

バイオインフォマティクスとは？ What is bioinformatics?

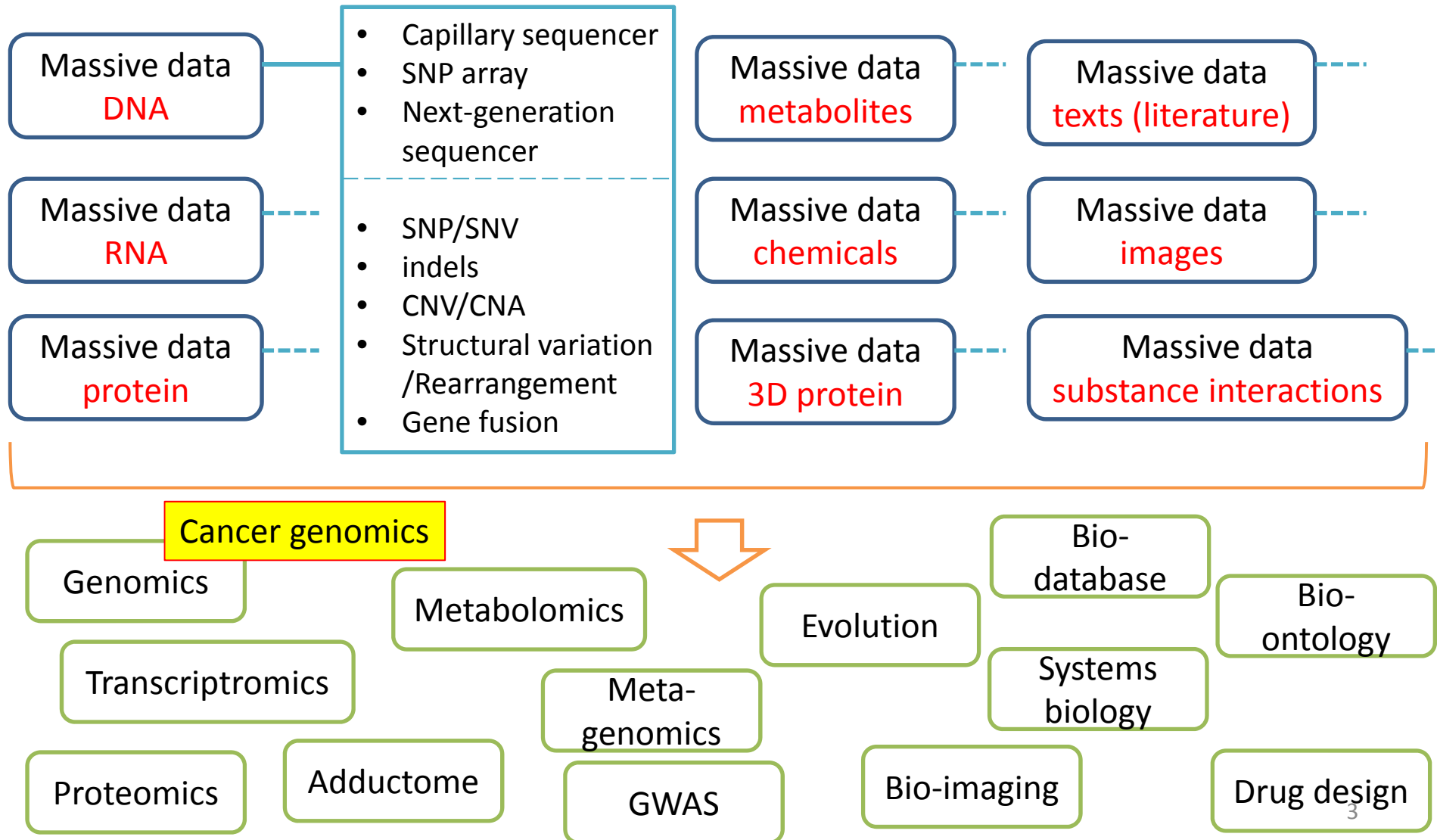
* 日本語で行います

研究所
バイオインフォマティクス部門
加藤 護

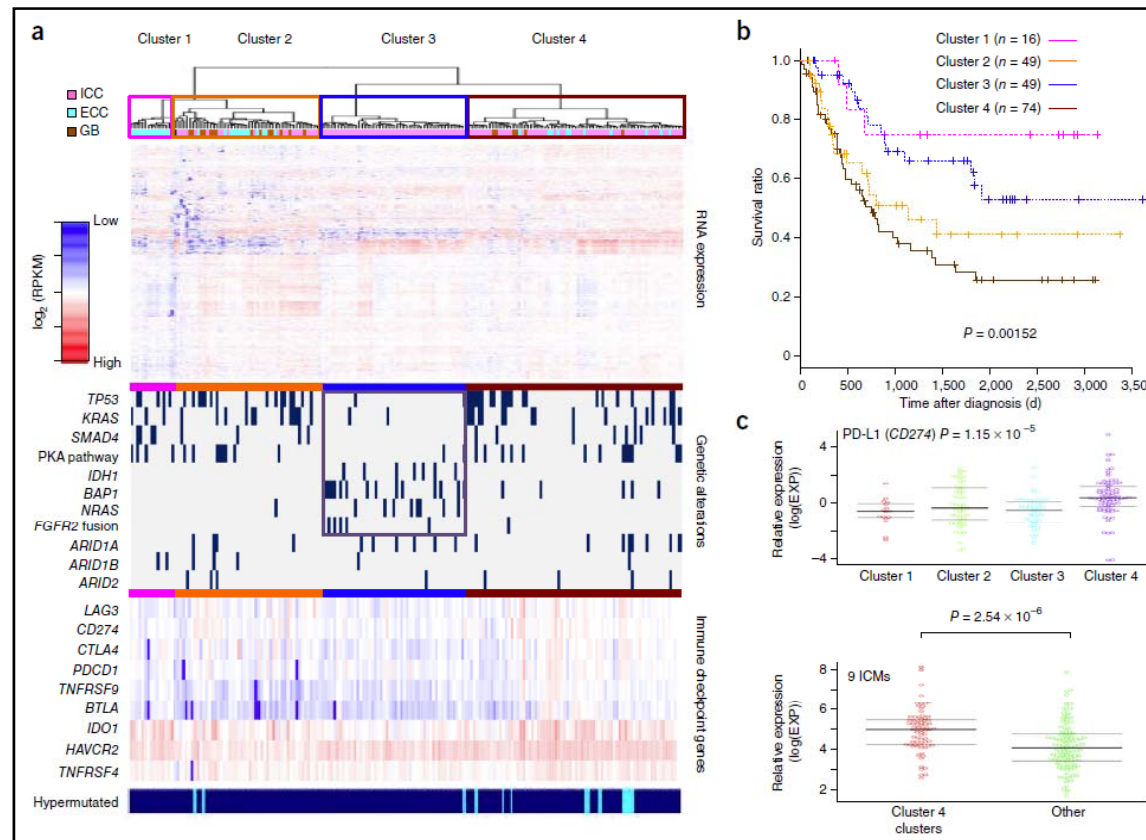
Outline

1. General concepts of bioinformatics
 - Within my limited experiences though...
2. Its application to genomic medicine
 - clinical sequencing

Bioinformatics: greedy and cloudy discipline



RNA analysis and the clinical relationships in bile-duct cancer genomics



(Nakamura et al, Nat Genet, 2015)

- Bioinformatics support for Shibata group in NCC

NGS data

6 TB data

```
@PERI8:9:45
CCCTCAGCTACGGGGGGGGGGTGGCTTCTTCTGTTACCTGGTG
GTGGCGGCTGTGACGCTCCTGCTGCTGCGCAGCCCCAGAACGGC
CGGAGCCATCCACGCGCTACCGTACCGGCGACATCGATCCAAT
GATACGCGGCTGAGCACA
+
/0(..0***0000000000%02..(15030111/322-***-(03/24)++-
22/+++230000.+++2.111---%***(**-1,1/*+(-
**2++**+/1,0(0..0.4%+++4223+++4*.)***+*024%++2+**+,
...
```



200 columns (samples)

Gene	BD003T	BD004T	BD005T	BD006T	BD007T
ENST0000	31.35851	81.2562	58.13853	35.76353	40.48326
ENST0000	10.01731	1.137802	32.82091	15.14492	2.095884
ENST0000	0	0	0	0	0
ENST0000	3.982066	1.120111	1.371183	5.04892	2.619011
ENST0000	0	0	0	0	0
ENST0000	0.241376	0.119728	0.021227	0.009749	0
ENST0000	0.061229	0.032396	0.057434	0.039568	0.093569
ENST0000	0	0.581962	0	0	0
ENST0000	146.4966	163.5045	205.3889	162.6099	96.99319
ENST0000	0	0	0	0	0
ENST0000	0	0	0	0	0
ENST0000	8.933542	1.986797	2.840649	4.501061	0.370228
ENST0000	3.923663	0.688758	2.00794	1.949078	0.777532

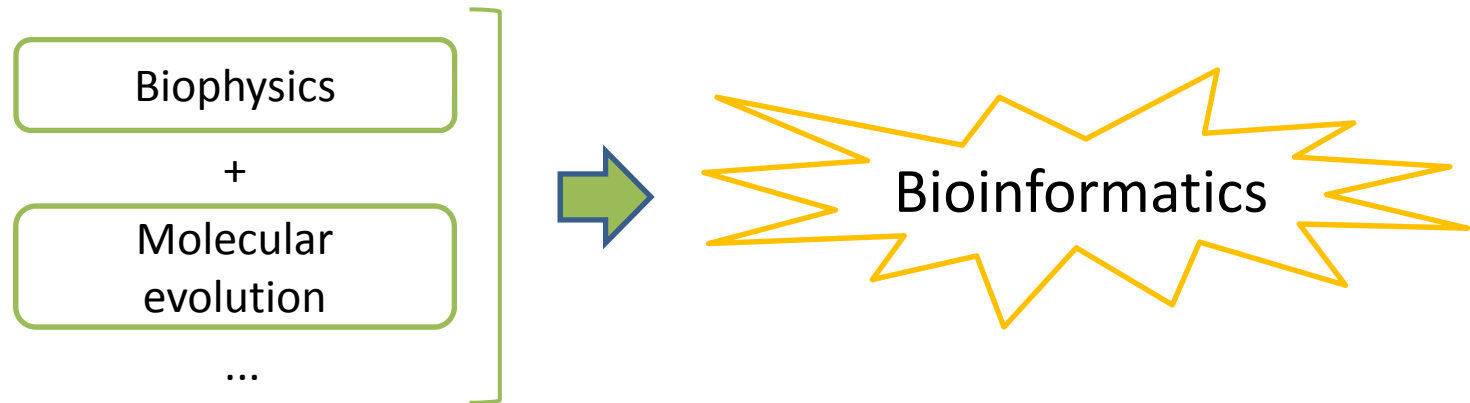
40,000
rows
(transcripts)



Clustering analysis

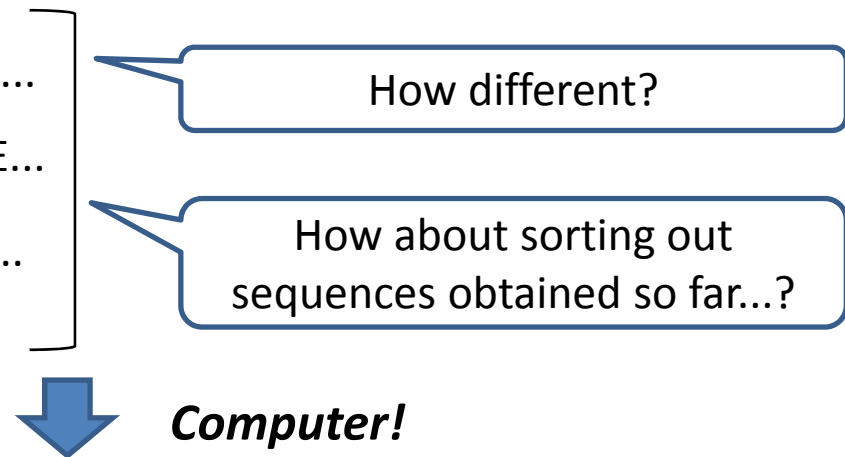
Birth of bioinformatics

- It's in 1970s



Protein sequence *1 letter = 1 amino acid

1. MKILETPFASGDLSMLVLLPDEVSDLERIEKTINFE...
2. MKILETPFASGDLSMLVLNPDEVSDLERIEKFINFE...
3. MKILETPFSSGDLSMLVLIPDEVSDLERIEKTINFE...
- ...



Dayhoff matrix

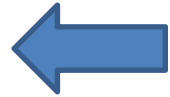


Dayhoff matrix & homology search

Homologous - similar
⇒ similar proteins

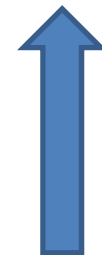
(Dayhoff et al, 1978)

Matrix



1. MKILETPFASGDLSMLVLLPDEVSDLERIEKTINFE...
2. MKILETPFASGDLSMLVLNPDEVSDLERIEKFINFE...
3. MKILETPFSSGDLSMLVLIPDEVSDLERIEKTINFE...

➡ Homology search
- **similar sequence**
⇒ **similar protein function**



New. MKILETPGASGDLSMLVLLPDE...

- Which sequence does this new resemble?

Homology search

- DNA

Matrix

	A	T	G	C
A	5	-4	-4	-4
T	-4	5	-4	-4
G	-4	-4	5	-4
C	-4	-4	-4	5

New: A T G C

Seq1: T T G C

$$-4 + 5 + 5 + 5 = 11$$

Seq2: T C G C

$$-4 - 4 + 5 + 5 = 2$$



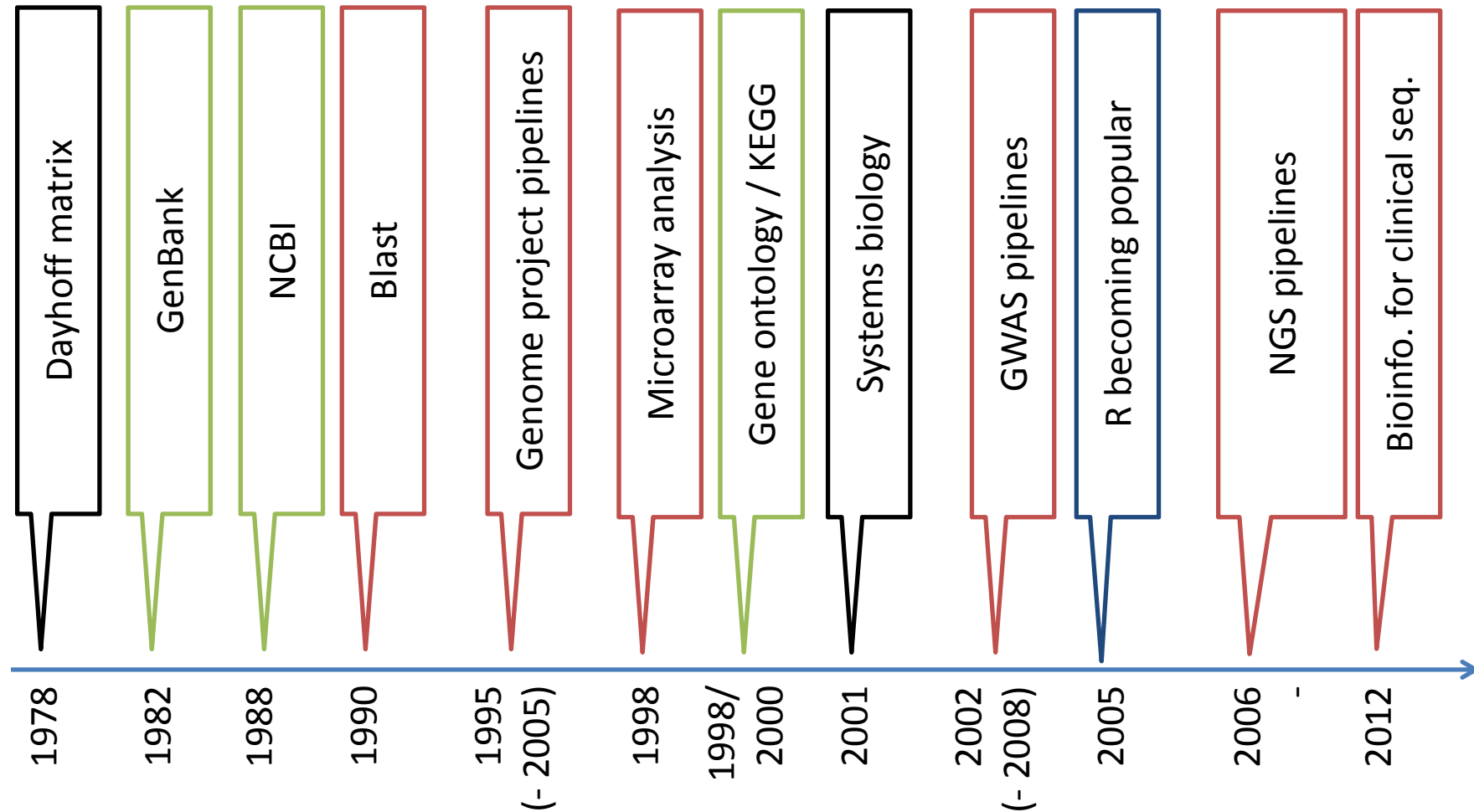
*New is more similar to seq1
even in the protein function.*

- Protein (amino acid sequence)

*(Say, seq1 function is
well-known.)*

Same idea

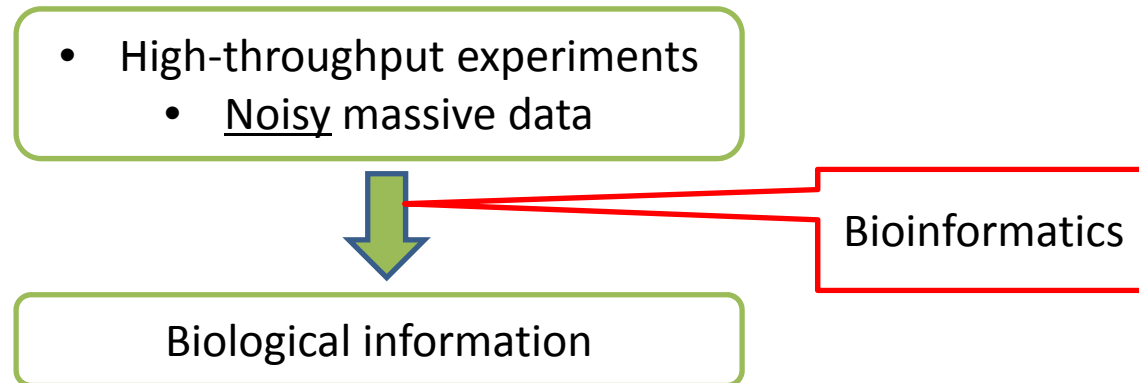
Brief history of bioinformatics



The essence

Signal from Noise

– a needle in a haystack –

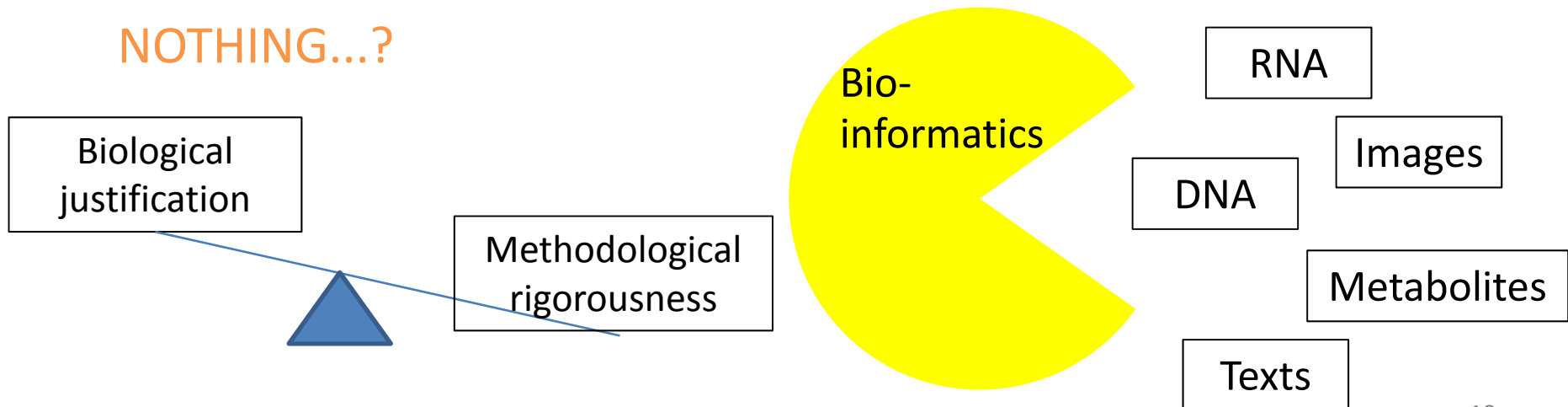


	Data type: phenotypic or molecular?	Data quantity	Emphasis: discovery or proof?
Bioinformatics	Molecular	Great many (>millions)	Discovery

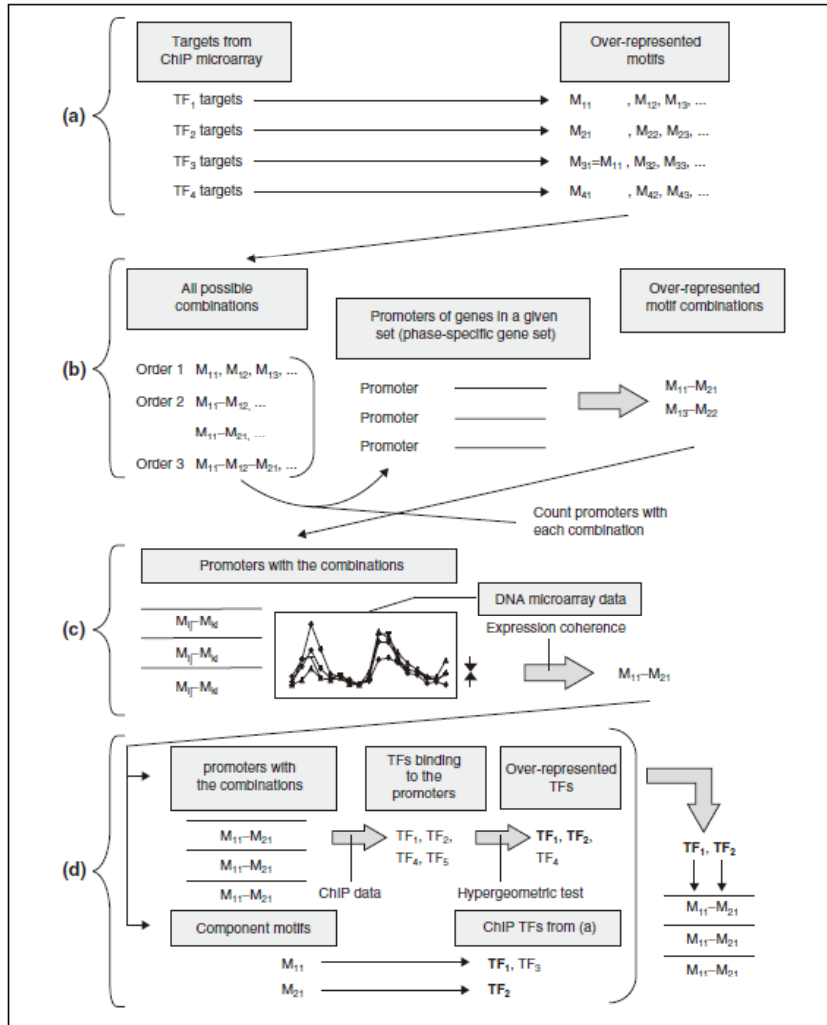
The methodological aspect

The theoretical basis

- Mechanics:
Newton's three laws
- Statistics:
population and sampling
- Bioinformatics:
NOTHING...?
- The spirit of “**pragmatism**”
✓ Whatever method,
as long as useful



Discovery, relaxing methodological rigorousness

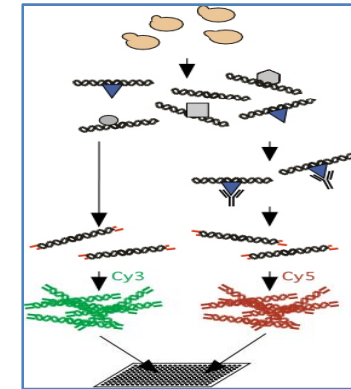


```
>YOR128C
TCGAAAAAGATCATT...
>YOR119W
CAGTTTATATAAATT...
> ...
```

Promoter sequences
represented by
200 * 6000 letters

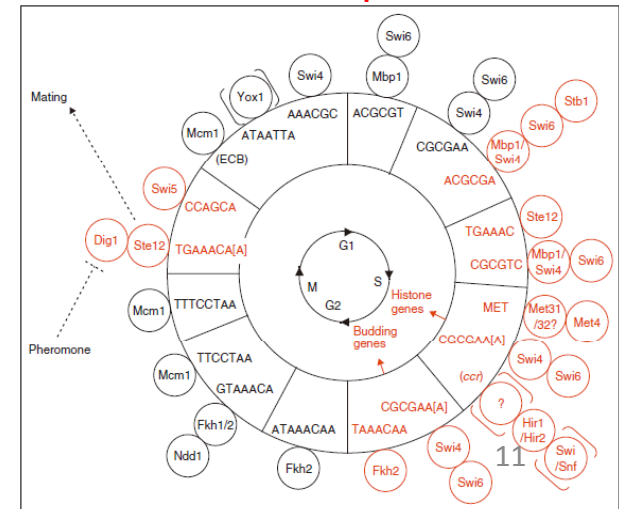
In the step (a): search TF-binding motif
200,000 statistical tests! $\Rightarrow$ 200,000 p-values

As, simply,
score...



(Lee
et al.,
Cell,
2002)

Binding data
represented by
113 * 6000 p-values



(Kato et al, Genome Biol., 2004)

Therefore, collection of arts

SKILL

Level I

Programming:
Linux, Perl, R, SQL

Practically
enough...

Level II

AND

Molecular
biology

Genome
biology

Hopefully...

Level III

AND

Classical
statistics

Computer-
intensive statistics

Expertise level

OR

Bayes
statistics

Graph theory

Machine
learning

Molecular
evolution

Lexical
analysis

Pattern
recognition

Statistical
genetics

Syntactic
analysis

Information
theory

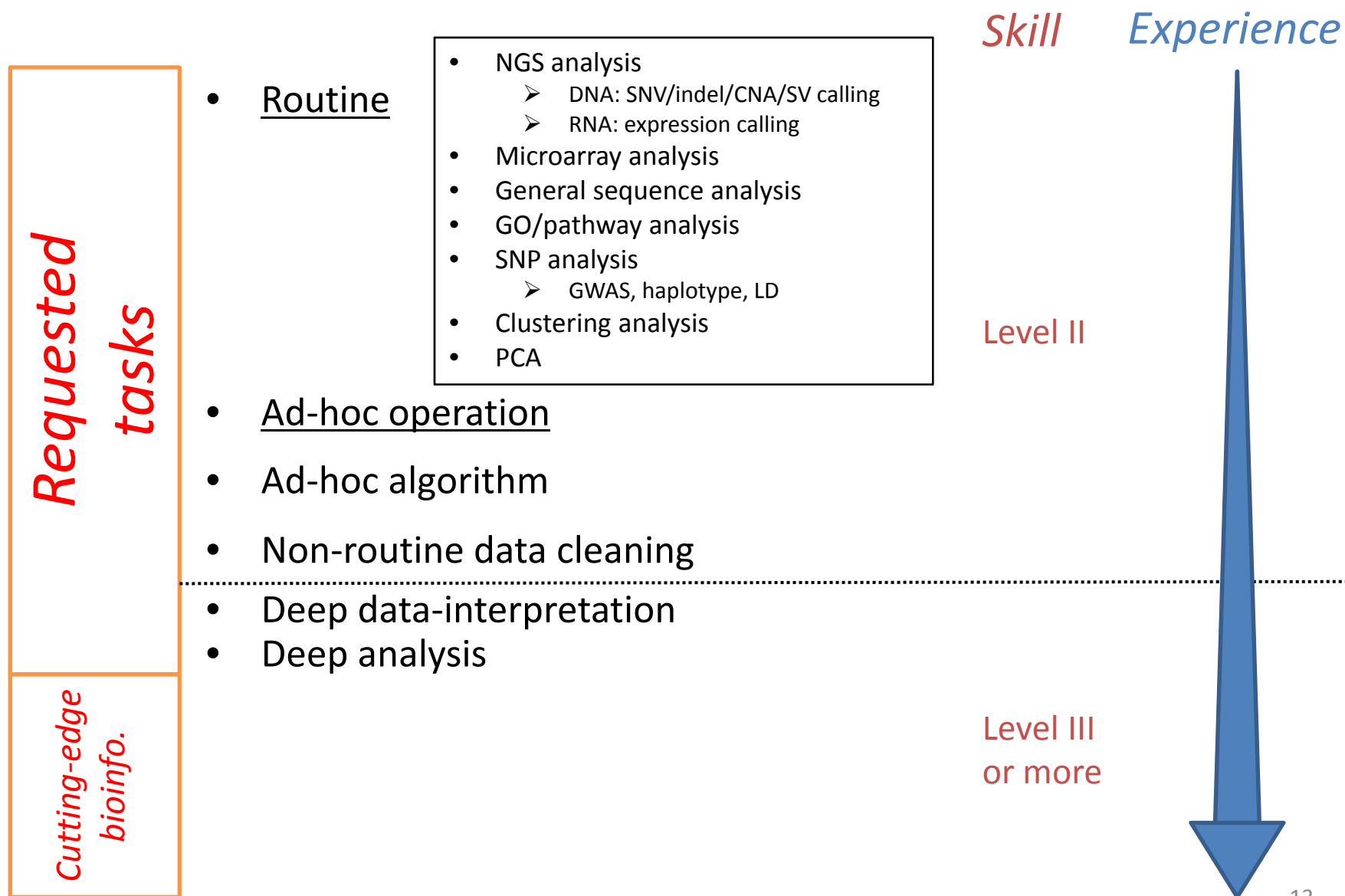
Combinatorics

Algorithm

Data mining

More...

Is level II enough?



Routine/Ad-hoc operation?

- NGS routine

On Linux

```
# make index
bwa index -p human_chrs.fa -a bwts human_chrs.fa

# aln
bwa aln -t 6 human_chrs.fa test.fastq 1> test.aln 2> test.aln.err
# samse
bwa samse human_chrs.fa test.aln test.fastq 1> test.sam 2> test.sam.err

# bwasw
bwa bwasw -t 6 human_chrs.fa test.fastq 1>| tmp.sw.1 2>| tmp.sw.2

# sam -> bam
## with index
samtools view -bS test.sam > test.bam

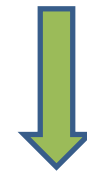
## no index
samtools faidx human_chrs.fa (--> .fai) # fasta index
samtools view -bt human_chrs.fa.fai test.sam > test.bam

# bam -> sam
samtools view -h test.bam > test.sam

...
```

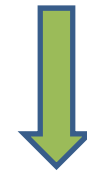
- Ad-hoc operation

1. Installation of a public bioinformatics tool



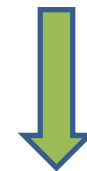
- Eg, “bwa” on the left

2. Understand the manual



- Though, sometimes

3. Apply to your own data



- Eg, command lines on the left

4. Get results

Advanced type of working: division of labor

- When bioinformatics tasks are many and complex, ...



Algorithm design

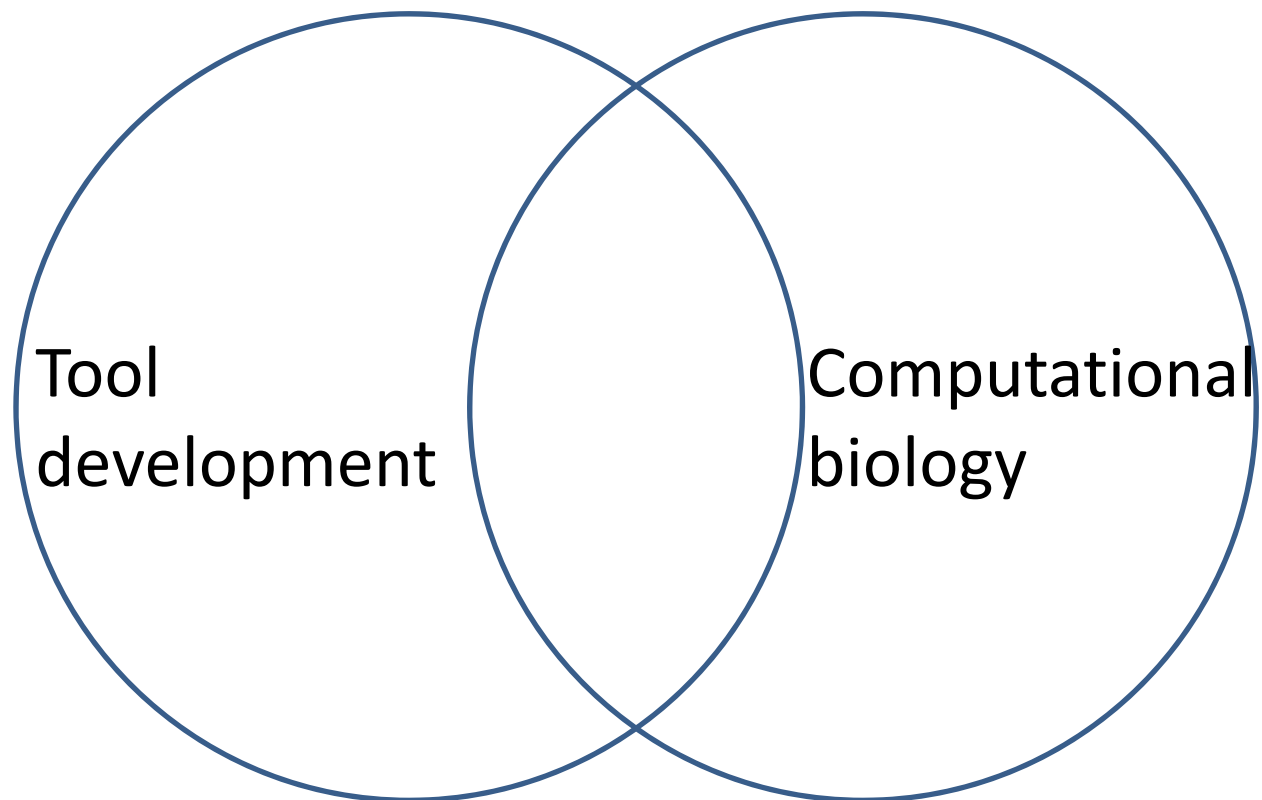
~ = experimental design
by scientists



Implementation (coding)

~ = perform experiment
by technical staff / technicians

Two extreme types of studies



Example of tool development

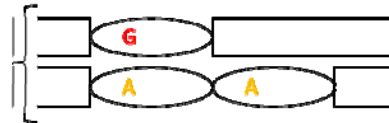
Problem in RETINA data

- What is the configuration of haplotypes(/genotypes)?

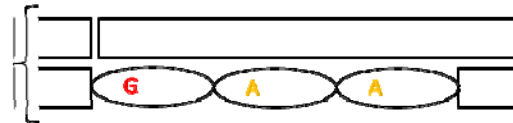
RETINA data: total number of bases over homologous chromosomes

Sample name	rs1065852 FAM (G)	rs3892087 Yellow (A)	...
NA10860	1	2	1 2 ...
NA11992	0	3	0 3 ...
NA12878	1	1	1 1 ...
NA12239	2	0	2 0 ...
...	...	...	...

rs1065852 for NA10860



OR?



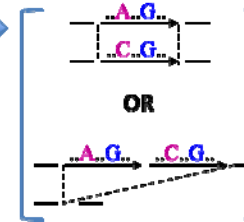
Principle of the algorithms

- List all possible diplotypes (pairs of haplotypes) that are consistent with the total numbers

Sample	SNVC 1 # of A	SNVC 1 # of C	SNVC 2 # of G	SNVC 2 # of T
Ind. 1	1	1	2	0
Ind. 2	2	1	2	1
Ind. 3	1	0	0	1

SNVC: SNV In a CNV

Individual 1



Sample	Possible diplotype (haplotype / haplotype)
Ind. 1	AG / CG
Ind. 2	AG, CG / -
...	...

"/": separator between copy units
"-": deletion

Principle of the algorithms

- Repeat E- and M-steps to estimate haplotype frequencies
 - Using possible diplotypes obtained at the previous step

Sample	Possible diplotype	Weight (probability that the sample takes this diplotype)	Diplotype frequency under Hardy-Weinberg equilibrium
Ind. 1	haplotype 1 / haplotype 1	$w_{11} \propto F(h1 / h1)$	$F(h1 / h1) = 1 \cdot F(h1) \cdot F(h1)$
	haplotype 1 / haplotype 2	$w_{12} \propto F(h1 / h2)$	$F(h1 / h2) = 2 \cdot F(h1) \cdot F(h2)$
	haplotype 2 / haplotype 3	$w_{13} \propto F(h2 / h3)$	$F(h2 / h3) = 2 \cdot F(h2) \cdot F(h1)$
Ind. 2	haplotype 1 / haplotype 1	$w_{21} \propto F(h1 / h1)$	$F(h1 / h1) = 1 \cdot F(h1) \cdot F(h1)$
	...	...	...

$F(x)$: frequency of x

Giving arbitrary values to haplotype frequencies

E step:
Haplotype frequencies
→ Diplotype frequencies
→ Update the weights

Iteration

M step:
Number of haplotypes, considering the weights
→ Haplotype frequencies

Haplotype (14 SNVC sites)	True count	True frequency	Estimated frequency
[AA-AAAAAAAAAA]	428	0.3639	0.3653
[AA-AAAAAAAAAB]	382	0.3248	0.3235
[BA-AAAABAAAAA]	244	0.2075	0.2066
[AA-AAAAA-AAA]	24	0.0204	0.0194
[A-AAAA-AAAA]	23	0.0198	0.0199
[AA-AAAA-AAAA]	21	0.0179	0.0167
[BA-AAAAAAAAAB]	18	0.0153	0.0153
[AA-AAAAAABAB, AA-AAAAAABAB]	18	0.0138	0.0118
[AA-A-AAAAAAAA]	11	0.0094	0.0079
[AA-AAAAAAAAAA, AA-AAAAAAAAAA]	6	0.0051	0.0050
[BAAAAABAAAAAB]	-	-	0.0009
...	...	...	...

Allelic copy number	True count	True frequency	Estimated frequency
1 (copy)	1130	0.9609	0.9601
0 (copies)	23	0.0198	0.0200
2 (copies)	23	0.0198	0.0200
3 (copies)	-	-	<10 ⁻¹⁰

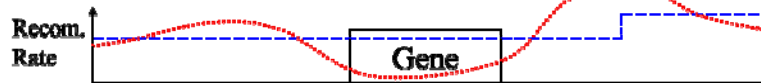
CNV-phaser
(Kato et al, Am. J. Hum. Genet., 2008)

MOCS-phaser
(Kato et al, Bioinformatics, 2008)

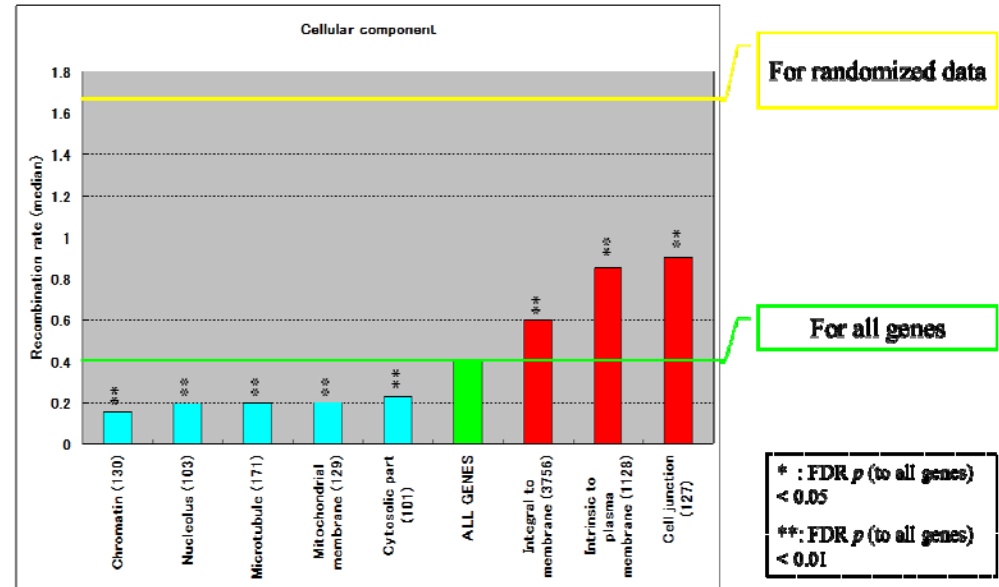
Example of computational biology

Background

- 2003 – The complete sequence of the human genome
 - an “average” sequence
- 2005 – The International HapMap Project, Phase I
 - Polymorphism in human genomes
 - Focused on single nucleotide polymorphism (SNP)
 - ✓ Catalogued SNP genotypes for 270 individuals in three ethnic populations (Asian, African, European) at 3 million SNP loci
 - Medical application as well as biological investigation
- The first data on human recombination rates at a high resolution
 - New statistical methods (by a HapMap group) to infer recombination rates from large-scale SNP genotyping data
 - kilo-base level high-resolution

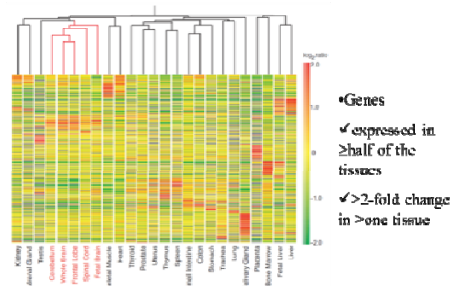


Gene Ontology: *cellar component*

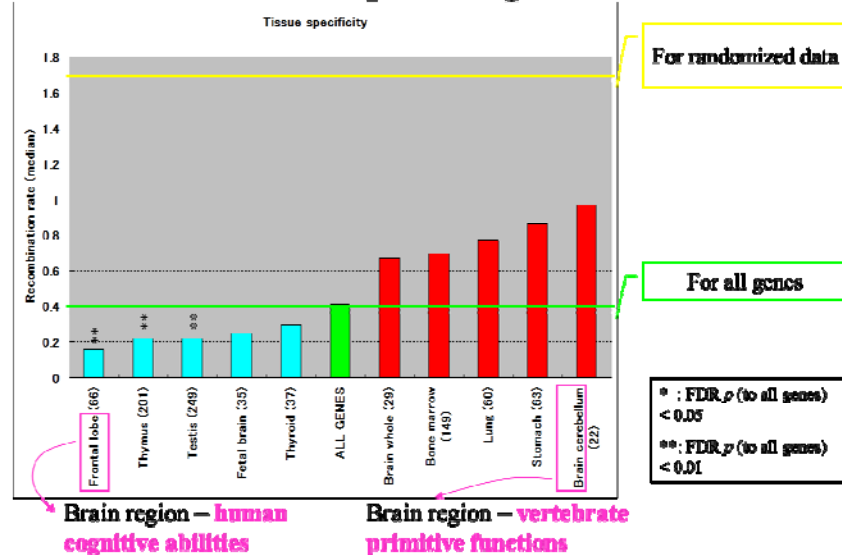


Tissue-specific genes

- Tissue-specific genes
 - We performed microarray experiments.
 - Microarrays with ~30,000 probes for 25 tissue samples
 - Genes highly expressed (FDR p < 0.001) in each tissue
 - The clustering analysis confirmed that we successfully measured expression levels.
 - Genes highly expressed in only one tissue → tissue-specific genes

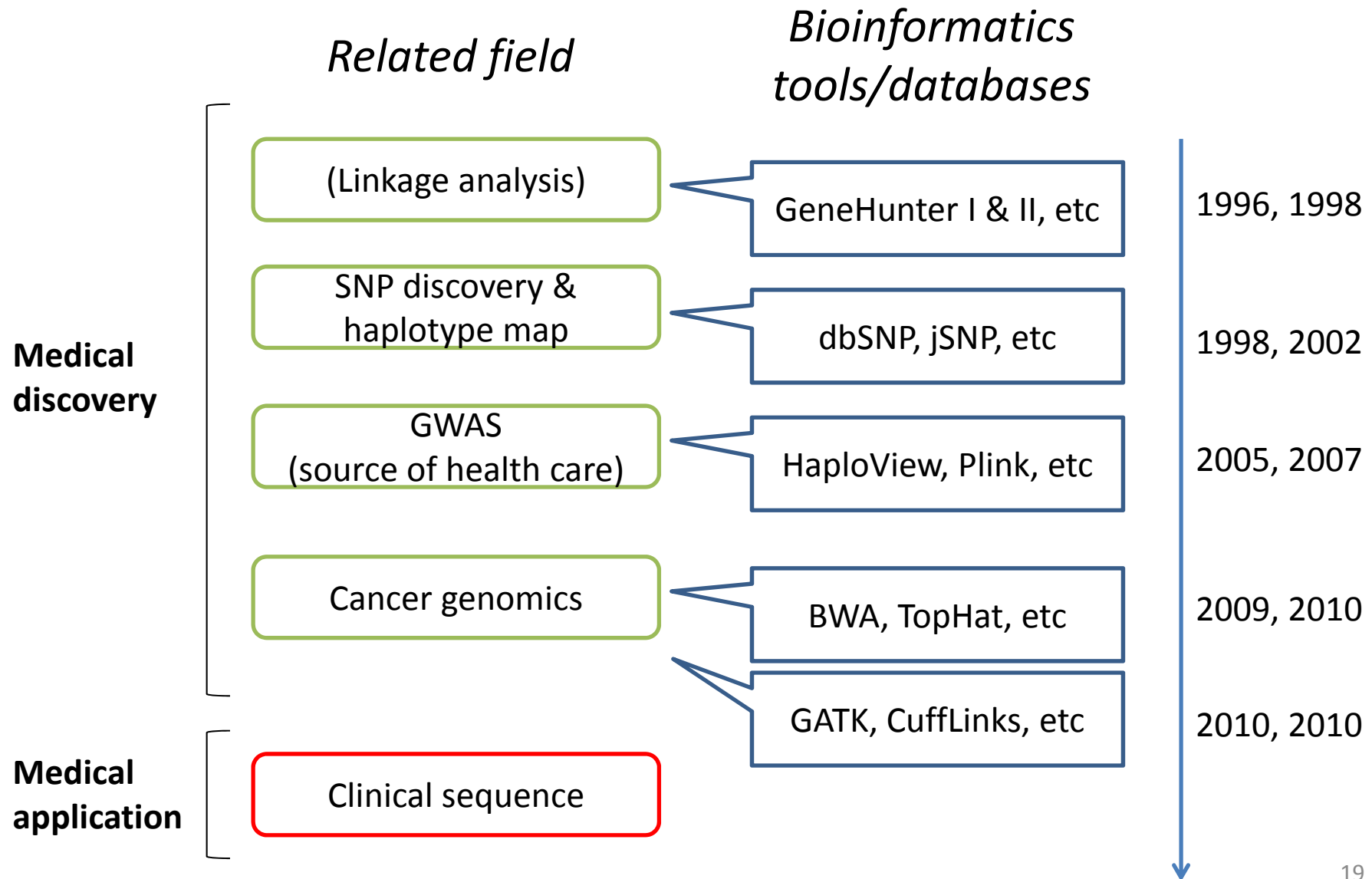


Tissue-specific genes



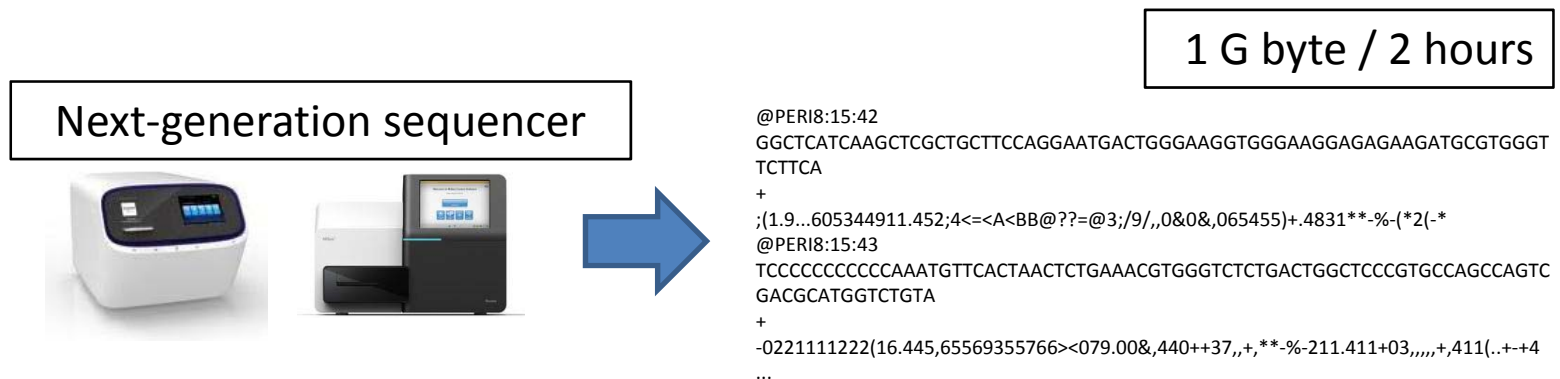
(Kato et al, Hum. Mol. Genet., 2008)

Genomic medicine & bioinformatics



The latest – clinical sequencing

- Detection of **different mutations in just one assay**
 - *Multiplex PCR, mass spectrometry, FISH*
 - ✓ Mostly, single type and single to several mutations
 - *Next-generation sequencing*
 - ✓ Point mutations, fusions, amplifications, deletions
 - ✓ All exons of ~100 genes (in our case)
 - ✓ Potential for research discovery



Gene alterations and drugs

Table 1 | Genomic alterations as putative predictive biomarkers for cancer therapy

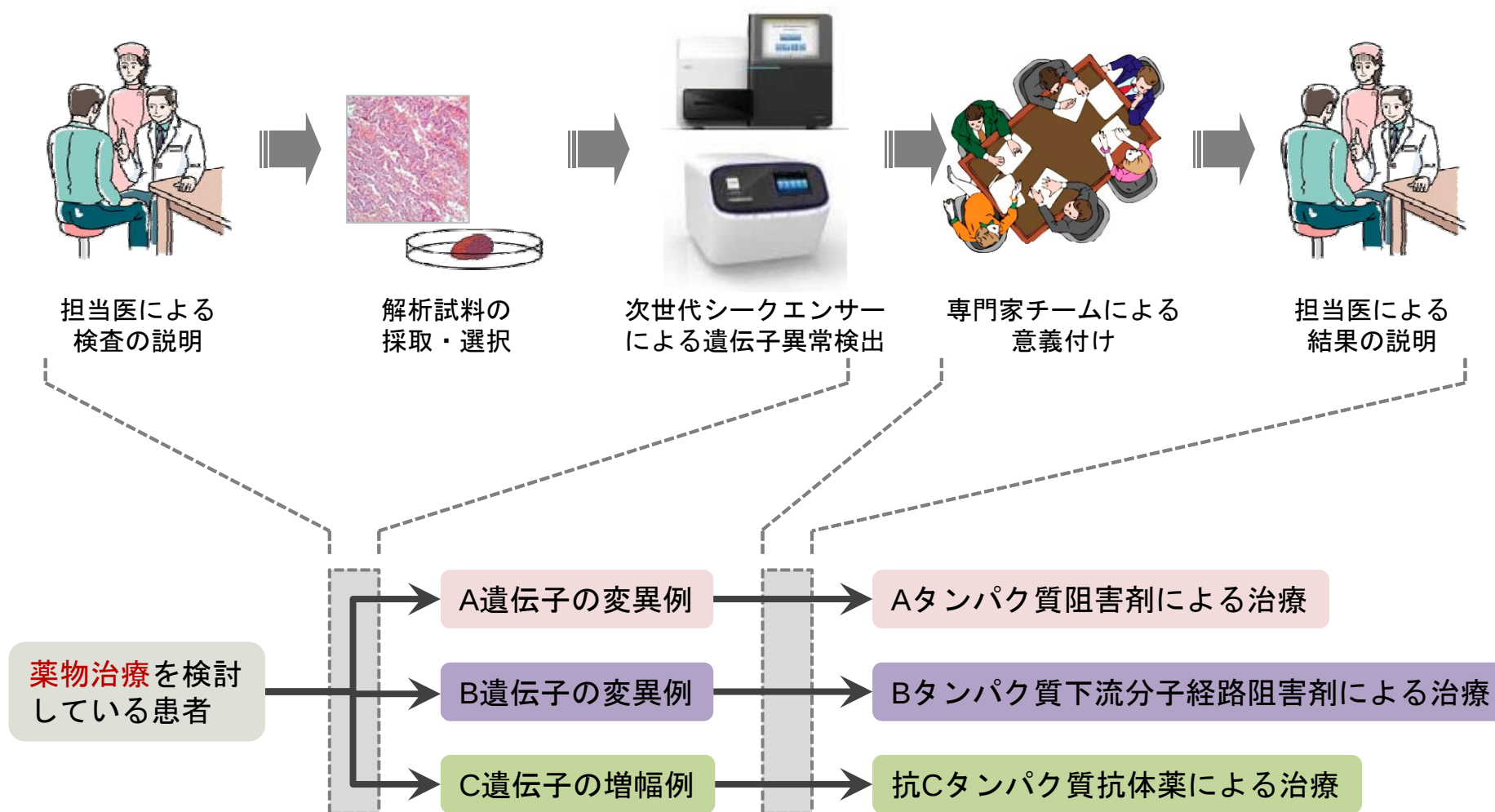
Genes	Pathways	Aberration type	Disease examples	Putative or proven drugs
PIK3CA ^{53,54} , PIK3R1 (REF. 53), PIK3R2, AKT1, AKT2 and AKT3 (REFS 54,55)	Phosphoinositide 3-kinase (PI3K)	Mutation or amplification	Breast, colorectal and endometrial cancer	* PI3K inhibitors * AKT inhibitors
PTEN ⁵⁶	PI3K	Deletion	Numerous cancers	* PI3K inhibitors
MTOR ⁵⁷ , TSC1 ⁵⁸ and TSC2 (REF. 59)	mTOR	Mutation	Tuberous sclerosis and Bladder cancer	* mTOR inhibitors
RAS family (HRAS, NRAS, KRAS), BRAF ⁶⁰ and MEK1	RAS-MEK	Mutation, rearrangement or amplification	Numerous cancers, including melanoma and prostate cancer	* RAF inhibitors * MEK inhibitors * PI3K inhibitors
Fibroblast growth factor receptor 1 (FGFR1), FGFR2, FGFR3, FGFR4 (REF. 36)	FGFR	Mutation, amplification or rearrangement	Myeloma, sarcoma and bladder, breast, ovarian, lung, endometrial and myeloid cancers	* FGFR inhibitors * FGFR antibodies
Epidermal growth factor receptor (EGFR)	EGFR	Mutation, deletion or amplification	Lung and gastrointestinal cancer	* EGFR inhibitors * EGFR antibodies
ERBB2 (REF. 61)	ERBB2	Amplification or mutation	Breast, bladder, gastric and lung cancer	* ERBB2 inhibitors * ERBB2 antibodies
SMO ^{62,63} and PTCH1 (REF. 64)	Hedgehog	Mutation	Basal cell carcinoma	* Hedgehog inhibitor
MET ⁶⁵	MET	Amplification or mutation	Bladder, gastric and renal cancer	* MET inhibitors * MET antibodies
JAK1, JAK2, JAK3 (REF. 66), STAT1, STAT3	JAK-STAT	Mutation or rearrangement	Leukaemia and lymphoma	* JAK-STAT inhibitors * STAT decoys
Discoidin domain-containing receptor 2 (DDR2)	RTK	Mutation	Lung cancer	* Some tyrosine kinase inhibitors
Erythropoietin receptor (EPOR)	JAK-STAT	Rearrangement	Leukaemia	* JAK-STAT inhibitors
Interleukin-7 receptor (IL7R)	JAK-STAT	Mutation	Leukaemia	* JAK-STAT inhibitors
Cyclin-dependent kinases (CDKs; ⁶⁷ CDK4, CDK6, CDK8), CDKN2A and cyclin D1 (CCND1)	CDK	Amplification, mutation, deletion or rearrangement	Sarcoma, colorectal cancer, melanoma and lymphoma	* CDK inhibitors
ABL1	ABL	Rearrangement	Leukaemia	* ABL inhibitors
Retinoic acid receptor-α (RARα)	RARα	Rearrangement	Leukaemia	* All-trans retinoic acid
Aurora kinase A (AURKA) ⁶⁸	Aurora kinases	Amplification	Prostate cancer and breast cancer	* Aurora kinase inhibitors
Androgen receptor (AR) ⁶⁹	Androgen	Mutation, amplification or splice variant	Prostate cancer	* Androgen synthesis inhibitors * Androgen receptor inhibitors
FLT3 ⁷⁰	FLT3	Mutation or deletion	Leukaemia	* FLT3 inhibitors
MET	MET-HGF	Mutation or amplification	Lung cancer and gastric cancer	* MET inhibitors
Myeloproliferative leukaemia (MPL)	THPO, JAK-STAT	Mutation	Myeloproliferative neoplasms	* JAK-STAT inhibitors
MDM2 (REF. 71)	MDM2	Amplification	Sarcoma and adrenal carcinoma	* MDM2 antagonist
KIT ⁷²	KIT	Mutation	GIST, mastocytosis, leukaemia	* KIT inhibitors
PDGFRA and PDGFRB	PDGFR	Deletion, rearrangement or amplification	Haematological cancer, GIST, sarcoma and brain cancer	* PDGFR inhibitors
Anaplastic lymphoma kinase (ALK) ^{73,74,75}	ALK	Rearrangement or mutation	Lung cancer and neuroblastoma	* ALK inhibitors
RET	RET	Rearrangement or mutation	Lung cancer and thyroid cancer	* RET inhibitors
ROS1 (REF. 75)	ROS1	Rearrangement	Lung cancer and cholangiocarcinoma	* ROS1 inhibitors
NOTCH1 and NOTCH2	Notch	Rearrangement or mutation	Leukaemia and breast cancer	* Notch signalling pathway inhibitors

Gene alterations and molecularly targeted drugs
(Reviewed by Simon et al, 2013)

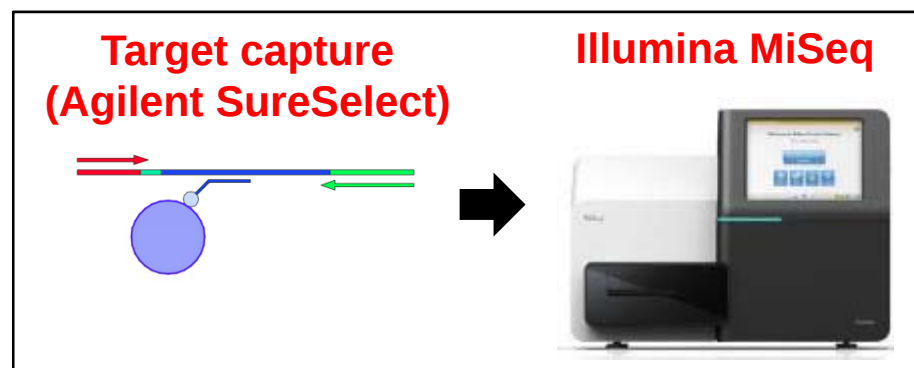
- If one assay simply detects these, ...

Clinical sequencing for cancer in National Cancer Center, Japan

– Collaboration by doctors and scientists –



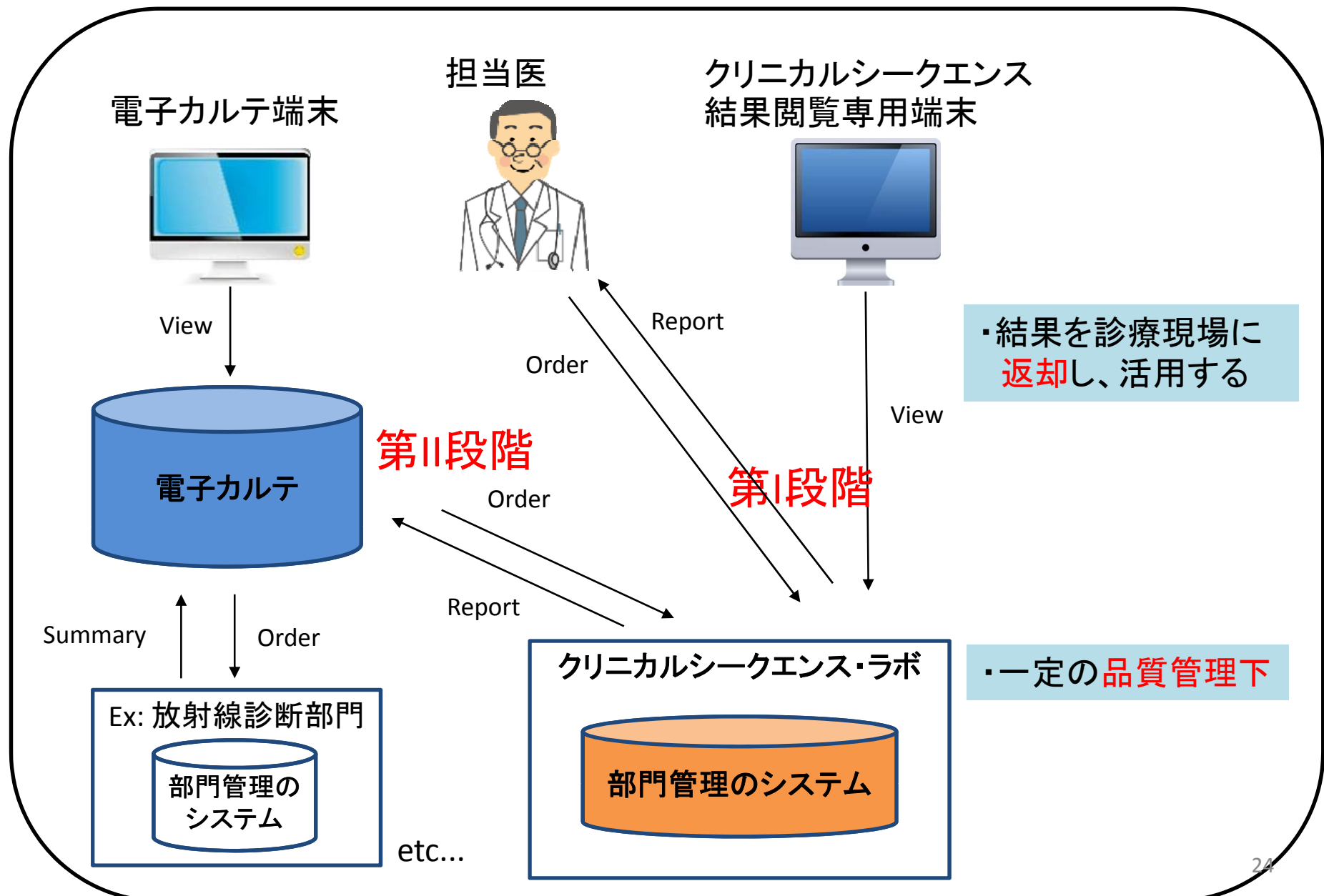
Platform and genes



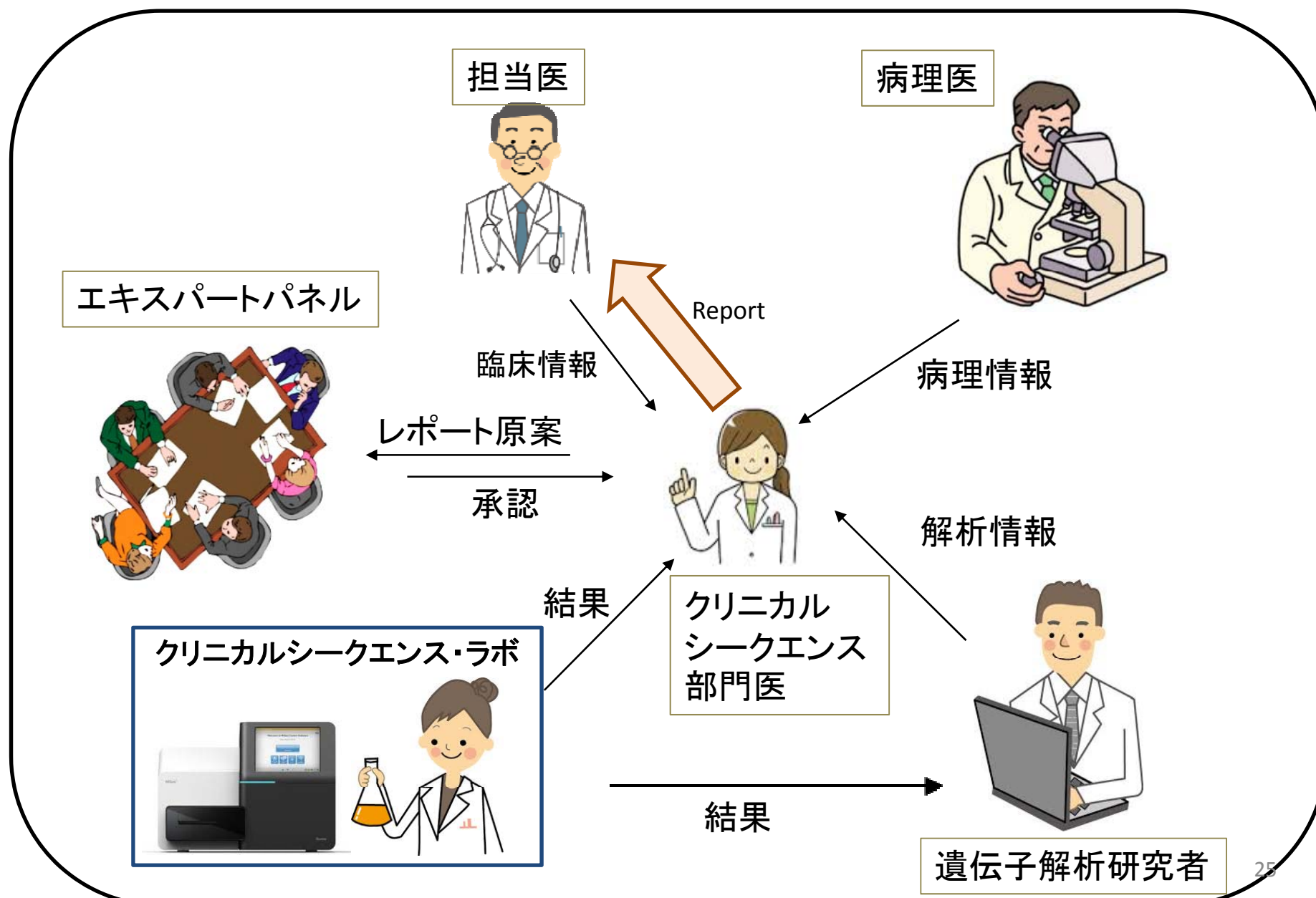
NCC oncopanel v2

Mutation / Amplification				Fusion	
ABL1	CREBBP	IGF1R	NOTCH1	ROS1	ALK
AKT1	CTNNB1	IGF2	NOTCH2	SETD2	RET
AKT2	CUL3	IL7R	NOTCH3	SMAD4	ROS1
AKT3	DDR2	JAK1	NRAS	SMARCA4	FGFR2
ALK	EGFR	JAK2	NRG1	SMO	FGFR3
APC	ENO1	JAK3	NT5C2	STAT3	AKT3
ARID1A	EP300	KEAP1	PALB2	STK11	BRAF
ARID2	ERBB2	KIT	PBRM1	TP53	RAF1
ATM	ERBB3	KRAS	PDGFRA	TSC1	NOTCH1
AXIN1	ERBB4	MAP2K1	PDGFRB	VHL	NRG1
BAP1	EZH2	MAP2K4	PIK3CA		
BARD1	FBXW7	MAP3K1	PIK3R1		
BCL2L11	FGFR1	MAP3K4	PTCH1		
BRAF	FGFR2	MDM2	PTEN		
BRCA1	FGFR3	MET	RAC1		
BRCA2	FGFR4	MTOR	RAC2		
CCND1	FLT3	MYC	RAD51C		
CDK4	HRAS	MYCN	RAF1		
CDKN2A	IDH1	NF1	RB1		
CHEK2	IDH2	NFE2L2	RET		

ゲノム医療に向けて： 中央病院内 クリニカルシーケンス・ラボ(仮称) 運用計画(案)

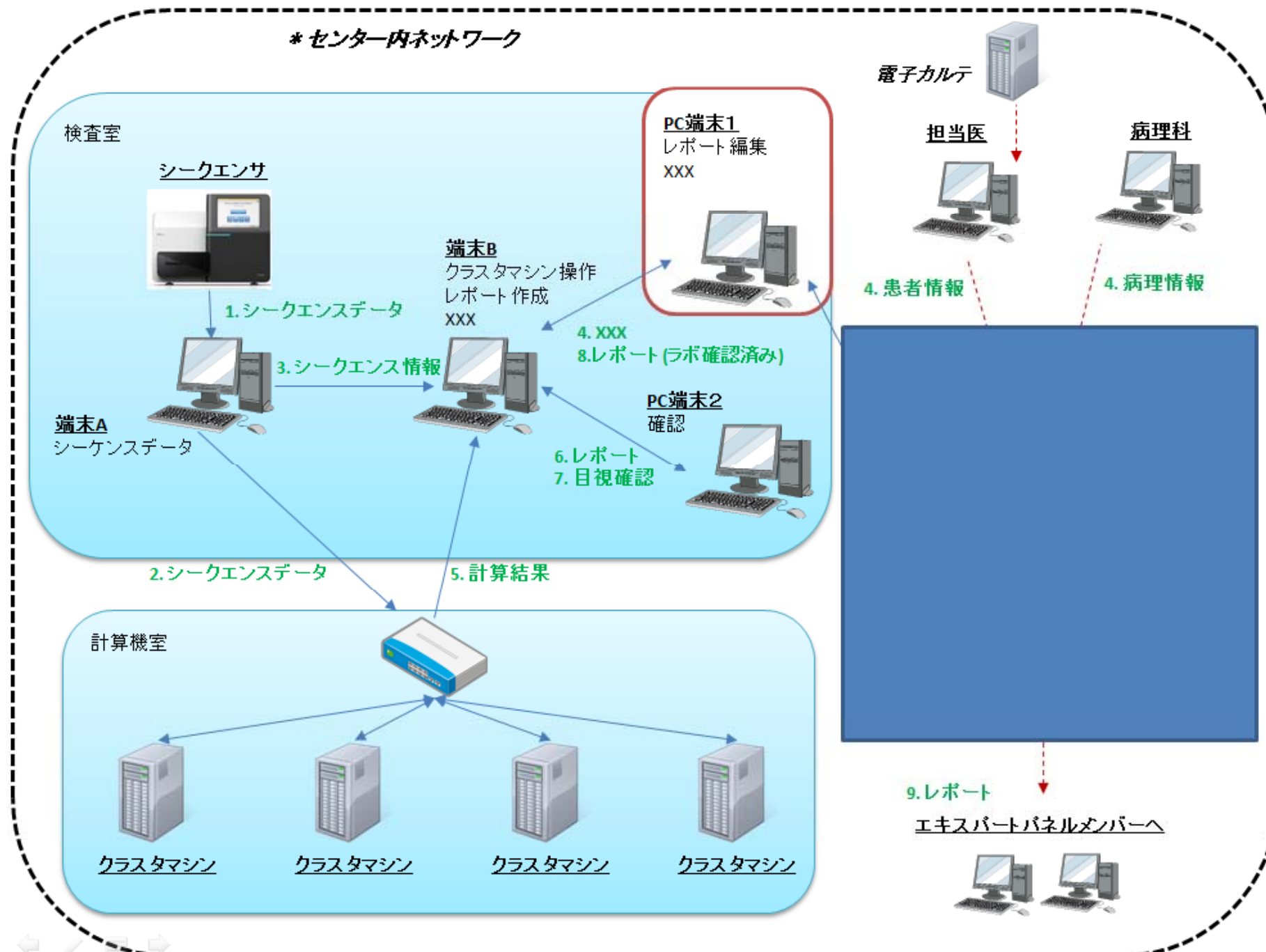


ゲノム医療に向けて:実施体制(暫定案)

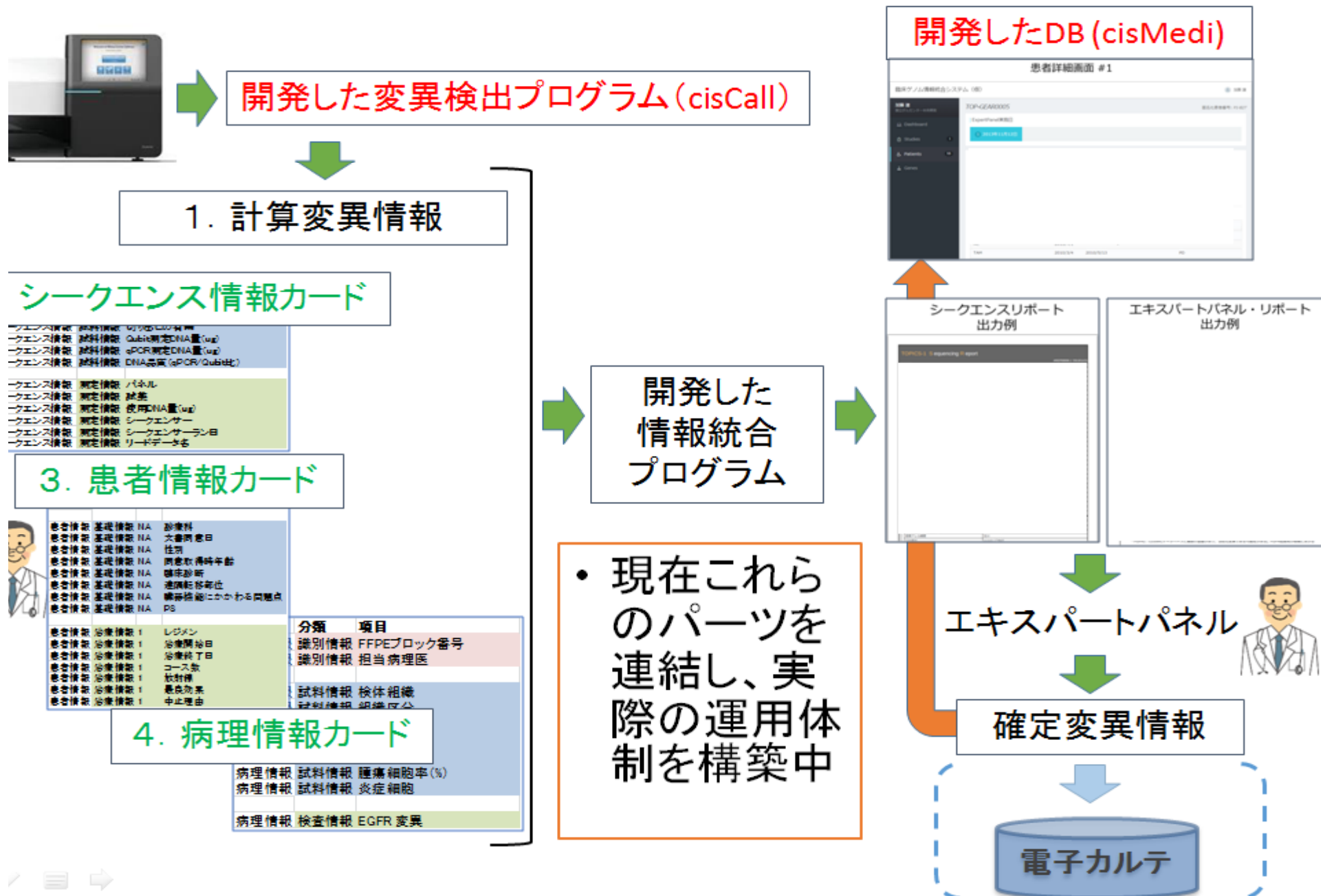


■データフロー(暫定案) * バイオインフォマティクス部門設計

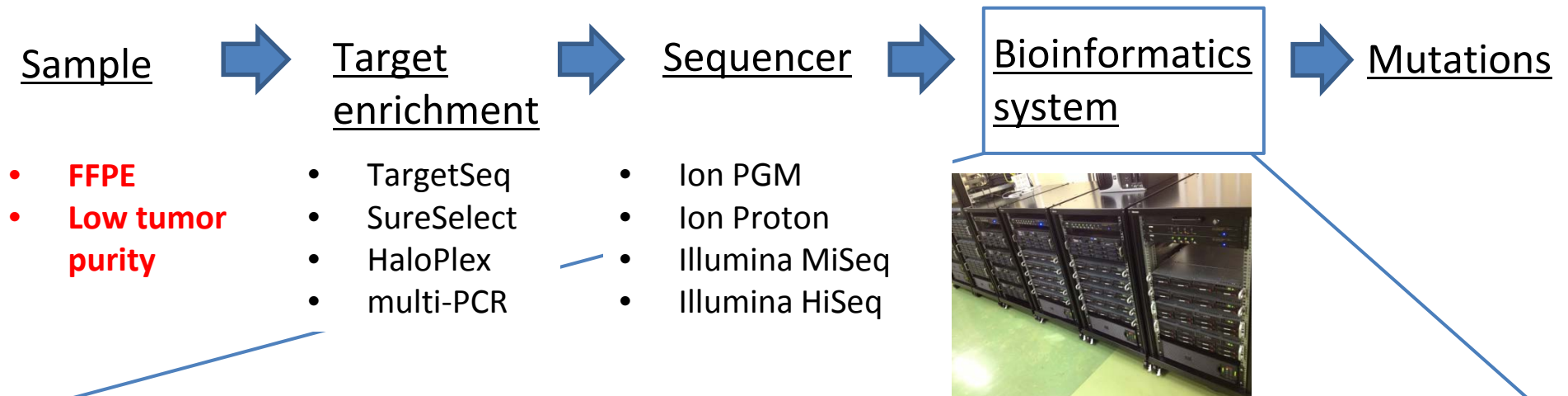
* センター内ネットワーク



システムモデル Ver. 1



開発した臨床シーケンス用変異検出プログラム (cisCall)



Clinical Sequence Call (*cisCall*) System

- Sequence data



- Read-QC report

cisMuton



- **SNV/indel** call table
 - Call-QC report

cisFusion



- **Fusion** call table
 - with breakpoints

cisCton



- **CNA** call table
 - with visualization

SNV

– single nucleotide variation –
and indel

Principle of the detection method

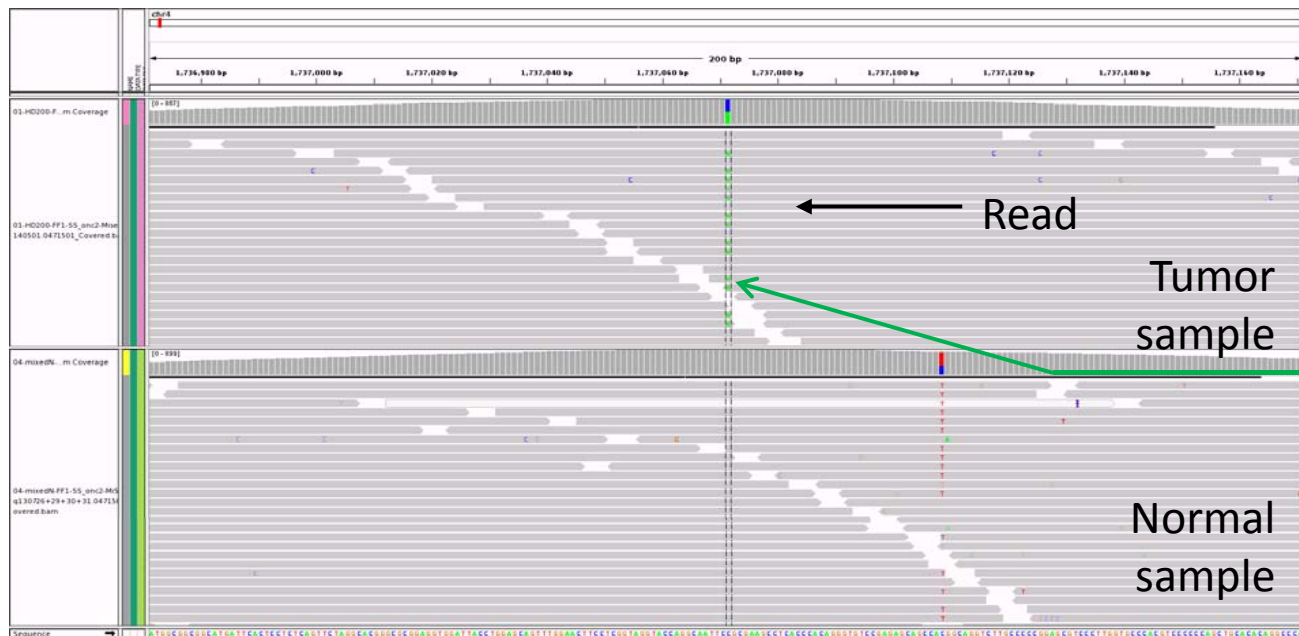
Fragmented DNAs are sequenced by NGS

G C C C G G
A G C C C G
A A G C C C

Variant in
tumor
sample

Human genome DNA sequence

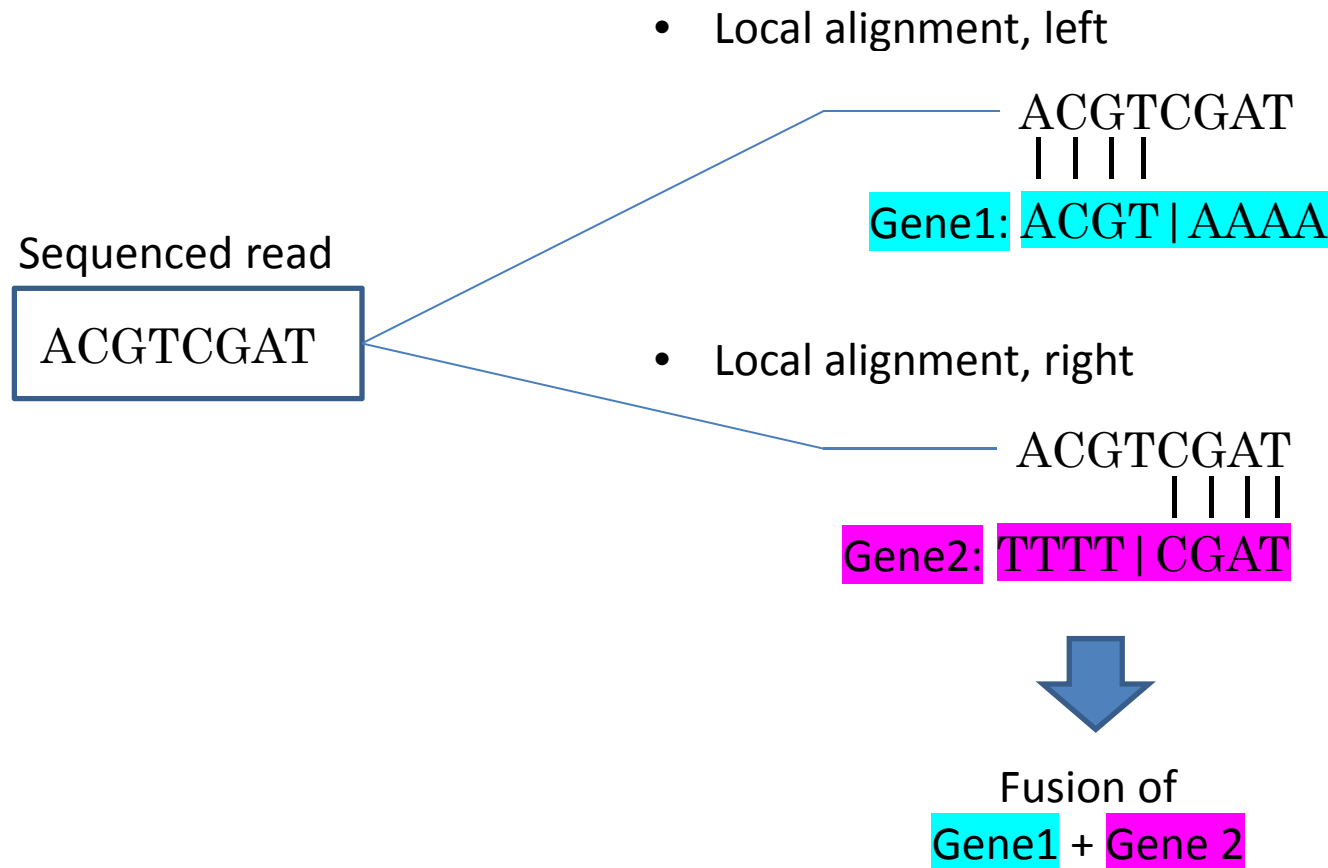
....AAAACCCCGGGGTTTT....



- Identify bases present in the tumor but absent in the (unmatched) normal

Fusion genes

Principle of the detection method

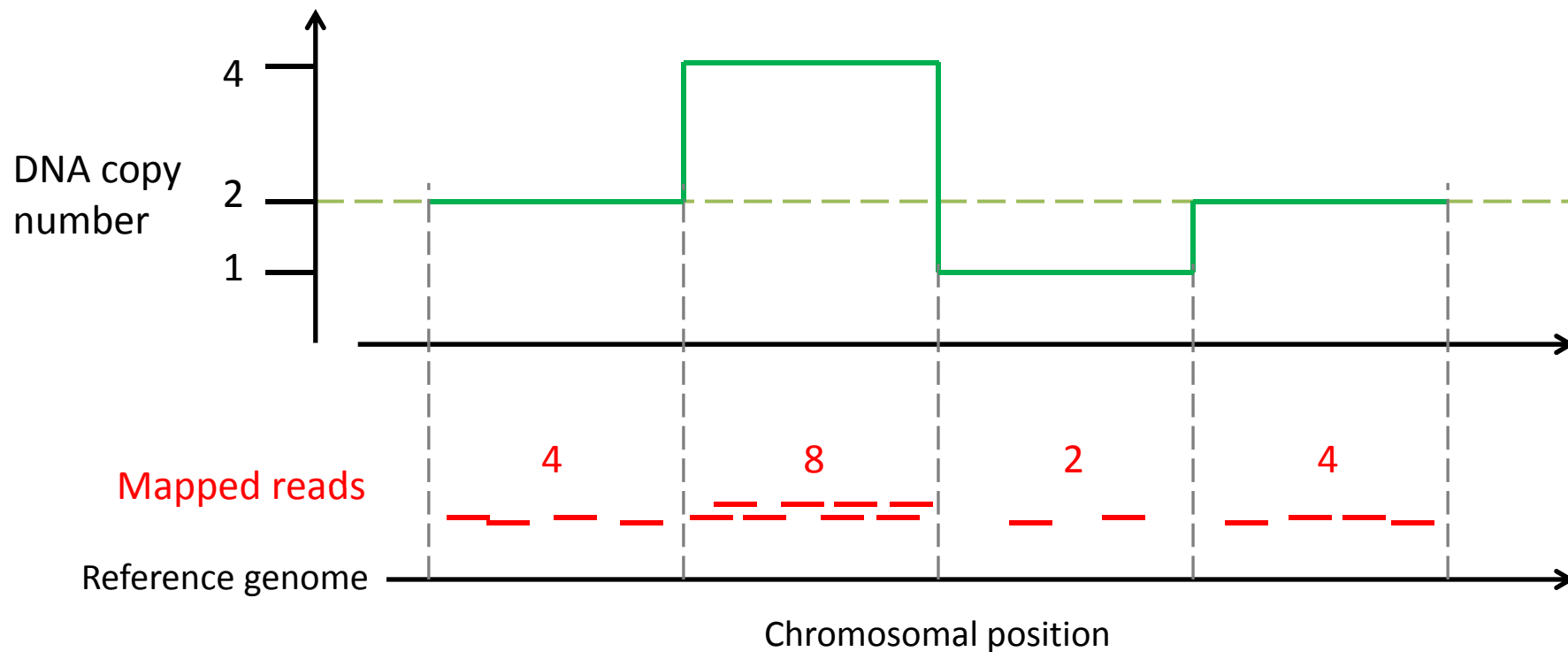


CNA

– copy number alteration –

Principle of the detection method

- Mapped sequence reads ↗ , copy numbers ↗
- Difference in depth between the tumor and normal

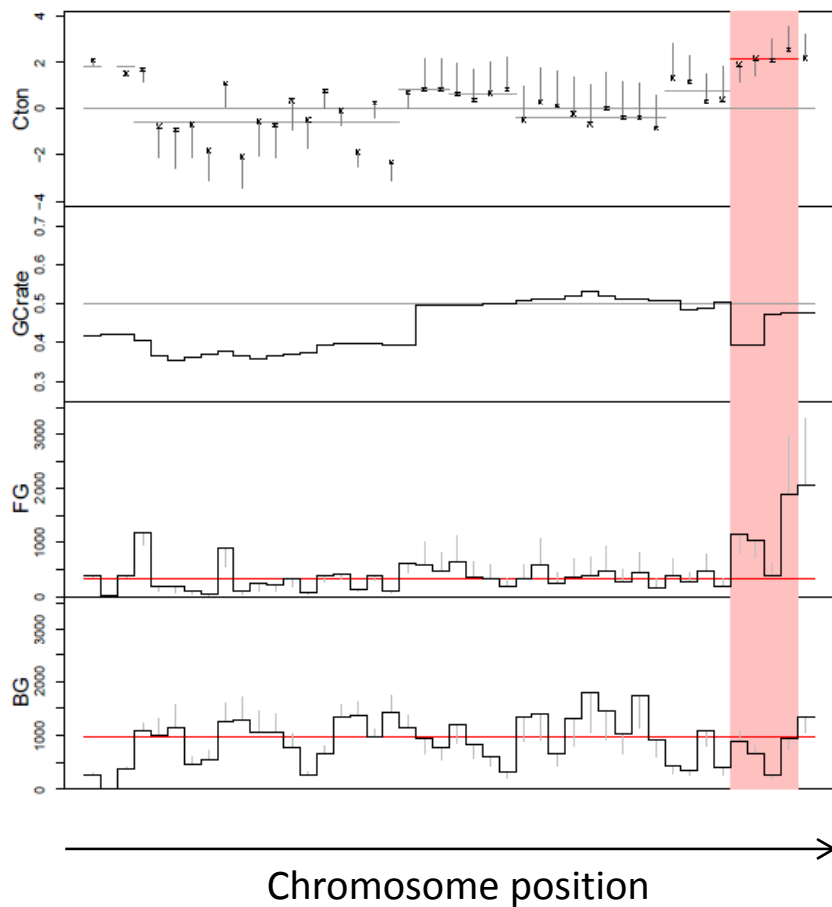


In practice, ...

This is the very place of bioinformatics

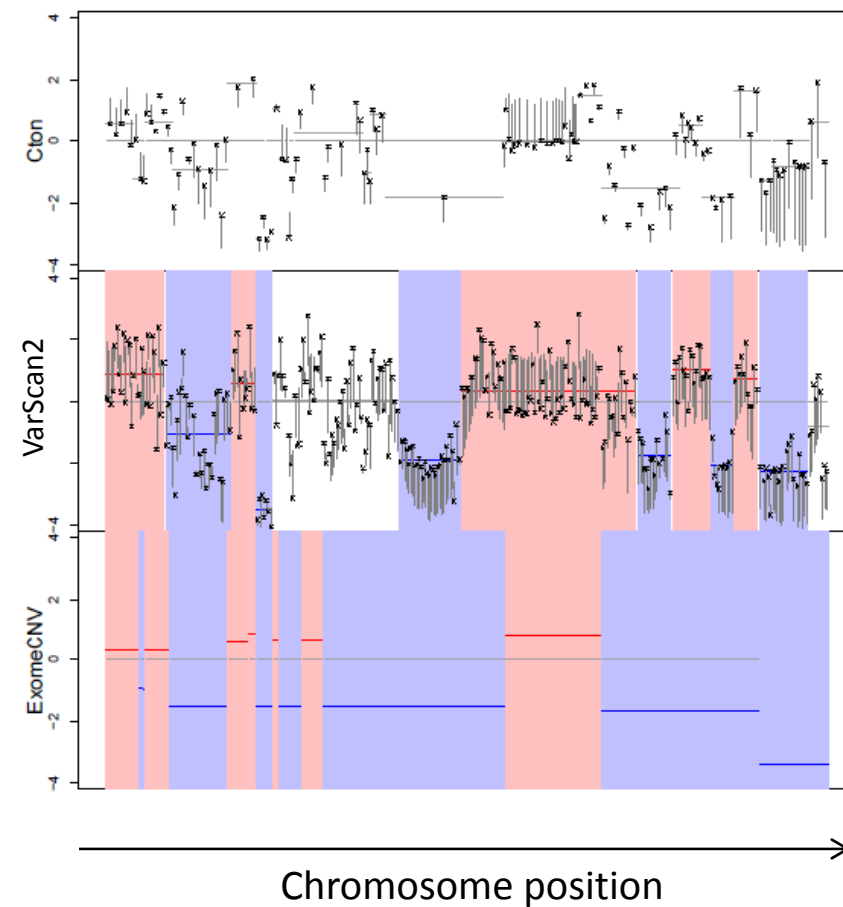
ポジコン

qPCR 増幅 LogRatio: 3.14



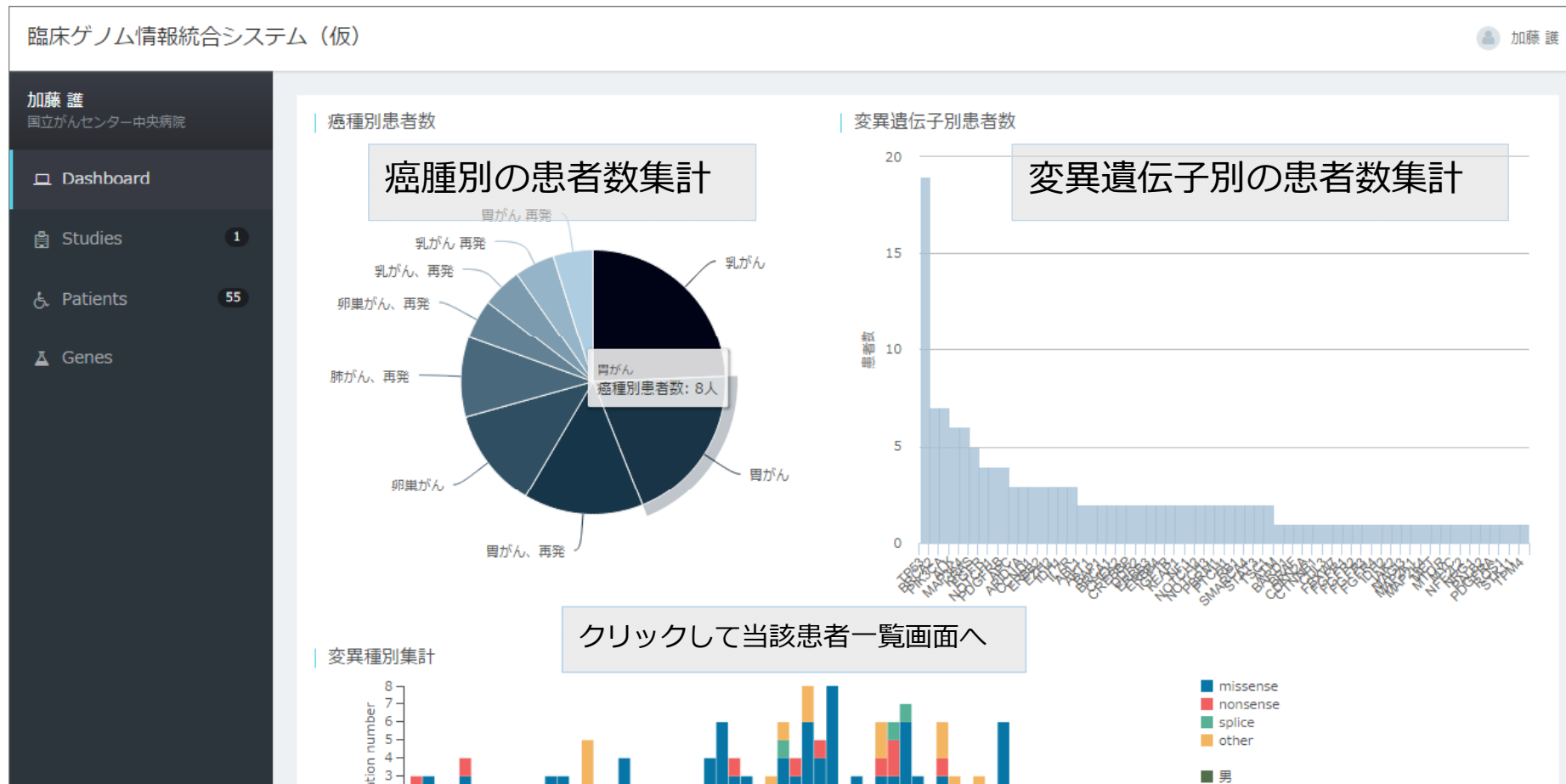
ネガコン

同一検体 FFPE-Frozen



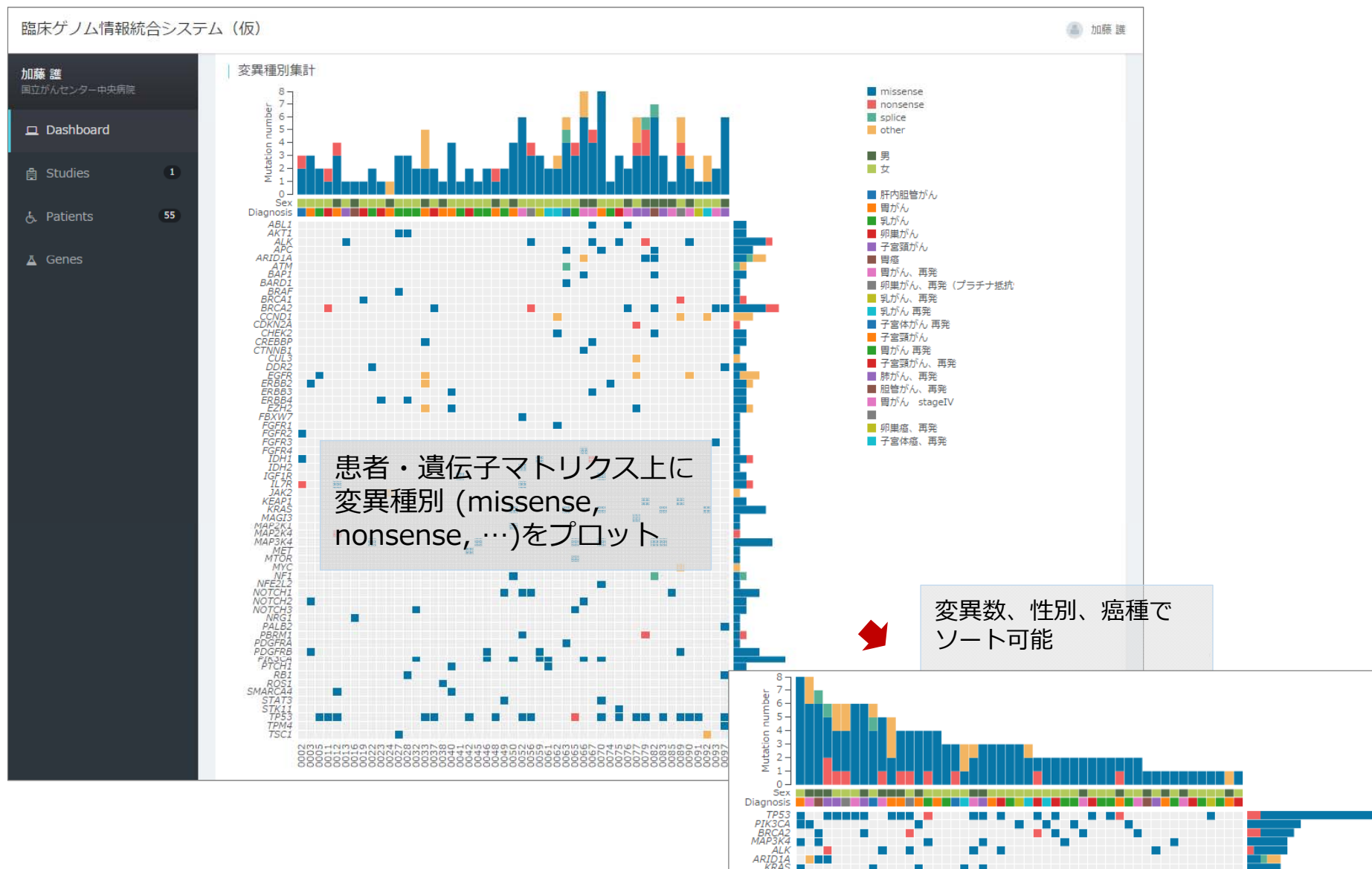
開発したDB (cisMedi): Database & post-calling system

ダッシュボード画面 #1 集計情報を表示



With Dr. Tsuchihara

ダッシュボード画面 #2



患者個別の情報を表示

臨床ゲノム情報統合システム（仮） 加藤 護

加藤 護
国立がんセンター中央病院

- Dashboard
- Studies 1
- Patients 55
- Genes

TOP-GEAR番号	患者実名	患者ID	匿名化患者番号	性別	臨床診断名	変異数	Mutated Genes
					肝内胆管がん	3	FGFR2, IDH1, IL
					胃がん	3	ERBB2, NOTCH2
					乳がん	2	EGFR, TP53
					卵巣がん	2	BRCA2, TP53
					胃がん	4	IL7R, MAP2K4, S
					子宮頸がん	1	ALK
					胃癌	1	NRG1
					卵巣がん	1	BRCA1
					乳がん	2	DDR2, MAP3K4
					卵巣がん		
					胃がん	1	JAK2
					乳がん	3	AKT1, BRAF, TS
					乳がん	3	AKT1, ERBB4, R
					乳がん	2	NOTCH3, PIK3C
					胃がん	5	CREBBP, EGFR, I
					卵巣がん	2	BRCA2, TP53
					胃がん	1	ROS1
					胃がん	4	ERBB3, EZH2, P
					乳がん	1	IGF1R
					卵巣がん	2	MET, TP53
					乳がん	1	MAP3K4

クリックして各変異遺伝子の
詳細画面へ

クリックして各患者
の詳細画面へ

患者詳細画面 #1

臨床ゲノム情報統合システム（仮）

加藤 謙

加藤 謙
国立がんセンター中央病院

Dashboard

Studies 1

Patients 55

Genes

匿名化患者番号:

ExpertPanel実施日

2013年11月12日

エキスパートパネル・リポートを出力

臨床情報

📄 Expert Panel報告書を出力

性別: 女

同意取得時年齢:

診療科:

担当医:

検体番号:

文書同意日:

臨床診断:

TNM分類: 再発

遠隔転移部位:
☒肝 ☐肺 ☐腹膜 ☒リンパ節 ☐骨 ☐脳 ☐その他()

PS: 1

保険:

臓器機能にかかわる問題点: なし

前治療

レジメン	放射線	治療開始日	治療終了日	コース数	最良総合効果	中止理由 ▼
PD						

エキスパートパネル・レポート 出力例

Expert Panel報告書		Expert Panel日:	
TOP-GEAR番号:	検体番号:	性別: 女	
同意取得時年齢:	診療科:		
文書同意日: ↓	臨床診断:		

遠隔転移部位:	☒ 肝臓	☐ 肺	☐ 腹膜	☐ LN	☐ 骨	☐ 脳	☐ その他 ()
---------	--	----------------------------	-----------------------------	-----------------------------	----------------------------	----------------------------	----------------------------------

PS: 1	保険:	臓器機能にかかわる問題点:
-------	-----	---------------

前治療

レジメン	放射線	治療開始日	治療終了日	コース数	最良総合効果	中止理由

検体情報				
検体組織	採取法	組織型	切片の大きさ (cm)	腫瘍細胞率 (%)
肝臓 (原発)	手術	腺癌		50

* 広範な壊死や固定不良は認められない。

Qubit測定DNA量 (ug)	DNA品質 (qPCR/Qubit比)
1.62	0.794499

遺伝子異常情報				
変異遺伝子	変異アレル頻度	CDS変化	アミノ酸変化	COSMIC ID
IDH1	33.4	Exon4:c.C394T	p.R132C	28747
IL7R	52.7	Exon5:c.T603A	p.Y201X	
FGFR2	69.4	Exon7:c.G870T	p.W290C	1346285

Expert Panelからの意見

- IDH1 : 既知の機能獲得変異である。対応する治療なし。
- IL7R : 短縮型変異のため、機能欠失変異と考えられる。対応する治療なし。
- FGFR2 : COSMICデータベースに複数の登録があり、活性化変異である可能性がある。FGFR阻害剤が候補にあがる

Conclusions

- **Bioinformatics – greedy and cloudy discipline**
 - ✓ Originally, it's generated from biophysics + molecular evolution
 - Alignment of protein (amino acid) sequences
 - Homology search
 - ✓ Expanding into many types of massive biological data
- **The essence – signal from noise** (a needle in a haystack)
 - ✓ Discovery > proof
 - Relaxing methodological rigorousness
 - ✓ **Pragmatism** – take in whatever method if useful
 - No theoretical basis...?
- Minimally required skills are NOT many
 - ✓ **Programming: linux, perl, R, SQL**
 - ✓ **Biology: molecular biology, genome biology**
- Can divide tasks into: algorithm design + program implementation
- Two study types: tool development + computational biology
- **Expanding from science to medicine**
 - ✓ Bioinformatics for clinical sequencing

END