Single Cell
-USB Sequencer
-Long Read

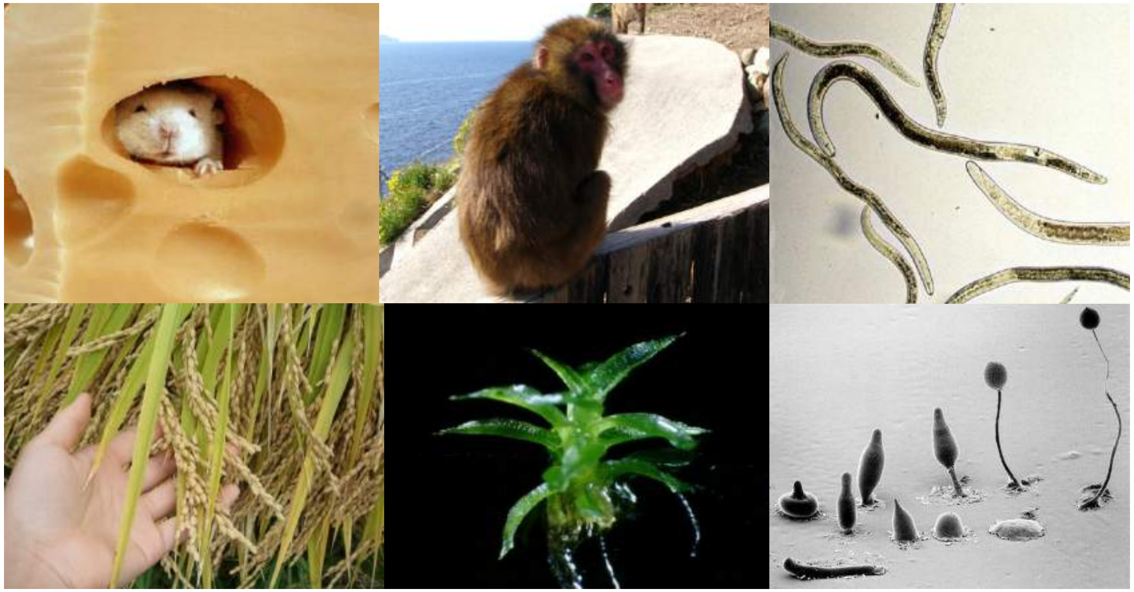
Bioniformatics



Objectives

Humans for Biomedical Applications

+Model/Non-model Organisms for Basic Sciences



Towards Next Generation Genome Sciences

Next Generation Sequencing for Medical Applications

Some basics for Illumina sequencing

Using cancer genomics as an example

- **Genome sequencing (COSMIC)**
- **"Multi-Omcs" applications (IHEC/ENCODE/ROADMAP)**

(BREAK)

Recent topics

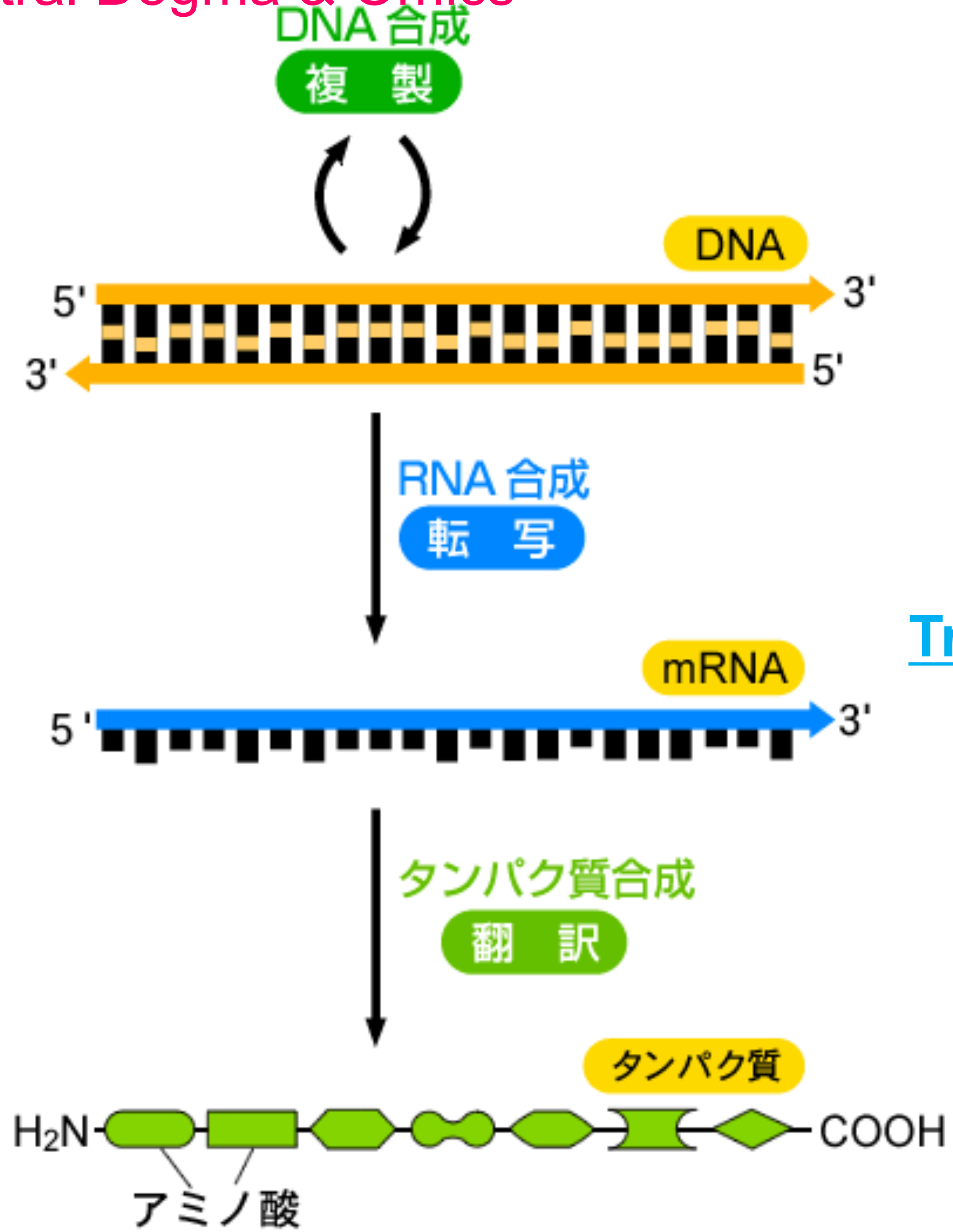
- **Single cell applications (Human Cell Atlas)**
- **MinION sequencing (One Health)**

Some Basics

Yutaka Suzuki
Department of Computational Biology and Medical Sciences
Graduate School of Frontier Sciences
The University of Tokyo

東京大学
THE UNIVERSITY OF TOKYO

Central Dogma & Omics



Analysis of the "entity"

Genome

Transcriptome

Proteinome

Towards medical appication

Disease

Clinical trial

Cell/Animal model

??

Mal-function of genes

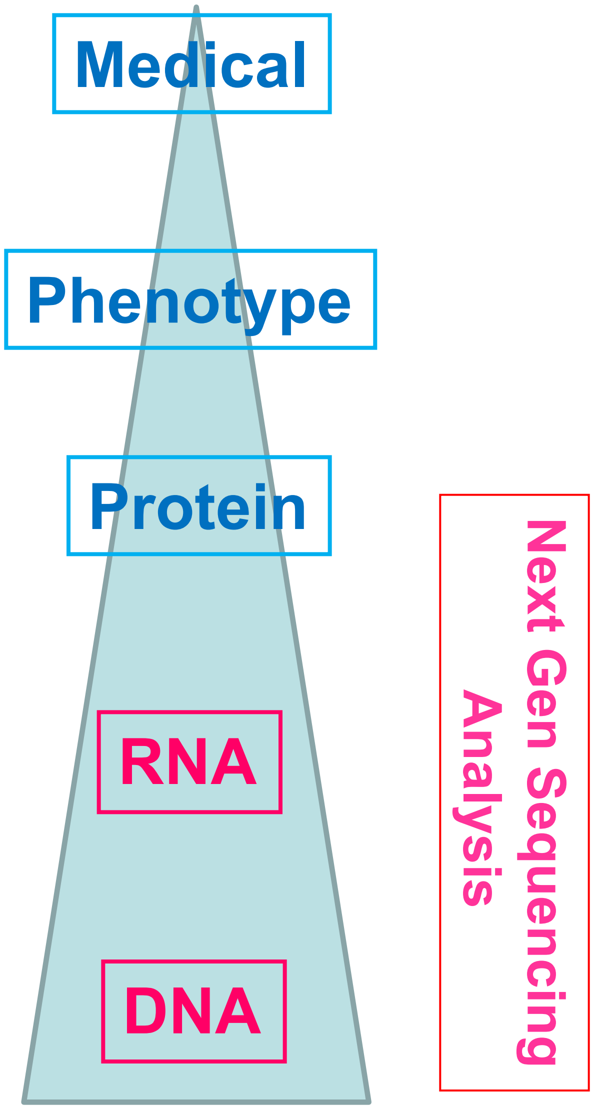
??

Expression disorder

RNA Seq, ChIP Seq

DNA mutation

Exome, Whole genome resequencing



Medical

Phenotype

Protein

RNA

DNA

Next Gen Sequencing Analysis

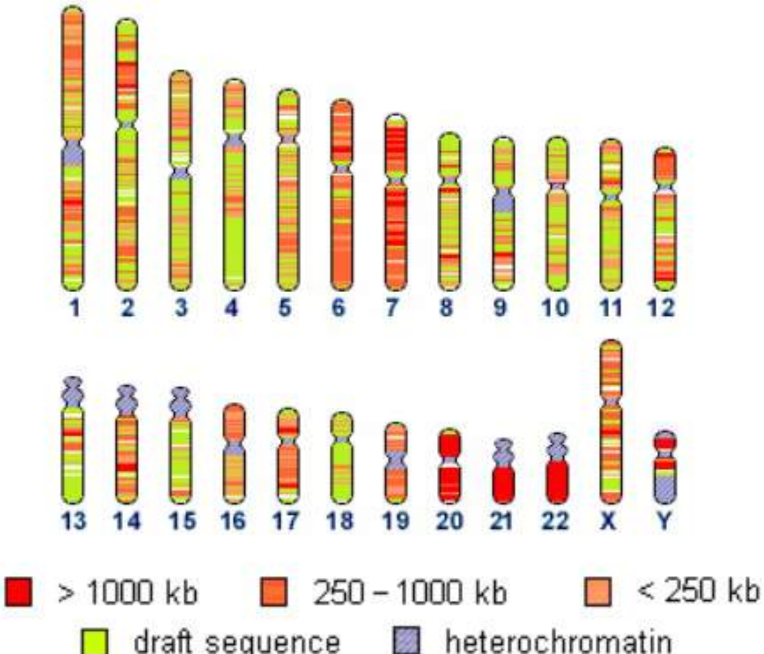
Data compilation

Human Genome Sequence Revealed

-22x2+ XY Chromosomes

-3Gb

-20,000 genes



Human Full-length cDNA Annotation Meeting~ H-invitational

August 25 ~ September 3, 2002.

>120 researchers from around the world @ Tokyo Bay

Annotated <u>41,421</u> Full-length cDNAs

Full-length cDNAs

- FLJ (Japan)
- MGC (USA)
- DKFZ (Germany)
- HSPC (China)

+ RefSeq

+ Ensembl genes

↓

Comprehensive dataset of Full-length cDNAs

H-Invitational

Welcome to H-Invitational Homepage

Human Full-Length cDNA Annotation Invitational

H-Invitational Discussion Forum (BBS)
Member:Welcome to Photos List Page

Download Final Data Set for Publications

H-Invitational Functional Annotation Revision Meeting
October 28 - November 1, 2002
H-Invitational Functional Annotation Revision Meeting

Manual to revise functional annotation ------------ PDF NEW
To all participants concerned with functional annotation:
Especially to members of team 1 & 3 NEW

H-Invitational Annotation Jamboree on
August 25 - September 3. 2002
Progress Up To Now (Sept. 20)
Download Presentation Slides UPDATED on Sept. 3
Proposal of Annotation Themes
Program of August 25
General Announcements
Annotation Teams
Annotation Team Assignments
cDNA Data Sets
Download Data Sets
Explanation of Data Sets for Downloading (above)
Participants list
Meeting Schedule

Pre-Computing Results
Download H-Invitational Documents
SOUP: H-Invitational Annotation System
Online Annotation Manual Ver.2.0 ------------ PDF NEW
GTOP (NIG) (click "Organisms" and then "hsap0" to see all results)
G-integra
Pre-Computing Guide (in prep)

Sold short
Original yardstick shows the metre is not long enough
p8

ns for
tabase

D. CYRANOSKI

are expressed by looking for
erns in the sequence. cDNAs,
mRNA expressed in cells, offer a
route. "This will be a real human-
gue — not predicted from the
ome sequence. These are real
says meeting organizer Takashi
rector of the DDBJ in Mishima.

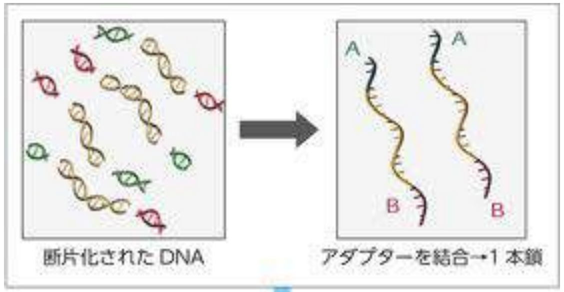
work with than mRNA itself.

genome sequence often means guessing at

Most of the cDNAs are already publicly

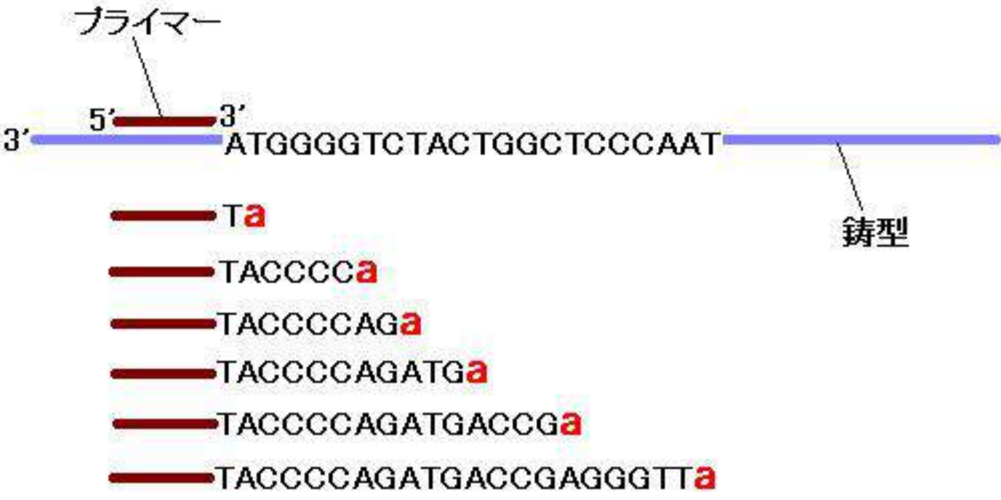
Sanger sequencing (di-deoxy method)

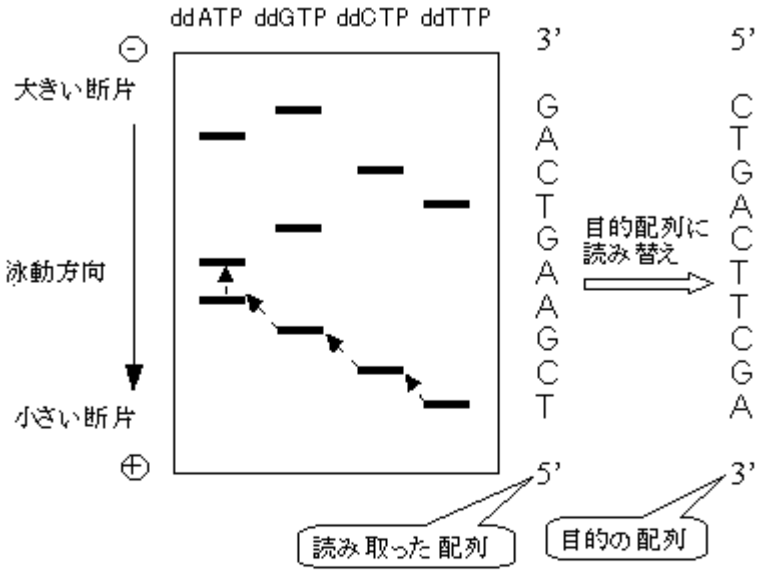
Fragmented DNA

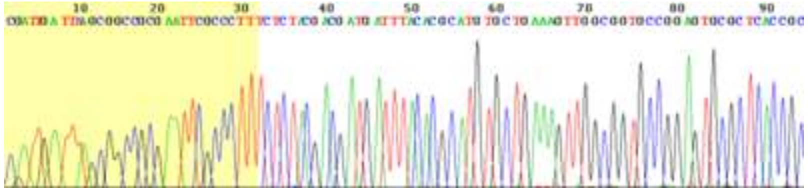


"Cloning"

Reaction by tubes







“NextGen sequencing”

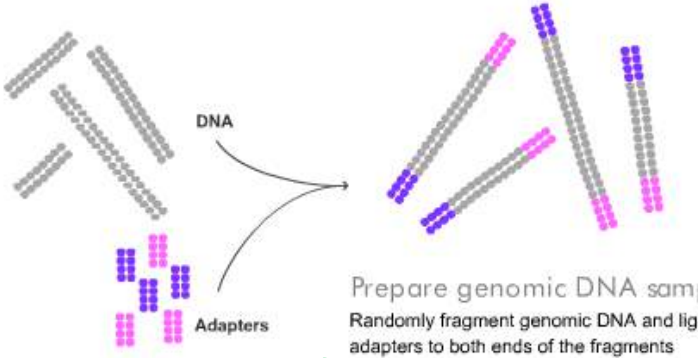
Maker	Products	Read base	length	Time	cost
Roche	GS Titanium	1Gb	330 bp	0.4 day	**1.5000 USD**
Illumina	Genome Analyzer	600Gb	150bp	10 day	**1,2000 USD**
AB	SOLiD4	200Gb	50 bp	7 day	**10,000 USD**
PacBio	RSII	?	5000 bp	1 day	**1000 USD**

※Gb=1,000,000,000 base: Human Genome=3Gb

Next Generation Sequencer, “Illumina” (Google “Illumina, method”)

Bridge (Cluster) PCR

1

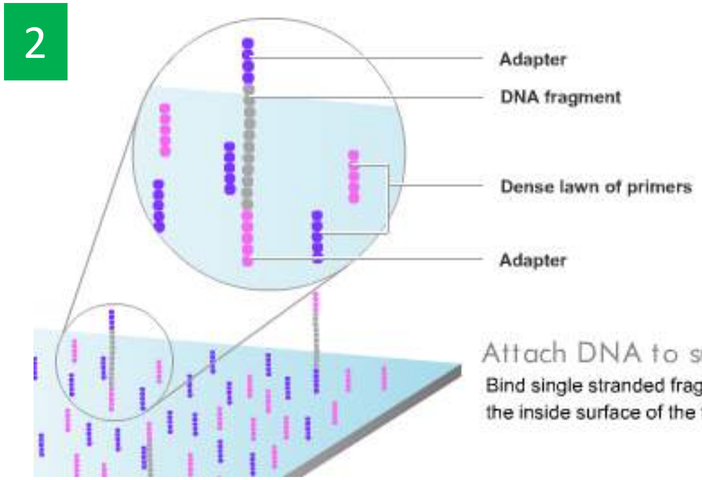


Prepare genomic DNA sample

Randomly fragment genomic DNA and ligate adapters to both ends of the fragments

any kinds of DNA

2



Attach DNA to surface

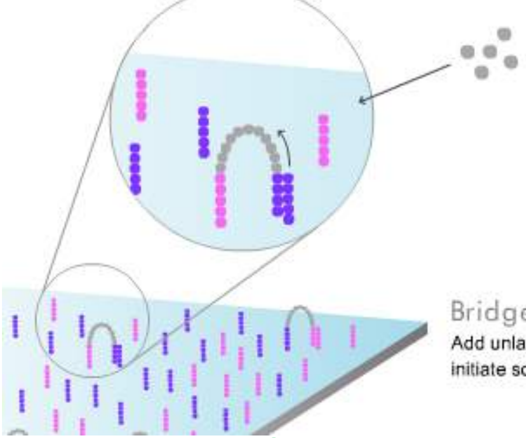
Bind single stranded fragments randomly to the inside surface of the flow cell channels.

Hybridize to complementary oligo attached to the surface

Make a copy of the template DNA

3

Template could not spread freely

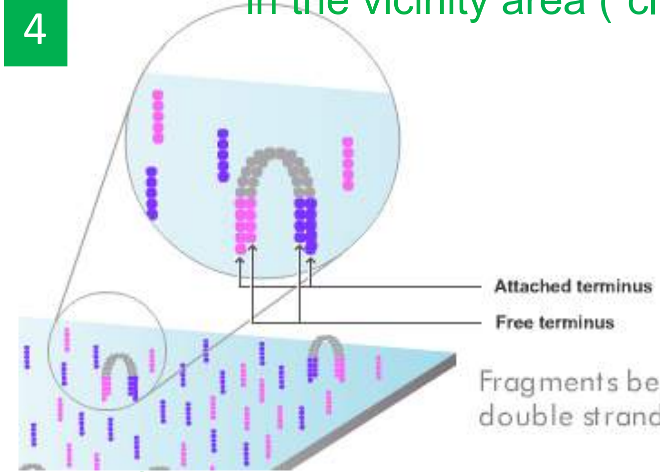


Bridge amplification

Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4

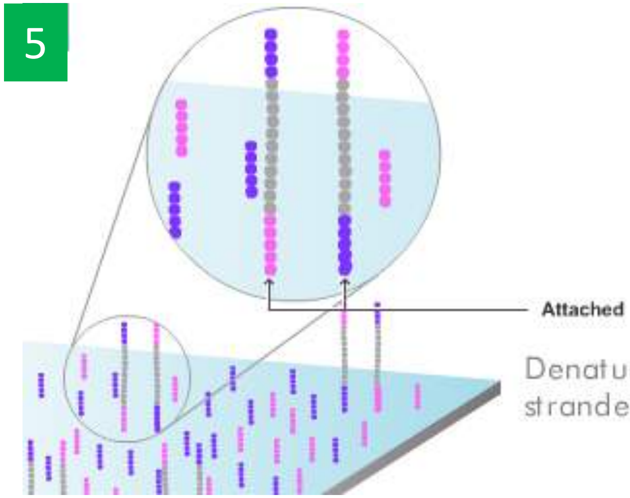
PCR using the primers
in the vicinity area (“cluster” PCR)



Fragments become double stranded

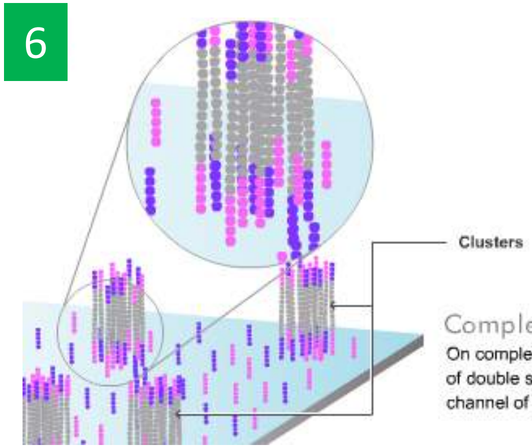
Reversible Dye Terminator

5



Denature the double stranded molecules

6

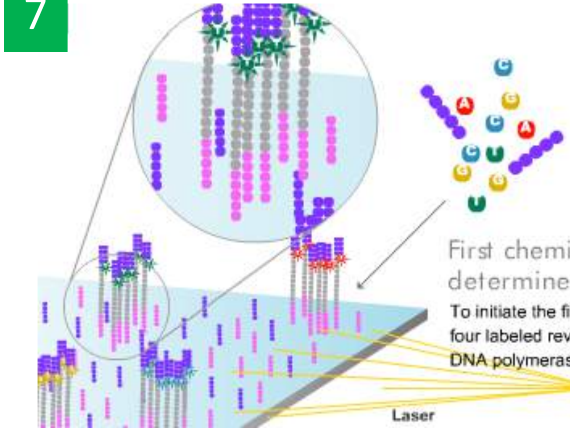


Completion of amplification

On completion, several million dense clusters of double stranded DNA are generated in each channel of the flow cell.

Cluster (or colony) of the template DNA
Is formed

7



First chemistry cycle: determine first base

To initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8

Di-deoxy sequencing using 4-colored dye

Image of first chemistry cycle

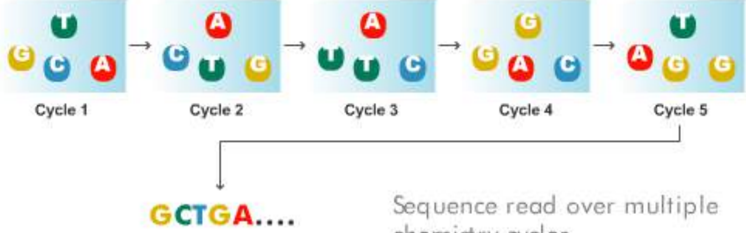
After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

Before initiating the next chemistry cycle

The blocked 3' terminus and the fluorophore from each incorporated base are removed.

9

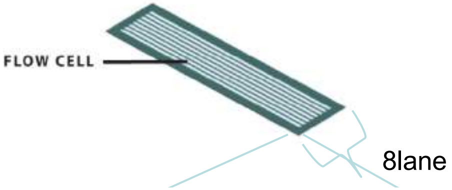
-36 cycle reactions (36 base can be read)



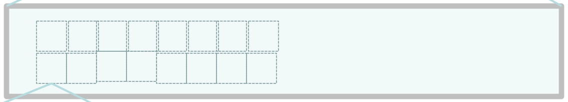
Sequence read over multiple chemistry cycles

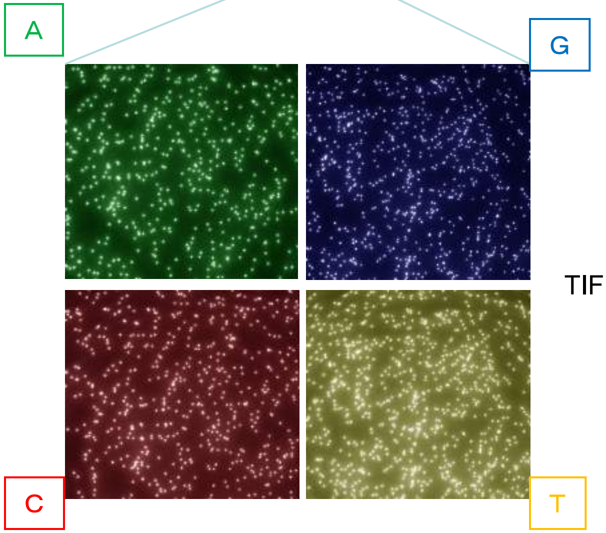
Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at a time.

Illumina



15000-25000cluster×300 ”tile”× 8レーン= 40-60 million cluster





TIFF images

(Base can be called
from about ~50 % of the clusters)

Sequence generation ~36-mer FASTA sequence

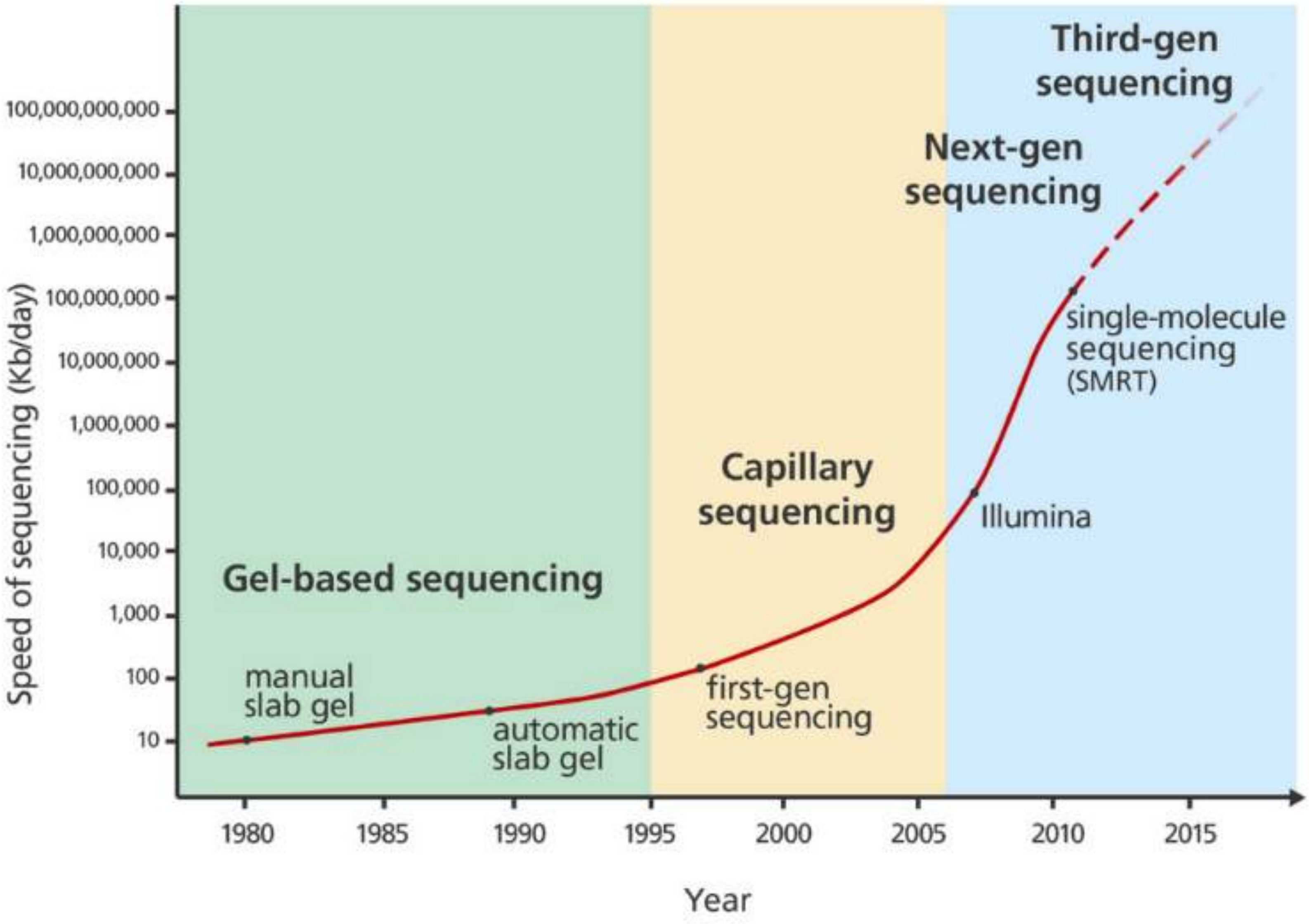
344	752	TTCTAATTGAGTTTATTTGAATCTT
270	793	TGAGAGAAACTTGGTCGTTCCGTCT
282	788	TTGGAAATTGTTGAGGGTGTGGGGT
254	708	TTTGATCTTCTTGTCCTTCTTCCCT
215	317	TGGACAGCCTTTATACTTAATTCCT
284	778	TTTTACAGAAATCTGATTGGTTCAT
242	782	TAACTTACATATGCATTCCTGTTAT
56	582	TGGGCCCTATGTTCCATCTCTTCCT

Mapping to the reference genome

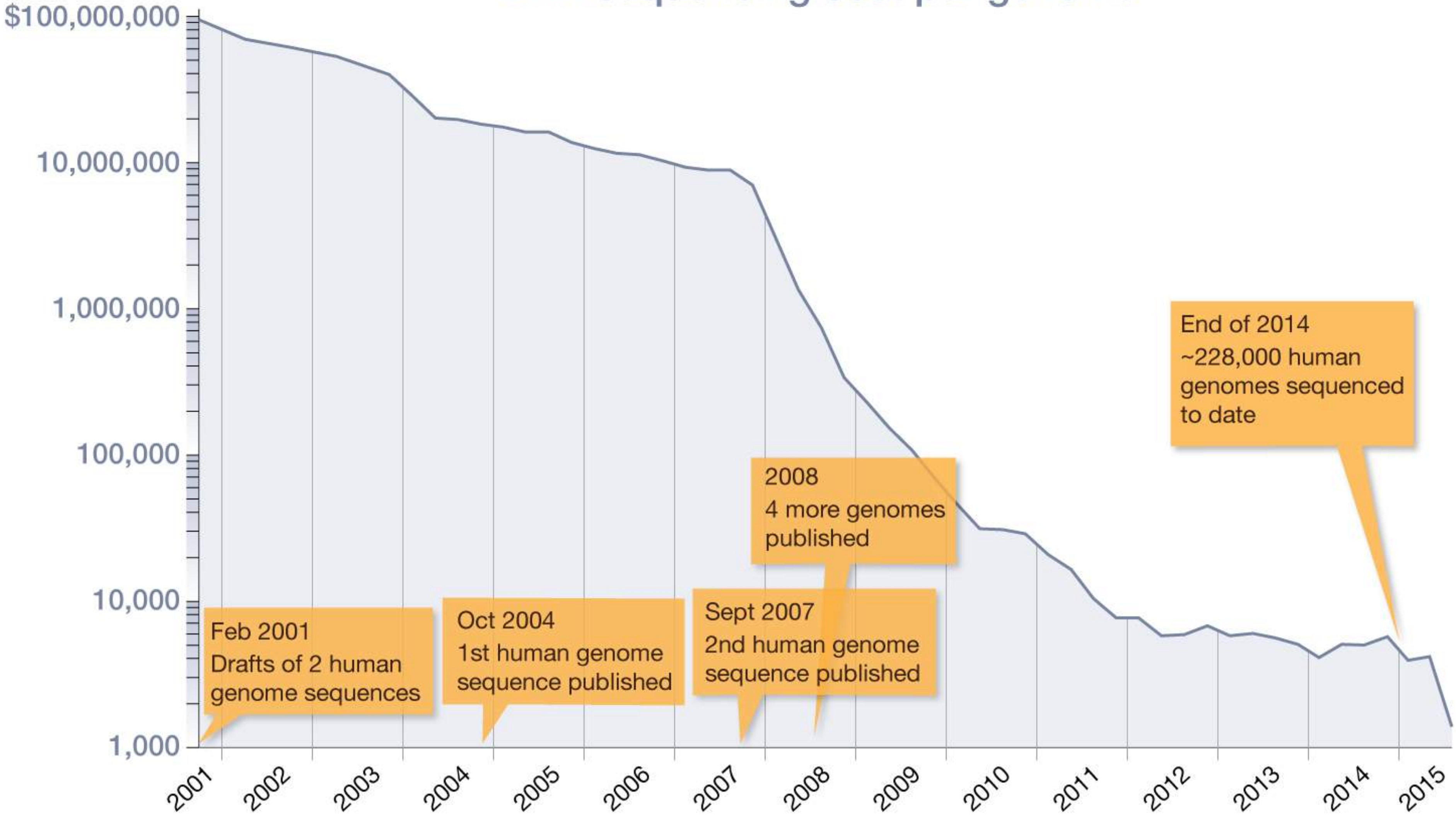
```
GGGATAATTCTGGGTTCTTTTCTCT 11453 1 BAC_plus_vector:35117 F
 GGGATAATTCTGGGTTCTTTTCTCC 9359

TAAAAATCGCCTTCAAAATTGCTTT 11453 1 BAC_plus_vector:155924 F
TAAAAATCGCCTTCAAAATTGCTTA 9359

GTCAGAGGCAAGAGGTGGGGGGCAT 0 0

TTTCATATTTTTTGCAATTGCTTCT 10406 1 BAC_plus_vector:161947 F
TTTCATATTTTCTGCAATTGCTTCC 9359

TTTTTAAAAAATTGCAAGTTTTTAT 0 0

TGTAAATGATGTAACCTTGTCTTCT 11453 1 BAC_plus_vector:35411 F
TGTAAATGATGTAACCTTGTCTTCA 9359
```

1 run (8 lanes)

Basic requirement is any kinds of DNA fragments having Solexa-tags at the both ends


Third-gen sequencing
Next-gen sequencing
Capillary sequencing
Gel-based sequencing
single-molecule sequencing (SMRT)
Illumina
first-gen sequencing
manual slab gel
automatic slab gel
Speed of sequencing (Kb/day)
100,000,000,000
10,000,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
100,000
10,000
1,000
100
10
1980
1985
1990
1995
2000
2005
2010
2015
Year

DNA sequencing cost per genome



Feb 2001
Drafts of 2 human genome sequences

Oct 2004
1st human genome sequence published

Sept 2007
2nd human genome sequence published

2008
4 more genomes published

End of 2014
~228,000 human genomes sequenced to date

HiSeq X Ten – 1000 USD genome

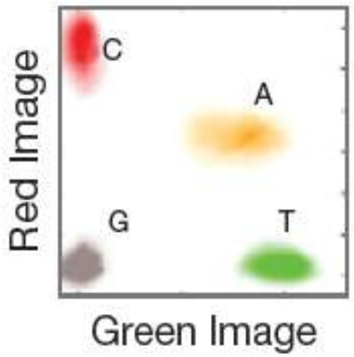
- 1.8 Tb/run x 10 instruments
  - 6 Billion reads
- 16 human genome/run
  - 3 days
- データクオリティ
  - 150bp x2
75% has >Q30

illumina

NextSeq 500

- Cost-performance
  - 30x WGS 1sample/run
  - WES 10 samples/run
    25~120Gb
  - 25~120 Gb
- Two-color detection

Figure 2: Two-Channel SBS Imaging



Accelerated detection of all four DNA bases is performed on the NextSeq 500 System using only two images to capture red and green filter wavelength bands. A bases will be present in both images (yellow cluster), C bases in red only, T bases in green only, and G bases in neither.

illumina

Illumina Sequencers for versatile uses

High Throughput
WES, target sequencing

National project level

MiniSeq

HiSeq 2500

HiSeq X Ten

Desktop types

MiSeq

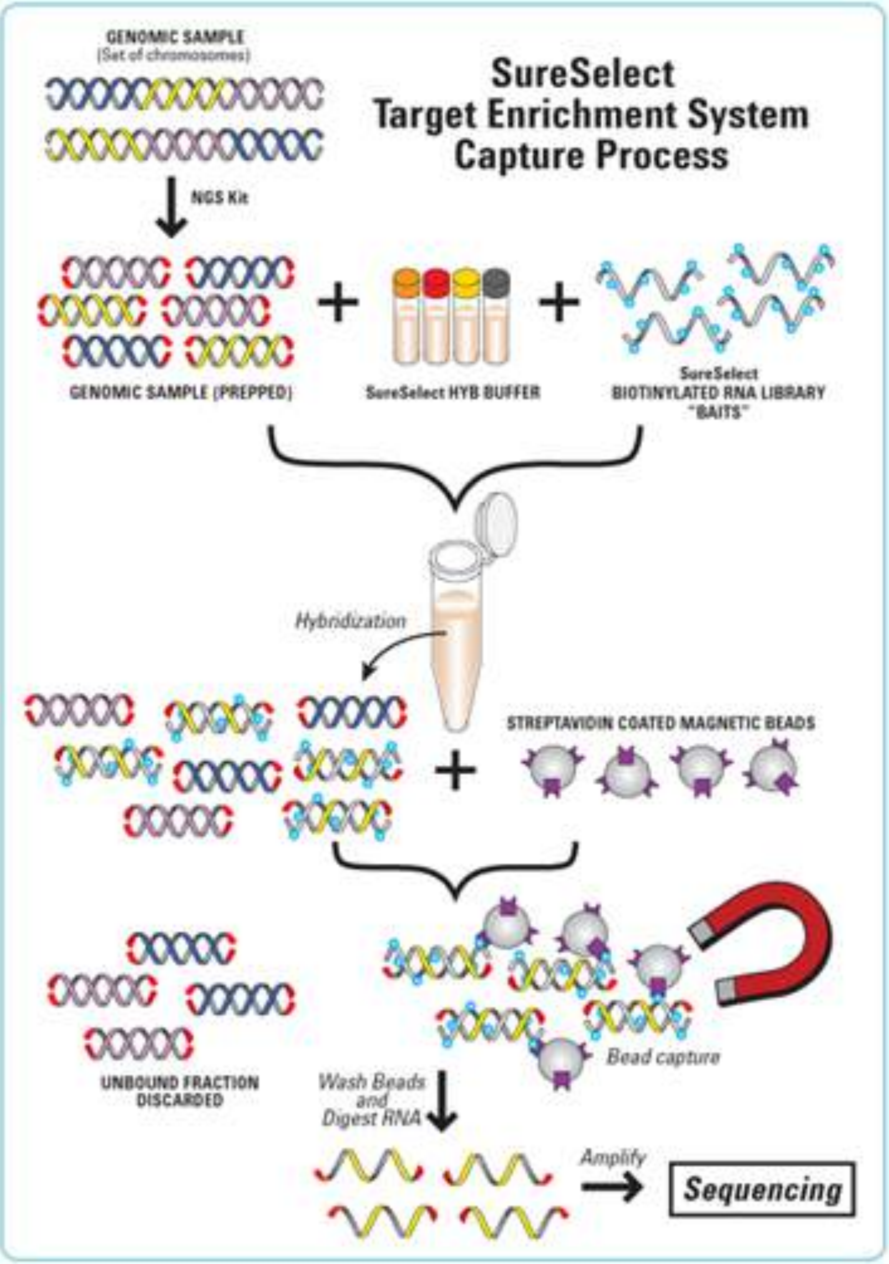
NextSeq 500

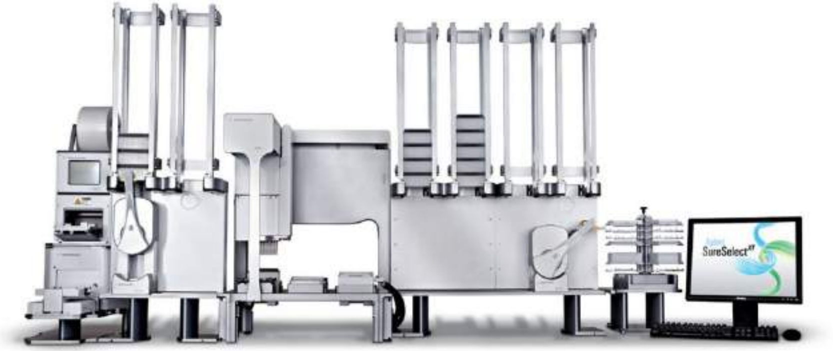
NovaSeq

Low throughput;
Micro-organism

Population level WGS

Targeted capture scheme, “Exome”





Hybri-captureによるPromoter Re-sequencing

Chromosome Browser: 1

Gene Labels: GeneSymbol

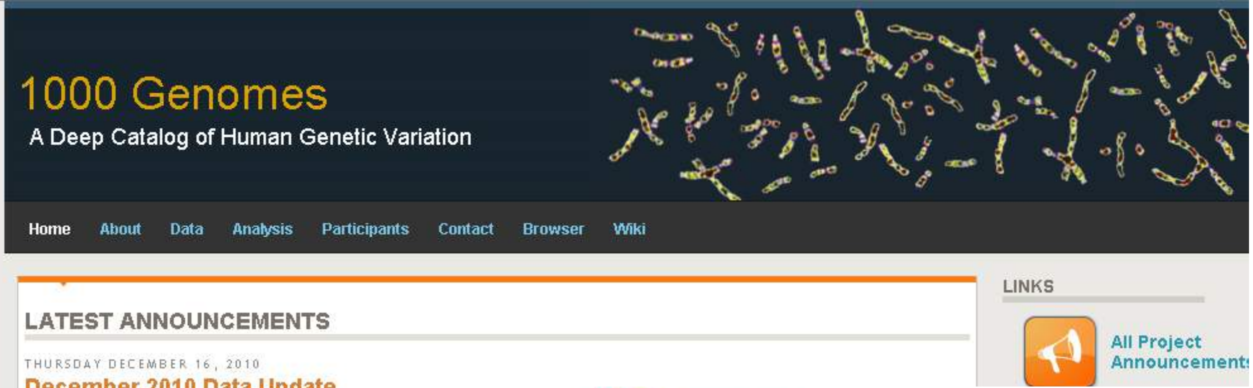
Chr: 1

883,940 - 893,538

2X

36base、ミスマッチ0でマップされたタグ：6,322,381
このうち、ハイブリキャプチャ領域にあるタグ：5,289,381 (84%)
ハイブリキャプチャ領域の全base数：3,986,164
このうち×5以上でカバー：3,403,145 (85%)

Ethnic Diversity



LATEST ANNOUNCEMENTS

THURSDAY DECEMBER 16, 2010

December 2010 Data Update

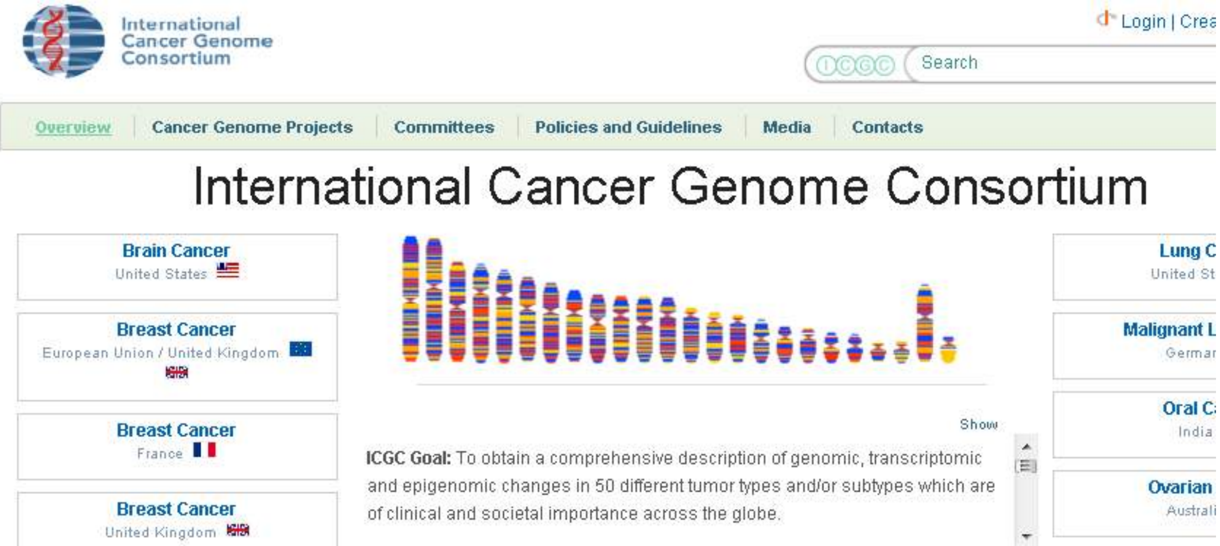
Full Project Genotype Release

Genotypes and haplotypes have been added to the SNP calls released ir 20100804 sequence and alignment release of the 1000 genomes projec genome and are released in the format VCF 4.0

Data access links: EBI / NCBI

Link to additional information: README file

Somatic mutations in cancer genomes

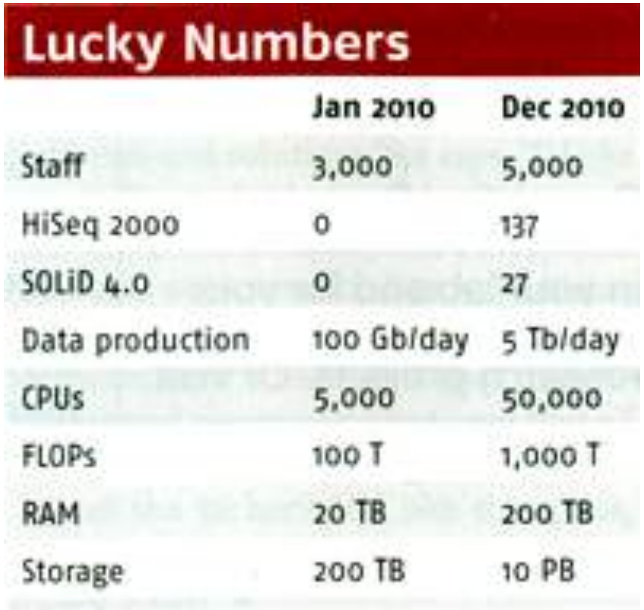


International Cancer Genome Consortium

Brain Cancer	United States
Breast Cancer	European Union / United Kingdom
Breast Cancer	France
Breast Cancer	United Kingdom
Lung C	United St
Malignant L	Germar
Oral C	India
Ovarian	Australi

ICGC Goal: To obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe.

Beijin Genome Institute (China)



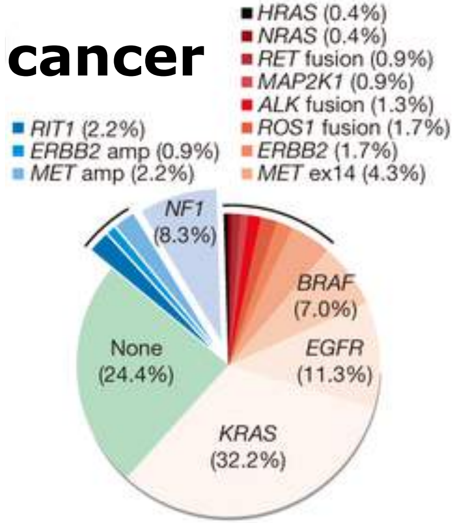
Lucky Numbers

	Jan 2010	Dec 2010
Staff	3,000	5,000
HiSeq 2000	0	137
SOLiD 4.0	0	27
Data production	100 Gb/day	5 Tb/day
CPUs	5,000	50,000
FLOPs	100 T	1,000 T
RAM	20 TB	200 TB
Storage	200 TB	10 PB

~Cancer Genomics~

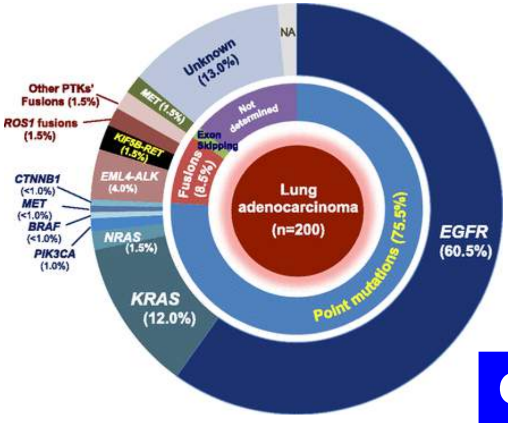
Genome sequencing at National Cancer Center East Hospital:

Lung cancer



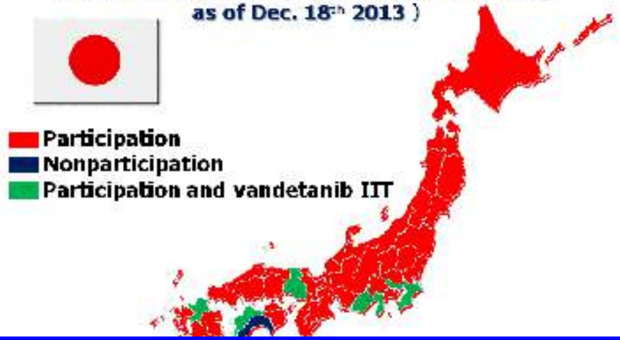
(Fig.3d, TCGA 2014 *Nature*)

Clinical Trial of the molecular targeting drugs



(Fig.6, Seo et al. 2012 *Genome Res*)

LC-SCRUM-Japan
(151 institutes in 46 prefectures participating as of Dec. 18th 2013)

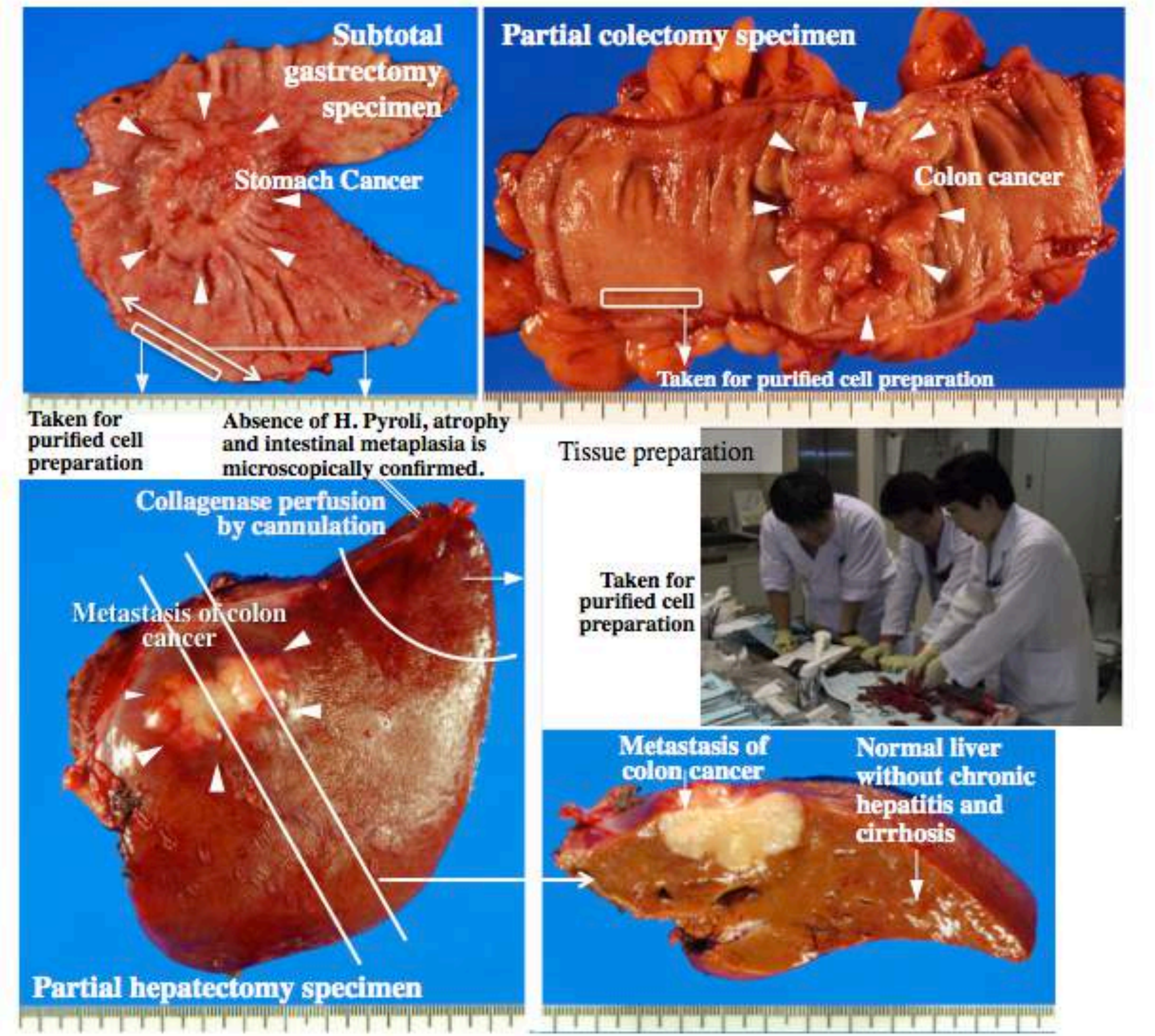


Clinical sequencing also

EGFR mutants were enriched in non-smokers, females and Asian.

Materials

Colon cancers



http://crest-ihec.jp/project/index.html#kanai2

SNP call using GATK+ MuTect (BROAD pipeline)

APPISTRY WEBINAR

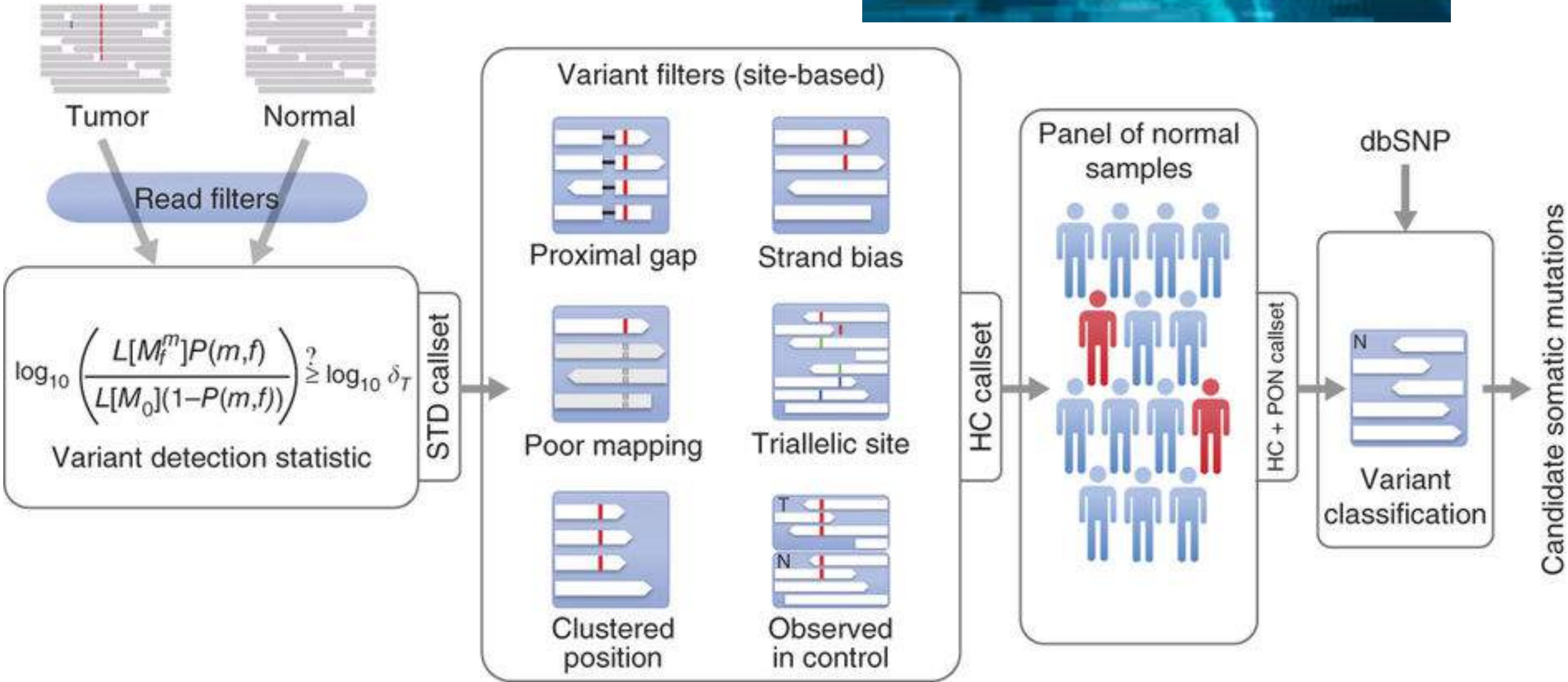
Using the GATK + MuTect

A Best Practices Workflow for Somatic Mutation Studies

Thursday, August 8th
12 PM EDT

REGISTER NOW

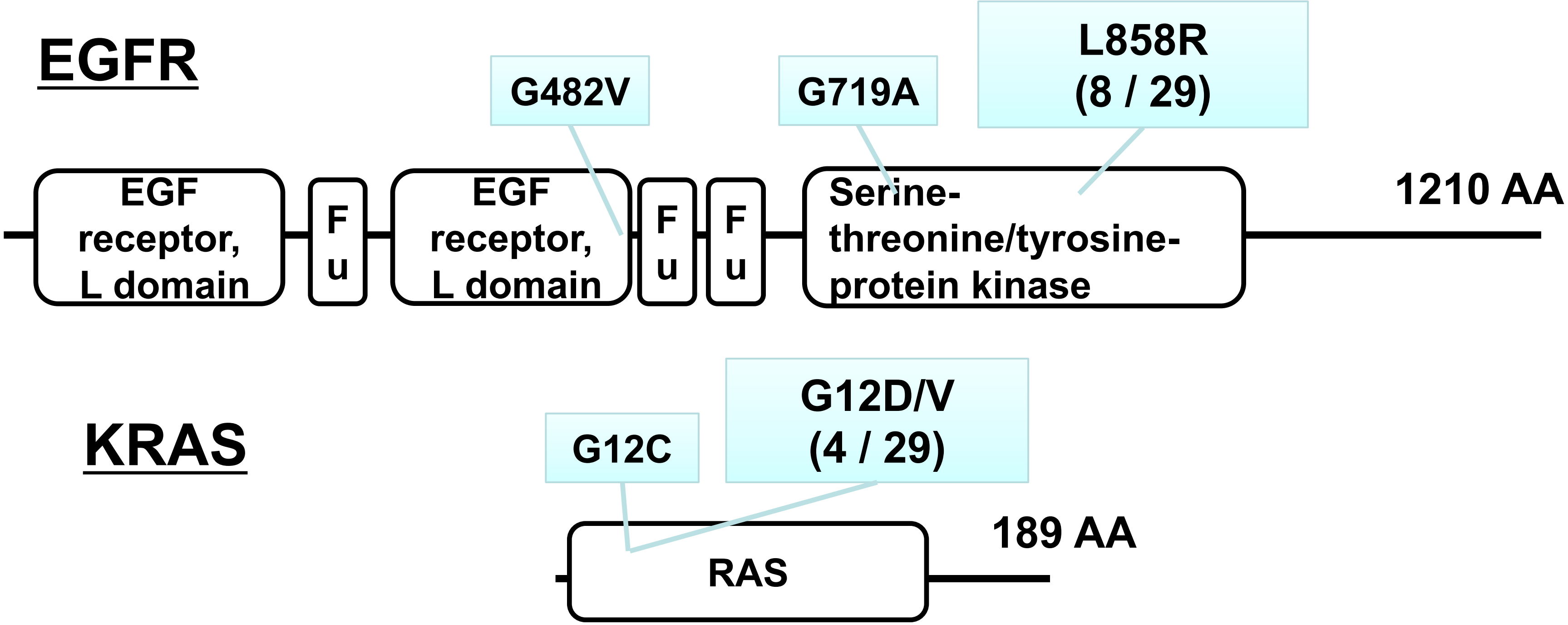
Appistry



Recursive SNVs in 29 T/N paired sequencing

Three recursive SNVs were identified.

Of these, two were well characterized SNVs in…



EGFR

G482V

G719A

L858R
(8 / 29)

EGF receptor, L domain

Fu

EGF receptor, L domain

Fu

Fu

Serine-threonine/tyrosine-protein kinase

1210 AA

KRAS

G12C

G12D/V
(4 / 29)

RAS

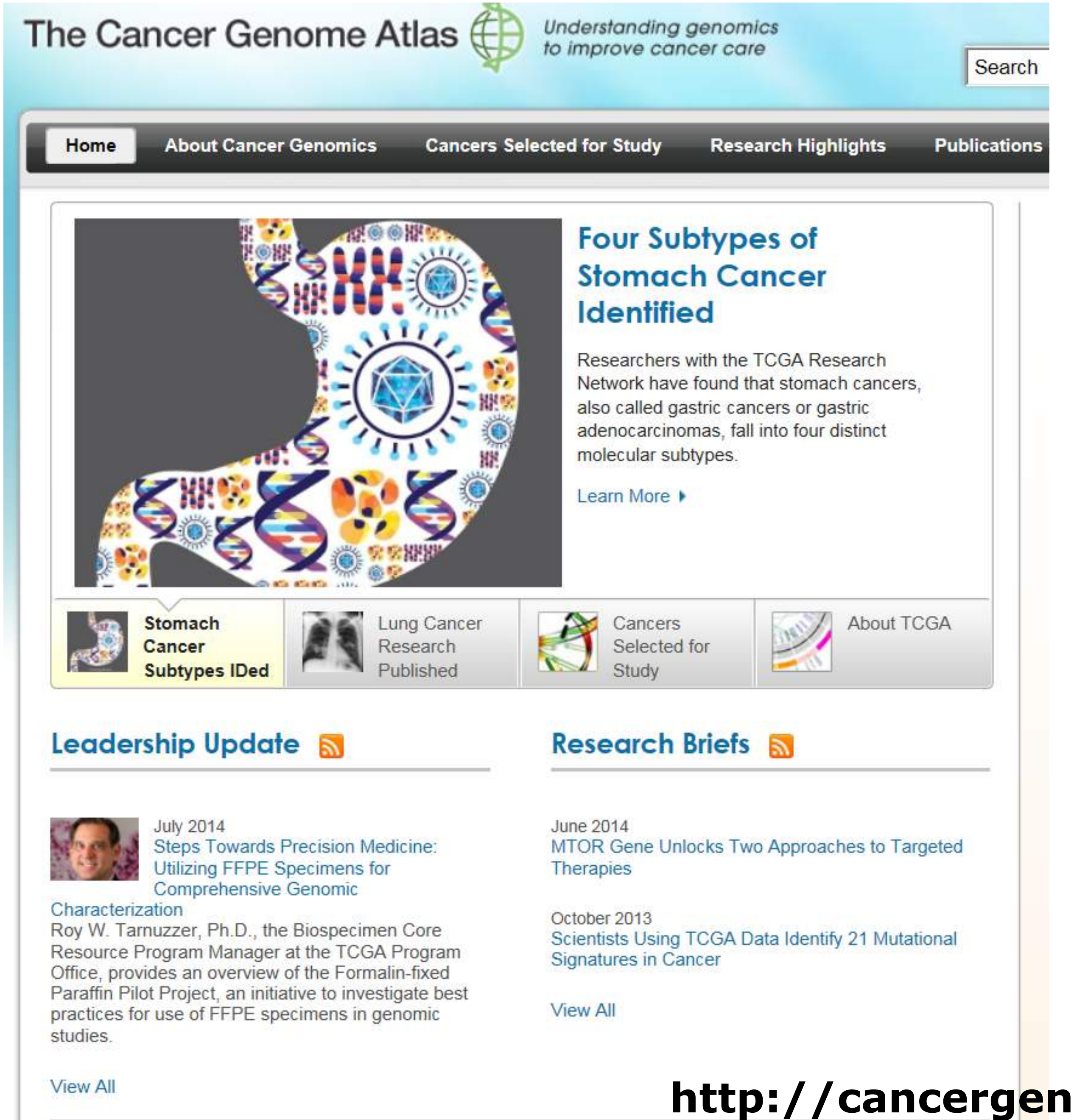
189 AA

**→ There were very rare “somatic” recursive SNVs,
except for well-known mutations, such as EGFR L858R and RAS G12V**

**Problems are not in the presence of called SNVs in cancer,
but in the absence of SNVs in the normal tissue (mostly because of loose SNV calls in normal tissues);**

Most seemingly recursive SNVs were turned out to be germ line

The Cancer Genome Atlas (TCGA)



Four Subtypes of Stomach Cancer Identified

Researchers with the TCGA Research Network have found that stomach cancers, also called gastric cancers or gastric adenocarcinomas, fall into four distinct molecular subtypes.

Learn More ▸

Leadership Update

July 2014
Steps Towards Precision Medicine: Utilizing FFPE Specimens for Comprehensive Genomic Characterization
Roy W. Tarnuzzer, Ph.D., the Biospecimen Core Resource Program Manager at the TCGA Program Office, provides an overview of the Formalin-fixed Paraffin Pilot Project, an initiative to investigate best practices for use of FFPE specimens in genomic studies.

View All

Research Briefs

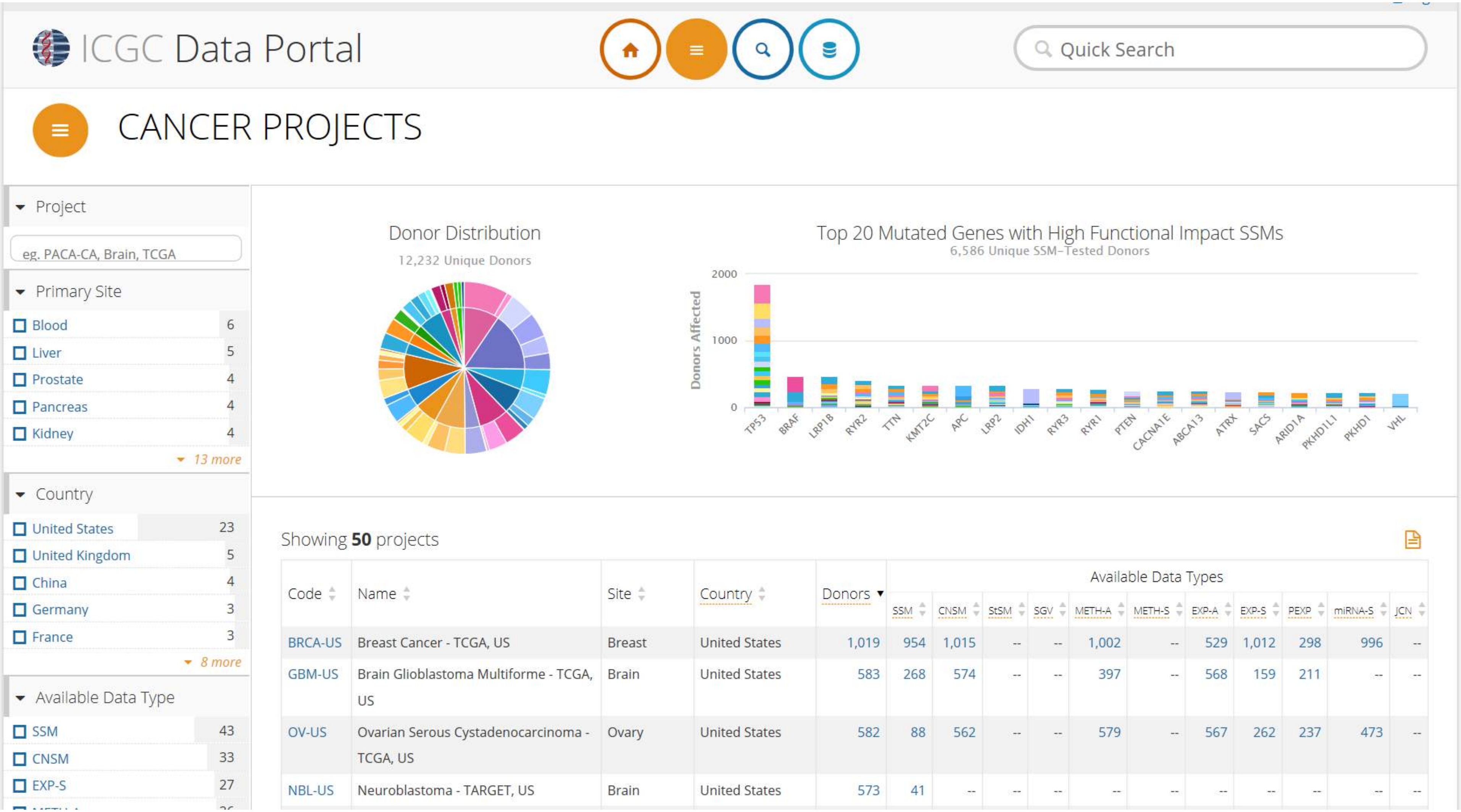
June 2014
MTOR Gene Unlocks Two Approaches to Targeted Therapies

October 2013
Scientists Using TCGA Data Identify 21 Mutational Signatures in Cancer

View All

http://cancergenome.nih.gov/

International Cancer Genome Consortium (ICGC)



Project

eg. PACA-CA, Brain, TCGA

Primary Site

☐ Blood	6
☐ Liver	5
☐ Prostate	4
☐ Pancreas	4
☐ Kidney	4

13 more

Country

☐ United States	23
☐ United Kingdom	5
☐ China	4
☐ Germany	3
☐ France	3

8 more

Available Data Type

☐ SSM	43
☐ CNSM	33
☐ EXP-S	27

Donor Distribution

12,232 Unique Donors

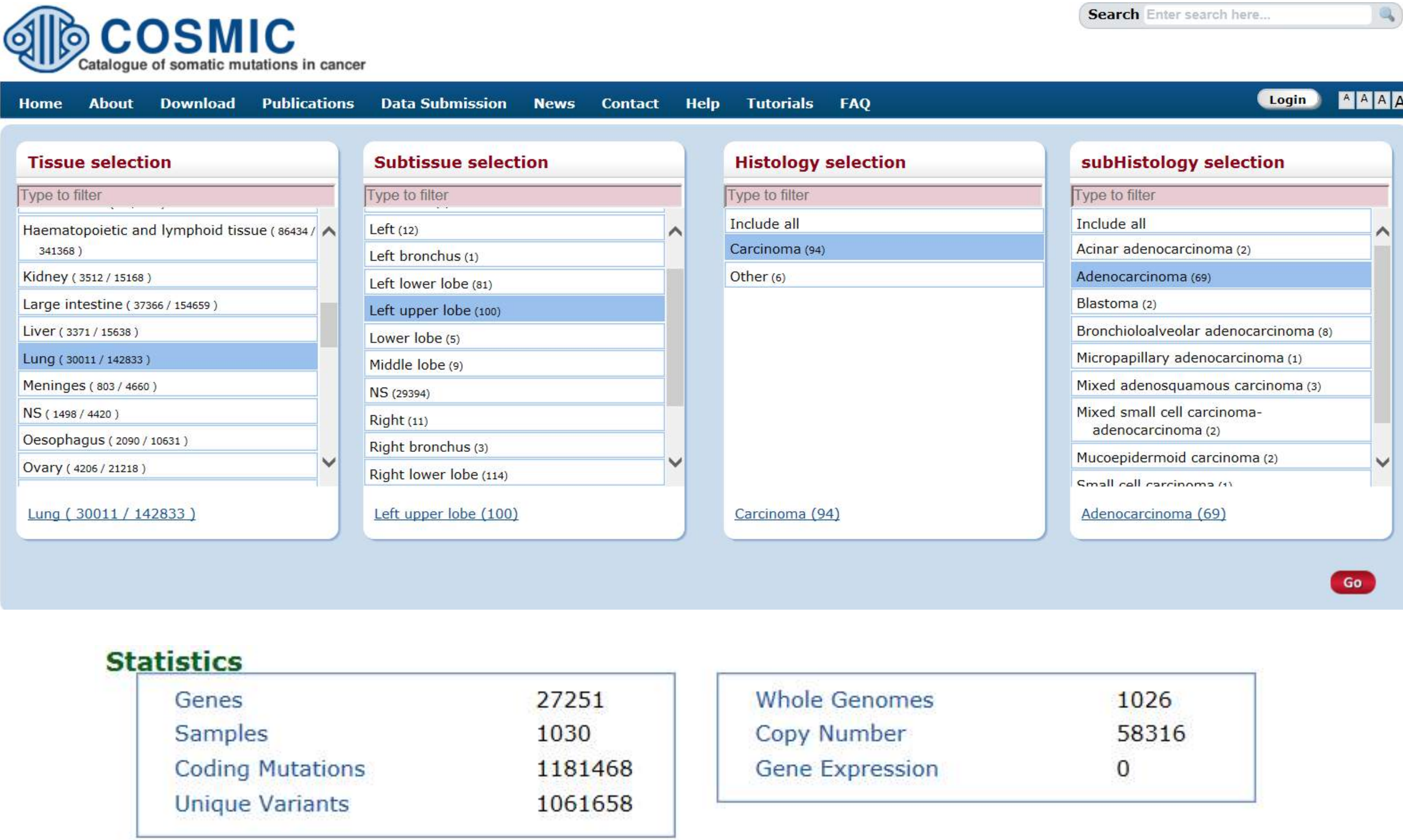
Top 20 Mutated Genes with High Functional Impact SSMs

6,586 Unique SSM-Tested Donors

Showing **50** projects

Code	Name	Site	Country	Donors	Available Data Types										
					SSM	CNSM	StSM	SGV	METH-A	METH-S	EXP-A	EXP-S	PEXP	miRNA-S	JCN
BRCA-US	Breast Cancer - TCGA, US	Breast	United States	1,019	954	1,015	--	--	1,002	--	529	1,012	298	996	--
GBM-US	Brain Glioblastoma Multiforme - TCGA, US	Brain	United States	583	268	574	--	--	397	--	568	159	211	--	--
OV-US	Ovarian Serous Cystadenocarcinoma - TCGA, US	Ovary	United States	582	88	562	--	--	579	--	567	262	237	473	--
NBL-US	Neuroblastoma - TARGET, US	Brain	United States	573	41	--	--	--	--	--	--	--	--	--	--

https://dcc.icgc.org/projects



Tissue selection

Type to filter

- Haematopoietic and lymphoid tissue (86434 / 341368)
- Kidney (3512 / 15168)
- Large intestine (37366 / 154659)
- Liver (3371 / 15638)
- Lung (30011 / 142833)
- Meninges (803 / 4660)
- NS (1498 / 4420)
- Oesophagus (2090 / 10631)
- Ovary (4206 / 21218)

Lung (30011 / 142833)

Subtissue selection

Type to filter

- Left (12)
- Left bronchus (1)
- Left lower lobe (81)
- Left upper lobe (100)
- Lower lobe (5)
- Middle lobe (9)
- NS (29394)
- Right (11)
- Right bronchus (3)
- Right lower lobe (114)

Left upper lobe (100)

Histology selection

Type to filter

- Include all
- Carcinoma (94)
- Other (6)

Carcinoma (94)

subHistology selection

Type to filter

- Include all
- Acinar adenocarcinoma (2)
- Adenocarcinoma (69)
- Blastoma (2)
- Bronchioloalveolar adenocarcinoma (8)
- Micropapillary adenocarcinoma (1)
- Mixed adenosquamous carcinoma (3)
- Mixed small cell carcinoma-adenocarcinoma (2)
- Mucoepidermoid carcinoma (2)
- Small cell carcinoma (1)

Adenocarcinoma (69)

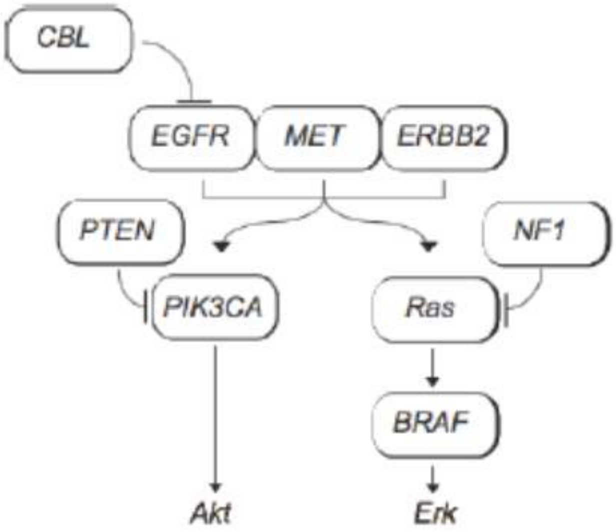
Statistics

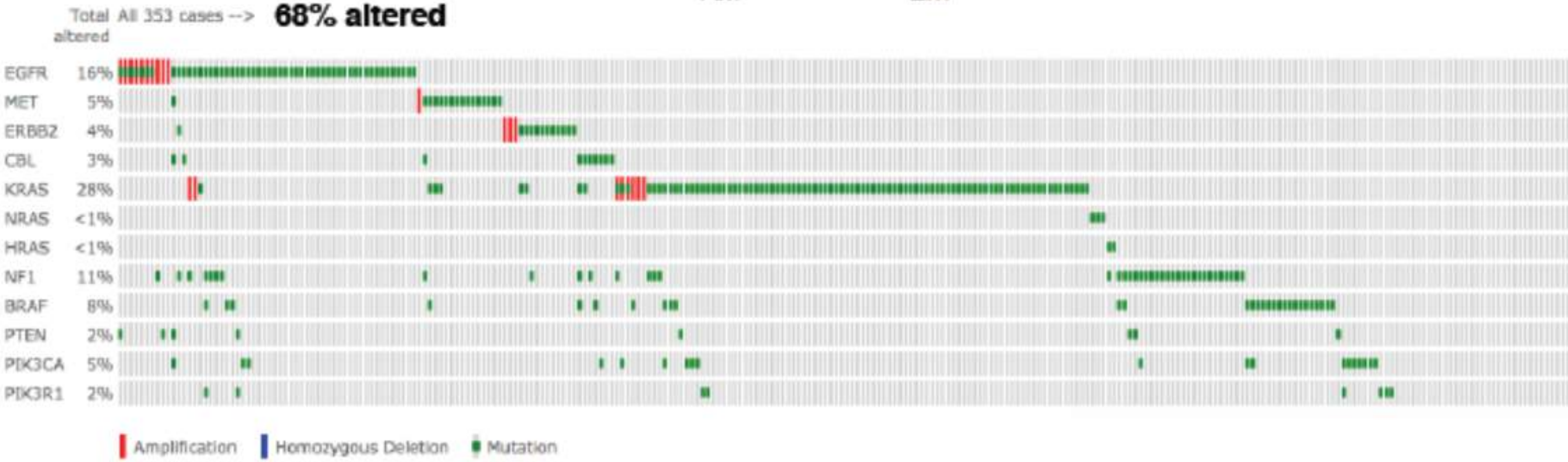
Genes	27251
Samples	1030
Coding Mutations	1181468
Unique Variants	1061658

Whole Genomes	1026
Copy Number	58316
Gene Expression	0

http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/

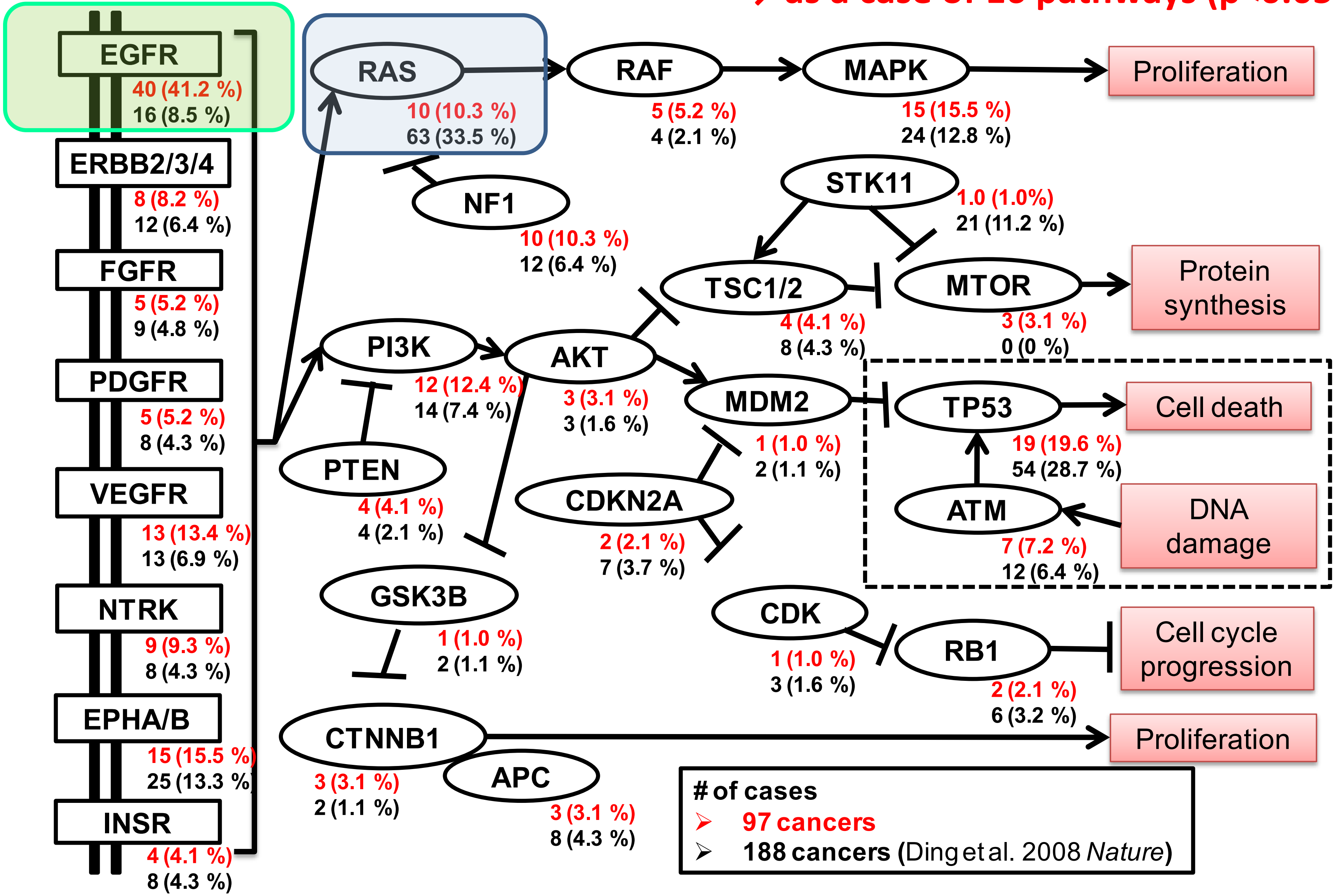
Major deregulation of RTK/RAS/RAF and PI3K/AKT in lung adenocarcinoma





Screening for "cancer-related pathway" candidates (I)
~SNVs in the EGFR-MAPK pathway

→ as a case of 16 pathways ($p<0.05$)



EGFR
40 (41.2 %)
16 (8.5 %)

ERBB2/3/4
8 (8.2 %)
12 (6.4 %)

FGFR
5 (5.2 %)
9 (4.8 %)

PDGFR
5 (5.2 %)
8 (4.3 %)

VEGFR
13 (13.4 %)
13 (6.9 %)

NTRK
9 (9.3 %)
8 (4.3 %)

EPHA/B
15 (15.5 %)
25 (13.3 %)

INSR
4 (4.1 %)
8 (4.3 %)

RAS
10 (10.3 %)
63 (33.5 %)

NF1
10 (10.3 %)
12 (6.4 %)

RAF
5 (5.2 %)
4 (2.1 %)

MAPK
15 (15.5 %)
24 (12.8 %)

Proliferation

STK11
1.0 (1.0%)
21 (11.2 %)

TSC1/2
4 (4.1 %)
8 (4.3 %)

MTOR
3 (3.1 %)
0 (0 %)

Protein synthesis

PI3K
12 (12.4 %)
14 (7.4 %)

AKT
3 (3.1 %)
3 (1.6 %)

MDM2
1 (1.0 %)
2 (1.1 %)

TP53
19 (19.6 %)
54 (28.7 %)

Cell death

ATM
7 (7.2 %)
12 (6.4 %)

DNA damage

PTEN
4 (4.1 %)
4 (2.1 %)

CDKN2A
2 (2.1 %)
7 (3.7 %)

GSK3B
1 (1.0 %)
2 (1.1 %)

CDK
1 (1.0 %)
3 (1.6 %)

RB1
2 (2.1 %)
6 (3.2 %)

Cell cycle progression

CTNNB1
3 (3.1 %)
2 (1.1 %)

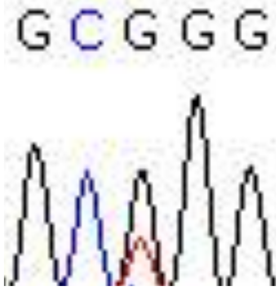
APC
3 (3.1 %)
8 (4.3 %)

Proliferation

# of cases
- **97 cancers**
- **188 cancers** (Ding et al. 2008 *Nature*)

Called SNVs were sufficiently solid to be validated by Sanger in most case

AD003
(cancer)

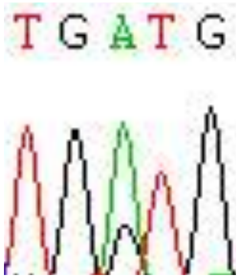


EGFR
c.2573T>G
p.L858R

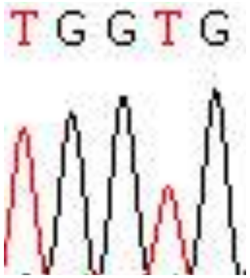
AD084
(cancer)

EGFR
c.2236_2250delGAATTAAGAGAAGCA
p.E746_A750del (exon19 deletion)

AD030
(cancer)

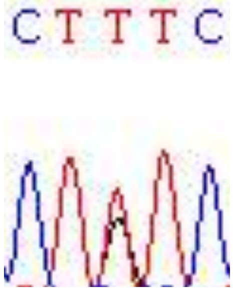


AD030
(normal tissue)

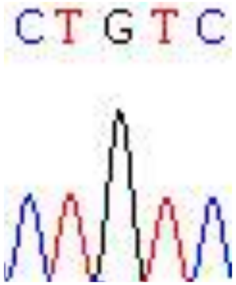


KRAS
c.35G>A
p.G12D

AD002
(cancer)

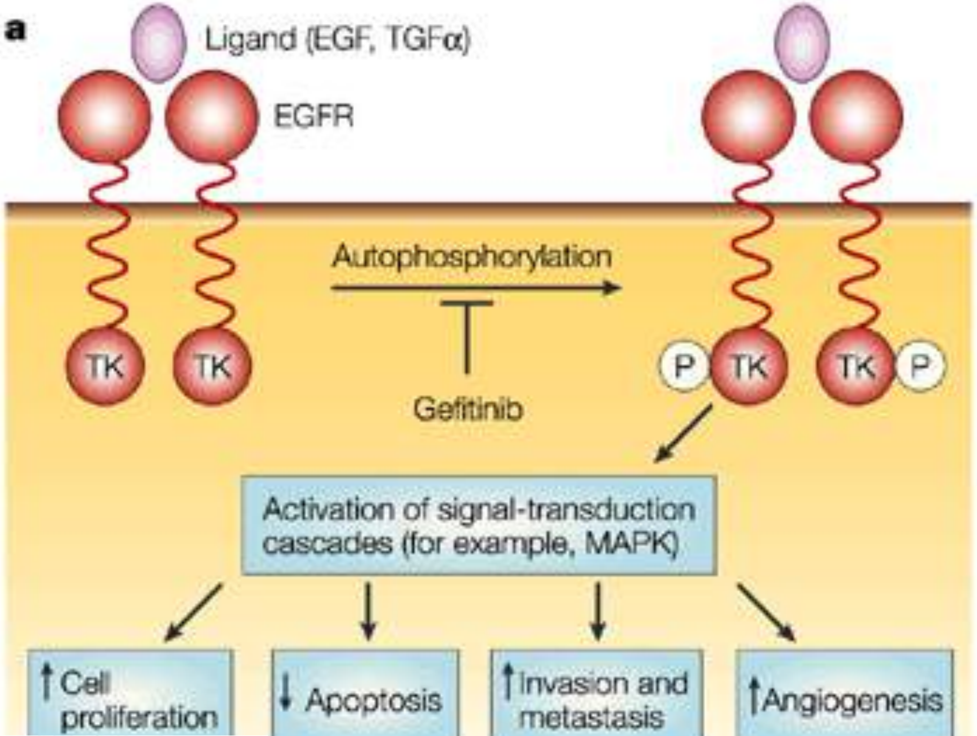


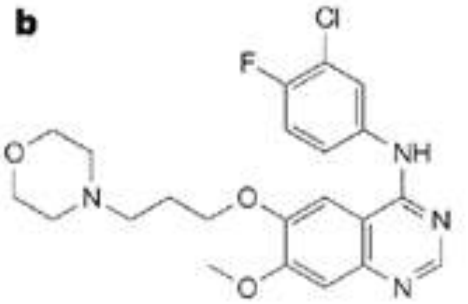
AD002
(normal tissue)



TP53
c.830G>T
p.C277F

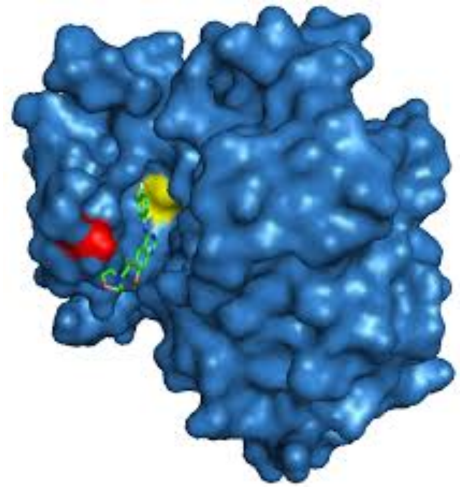
Gefitnib



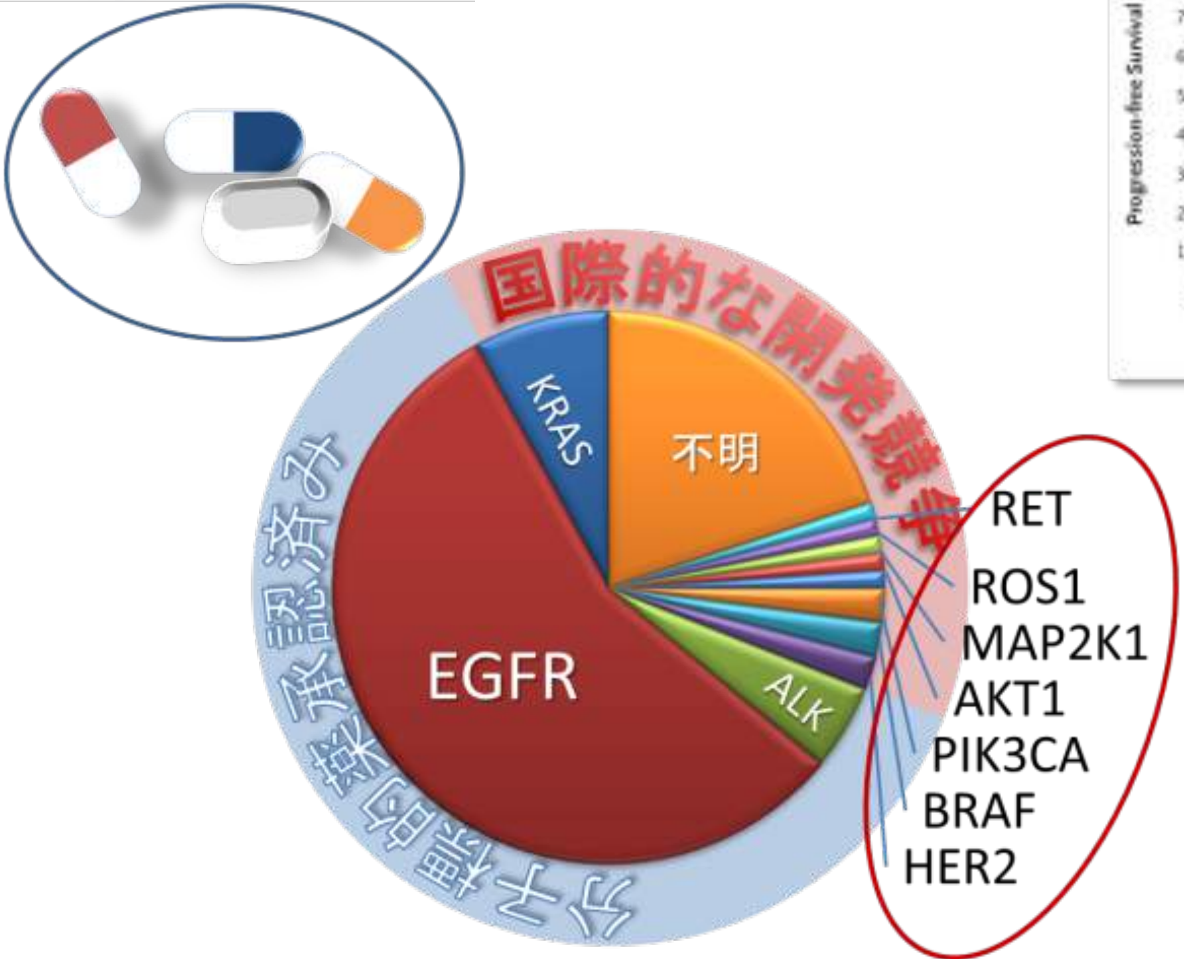


Gefitinib (ZD1839)

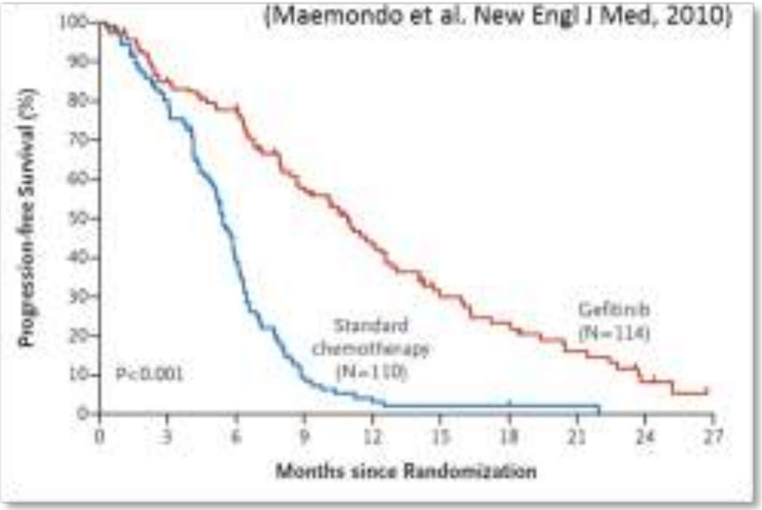
Nature Reviews | Drug Discovery



“Druggable” mutations



EGFR変異肺がんに対するEGFR阻害剤の効果



(Maemondo et al. New Engl J Med, 2010)

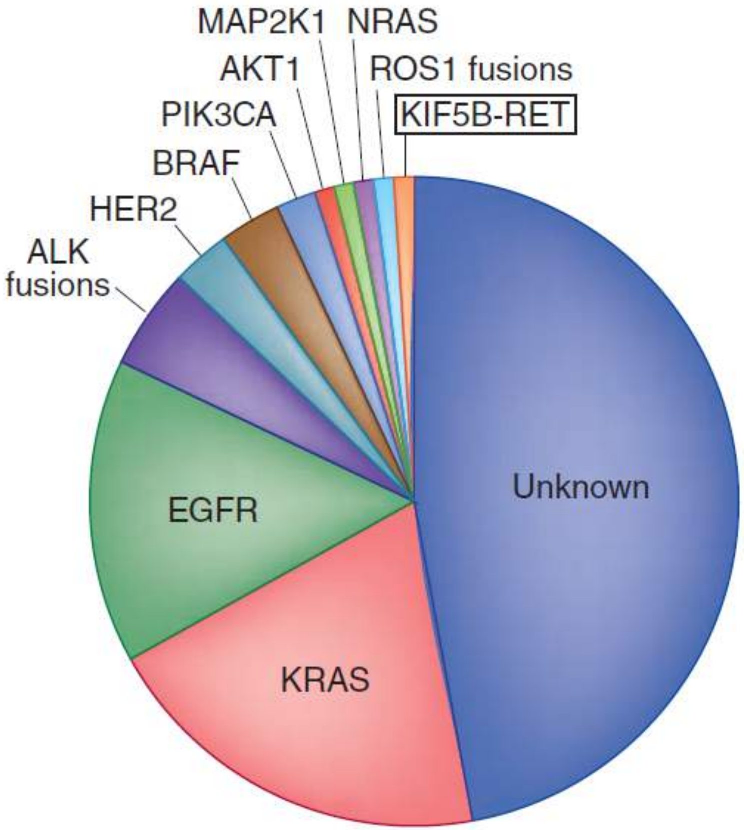
Druggable mutations in Lung Cancers

Drug target genes and Biomarkers

Lung adenocarcinoma therapeutic targets: 2012

Lung adenocarcinoma drivers

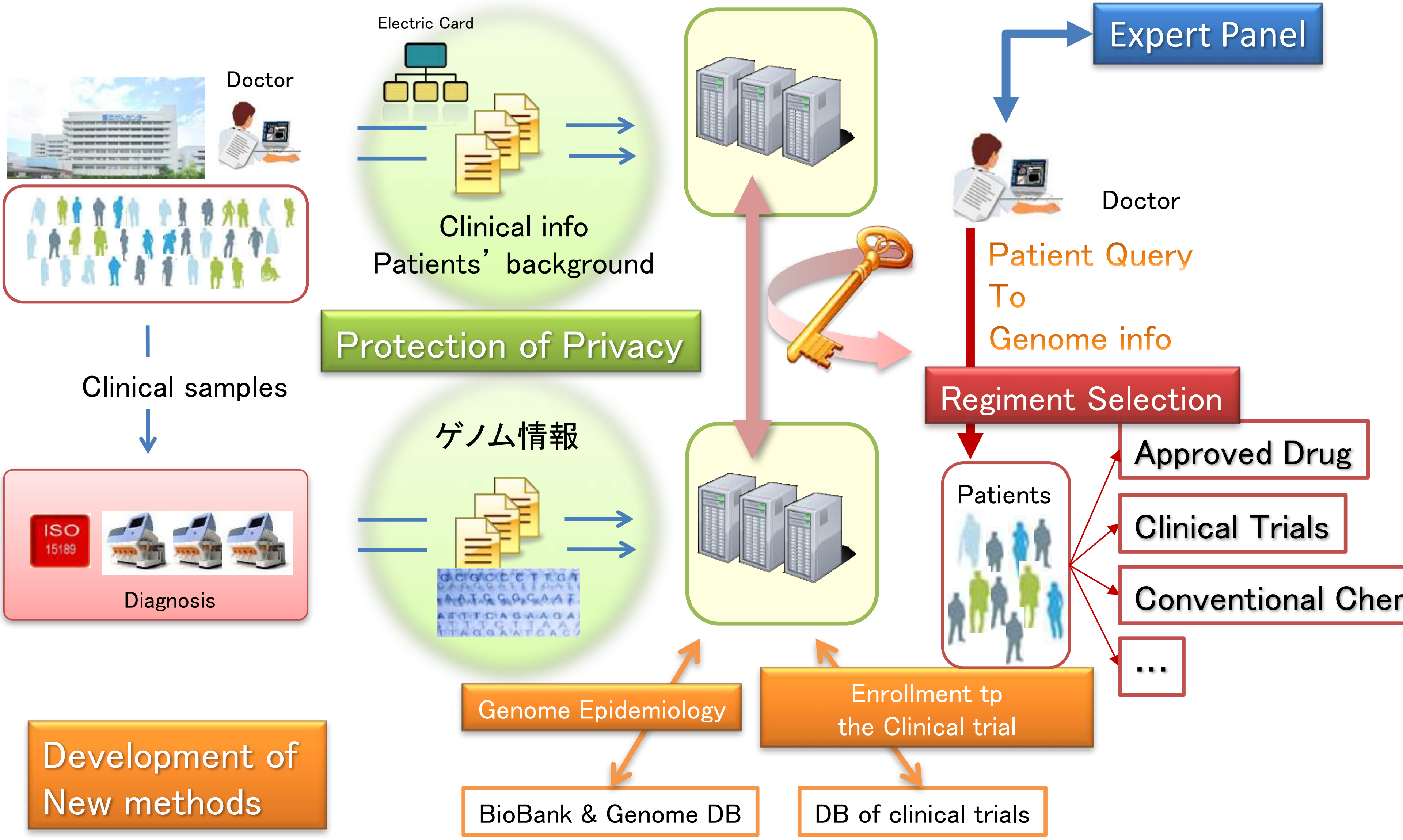
Despite the identification of molecular subsets, more than half of all lung adenocarcinomas lack an identifiable driver mutation.



Adopted from Pao and Hutchinson, 2012

NCC NATIONAL CANCER CENTER JAPAN
EPOC
Exploratory Oncology Research & Clinical Trial Center
早期・探索臨床研究センター

“Omics” information to the clinical applications



Doctor

Clinical samples

ISO 15189

Diagnosis

Electric Card

Clinical info
Patients’ background

Protection of Privacy

ゲノム情報

Expert Panel

Doctor

Patient Query
To
Genome info

Regiment Selection

Patients

Approved Drug

Clinical Trials

Conventional Cher

…

Genome Epidemiology

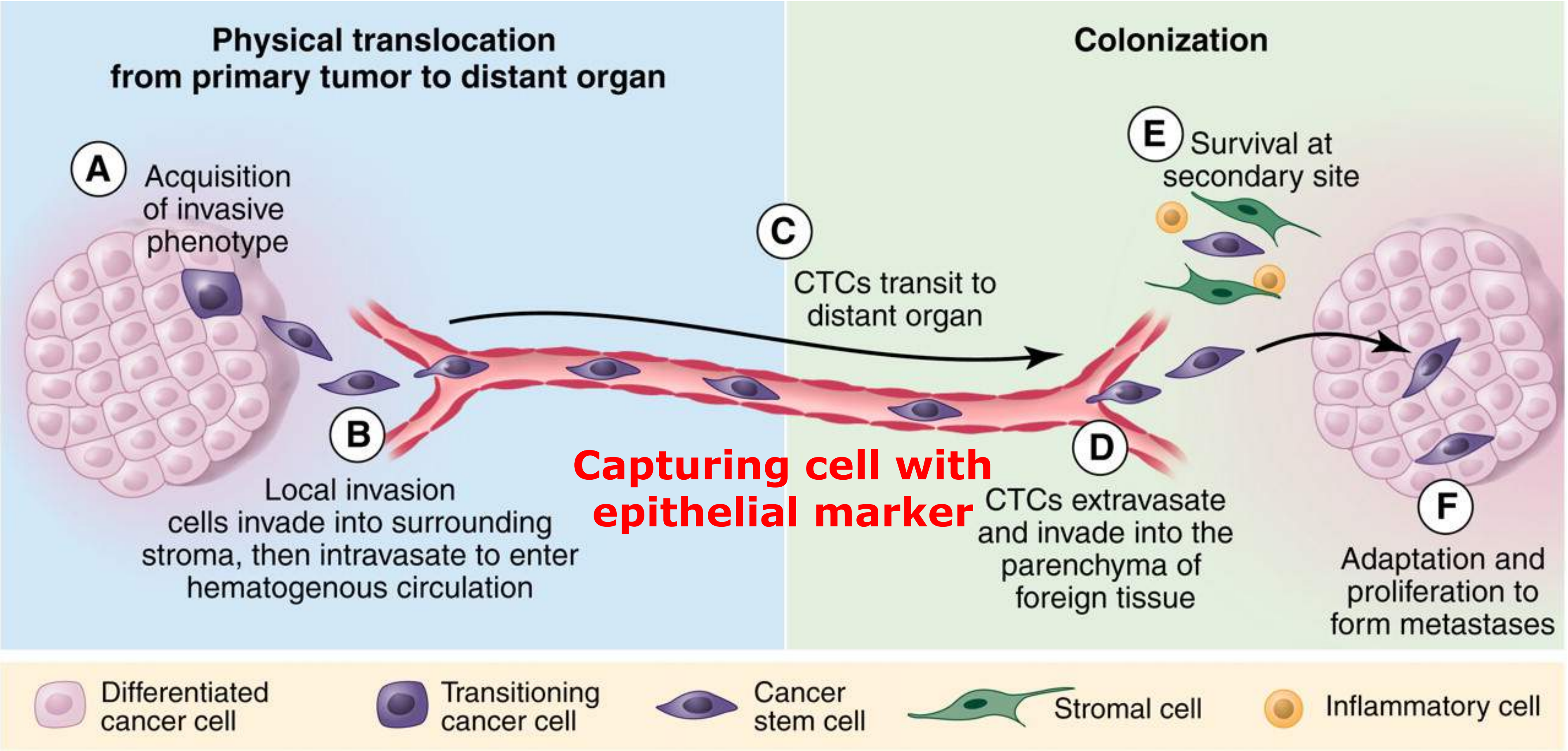
Enrollment tp
the Clinical trial

BioBank & Genome DB

DB of clinical trials

Development of
New methods

Circulating tumor cells (CTC)

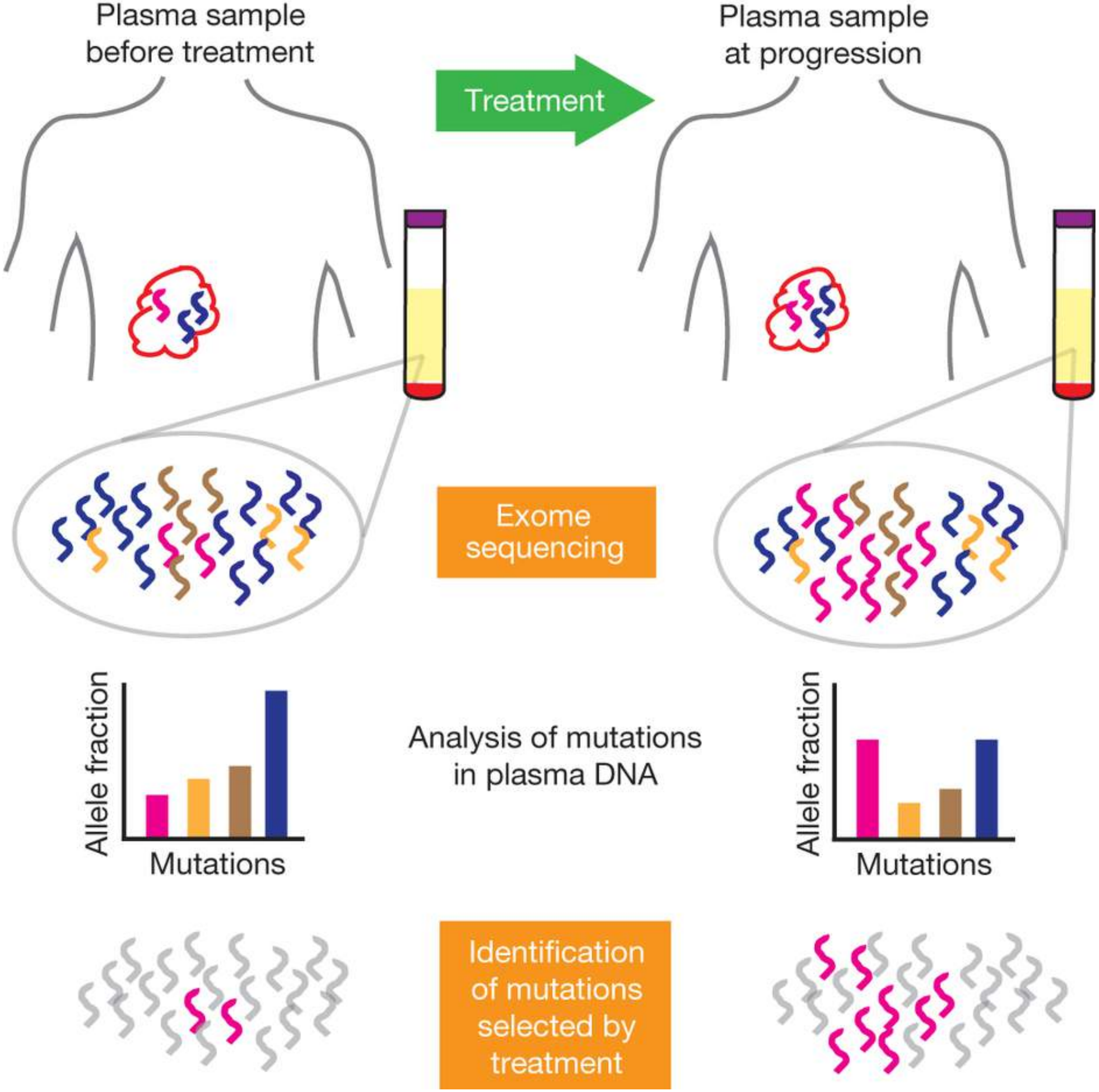


Capturing cell with epithelial marker

Primary culture: on dish or "xenograft"

Patient derived xenograft (PDX)

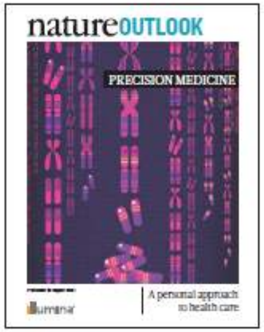
“Liquid Biopsy”: Cell-free (plasma) DNA



nature OUTLOOK

PRECISION MEDICINE

8 September 2016 / Vol 537 / Issue No 7619

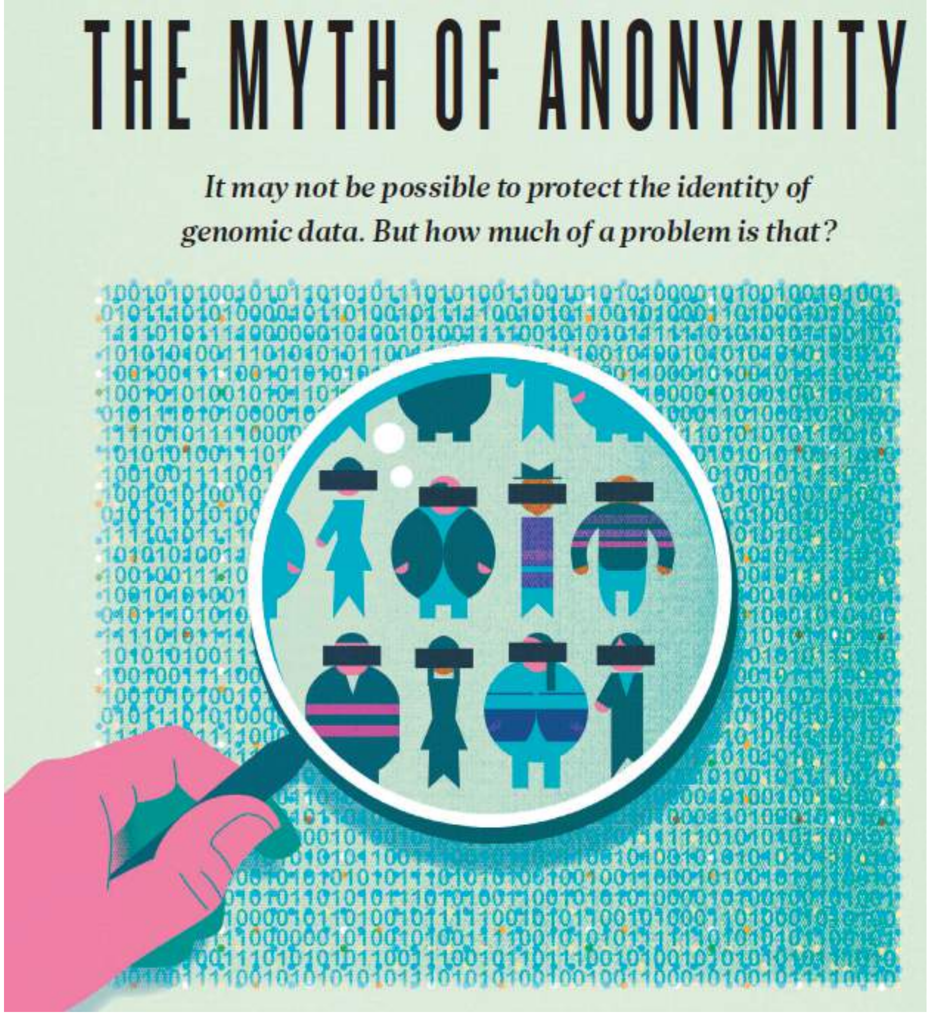


Cover art: Andrew Baker

The underlying concept of precision medicine, in which health care is individually tailored on the basis of a person's genes, lifestyle and environment, is not new: transfusion patients have been matched with donors according to blood type for more than a century (see page S52). But advances in genetics, and the growing availability of health data, present an opportunity to make precise personalized patient care a clinical reality.

Since the first human genome was sequenced in 2001, after more than a decade and at a cost of around US$3 billion,

CONTENTS



THE MYTH OF ANONYMITY

It may not be possible to protect the identity of genomic data. But how much of a problem is that?

The people who set up databases need to take a long view when making promises and asking for consent as they collect the data,

Not being clear about how participation in a study could lead to privacy breaches creates the risk that any problems that arise may make potential donors less willing to have their DNA sequenced.

We can't do research on human beings and look people in the eye and promise them that nothing bad will ever happen," Angrist says. "If we reassure people and something bad happens, then it's that much worse."

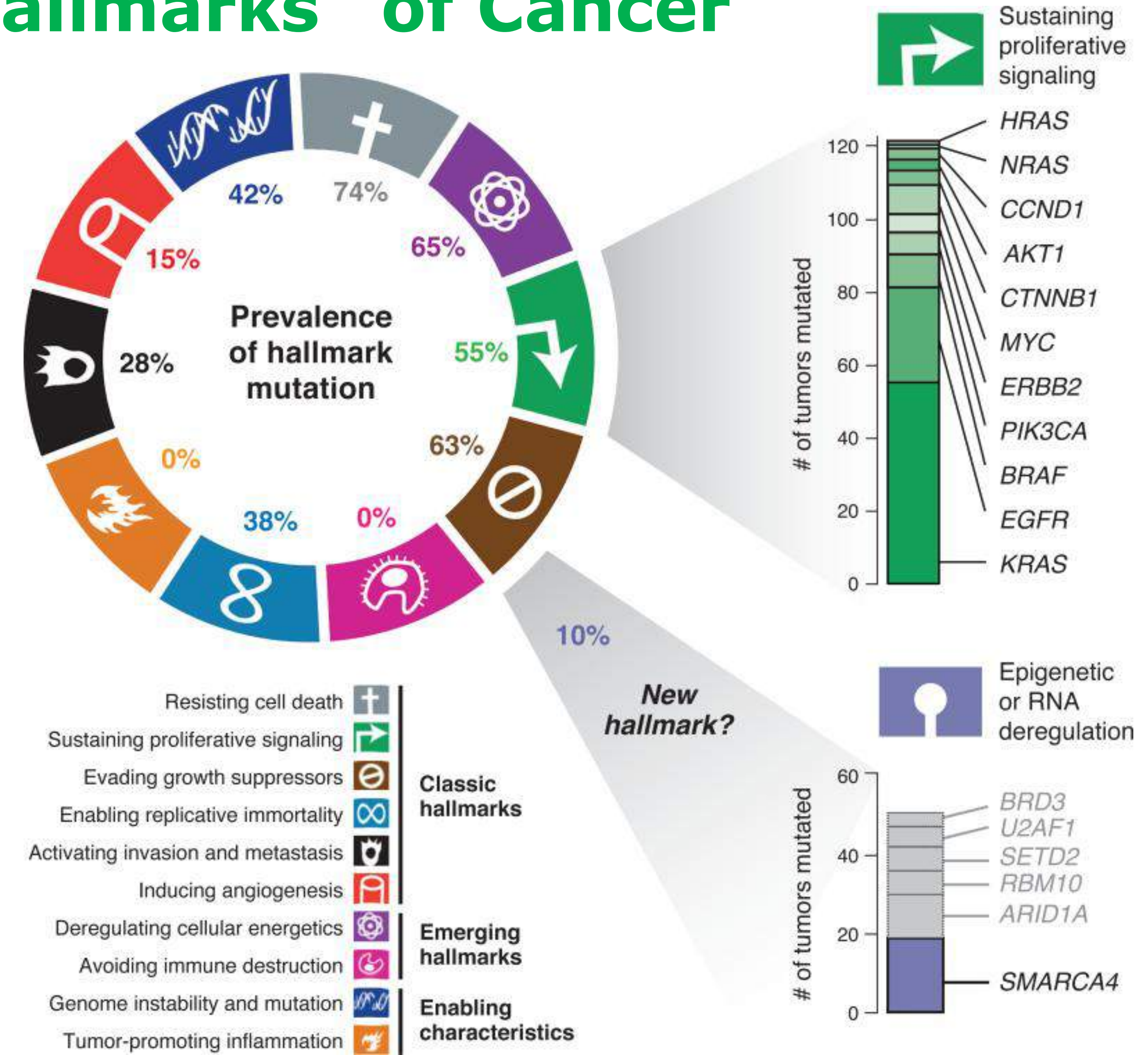
Instead, he argues, engaging with donors and spelling out the risks and benefits can change the privacy equation. "If you talk to people who have children with undiagnosed diseases, they would tell you: 'We would gladly forgo privacy in the interest of accelerated research'."

Multi-omics applications; Transcriptome and Epigenome Analysis of cancer cells

Yutaka Suzuki
Department of Computational Biology and Medical Sciences
Graduate School of Frontier Sciences
The University of Tokyo

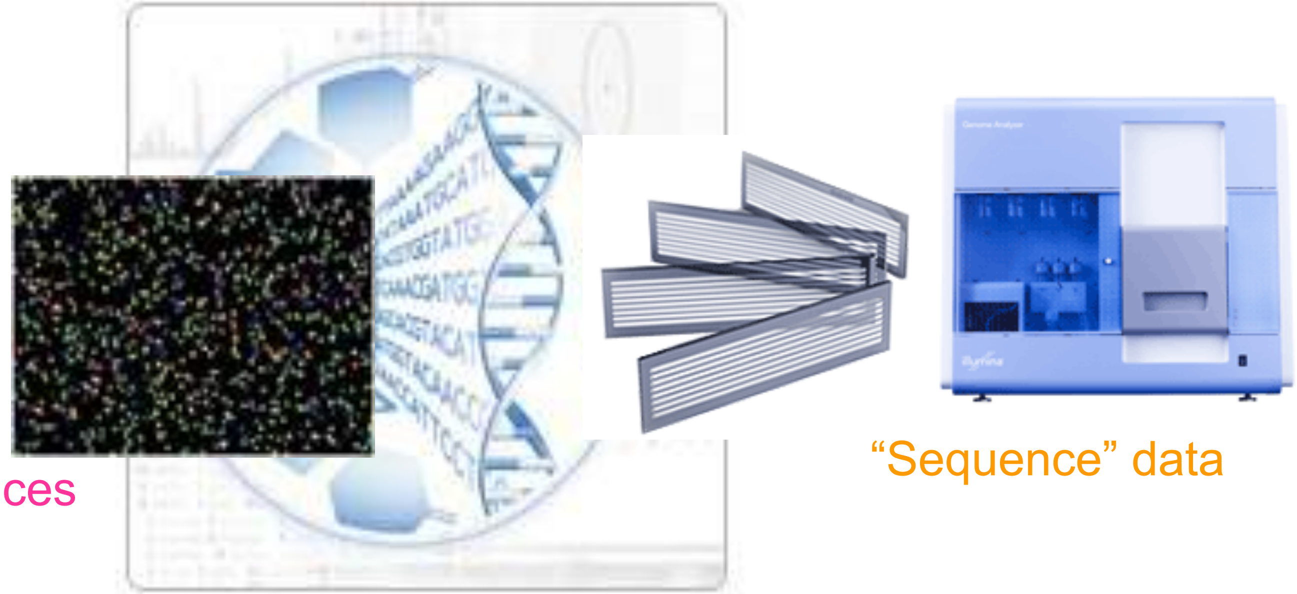
"Hallmarks" of Cancer

Imielinski et al Cell (2012)



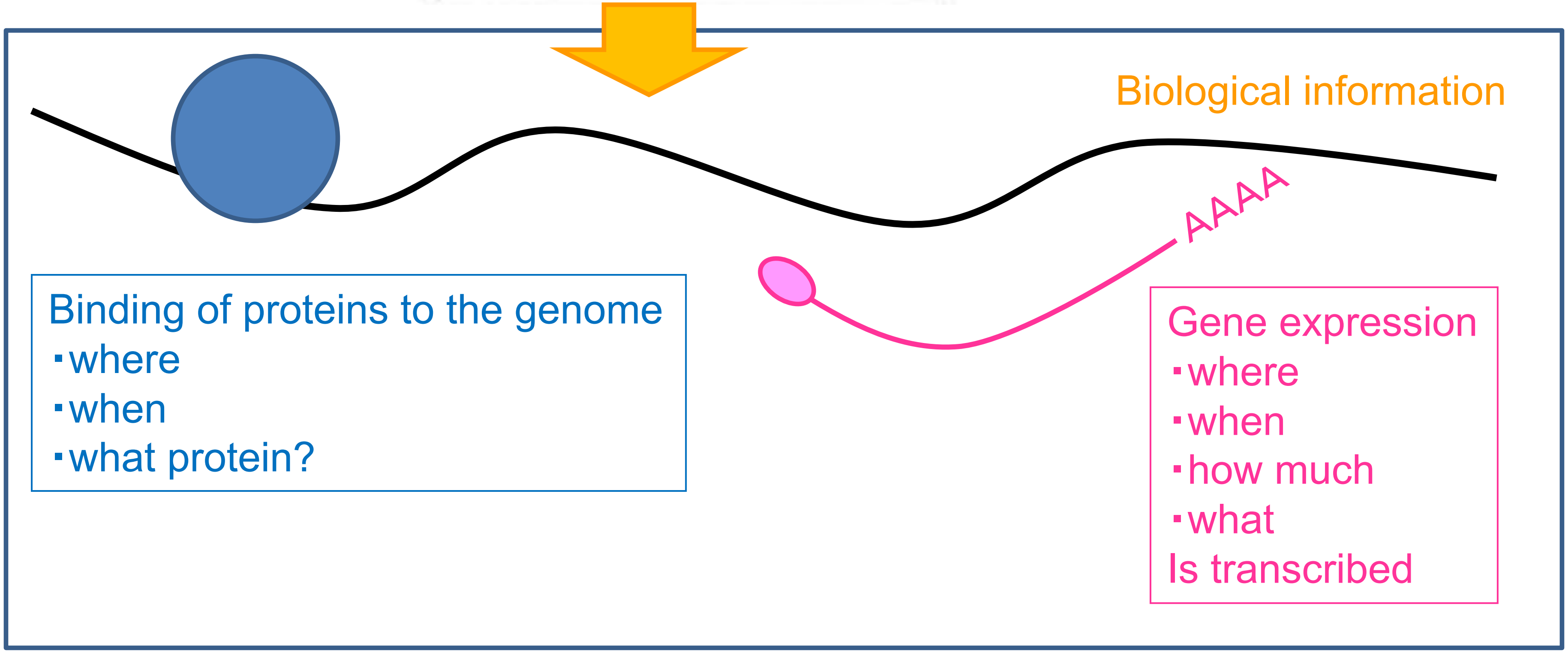
Biological information can be also read by NGS

“NGS”



Numerous sequences

“Sequence” data



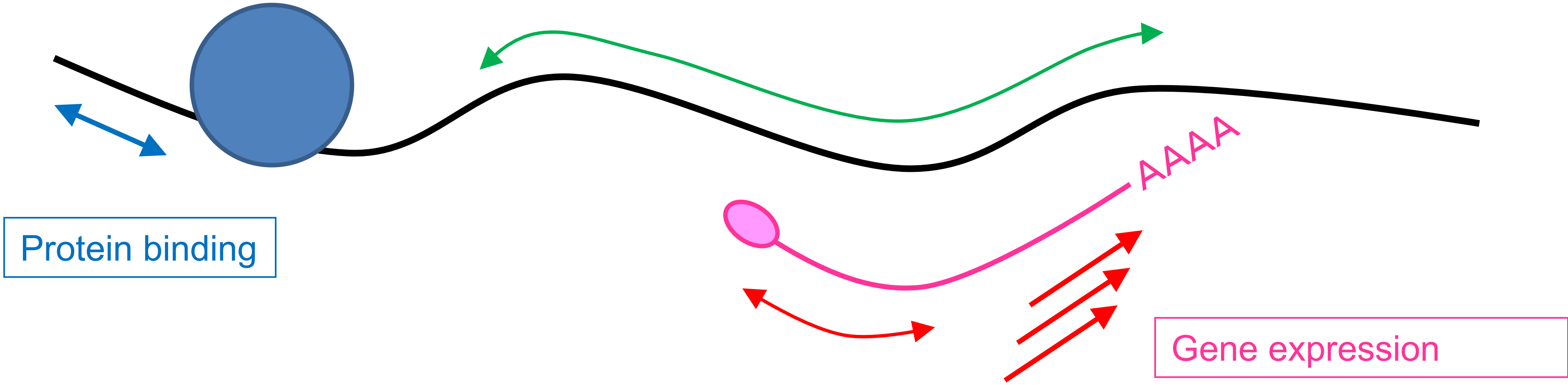
Biological information

Binding of proteins to the genome
- where
- when
- what protein?

Gene expression
- where
- when
- how much
- what

Is transcribed

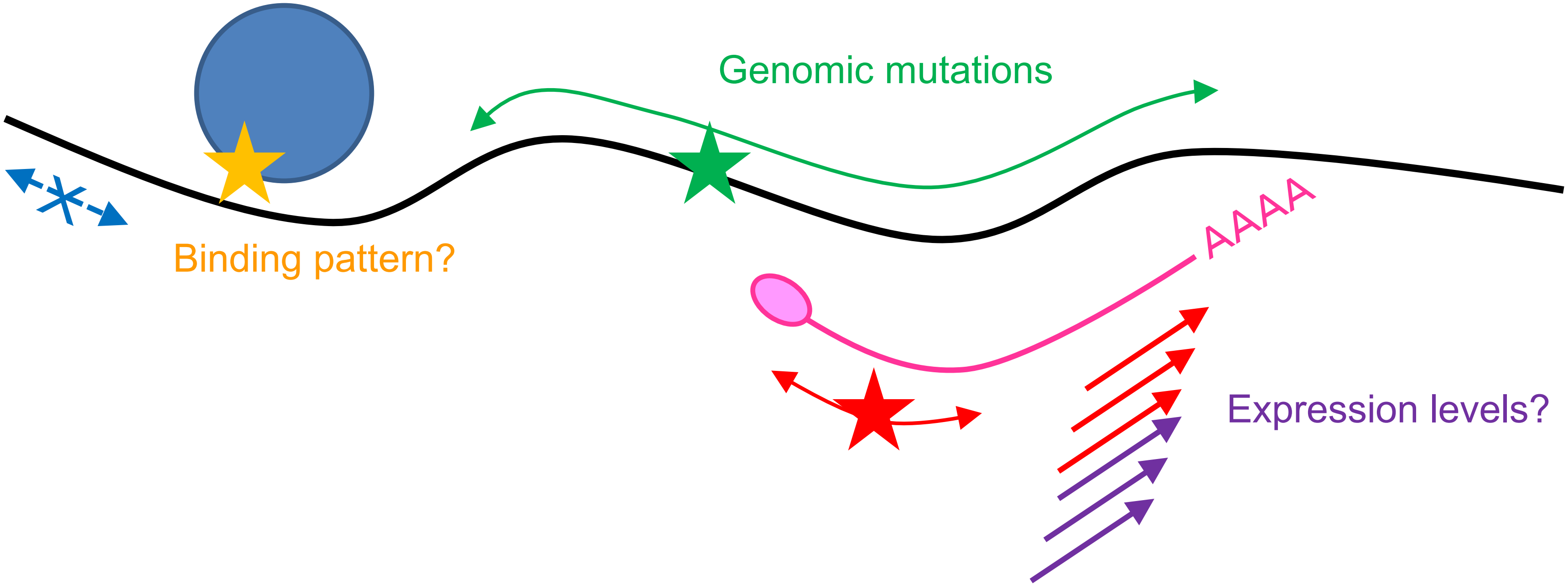
“Normal cells”



Protein binding

Gene expression

“Diseased” cells



Genomic mutations

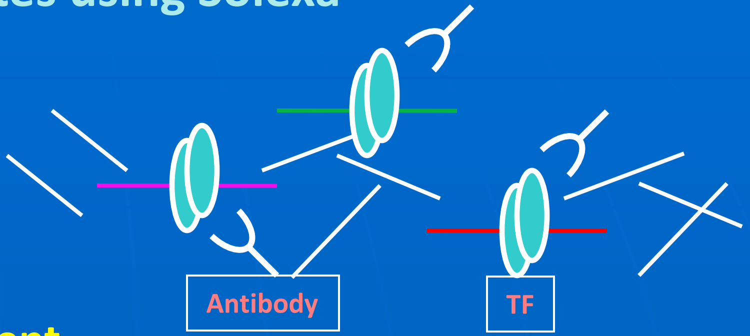
Binding pattern?

Expression levels?

ChIP-"Seq": Identification of TF binding sites using Solexa

Chromatin IP

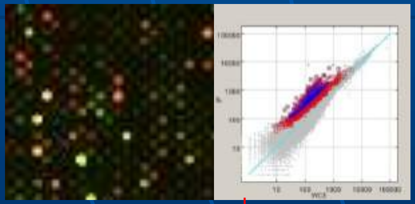
Recovery of DNA fragment



ChIP on CHIP

Linker ligation & PCR

Hybridize to Genome/ Promoter Array



ChIP on SEQ

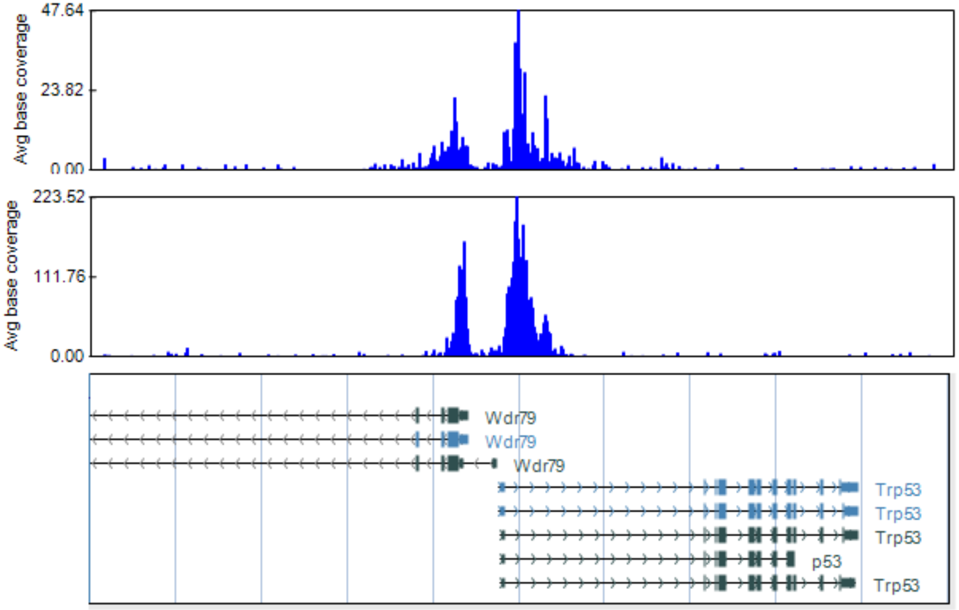
Solexa-linker ligation & Sequencing

Mapping to Genome and TFBS identification

Whole-Genome & Quantitative

Transcription Fctors

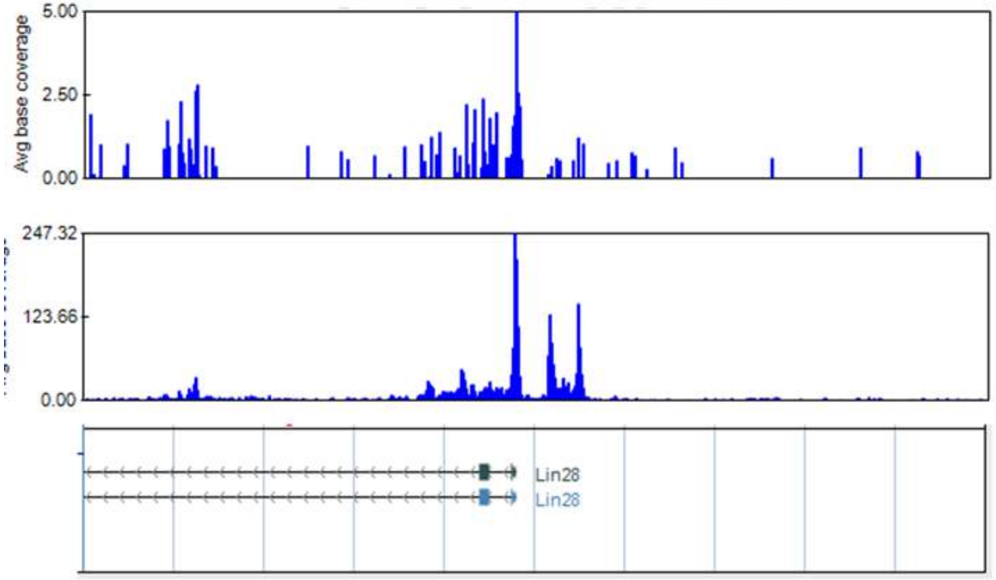
TFx



18,950,096

66,613,786

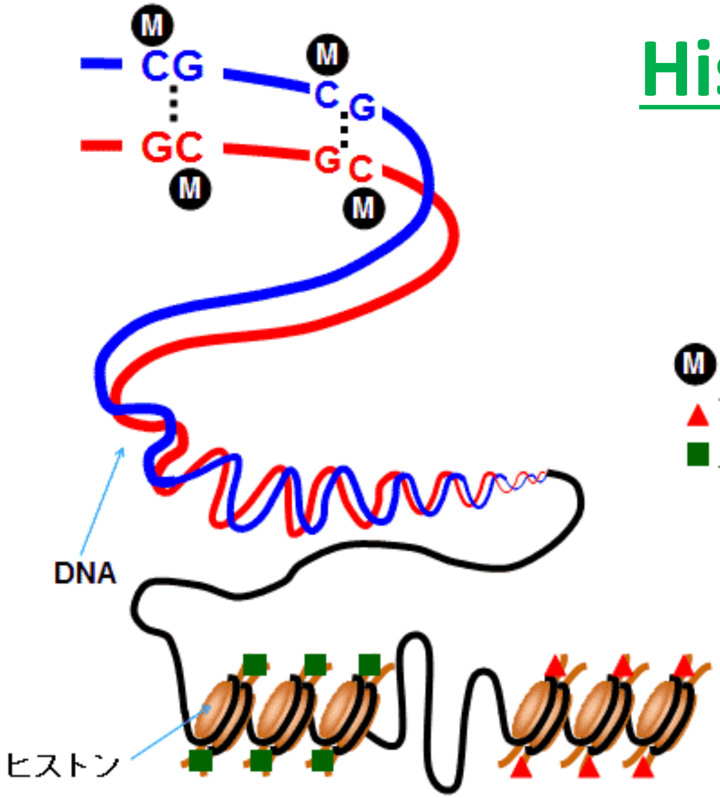
Pol2



7,118,567 tags

71,172,337 tags

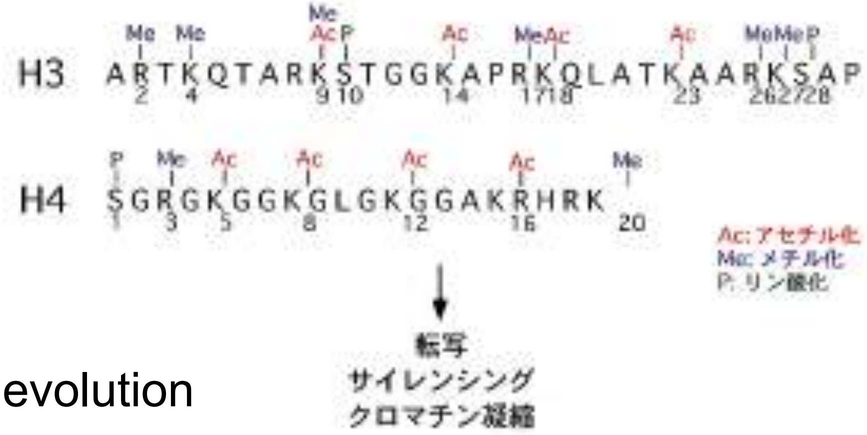
Histone Modifications



● DNAにつく目印（DNAメチル化）

▲ 🟩 ヒストンにつく目印（ヒストン修飾）

Frequently analyzed sites：K4, 9, 27, 36



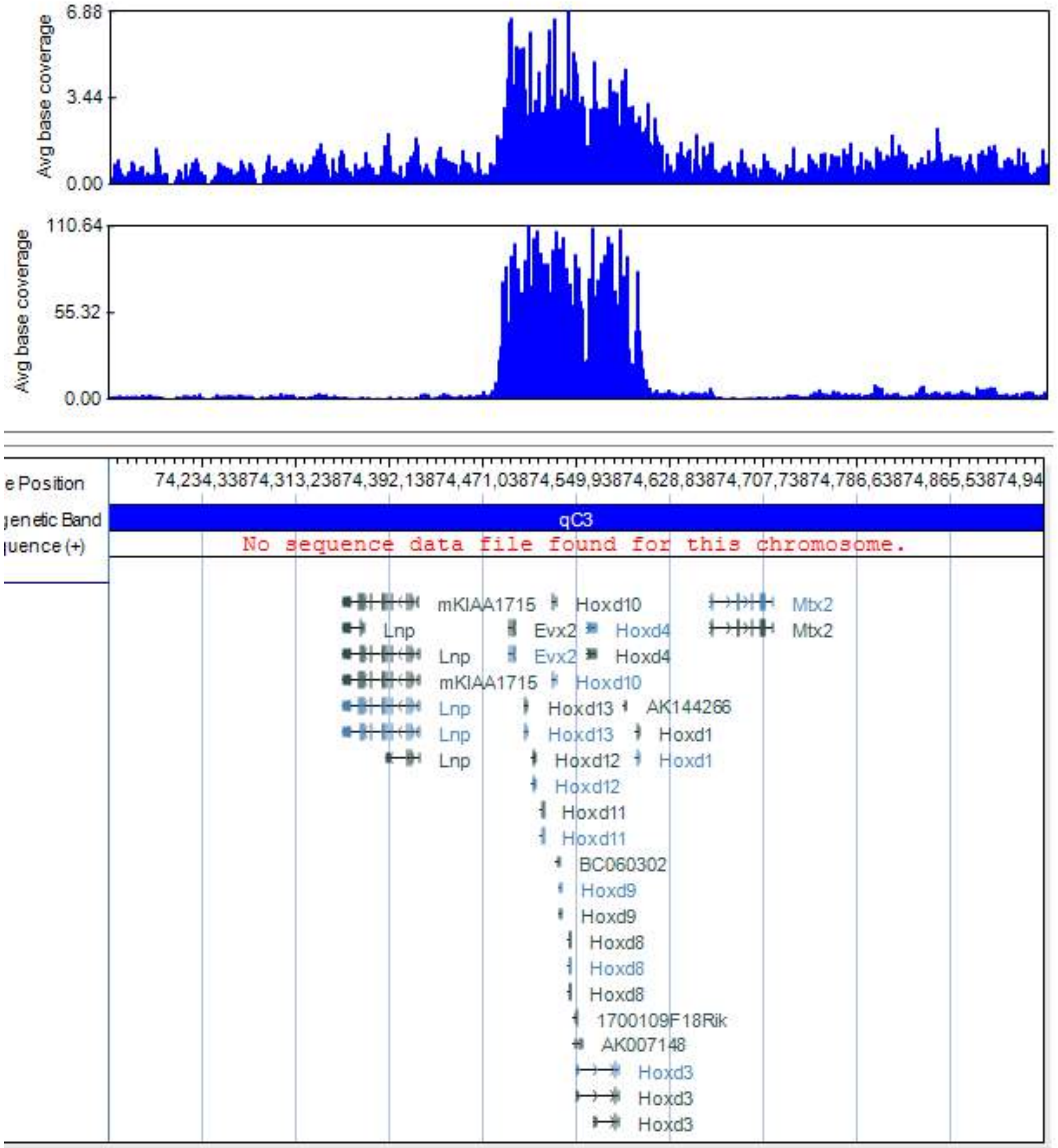
Amino Acids conserved throughout the evolution

“Epigenome” Markers

IP	Marker
H3K4me3	Active, Promoter
H3K4/9ac	Active
Pol II	Active
H3K36me3	Active, Elongation
H3K9me3	Silent, Heterochromatin
H3K27me3	Silent, PRC2
H3K4me1	Active, Enhancer
H3K27ac	Active, Enhancer/Promoter

Repressive marks

H3K27me3



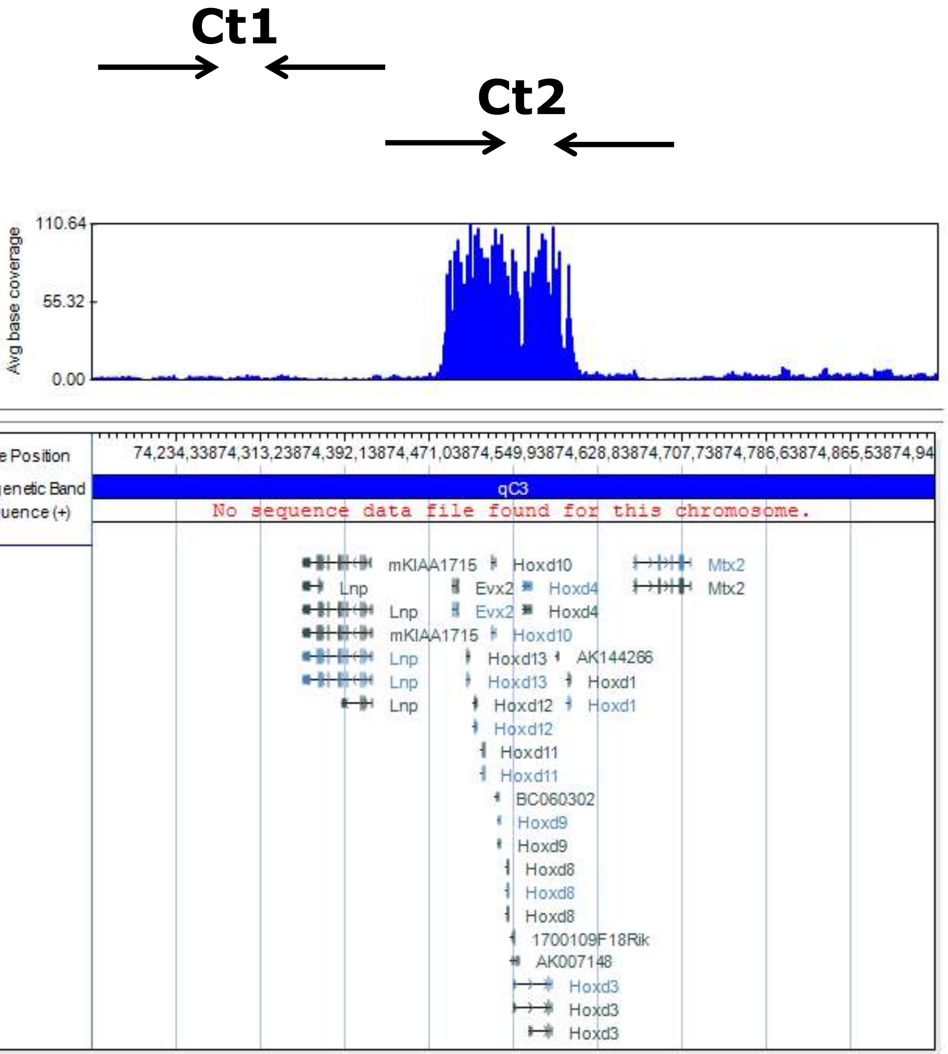
21,613,561 tags

68,149,639 tags

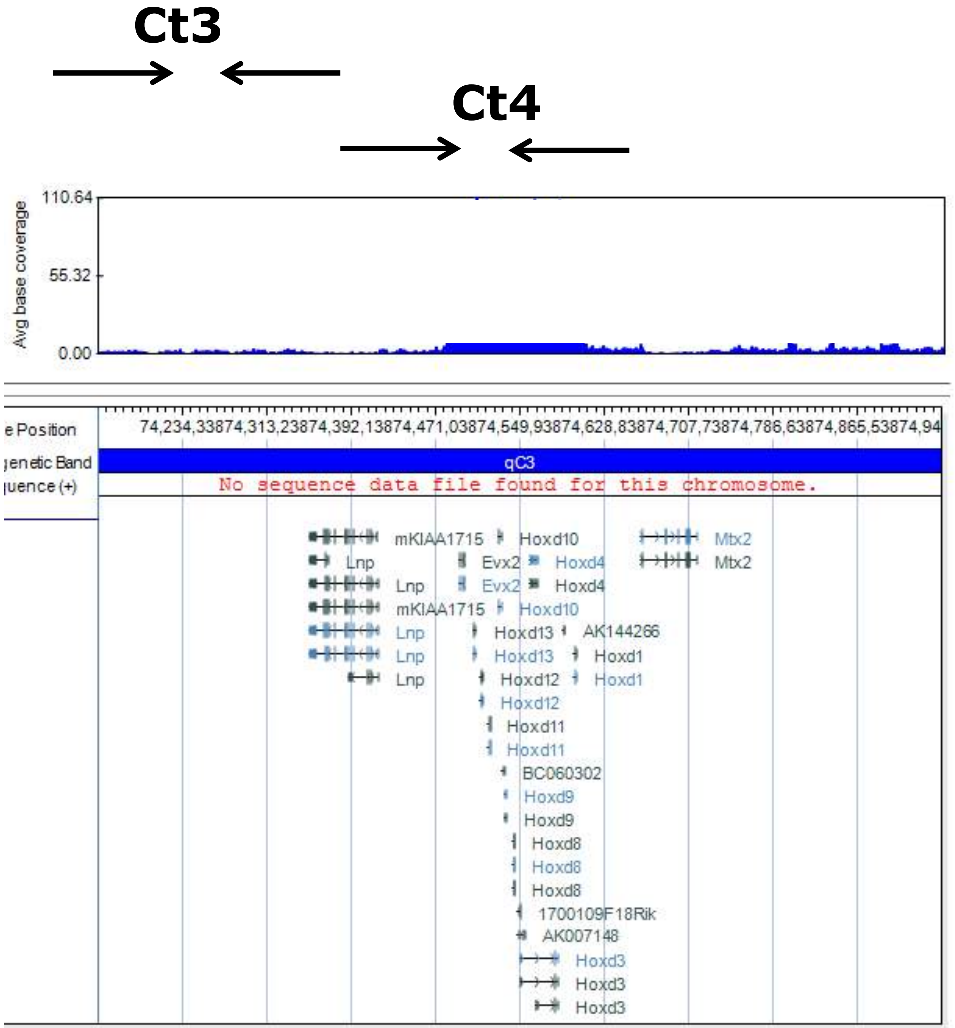
qRT-PCR validation

※: Ct= log2(intensity)

$$\text{Fold enrich} = \frac{(\Delta Ct_{\text{Peak}})}{(\Delta Ct_{\text{Random}})} = \frac{(Ct4-Ct2)}{(Ct3-Ct1)}$$



IP sample

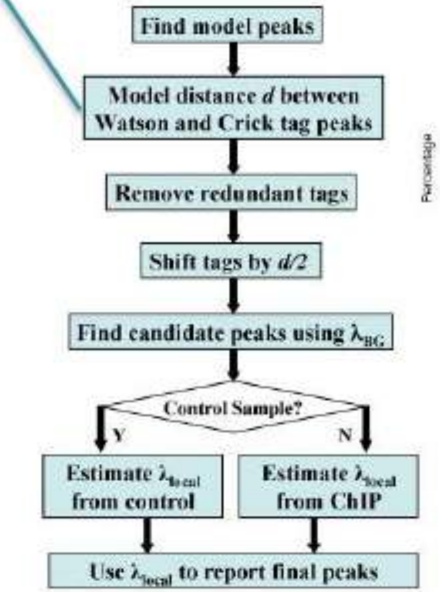


WCE (input) sample

ChIP-seq analysis

Peak calling – MACS

- MACS select at least 1,000 "model peaks" for calculating the distance "d" between paired peaks.



Find model peaks

Model distance *d* between Watson and Crick tag peaks

Remove redundant tags

Shift tags by *d/2*

Find candidate peaks using λ_{BG}

Control Sample?

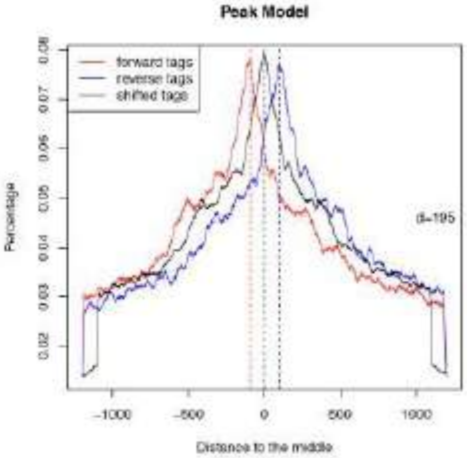
Y

N

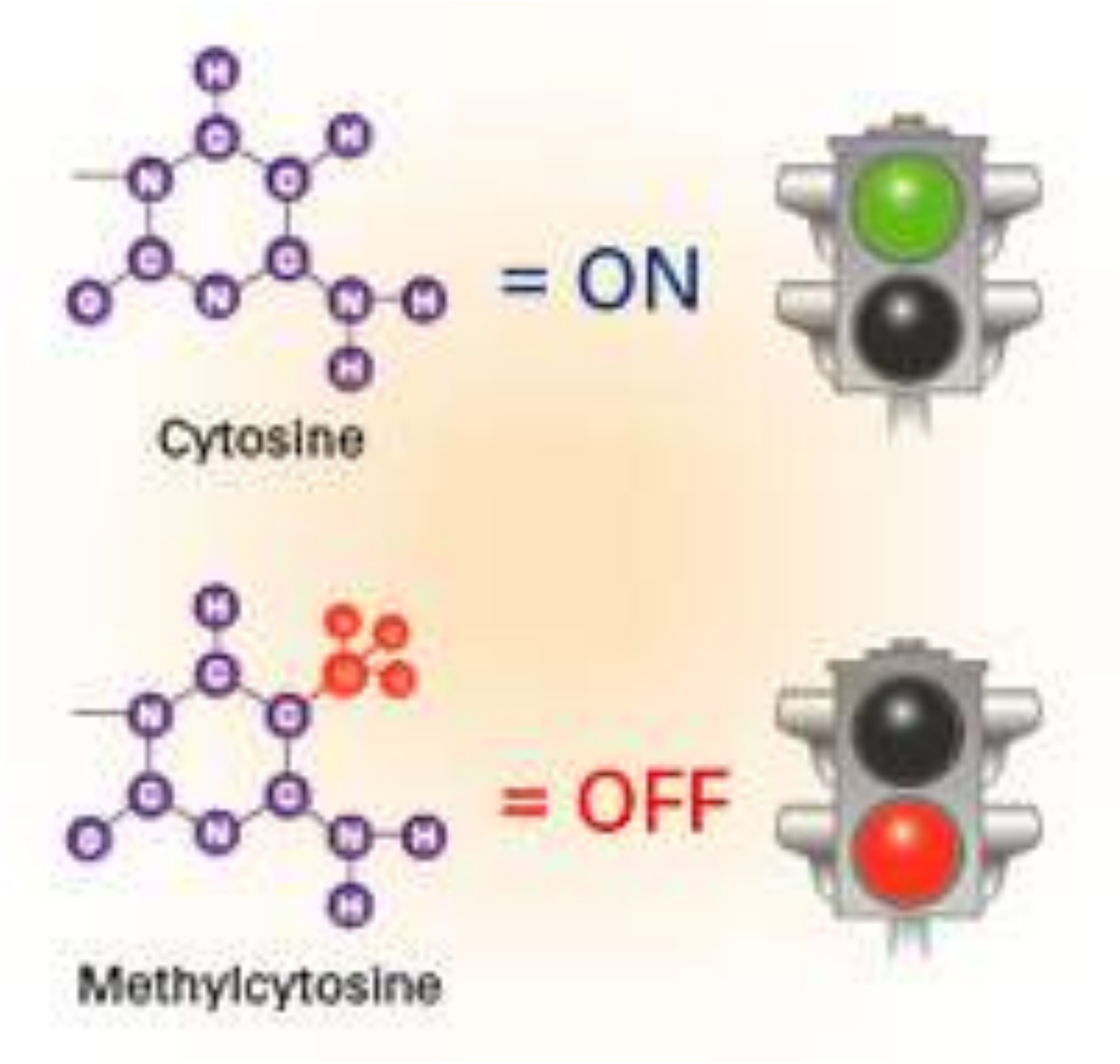
Estimate λ_{local} from control

Estimate λ_{local} from ChIP

Use λ_{local} to report final peaks

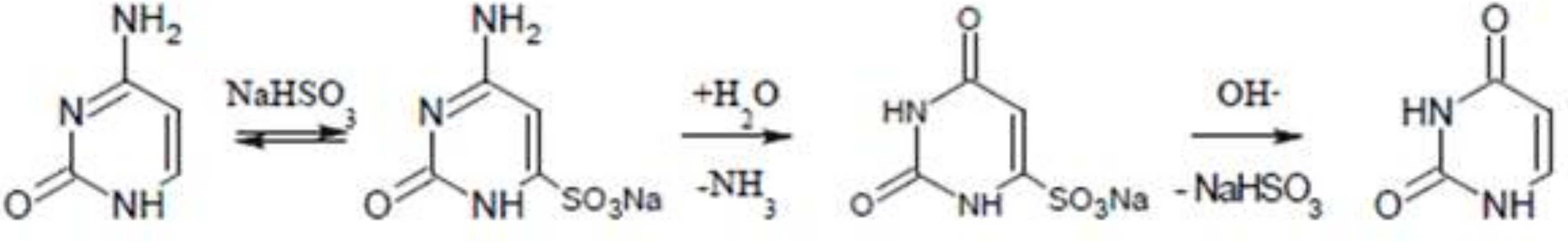


DNA methylation（CpG methylation）



Bisulfite Sequencing

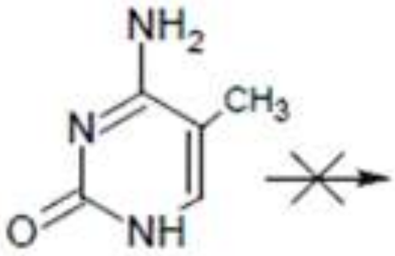
Non-methylated C



Cytosine

Uracil

Methylated C



5-Methyl Cytosine

Bisulfite Sequencing cont.

Bisulfite Treatment: C->U conversion

- Non-methyl C: C-> T
- Methyl CX: C-> C

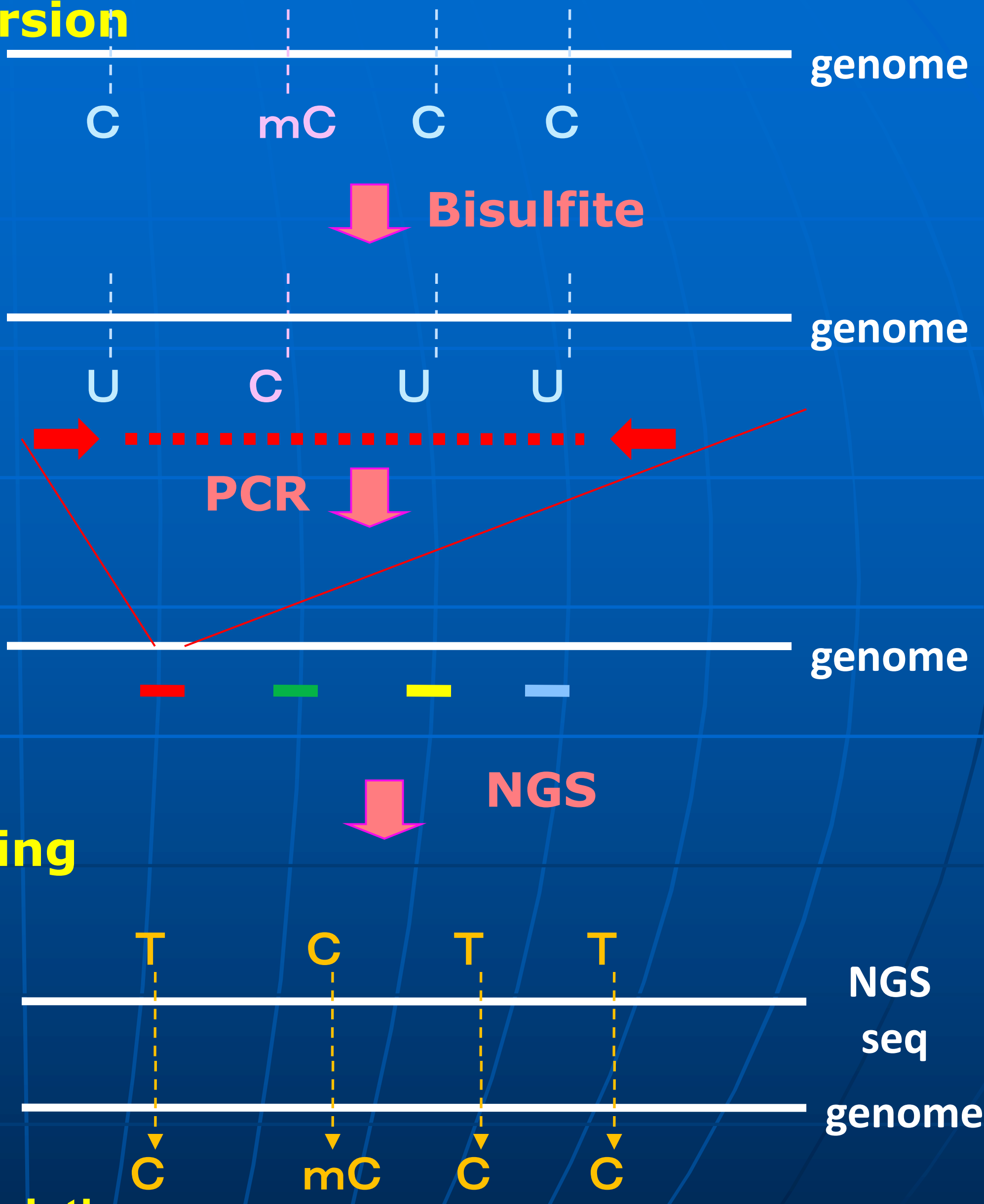
(PCR for the Target region)

(Mix 100-1000 fragments)

NGS Adaptor ligation & Sequencing

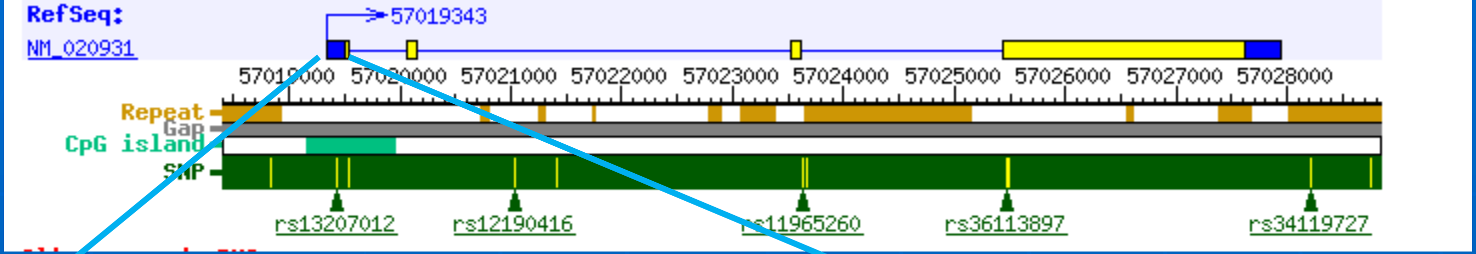
Mapping to Genome

Identification of methylation/non-methylation

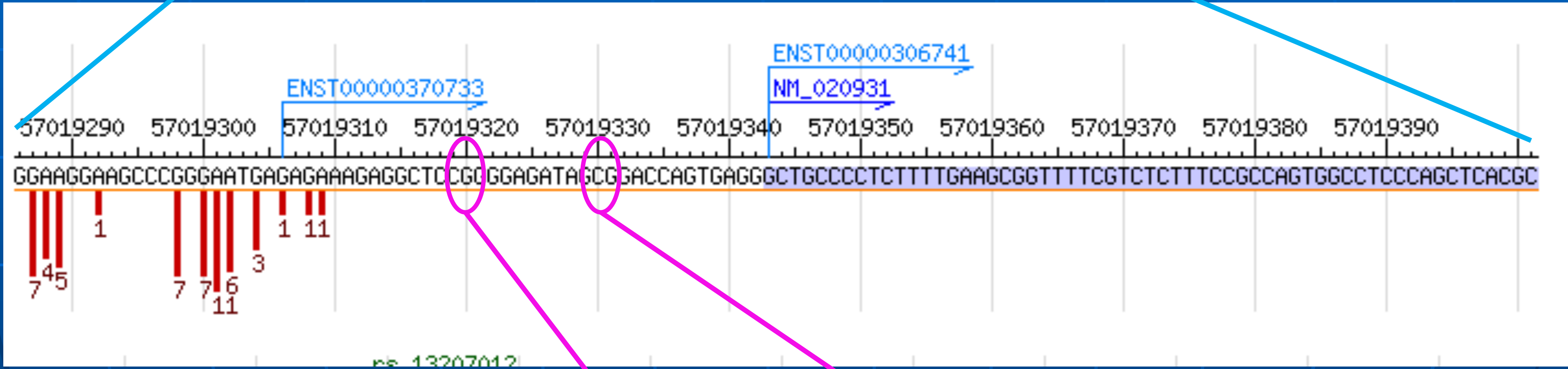


DNA Methylation by BS sequencing

DNA Methylation in the KIAA1586 gene



HEK293 cell

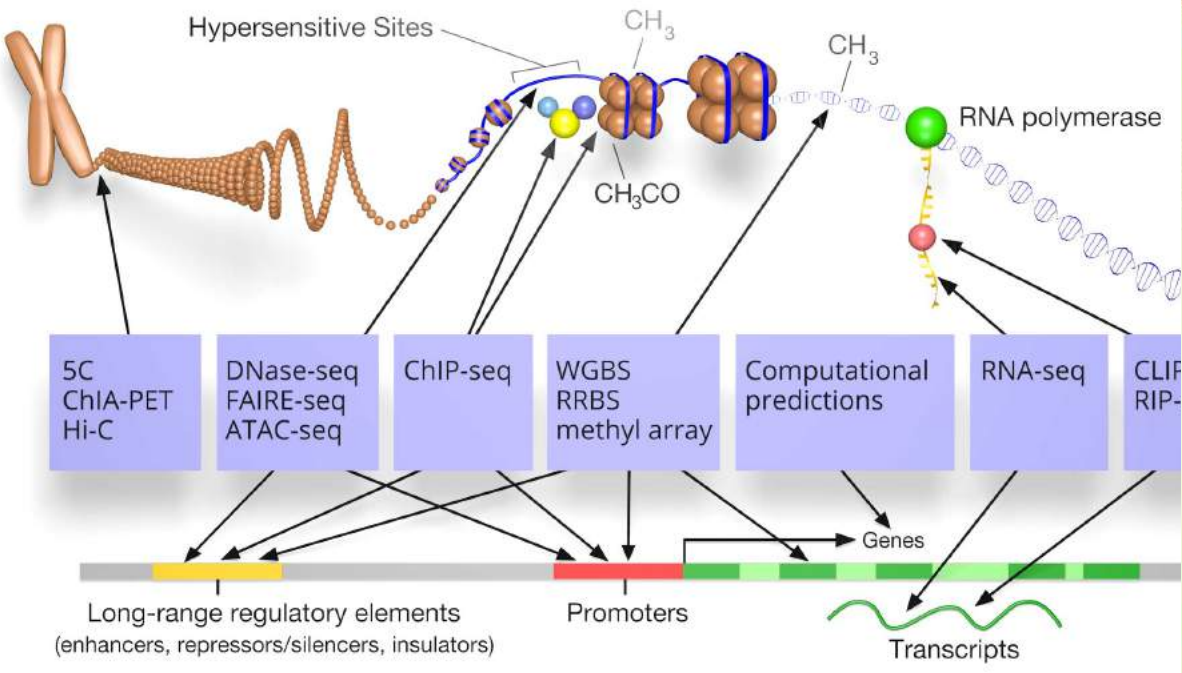


mC/C (%)

0/16 (0 %)

1/15(7 %)

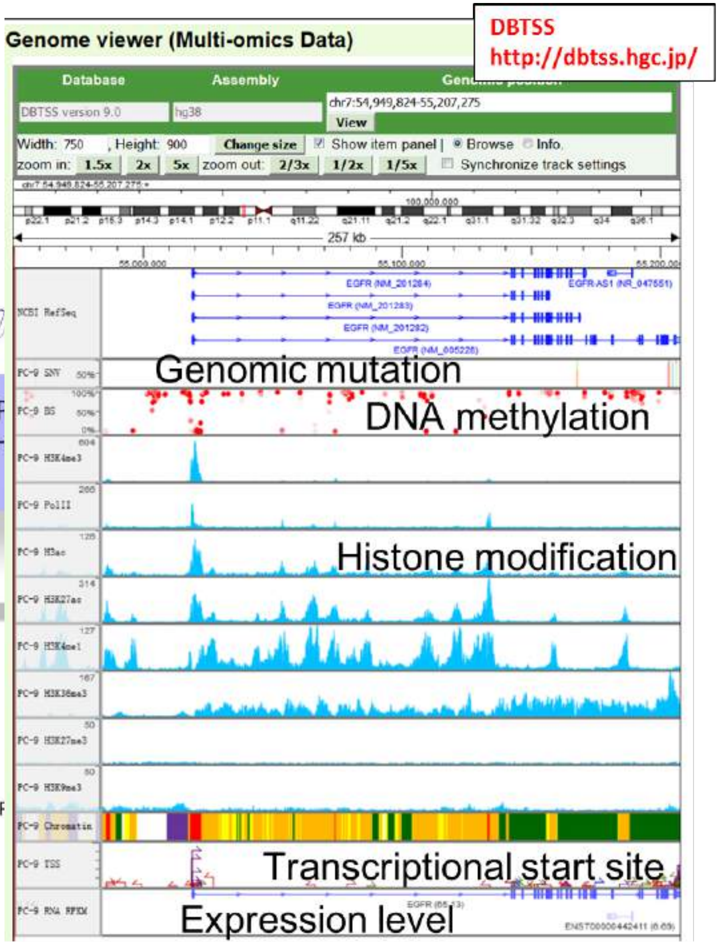
Epigenome Catalogue in public databases



ENCODE

Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael P

DBTSS
http://dbtss.hgc.jp/



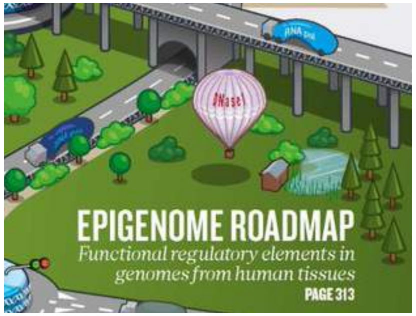
Genome viewer (Multi-omics Data)

Database: DBTSS version 9.0
Assembly: hg38
Gen… position: chr7:54,949,824-55,207,275

Width: 750, Height: 900 ☑ Show item panel | ● Browse ○ Info
zoom in: 1.5x 2x 5x zoom out: 2/3x 1/2x 1/5x ☐ Synchronize track settings

Suzuki et al. 2014 *Nucleic Acids Research*
Suzuki et al. 2015 *Nucleic Acids Research DB issu*

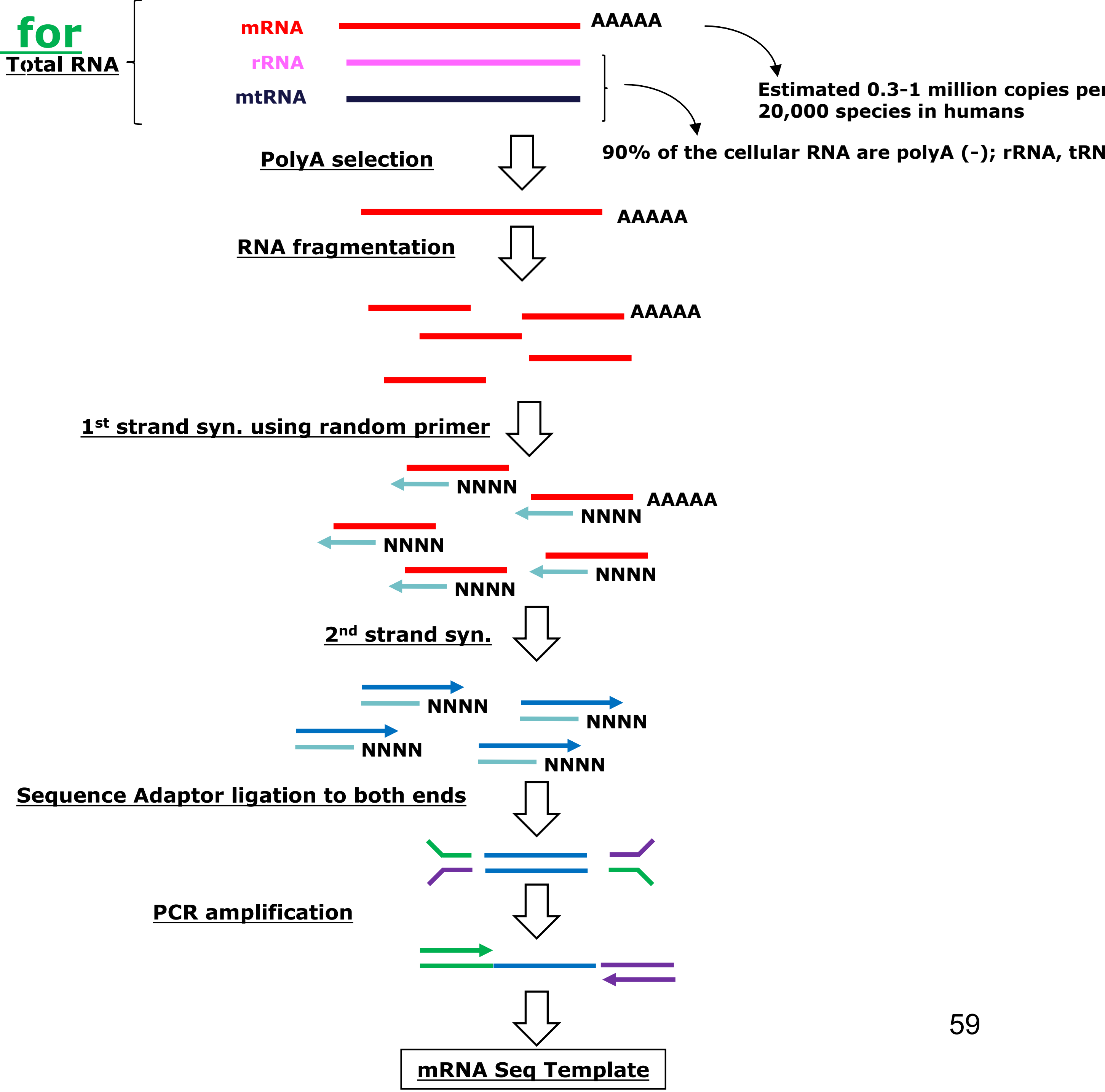
ational Human Genome Research Institute
ENCODE



IHEC celebrates major coordinated paper release

Cell Press Special Edition
International Human Epigenome Consortium Collection

Template Prep. for RNA Seq



Total RNA

mRNA AAAAA

rRNA

mtRNA

Estimated 0.3-1 million copies per
20,000 species in humans

90% of the cellular RNA are polyA (-); rRNA, tRN

PolyA selection

AAAAA

RNA fragmentation

AAAAA

1st strand syn. using random primer

NNNN AAAAA NNNN NNNN NNNN NNNN

2nd strand syn.

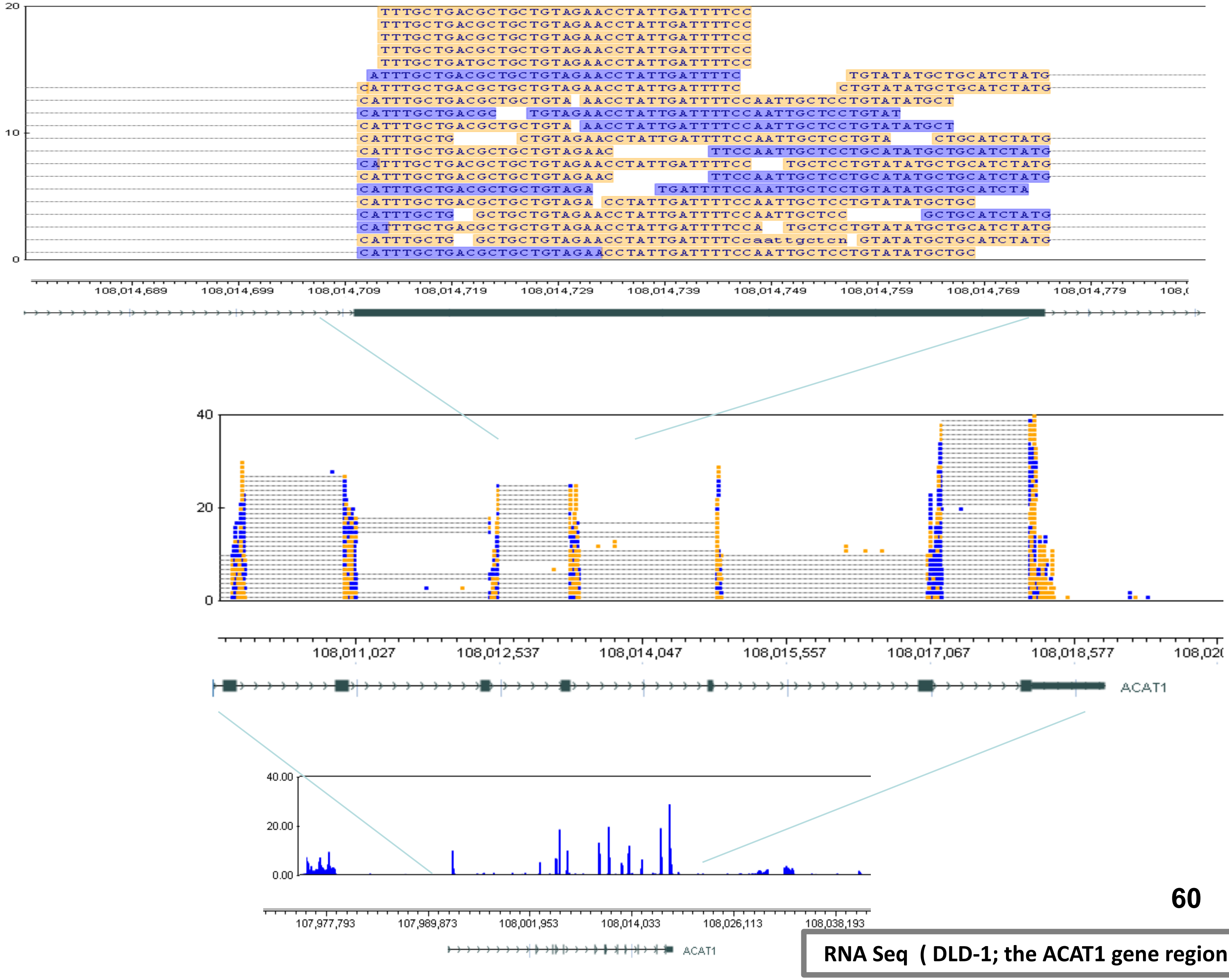
NNNN NNNN NNNN NNNN

Sequence Adaptor ligation to both ends

PCR amplification

mRNA Seq Template

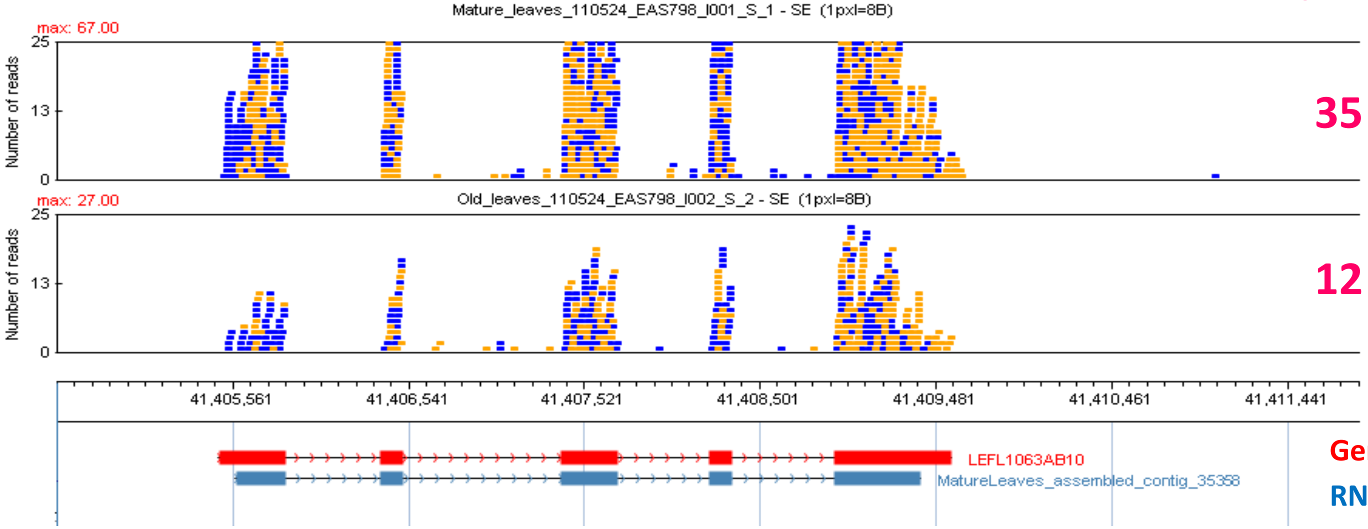
Examples of NGS data (RNA Seq on Genome Studio Viewer)



RNA Seq (DLD-1; the ACAT1 gene region)

RNA Sequencing for gene expression analysis

Expression level



35 rpkm

12 rpkm

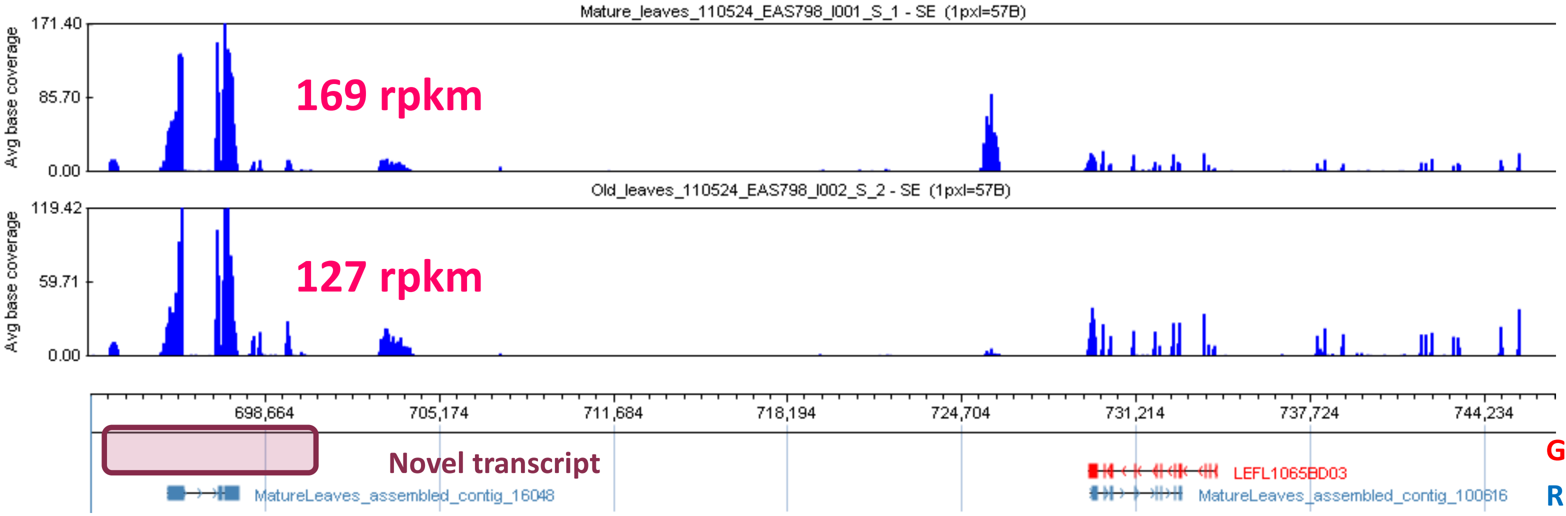
Gene model

RNA Seq assembled contig

(rpkm: read per million tags per kb mRNA)

Novel transcripts

Expression level



169 rpkm

127 rpkm

Novel transcript

Gene model

RNA Seq assembled contig

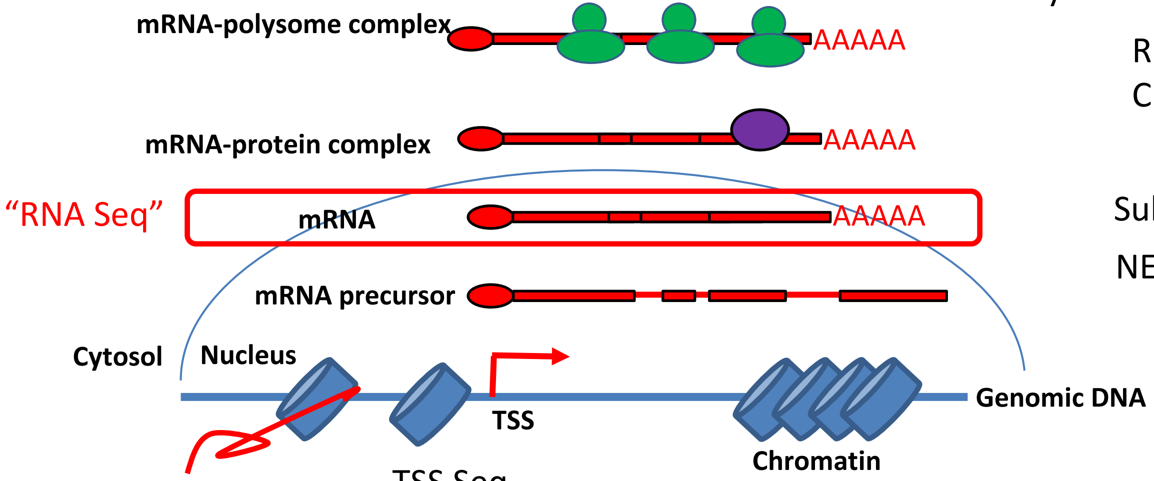
RNA Seq for various purposes

Sub-population of RNAs

Polysome for translational control

Ribosome Profiling
Polysome Seq

RIP Seq
CLIP Seq



SubCellular Seq

NET Seq

TSS Seq

GRO Seq

RNA-Chromatin factor interactions

RNA sub-structures

RNA length
RNA Seq
miRNA Seq

RNA sequence
rRNA depletion

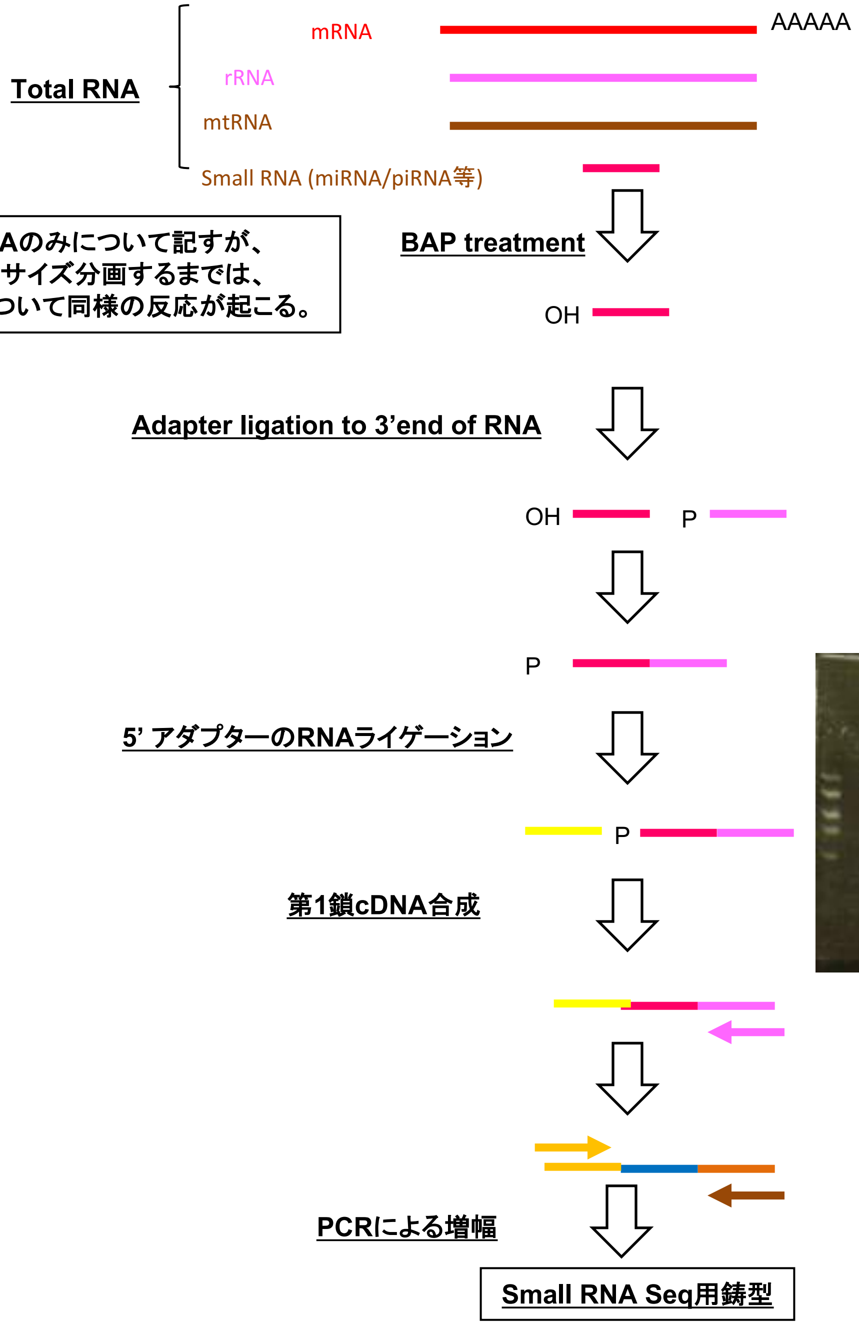
Modified RNA base

Transcriptional Start Sites

CAGE
TSS Seq
START Seq

polyA
(transcriptional termination sites)

CEL Seq
polyA Seq


Total RNA

mRNA AAAAA

rRNA

mtRNA

Small RNA (miRNA/piRNA等)

BAP treatment

OH

Adapter ligation to 3’end of RNA

OH P

P

5’ アダプターのRNAライゲーション

P

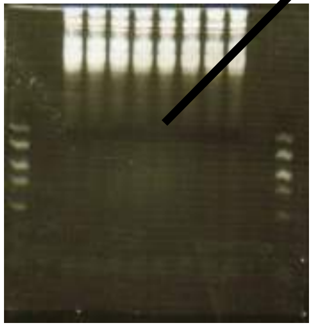
第1鎖cDNA合成

PCRによる増幅

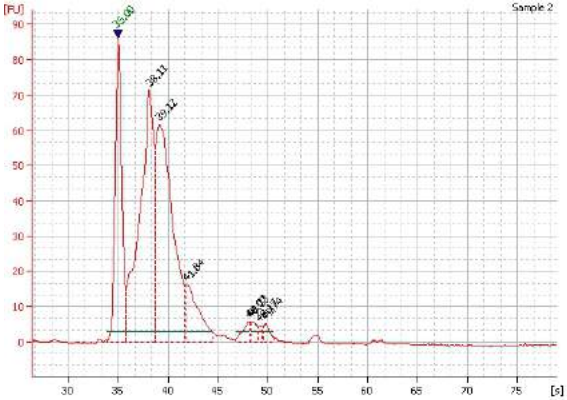
Small RNA Seq用鋳型

※図はsmall RNAのみについて記すが、
最後のステップでサイズ分画するまでは、
すべてのRNAについて同様の反応が起こる。

	Takara Protocol	Illumina protocol (v1.5)
Total RNA input	**100ug**	**1ug**
Size selection	**Needed**	**Not needed**

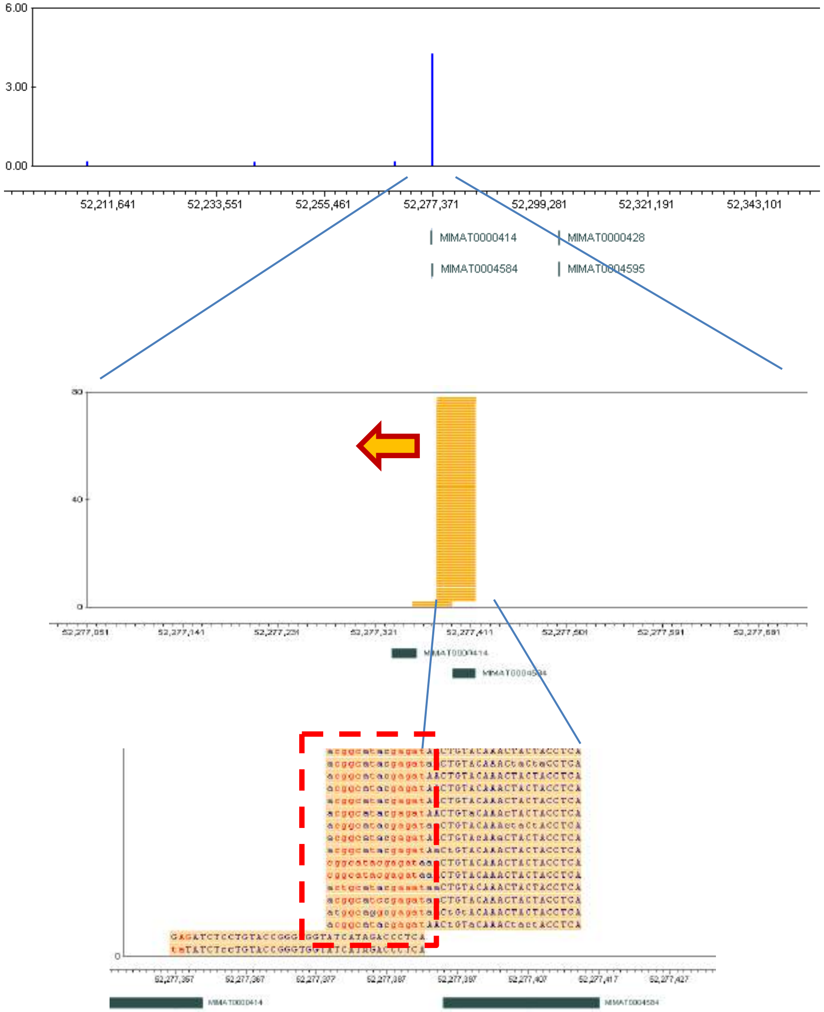


約18 nt~30 nt分画のSmall RNAを単離



Kanematsu et al Gene 2012

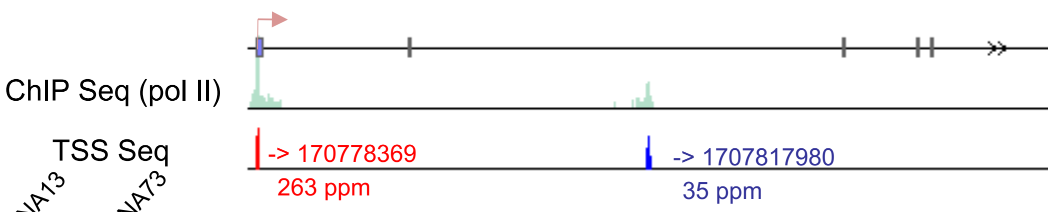
small RNA Seq (DLD-1; the MIMAT0004584 gene region)



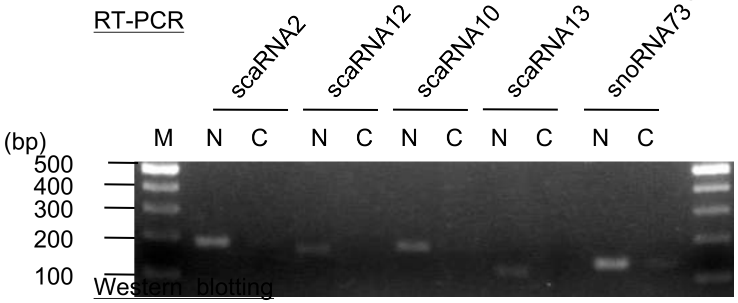
Filtration of TSSs using various types of NGS analysis

TSS Seq/ChIP Seq

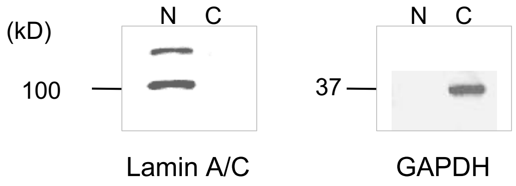
NM_001829 (the chloride channel 3 gene)

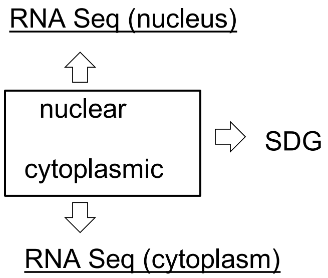


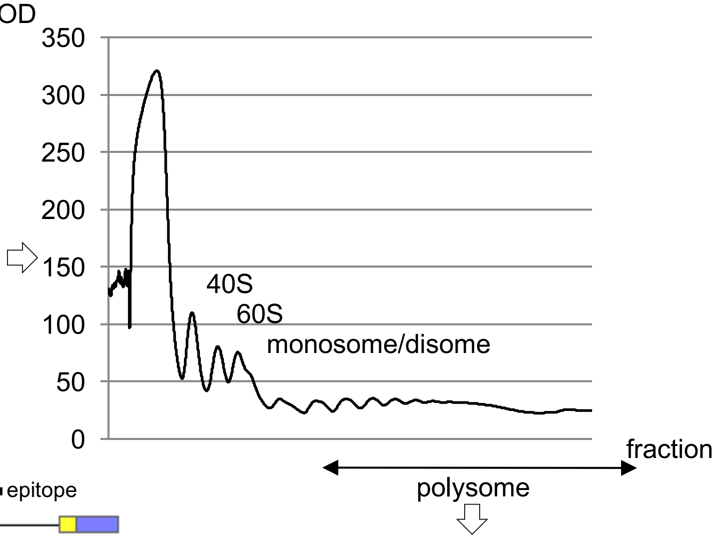
RT-PCR

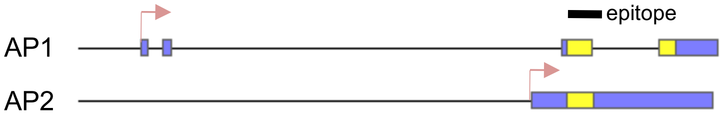


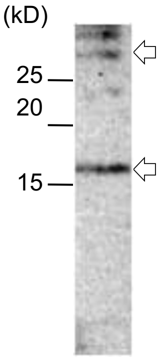
Western blotting







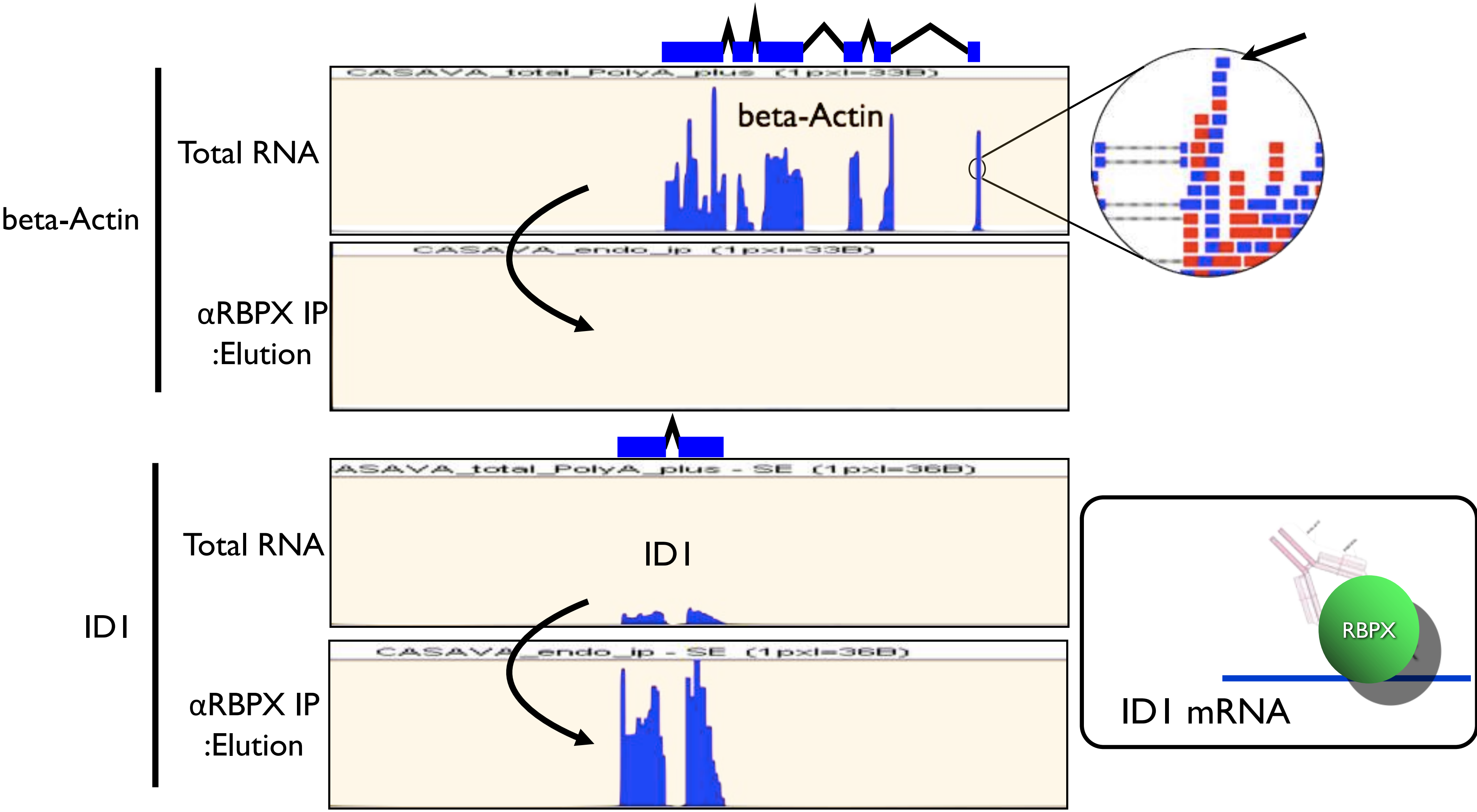




NM_018952 (HoxB6)	AP1	AP2
TSS tag	1170ppm	25ppm
pol II binding	bound	bound
Corresponding cDNA	BC014651	X58431 (HIT00195371)*1
Longest ORF in the cDNA	224 aa	140 aa
Expected molecular weight of protein product	25.4 kD	15.2 kD
Translation caveat	no	no
polysome tag enrich (p value)	1e−7	2e−3
RNA Seq tag (polysome)	9ppm	5ppm

=>Identification of TSS having protein consequences

Identification of RNA binding protein target mRNAs



beta-Actin

Total RNA

beta-Actin

αRBPX IP
:Elution

ID1

Total RNA

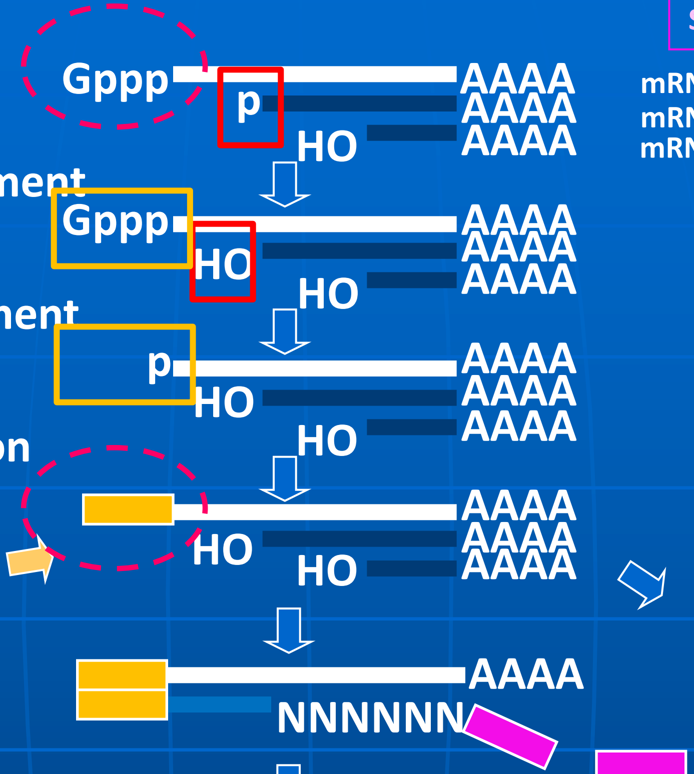
ID1

αRBPX IP
:Elution

ID1 mRNA

“oligo-capping” + “Illumina GA” =TSS Seq: massive generation of TSS-tags

Starting from 50 microG total RNA



mRNA with CAP
mRNA without CAP
mRNA without CAP

“Oligo-Capping”
- BAP treatment
- TAP treatment
- RNA ligation

5’-Oligo＝Solexa primer A

First Strand Synthesis

1st strand primer＝Solexa primer B

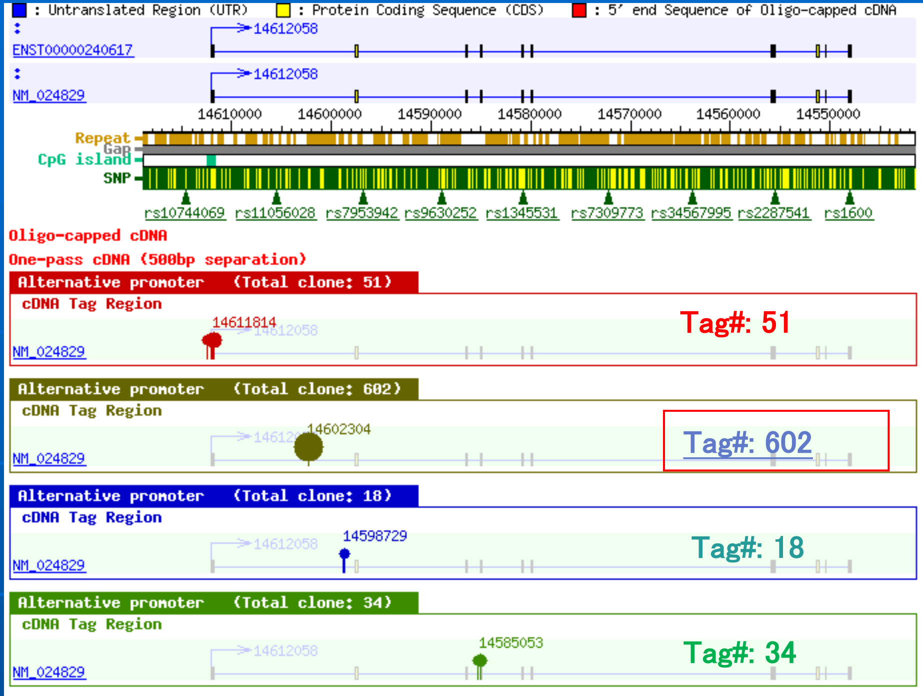
15 cycle PCR->150- 250bp size fractionation

Used as a Template for Next Gen Sequencing (NGS)

36 bp data sufficiently identifies
TSS positions in the human genome

Advantages in Solexa: Detecting Promoter-specific induction

Stimulation (+)



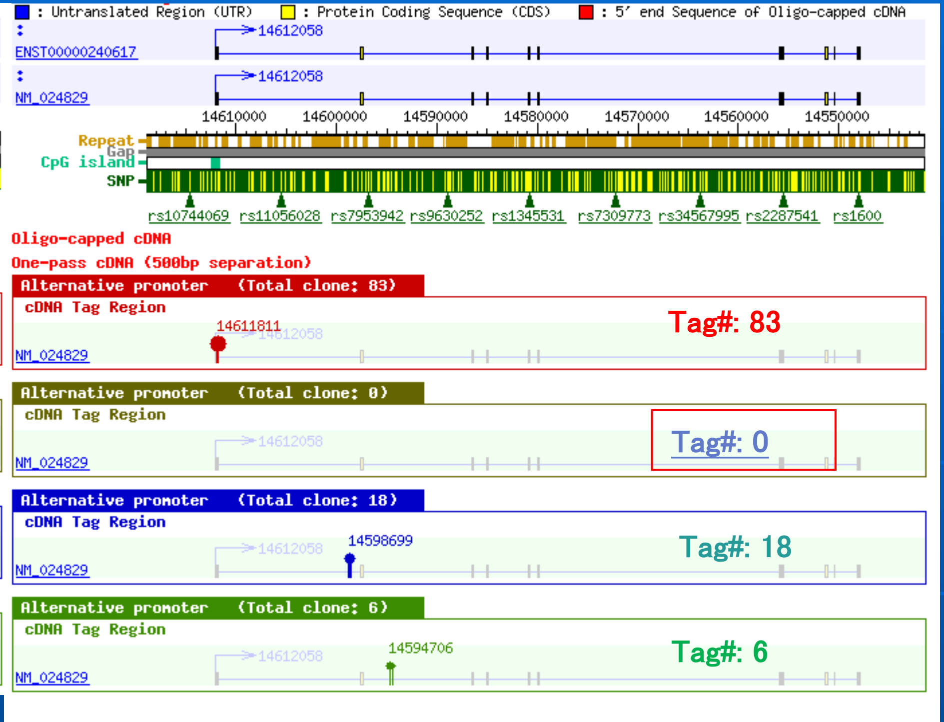
Tag#: 51

Tag#: 602

Tag#: 18

Tag#: 34

Stimulation (−)



Tag#: 83

Tag#: 0

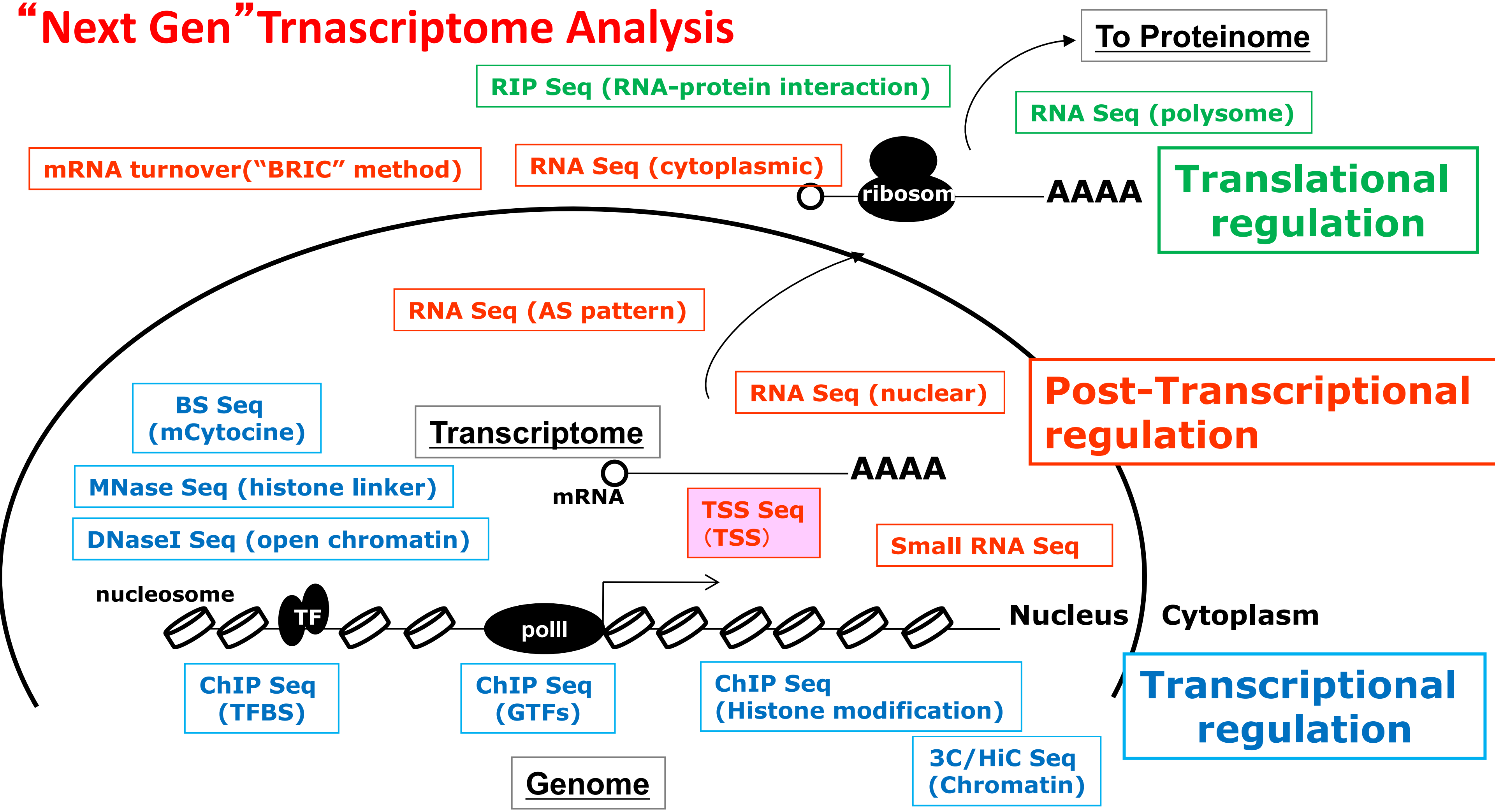
Tag#: 18

Tag#: 6

Second AP is selectively induced

Could be masked by other APs
in microarray or RT-PCR analyses

“Next Gen”Trnascriptome Analysis



To Proteinome

RIP Seq (RNA-protein interaction)

RNA Seq (polysome)

mRNA turnover(“BRIC” method)

RNA Seq (cytoplasmic)

ribosom

AAAA

Translational regulation

RNA Seq (AS pattern)

RNA Seq (nuclear)

Post-Transcriptional regulation

BS Seq (mCytocine)

Transcriptome

MNase Seq (histone linker)

mRNA

AAAA

DNaseI Seq (open chromatin)

TSS Seq (TSS)

Small RNA Seq

nucleosome

TF

polII

Nucleus

Cytoplasm

ChIP Seq (TFBS)

ChIP Seq (GTFs)

ChIP Seq (Histone modification)

Transcriptional regulation

3C/HiC Seq (Chromatin)

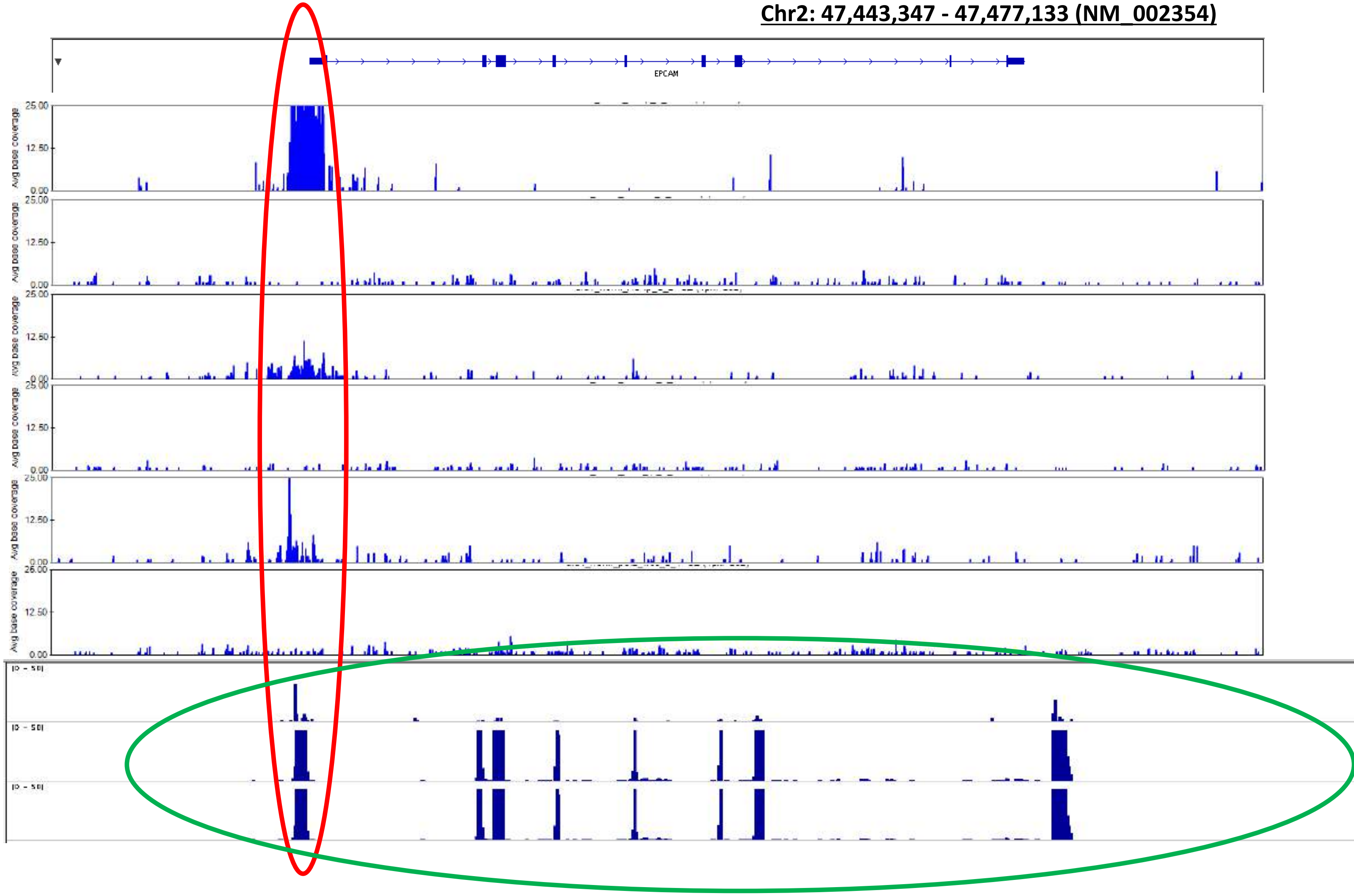
Genome

Next Gen Sequencer as a common platform

Data Integration

DLD-1 cell (colon caner)

Chr2: 47,443,347 - 47,477,133 (NM_002354)



Annotated mRNA

DLD-1_H3K4me3 (IP)

DLD-1_H3K4me3 (background)

DLD-1_H3Ac (IP)

DLD-1_H3Ac (background)

DLD-1_pol II (IP)

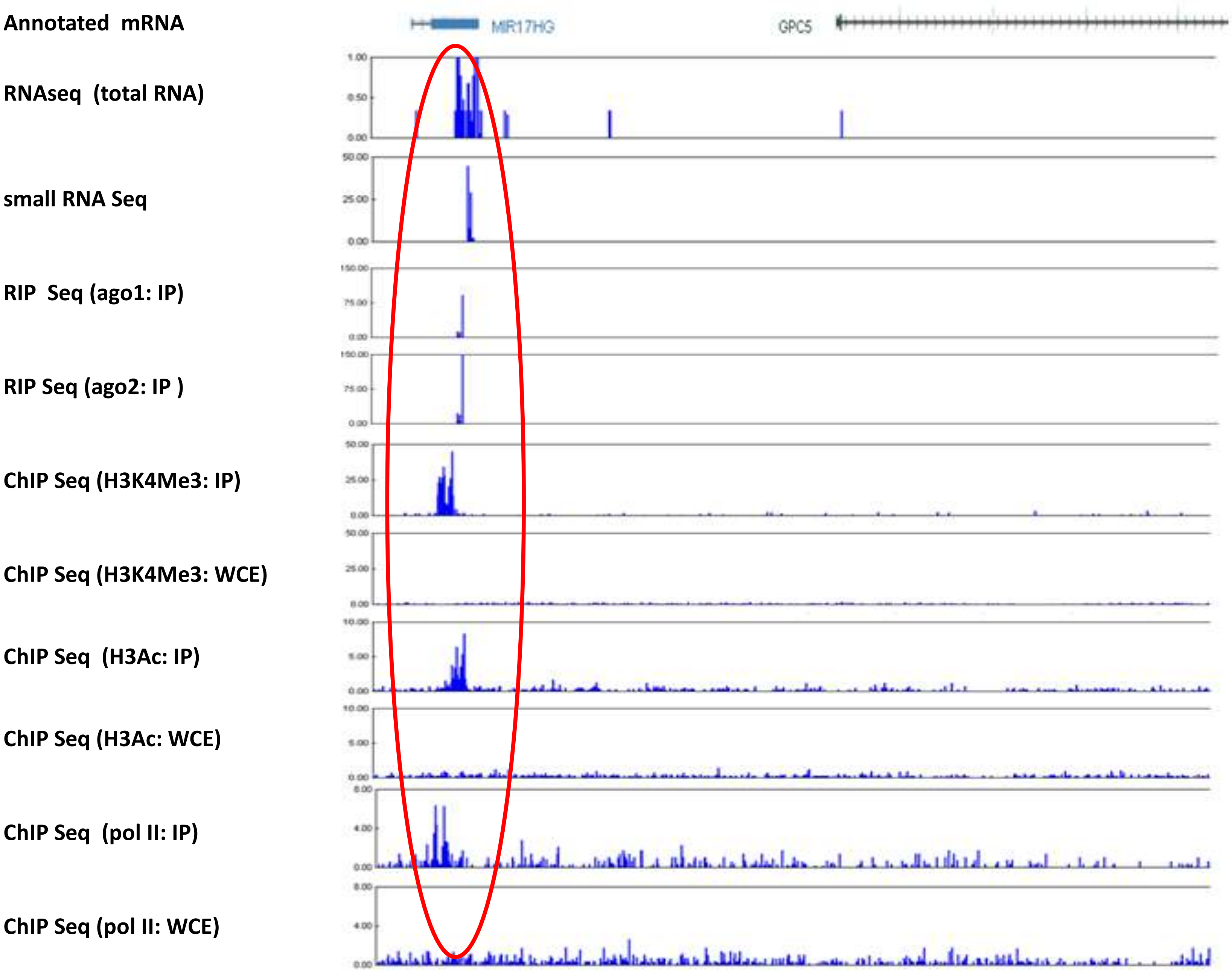
DLD-1_pol II (background)

DLD-1_TSSseq

DLD-1_RNAseq

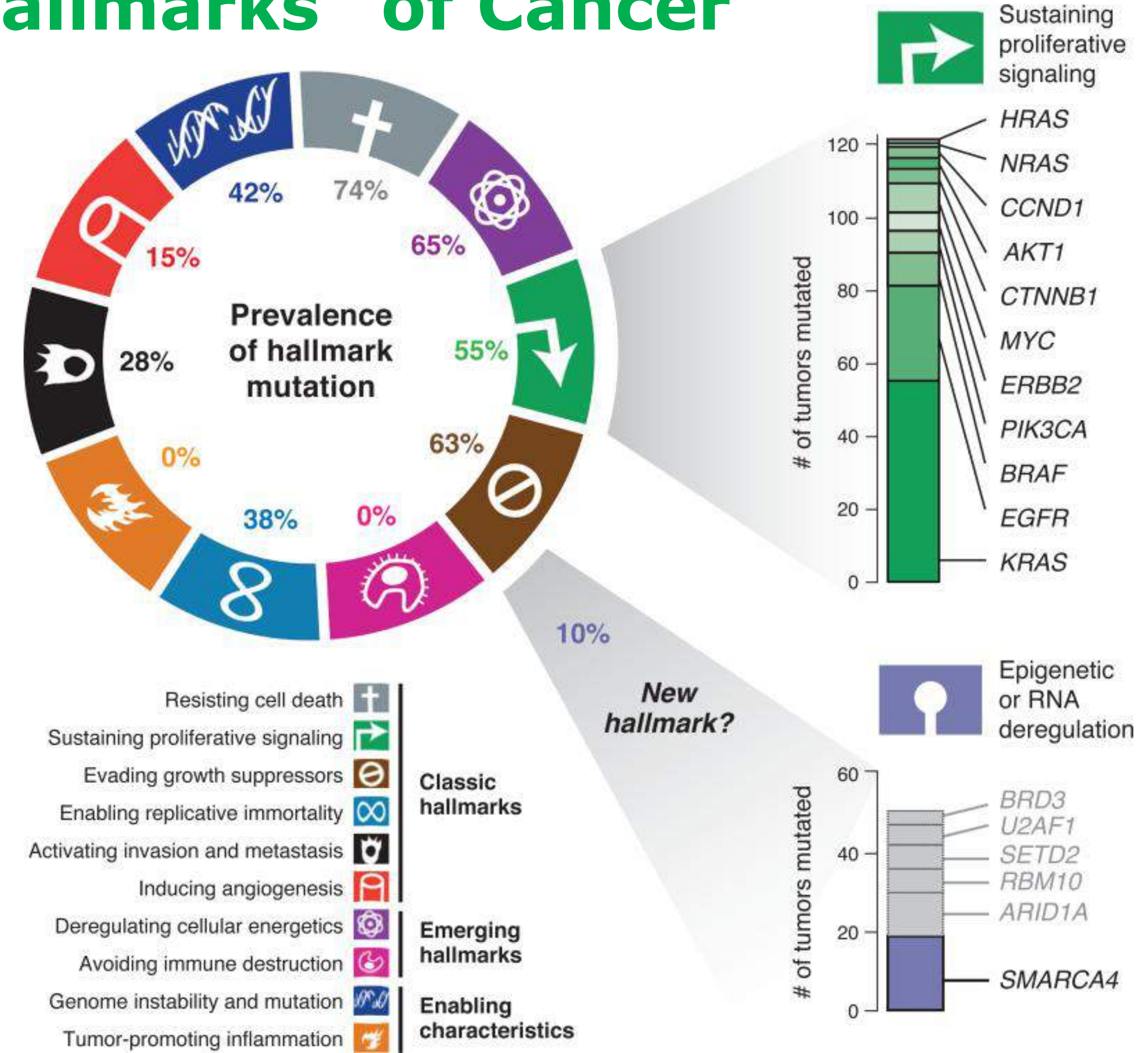
DLD-1_Polysome

B The MIR17HG_gene region (DLD-1 cells)



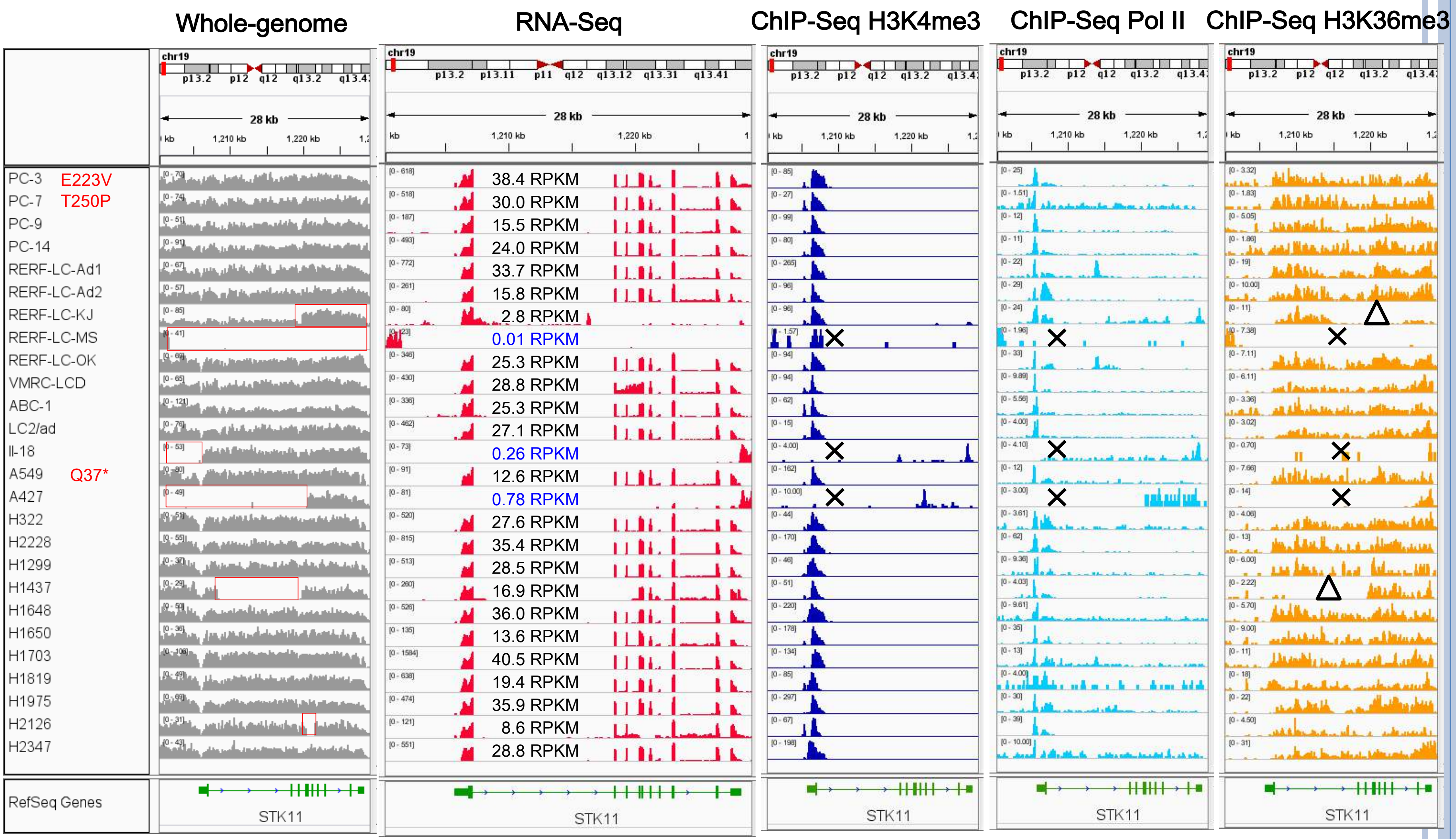
“Hallmarks” of Cancer

Imielinski et al Cell (2012)



資料２－２

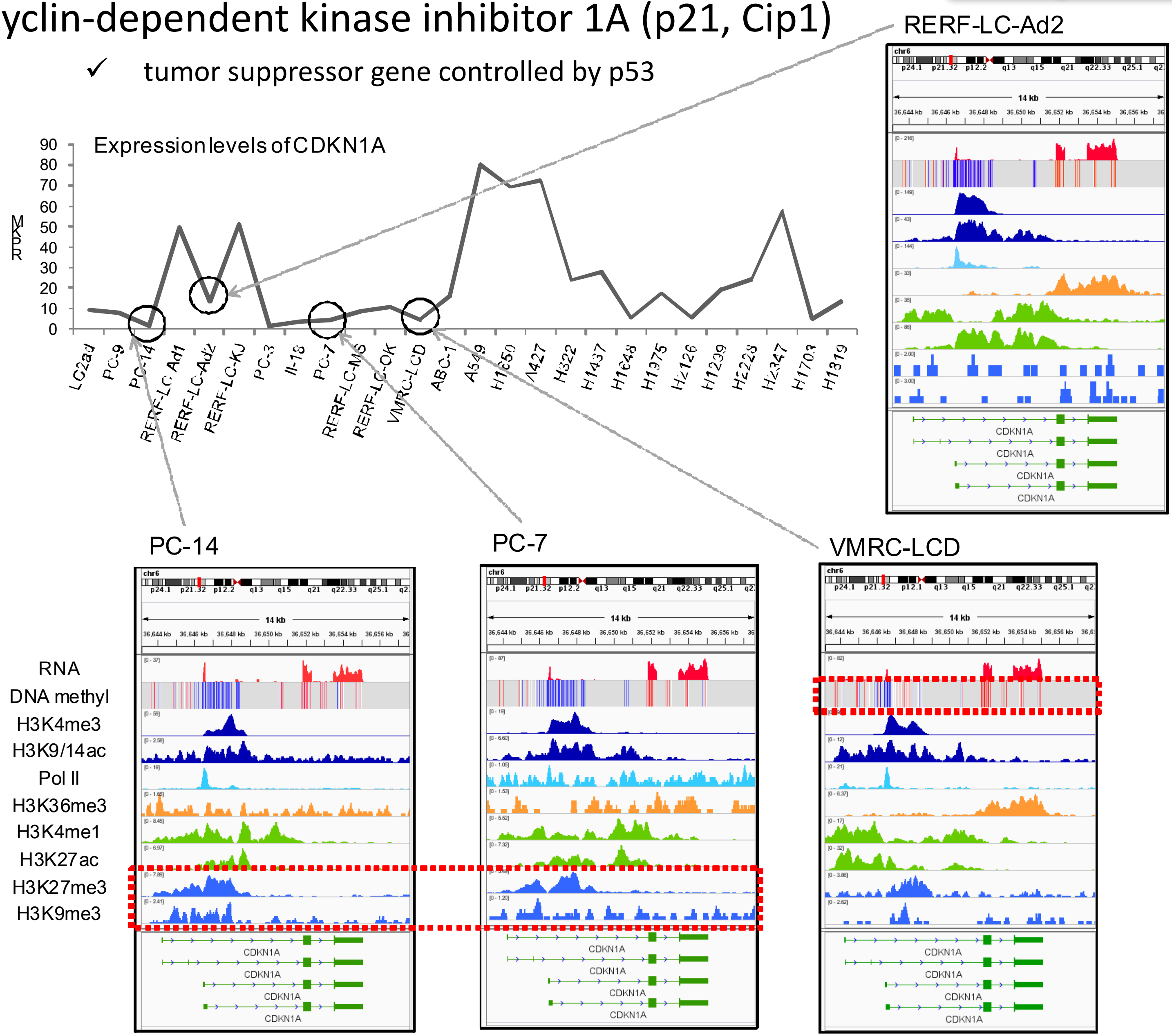
STK11遺伝子についての遺伝子発現異常パターン



ゲノム異常　　遺伝子発現異常　　エピゲノム異常

CDKN1A cyclin-dependent kinase inhibitor 1A (p21, Cip1)

- ✓ tumor suppressor gene controlled by p53



No Genomic Loss; but Epigenomic Loss at the Histone level

Integrated analyses

CDKN2A cyclin-dependent kinase inhibitor 2A

G67V ($\mathrm{p16^{INK4a}}$)

62-base deletion ($\mathrm{p16^{INK4a}/p14^{ARF}}$)

D84V ($\mathrm{p16^{INK4a}}$)

E69* ($\mathrm{p16^{INK4a}}$)

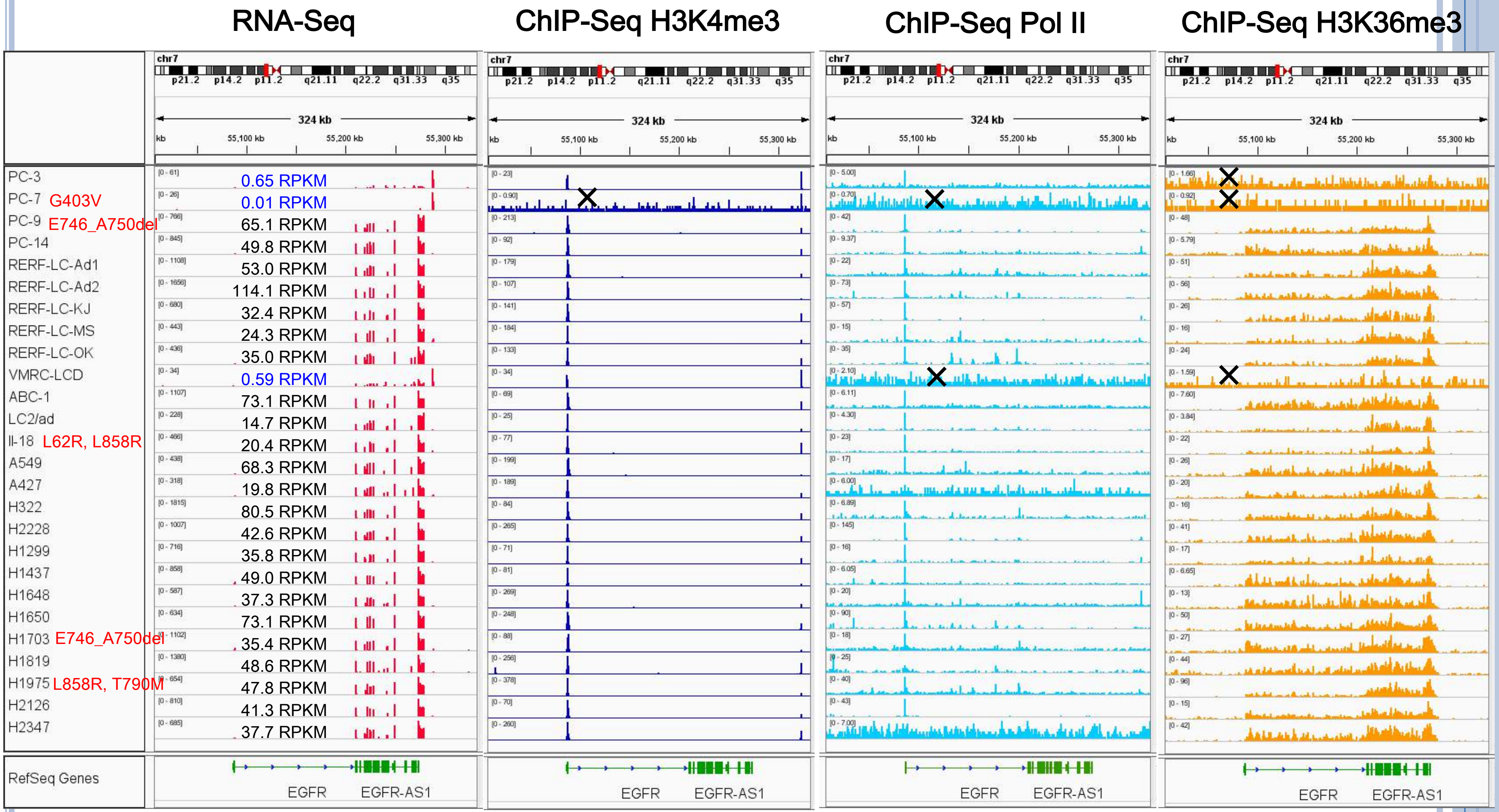
$\mathrm{p16^{INK4a}}$の異常

Genomic deletion:
13 cell lines
SNVs/indels:
4 cell lines
DNA methylation:
6 cell lines

ゲノム変異と
DNAメチル化が
発現量に大きく
寄与している

EGFR epidermal growth factor receptor

PC-7: Non-adherent cell



Cell line	H3K4me3	Pol II	H3K36me3
PC-7	×	×	×
VMRC-LCD	○	×	×
PC-3	○	△	×

Small-molecule inhibitors to chromatin-associated factors

Table 1: Small molecule inhibitors to chromatin-associated proteins

Chromatin-binding protein	Compound
Histone methyltransferases	
DOT1L	EPZ004777 (ref. 21), EPZ-5676 (ref. 24), SGC0946 (ref. 86)
EZH2	GSK126 (ref. 37), GSK343 (refs 87,88), EPZ005687 (ref. 38), EPZ-6438 (ref. 44), EI1 (ref. 39), UNC1999 (ref. 89)
G9A	BIX01294 (ref. 90), UNC0321 (ref. 91), UNC0638 (ref. 92), NC0642 (ref. 88), BRD4770 (ref. 93)
PRMT3	14u (ref. 94)
PRMT4 (CARM1)	17b (Bristol-Myers Squibb) (refs 95,96), MethylGene (ref. 97)
Histone demethylases	
LSD1	Tranylcypromine (ref. 62), ORY-1001 (ref. 63)
Bromodomains	
BET	JQ1 (ref. 73), IBET762 (ref. 72), IBET151 (refs 76,98), PFI-1 (ref. 99)
BAZ2B	GSK2801 (ref. 88)
Chromodomains	
L3MBTL1	UNC669 (ref. 100)
L3MBTL3	UNC1215 (ref. 101)

Helin & Dhanak. 2013 *Nature*
Chromatin proteins and modifications as drug targets

JQ1:
a small-molecule bromodomain inhibitor

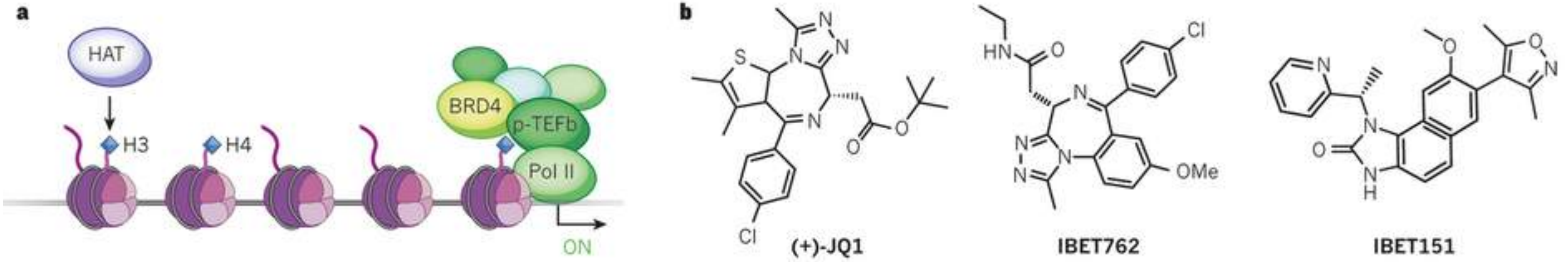
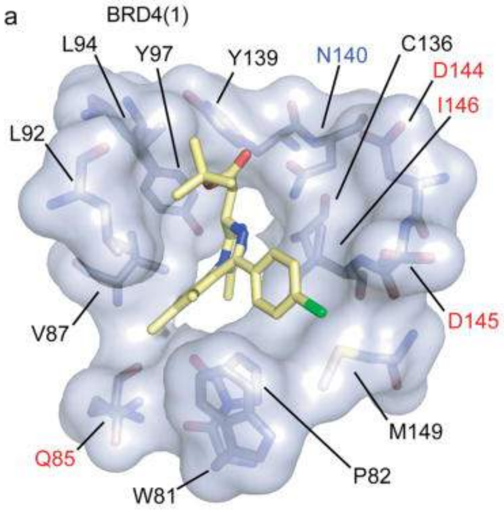


Fig. 4 Bromodomain proteins and their inhibitors.

Helin & Dhanak. 2013 *Nature*
Chromatin proteins and modifications as drug targets



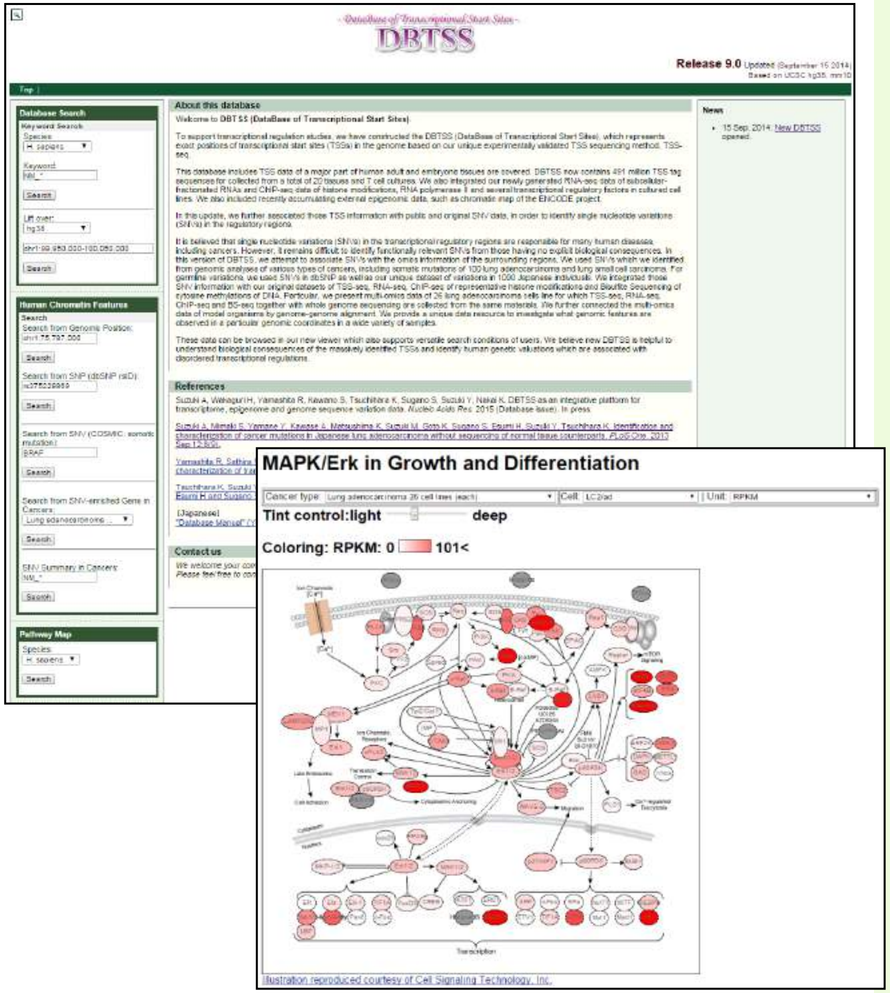
Filippakopoulos et al. 2010 *Nature*
Selective inhibition of BET bromodomains

Fig. 3a The acetyl-lysine binding pocket of BRD4(1) is shown as a semi-transparent surface with contact residues labelled and depicted in stick representation. Carbon atoms in (+)-JQ1 are coloured yellow to distinguish them from protein residues. Distinguishing surface residues are shown in red; the family conserved asparagine is shown in blue.

All data will be publicly open from our database

DBTSS

http://dbtss.hgc.jp/



DataBase of Transcriptional Start Sites

DBTSS

Release 9.0 Updated (September 15 2014)

About this database

Welcome to **DBTSS (DataBase of Transcriptional Start Sites)**

To support transcriptional regulation studies, we have constructed the DBTSS (DataBase of Transcriptional Start Sites), which represents exact positions of transcriptional start sites (TSSs) in the genome based on our unique experimentally validated TSS sequencing method, TSS-seq.

This database includes TSS data of a major part of human adult and embryonic tissues are covered. DBTSS now contains 491 million TSS tag sequences for collected from a total of 20 tissues and T cell cultures. We also integrated our newly generated RNA-seq data of subcellular-fractionated RNAs and ChIP-seq data of histone modifications, RNA polymerase II and several transcriptional regulatory factors in cultured cell lines. We also included recently accumulating external epigenomic data, such as chromatin map of the ENCODE project.

In this update, we further associated those TSS information with public and original SNV data, in order to identify single nucleotide variations (SNVs) in the regulatory regions.

It is believed that single nucleotide variations (SNVs) in the transcriptional regulatory regions are responsible for many human diseases including cancers. However, it remains difficult to identify functionally relevant SNVs from those having no explicit biological consequences. In this version of DBTSS, we attempt to associate SNVs with the omics information of the surrounding regions. We used SNVs which we identified from genomic analyses of various types of cancers, including somatic mutations of 100 lung adenocarcinoma and lung small cell carcinoma. For germline variations, we used SNVs in dbSNP as well as our unique dataset of variations in 1000 Japanese individuals. We integrated those SNV information with our original datasets of TSS-seq, RNA-seq, ChIP-seq of representative histone modifications and Bisulfite Sequencing of cytosine methylations of DNA. Particular, we present multi-omics data of 26 lung adenocarcinoma cells line for which TSS-seq, RNA-seq, ChIP-seq and BS-seq together with whole genome sequencing are collected from the same materials. We further connected the multi-omics data of model organisms by genome-genome alignment. We provide a unique data resource to investigate what genomic features are observed in a particular genomic coordinates in a wide variety of samples.

These data can be browsed in our new viewer which also supports versatile search conditions of users. We believe new DBTSS is helpful to understand biological consequences of the massively identified TSSs and identify human genetic valuations which are associated with disordered transcriptional regulations.

References

Suzuki A, Wakaguri H, Yamashita R, Kawano S, Tsuchihara K, Sugano S, Suzuki Y, Nakai K. DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. *Nucleic Acids Res.* 2015 (Database issue). In press.

Suzuki A, Mimaki S, Yamane Y, Kawase A, Matsushima K, Suzuki M, Goto K, Sugano S, Esumi H, Suzuki Y, Tsuchihara K. Identification and characterization of cancer mutations in Japanese lung adenocarcinoma without sequencing of normal tissue counterparts. *PLoS One*. 2013

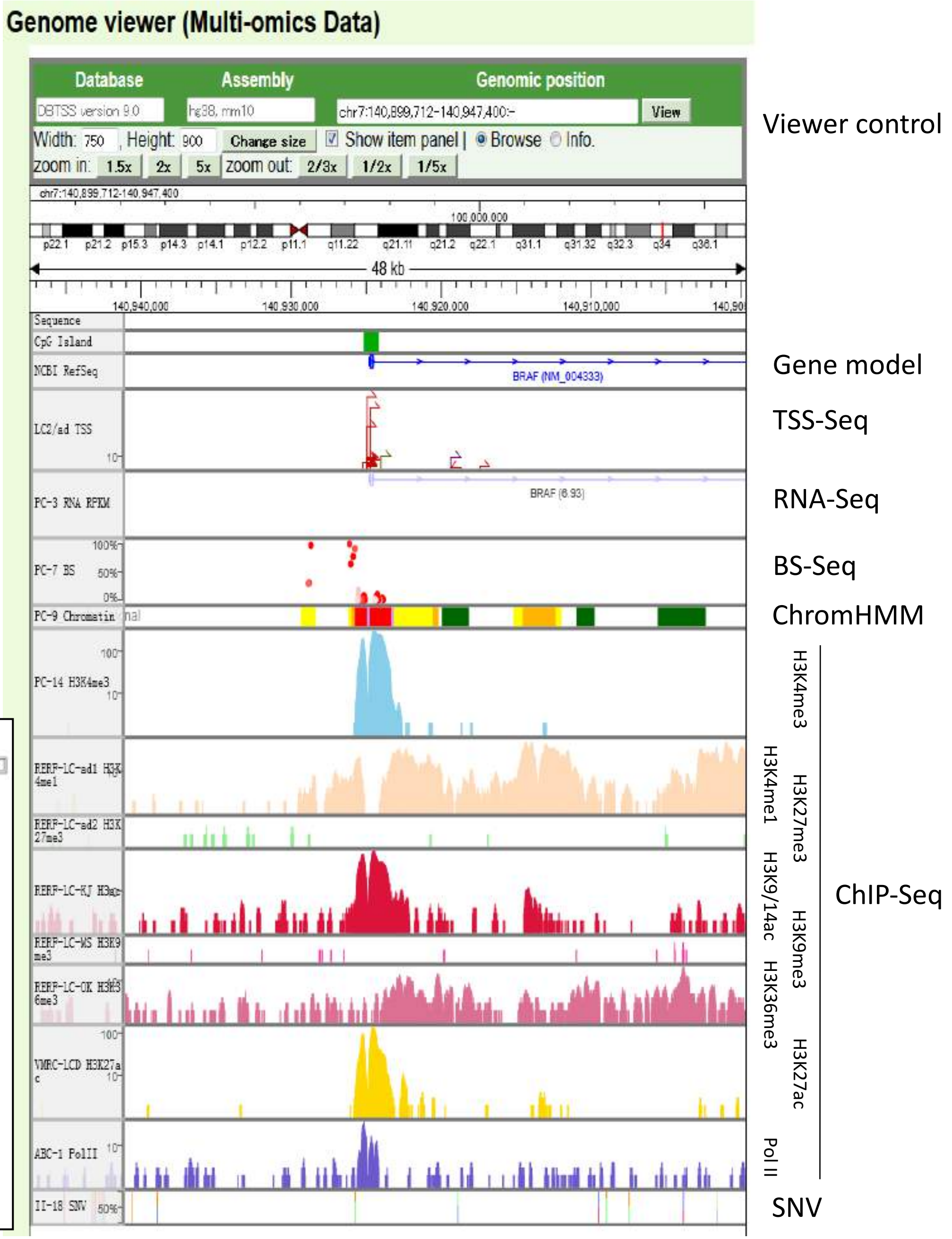
MAPK/Erk in Growth and Differentiation

Cancer type: Lung adenocarcinoma 26 cell lines (each) | Cell: LC2/ad | Unit: RPKM

Tint control: light — deep

Coloring: RPKM: 0 — 101<

Illustration reproduced courtesy of Cell Signaling Technology, Inc.



Genome viewer (Multi-omics Data)

Database: DBTSS version 9.0

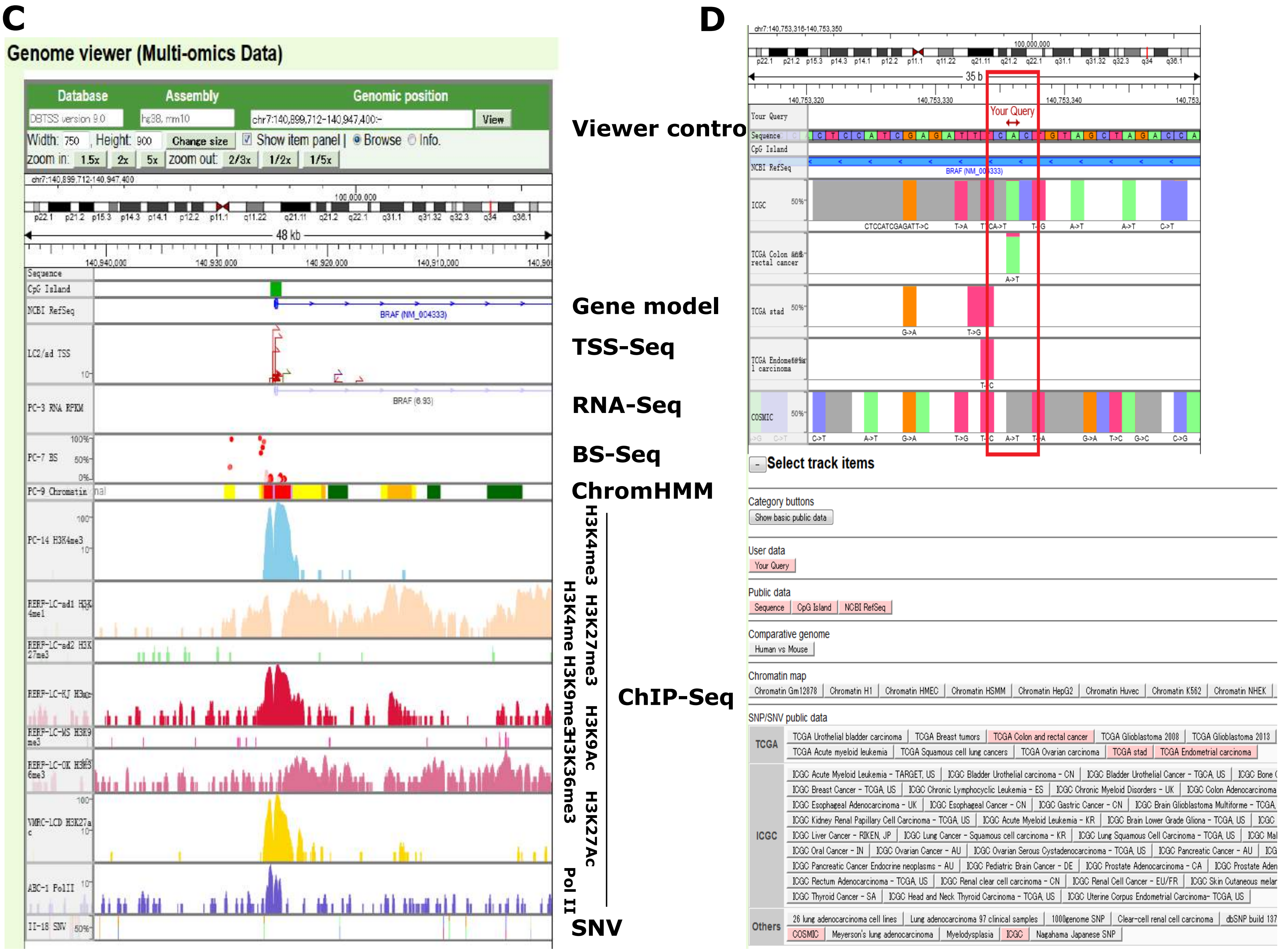
Assembly: hg38, mm10

Genomic position: chr7:140,899,712-140,947,400:-

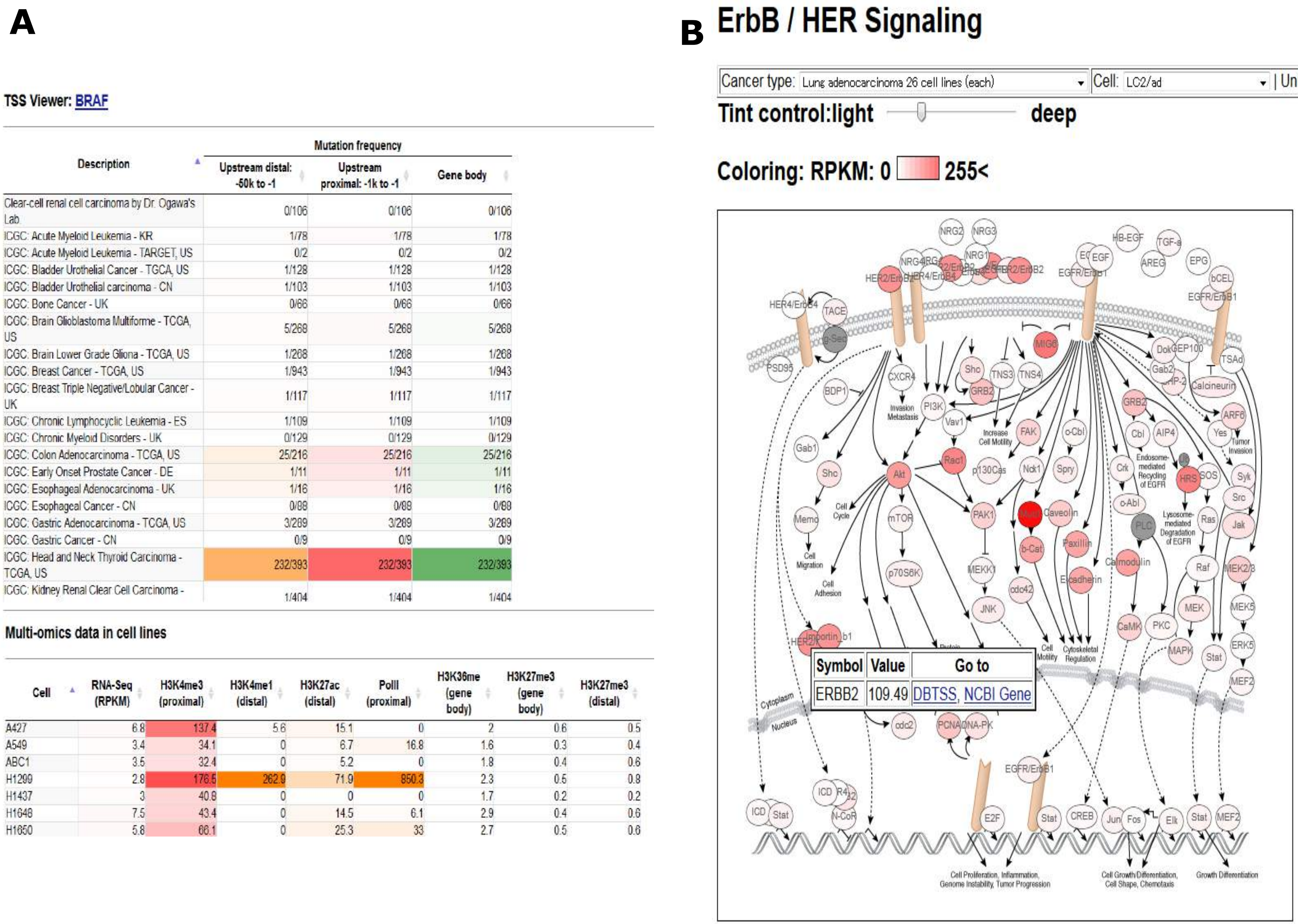
Width: 750, Height: 900

☑ Show item panel | ● Browse ○ Info.

(Suzuki et al. 2015 *Nucleic Acids Research*)


C
Genome viewer (Multi-omics Data)
Database
Assembly
Genomic position
DBTSS version 9.0
hg38, mm10
chr7:140,899,712-140,947,400:-
View
Width: 750
Height: 900
Change size
☑ Show item panel
● Browse
○ Info.
zoom in: 1.5x 2x 5x
zoom out: 2/3x 1/2x 1/5x
48 kb
Sequence
CpG Island
NCBI RefSeq
BRAF (NM_004333)
LC2/ad TSS
PC-3 RNA RPKM
BRAF (6.93)
PC-7 BS
PC-9 Chromatin
PC-14 H3K4me3
RERF-LC-ad1 H3K4me1
RERF-LC-ad2 H3K27me3
RERF-LC-KJ H3ac
RERF-LC-MS H3K9me3
RERF-LC-OK H3K36me3
VMRC-LCD H3K27ac
ABC-1 PolII
II-18 SNV
Viewer contro
Gene model
TSS-Seq
RNA-Seq
BS-Seq
ChromHMM
H3K4me3
H3K27me3
H3K9Ac
H3K27Ac
H3K4me
H3K9me3
H3K36me3
Pol II
ChIP-Seq
SNV
D
chr7:140,753,316-140,753,350
35 b
Your Query
Sequence
CpG Island
NCBI RefSeq
BRAF (NM_004333)
ICGC
TCGA Colon and rectal cancer
TCGA stad
TCGA Endometrial carcinoma
COSMIC
Select track items
Category buttons
Show basic public data
User data
Your Query
Public data
Sequence
CpG Island
NCBI RefSeq
Comparative genome
Human vs Mouse
Chromatin map
Chromatin Gm12878
Chromatin H1
Chromatin HMEC
Chromatin HSMM
Chromatin HepG2
Chromatin Huvec
Chromatin K562
Chromatin NHEK
SNP/SNV public data
TCGA
TCGA Urothelial bladder carcinoma
TCGA Breast tumors
TCGA Colon and rectal cancer
TCGA Glioblastoma 2008
TCGA Glioblastoma 2013
TCGA Acute myeloid leukemia
TCGA Squamous cell lung cancers
TCGA Ovarian carcinoma
TCGA stad
TCGA Endometrial carcinoma
ICGC
Others
26 lung adenocarcinoma cell lines
Lung adenocarcinoma 97 clinical samples
1000genome SNP
Clear-cell renal cell carcinoma
dbSNP build 137
COSMIC
Meyerson's lung adenocarcinoma
Myelodysplasia
ICGC
Nagahama Japanese SNP

Figu



TSS Viewer: BRAF

Description	Mutation frequency		
	Upstream distal: -50k to -1	Upstream proximal: -1k to -1	Gene body
Clear-cell renal cell carcinoma by Dr. Ogawa's Lab	0/106	0/106	0/106
ICGC: Acute Myeloid Leukemia - KR	1/78	1/78	1/78
ICGC: Acute Myeloid Leukemia - TARGET, US	0/2	0/2	0/2
ICGC: Bladder Urothelial Cancer - TGCA, US	1/128	1/128	1/128
ICGC: Bladder Urothelial carcinoma - CN	1/103	1/103	1/103
ICGC: Bone Cancer - UK	0/66	0/66	0/66
ICGC: Brain Glioblastoma Multiforme - TCGA, US	5/268	5/268	5/268
ICGC: Brain Lower Grade Gliona - TCGA, US	1/268	1/268	1/268
ICGC: Breast Cancer - TCGA, US	1/943	1/943	1/943
ICGC: Breast Triple Negative/Lobular Cancer - UK	1/117	1/117	1/117
ICGC: Chronic Lymphocyclic Leukemia - ES	1/109	1/109	1/109
ICGC: Chronic Myeloid Disorders - UK	0/129	0/129	0/129
ICGC: Colon Adenocarcinoma - TCGA, US	25/216	25/216	25/216
ICGC: Early Onset Prostate Cancer - DE	1/11	1/11	1/11
ICGC: Esophageal Adenocarcinoma - UK	1/16	1/16	1/16
ICGC: Esophageal Cancer - CN	0/88	0/88	0/88
ICGC: Gastric Adenocarcinoma - TCGA, US	3/289	3/289	3/289
ICGC: Gastric Cancer - CN	0/9	0/9	0/9
ICGC: Head and Neck Thyroid Carcinoma - TCGA, US	232/393	232/393	232/393
ICGC: Kidney Renal Clear Cell Carcinoma -	1/404	1/404	1/404

Multi-omics data in cell lines

Cell	RNA-Seq (RPKM)	H3K4me3 (proximal)	H3K4me1 (distal)	H3K27ac (distal)	PolII (proximal)	H3K36me (gene body)	H3K27me3 (gene body)	H3K27me3 (distal)
A427	6.8	137.4	5.6	15.1	0	2	0.6	0.5
A549	3.4	34.1	0	6.7	16.8	1.6	0.3	0.4
ABC1	3.5	32.4	0	5.2	0	1.8	0.4	0.6
H1299	2.8	176.5	262.9	71.9	850.3	2.3	0.5	0.8
H1437	3	40.8	0	0	0	1.7	0.2	0.2
H1648	7.5	43.4	0	14.5	6.1	2.9	0.4	0.6
H1650	5.8	66.1	0	25.3	33	2.7	0.5	0.6

ErbB / HER Signaling

Cancer type: Lung adenocarcinoma 26 cell lines (each) | Cell: LC2/ad | Unit: RF

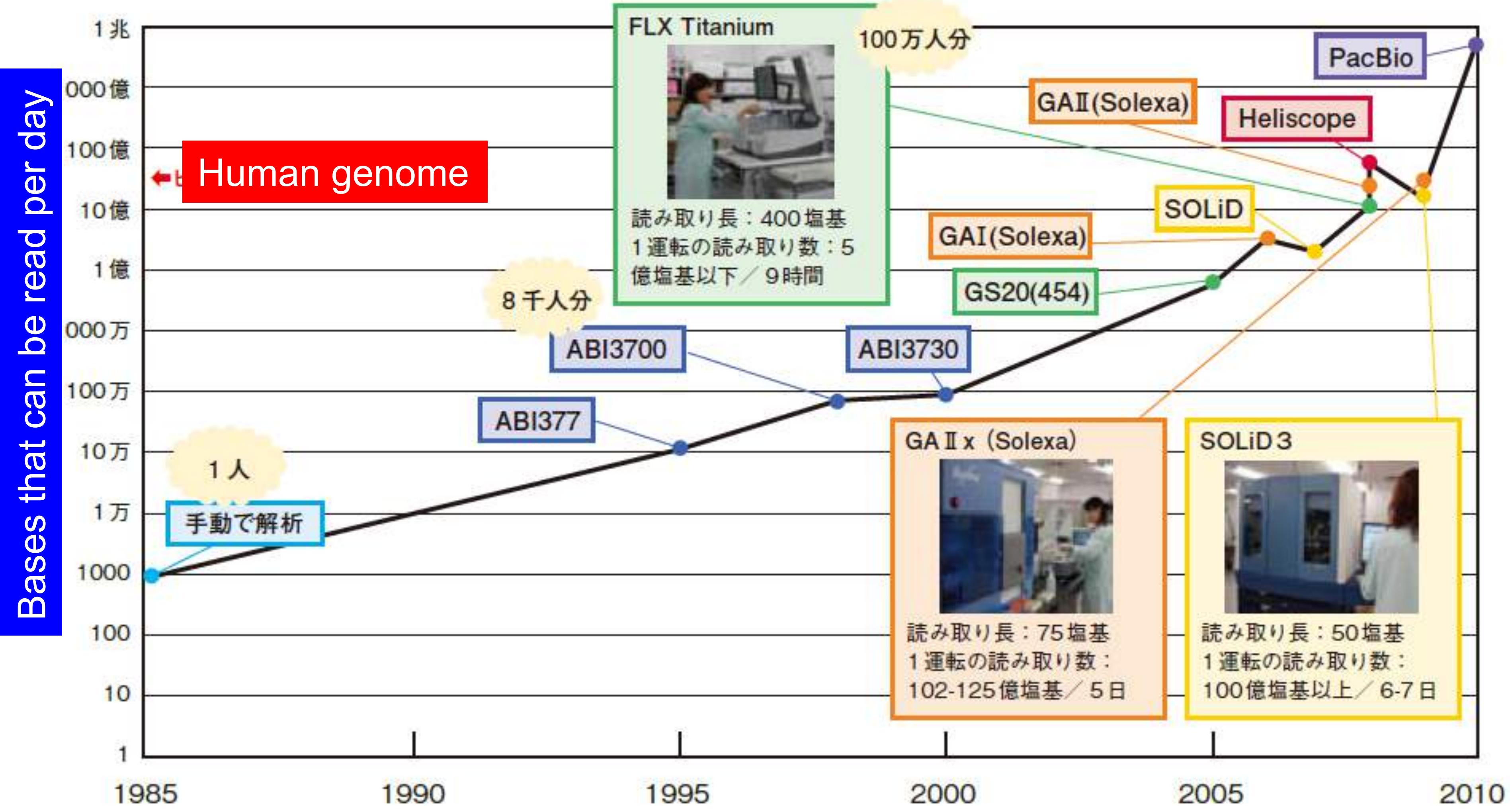
Tint control:light deep

Coloring: RPKM: 0 🟥 255<

Symbol	Value	Go to
ERBB2	109.49	DBTSS, NCBI Gene

Figur

Power of NGS; their sequence capacity



Source: Cell Innovation program pamphlet

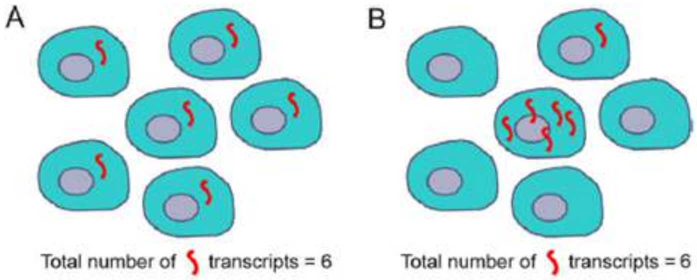
Single cell Sequencing

Yutaka Suzuki
partment of Computational Biology and Medical Scienc
Graduate School of Frontier Sciences
The University of Tokyo

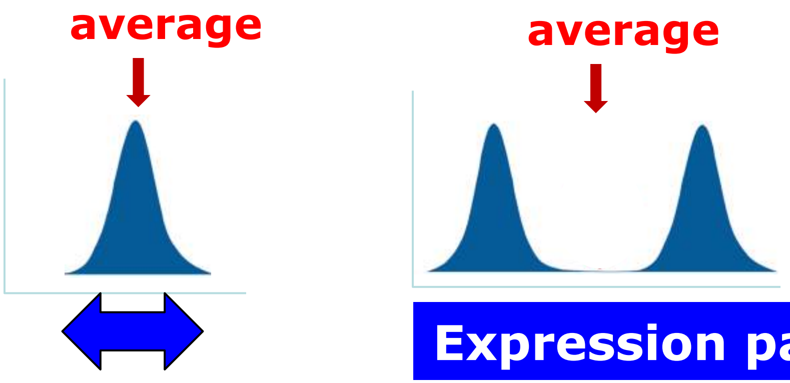
東京大学
THE UNIVERSITY OF TOKYO

Why single cell?
(Stahlberg et al., Methods 50 (2010) 282-288

Homogeneous population **Heterogeneous population**



Bulk cells or single cells?

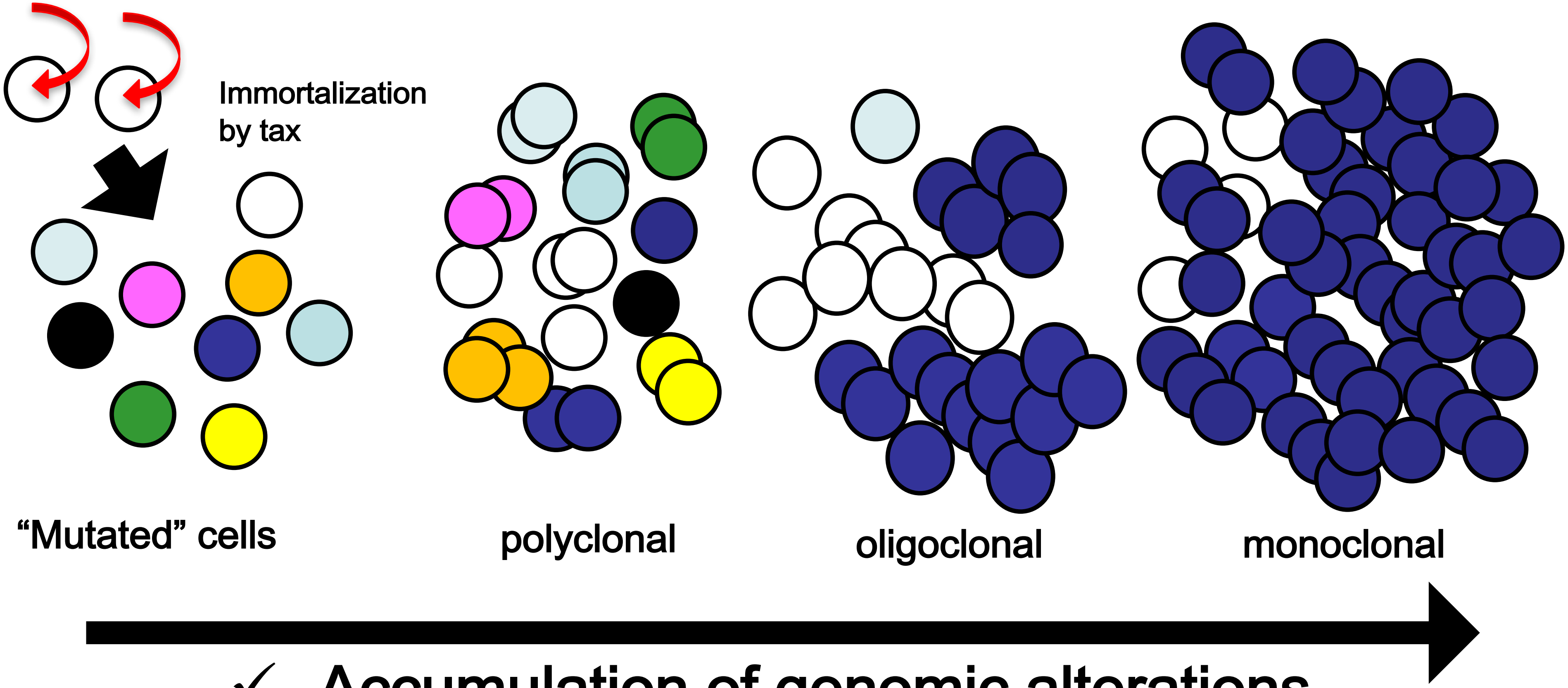


Expression patterns of sub-po

Difergence in gene expr

ATL: Adult T-cell leukemia/lymphoma

HTLV-1: Human T-cell leukemia virus
Approx. 50 years after infection (via maternal milk)

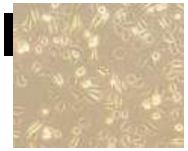


- ✓ Accumulation of genomic alterations
- ✓ Clonal and genomic heterogeneity

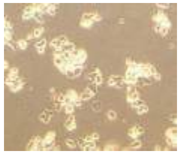
Single-cell RNA-Seq of LAD cell l using C1 system

Single-cell RNA-Seq in C1 system (Fluidigm):

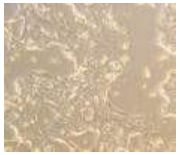
- ✓ **Automatic cell isolation and cDNA amplification**
- ✓ **96 individual cells**
- ✓ **Cell size: 5-25 µm**



PC9



VMRC-LCD

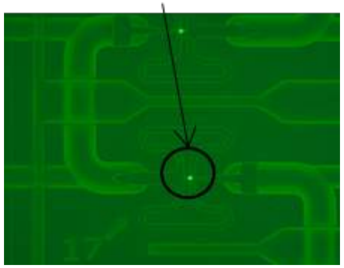


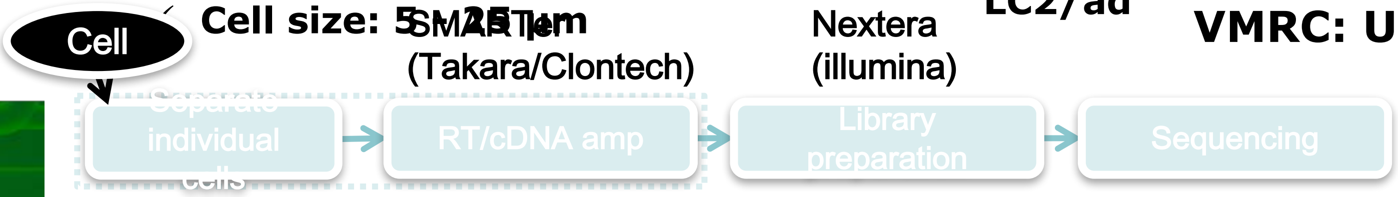
LC2/ad

Drivers

LC2ad: RET
PC9: EGFR
VMRC: Unkno

Captured single cell

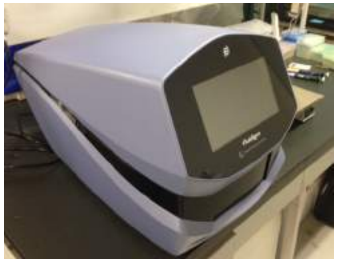




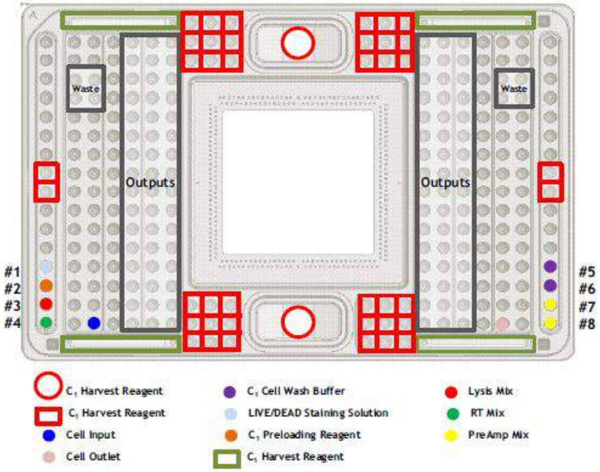
SMARTer (Takara/Clontech)

Nextera (illumina)

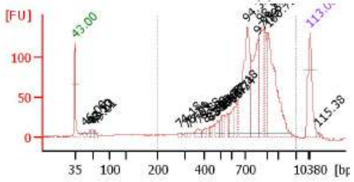
C1 Single-Cell Auto Prep system (Fluidigm)



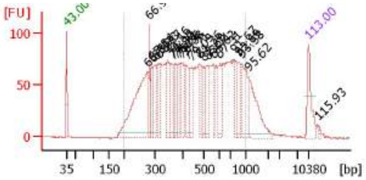
HiSeq2500 (illumina)



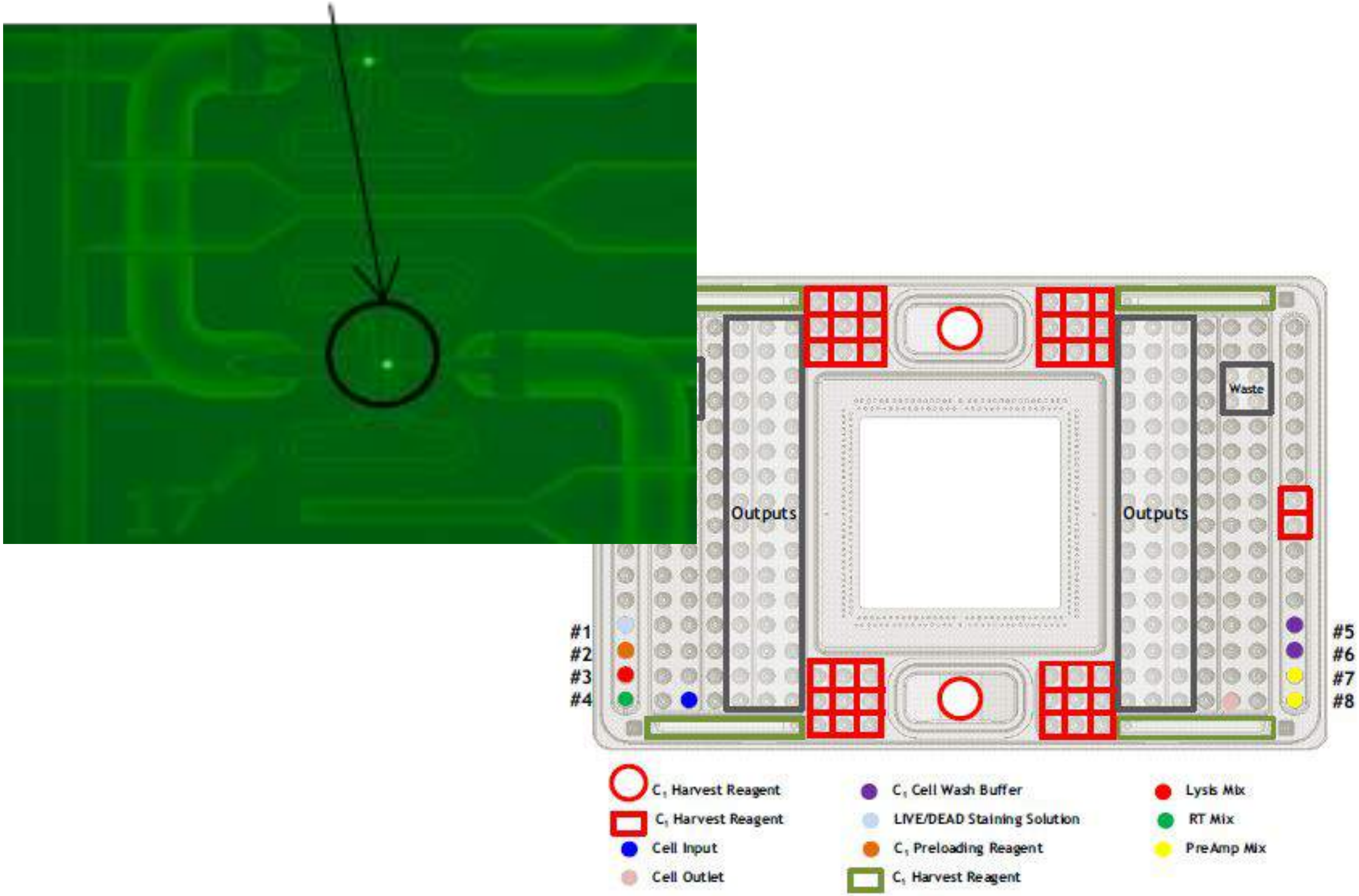
Amplified cDNA



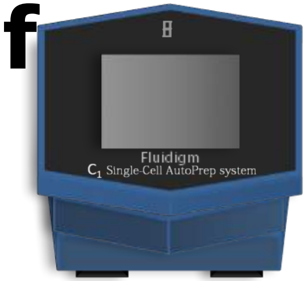
Sequencing library (12-plex)

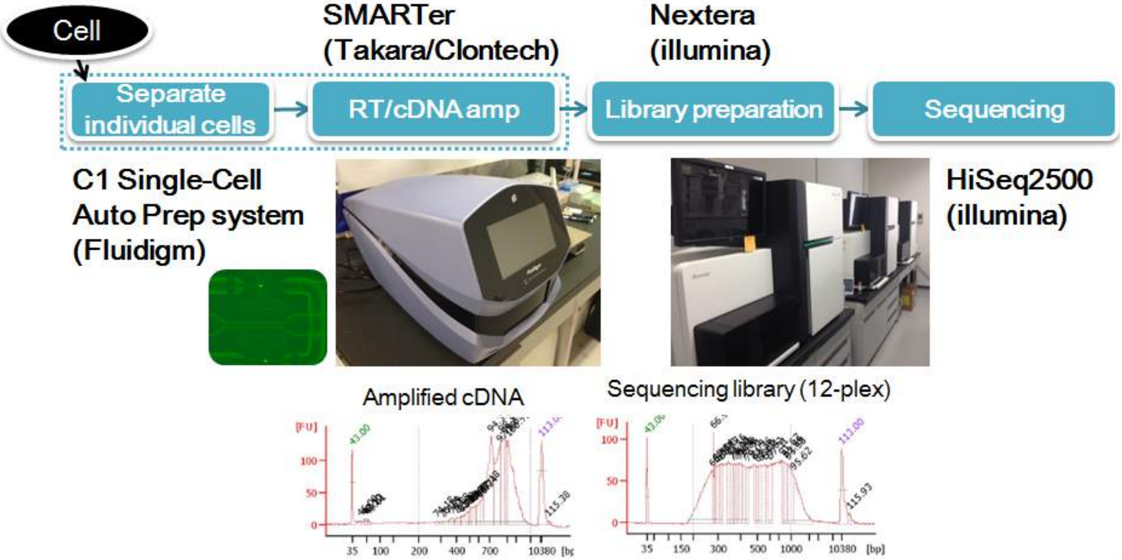


Captured single cell



Understanding heterogeneity of transcriptomic/epigenomic perturbations by single-cell





Cell

SMARTer (Takara/Clontech)

Nextera (illumina)

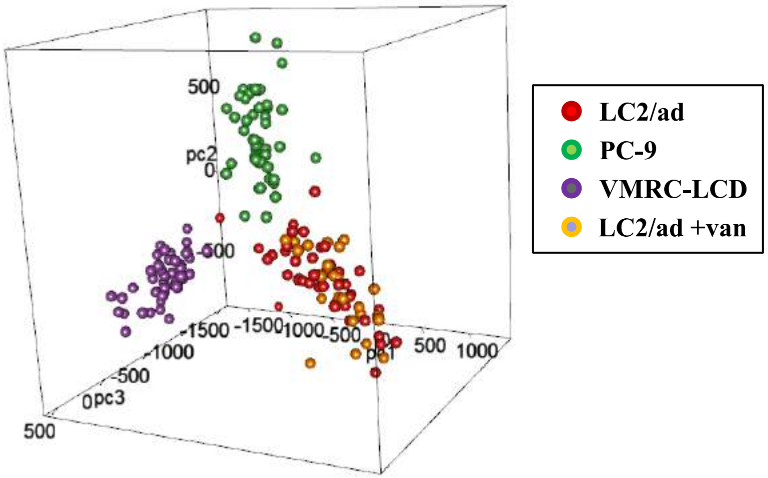
Separate individual cells → RT/cDNA amp → Library preparation → Sequencing

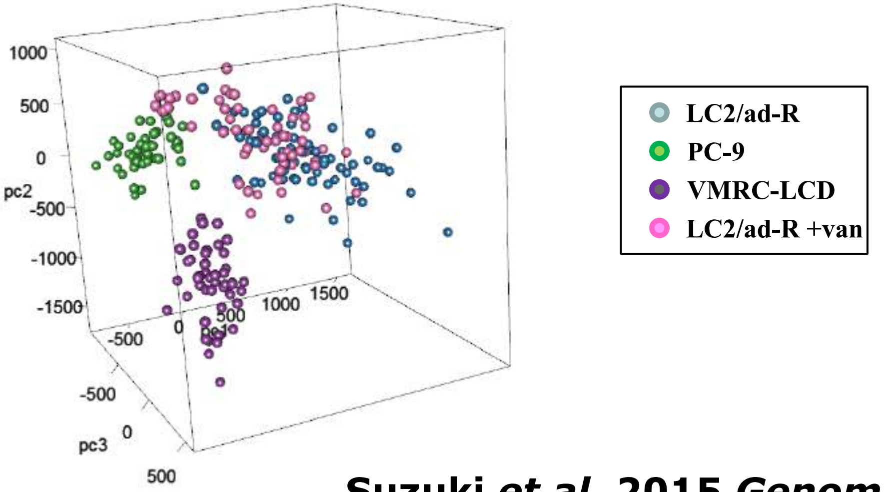
C1 Single-Cell Auto Prep system (Fluidigm)

HiSeq2500 (illumina)

Amplified cDNA

Sequencing library (12-plex)





Suzuki *et al.* 2015 *Genome*

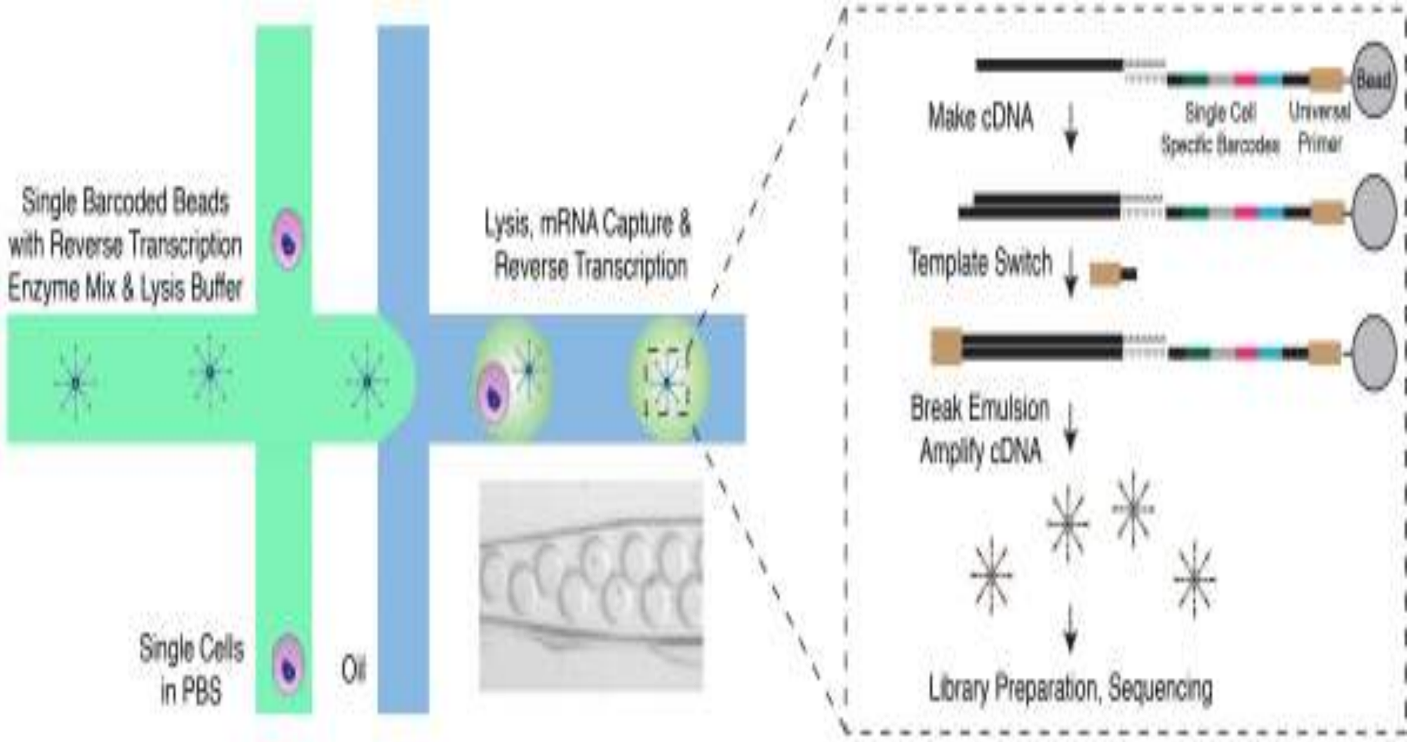
We have previously analyzed single-cell RNA-Seq data of lung cancer cell lines with drug treatments using C1 system.

We also plan to conduct

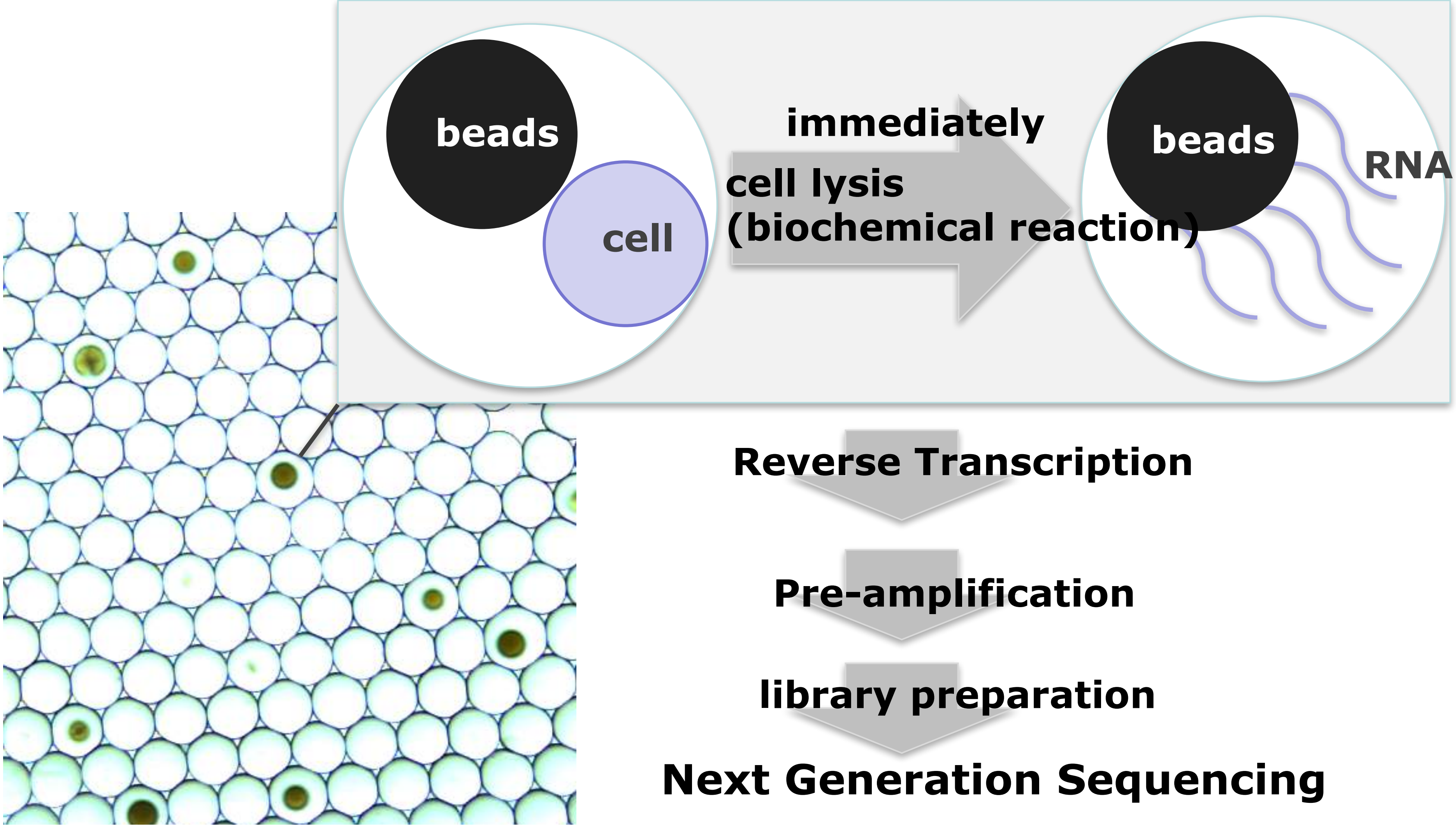
Macosko et al Cell (2015

Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets

Evan Z. Macosko,[1,2,3,*] Anindita Basu,[4,5] Rahul Satija,[4,6,7] James Nemesh,[1,2,3] Karthik Shekhar,[4] Melissa Goldman,[1,2]
Itay Tirosh,[4] Allison R. Bialas,[8] Nolan Kamitaki,[1,2,3] Emily M. Martersteck,[9] John J. Trombetta,[4] David A. Weitz,[5,10]
Joshua R. Sanes,[9] Alex K. Shalek,[4,11,12] Aviv Regev,[4,13,14] and Steven A. McCarroll[1,2,3,*]



Drop-seq



immediately
cell lysis
(biochemical reaction)

Reverse Transcription

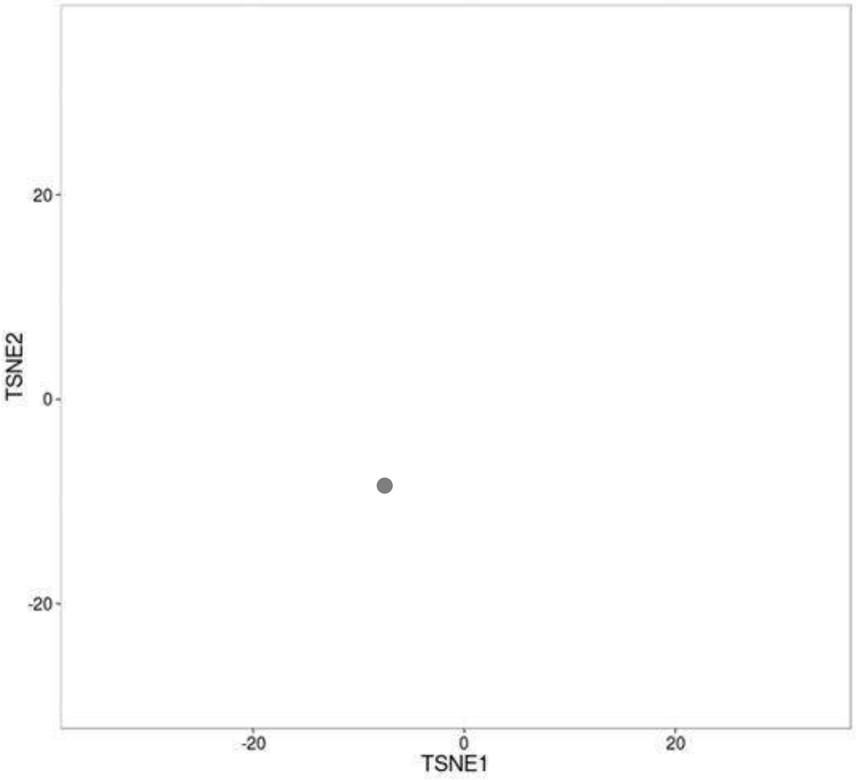
Pre-amplification

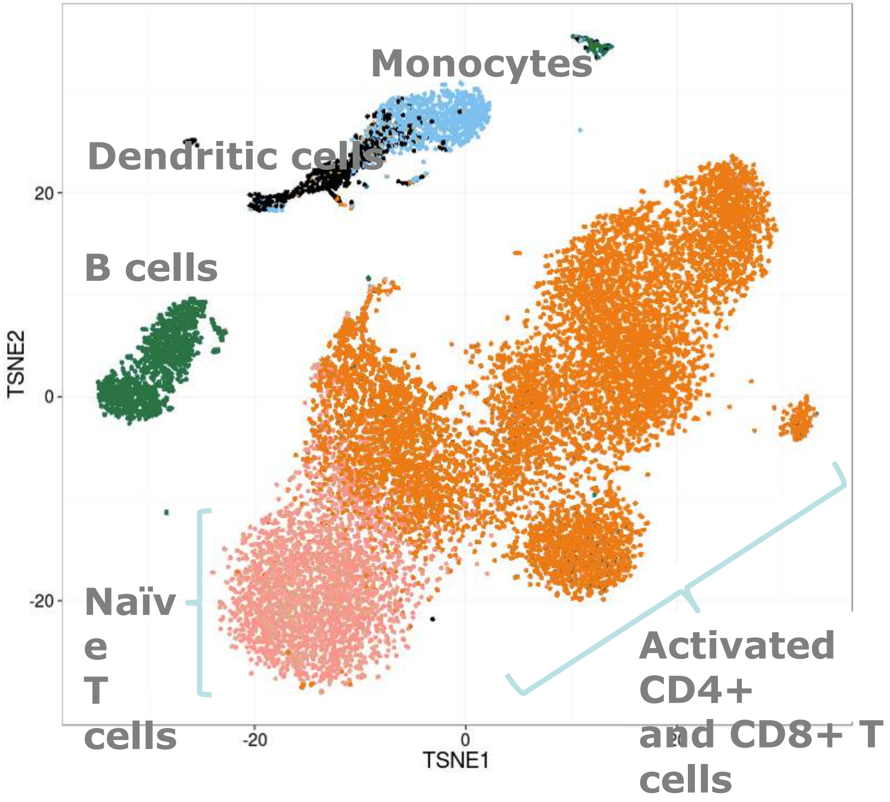
library preparation

Next Generation Sequencing

Gene Expression of PBMC for detailed analyses of infectious diseases

Conventional Gene Expression **10x Genomics**





One “average” data point from a mixture of cells

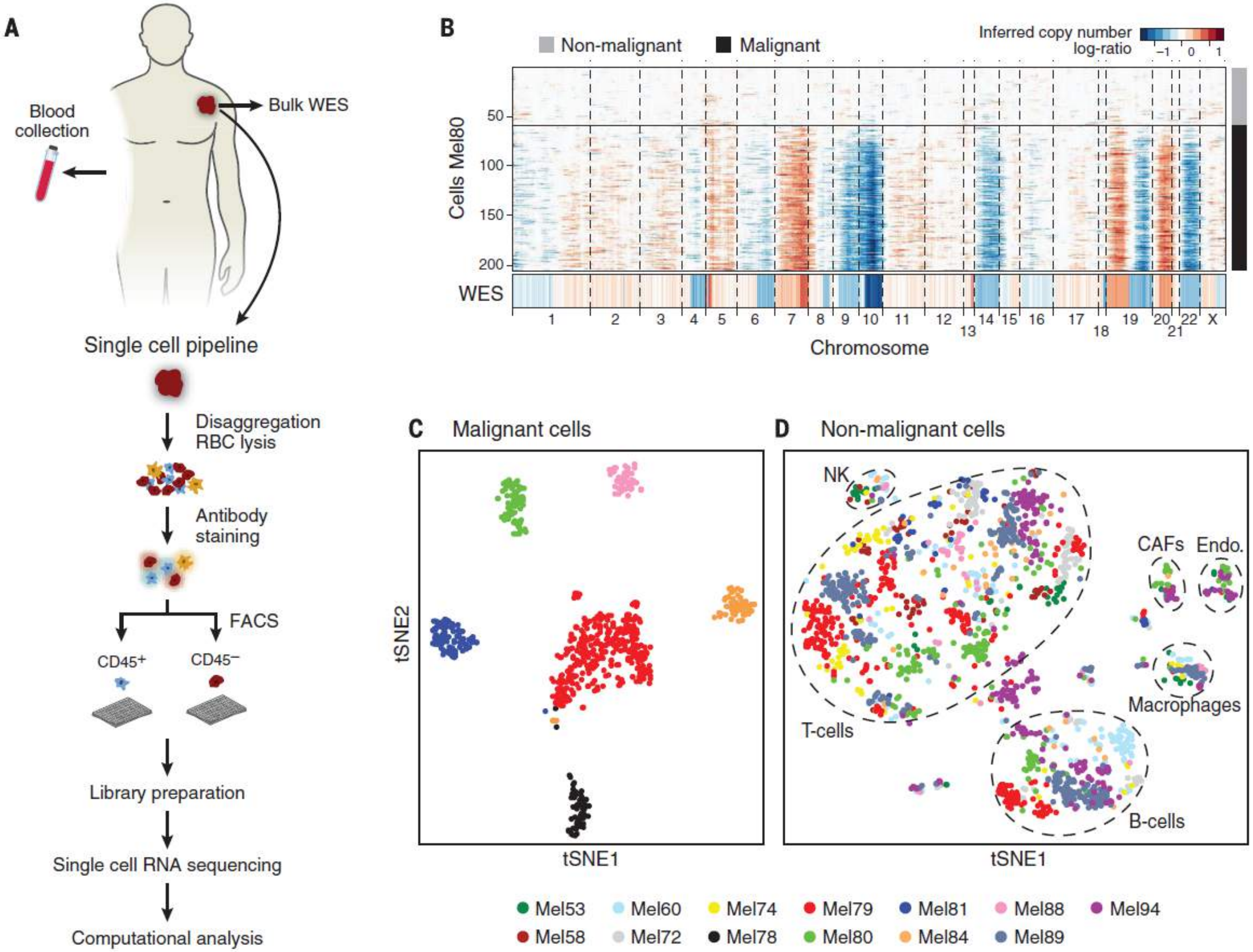
- Number of cells detected: ~16,000 cells, Number of raw reads per cell: ~15k

Confidential; Courtesy of 10X genomics co.

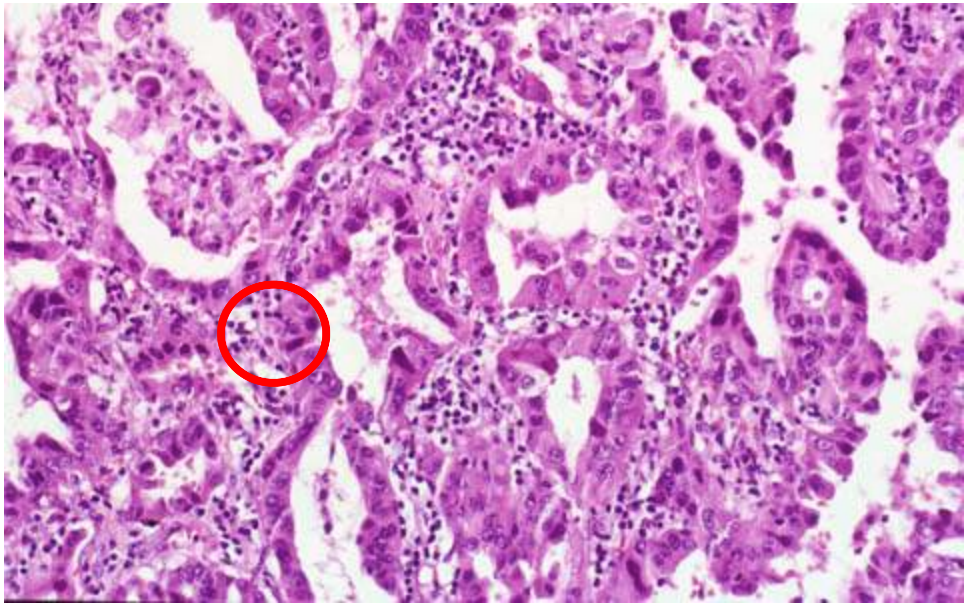
CANCER GENOMICS

Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq

**Itay Tirosh,[1]* Benjamin Izar,[1,2,3]*†‡ Sanja
Marc H. Wadsworth II,[1,4,5,6] Daniel Trea
Christopher Rodman,[1] Christine Lian,[7] G
Ken Dutton-Regester,[1,2,9] Jia-Ren Lin,[10]
Alex S. Genshaft,[1,4,5,6] Travis K. Hughes,
Samuel W. Kazer,[1,4,5,6] Aleth Gaillard,[1,4,
Alexandra-Chloé Villani,[1] Cory M. Johan
Eliezer M. Van Allen,[1,2,3] Monica Bertagn
Ryan J. Sullivan,[15] Keith T. Flaherty,[15] D
Charles H. Yoon,[12,13]† Orit Rozenblatt-Ro
Aviv Regev,[1,17,18]†‡ Levi A. Garraway[1,2,3,**



(in Vivo)Single Cell Biology of Cancer



Lung Adenocarcinoma

Cancer cells

- **Micro enviromnet**
- **Cancer evoluion**
- **EMT**
- **Cancer Stem Cell**
- **Metastasis/ drug resistance**

Scaffold Cel

- **Cancer Associated Fibroblas**

Immune Cells

- **Macrophage**
- **Cytotoxic T Cells**

Hopefully 1,000- 10,000 cells

Keeping heterogeneity

Collabirations

TRANSCRIPTION

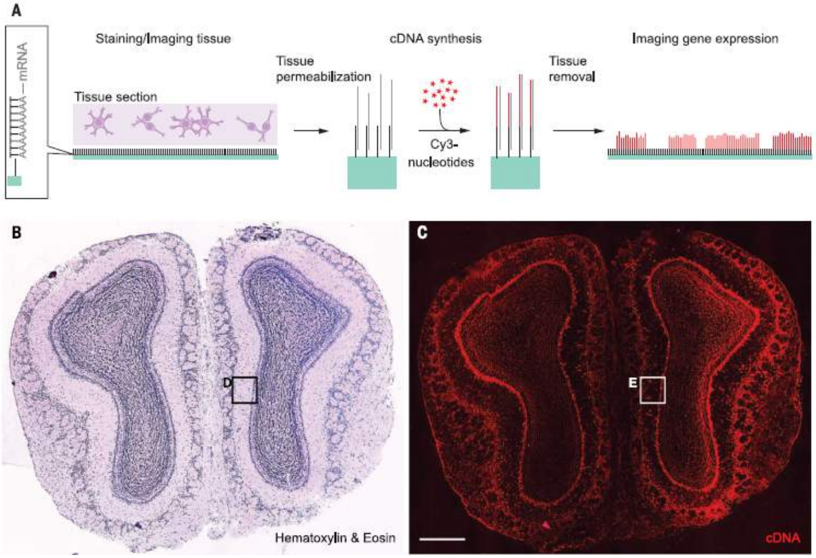
Visualization and analysis of gene expression in tissue sections by spatial transcriptomics

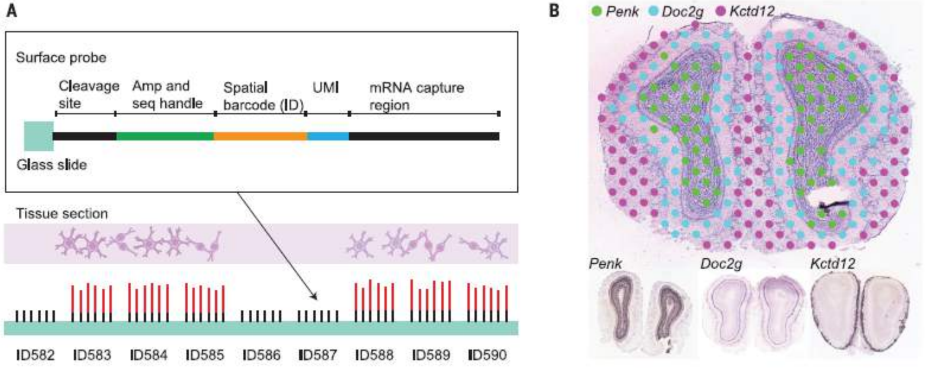
Patrik L. Ståhl,[1,2]* Fredrik Salmén,[2]* Sanja Vickovic,[2]† Anna Lundmark,[2,3]† José Fernández Navarro,[1,2] Jens Magnusson,[1] Stefania Giacomello,[2] Michaela Asp,[2] Jakub O. Westholm,[4] Mikael Huss,[4] Annelie Mollbrink,[2] Sten Linnarsson,[5] Simone Codeluppi,[5,6] Åke Borg,[7] Fredrik Pontén,[8] Paul Igor Costea,[2] Pelin Sahlén,[2] Jan Mulder,[9] Olaf Bergmann,[1] Joakim Lundeberg,[2]‡ Jonas Frisén[1]

Analysis of the pattern of proteins or messenger RNAs (mRNAs) in histological tissue sections is a cornerstone in biomedical research and diagnostics. This typically involves the visualization of a few proteins or expressed genes at a time. We have devised a strategy, which we call "spatial transcriptomics," that allows visualization and quantitative analysis of the transcriptome with spatial resolution in individual tissue sections. By positioning histological sections on arrayed reverse transcription primers with unique positional barcodes, we demonstrate high-quality RNA-sequencing data with maintained two-dimensional positional information from the mouse brain and human breast cancer. Spatial transcriptomics provides quantitative gene expression data and visualization of the distribution of mRNAs within tissue sections and enables novel types of bioinformatics analyses, valuable in research and diagnostics.

Science AAAS

Visualization and analysis of gene expression in tissue sections by spatial transcriptomics
Patrik L. Ståhl, Fredrik Salmén, Sanja Vickovic, Anna Lundmark, José Fernández Navarro, Jens Magnusson, Stefania Giacomello, Michaela Asp, Jakub O. Westholm, Mikael Huss, Annelie Mollbrink, Sten Linnarsson, Simone Codeluppi, Åke Borg, Fredrik Pontén, Paul Igor Costea, Pelin Sahlén, Jan Mulder, Olaf Bergmann, Joakim Lundeberg and Jonas Frisén (June 30, 2016)
Science **353** (6294), 78-82. [doi: 10.1126/science.aaf2403]





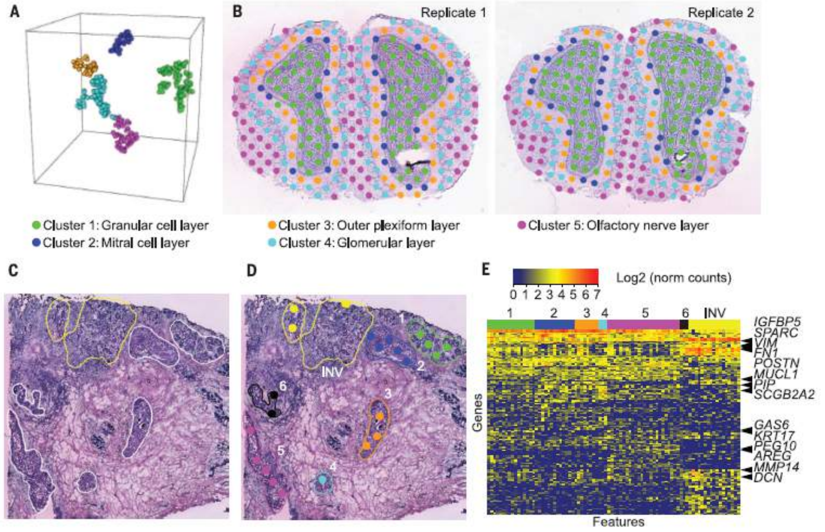
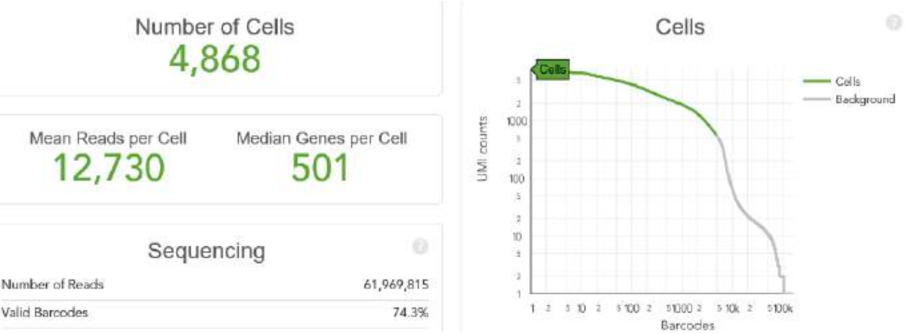


Fig. 4. Comparative analyses of tissue domains. (**A**) t-SNE analysis and hierarchical clustering of 551 features from two replicates creates five distinct clusters. (**B**) The features placed back onto the two tissue images. (**C** and **D**) Histological section of a breast cancer biopsy (C) containing invasive ductal cancer (INV) and six separate areas of ductal cancer in situ (1 to 6), with analyzed spatial transcriptomics features in (D). INV areas without, or with minimal, stromal infiltration were selected. (**E**) Gene expression heat map over the different areas in four adjacent sections (D) and (fig. S11).

Single cell RNA Seq of PBMC (mouse)

Condition 1



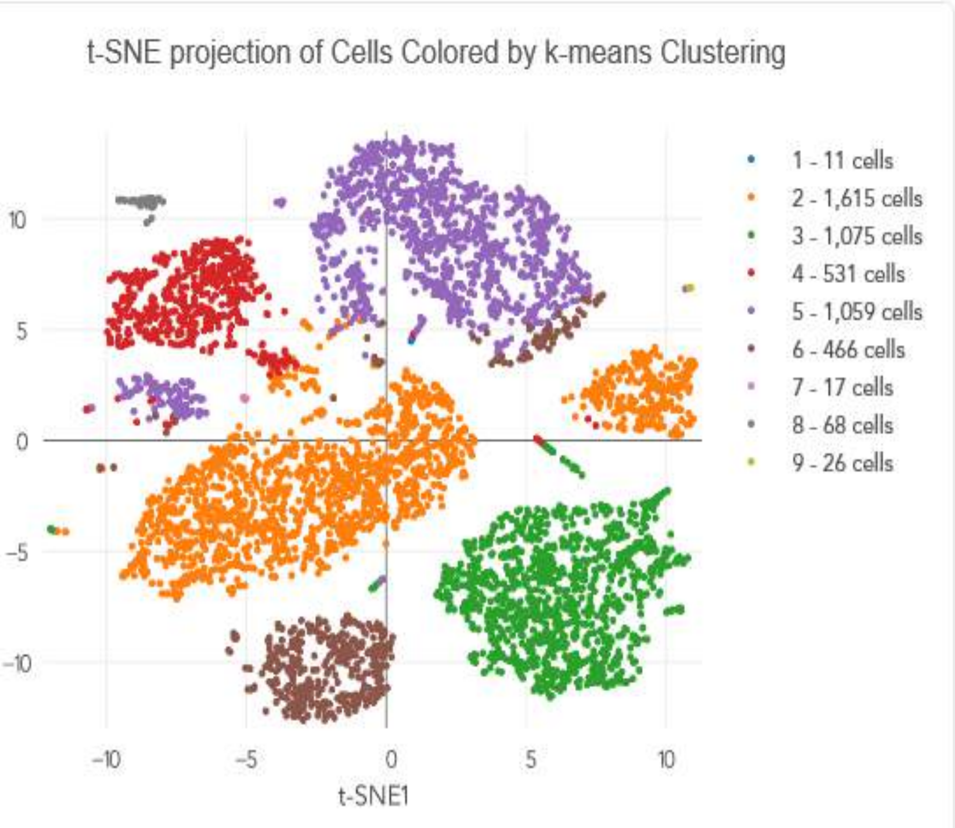
Number of Cells
4,868

Mean Reads per Cell
12,730

Median Genes per Cell
501

Sequencing

Number of Reads	61,969,815
Valid Barcodes	74.3%



t-SNE projection of Cells Colored by k-means Clustering

Condition 2

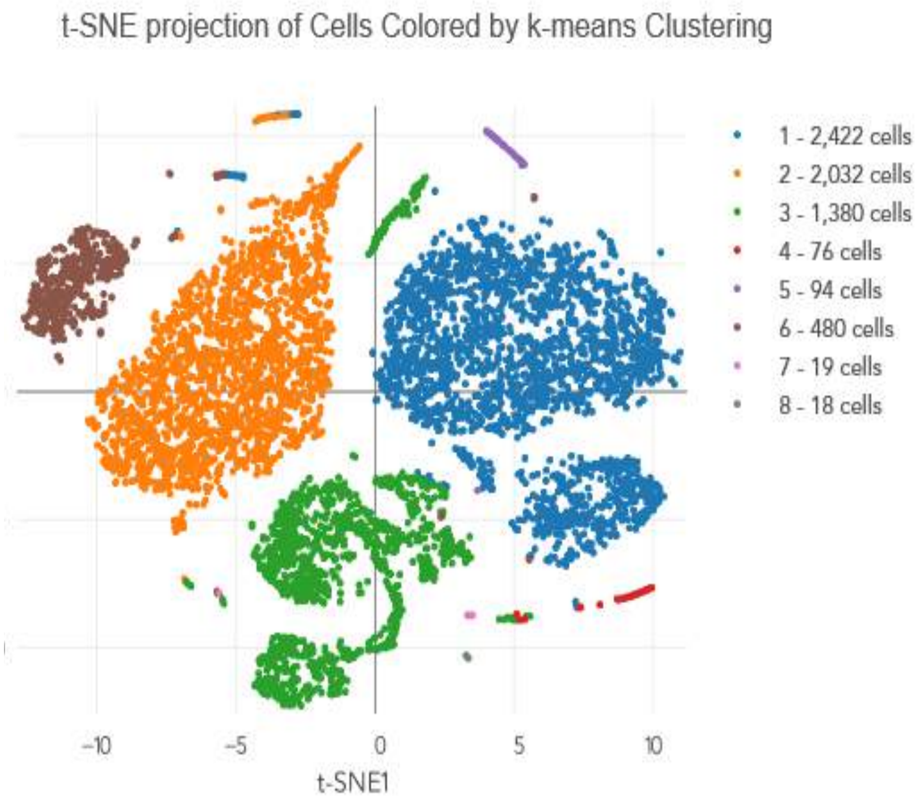
Number of Cells
6,521

Mean Reads per Cell
9,199

Median Genes per Cell
464

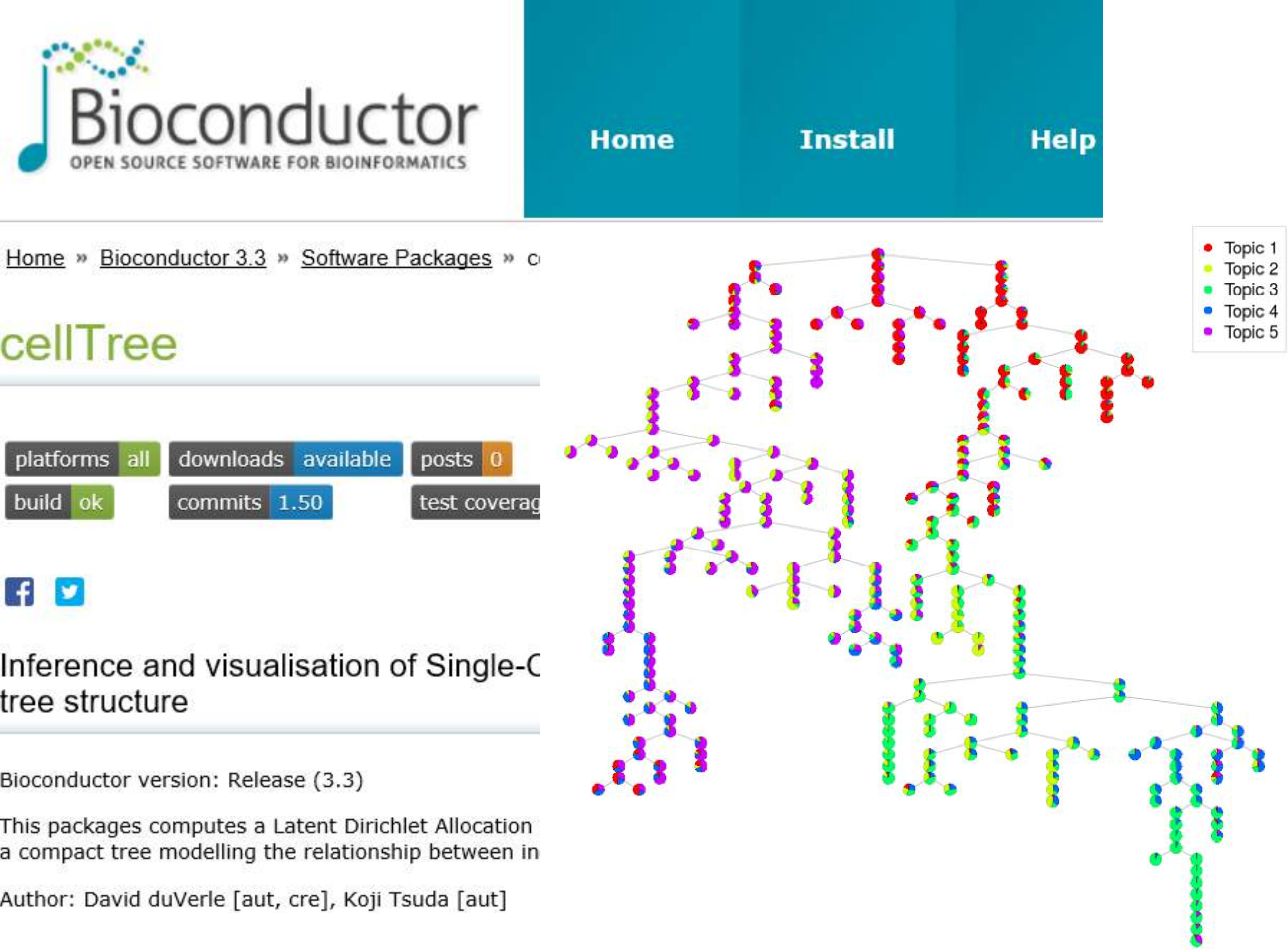
Sequencing

Number of Reads	59,988,774
Valid Barcodes	77.3%



t-SNE projection of Cells Colored by k-means Clustering

Needs for RNA capture to enrich tags for the ge



Home » Bioconductor 3.3 » Software Packages » c

cellTree

Inference and visualisation of Single-C
tree structure

Bioconductor version: Release (3.3)

This packages computes a Latent Dirichlet Allocation
a compact tree modelling the relationship between in

Author: David duVerle [aut, cre], Koji Tsuda [aut]

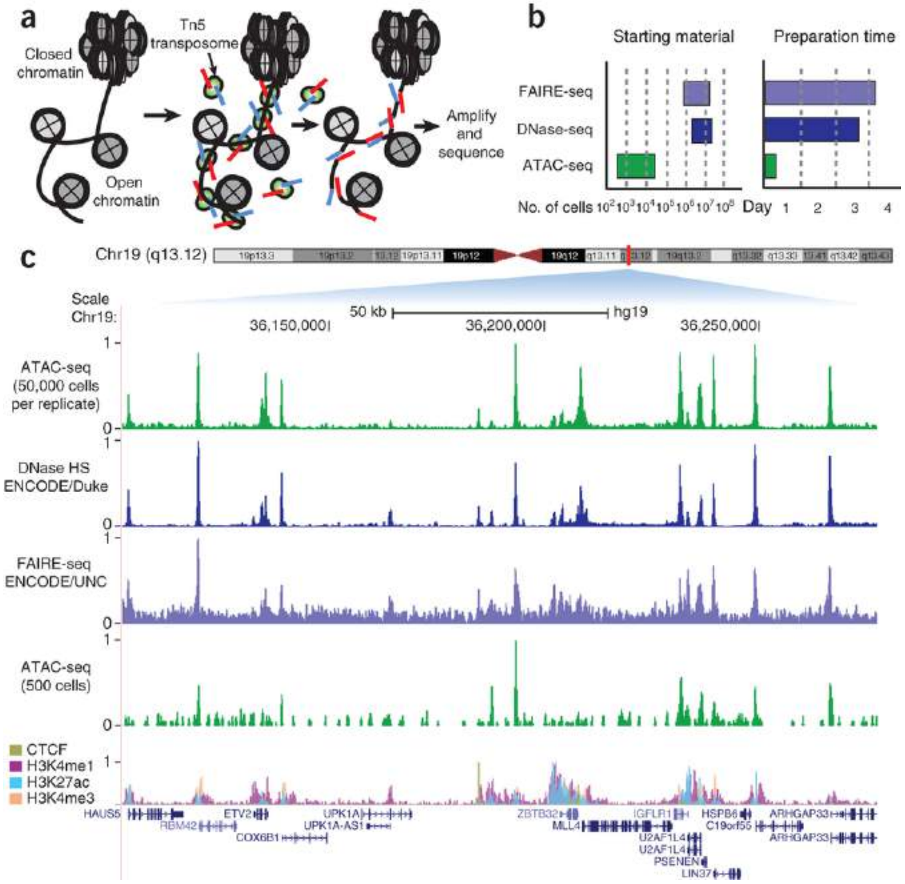
PyClone Nat Methods (2014)

Clonatiry 解析

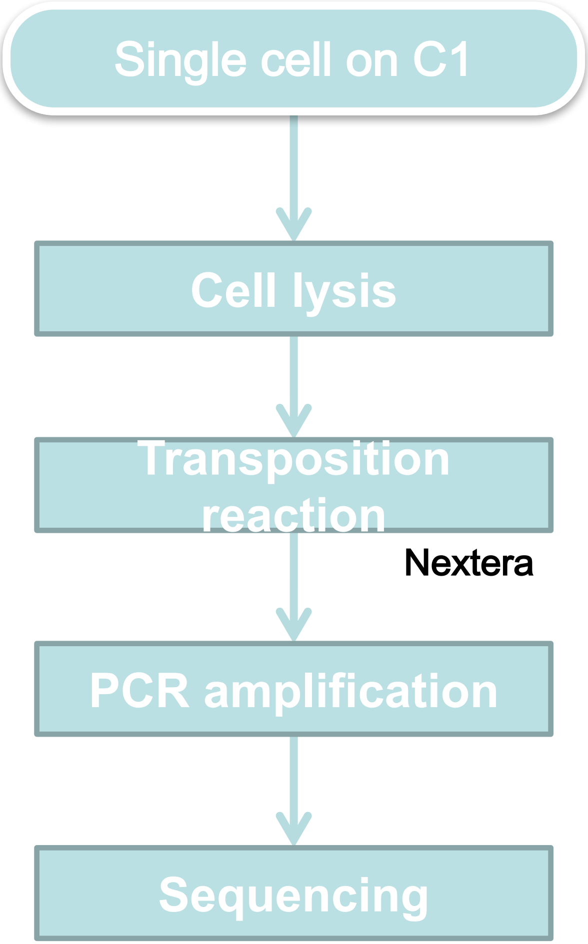
Diversity in Epigenome? Explaining the Diversity in Transcriptomes??

Epigenomic analysis at low input cell numbe

Assay for transposase-accessible chromatin using sequencin



Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome



Single cell on C1

Cell lysis

Transposition reaction

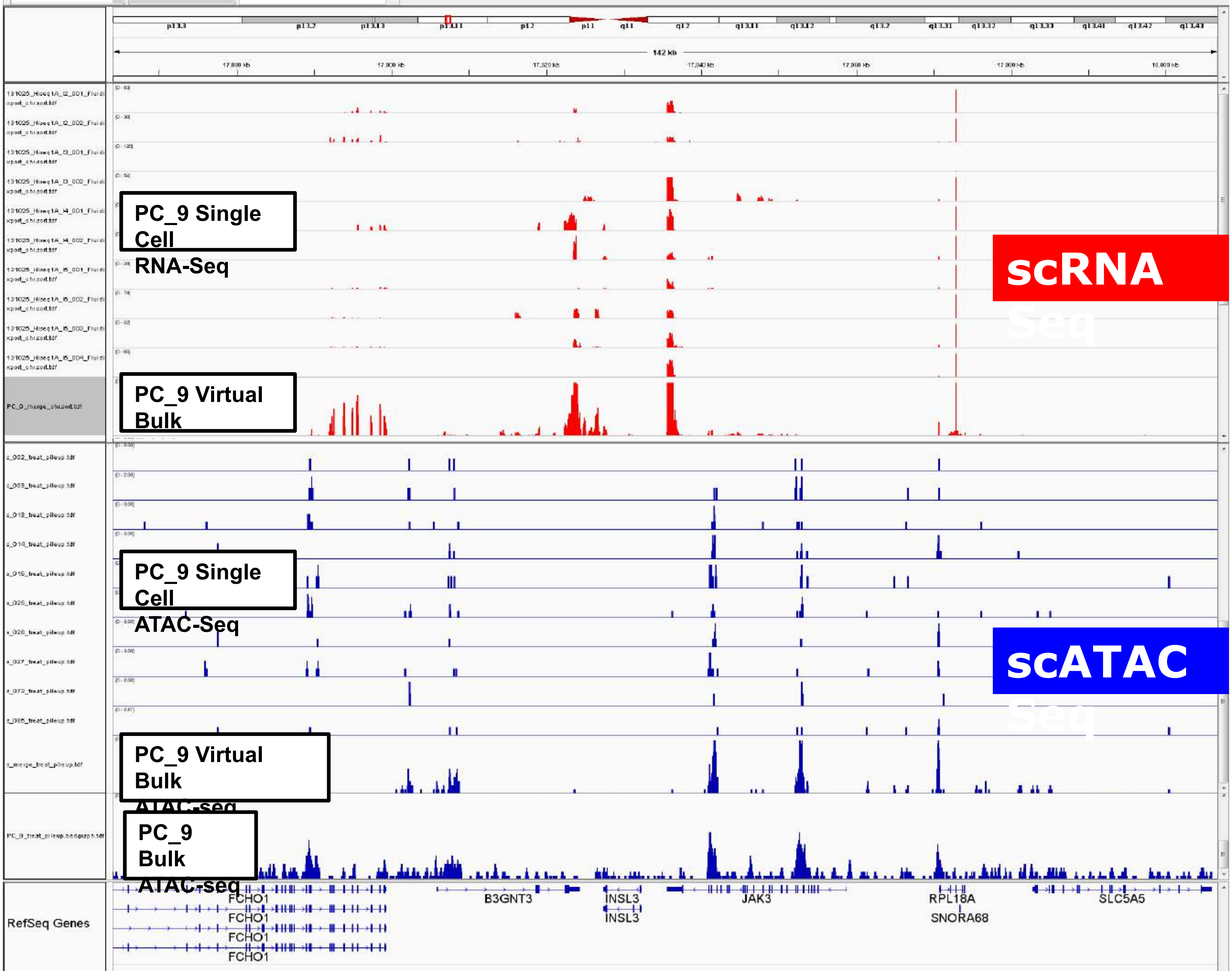
Nextera

PCR amplification

Sequencing

HiSeq2500, PE

Greenleaf et al Nature 2015



PC_9 Single Cell RNA-Seq

PC_9 Virtual Bulk

scRNA Seq

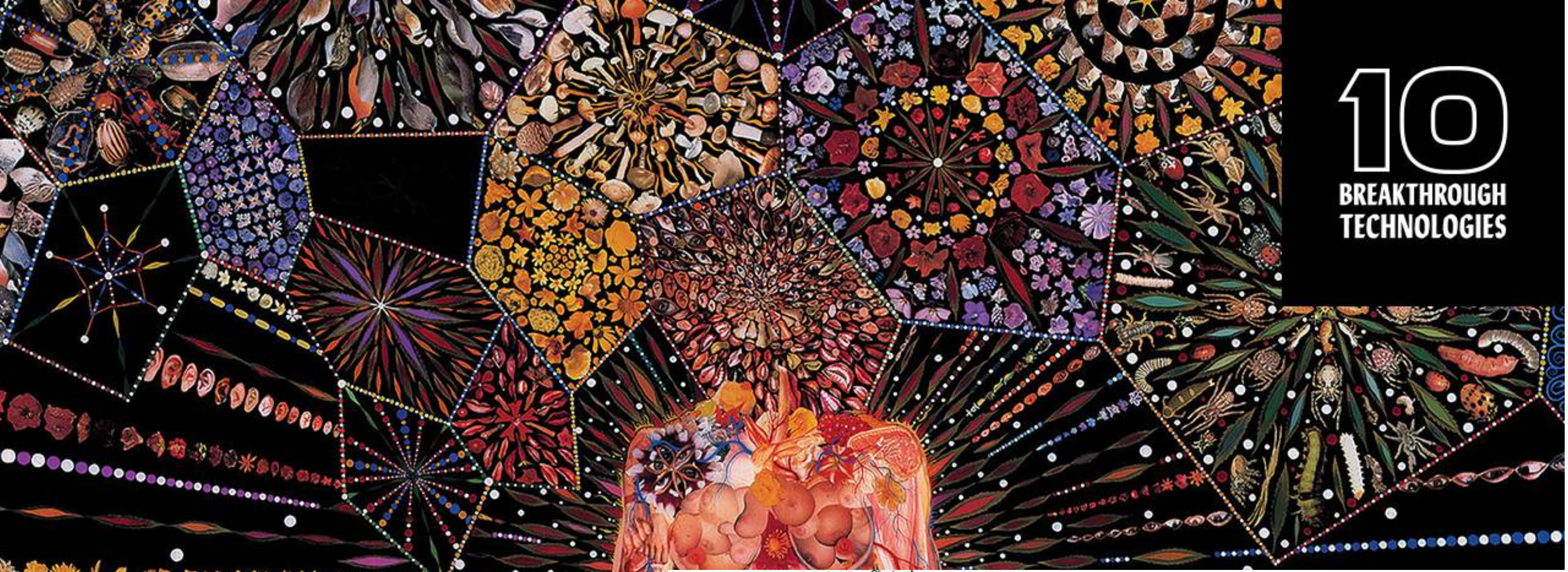
PC_9 Single Cell ATAC-Seq

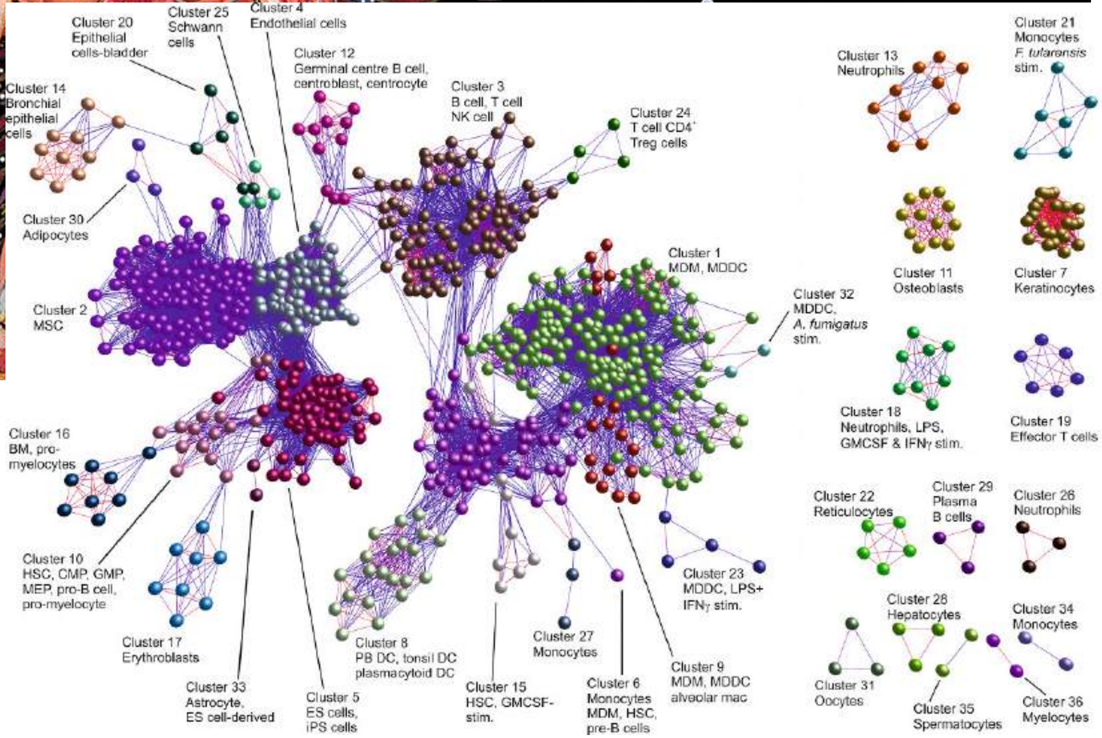
PC_9 Virtual Bulk ATAC-seq

PC_9 Bulk ATAC-seq

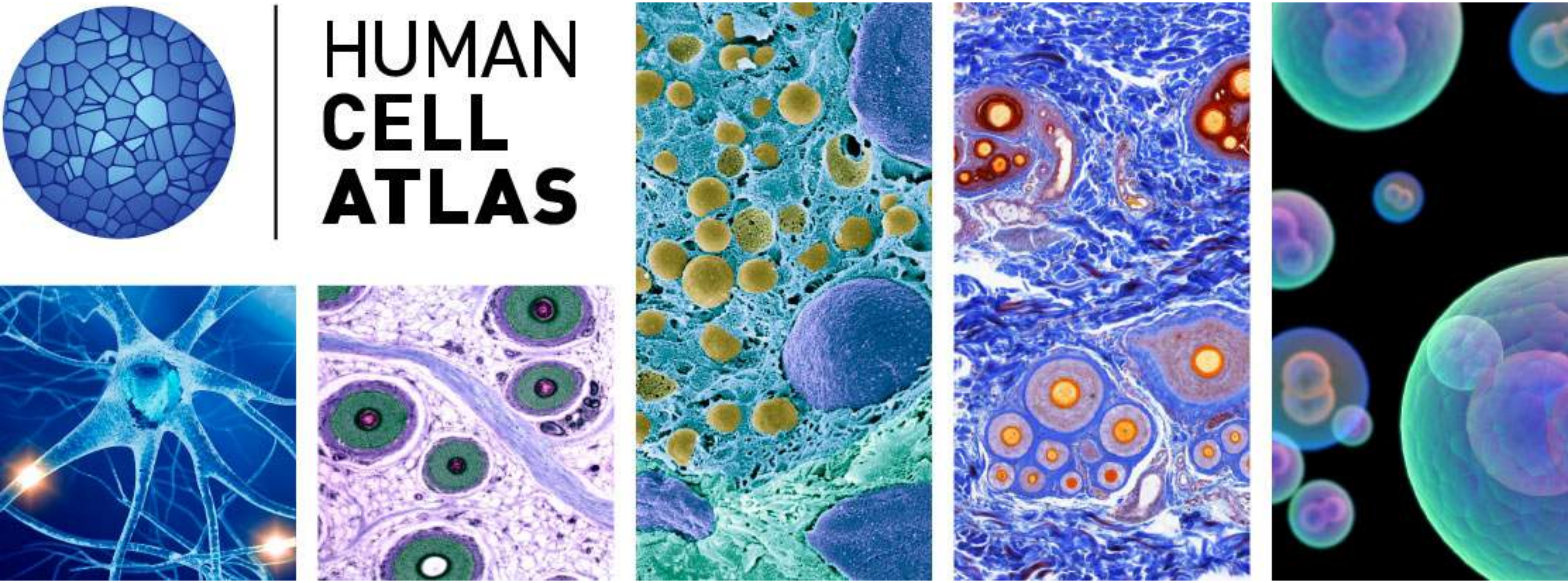
scATAC Seq

10 BREAKTHROUGH TECHNOLOGIES





From Bulk Cells to Single Cells



Human Cell Atlas:
http://www.ebi.ac.uk/about/jobs/human-cell-atlas-opportunities

Genotyping Pathogens of Tropical Diseases; Making use of a portable sequencer, MinION

Yutaka Suzuki
partment of Computational Biology and Medical Scienc
Graduate School of Frontier Sciences
The University of Tokyo

東京大学
THE UNIVERSITY OF TOKYO

Today’s topic:
Liberation of the Sequencing

Illumina HiSeq

Instrument: 1.5M USD (

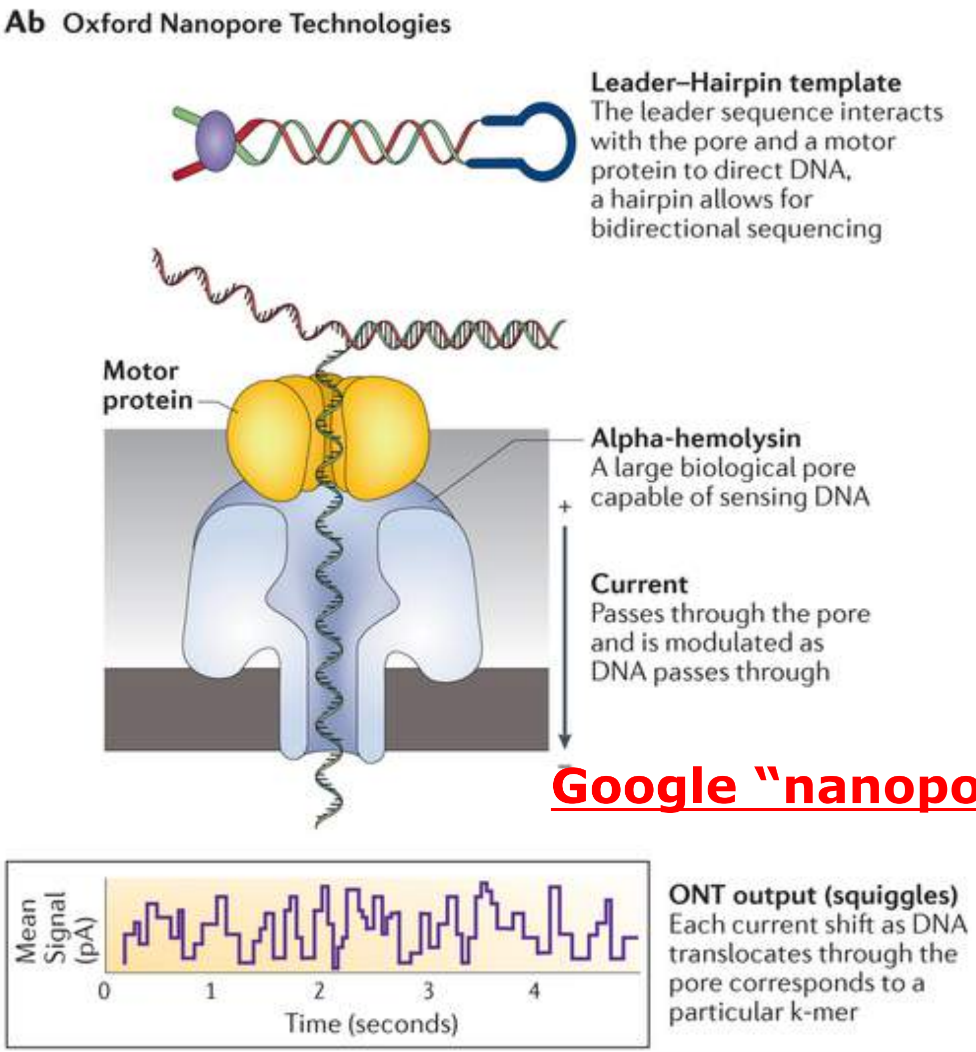
Reagents: 40,000 US

Oxford Nanopore MinION:
Portable, one time use
1000 USD/per use

Less powerful
But
Easy to start

New challenges are starting using a tin sequence...

MinION from Oxford Nanopore Technologies



Leader–Hairpin template
The leader sequence interacts with the pore and a motor protein to direct DNA, a hairpin allows for bidirectional sequencing

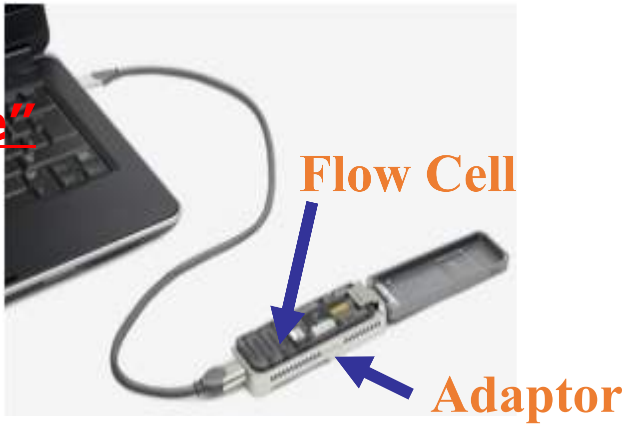
Alpha-hemolysin
A large biological pore capable of sensing DNA

Current
Passes through the pore and is modulated as DNA passes through

ONT output (squiggles)
Each current shift as DNA translocates through the pore corresponds to a particular k-mer

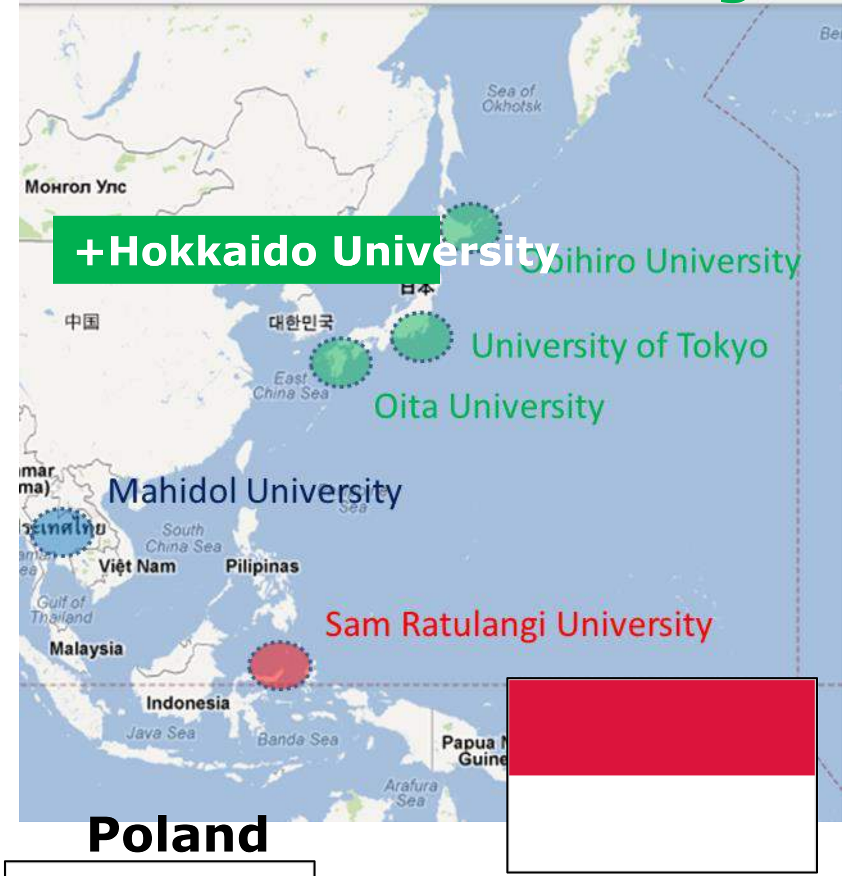
Google "nanopore"

- Read length: > 10 Kb
  - (λ-Phage DNA, 50 Kb)
- # of pore : 512 (MinION at early access)
- Run time: 6-48 hours
- Error rate =~ 10%
- >1M reads/run
- Publicly available



For Tropical Diseases, the collaboration

JSPS Core-to-Core Program (2012-2015; 2016-2018)



JAPAN

- **Yutaka Suzuki; Professor, University of Tokyo**
- **Ryuichiro Maeda, Professor, Obihiro University of Agriculture and Veterinary Sciences**
- **Junya Yamagishi; Associate Professor, Research Center for Zoonosis Control, Hokkaido University**
- **Yuki Eshita; Associate Professor, Oita University**
- **Fujihiko Minamoto, Technical staff, IMS> University of Tokyo**

Indonesia

Poland

Vertical inversion

UNSRAT

- **Josef Tuda; Professor, School of Medicine**
- **Arthur E. Mongan; School of Medicine**
- **Sarah Warouw; School of Medicine, Dean**

Members: Marie-Culie Research Promotion Program (2012-2

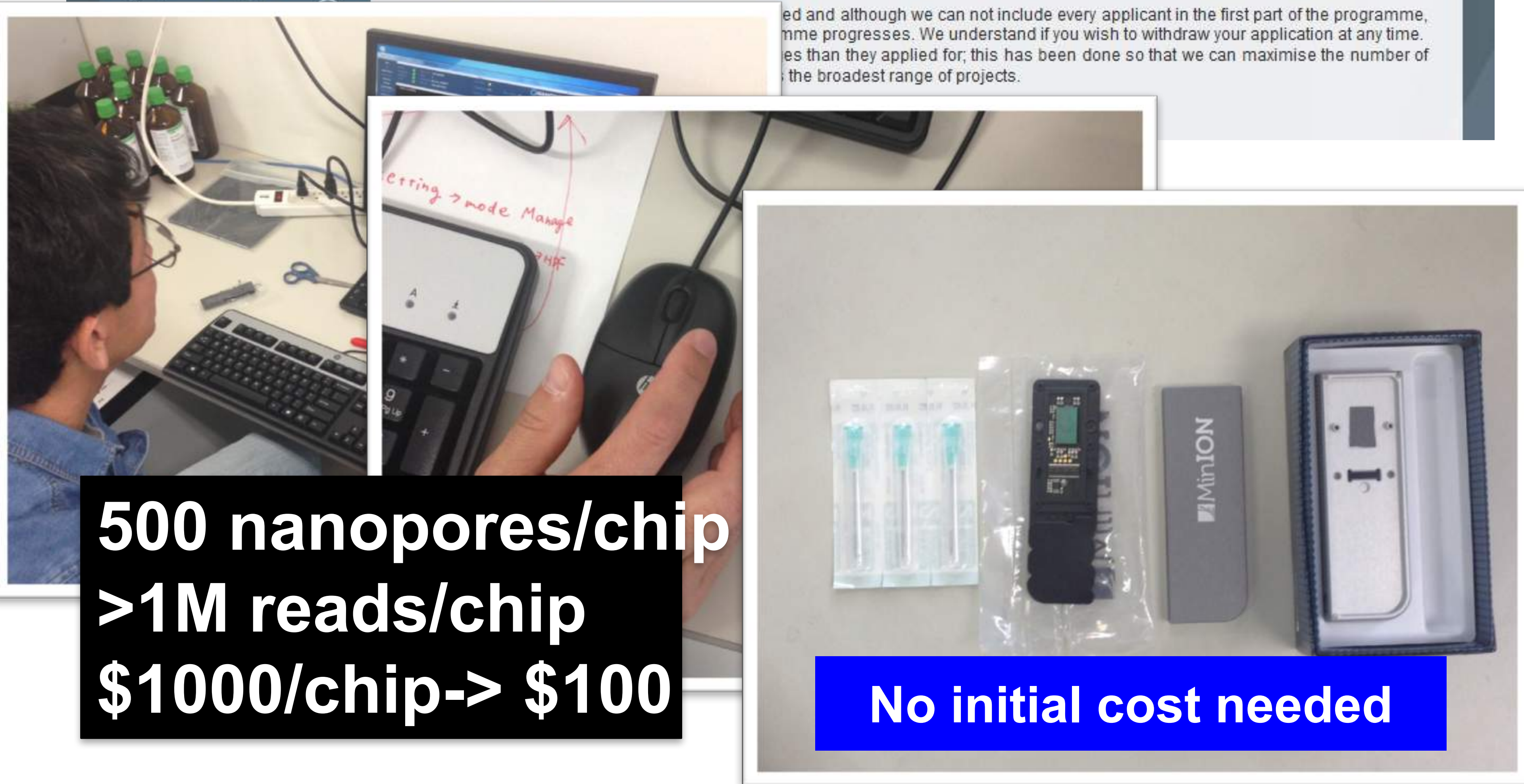
EU

- **Wojciech Makalowski, Munster University, Germany**

MinION™ Access Programme

Status update 14th February 2014: Oxford Nanopore has started to issue invitations to a number of applicants to take part in the first part of the MinION Access Programme (MAP). Applicants should receive an email within the next few days, either inviting the applicant to review some information and register for the MAP, or notifying the applicant that we are unable to invite them into the first wave of participants.

ed and although we can not include every applicant in the first part of the programme,
mme progresses. We understand if you wish to withdraw your application at any time.
es than they applied for; this has been done so that we can maximise the number of
the broadest range of projects.



500 nanopores/chip
>1M reads/chip
$1000/chip-> $100

No initial cost needed

Nature 7 May, 20

BIOTECHNOLOGY

Pint-sized DNA sequencer impresses first users

Portable device offers on-the-spot data to fight disease, catalogue species and more.

BY ERIKA CHECK HAYDEN

In April, Joshua Quick boarded a plane to Guinea with three genetic sequencers packed in his luggage. That fact alone is astonishing: most sequencing machines are much too heavy and delicate to travel as checked baggage in the hold of a commercial airliner. What came next was even more impressive. For 12 days, Quick used these

be used

“This

says Qu

Nichola

at the U

don't ha

and cost

Quic

what ex

palm-si

uz. “We're all

suggested that it was not nearly as fast as ▶

International Space Station

Biomolecule Sequencer (Biomolecule Sequencer) - 11.02.16

ISS Science for Everyone

Science Objectives for Everyone

Living organisms contain DNA, or deoxyribonucleic acid, and sequencing DNA is a powerful way to understand how they respond to changing environments. The Biomolecule Sequencer investigation seeks to demonstrate, for the first time, that DNA sequencing is feasible in an orbiting spacecraft. A space-based DNA sequencer could identify microbes, diagnose diseases and understand crew member health, and potentially help detect DNA-based life elsewhere in the solar system.

Target sequencing (entire gene regions by R

Four cancer-related genes

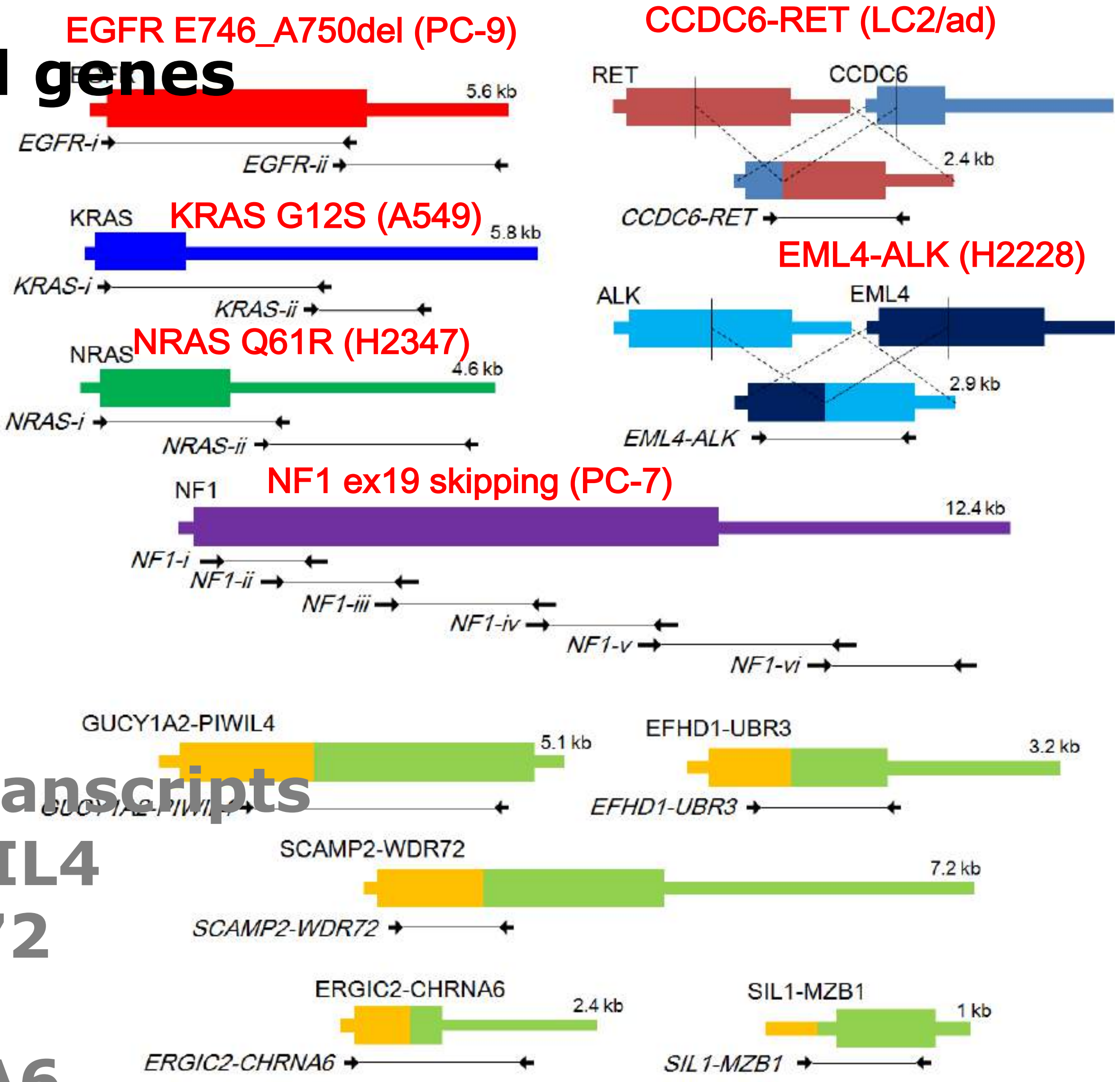
- **EGFR**
- **KRAS**
- **NRAS**
- **NF1**

Two fusion genes

- **EML4-ALK**
- **CCDC6-RET**

Five novel fusion transcripts

- GUCY1A2-PIWIL4
- SCAMP2-WDR72
- EFHD1-UBR3
- ERGIC2-CHRNA6
- SIL1-MZB1



EGFR E746_A750del (PC-9)

CCDC6-RET (LC2/ad)

KRAS G12S (A549)

EML4-ALK (H2228)

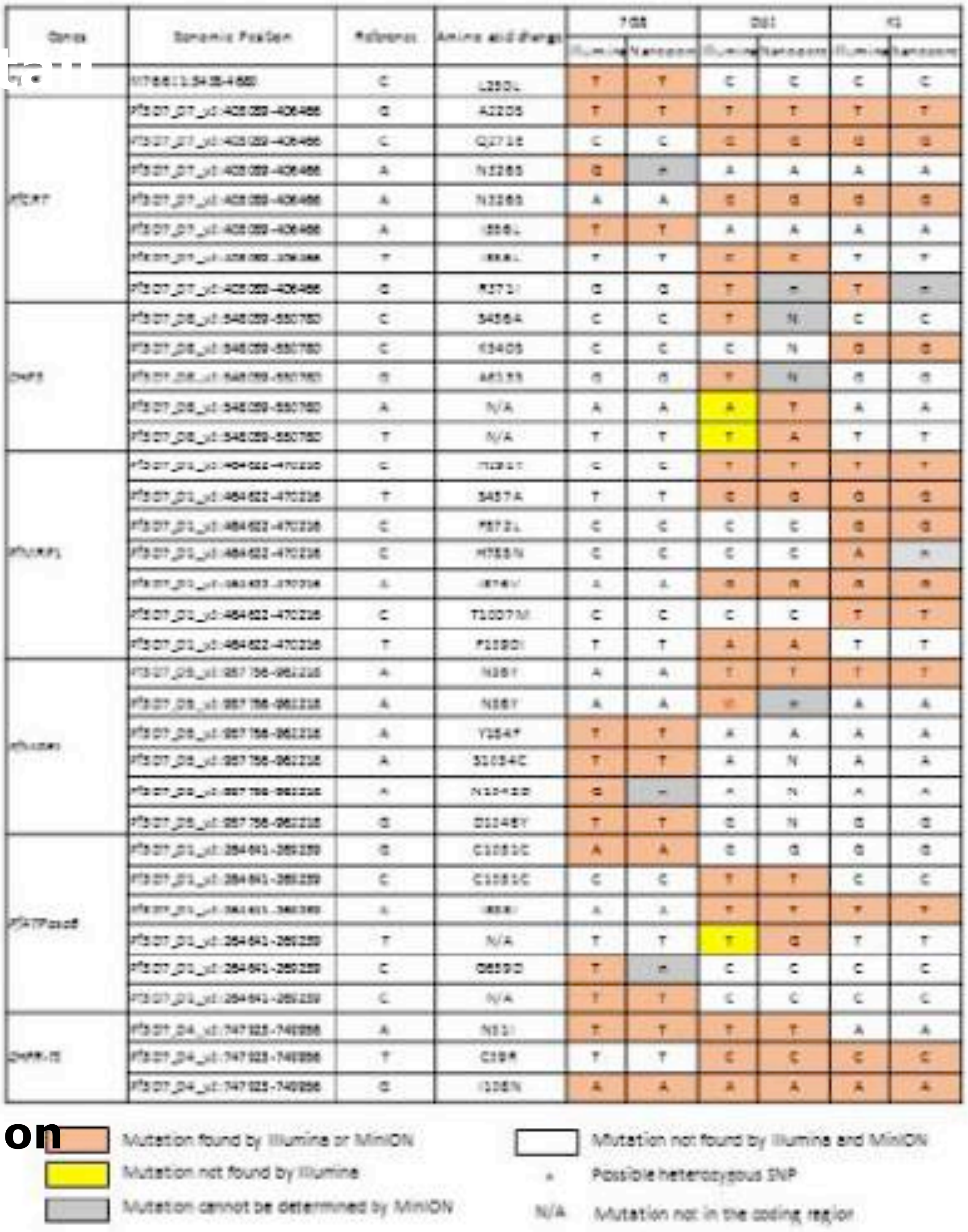
NRAS Q61R (H2347)

NF1 ex19 skipping (PC-7)

Target length: 640 - 34

Pf target gene co tail

B



Genes	Genomic Position	Reference	Amino acid change	7G8 Illumina	7G8 Nanopore	Dd2 Illumina	Dd2 Nanopore	K1 Illumina	K1 Nanopore
[illegible]	[illegible]	C	L250L	T	T	C	C	C	C
PfCRT	Pf3D7_07_v3:[illegible]	G	A220S	T	T	T	T	T	T
	Pf3D7_07_v3:[illegible]	C	Q271E	C	C	G	G	G	G
	Pf3D7_07_v3:[illegible]	A	N326S	G	–	A	A	A	A
	Pf3D7_07_v3:[illegible]	A	N326S	A	A	G	G	G	G
	Pf3D7_07_v3:[illegible]	A	I356L	T	T	A	A	A	A
	[illegible]	T	[illegible]	T	T	C	C	T	T
	Pf3D7_07_v3:[illegible]	G	R371I	G	G	T	–	T	–
DHPS	Pf3D7_08_v3:[illegible]	C	S436A	C	C	T	N	C	C
	Pf3D7_08_v3:[illegible]	C	K540E	C	C	C	N	G	G
	Pf3D7_08_v3:[illegible]	G	A613S	G	G	T	N	G	G
	Pf3D7_08_v3:[illegible]	A	N/A	A	A	A	T	A	A
	Pf3D7_08_v3:[illegible]	T	N/A	T	T	T	A	T	T
PfUBP1	Pf3D7_01_v3:[illegible]	C	[illegible]	C	C	T	T	T	T
	Pf3D7_01_v3:[illegible]	T	S457A	T	T	G	G	G	G
	Pf3D7_01_v3:[illegible]	C	F572L	C	C	C	C	G	G
	Pf3D7_01_v3:[illegible]	C	H755N	C	C	C	C	A	–
	[illegible]	A	I876V	A	A	G	G	G	G
	Pf3D7_01_v3:[illegible]	C	T1007M	C	C	C	C	T	T
	Pf3D7_01_v3:[illegible]	T	F1190I	T	T	A	A	T	T
PfMDR1	Pf3D7_05_v3:[illegible]	A	N86Y	A	A	T	T	T	T
	Pf3D7_05_v3:[illegible]	A	N86Y	A	A	Y	–	A	A
	Pf3D7_05_v3:[illegible]	A	Y184F	T	T	A	A	A	A
	Pf3D7_05_v3:[illegible]	A	S1034C	T	T	A	N	A	A
	Pf3D7_05_v3:[illegible]	A	N1042D	G	–	A	N	A	A
	Pf3D7_05_v3:[illegible]	G	D1246Y	T	T	G	N	G	G
PfATPase6	Pf3D7_01_v3:[illegible]	G	C1081C	A	A	G	G	G	G
	Pf3D7_01_v3:[illegible]	C	C1081C	C	C	T	T	C	C
	[illegible]	A	[illegible]	A	A	T	T	T	T
	Pf3D7_01_v3:[illegible]	T	N/A	T	T	T	G	T	T
	Pf3D7_01_v3:[illegible]	C	G639D	T	–	C	C	C	C
	Pf3D7_01_v3:[illegible]	C	N/A	T	T	C	C	C	C
DHFR-TS	Pf3D7_04_v3:[illegible]	A	N51I	T	T	T	T	A	A
	Pf3D7_04_v3:[illegible]	T	C59R	T	T	C	C	C	C
	Pf3D7_04_v3:[illegible]	G	S108N	A	A	A	A	A	A

🟧 Mutation found by Illumina or MiniON

🟨 Mutation not found by Illumina

▨ Mutation cannot be determined by MiniON

☐ Mutation not found by Illumina and MiniON

* Possible heterozygous SNP

N/A Mutation not in the coding region

nina)

5)

Runtuwene et al in revision

Sequencing by MinION

Contact me at: ysuzuki@hgc

1 µg amplicon mixture (80 µl)

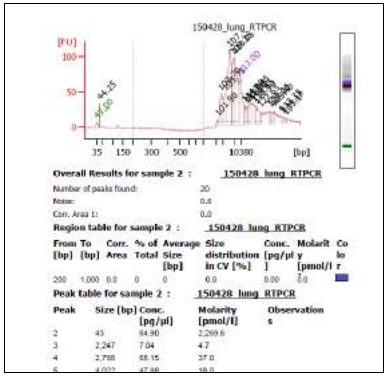
2-3h

End Repair

dA-tailing

Adapter Ligation

Sequencing library



Overall Results for sample 2 : 150428_lung_RTPCR

Number of peaks found:	20
Noise:	0.0
Corr. Area 1:	0.0

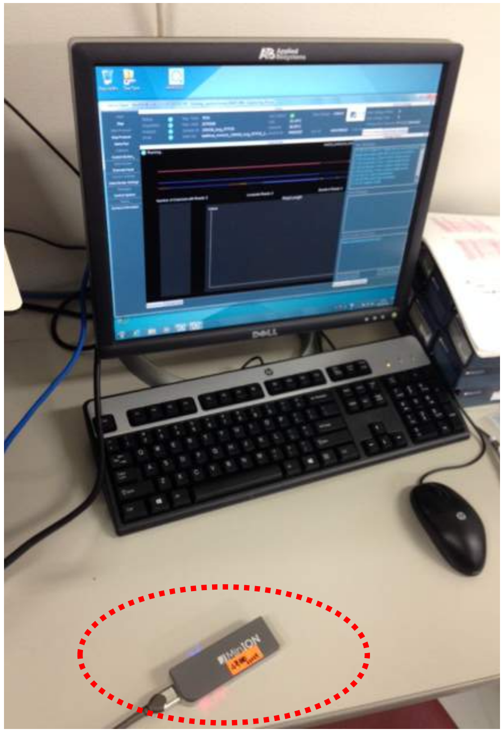
Region table for sample 2 : 150428_lung_RTPCR

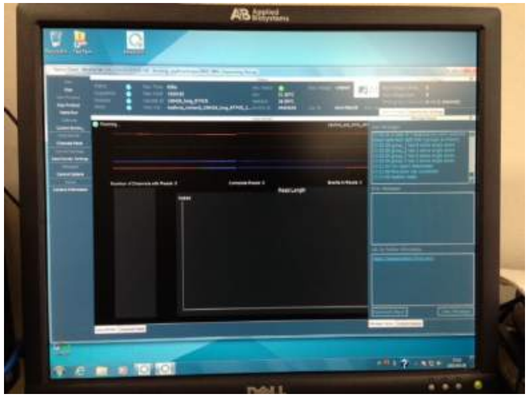
From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
200	1,000	0.0	0	0	0.0	0.00	0.0	🟦

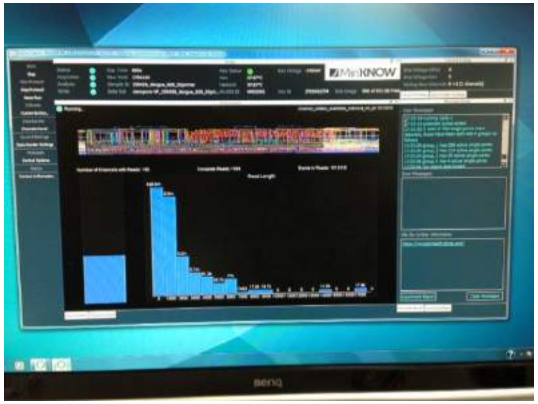
Peak table for sample 2 : 150428_lung_RTPCR

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
2	45	64.90	2,269.6	
3	2,247	7.04	4.7	
4	2,788	68.15	37.0	

Library loading







48h

Sequencing

R7

An example of alignments

Download GenBank Graphics

Homo sapiens epidermal growth factor receptor (EGFR), transcript variant 1, mRNA
Sequence ID: ref|NM_005228.3| Length: 5616 Number of Matches: 1

Range 1: 189 to 3418 GenBank Graphics

3 kb, 90% identity

Score	Expect	Identities	Gaps	Strand
3904 bits(2373)	0.0	2953/3295(90%)	252/3295(7%)	Plus/Plus

```
y  304  CTTTCCTCTTTAAAGACCATCCAGGA-GTGGCTGGTTATGTCCTCATTG-CCTCAACACA  361
t  472  CTTTCCTTCTTAAAGACCATCCAGGAGGTGGCTGGTTATGTCCTCATTGCCCTCAACACA  531

y  362  GTGGAGCGAATTCCTTTGG-AAACCTGCAGATCATCAGAGGAAATA-G------GAAAAT  413
t  532  GTGGAGCGAATTCCTTTGGAAAACCTGCAGATCATCAGAGGAAATATGTACTACGAAAAT  591

y  414  TCCTATGCCTTAGCAGTCTTATCTAACTATGATGCAAATAAAACC-GACTGAAGGAGCTG  472
t  592  TCCTATGCCTTAGCAGTCTTATCTAACTATGATGCAAATAAAACCGGACTGAAGGAGCTG  651

y  473  CCCATGAGAAATTTACAGGAAATCCTGCATGGCGCTGTCCGCGGTTCAGCAACAACCCTG  532
t  652  CCCATGAGAAATTTACAGGAAATCCTGCATGGCGCCGT--GCGGTTCAGCAACAACCCTG  709

y  533  CGCTGTGCAATGTGGAGAGCATCCAGTGGC--GACATAGTCAGCAGTGACTTTCTCAGCA  590
t  710  CCCTGTGCAACGTGGAGAGCATCCAGTGGCGGGACATAGTCAGCAGTGACTTTCTCAGCA  769

y  591  ACATGTCGATGGACTTCCAGAACCACCTGGGGCAGCTGCCAAAA---TGATCTCAAGCTG  647
t  770  ACATGTCGATGGACTTCCAGAACCACCT-GGGCAGCTGCCAAAAGTGTGATC-CAAGCTG  827

y  648  TCCCAATGGGAGCTGCTGGGGTGCAGGAGAGGAAAACTGCCAGAAACTGACC-AAATCAT  706
t  828  TCCCAATGGGAGCTGCTGGGGTGCAGGAGAGGAGAACTGCCAGAAACTGACCAAAATCAT  887

y  707  CTGTGCCCAGCAGTGCTCCGGGCGCTGCCGTGGC-ATTCCCCCAGTGACTGCTGC--CAA  763
```

nt

Nucleotide BLAST
Query: d0d98a04-22e2-4559-b360-92deaae0e929_Basecall_2D_000_2d pass/nanopore_HP_150910_LungAdenocarcinoma_PCRmix_1_5516_1_ch92_file318_strand.fast5
Database: Human genomic plus transcript

Development of a bioinformatics pipeline

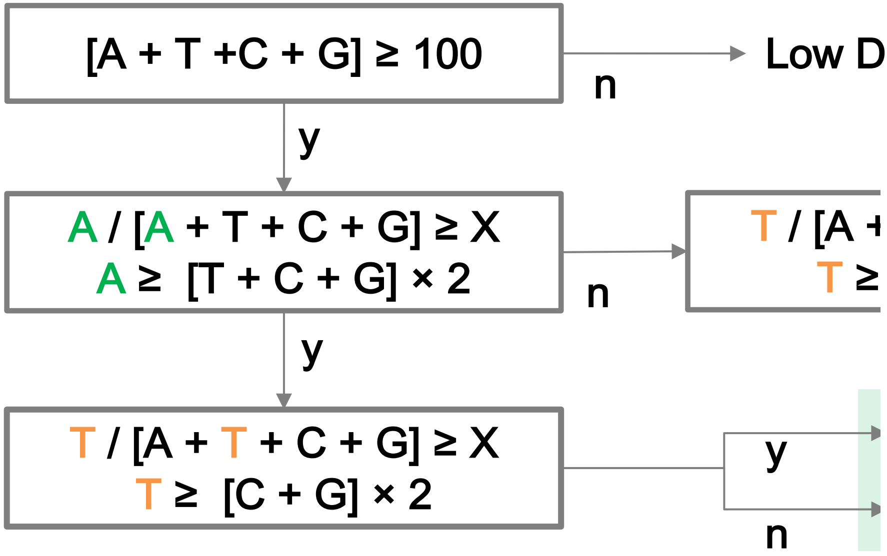
Four kinds of bases
→ A, T, C, G
Number of supporting reads
→ A ≥ T ≥ C ≥ G

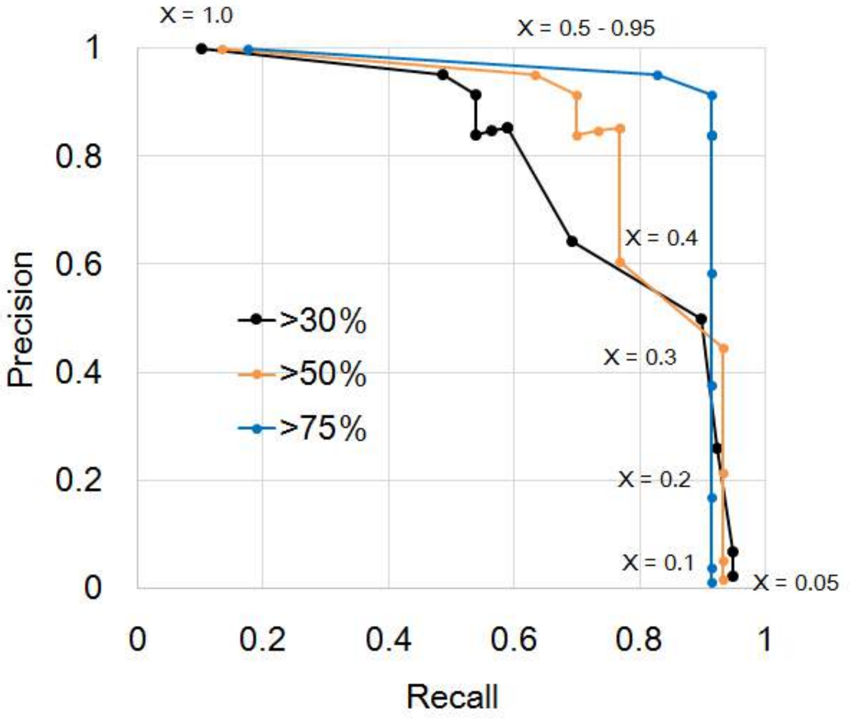
A A/T

A A
A A
A A

Step 1: Alignment of MinION reads to reference sequences.

Step 2: Constructing consensus in each position of the targets.
Only counting each base without mismatches in ±3 bp.
X: 0.1 – 1.0


[A + T +C + G] ≥ 100
n
Low Depth/Unknown
y
A / [A + T + C + G] ≥ X
A ≥ [T + C + G] × 2
n
T / [A +
T ≥
y
T / [A + T + C + G] ≥ X
T ≥ [C + G] × 2
y
n

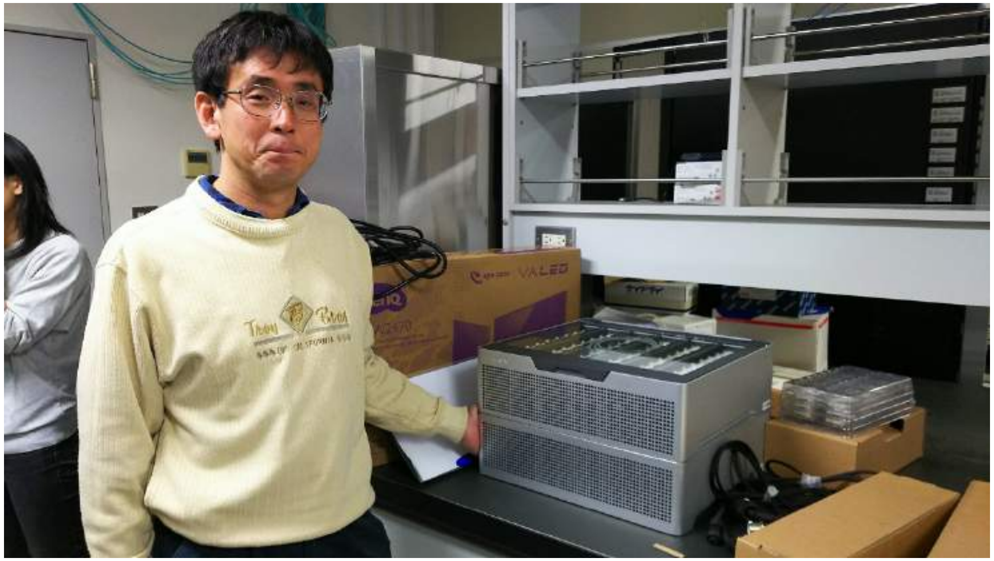


Detecting single nucleot
comparing with referenc

Suzuki et al DNA Re

Large Scale MinION sequencing “PromethION”

Ver>9 Flow Cell x 48枚; independen



3000 pores

120Gb/cell/run

120Gb x 48=~ 6Tb/per run->

>95% fidelity

$75,000; long waiting lis

Human WGS

De novo assemble…

LAD cell; H1975 for

5μg sta

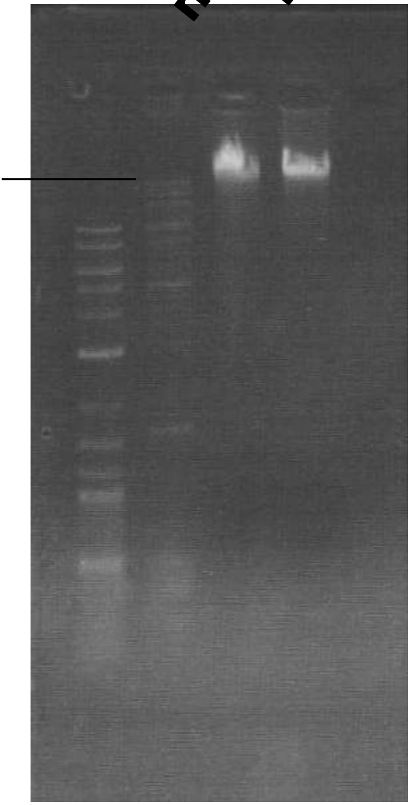
BluePippin

2log

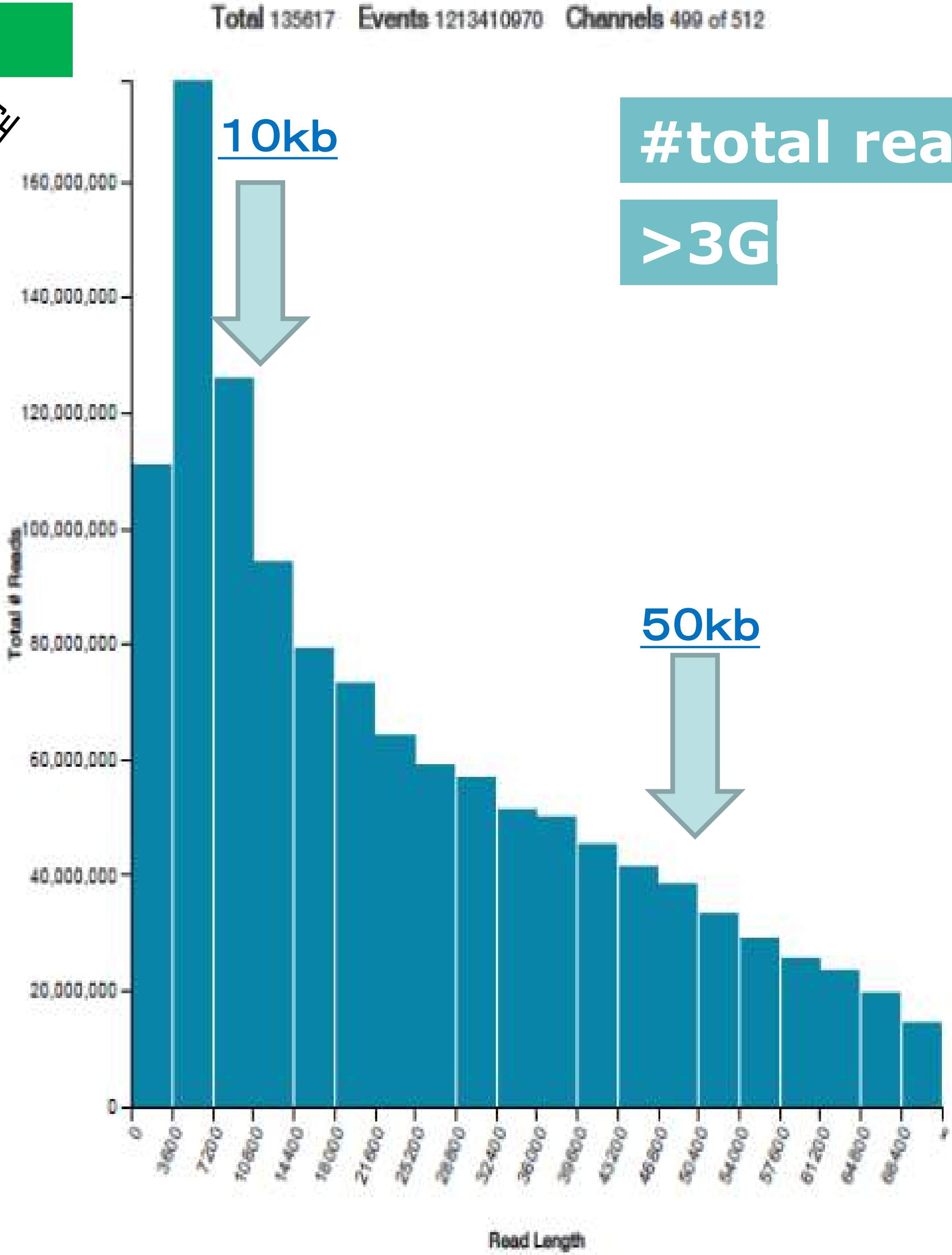
Extensio n

元 DNA

Nanopore鋳型



40kb



10kb

50kb

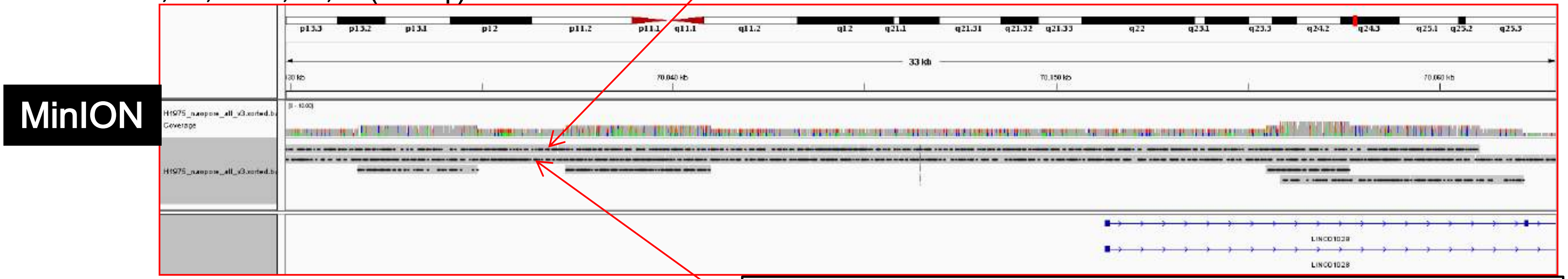
#total reads:

>3G

H1975
Around LINC01028 region (UCSC hg38)

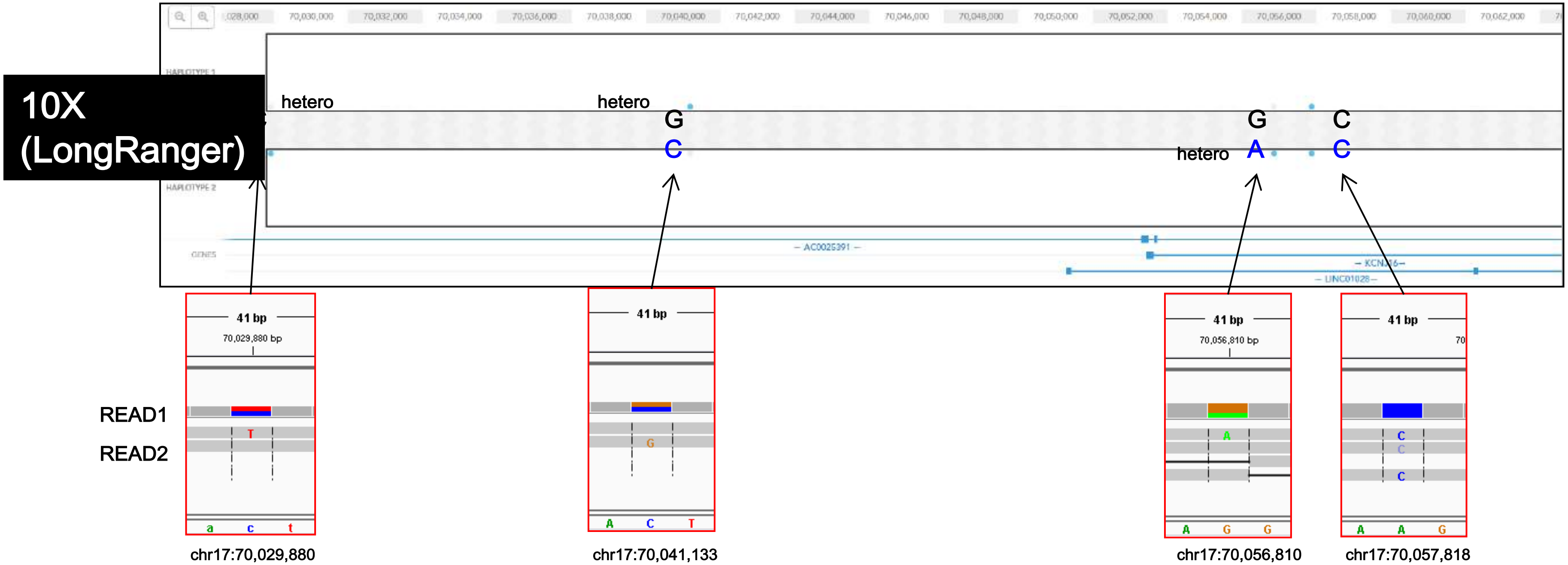
overlap
chr17:70,029,880-70,060,060(31.2kbp)

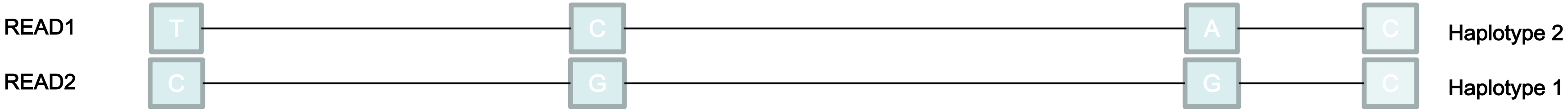
0de7d400-ad4d-4395-b5b6-af851096cd66_Basecall_2D_2d = read1
(length: 36,815 bp, alignment region: 38,297 bp)



6be71e8b-62e0-4a73-b6fd-e01d5f7b0460_Basecall_2D_2d = read2
(length: 36,262 bp, alignment region: 37,619 bp)

10x Phase Block chr17: 70,029,880-70,074,394 (38.3kbp)





Haplotype 1 can be phased by MinION Read2. Haplotype 2 can be phased by MinION Read1.

Sequencing using MinION as a convenient method for genotyping

PCR amplicon sequencing

Cancerous Mutations in Humans
(PCR cocktails of Driver genes in Lung cancers)

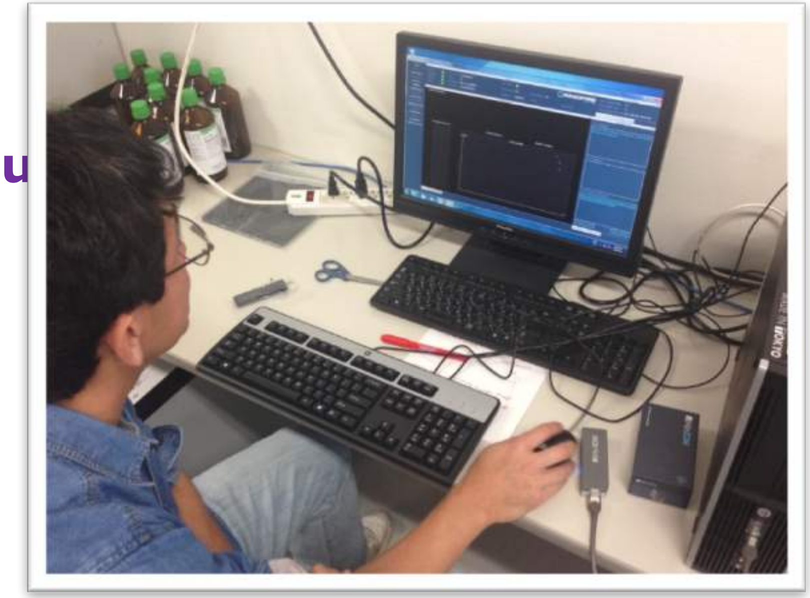
SNVs, short indels, exon skipping,
gene fusions and their "phasing"
(K-Ras, EGFR, NF1, ALK, RET etc)

SNPs in Pathogens of Tropical Diseases **2014-2016; Initial Eva**
(PCR cocktails of drug-resistant genes in malaria parasites)
(Chloroquine, Artemisinin, Sulfadoxin-resistant genes in Pf)

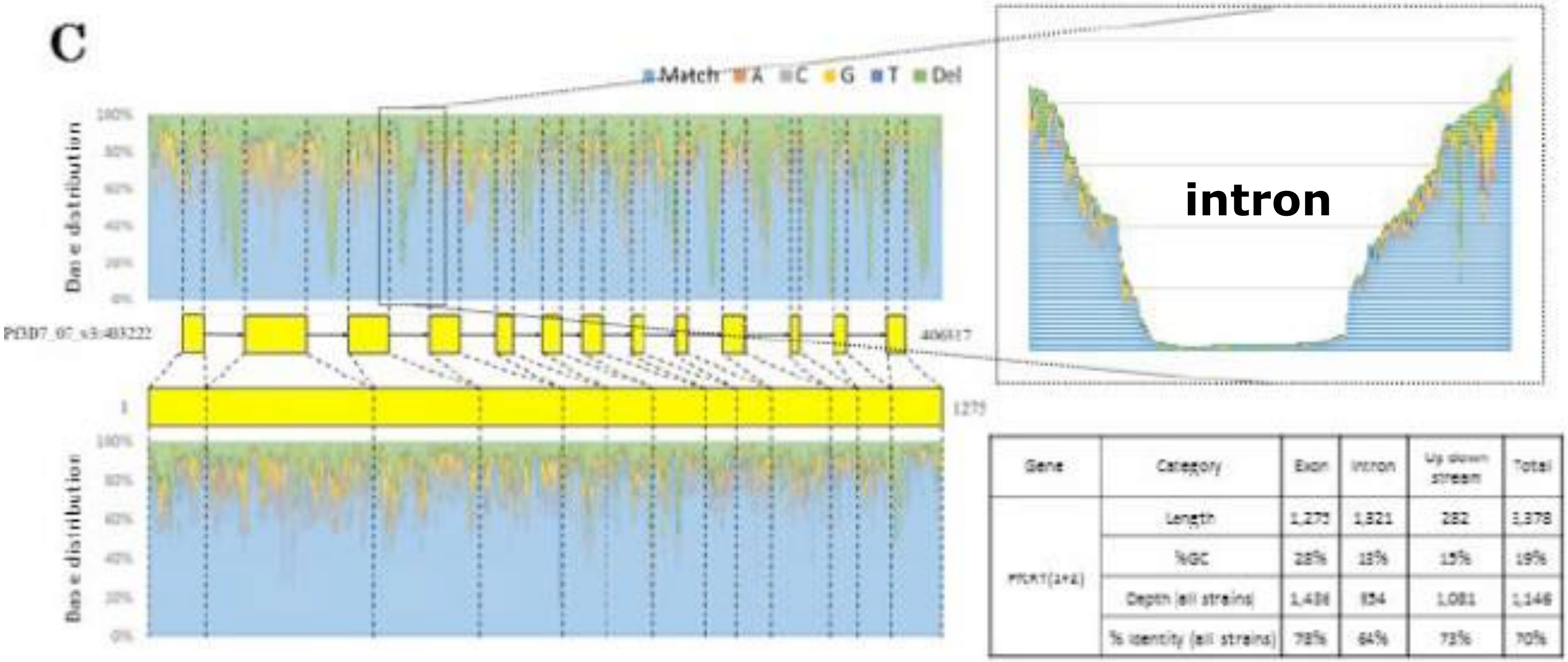
LAMP amplification

Pathogen detection

(Serotyping of Dengue Viru



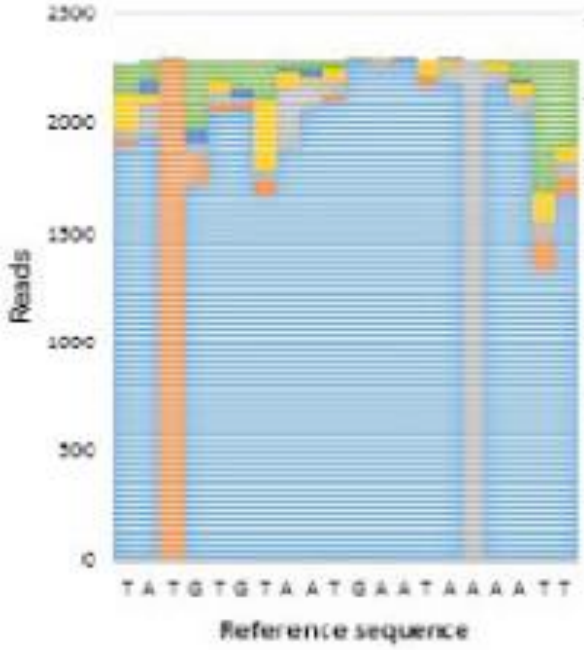
Pf CRT ge



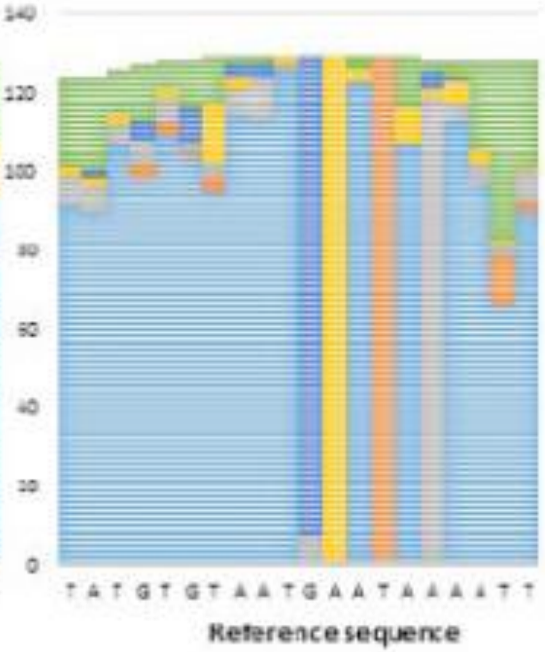
Gene	Category	Exon	Intron	Up down stream	Total
PfCRT(1+2)	Length	1,275	1,821	282	3,378
	%GC	28%	13%	15%	19%
	Depth (all strains)	1,438	854	1,081	1,146
	% identity (all strains)	78%	64%	73%	70%

Drug resistance mu

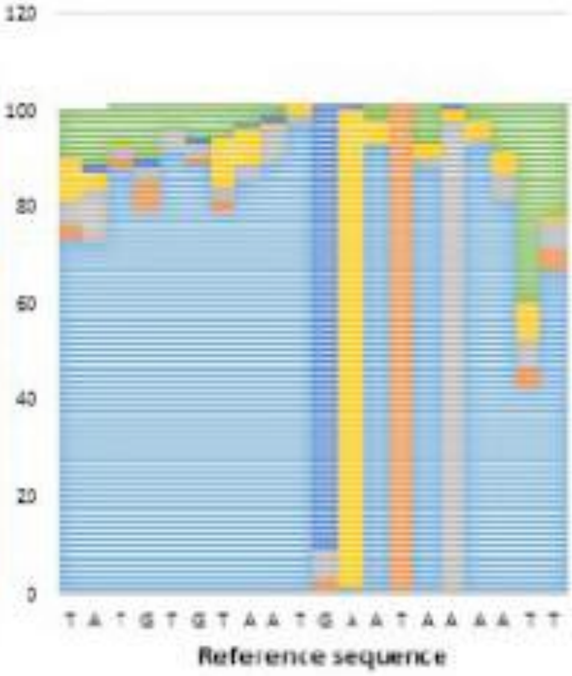
Sample#1



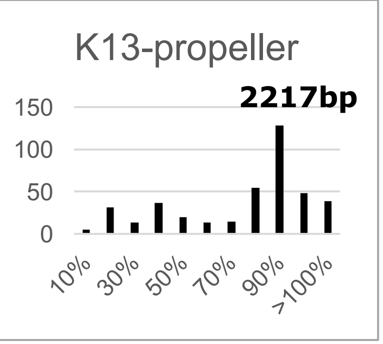
Sample#2

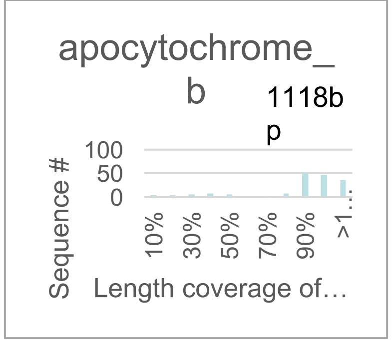


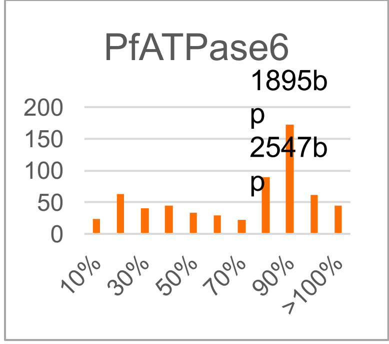
Sample#3

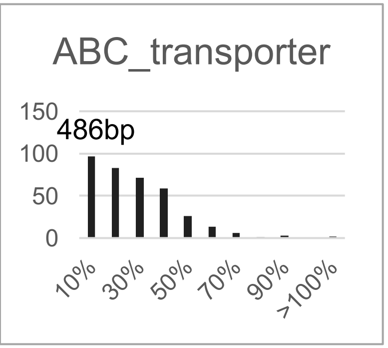


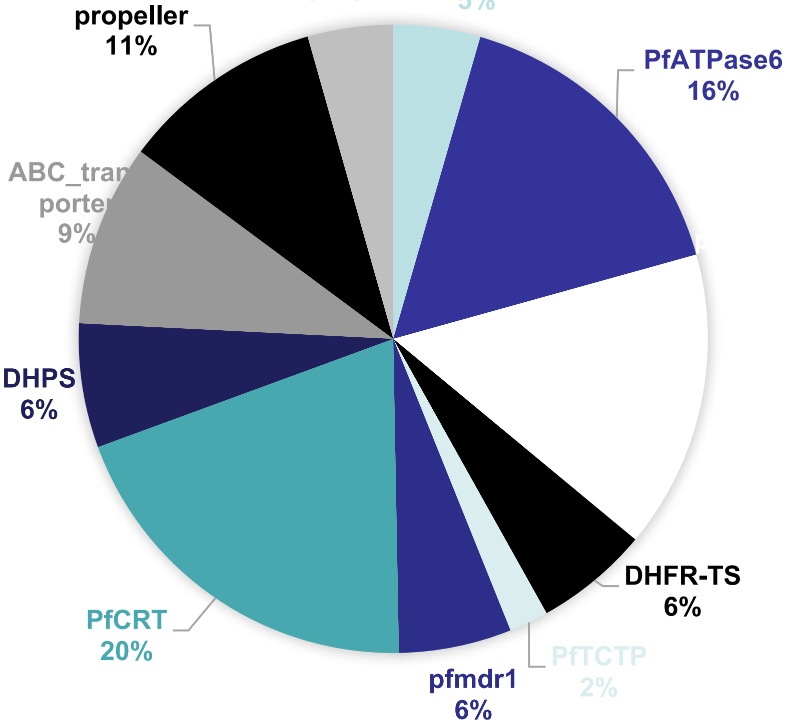
Amplicon coverage of the targets (the case of pateints

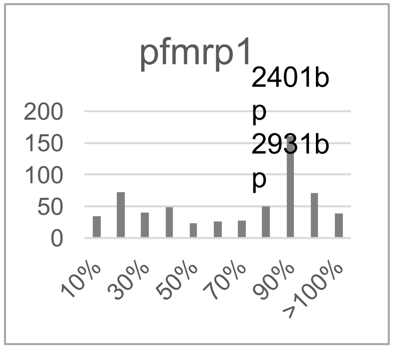


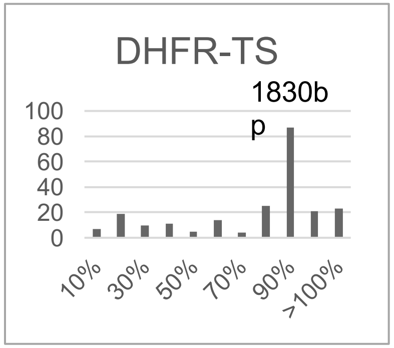


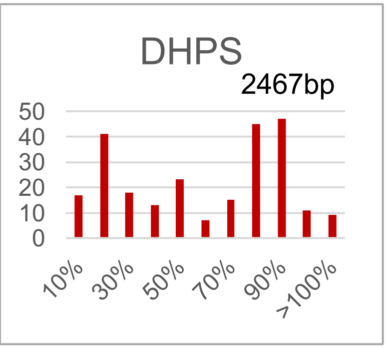


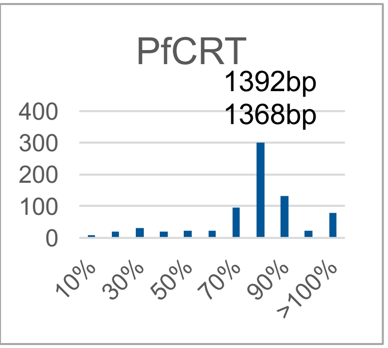


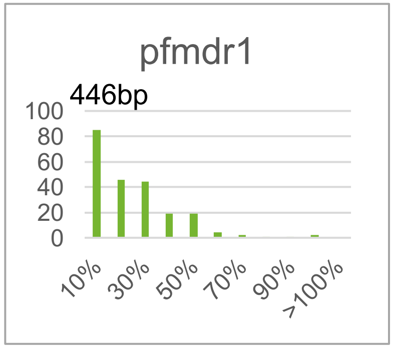


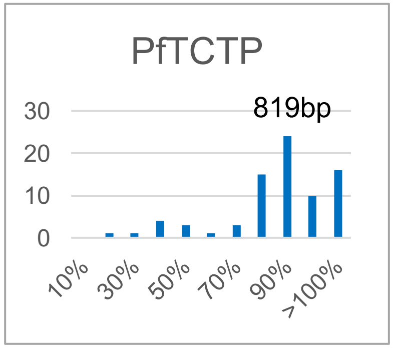










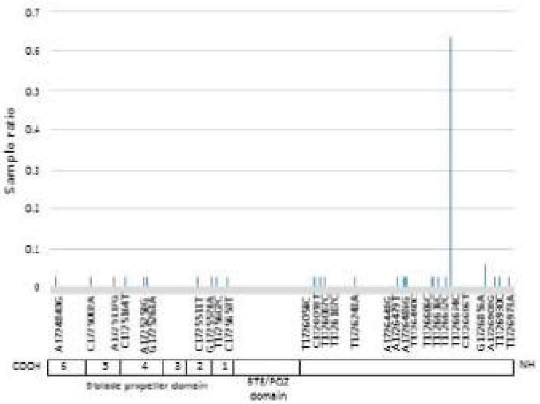


Inferred Drug Resi

Drugs	Patients									
	2-#2	2-#5	5-#1	5-#3	5-#6	5-#39	5-#40	5-#41	5-#43	5-#46
Atovaquone	-	-	-	-	-	-	-	-	-	-
Chloroquine/ mefloquine	+	+	+	+	+	+	+	+	+	+
SP	+	+	+	+	+	+	+	+	+	+
Artemisinin	-	-	?	?	?	?	?	-	?	?
ATPase6 mutations	-	-	+	?	+	+	+	-	+	+
K13-prop mutations	-	-	-	-	-	-	-	-	-	-

A novel SNV in the Pf K13 gene in M

A K13 gene (n=54)

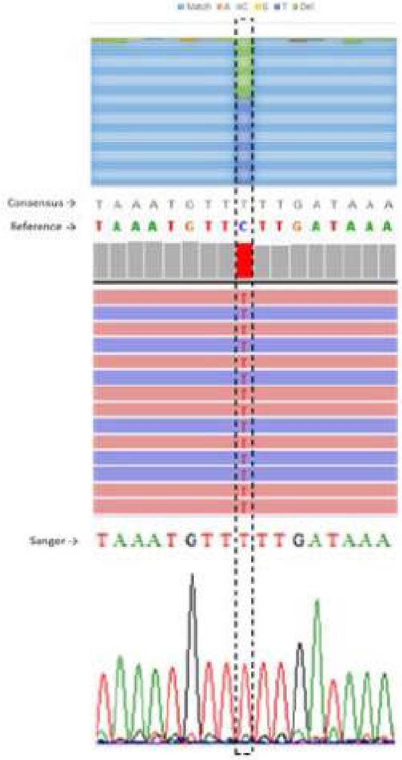


	Harboring a SNP(s) (%)	Reference type (%)	Typing failed (%)	Total (%)
Number of samples	33 (61%)	17 (31%)	4 (7%)	54 (100%)
Number of SNP positions	28	-	-	29

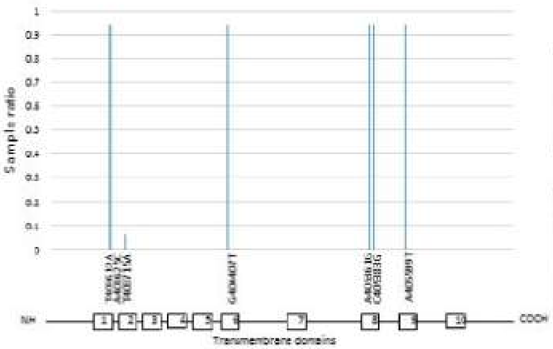
	C1726696T only (%)	C1726696T and other (%)	Other positions (%)	Reference type (%)	Total (%)
Number of samples	21 (42%)	1 (2%)	11 (22%)	17 (34%)	50 (100%)
Number of SNP positions	1	2	27	-	29

B

	Number of samples validated	Consistent	Inconsistent
C1726696T mutant allele	12	12	0
Other mutant alleles	9	8	1
Wildtype	17	17	0



C PfCRT (n=17)



	Harboring a SNP(s) (%)	Reference type (%)	Typing failed (%)	Total (%)
Number of samples	17 (100%)	0 (0%)	0 (0%)	17 (100%)
Number of SNP positions	7	-	-	7

	Haplotype 1 (%)	Haplotype 2 (%)	Total (%)
Number of samples	16 (94%)	1 (6%)	17 (100%)
Number of SNP positions	6	7	7

Entire gene region; starting from FTA (filter paper)

16S rRNA sequencing (full-length; PCR amplicon)

Clinical samples

1-1	A.baumanii
1-2	A.baumanii
1-3	A.baumanii
1-4	A.baumannii
1-5	A.baumannii
1-6	A.baumannii
2-1*	E.cloacae
2-2	E.cloacae
2-3	E.cloacae
2-4	E.cloacae
2-5	E.cloacae
2-6	E.cloacae
3-1*	E.coli
3-2	E.coli
3-3	E.coli
3-4	E.coli
3-5	E.coli
3-6	E.coli
4-1	E.faecium
4-2	E.faecium
4-3	E.faecium

9-1	N.gonorrhoeae
9-2	N.gonorrhoeae
9-3	N.gonorrhoeae
9-4	N.gonorrhoeae
9-5	N.gonorrhoeae
9-6	N.gonorrhoeae
10-1	P.aeruginosa
10-2	P.aeruginosa
10-3	P.aeruginosa
10-4	P.aeruginosa
10-5	P.aeruginosa
10-6	P.aeruginosa
11-1	P.mirabilis
11-2	P.mirabilis
11-3	P.mirabilis
11-4	P.mirabilis
11-5	P.mirabilis
11-6	P.mirabilis
12-1	S.agalactiae
12-2	S.agalactiae
12-3	S.agalactiae

⋮

Reference strains (for Error patterning/Squiggle matching

11	*Streptococcus*	*pneumoniae*
12	*Staphylococcus*	*aureus*
13	*Klebsiella*	*pneumoniae*
17	*Haemophilus*	*influenzae*
18	*Acinetobactor*	*baumanii*
19	*Legionella*	*pneumophila*
20	*Burkholderia*	*pseudomallei*
23	*Escherichia*	*coli*
24	*Pseudomonas*	*aeruginosa*
25	*Proteus*	*mirabilis*
26	*Mycoplasma*	*pneumonia*
27	*Streptococcus*	*mutans*
28	*Streptococcus*	*pyogenes*
29	*Streptococcus*	*agalactiae*
30	*Staphylococcus*	*aureus(MRSA)*
31	*Staphylococcus*	*epidermidis*
32	*Moraxella*	*catarrhalis*
33	*Corynebacterium*	*diphtheriae*
34	*Stenotrophomonas*	*maltophilia*
35	*Serratia*	*marcescens*
36	*Enterobacter*	*cloacae*

>645610 Enterobacter subsp. cloacae str. NCTC 9394 FP929040.1 complement(119488..121031) Gc01486_1of1
>gi|444439576|ref|NR_074891.1| Escherichia coli O157:H7 str. Sakai strain Sakai 16S ribosomal RNA, complete sequence

■ E.coli PCR product : 4/22 run
input reads : 500
E.cloacae map : 5 (1%)
E.coli map : 495 (99%)

■ E.cloacae PCR product : 5/10run
input reads : 500
E.cloacae map : 497 (99%)
E.coli map : 3 (1%)

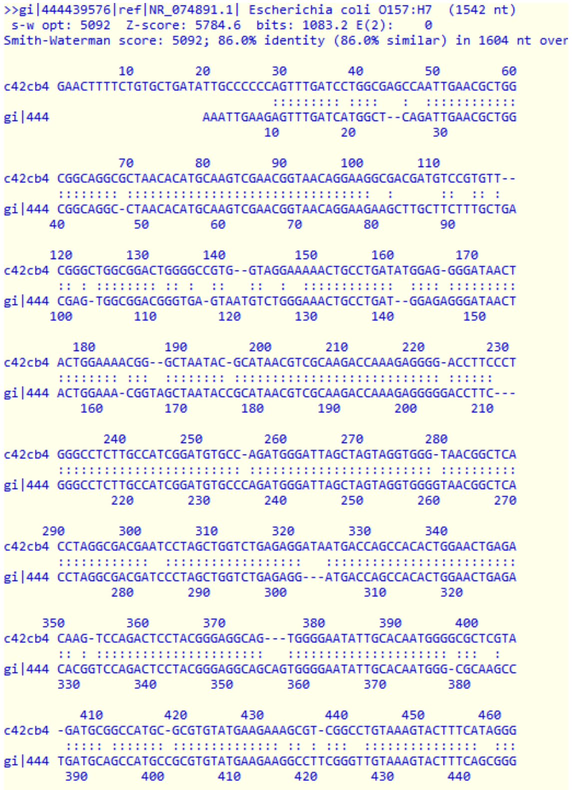
11件の詳細は下記の通りでした。

■ E.cloacae PCR product : 5/10run　について、

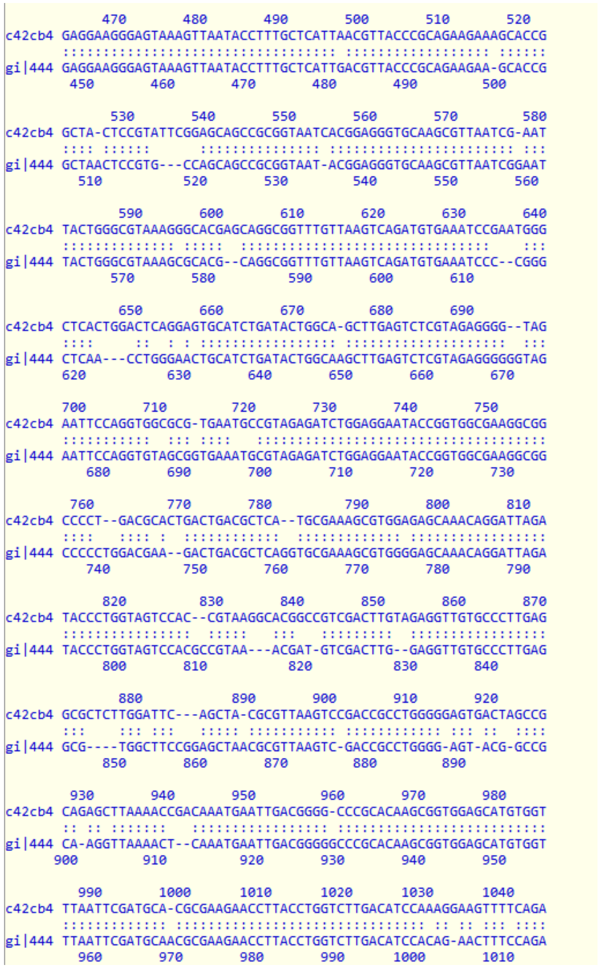
Case1：210269 Enterobacter cloacae str. AW3 EF446900.1 2..1357 Gi04667_1of1
input reads : 500
E.cloacae map : 5 (1%)
E.coli map : 495 (99%)

■ E.coli PCR product : 4/22 run

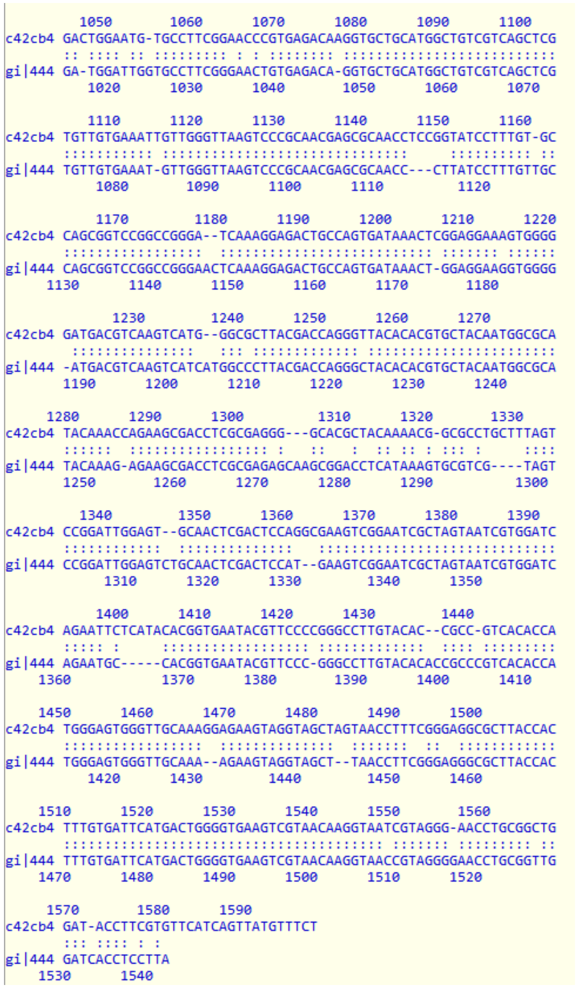
1,600 bp



```
>>gi|444439576|ref|NR_074891.1| Escherichia coli O157:H7  (1542 nt)
 s-w opt: 5092  Z-score: 5784.6  bits: 1083.2 E(2):    0
Smith-Waterman score: 5092; 86.0% identity (86.0% similar) in 1604 nt over

               10        20        30        40        50        60
c42cb4 GAACTTTTCTGTGCTGATATTGCCCCCCAGTTTGATCCTGGCGAGCCAATTGAACGCTGG
                                   ::::::::: ::::   :  ::::::::::::
gi|444                    AAATTGAAGAGTTTGATCATGGCT--CAGATTGAACGCTGG
                                  10        20          30

               70        80        90       100       110
c42cb4 CGGCAGGCGCTAACACATGCAAGTCGAACGGTAACAGGAAGGCGACGATGTCCGTGTT--
       :::::::: ::::::::::::::::::::::::::::::::  :      ::  :: :
gi|444 CGGCAGGC-CTAACACATGCAAGTCGAACGGTAACAGGAAGAAGCTTGCTTCTTTGCTGA
      40         50        60        70        80        90

      120       130       140         150       160        170
c42cb4 CGGGCTGGCGGACTGGGGCCGTG--GTAGGAAAAACTGCCTGATATGGAG-GGGATAACT
       :: : :::::::: :: :  ::   ::  :  ::::::::::::  :::: :::::::::
gi|444 CGAG-TGGCGGACGGGTGA-GTAATGTCTGGGAAACTGCCTGAT--GGAGAGGGATAACT
      100        110        120       130       140         150

         180         190        200       210       220        230
c42cb4 ACTGGAAAACGG--GCTAATAC-GCATAACGTCGCAAGACCAAAGAGGGG-ACCTTCCCT
       :::::::: :::  :::::::: ::::::::::::::::::::::::::: ::::::
gi|444 ACTGGAAA-CGGTAGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTC---
          160        170       180       190       200       210

             240       250        260       270        280
c42cb4 GGGCCTCTTGCCATCGGATGTGCC-AGATGGGATTAGCTAGTAGGTGGG-TAACGGCTCA
       :::::::::::::::::::::::: :::::::::::::::::::::::: ::::::::::
gi|444 GGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCA
              220       230       240       250       260       270

     290       300       310       320       330       340
c42cb4 CCTAGGCGACGAATCCTAGCTGGTCTGAGAGGATAATGACCAGCCACACTGGAACTGAGA
       ::::::::::::  ::::::::::::::::::   :::::::::::::::::::::::::
gi|444 CCTAGGCGACGATCCCTAGCTGGTCTGAGAGG---ATGACCAGCCACACTGGAACTGAGA
              280       290       300          310       320

     350        360       370          380       390       400
c42cb4 CAAG-TCCAGACTCCTACGGGAGGCAG---TGGGGAATATTGCACAATGGGGCGCTCGTA
       :: : ::::::::::::::::::::::   ::::::::::::::::::::: :::  :
gi|444 CACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGG-CGCAAGCC
       330       340       350       360       370        380

          410        420       430        440       450       460
c42cb4 -GATGCGGCCATGC-GCGTGTATGAAGAAAGCGT-CGGCCTGTAAAGTACTTTCATAGGG
        ::::: ::::::: :::::::::::::: :: : :::  :::::::::::::::  :::
gi|444 TGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGCGGG
        390       400       410       420       430       440
```



```
            470       480       490       500       510       520
c42cb4 GAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTAACGTTACCCGCAGAAGAAAGCACCG
       :::::::::::::::::::::::::::::::::: :::::::::::::::::: ::::::
gi|444 GAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAA-GCACCG
        450       460       470       480       490        500

             530       540       550       560       570        580
c42cb4 GCTA-CTCCGTATTCGGAGCAGCCGCGGTAATCACGGAGGGTGCAAGCGTTAATCG-AAT
       :::: ::::::      ::::::::::::::: ::::::::::::::::::::::: :::
gi|444 GCTAACTCCGTG---CCAGCAGCCGCGGTAAT-ACGGAGGGTGCAAGCGTTAATCGGAAT
         510          520       530        540       550       560

              590       600       610       620       630       640
c42cb4 TACTGGGCGTAAAGGGCACGAGCAGGCGGTTTGTTAAGTCAGATGTGAAATCCGAATGGG
       :::::::::::::: :::::  :::::::::::::::::::::::::::::::    :::
gi|444 TACTGGGCGTAAAGCGCACG--CAGGCGGTTTGTTAAGTCAGATGTGAAATCCC--CGGG
             570       580         590       600       610

              650       660       670        680       690
c42cb4 CTCACTGGACTCAGGAGTGCATCTGATACTGGCA-GCTTGAGTCTCGTAGAGGGG--TAG
       ::::     ::  : : ::::::::::::::::: ::::::::::::::::::::  :::
gi|444 CTCAA---CCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAG
       620          630       640       650       660       670

       700       710        720       730       740       750
c42cb4 AATTCCAGGTGGCGCG-TGAATGCCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGG
       :::::::::::  ::: ::::   ::::::::::::::::::::::::::::::::::::
gi|444 AATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGG
          680       690       700       710       720       730

        760         770       780         790       800       810
c42cb4 CCCCT--GACGCACTGACTGACGCTCA--TGCGAAAGCGTGGAGAGCAAACAGGATTAGA
       ::::   :::: :  ::::::::::::  ::::::::::::: :::::::::::::::::
gi|444 CCCCCTGGACGAA--GACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGA
          740         750       760       770       780       790

            820         830       840       850       860       870
c42cb4 TACCCTGGTAGTCCAC--CGTAAGGCACGGCCGTCGACTTGTAGAGGTTGTGCCCTTGAG
       ::::::::::::::::  :::::   :::   :::::::::  :::::::::::::::::
gi|444 TACCCTGGTAGTCCACGCCGTAA---ACGAT-GTCGACTTG--GAGGTTGTGCCCTTGAG
            800       810          820          830       840

              880           890       900       910       920
c42cb4 GCGCTCTTGGATTC---AGCTA-CGCGTTAAGTCCGACCGCCTGGGGGAGTGACTAGCCG
       :::    ::: :::   ::::: ::::::::::: :::::::::::: ::: ::  ::::
gi|444 GCG----TGGCTTCCGGAGCTAACGCGTTAAGTC-GACCGCCTGGGG-AGT-ACG-GCCG
            850       860       870        880        890

        930       940       950        960       970       980
c42cb4 CAGAGCTTAAAACCGACAAATGAATTGACGGGG-CCCGCACAAGCGGTGGAGCATGTGGT
       :: :: :::::::   ::::::::::::::::: ::::::::::::::::::::::::::
gi|444 CA-AGGTTAAAACT--CAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGT
      900        910         920       930       940       950

         990       1000      1010      1020      1030      1040
c42cb4 TTAATTCGATGCA-CGCGAAGAACCTTACCTGGTCTTGACATCCAAAGGAAGTTTTCAGA
       ::::::::::::: ::::::::::::::::::::::::::::::: :: :: ::: ::::
gi|444 TTAATTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCACAG-AACTTTCCAGA
         960       970       980       990      1000       1010
```



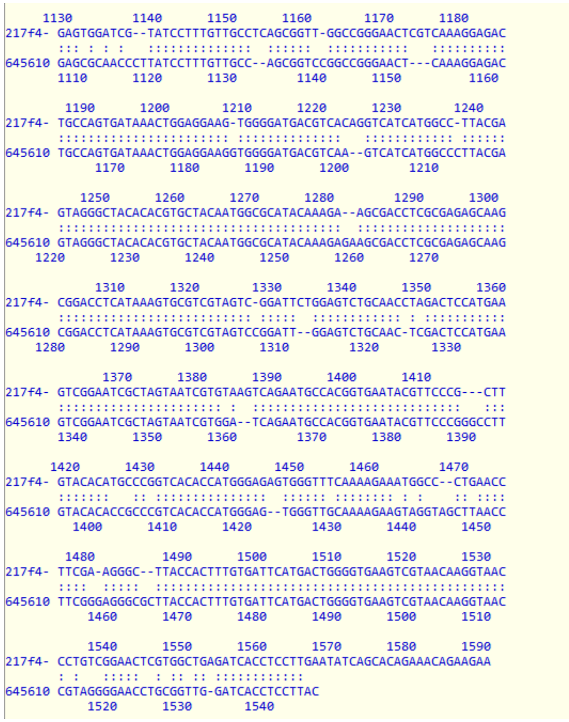
```
         1050       1060      1070      1080      1090      1100
c42cb4 GACTGGAATG-TGCCTTCGGAACCCGTGAGACAAGGTGCTGCATGGCTGTCGTCAGCTCG
       :: :::: :: ::::::::: : : :::::::: ::::::::::::::::::::::::::
gi|444 GA-TGGATTGGTGCCTTCGGGAACTGTGAGACA-GGTGCTGCATGGCTGTCGTCAGCTCG
          1020      1030      1040       1050      1060      1070

          1110      1120      1130      1140      1150      1160
c42cb4 TGTTGTGAAATTGTTGGGTTAAGTCCCGCAACGAGCGCAACCTCCGGTATCCTTTGT-GC
       ::::::::::: ::::::::::::::::::::::::::::::     :::::::::: ::
gi|444 TGTTGTGAAAT-GTTGGGTTAAGTCCCGCAACGAGCGCAACC---CTTATCCTTTGTTGC
           1080       1090      1100      1110         1120

           1170        1180      1190      1200      1210      1220
c42cb4 CAGCGGTCCGGCCGGGA--TCAAAGGAGACTGCCAGTGATAAACTCGGAGGAAAGTGGGG
       :::::::::::::::::  :::::::::::::::::::::::::: ::::::: ::::::
gi|444 CAGCGGTCCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACT-GGAGGAAGGTGGGG
     1130      1140      1150      1160      1170       1180

             1230        1240      1250      1260      1270
c42cb4 GATGACGTCAAGTCATG--GGCGCTTACGACCAGGGTTACACACGTGCTACAATGGCGCA
        :::::::::::::::   ::: ::::::::::::: :::::::::::::::::::::::
gi|444 -ATGACGTCAAGTCATCATGGCCCTTACGACCAGGGCTACACACGTGCTACAATGGCGCA
       1190      1200      1210      1220      1230      1240

     1280      1290      1300         1310      1320       1330
c42cb4 TACAAACCAGAAGCGACCTCGCGAGGG---GCACGCTACAAAACG-GCGCCTGCTTTAGT
       ::::::  ::::::::::::::::: :   ::   :  :: :: : ::: :     ::::
gi|444 TACAAAG-AGAAGCGACCTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCG----TAGT
       1250       1260      1270      1280      1290          1300

         1340        1350      1360      1370      1380      1390
c42cb4 CCGGATTGGAGT--GCAACTCGACTCCAGGCGAAGTCGGAATCGCTAGTAATCGTGGATC
       ::::::::::::  ::::::::::::::   :::::::::::::::::::::::::::::
gi|444 CCGGATTGGAGTCTGCAACTCGACTCCAT--GAAGTCGGAATCGCTAGTAATCGTGGATC
            1310      1320      1330        1340      1350

           1400      1410      1420      1430        1440
c42cb4 AGAATTCTCATACACGGTGAATACGTTCCCCGGGCCTTGTACAC--CGCC-GTCACACCA
       ::::: :     :::::::::::::::::: :::::::::::::  :::: :::::::::
gi|444 AGAATGC-----CACGGTGAATACGTTCCC-GGGCCTTGTACACACCGCCCGTCACACCA
    1360           1370      1380       1390      1400      1410

    1450      1460      1470      1480      1490      1500
c42cb4 TGGGAGTGGGTTGCAAAGGAGAAGTAGGTAGCTAGTAACCTTTCGGGAGGCGCTTACCAC
       :::::::::::::::::  ::::::::::::::  :::::::  ::  ::::::::::::
gi|444 TGGGAGTGGGTTGCAAA--AGAAGTAGGTAGCT--TAACCTTCGGGAGGGCGCTTACCAC
          1420      1430        1440        1450      1460

    1510      1520      1530      1540      1550       1560
c42cb4 TTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAATCGTAGGG-AACCTGCGGCTG
       ::::::::::::::::::::::::::::::::::::::: ::::::: ::::::::: ::
gi|444 TTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTG
    1470      1480      1490      1500      1510      1520

     1570       1580      1590
c42cb4 GAT-ACCTTCGTGTTCATCAGTTATGTTTCT
       ::: :::: : :
gi|444 GATCACCTCCTTA
    1530      1540
```

■ E.cloacae PCR product : 5/10run

1,600 bp

```
>>645610 Enterobacter subsp. cloacae str. NCTC 9394 FP92  (1546 nt)
rev-comp s-w opt: 5532  Z-score: 6720.0  bits: 1256.2 E(2):    0
Smith-Waterman score: 5532; 88.4% identity (88.4% similar) in 1563 nt ove

            20        30        40        50          60        70
217f4- ACTGCTAGTTGTACTGGCCTGAGATTGTGAACTGATCGCT--CTACAGGCCTATAACACA
                                     : ::: ::::  :  :::::::: : ::::
645610  TTAAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTA-ANCACA
                10        20        30        40        50

              80        90       100       110       120       130
217f4- GTGGCAAGTCGAACGGGCAACAGGAAGCAGCTTGCTGCTTCGCTGACGAGTGGCGAGACG
          :::::::::::::  :  ::::::::::::::::::: :::::::::::::: ::::
645610 T--GCAAGTCGAACGGTAACNAGGAAGCAGCTTGCTGCTTTGCTGACGAGTGGCG-GACG
         60        70        80        90       100       110

             140        150       160       170       180
217f4- GGTGAGTAATGTCTGG-AAACTGCTTTAGTGGATGGAGGGGGATAACTACTGGAAACG-T
       :::::::::::::::: ::::::: :     ::::::::::::::::::::::::::: :
645610 GGTGAGTAATGTCTGGGAAACTGCCT-----GATGGAGGGGGATAACTACTGGAAACGGT
         120       130       140            150       160       170

     190       200       210       220        230       240
217f4- AGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGAC-TTCGGGCCTCTTGCCATCG
       :::::::::::::::::::::::::::::::::::::::: :::::::::::::::::::
645610 AGCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCG
              180       190       200       210       220       230

      250       260       270       280          290       300
217f4- TATCTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAACGG---AGCTACGGCGACCATTC
        :: ::::::::::::::::::::::::::::::::::::   : ::: :::::: :: :
645610 GATGTGCCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTA-GGCGACGATCC
              240       250       260       270        280

         310        320       330       340        350       360
217f4- CCCGCTGGTCTC-GAGGATGACCAGCCACACTGGAA-TGAGACACGGTCCAGACTCCTAC
       :  ::::::::  ::::::::::::::::::::::: :::::::::::::::::::::::
645610 CTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGAGACACGGTCCAGACTCCTAC
     290       300       310       320       330       340

           370       380       390        400       410       420
217f4- GGGAGGCAGCAGTGGGGAATACTTGCACAATGG-CGCAAGCCTGATGCAGCCATGCCGTG
       ::::::::::::::::::::: ::::::::::: :::::::::::::::::::::::: :
645610 GGGAGGCAGCAGTGGGGAATA-TTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCG
     350       360       370        380       390       400

            430       440       450        460       470       480
217f4- TATATGAAGAAGAAGCCCCTTCGGGTTGTAAA-TACTTTCAGCGGGGAGGAAGGCGATAA
       : ::::::::::     ::::::::::::::: :::::::::::::::::::::::::::
645610 TGTATGAAGAAGG----CCTTCGGGTTGTAAAGTACTTTCAGCGGGGAGGAAGGCGATAA
      410       420           430       440       450       460

             490       500       510        520       530       540
217f4- GGTTAATAACCTTGTCGATTGACGTTACC-GCAGAAAAGAAGCACCCCGGCTAACTCCGT
       ::::::::::::::::::::::::::::: ::::::  ::::::::  ::::::::::::
645610 GGTTAATAACCTTGTCGATTGACGTTACCCGCAGAA--GAAGCACC--GGCTAACTCCGT
          470       480       490       500           510       520
```

```
              550       560       570       580       590       600
217f4- GCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTGGACT
       :::::::::::::::::::::::::::::::::::::::::::::::::::::::  :
645610 GCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGC
              530       540       550       560       570       580

              610       620       630       640       650
217f4- AGCACGCAGGCGGTCTGTCAAGTCGGATGTGGAACCCCCTATGGGCTCATCCTGGGAA-T
        :::::::::::::::::::::::::::::: :: ::::   ::::::: :::::::: :
645610 -GCACGCAGGCGGTCTGTCAAGTCGGATGTGAAATCCCC---GGGCTCAACCTGGGAACT
               590       600       610          620       630

     660        670       680       690        700          710
217f4- GCATTCGAA-CTGGCAGGCTAGAGTCTTGTAGAGGGGG-TAGAATTCCATG---AGGTGT
       ::::::::: :::::::::::::::::::::::::::: :::::::::: :   ::  ::
645610 GCATTCGAAACTGGCAGGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAGGTGTAGCGGT
        640       650       660       670       680       690

          720       730       740       750       760       770
217f4- GAAATGCGGTGAGATCTCGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACT
       ::::::::  ::::::: ::::::::::::::::::::::::::::::::::::::::::
645610 GAAATGCGTAGAGATCT-GGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAAAGACT
        700       710        720       730       740       750

          780       790       800       810        820       830
217f4- GACGCTCAGGTGCCGAACCGTGGGGAGCAAACAGGATC-GATACCCTGGTAGTCCACGCC
       :::::::::::::  :: :::::::::::::::::::  :::::::::::::::::::::
645610 GACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCC
         760       770       780       790       800       810

            840       850       860       870       880       890
217f4- GTAA-CGATGTCGACTTTGTGAGAGGTTGTGCCCTGTGAGGCGTGGCTTCCGGCTACTAC
       :::: :::::::::::: :    :::::::::::: :::::::::::::::::     ::
645610 GTAAACGATGTCGACTTGG----AGGTTGTGCCCT-TGAGGCGTGGCTTCCGGAGCTAAC
         820       830           840        850       860       870

             900       910       920       930       940       950
217f4- GCGT-AAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAACTGCAAATGAATTGACGG
       :::: :::::::::::::::::::::::::::::::::::::   :::::::::::::::
645610 GCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGG
              880       890       900       910       920       930

             960       970       980       990      1000      1010
217f4- GGCCGCAGGGTAACCGGTGGAGCATGTGGTTTAATTTCGATGCCAACGCGAAGAACCTTA
       :: : : :   :: :::::::::::::::::::::: :::::: ::::::::::::::::
645610 GGGC-CCGCACAAGCGGTGGAGCATGTGGTTTAATT-CGATGC-AACGCGAAGAACCTTA
               940       950       960        970        980

            1020      1030      1040      1050      1060      1070
217f4- CCTACTCTTGACATCCAGAGGACTTTCCAGAGTCGATTGGTGCCTTCGAGAAACTCTGAG
       :::::::::::::::::::: ::::  :::::  : :::::::::::: ::: :::::::
645610 CCTACTCTTGACATCCAGAGAACTTAGCAGAGATGCTTGGTGCCTTCGGGAA-CTCTGAG
       990      1000      1010      1020      1030       1040

            1080      1090      1100         1110      1120
217f4- ACAGGTGCTGCATCGCTTATGTCAGCTCGT--TGTGAA-TGTTGGGTTAAGTCCCGCAAC
       ::::::::::::: :::   ::::::::::  :::::: :::::::::::::::::::::
645610 ACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGTCCCGCAAC
       1050      1060      1070      1080      1090      1100
```



```
     1130        1140      1150      1160       1170      1180
217f4- GAGTGGATCG--TATCCTTTGTTGCCTCAGCGGTT-GGCCGGGAACTCGTCAAAGGAGAC
       ::: : : :   ::::::::::::::  ::::::  :::::::::::   ::::::::::
645610 GAGCGCAACCCTTATCCTTTGTTGCC--AGCGGTCCGGCCGGGAACT---CAAAGGAGAC
       1110      1120      1130        1140      1150         1160

        1190      1200       1210      1220      1230       1240
217f4- TGCCAGTGATAAACTGGAGGAAG-TGGGGATGACGTCACAGGTCATCATGGCC-TTACGA
       ::::::::::::::::::::::: ::::::::::::::   :::::::::::: ::::::
645610 TGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAA--GTCATCATGGCCCTTACGA
            1170      1180      1190      1200        1210

          1250      1260      1270      1280        1290      1300
217f4- GTAGGGCTACACACGTGCTACAATGGCGCATACAAAGA--AGCGACCTCGCGAGAGCAAG
       ::::::::::::::::::::::::::::::::::::::  ::::::::::::::::::::
645610 GTAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCGAGAGCAAG
    1220      1230      1240      1250      1260      1270

            1310      1320       1330      1340      1350      1360
217f4- CGGACCTCATAAAGTGCGTCGTAGTC-GGATTCTGGAGTCTGCAACCTAGACTCCATGAA
       :::::::::::::::::::::::::: :::::  :::::::::::: : :::::::::::
645610 CGGACCTCATAAAGTGCGTCGTAGTCCGGATT--GGAGTCTGCAAC-TCGACTCCATGAA
    1280      1290      1300      1310        1320       1330

             1370      1380      1390      1400      1410
217f4- GTCGGAATCGCTAGTAATCGTGTAAGTCAGAATGCCACGGTGAATACGTTCCCG---CTT
       :::::::::::::::::::::: :  ::::::::::::::::::::::::::::   :::
645610 GTCGGAATCGCTAGTAATCGTGGA--TCAGAATGCCACGGTGAATACGTTCCCGGGCCTT
       1340      1350      1360        1370      1380      1390

      1420      1430      1440      1450      1460        1470
217f4- GTACACATGCCCGGTCACACCATGGGAGAGTGGGTTTCAAAAGAAATGGCC--CTGAACC
       :::::::   :: :::::::::::::::  :::::: :::::::: : :    :: ::::
645610 GTACACACCGCCCGTCACACCATGGGAG--TGGGTTGCAAAAGAAGTAGGTAGCTTAACC
         1400      1410      1420        1430      1440      1450

        1480         1490      1500      1510      1520      1530
217f4- TTCGA-AGGGC--TTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAAC
       ::::  :::::  :::::::::::::::::::::::::::::::::::::::::::::::
645610 TTCGGGAGGGCGCTTACCACTTTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAAC
           1460      1470      1480      1490      1500      1510

           1540      1550      1560      1570      1580      1590
217f4- CCTGTCGGAACTCGTGGCTGAGATCACCTCCTTGAATATCAGCACAGAAACAGAAGAA
       : :   :::::  : :: :: ::::::::::::
645610 CGTAGGGGAACCTGCGGTTG-GATCACCTCCTTAC
           1520      1530       1540
```

Reaction Scheme (total reaction

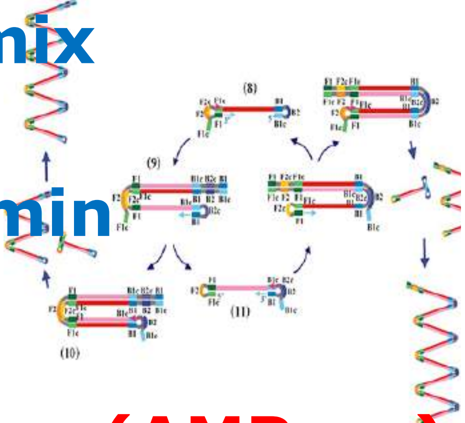
A

LAMP Amplification

Serum (1-5 microL)

Mix with Dry Lamp reagent mix

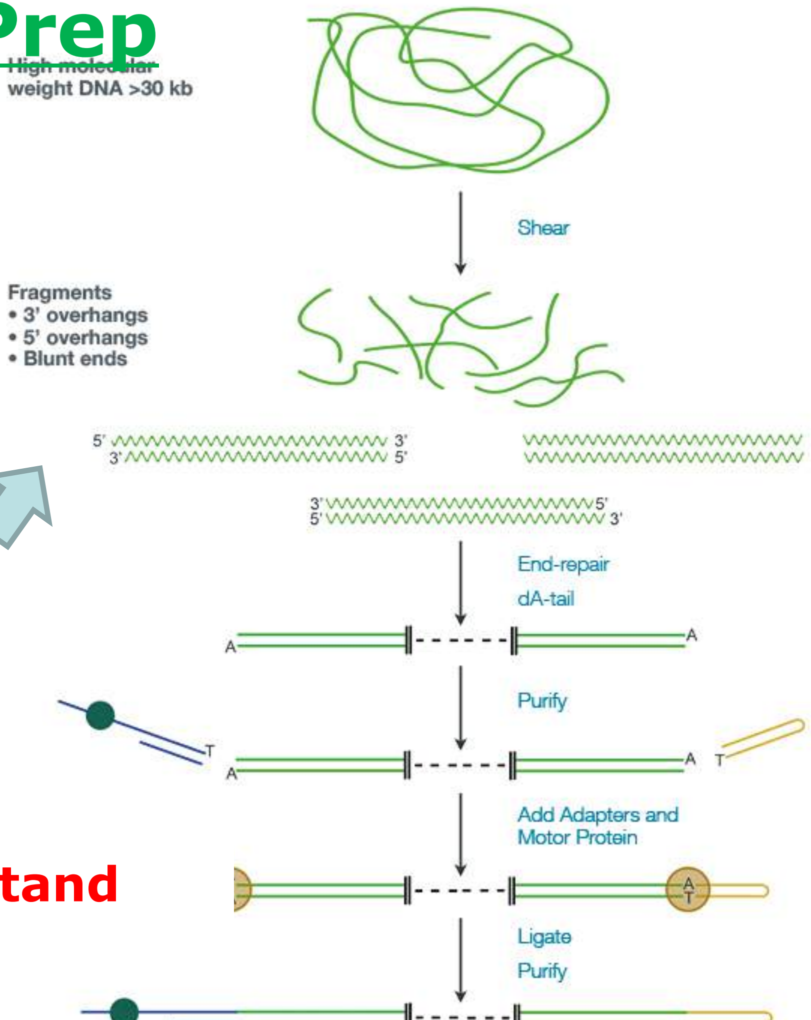
65oC 60 min



Purification (AMPure)

Needing only a magnetic stand

Nanopore Sample Prep



Purification

Yamagishi et al Sci Rep 2017

Summary of the results from the Control templates (D1- D4; with the fixed parameters of LAST)

Templates	**#of total reads (2D)**	**#reads (mapped to the corresponding serotype (score=150, mismap=1))**				
		D1	D2	D3	D4	%
D1	3733	1025	1	3	1	100%
D2	9260	18	6252	23	219	96%
D3	9750	8	4	7707	0	100%
D4	2758	1	0	7	1923	100%

Precise detection of serotypes wer possible

Use of MinION in Indonesia (U

A clinic in a town area

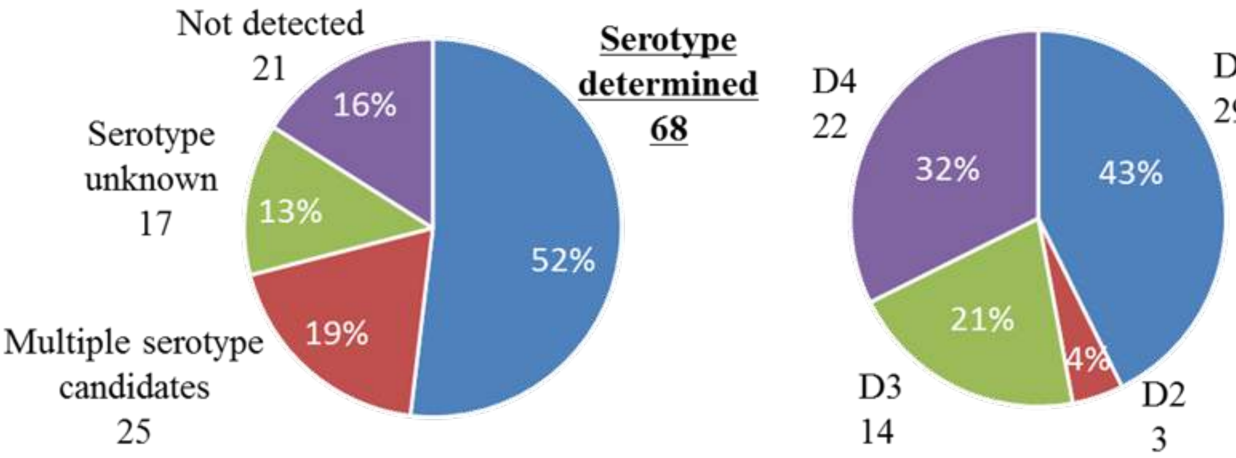
A clinic in a countryside

A clinician in operation

A storage for the reage

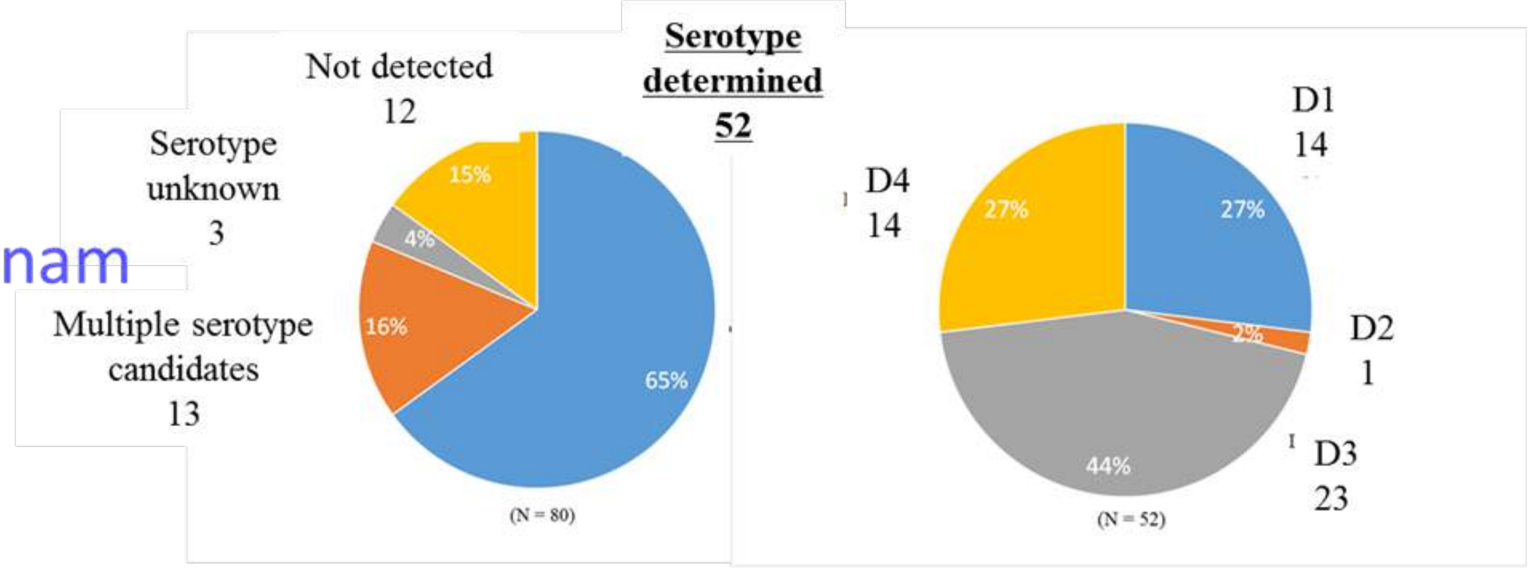
Large-scale serotyping of DENVs in South Ea

Manado, Indonesia

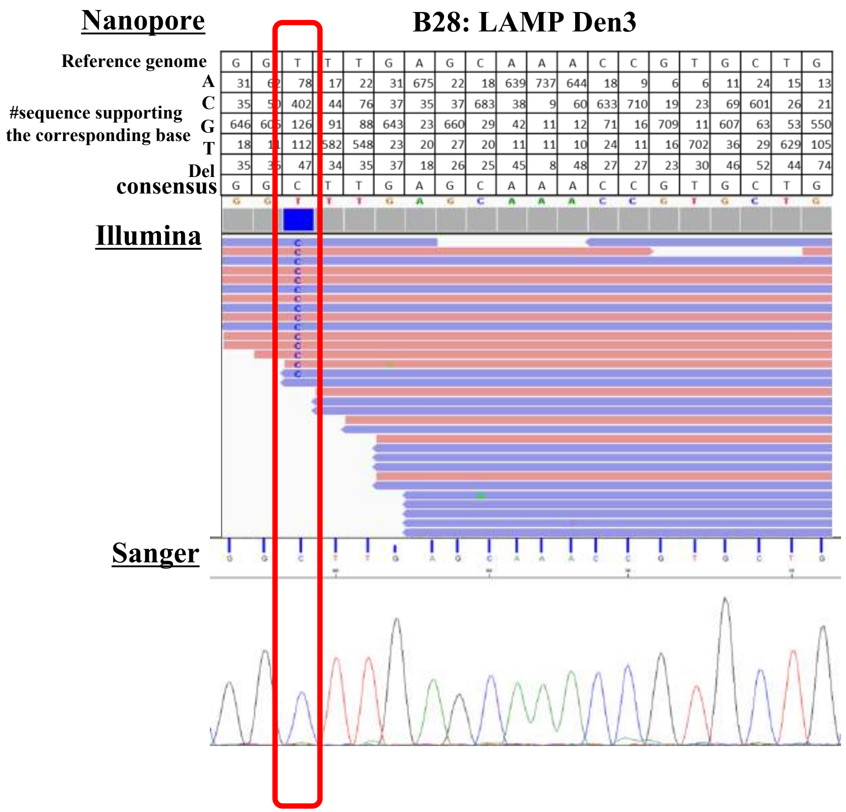


Population of serotypes (n= 131)

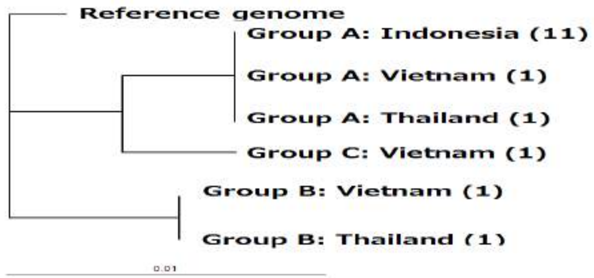
Hanoi, Vietnam

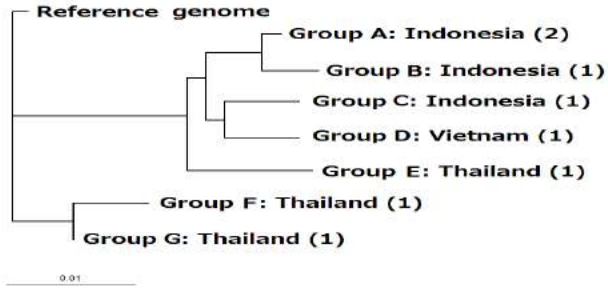


Population of serotypes (n= 80)

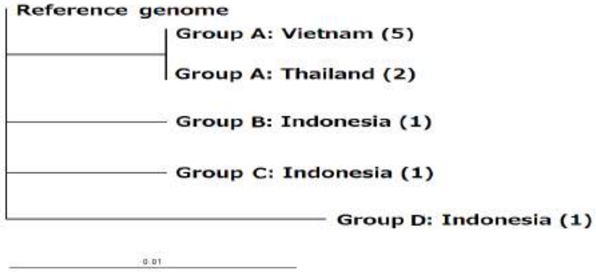


Reference genome	G	G	T	T	T	G	A	G	C	A	A	A	C	C	G	T	G	C	T	G
#sequence supporting the corresponding base A	31	62	78	17	22	31	675	22	18	639	737	644	18	9	6	6	11	24	15	13
C	35	50	402	44	76	37	35	37	683	38	9	60	633	710	19	23	69	601	26	21
G	646	605	126	91	88	643	23	660	29	42	11	12	71	16	709	11	607	63	53	550
T	18	11	112	582	548	23	20	27	20	11	11	10	24	11	16	702	36	29	629	105
Del	35	36	47	34	35	37	18	26	25	45	8	48	27	27	23	30	46	52	44	74
consensus	G	G	C	T	T	G	A	G	C	A	A	A	C	C	G	T	G	C	T	G



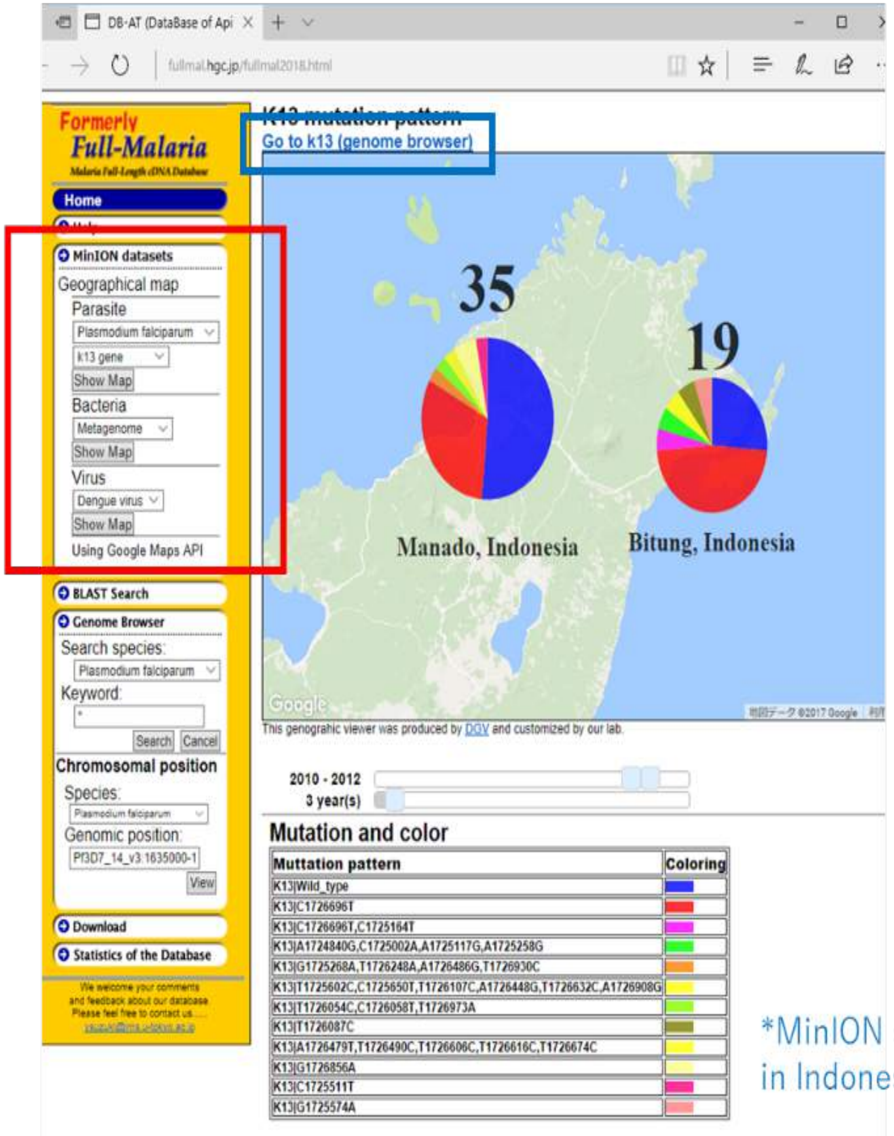


	DENV1	DENV2	DENV3	DENV4
Indonesia	2 (19)	5 (20)	5 (23)	4 (4)
Vietnam	2 (4)	4 (4)	1 (13)	1 (5)
Thailand	2 (2)	6 (7)	1 (1)	1 (2)



PCR or LAMP-based Data

Screen shot of the database

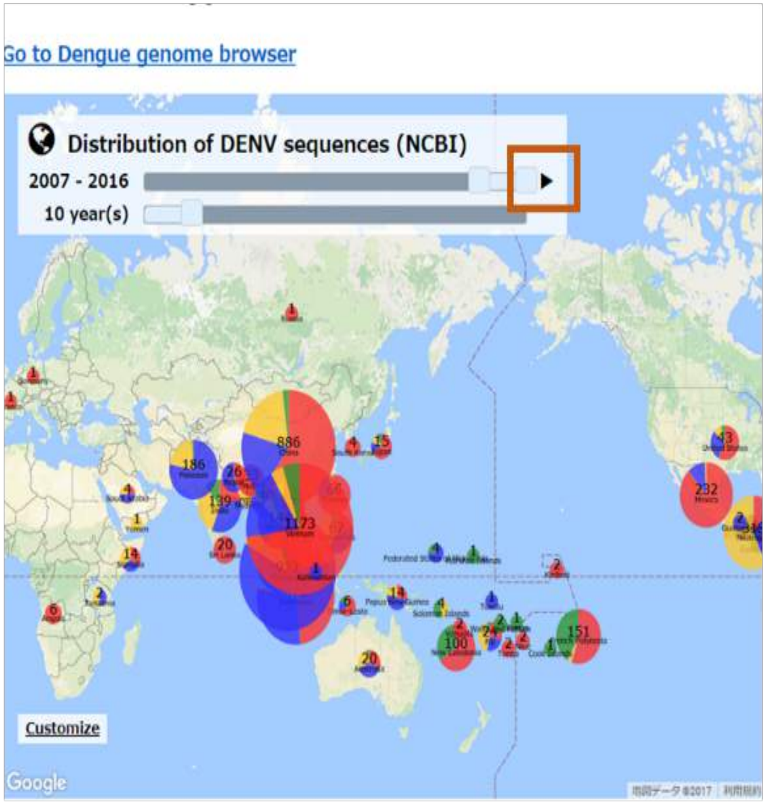


We welcome your comments and feedback about our database. Please feel free to contact us.

This genographic viewer was produced by DGV and customized by our lab.

Mutation and color

Muttation pattern	Coloring
K13|Wild_type	🟦
K13|C1726696T	🟥
K13|C1726696T,C1725164T	🟪
K13|A1724840G,C1725002A,A1725117G,A1725258G	🟩
K13|G1725268A,T1726248A,A1726486G,T1726930C	🟧
K13|T1725602C,C1725650T,T1726107C,A1726448G,T1726632C,A1726908G	🟨
K13|T1726054C,C1726058T,T1726973A	🟩
K13|T1726087C	🟫
K13|A1726479T,T1726490C,T1726606C,T1726616C,T1726674C	🟨
K13|G1726856A	🟨
K13|C1725511T	🟥
K13|G1725574A	🟥



*Not all dengue samples are from MinION but from Illumina sequencing

*MinION sequencing of Pf in Indonesia

http://fullmal.hgc.jp

Developed based on
DGV: Dengue Genographic Vi
Publication: *Front microbiol.*
doi:10.3389/fmicb.2016.008

ARTICLE

doi:10.1038/nature22040

Virus genomes reveal factors that spread and sustained the Ebola epidemic

A list of authors and their affiliations appears at the end of the paper

Video

The 2013–2016 West African epidemic caused by the Ebola virus was of unprecedented magnitude, duration and impact. Here we reconstruct the dispersal, proliferation and decline of Ebola virus throughout the region by analysing 1,610 Ebola virus genomes, which represent over 5% of the known cases. We test the association of geography, climate and demography with viral movement among administrative regions, inferring a classic 'gravity' model, with intense dispersal between larger and closer populations. Despite attenuation of international dispersal after border closures, cross-border transmission had already sown the seeds for an international epidemic, rendering these measures ineffective at curbing the epidemic. We address why the epidemic did not spread into neighbouring countries, showing that these countries were susceptible to substantial outbreaks but at lower risk of introductions. Finally, we reveal that this large epidemic was a heterogeneous and spatially dissociated collection of transmission clusters of varying size, duration and connectivity. These insights will help to inform interventions in future epidemics.

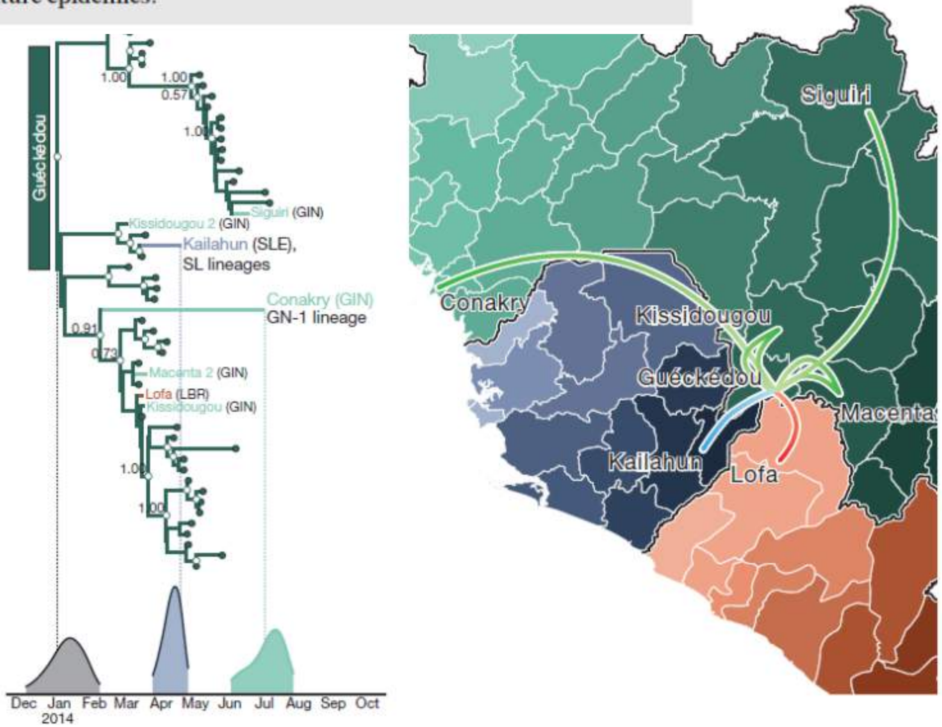


Figure 1 | Summary of early epidemic events. a, Temporal phylogeny of earliest sampled EBOV lineages in Guéckédou Prefecture, Guinea. 95% posterior densities of most recent common ancestor estimates for all lineages (grey) and lineages into Kailahun District, Sierra Leone (SLE; blue) and to Conakry Prefecture, Guinea (GIN; green) are shown at the bottom. Posterior probabilities >0.5 are shown for lineages with >5 descendent sequences. LBR, Liberia. **b,** Dispersal events marked by coloured lineages and labelled by name on the phylogeny are projected on a map with directionality indicated by colour intensity (from light to dark). Lineages that migrated to Conakry Prefecture (labelled as GN-1 lineage) and Kailahun District (labelled as SL lineages) have led to the vast majority of EVD cases throughout the region.

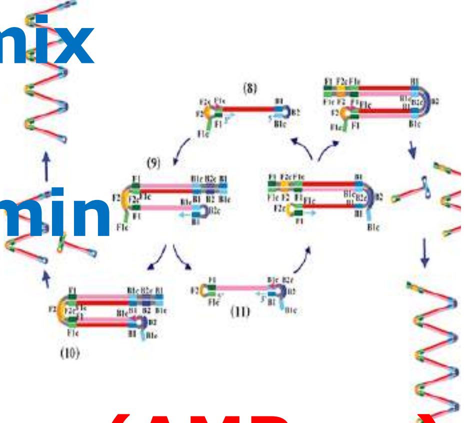
Sample Prep without a PCR

LAMP Amplification

Serum (1-5 microL)

Mix with Dry Lamp reagent mix

65oC 60 min



Purification (AMPure)

Needing only a magnetic stan

Nanopore Sample Pre

SIM card-like attachment

Purification (AMPure) → MinION

Global Research Alliance for Infectious Disease

IRELAND
University of Dublin
(W. Hall)

GERMANY
Munster University
(W. Makalowski)

POLAND
Poznan University
(I. Makalowski*)

EGYPT
Assiut University
(T. Mohammed)

Saudi Arabia
KAUST (A. Pain)

THAILAND
Mahidol University
(N. Komalamisra)

University of Tokyo (Y. Suzuki)

Obihiro University
(R. Maeda)

Hokkaido University
(J. Yamagishi)

Osaka University
(TBP*)

JAPAN

VIETNAM
National Institute of Health and Epidemiology
(L. A. Nguyen)

INDONESIA
Sam Ratulangi University
(J. Tuda, E)

PCR/LAN x

Targets (tentativ
Malaria Parasites
Dengue Virus
Chikungunya Virus
Japanese Encephaliti
(Zika Virus)
HBV

-Acceler
-Facilita

Oxford Nanopore @nanopore · 7月5日
International Science! Researchers frm Thailand, Indonesia, Iran, Poland, Russia, India, Korea, Japan in Thailand today exploring pathogens

ti

***JSPS Asia Africa Core to Cor**


2012 09 25

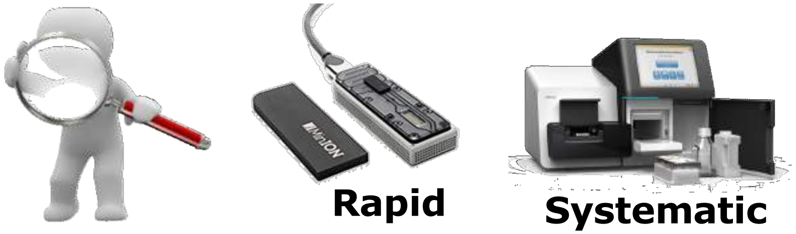
Technical Training in Japan

G-RAID III; Call for Par

Third meeting of Global Research Alliance for Infectious Diseases: Sam Ratulangi University

Manado, Indonesia, 2018. 7. 9-12 (tentative)

Integration of Genomics and One Health Concept aiming innovative disease control



Rapid **Systematic**

Monitoring with genomics (NGS)
- Genome epidemiology -

[What we can achieve with this project in particular]

- **Emergency sequencing in outbreak (ex: Ebolavirus)**
- **Rapid and intensive monitoring of specific genotype (ex: malaria, AMR)**
- **Environmental sequencing to [redacted] ex: livestock, B. anthracis)**
- **and** **capacity building**

MinION sequencing of pathogens of tropical diseases

Sam Ratulangi University

Sorong, Papua
(as of 22nd Aug, 2017)

(“現場”写真)

G-RAID II; Mahidol U, Bangkok, Thai; 2017.7.4-7

Conclusions:

i) Illumina Sequencing has given diverse

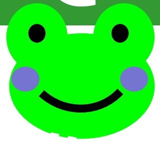
ii) MinION Sequencing has started to dive
the Venues of the analyses/researchers.

Acknowledgments

東京大学
THE UNIVERSITY OF TOKYO

Department of Computational Biology and Me
Sciences,
the University of Tokyo
Kazumi Abe, Kiyomi Imamura
Sarun Sereewattanawoot
kai

ysuzuki@h



Calling for challenging white papers!

NCC EPOC

Division of Genome Translational rch,
EPOC, National Cance
Ayako Suzuki, Hir umi,
Katsuya Tsuchihar

KE

システム癌新次元

Soga

MEXT KAKENHI Grant Number 16H01582 and 16H06279

Thank you for your kind at

Funding

CREST

Single Cell Analysis

PAGS (ゲノム支援)

LSDB 統合データベースプロジェクト

ToGo Database

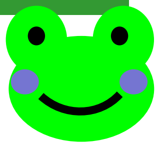
JSPS Asia-Africa Core-to-Core Pr

Acknowledgments

Dept of CBMS, the University of Tokyo

Kiyomi Imamura
Kazumi Abe
Uni Ishikawa
Megumi Konbu
Sachie Shima
Terumi Horiuc
Yuuta Kuze
Hiroyuki Wakaguri
Masahide Seki
Sumio Sugano

ysuzuki@h



Funding

Single Cell Analysis   PAGS

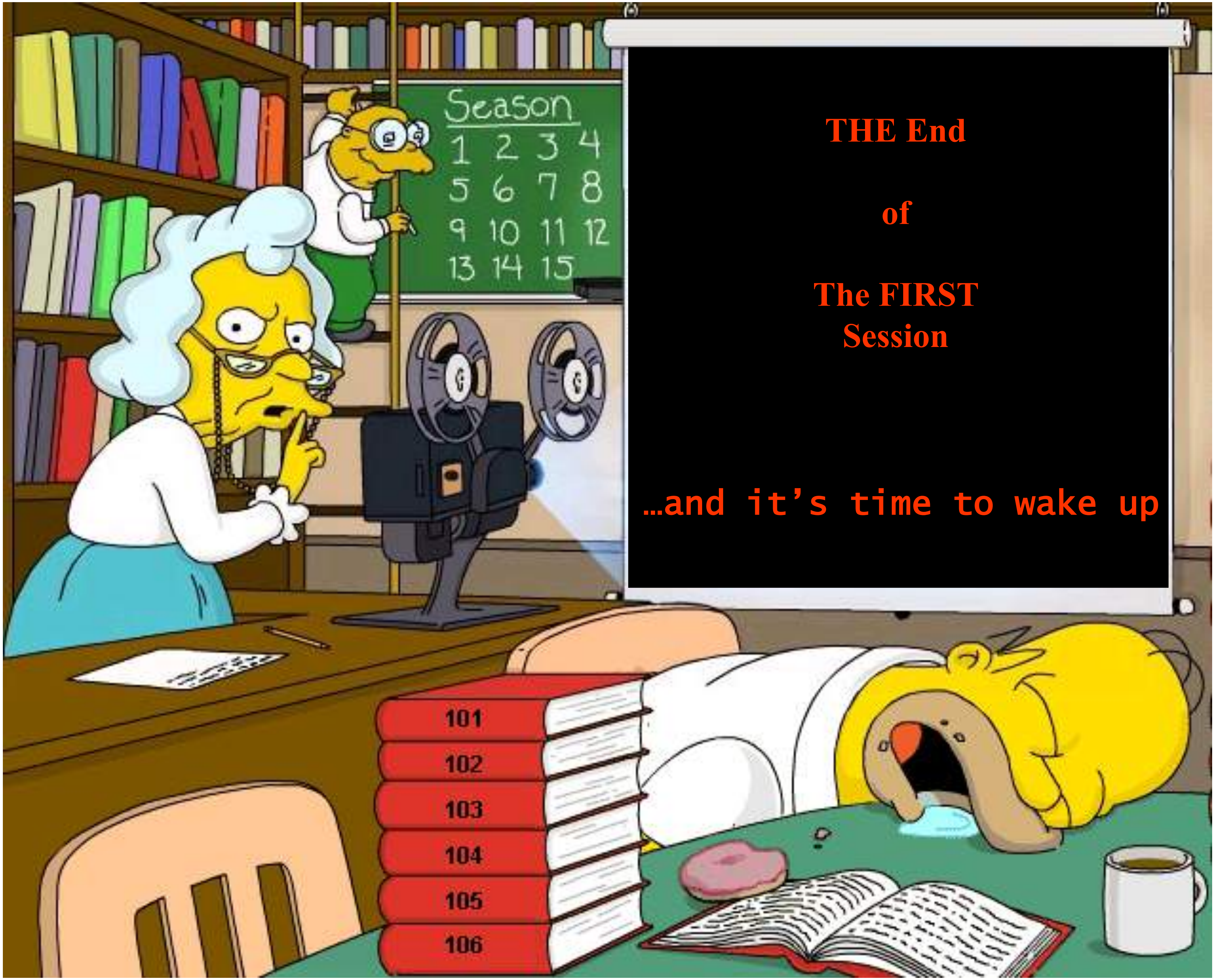
ToGo Database

JSPS Asia-Africa Core-to-Core Pr

Database

DBTSS/KERO @http://dbtss.hgc.jp/

FullParasite/DBAT @http://fullmal.hgc.jp/



THE End

of

The FIRST
Session

…and it’s time to wake up

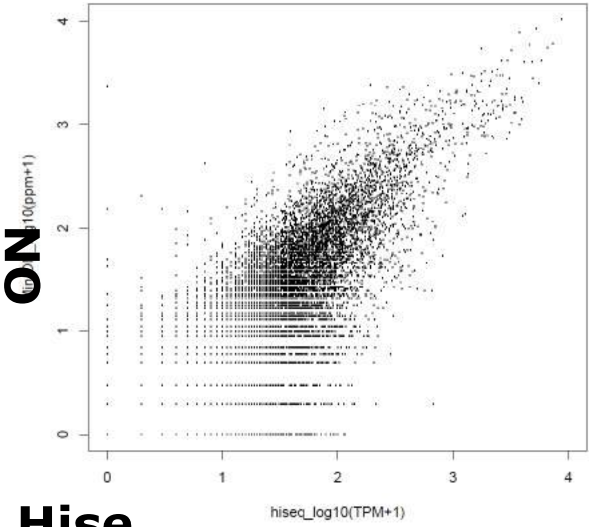
MinION Application expanded t

Comparison with MinION and TruSeq with va

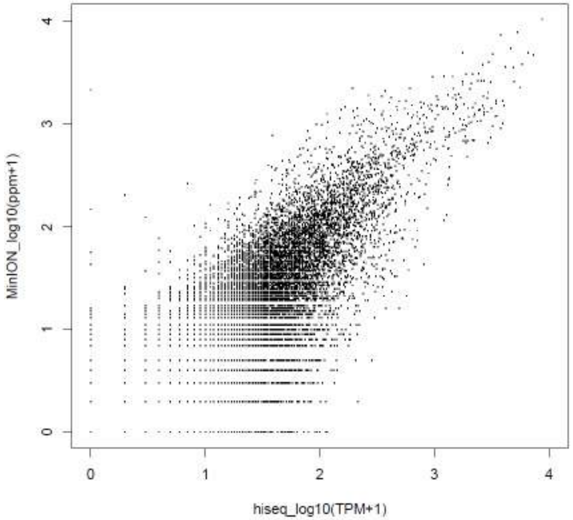
Mapped: 645808 / 723750 = 0.89

MinION

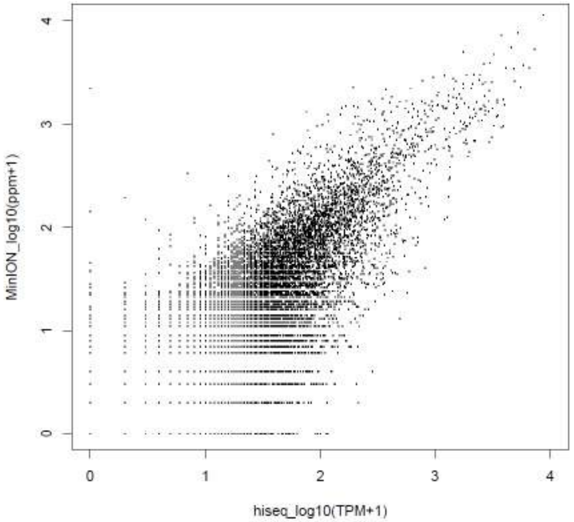
Hiseq



omfLAST
R=0.89



BWA-MEM
R=0.89



BWA-MEM
ont2d
R=0.88

Reasonalbe correlation between MinION

MMSC “Salc

This round is already closed.
But we will call the collaborators next year, too.
Hopefully, May, 2016.

(an advance announcement)

CALL FOR COLLABORATOR

for molecular medical science

To create a mutually beneficial relationship between Japanese and Indonesian science community, we keep scientific activities supported by a couple of grants such as JICA 2001-2003, Heiwa Nakajima Foundation 2004, Japan Society for the promotion of Science (JSPS) 2005-2007, Japan Science and Technology Agency (JST) 2008-2011, and JSPS Core-to-Core Program 2011-.
From 2013-2014, we recruited 6 students as collaborators of the first season for molecular diagnostic research for Malaria. With devoted supports from faculty staffs of UNSRAT, we have made successful achievements including oral presentation in the 55th Annual Meeting of the Japanese Society of Tropical Medicine held in Japan.

Now we are planning to start the second season.

The tentative research objects are

1. Molecular diagnostic of drug resistant Malaria parasites with portable and rapid next generation sequencer, Nanopore.
2. Molecular diagnostic of drug resistant Malaria parasites with classic PCR-RFLP method.
3. Evaluation of immune status in Malaria and other parasites by immune gene expression.
4. Epidemiological study for toxoplasma antibody prevalence.

Hopefully, we will open the recruitment in February 2015 and participants will be selected mainly by essay demonstrating how you will contribute the

arget is medical students in 5th

departments will be accepted.

molecular biology workshops

nany motivated candidates will

ssoc. prof. Junya Yamagishi

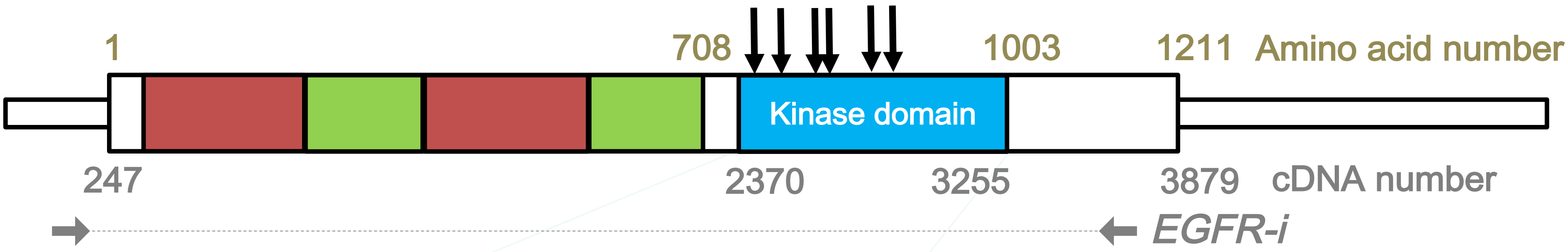
.jp

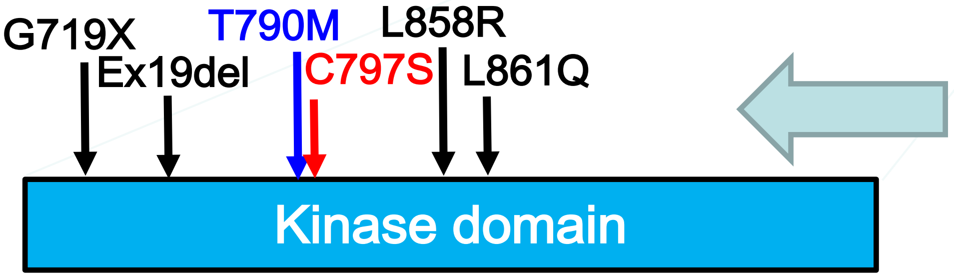
Thank you for your kind attention.

…Still it’s time for lunch.

Founding members of MMSC, UNSRAT Indonesia

Long reads for phasing:
EGFR kinase domain covered by MinION rea





Related in sensitivity and resistance to the molecular targeted drugs

Primary mutation: L858R. ex19del…

Sensitive to gefitinib/erlotinib

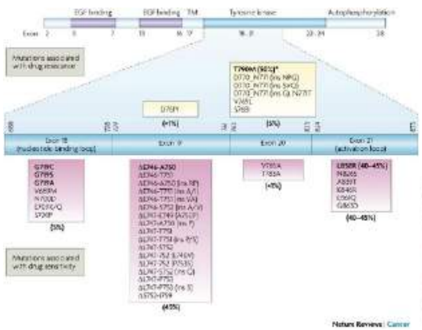
Secondary mutation: T790M

Resistant to gefitinib/erlotinib

Sensitive to AZD9291

Tertiary mutations: C797S

Resistant to AZD9291



Sharma *et al*.. 2007 *Nat Rev Cancer*

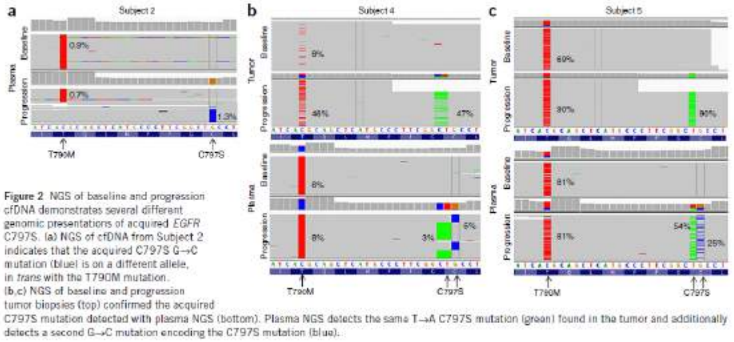


Figure 2 NGS of baseline and progression cfDNA demonstrates several different genomic presentations of acquired *EGFR* C797S. (a) NGS of cfDNA from Subject 2 indicates that the acquired C797S G→C mutation (blue) is on a different allele, in *trans* with the T790M mutation. (b,c) NGS of baseline and progression tumor biopsies (top) confirmed the acquired C797S mutation detected with plasma NGS (bottom). Plasma NGS detects the same T→A C797S mutation (green) found in the tumor and additionally detects a second G→C mutation encoding the C797S mutation (blue).

Thress *et al*. 2015
Nature Medice

We examined whether the mutations co-occurs in the same allele or not

ONT physical long-read & whole-genom

High-molecular weighted (HMW) gDNAs

End Repair

dA-tailing

Adapter Ligation

Tether attachment

Sequencing library

Library loading

Metrichor

Provide cloud-based analytics platform

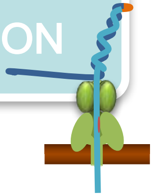
Basecalling

MinKNOW

Software that controls MinION

Sequencing (~48h)

MinION



Sequencing device

FASTQ extraction

poretools

Mapping

LAST
BWA

- Phasing
- Analysis of complex structural variants etc…

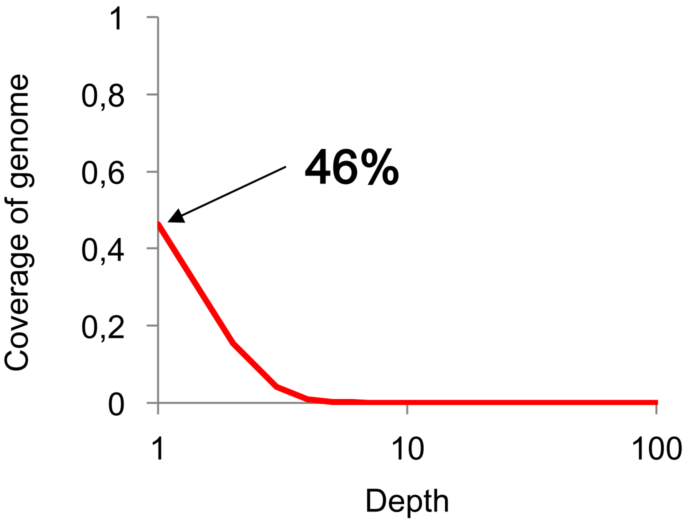
ONT physical long-read WGS for H1975 cells

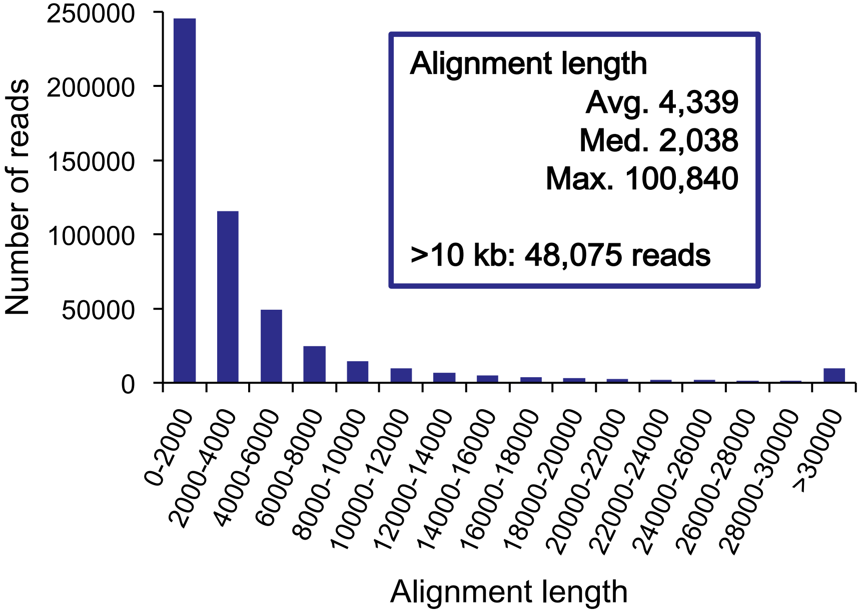
H1975, 2D

- 3 runs; R9.4
- bwa-0.7.12/bwa mem -t 4 -x ont2d
- Discarding supplementary alignment (0x800)

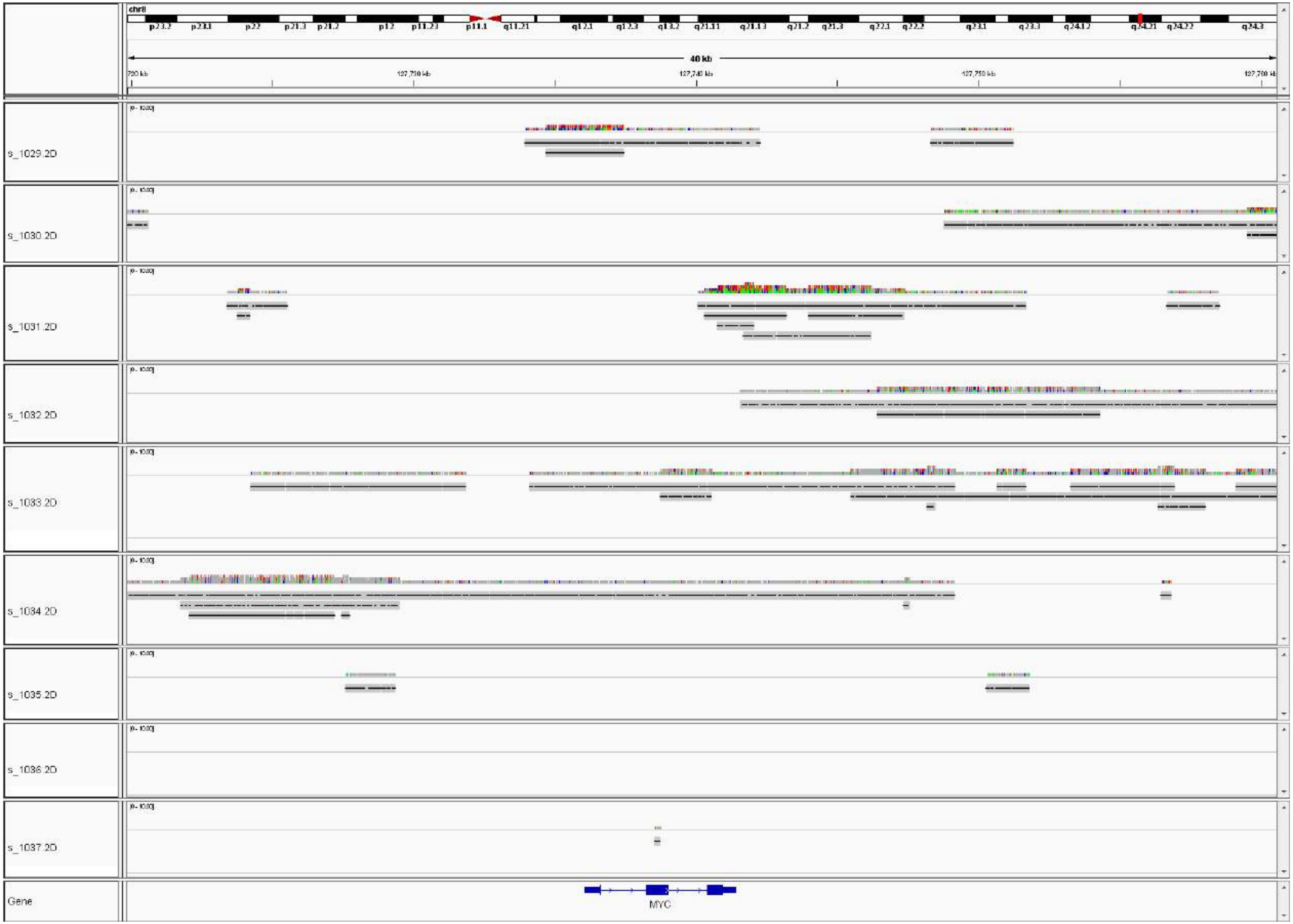
*Mapped to UCSC hg38; chr1-22 and XYM

Number	Run date	Cell	Number of reads								2D reads (pass + fail)				
			Raw		Template		Complement		2D		Total	Un-mapped	Mapped to human genome*	Avg. depth	Coverage (≥1×)
			pass	fail	pass	fail	pass	fail	pass	fail					
s_1038	161109	H1975	176143	141586	176143	132411	176143	54334	176143	13190	189333	3976	185357 (97.9%)	0.25	0.21
s_1039	161110	H1975	170608	125644	170608	115619	170608	47985	170608	12220	182828	4022	178806 (97.8%)	0.24	0.21
s_1040	161116	H1975	126835	96932	126835	91159	126835	33419	126835	10026	136861	2905	133956 (97.9%)	0.20	0.17
	Merge		473586	364162	473586	339189	473586	135738	473586	35436	509022	10903	498119 (97.9%)	0.69	0.46





Around MYC region (UCSC hg38)



160916 H1975

160921 H1975

160928 H1975

161021 RERF-LC-KJ

161022 H1975

161025 LC2/ad

161025 II-18

161028 H1975

161102 H1975

H1975 & LC2/ad → MYC amp

RNA Seq on MinION for long read

Cell
↓ <-RNeasy
RNA
↓ <-SMARTer
cDNA
↓ <-ONT prep kit
seq sample

→ sequence run(~48h)
fastq/fasta data
↓ <-BWA-MEM /
Graphmap / LAST
Expression analys

Poster #

Run date	CellType	Condition	FLOWCELL	PrepKit	Raw Reads	2d	
					ALL	pass	fail
16.08.31	LC2/ad	Bulk, SMARTer(UM dT, UM primer)	FLO_MIN104	SQK-NSK007 (custom)	169,401	49,593	20,915
16.09.02	H1975	Bulk, SMARTer(UM dT, UM primer)	FLO_MIN104	SQK-NSK007 (custom)	251,062	65,006	30,247
16.09.06	PC-9	Bulk, SMARTer(UM dT, UM primer)	FLO_MIN104	SQK-NSK007 (custom)	135,717	35,578	14,759

Stats of RNA Seq on MinION

FLOWCELL	**PrepKit**	**Raw Reads (1d+2d)**				
		Total (1d+2d)	**1d**	**2d**		
				Total (pass+fail)	**Pass**	**fail**
FLO_MIN106(R9.4)	SQK-LSK208	233,331	82,409	150,922	136,753	14,169
FLO_MIN106(R9.4)	SQK-LSK208	252,951	93,834	159,117	142,245	16,872
FLO_MIN106(R9.4)	SQK-LSK208	311,128	100,569	210,559	188,805	21,754
FLO_MIN106(R9.4)	SQK-LSK208	410,826	129,975	280,851	255,947	24,904
total		1,208,236	406,787	801,449	723,750	77,699

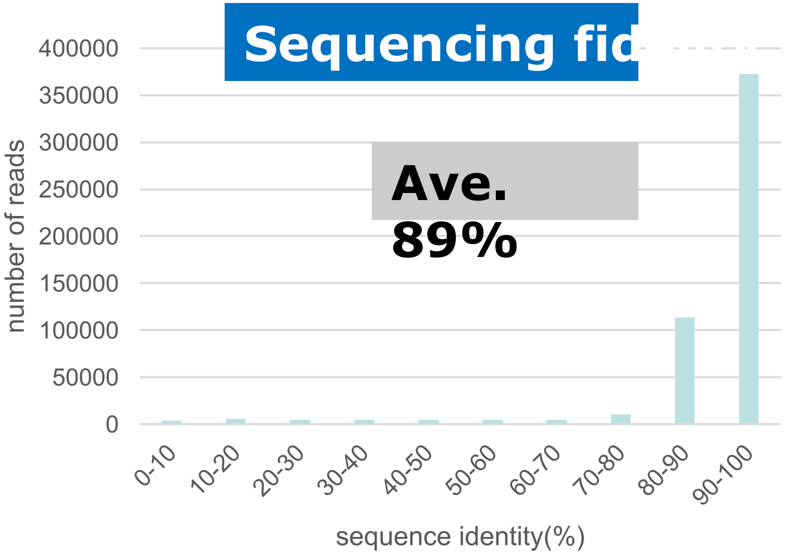
A MinION read mathced with the

Range 1: 117 to 1405 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1860 bits(2062)	0.0	1222/1304(94%)	59/1304(4%)	Plus/Plus

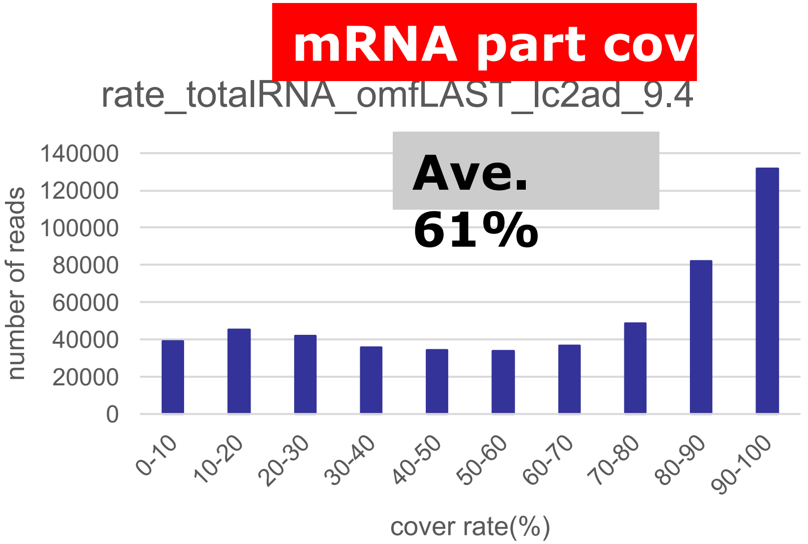
```
Query  60    TCTGCTCCTCCTGTTCGACAGTCAGCCGCATCTTCTTTTGCGTCGCCAGCCGAGCCACA  119
Sbjct  117   TCTGCTCCTCCTGTTCGACAGTCAGCCGCATCTTCTTTT-GCGTCGCCAGCCGAGCCACA  175

Query  120   TCGCTCAGGACACCATGGG-AAGGTGAAGGTCGGAGTCAACGGATTTG-TCGTATTGGGC
Sbjct  176   TCGCTCAG-ACACCATGGGGAAGGTGAAGGTCGGAGTCAACGGATTTGGTCGTATTGGGC

Query  178   GCCTGGTCACCAGGGCTGCTTTTAACTCTGGTAAAGTGGATATTGTTGCCATCACCGACC
Sbjct  235   GCCTGGTCACCAGGGCTGCTTTTAACTCTGGTAAAGTGGATATTGTTGCCATCAATGACC

Query  238   CCTTCATTGACCTCAACTAC--GGTTTACATATGTTCCAATATGATTCCACCCATGGCAA
Sbjct  295   CCTTCATTGACCTCAACTACATGGTTTACAT--GTTCCAATATGATTCCACCCATGGCAA

Query  296   ATTCCATGGCACCGTCAAGGCTGAGAACGG-AAGCTTGTCATCAATGGAAATC--ATCAC
Sbjct  353   ATTCCATGGCACCGTCAAGGCTGAGAACGGGAAGCTTGTCATCAATGGAAATCCCATCAC

Query  353   CATCTTCCAGGAGCGAGATCCCTCCAAAATCAGAGTGGGGCGATGCTGGCGCTGAGTACG
Sbjct  413   CATCTTCCAGGAGCGAGATCCCTCCAAAATCA-AGTGGGGCGATGCTGGCGCTGAGTACG

Query  413   TCGTGGAGTCCACTGGCGTCTTCACCACCATGGAGAAGGCTGGGGCTCATTTGCAGGGGG
Sbjct  472   TCGTGGAGTCCACTGGCGTCTTCACCACCATGGAGAAGGCTGGGGCTCATTTGCAGGGGG

Query  473   -AGCCAAA-GGGTCATCATCTCTGCCC--TCTGCTGACGCCCC-ATGTTCGTCGCGG-TG
Sbjct  532   GAGCCAAAAGGGTCATCATCTCTGCCCCCTCTGCTGATGCCCCCATGTTCGTCATGGGTG

Query  527   TGAACCATGAGAAGTATGACAACAGCCTCAAGATCATCAGCAATGCCTCCTGCACCACCA
Sbjct  592   TGAACCATGAGAAGTATGACAACAGCCTCAAGATCATCAGCAATGCCTCCTGCACCACCA

Query  587   ACTGCTTAGCACCCCTGGCCAAGGTCATCCGCGACAACTTTGGTATCGTGGAAGGACTCA
Sbjct  652   ACTGCTTAGCACCCCTGGCCAAGGTCATCCATGACAACTTTGGTATCGTGGAAGGACTCA

Query  647   TGACCACAGTCCATGCCATCACTGCCACCCAGAAGACTGTGGATGGCCCCTCCGGGAAAC  706
Sbjct  712   TGACCACAGTCCATGCCATCACTGCCACCCAGAAGACTGTGGATGGCCCCTCCGGGAAAC  771

Query  707   TGTGGCGTGATGGCCG-TGGGCTCTCCAGAACATCATCCCTGCCTCTACTGGCGCTGCCA  765
Sbjct  772   TGTGGCGTGATGGCCGCGGGGCTCTCCAGAACATCATCCCTGCCTCTACTGGCGCTGCCA  831

Query  766   AGGCTGTGGGCAAGGTCAT-CCTGAGCTGAACGGG--GCTCACTGGCGCGGCTTCAGGTG  822

Query  823   TCCCCACTGCCAACGTGTCAGGTGGTGGACCTGACCTGCCGTCTAGAAAAACCTGCCAAA  882
Sbjct  892   TCCCCACTGCCAACGTGTCA-GTGGTGGACCTGACCTGCCGTCTAGAAAAACCTGCCAAA  950

Query  883   TATGATGACATCAAGAAGGTGGTGAAGCAGGCGTCGGAGG---CCCTCAAGGGCATCCTG  939
Sbjct  951   TATGATGACATCAAGAAGGTGGTGAAGCAGGCGTCGGAGGGCCCCCTCAAGGGCATCCTG  101

Query  940   GGCTACACTGGAGCACCAGGTGGTCTC--CTGAC-TCAACAGCGACACCCA-TCCTCCAC  995
Sbjct  1011  GGCTACACT-GAGCACCAGGTGGTCTCCTCTGACTTCAACAGCGACACCCACTCCTCCAC  106

Query  996   CTTCGACGCTGGGGCTGGCATTGCCCTCAACGACCACTTTATCAAGCTCATTTCCTGGTA  105
Sbjct  1070  CTTTGACGCTGGGGCTGGCATTGCCCTCAACGACCACTTTGTCAAGCTCATTTCCTGGTA  112

Query  1056  TGACAAC--AGTTGACTCTACACAGCAACAGGGTGGTGGACCTCATGGCCCACATGGCCT  111
Sbjct  1130  TGACAACGAATTTGGCT----ACAGCAACAGGGTGGTGGACCTCATGGCCCACATGGCCT  118

Query  1114  ATCAAGGAGTAA---CCCT-GACCACCAGCCCCAGTAAG-GCACAAGAGGAAGAGAGAGA  116
Sbjct  1186  -CCAAGGAGTAAGACCCCTGGACCACCAGCCCCAGCAAGAGCACAAGAGGA--AGAGAGA  124

Query  1169  GACCCTCACTGCT-GGGAGTCC--GCCACACTCAGTCCCCCACCACACCGAATCT--CCT  122
Sbjct  1243  GACCCTCACTGCTGGGGAGTCCCTGCCACACTCAGTCCCCCACCACACTGAATCTCCCCT  130

Query  1224  CCTCACAGTTTCCATGTAGACCCC-TGAAGAGGGGA--GGCCT-GGGA--TGCACCTTTG  127
Sbjct  1303  CCTCACAGTTGCCATGTAGACCCCTTGAAGAGGGGAGGGGCCTAGGGAGCCGCACC-TTG  136

Query  1278  TCATGTACCATCAATAAAGTACCCTGTGCTCAACCAG--AAAAA  1319
Sbjct  1362  TCATGTACCATCAATAAAGTACCCTGTGCTCAACCAGTTAAAAA  1405
```

Stats of the performance in



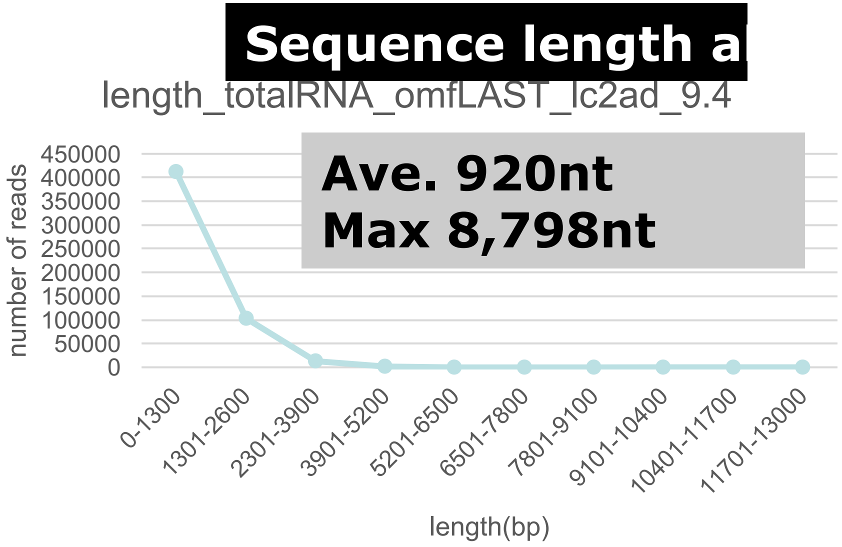
Sequencing fid

Ave. 89%



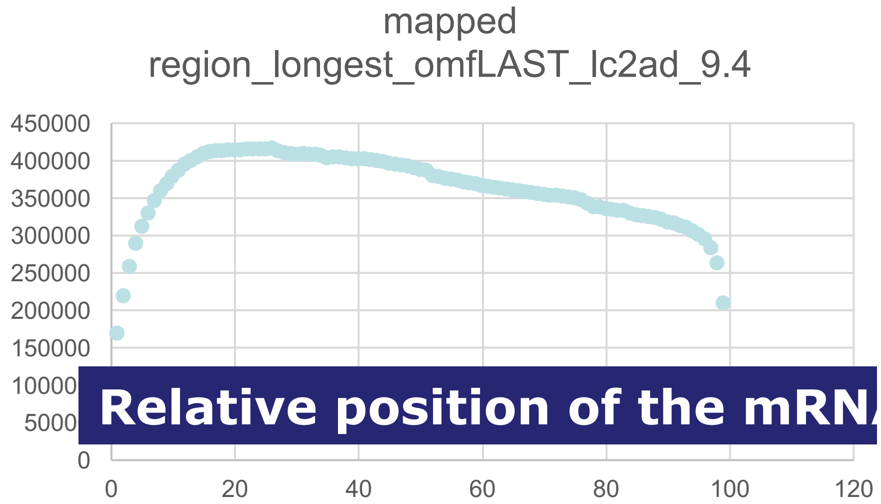
mRNA part cov

Ave. 61%



Sequence length a

Ave. 920nt
Max 8,798nt

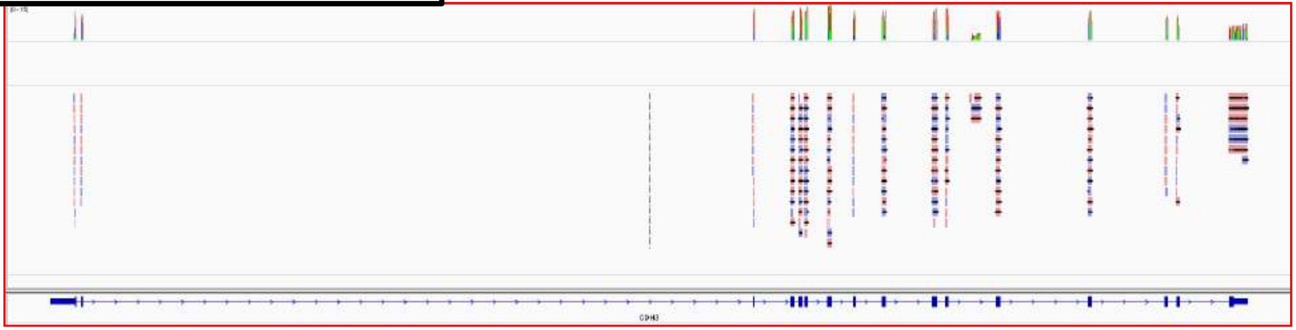


Relative position of the mRN

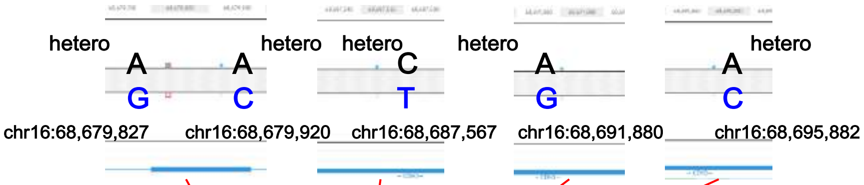
Phasing based on MinION

MinION Phase Block
chr16:68,679,827- chr16:68,695,882 (16kbp)

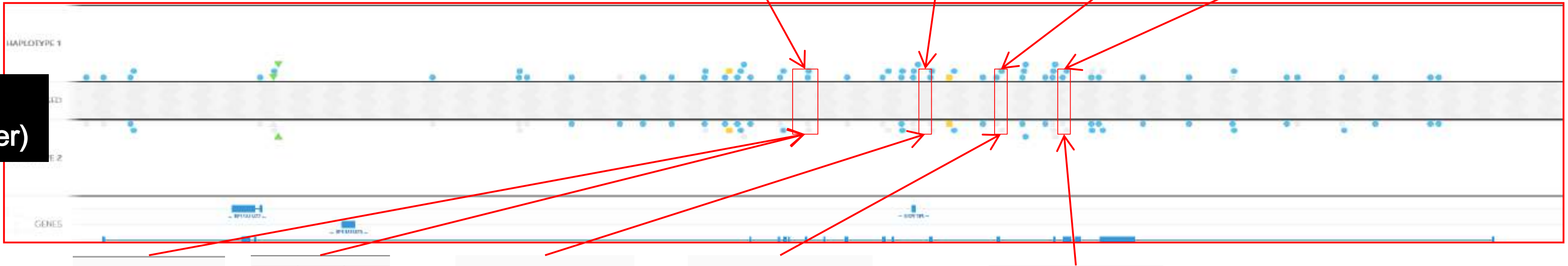
MinION
RNA Seq

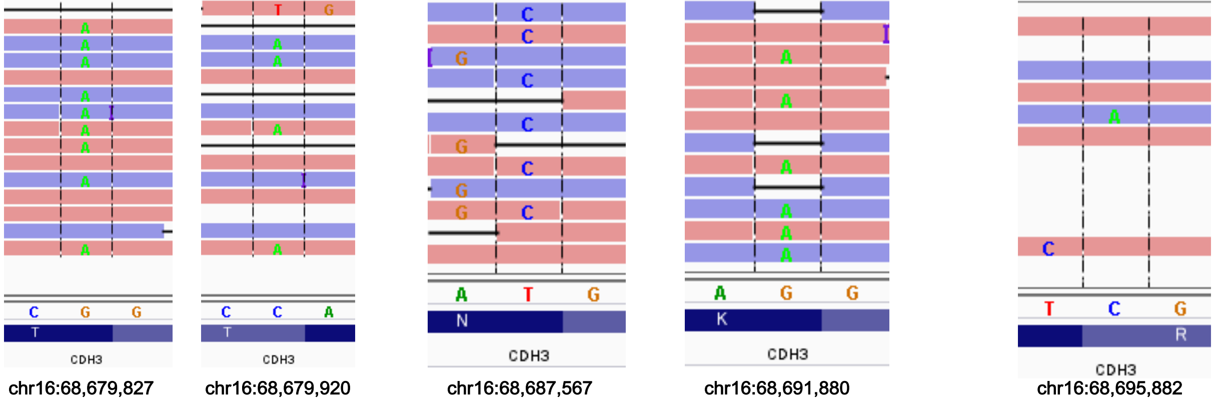


H1975 MinION RNA-seq
Around CDH3 region (UCSC hg38)



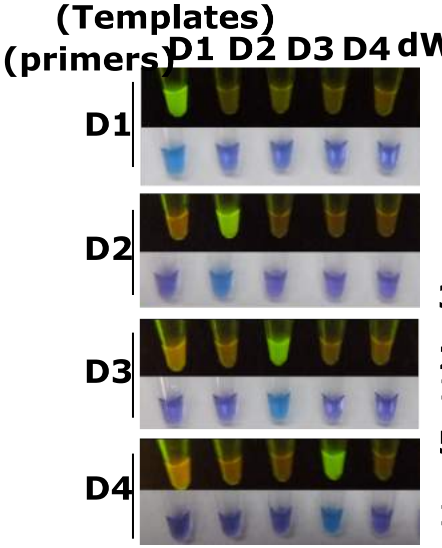
10X
(LongRanger)



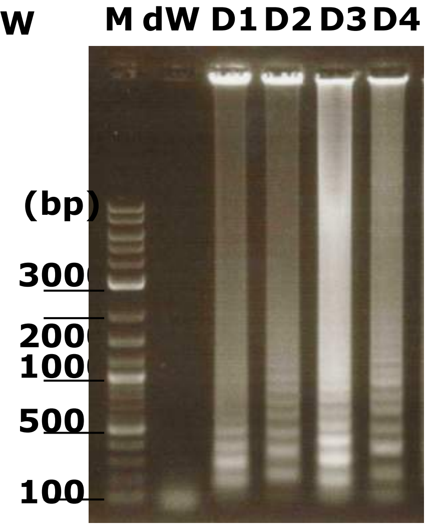


Template prep from in vitro cultured viruses; easy to be handled

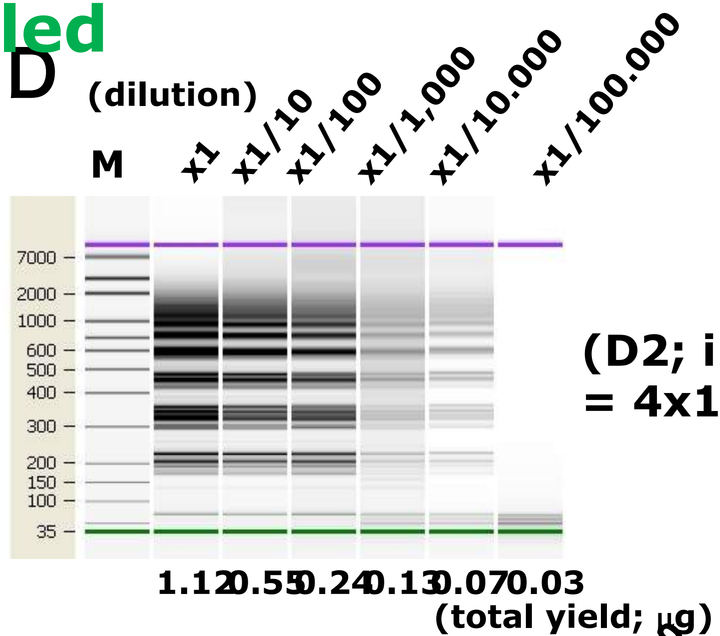
B



C

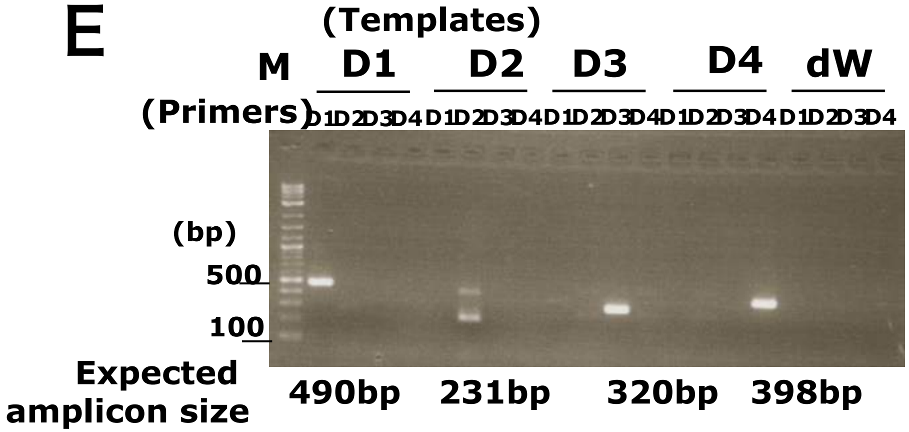


D

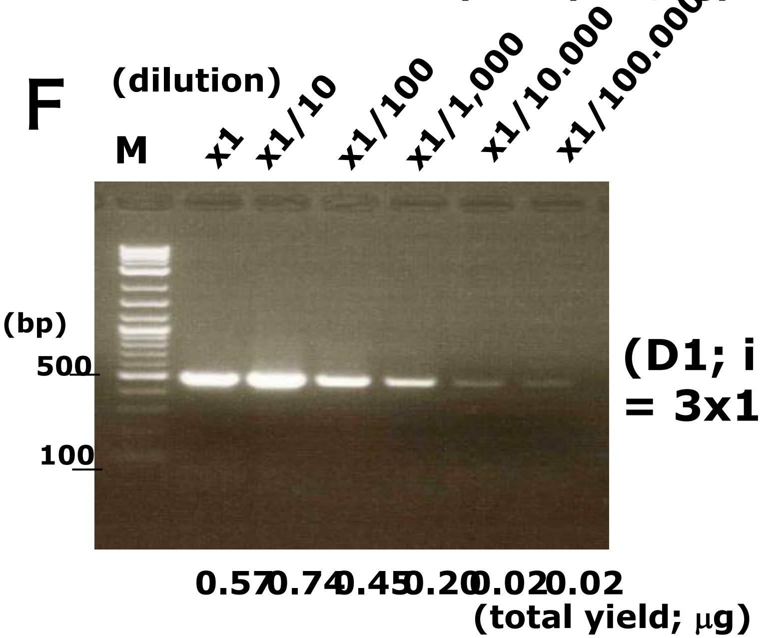


(D2; initial tit
= 4x10^7 PFU/

E



F



(D1; initial ti
= 3x10^8 PFU

- Serotypes could be separated on their own by LAMP

Inaccurate sequences may be used to separate serotypes,

Dengue Fever

-Transmitted by a bite of mosquito infected with dengue virus
(genome size=~ 11kb).

-Febrile illness that affects infants, young children and adults with symptoms appearing 3-14 days after the infective bite.

-There are four sero-types (D1~ D4), whose genomes were separated at 70% in the sequence identity.

	Den1	Den2	Den3
Den2	7385 / 10758		
Den3	7759 / 10743	7423 / 10740	
Den4	7168 / 10771	7216 / 10753	7214 / 10745

	Den1	Den2	Den3
Den2	69%		
Den3	72%	69%	
Den4	67%	67%	67%

-Second infection of the same sero-type may cause severe symptoms; dengue hemorrhagic fever, abdominal pain, persistent vomiting, bleeding and breathing difficulty and is a potentially lethal

http://www.who.int/topics/d