

# ビッグデータ解析の一例としての トランスクリプトーム解析とその周辺

東京大学・大学院農学生命科学研究科  
アグリバイオインフォマティクス教育研究プログラム

門田幸二(かどた こうじ)

kadota@iu.a.u-tokyo.ac.jp

<http://www.iu.a.u-tokyo.ac.jp/~kadota/>

# 自己紹介 (アグリバイオインフォマティクス)

バイオインフォマティクスという分野にいるヒトです。簡単にいうと、ビッグデータをサクサク解析する分野です。頂いたお題は「ビッグデータの扱い方」

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月～ 東京大学・大学院農学生命科学研究科  
 アグリバイオインフォマティクス人材養成プログラム(科学技術振興調整費: 2004/10-2009/3)  
 アグリバイオインフォマティクス教育研究プログラム(特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
<b>合計</b>	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20

1科目以上の合格者数

アグリ(agriculture)がついている理由は、拠点が農学部だから。

# 自己紹介 (アグリバイオインフォマティクス)

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月～ 東京大学・大学院農学生命科学研究科  
 アグリバイオインフォマティクス人材養成プログラム(科学技術振興調整費: 2004/10-2009/3)  
 アグリバイオインフォマティクス教育研究プログラム(特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
<b>合計</b>	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20

1科目以上の合格者数

2014(平成26年度)以降は外部  
予算が切れ、かなり厳しい状況

# 自己紹介 (アグリバイオインフォマティクス)

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月~ 東京大学・大学院農学生命科学研究科  
 アグリバイオインフォマティクス人材養成プログラム (科学技術振興調整費: 2004/10-2009/3)  
 アグリバイオインフォマティクス教育研究プログラム (特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
<b>合計</b>	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20

1科目以上の  
合格者数

①受講人数は増加傾向。  
ビッグデータを扱いたいとい  
う希望者数の増加を反映?!

# 自己紹介 (アグリバイオインフォマティクス)

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月～ 東京大学・大学院農学生命科学研究科  
アグリバイオインフォマティクス人材養成プログラム(科学技術振興調整費: 2004/10-2009/3)  
アグリバイオインフォマティクス教育研究プログラム(特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
<b>合計</b>	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コ マ数	3	?	?	8	8	5	5	11	13	14	18	20



# 自己紹介 (アグリバイオインフォマティクス)

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月～ 東京大学・大学院農学生命科学研究科  
アグリバイオインフォマティクス人材養成プログラム(科学技術振興調整費: 2004/10-2009/3)  
アグリバイオインフォマティクス教育研究プログラム(特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
<b>合計</b>	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20



③担当講義時間は着実に増加(20コマ×1.5hr = 30hr)。

# 自己紹介 (アグリバイオインフォマティクス)

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月～ 東京大学・大学院農学生命科学研究科  
 アグリバイオインフォマティクス人材養成プログラム(科学技術振興調整費: 2004/10-2009/3)  
 アグリバイオインフォマティクス教育研究プログラム(特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
合計	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20



# 自己紹介 (アグリバイオインフォマティクス)

ハンズオン講義(PCを用いた実習を含むバイオインフォマティクス講義)の実務派。大所高所的な話よりは実務的な話のほうが本当は得意です。

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月~ 東京大学・大学院農学生命科学研究科  
 アグリバイオインフォマティクス人材養成プログラム(科学技術振興調整費: 2004/10-2009/3)  
 アグリバイオインフォマティクス教育研究プログラム(特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
<b>合計</b>	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20

1科目以上の合格者数



# ハンズオン講義風景

講義室の後方はパイプ椅子。約130名。①私。



# ハンズオン...

臨床分野の方がイメージするハンズオンとは若干異なるかもしれませんが、実際に手を動かして学ぶという点では同じです。



## 日本臨床麻酔学会 第35回大会

The 35th Annual Meeting of the Japan Society For Clinical Anesthesia

HOME >

会長挨拶 >

開催概要 >

参加者へのご案内 >

事前登録 >

発表者へのご案内 >

日程表 >

**プログラム** >


演題登録 >


会場案内 >


宿泊案内 >


### プログラム


プログラムをPDFファイルにて公開しております。

> 会長講演・特別講演・海外招待講演 

> 招待講演・招請講演 


> 教育講演 


> シンポジウム 

> 日中麻酔会議 

> ハンズオンワークショップ 

> リフレresherコース 

> 共催セミナー 

> 一般演題 

> 特別企画 / 学会各賞授与式・講演 



# ビッグデータといえは

次世代シーケンサ (Next-Generation Sequencer; New Generation Sequencer; NGS) を多くのヒトがイメージします。



PacBio RS II



Illumina MiSeq



Illumina HiSeq

Togo picture gallery by DBCLS is Licensed under [a Creative Commons 表示 2.1 日本 \(c\)](https://creativecommons.org/licenses/by/2.1/ja/)

# ビッグデータといえは

シーケンサ(sequencer)とは、塩基配列を決定する機器(Platform)のことを指します。一度に大量の塩基配列を決定できる次世代の実験機器がNGSです。



PacBio RS II



Illumina MiSeq



Illumina HiSeq

NGS機器から得られる大量塩基配列データをNGSデータという。

# ビッグデータ ≡ NGSデータ

**DDBJ** (DNA Data Bank of Japan; 日本)  
**DRA** (DDBJ Sequence Read Archive)

**NCBI** (National Center for Biotechnology Information; 米国)  
**SRA** (Sequence Read Archive)

**EMBL-EBI** (英国)  
**ENA** (European Nucleotide Archive)



# ビッグデータ ≡ NGSデータ

NGS機器から得られる大量塩基配列データをNGSデータという。日米欧三極でアーカイブ(保管)する枠組み

**DDBJ** (DNA Data Bank of Japan; 日本)

**DRA** (DDBJ Sequence Read Archive)


国際塩基配列データベース  
のコンソーシアム (INSDC)

**NCBI** (National Center for  
Biotechnology Information; 米国)  
**SRA** (Sequence Read Archive)

**EMBL-EBI** (英国)  
**ENA** (European Nucleotide Archive)

# 最新情報はDDBJで...

データ量が膨大なので、ネットワーク的な観点からも、DDBJの利用が基本

 **DDBJ** (DNA Data Bank of Japan; 日本)  
**DRA** (DDBJ Sequence Read Archive)

国際塩基配列データベース  
のコンソーシアム (INSDC)

**NCBI** (National Center for  
Biotechnology Information; 米国)  
**SRA** (Sequence Read Archive)

**EMBL-EBI** (英国)  
**ENA** (European Nucleotide Archive)

# 最新情報はDDBJで...

DDBJのウェブサイト。①の講習会などでNGSデータに関する全般的な動向を把握可能。

English

Google™ カスタム検索 Search

DDBJの紹介 利用の手引き レポート・統計 FAQ お問い合わせ

第33回 DDBJing 講習会 in 東京 11月11日 × JST 東京本部別館

DDBJ Service

登録 Data Submission 検索・解析 Search / Analysis スパコン Super Computer アーカイブ ftp. ddbj.nig.ac.jp

Hot Topics ニュースアーカイブ

お知らせ データ公開 広報 メンテナンス 運用情報 一覧

- 2015.10.07 PDB 2015-09-26 公開
- 2015.10.02 getentry 不具合のお詫び
- 2015.10.02 Mass Submission System (MSS) 関連ツール更新のお知らせ
- 2015.09.29 DDBJ リリース 102.0, DAD リリース 72.0 完成
- 2015.09.24 「第33回 DDBJing 講習会 in 東京」開催のお知らせ
- 2015.09.24 UniProt 2015\_09 公開
- 2015.09.16 「初めての All-in-one 合同講習会」の様子が統合TVから公開されました

DDBJ DNA Data Bank of Japan

INSDC NCBI ENA/EBI International Nucleotide Sequence Database Collaboration

大学共同利用機関法人 情報・システム研究機構 国立遺伝学研究所

大学共同利用機関法人 情報・システム研究機構 Research Organization of Information and Systems

DBCLS Database Center for Life Science



# NGSデータのトレンド

①DBCLSという組織で提供しているDBCLS SRA上で、NGSデータの各種統計情報を眺める。

English

Google™ カスタム検索 Search

DDBJの紹介 利用の手引き レポート・統計 FAQ お問い合わせ

第33回 DDBJing 講習会 in 東京  
11月11日 × JST 東京本部別館

RSSを購読する  
DDBJ Twitter  
Mail Magazine

DDBJ  
INSDC  
NCBI ENA/EBI  
International Nucleotide Sequence Database Collaboration

大学共同利用機関法人 情報・システム研究機構  
NIG 国立遺伝学研究所

大学共同利用機関法人 情報・システム研究機構  
Research Organization of Information and Systems

DBCLS  
Database Center for Life Science

**1**

**DDBJ Service**

登録 Data Submission  
検索・解析 Search / Analysis  
スパコン Super Computer  
アーカイブ ftp. ddbj.nig.ac.jp

**Hot Topics** ニュースアーカイブ

お知らせ データ公開 広報 メンテナンス 運用情報 一覧

- 2015.10.07 PDB 2015-09-26 公開
- 2015.10.02 getentry 不具合のお詫び
- 2015.10.02 Mass Submission System (MSS) 関連ツール更新のお知らせ
- 2015.09.29 DDBJ リリース 102.0, DAD リリース 72.0 完成
- 2015.09.24 「第33回 DDBJing 講習会 in 東京」開催のお知らせ
- 2015.09.24 UniProt 2015\_09 公開
- 2015.09.16 「初めての All-in-one 合同講習会」の様子が統合TVから公開されました

# NGSデータのトレンド



DBCLS SRA

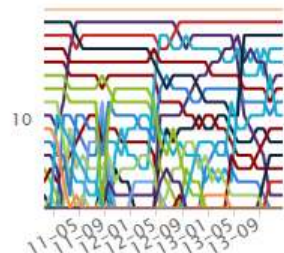
DISCOVER

Interesting & Available SRA Data

## Trends in SRA data

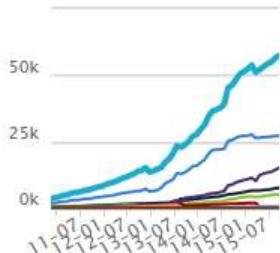
→ for more detail

### Species



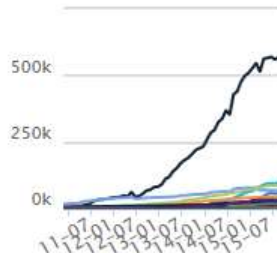
<i>Homo sapiens</i>	304177
<i>Mus musculus</i>	91408
human gut metagenome	40107
<i>Oryza sativa</i>	20649
soil metagenome	22369
human metagenome	20543
<i>Plasmodium falciparum</i>	19974
<i>Staphylococcus aureus</i>	19437
<i>Streptococcus pneumoniae</i>	19016
<i>Saccharomyces cerevisiae</i>	18142
<i>Danio rerio</i>	15847
<i>Drosophila melanogaster</i>	14720
<i>Salmonella enterica subsp. enterica</i>	14677
mouse gut metagenome	10822
gut metagenome	9410
others	442625
<b>Total</b>	<b>1002823</b> (experiments)

### Study Type



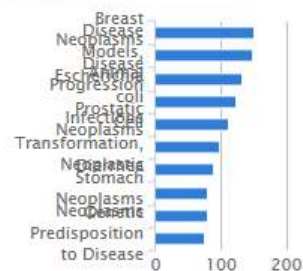
Whole Genome Sequencing	27525
Other	15807
Transcriptome Analysis	7976
Metagenomics	5238
Population Genomics	722
Epigenetics	669
Exome Sequencing	174
RNASeq	148
Cancer Genomics	87
Resequencing	57
Pooled Clone Sequencing	35
Gene Regulation Study	11
Synthetic Genomics	9
<b>Total</b>	<b>58458</b> (studies)

### Platform



Illumina HiSeq 2000	576935
Illumina MiSeq	98332
454 GS FLX Titanium	87703
Illumina Genome Analyzer II	68666
Illumina HiSeq 2500	62668
Illumina Genome Analyzer Ix	45046
unspecified	32244
454 GS FLX	27644
Illumina Genome Analyzer	16613
454 GS Junior	12782
...	
<b>Total</b>	<b>1002823</b> (experiments)

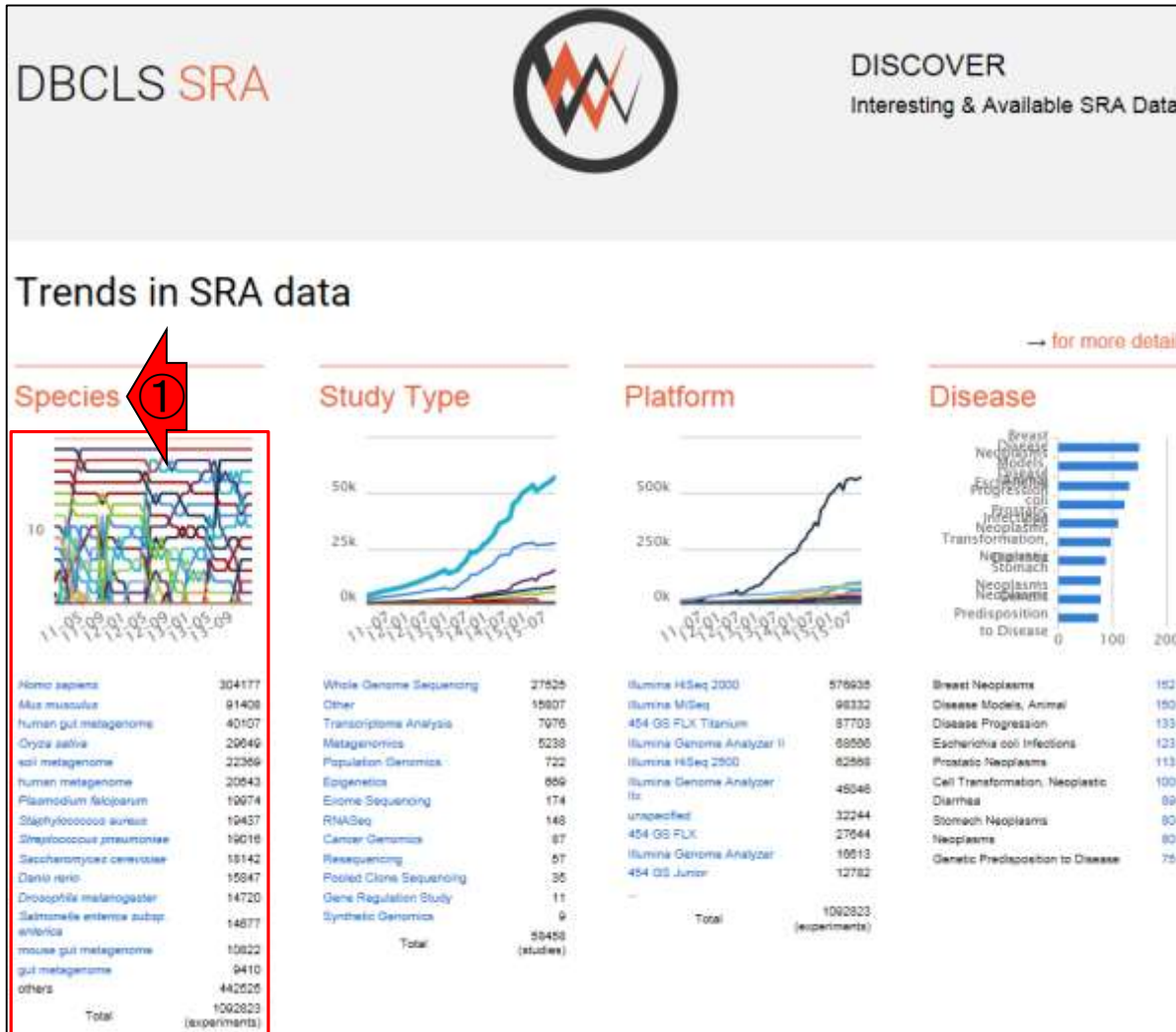
### Disease



Breast Neoplasms	152
Disease Models, Animal	150
Disease Progression	133
Escherichia coli Infections	123
Prostatic Neoplasms	113
Cell Transformation, Neoplastic	100
Diarrhea	89
Stomach Neoplasms	80
Neoplasms	80
Genetic Predisposition to Disease	75

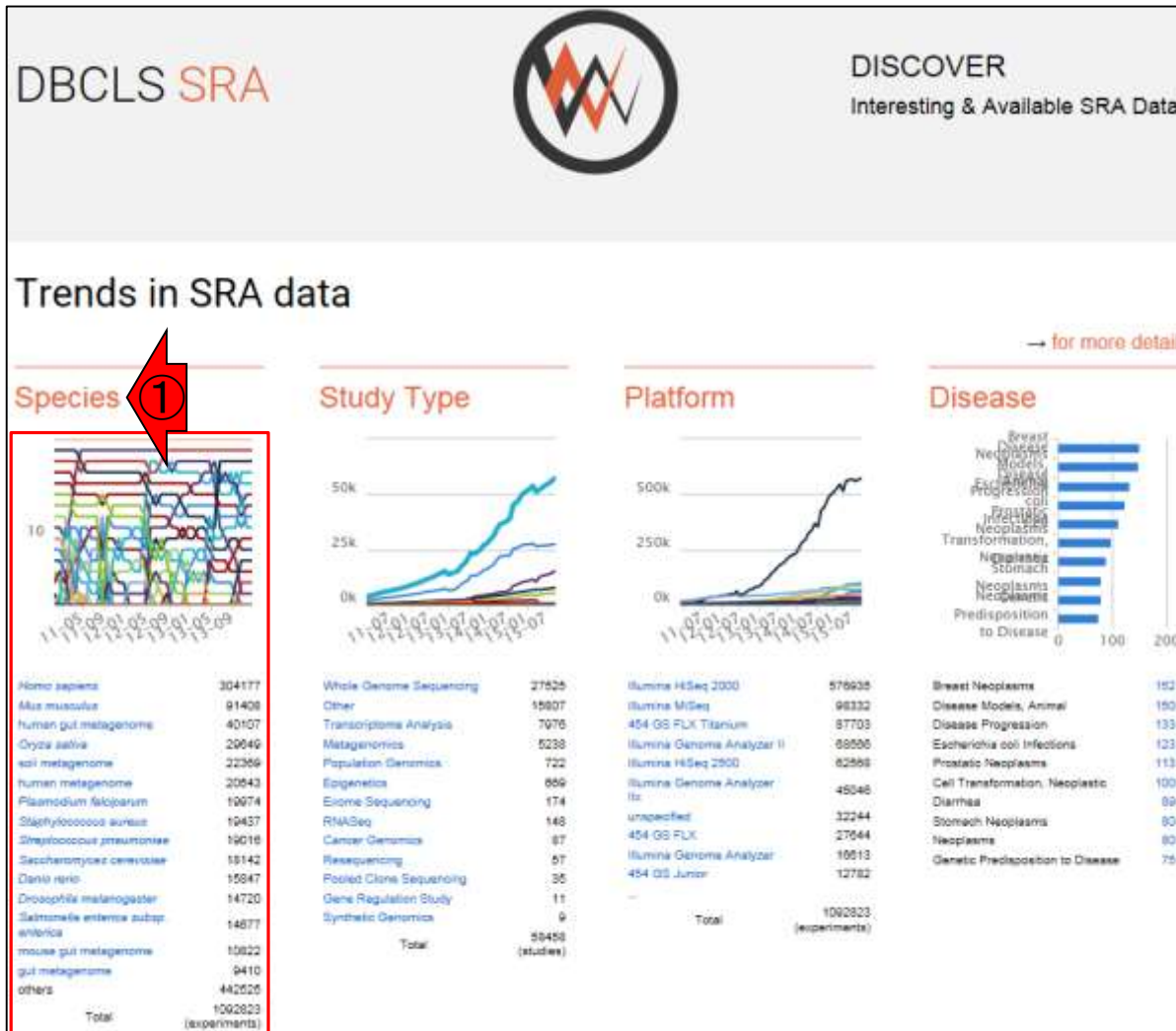
# NGSデータのトレンド

① Species (生物種) の項目で、どのような生物種のNGSデータが取得されているかを概観。



# NGSデータのトレンド

①Species(生物種)の項目で、どのような生物種のNGSデータが取得されているかを概観。②ヒト(*H. sapiens*)、③ヒトの腸内細菌のメタゲノム。腸内にどのような細菌のゲノムがどの程度いるのかを知る目的でNGSを利用。



①状態の違いを②特徴づける  
細菌の種類同定などに利用


# メタゲノム

術後障害		メタゲノム			
		細菌A	細菌B	細菌C	...
あり	患者1	10%	15%	8%	
あり	患者2	12%	13%	10%	
あり	患者3	11%	16%	7%	
なし	患者4	25%	18%	8%	
なし	患者5	27%	11%	9%	
なし	患者6	28%	16%	6%	

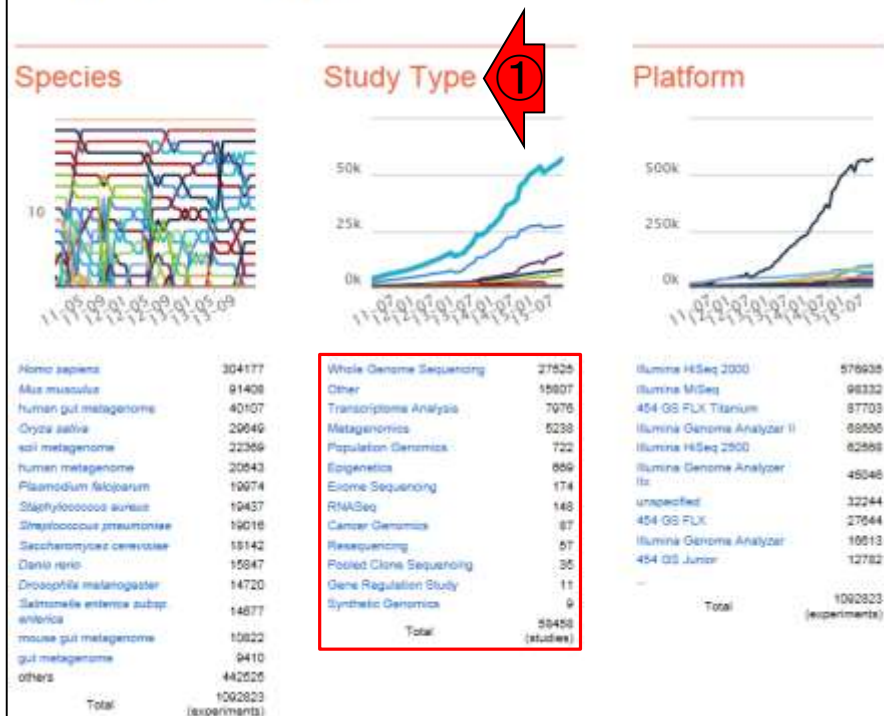


①どのような種類のNGSデータ解析がなされているか(Study Type)?

# NGSデータのトレンド

DBCLS SRA  DISCOVER  
Interesting & Available SRA Data

## Trends in SRA data



→ for more detail

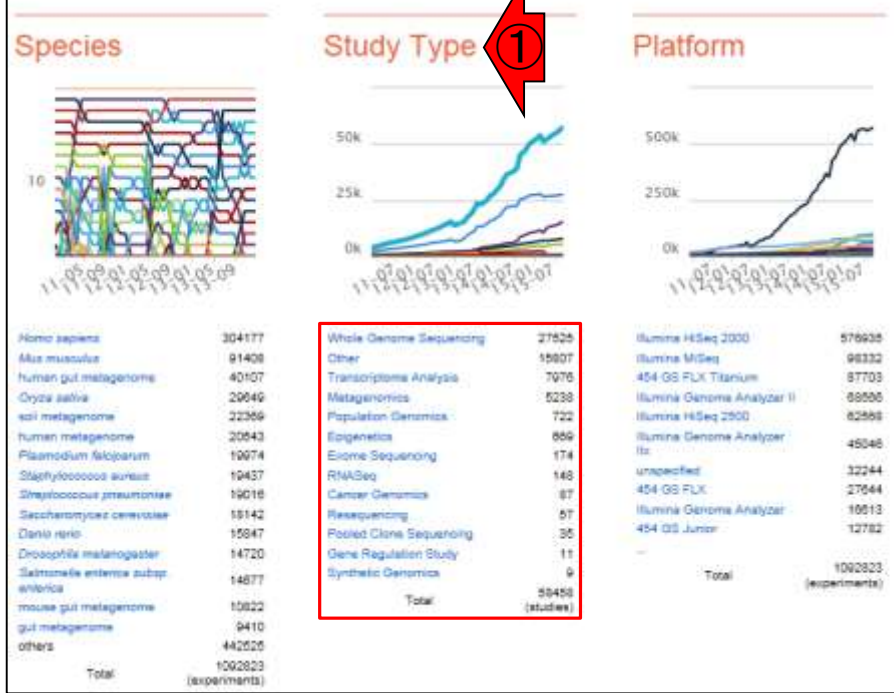
Whole Genome Sequencing	27525
Other	15807
Transcriptome Analysis	7976
Metagenomics	5238
Population Genomics	722
Epigenetics	669
Exome Sequencing	174
RNASeq	148
Cancer Genomics	87
Resequencing	57
Pooled Clone Sequencing	35
Gene Regulation Study	11
Synthetic Genomics	9
<b>Total</b>	<b>58458</b>

# NGSデータのトレンド

①どのような種類のNGSデータ解析がなされているか(Study Type)?  
 ②全ゲノム配列決定(Whole Genome Sequencing; WGS)を目的とした研究が圧倒的に多い。



## Trends in SRA data



→ for more detail

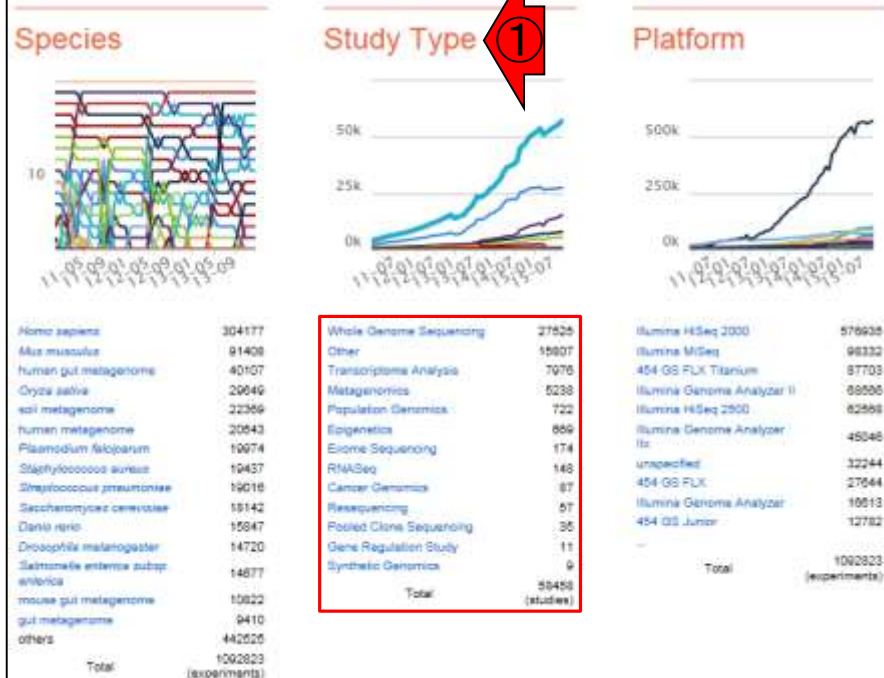
Whole Genome Sequencing	27525
Other	15807
Transcriptome Analysis	7976
Metagenomics	5238
Population Genomics	722
Epigenetics	669
Exome Sequencing	174
RNASeq	148
Cancer Genomics	87
Resequencing	57
Pooled Clone Sequencing	35
Gene Regulation Study	11
Synthetic Genomics	9
<b>Total</b>	<b>58458</b>

# NGSデータのトレンド

①どのような種類のNGSデータ解析がなされているか(Study Type)?  
 ②全ゲノム配列決定(Whole Genome Sequencing; WGS)を目的とした研究が圧倒的に多い。③その次に多いのが**トランスクリプトーム解析**。本講演のタイトル。メタゲノムではない。

DBCLS SRA  DISCOVER  
 Interesting & Available SRA Data

## Trends in SRA data



→ for more detail

② Whole Genome Sequencing	27525
Other	15807
③ Transcriptome Analysis	7976
Metagenomics	5238
Population Genomics	722
Epigenetics	669
Exome Sequencing	174
RNASeq	148
Cancer Genomics	87
Resequencing	57
Pooled Clone Sequencing	35
Gene Regulation Study	11
Synthetic Genomics	9
<b>Total</b>	<b>58458</b>

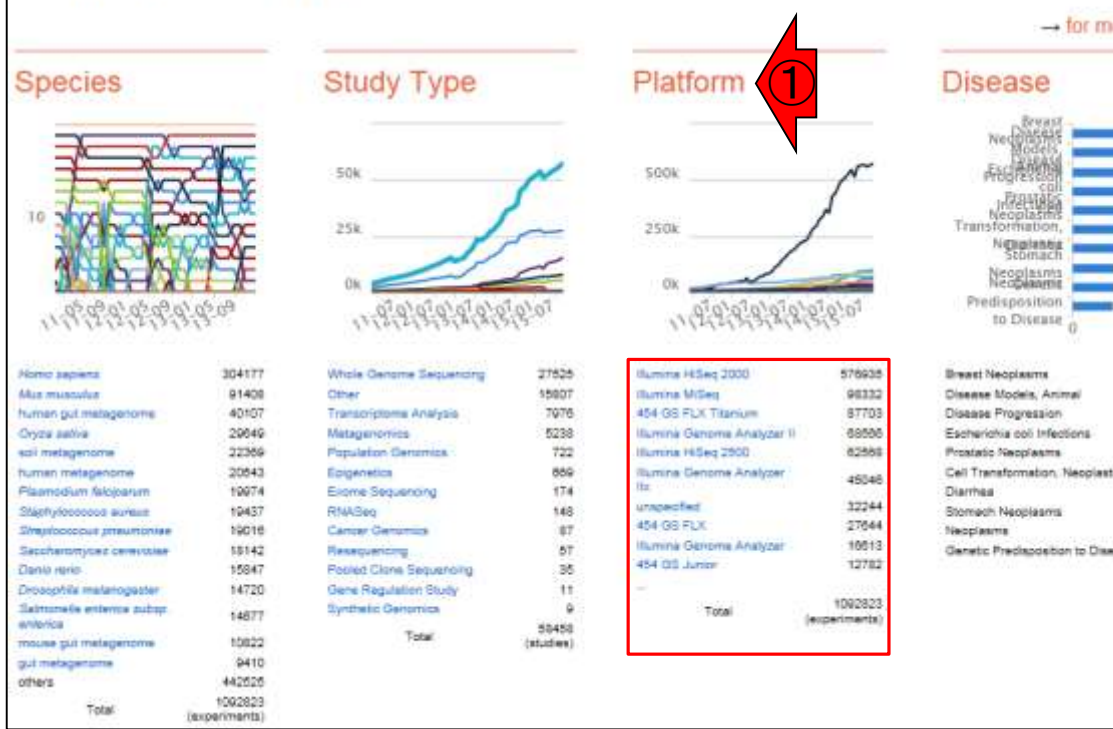


# NGSデータのトレンド

よく使われてきたNGS機器は、Illumina(イルミナ)社製のものがほとんどを占める

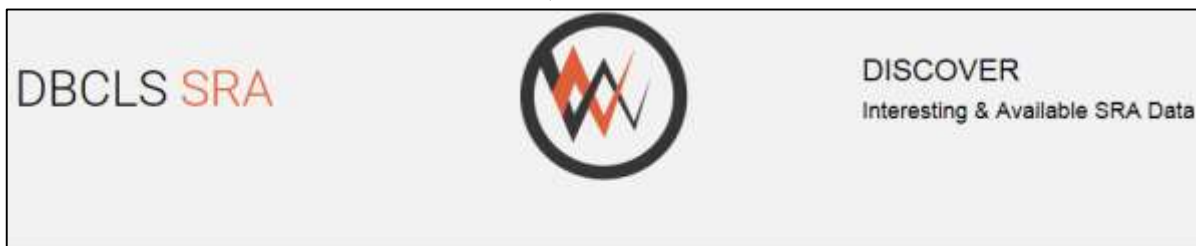
DBCLS SRA  DISCOVER  
Interesting & Available SRA Data

## Trends in SRA data



<u>Illumina HiSeq 2000</u>	576935
<u>Illumina MiSeq</u>	98332
454 GS FLX Titanium	87703
<u>Illumina Genome Analyzer II</u>	68566
<u>Illumina HiSeq 2500</u>	62568
<u>Illumina Genome Analyzer Ix</u>	45046
unspecified	32244
454 GS FLX	27644
<u>Illumina Genome Analyzer</u>	16613
454 GS Junior	12782
Ion Torrent PGM	7848
<u>Illumina HiSeq 1000</u>	6511
454 GS FLX+	6206
PacBio RS II	5747
PacBio RS	4049
Helicos HeliScope	3822
Complete Genomics	2811
AB 5500xl Genetic Analyzer	2685

# NGSデータのトレンド



## Trends in SRA data



①	Illumina HiSeq 2000	576935
	Illumina MiSeq	98332
②	454 GS FLX Titanium	87703
	Illumina Genome Analyzer II	68566
	Illumina HiSeq 2500	62568
	Illumina Genome Analyzer Ix	45046
	unspecified	32244
	454 GS FLX	27644
	Illumina Genome Analyzer	16613
	454 GS Junior	12782
	AB SOLiD 4 System	9941
	Ion Torrent PGM	7848
	Illumina HiSeq 1000	6511
	454 GS FLX+	6206
	PacBio RS II	5747
	PacBio RS	4049
	Helicos HeliScope	3822
	Complete Genomics	2811
	AB 5500xl Genetic Analyzer	2685

# ビッグデータといえは

これらです。①Illumina HiSeq  
というNGS機器の大きさは…



PacBio RS II



Illumina MiSeq



Illumina HiSeq



# Illumina HiSeqと私



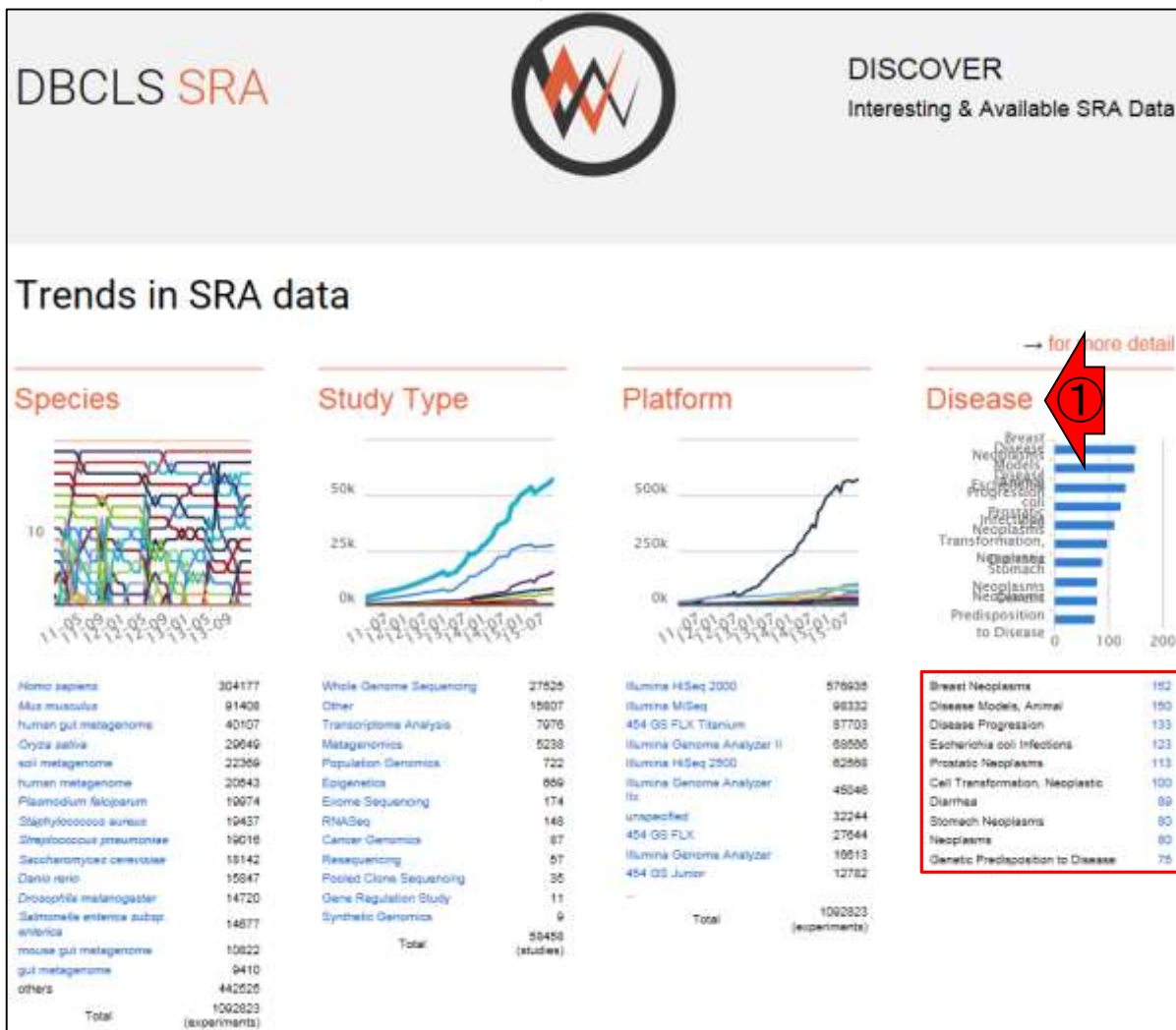
Illumina HiSeq



<https://www.facebook.com/illuminakk/>  
2014年7月22日のイルミナウェビナー

もちろんNGSは疾患(Disease)関連でも利用されています。

# NGSデータのトレンド



Breast Neoplasms	152
Disease Models, Animal	150
Disease Progression	133
Escherichia coli Infections	123
Prostatic Neoplasms	113
Cell Transformation, Neoplastic	100
Diarrhea	89
Stomach Neoplasms	80
Neoplasms	80
Genetic Predisposition to Disease	75

# NGS全般

NGS機器シェアトップのIlluminaのホームページ。①イルミナサポートプログラム。

illumina

予算申請用資料お申し込み

お問い合わせ

MyIllumina

Tools

アプリケーション システム インフォマティクス 臨床研究 受託サービス サイエンス サポート カンパニー

Search

メールニュース登録



Follow us:

## イルミナサポートプログラム

イルミナがお届けするサポートは、イルミナ School、ユーザーサポート、プロダクトケアの3つを軸に、専門知識と経験が豊富なスタッフが信頼をお届けします

詳細はこちら▶



イルミナ School



ユーザーサポート



プロダクトケア



Oncology Breakthrough  
イルミナブースイベント



日本人類遺伝学会 第60回大会  
イルミナ3つのブースイベント



最新疾患パネルが  
期間限定でお求めやすく



イルミナ  
サポートプログラム

製品検索

システムに適したキットを探す

ライブラリー調製&アレイキット セレクター  
試薬・消耗品一覧はこちら

ログイン

Username/Email

Password

オプション選択

ログイン

Facebook



イルミナ株式会社  
Facebook

# NGS全般

①イルミナiSchool。NGS関連全般の情報はここから入手できます。

illumina

予算申請用資料 お申し込み

お問い合わせ

Myillumina

Tools

アプリケーション

システム

インフォマティクス

臨床研究

受託サービス

サイエンス

サポート

カンパニー

Search



サポート » イルミナサポートプログラム

Follow us:

システム検討時から継続的なサポート

## イルミナのサポートプログラム

イルミナではシステム検討から導入後のサポートまで、様々な形でサポートをご提供します。専門知識を持つスタッフによるウェビナー、講習会、トレーニングに加え、日々の技術的質問に答えるテクニカルサポート、2年目以降の保守契約まで、お客様の製品ご使用の時期とニーズにあわせたサービスを心がけています。



コンサルティング  
ご提案



### イルミナ iSchool

- ウェビナー(新製品、技術紹介)
- 講習会、ワークショップ
- ユーザーによる発表/セミナー



### ユーザーサポート

- 製品キットセレクトター
- テクニカルサポート(電話およびメール)
- システム導入時オンサイトトレーニング
- 有償トレーニング

# NGS全般

①ウェビナーや②講習会など、サポート体制が充実していて素敵です。



illumina

予算申請用資料 お申し込み お問い合わせ Myllumina Tools

アプリケーション システム インフォマティクス 臨床研究 受託サービス サイエンス サポート カンパニー

Search

サポート » イルミナサポートプログラム » イルミナサポートプログラム - イルミナ iSchool

製品、技術、サイエンスの知識を強化する技術教育コース

## イルミナiSchool

ユーザーサポート プロダクトケア

イルミナiSchool はウェビナー、講習会、トレーニングの形式で、知識レベルにあわせた多彩なコースをご提供しています。製品導入をご検討されている方、および製品をお使いのユーザーを対象としています。ウェビナー と講習会は無料でご参加いただけます。

	① ウェビナー	② 講習会	トレーニング
概要	インターネットを介した短時間（30分から1時間）のオンラインセミナーです。録画しており、いつでも閲覧いただけます。	トピックにより半日から2日をめどに開催されるセミナー形式の座学コースです。内容によっては、ハンズオンで行うデータ解析実習を含みます。	システム操作からハンズオンのライブラリー調製まで、トピックにより2-3日のコースをご用意しています。
初級 NGSに興味がある、これから使いたいと思っている方	製品の使い方	ベーシック	システム操作
中級 製品を使い始めたユーザー	製品使用時のポイント	イルミナソリューションによる Sample to Answer	ライブラリー調製
上級 製品経験が豊富なユーザー	トラブルシューティング	3rdパーティツールを用いたアプリケーション	
プロフェッショナル 最先端のサイエンスにご興味をもつ方	ユーザーによるご講演	ユーザーによるご講演	



# NGS全般

①門田はウェビナーに計3回登場。「Illumina HiSeqと私」は、イルミナのfacebookから第3回ウェビナー後の写真を拝借させて頂きました。

illumina

アプリケーション システム インフォマティクス 臨床研究 受託サービス サイェンス サポート カンパニー

予算申請用資料お申し

illumina イルミナ株式会社さんが新しい写真2枚を追加しました  
2014年7月23日

7/22 16:00から行われた東京大学 門田先生の「トランスクリプトームデータ解析戦略2014」ウェビナー終了後のお写真です。  
130名を超える登録をいただきとても盛況でした。  
また、門田先生には、ウェビナー後にイルミナのラボで、NextSeqを見学いただきました。

先生ありがとうございました！  
先生のご講演録画、PDFはこちらから  
[http://www.illumina.co.jp/events/webinar\\_japan.ilmn?ws=ws](http://www.illumina.co.jp/events/webinar_japan.ilmn?ws=ws)

サポート » イルミナサポートプログラム » イルミナサポートプログラム - イルミナ iSchool



製品、技術、サイエンスの知識を強化する技術教育コース

## イルミナiSchool



ユーザーサポ

イルミナiSchool はウェビナー、講習会、トレーニングの形式で、知識レベルにあわせた多彩なコースをご提供 していただき製品をお使いのユーザーを対象としています。ウェビナー と講習会は無料でご参加いただけます。



ウェビナー

講習会

概要

インターネットを介した短時間（30分から1時間）のオンラインセミナーです。録画しており、いつでも閲覧いただけます。

トピックにより半日から2日をめどに開催されるセミナー形式の座学コースです。内容によっては、ハンズオンで行うデータ解析実習を含みます。

初級

NGSに興味がある、これから使いたいと思っている方

製品の使い方

ベーシック

中級

製品を使い始めたユーザー

製品使用時のポイント

イルミナソリューションによる Sample to Answer

上級

製品経験が豊富なユーザー

トラブルシュート

3rd パーティツールを用いたアプリケーション

プロフェッショナル

最先端のサイエンスにご興味をもつ方

ユーザーによるご講演

ユーザーによるご講演



# NGS原理

①NGS機器による塩基配列決定の原理を知りたい場合は、Youtubeもおススメ。

The screenshot shows a YouTube search interface for the query "ngs 原理". The search bar contains the text "ngs 原理" with a red arrow labeled "1" pointing to it. The search results are displayed below, with a red arrow labeled "2" pointing to the first result, "Illumina Sequencing Technology".

YouTube JP

ngs 原理

約 123 件の検索結果

フィルタ

**Illumina Sequencing Technology**

作者: Illumina Inc  
1 年前・視聴回数 192,473 回

This video provides an overview of the DNA sequencing workflow on an Illumina sequencer. The process begins with Nextera ...

5:04 HD

**Illumina Solexa Sequencing**

作者: Draven1983101  
5 年前・視聴回数 265,888 回

An animation describing the process of next generation sequencing on the Illumina platform.

1:38

**【NGS速習コース】3. データ解析基礎～3-1. R基礎1、3-2. R基礎2、3-3. R各種パッケージ**

作者: togotv  
10 か月前・視聴回数 461 回

本家はこちら <http://togotv.dbcls.jp/20141030.html> 本日の統合TVは、2014年9月1-12日に開催された、バイオインフォマティクス人材育成 ...

4:28:26

# NGS原理

①動画で原理を理解できるので便利な時代です。②関連動画も多数。

The image shows a YouTube video player interface. The video title is "Illumina Sequencing Technology" by Illumina Inc. The video progress is at 0:25 / 5:03. The video content includes a diagram of "Transposome" and "Double stranded DNA". A red arrow labeled "1" points to the diagram. The right sidebar shows a list of related videos, with a red arrow labeled "2" pointing to the search bar area. The video player includes standard controls like play/pause, volume, and full screen.

YouTube JP ngs 原理

Sample Prep  
Tagmentation

Transposome Transposome  
Double stranded DNA

①

次の動画 自動再生

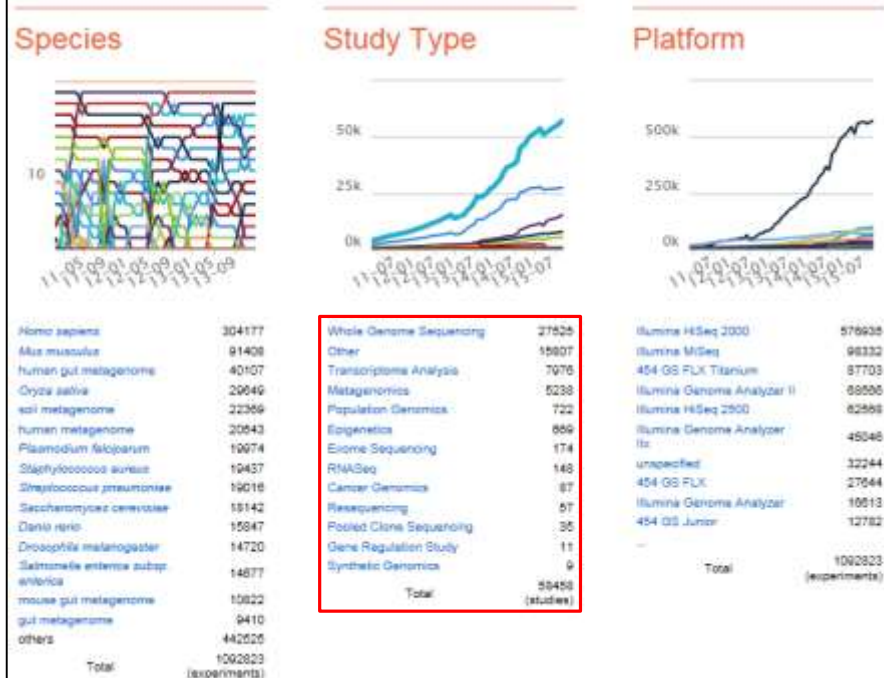
- Illumina sequencing | DNA sequencing by synthesis**  
作者: Suman Bhattacharjee  
視聴回数 13,926 回  
10:42
- Next-Generation Sequencing Technologies - Elaine Mardis**  
作者: GenomeTV  
視聴回数 156,395 回  
1:23:46
- Next-Generation Sequencing Technologies - Elaine Mardis**  
作者: GenomeTV  
視聴回数 51,369 回  
1:34:36
- Next Generation Sequencing (NGS) - An Introduction**  
作者: Applied Biological Materials - abm  
視聴回数 10,553 回  
9:30
- Statistics for Genomics: Introduction to RNAseq**  
作者: Rafael Irizarry  
視聴回数 32,988 回  
1:17:12
- Jonathan Weissman (UCSF/HHMI): DNA Sequencing**  
作者: iBiology

Illumina Inc  
チャンネル登録 2,601  
193,874  
+ 追加 共有 ... その他 635 20

# NGSデータのトレンド

DBCLS SRA  DISCOVER  
Interesting & Available SRA Data

## Trends in SRA data

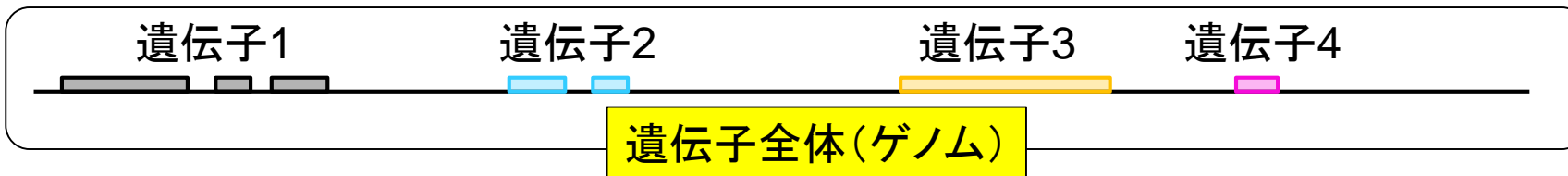


Dis	Count
Whole Genome Sequencing	27525
Other	15807
Transcriptome Analysis	7976
Metagenomics	5238
Population Genomics	722
Epigenetics	669
Exome Sequencing	174
RNASeq	148
Cancer Genomics	87
Resequencing	57
Pooled Clone Sequencing	35
Gene Regulation Study	11
Synthetic Genomics	9
<b>Total</b>	<b>58458</b>

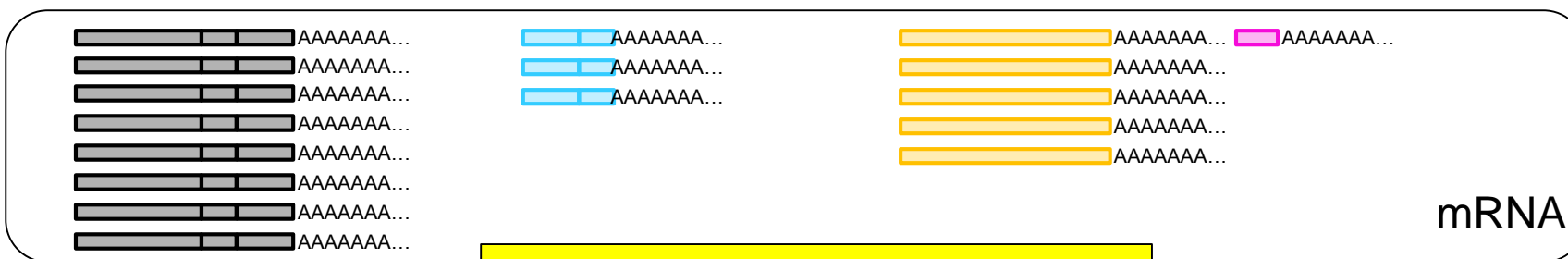
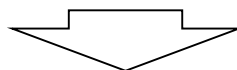


# トランスクリプトームとは

- ある状態のあるサンプル(例:目)のあるゲノムの領域



- ・どの染色体上のどの領域にどの遺伝子があるかは調べる個体(例:ヒト)が同じなら不変(目だろうが心臓だろうが...)



## 転写物全体(トランスクリプトーム)

- ・遺伝子1は沢山転写されている(発現している)
- ・遺伝子4はごくわずかしか転写されてない
- ・...

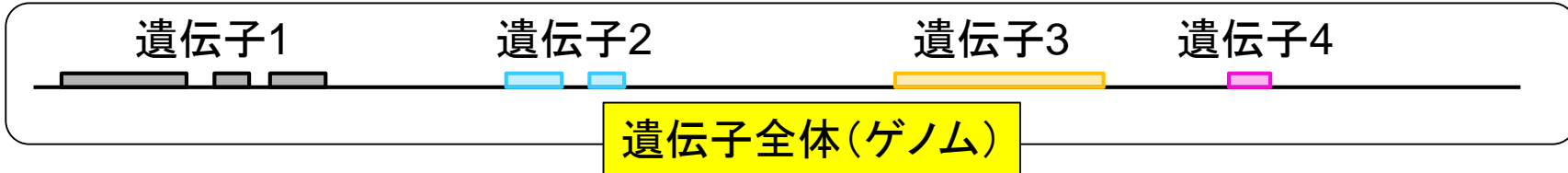
光刺激

ヒト

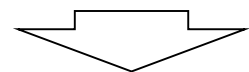


# トランスクリプトームとは

- ある状態のあるサンプル(例:目)のあるゲノムの領域



・どの染色体上のどの領域にどの遺伝子があるかは調べる個体(例:ヒト)が同じなら不変(目だろうが心臓だろうが...)

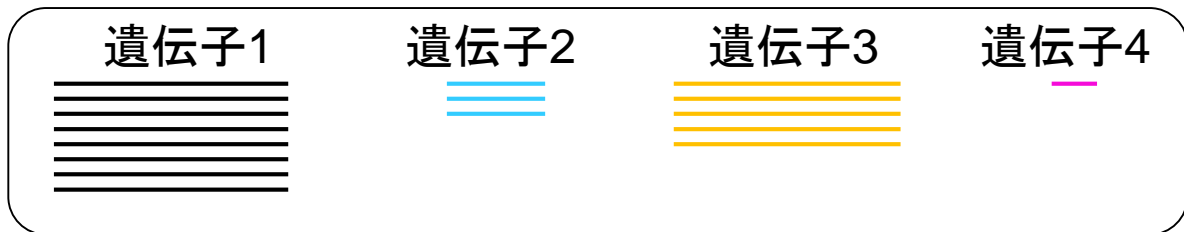


- ・遺伝子2は光刺激に应答して発現亢進
- ・遺伝子4も光刺激に应答して発現亢進

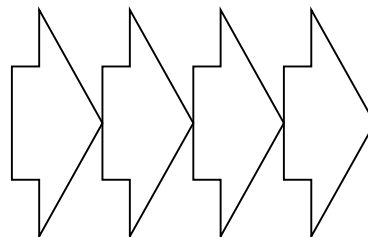
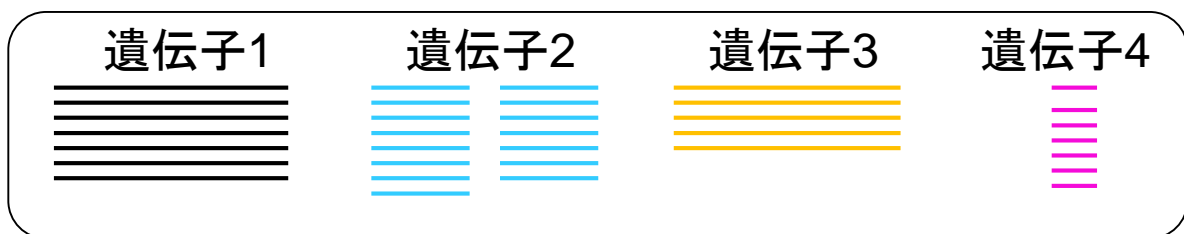
# トランスクリプトーム解析

状態の異なる複数サンプルのデータを取得して解析するのが一般的。サンプル間比較。

## ■ 光刺激前 (T1) の目のトランスクリプトーム



## ■ 光刺激後 (T2) の目のトランスクリプトーム

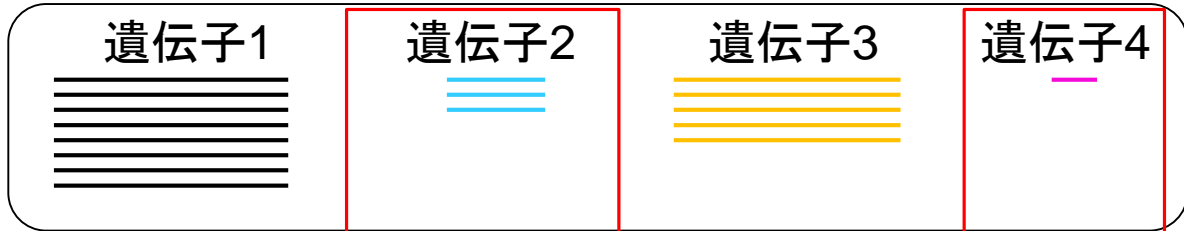


	T1	T2
遺伝子1	8	7
遺伝子2	3	15
遺伝子3	5	5
遺伝子4	1	7
...	...	...

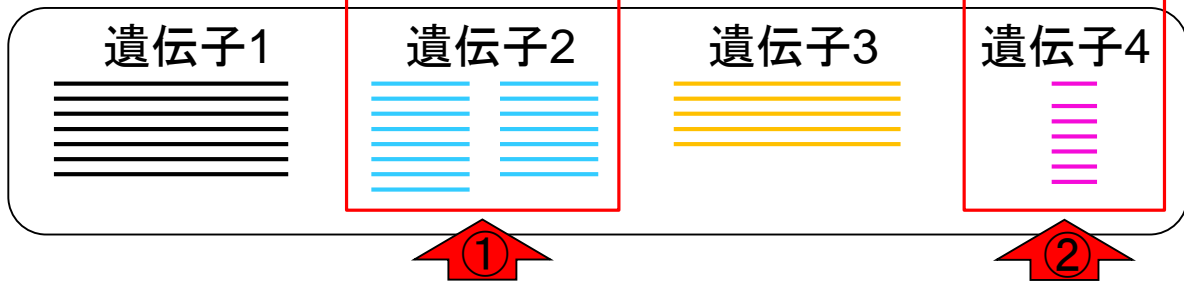
具体的な目的は、①や②の発現変動遺伝子同定など。

# トランスクリプトーム解析

## ■ 光刺激前 (T1) の目のトランスクリプトーム



## ■ 光刺激後 (T2) の目のトランスクリプトーム



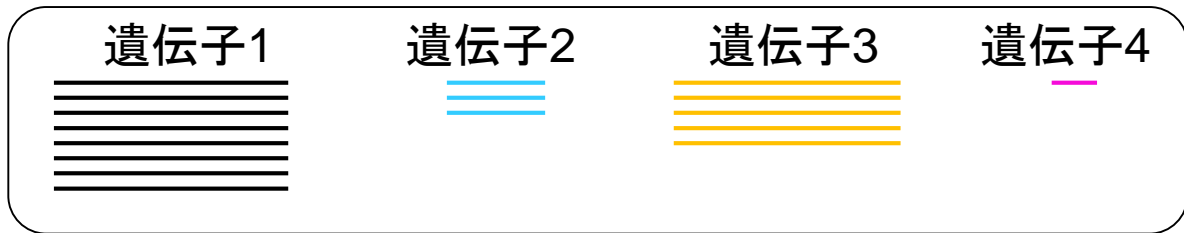
これがいわゆる「遺伝子発現行列」

	T1	T2
遺伝子1	8	7
遺伝子2	3	15
遺伝子3	5	5
遺伝子4	1	7
...	...	...



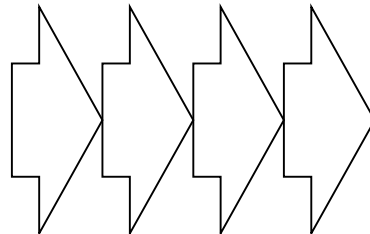
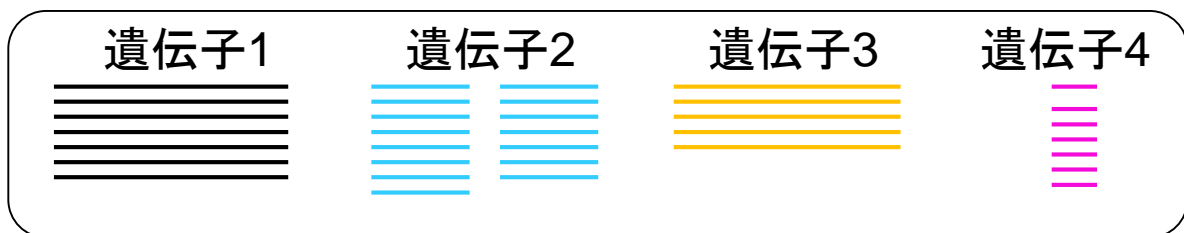
# トランスクリプトームデータ取得

## ■ 光刺激前 (T1) の目のトランスクリプトーム



これがいわゆる「遺伝子発現行列」

## ■ 光刺激後 (T2) の目のトランスクリプトーム



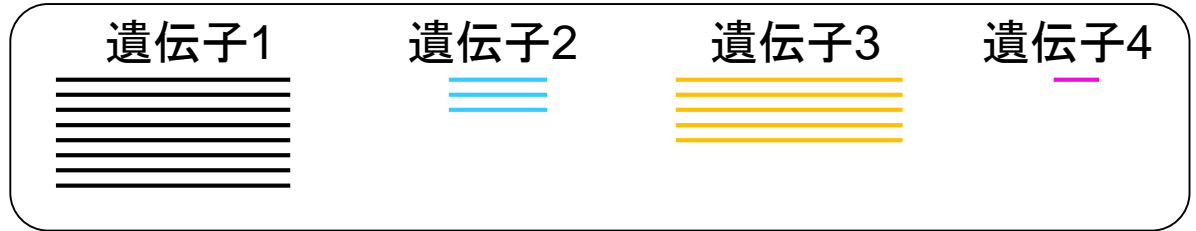
	T1	T2
遺伝子1	8	7
遺伝子2	3	15
遺伝子3	5	5
遺伝子4	1	7
...	...	...



現在はNGSの利用が主流。  
NGSを用いたRNAの配列決定  
(sequencing)なので、RNA-seq

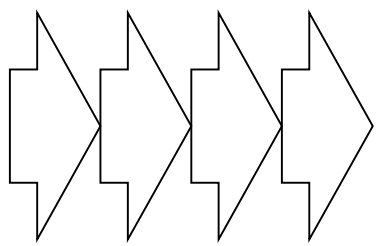
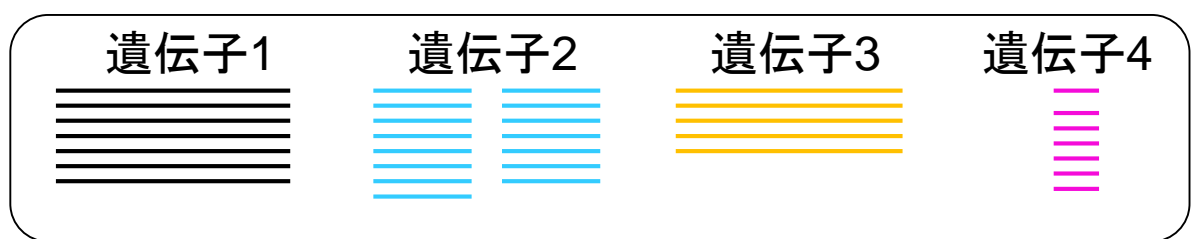
# トランスクリプトームデータ取得

## ■ 光刺激前 (T1) の目のトランスクリプトーム



これがいわゆる  
「遺伝子発現行列」

## ■ 光刺激後 (T2) の目のトランスクリプトーム

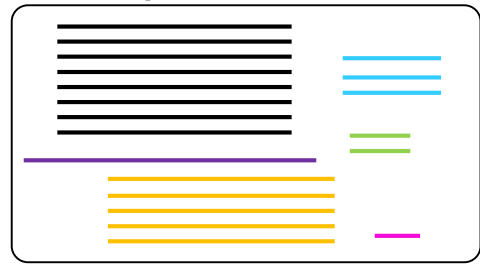


	T1	T2
遺伝子1	8	7
遺伝子2	3	15
遺伝子3	5	5
遺伝子4	1	7
...	...	...

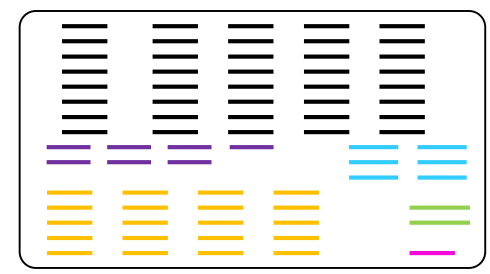


# RNA-seq概略

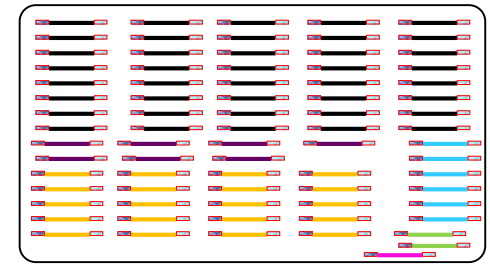
入力: 抽出されたRNA



断片化  
→



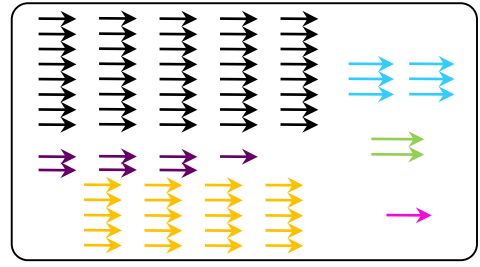
アダプター付加  
↓



NGSで  
配列決定  
←



出力: 塩基配列



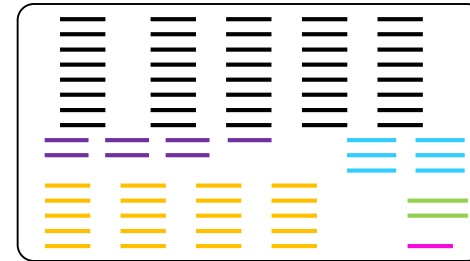
# RNA-seq概略

NGSの出力は、リードと呼ばれる100塩基程度の短い配列が延々と続く巨大なファイル。各矢印が1つのリードに相当。この段階では、まだどのリードがどの転写物由来かは不明(なので灰色一色)。

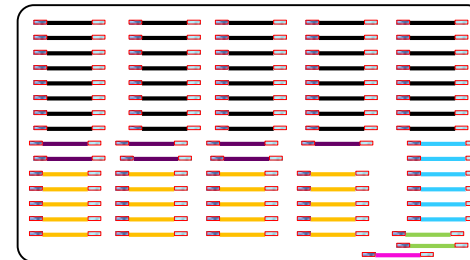
入力: 抽出されたRNA



断片化



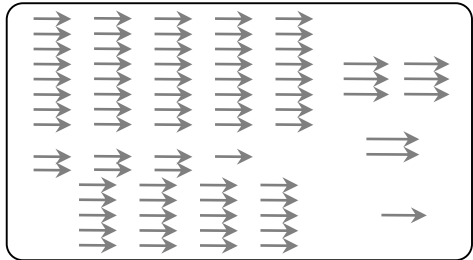
アダプター付加



NGSで  
配列決定



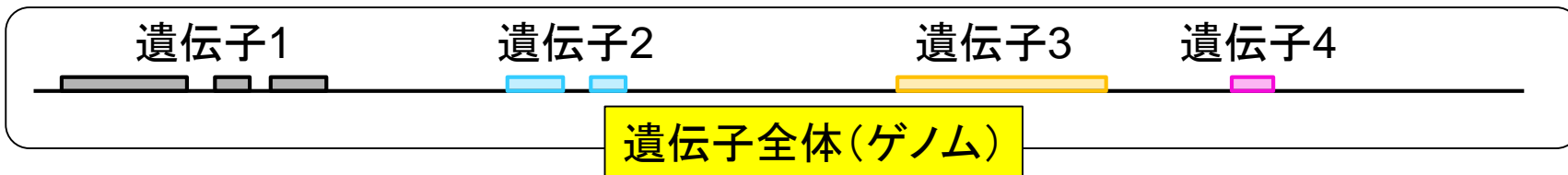
出力: 塩基配列



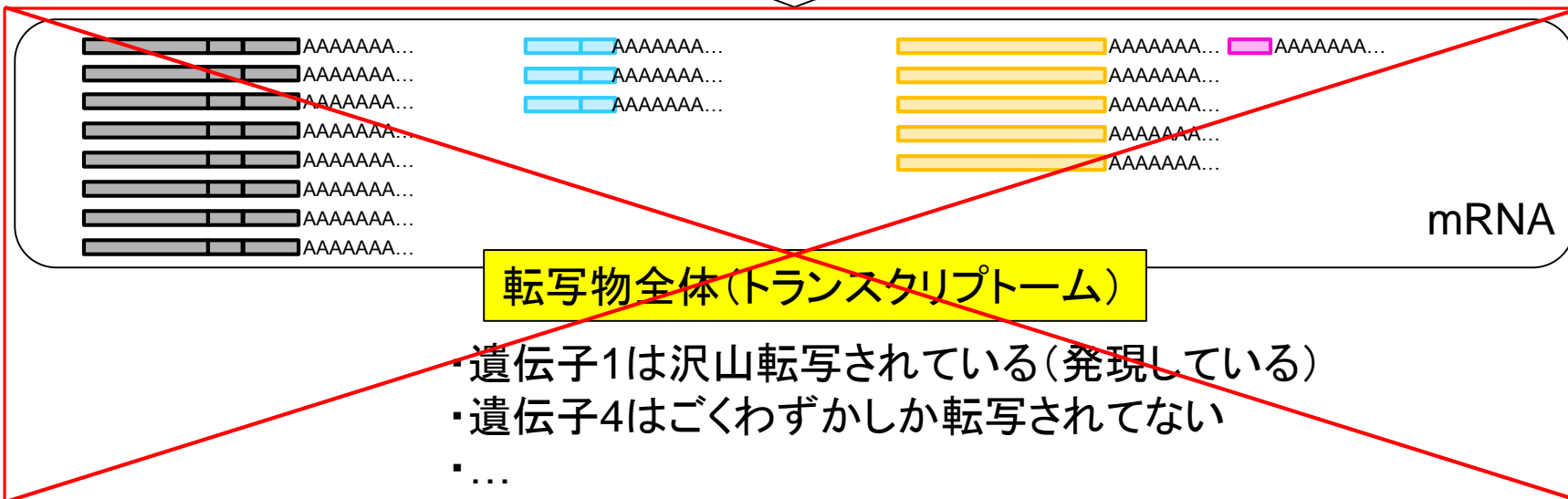
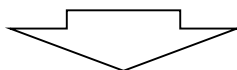
# 遺伝子 ≠ 転写物

赤枠部分の表現は、本当は不正確。昔は実験機器の解像度が事実上遺伝子レベルだった。遺伝子発現解析という表現はその名残り。

- ある状態のあるサンプル(例:目)のあるゲノムの領域



- ・どの染色体上のどの領域にどの遺伝子があるかは調べる個体(例:ヒト)が同じなら不変(目だろうが心臓だろうが...)



ある遺伝子領域から転写 (transcription) されている転写物 (transcript) は、1種類とは限らない

# 遺伝子 ≠ 転写物

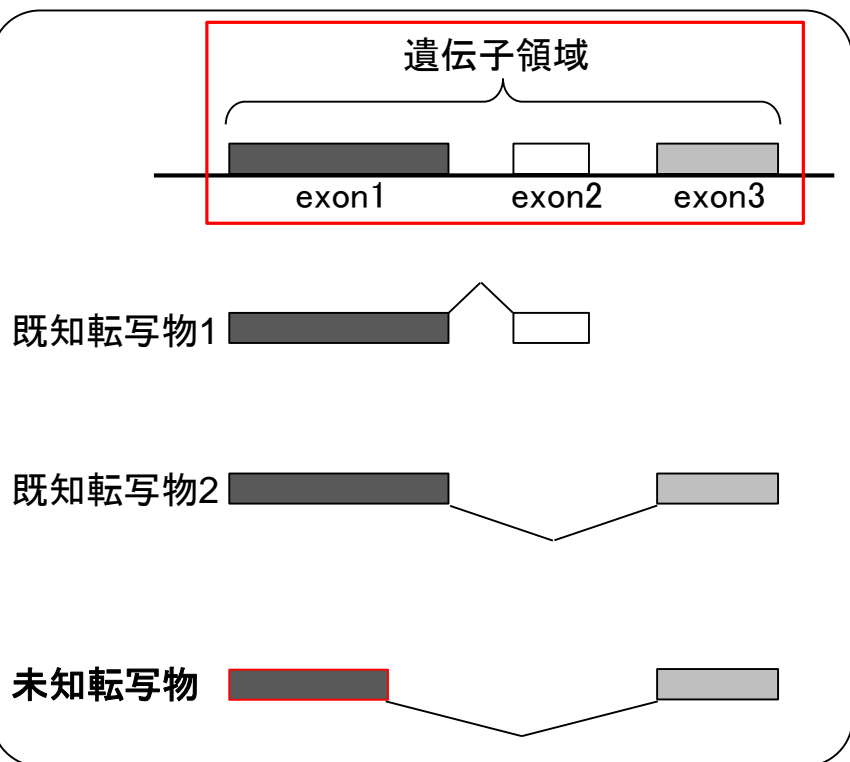
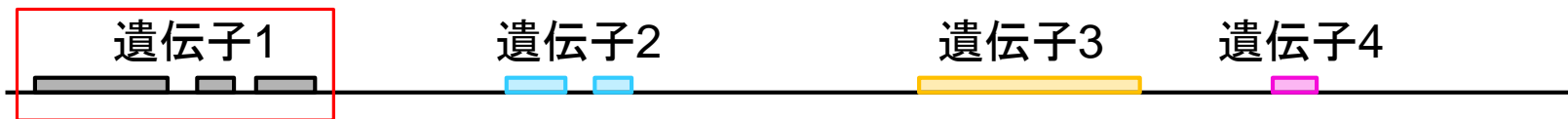
- ある状態のあるサンプル (例: 目) のあるゲノムの領域



# 遺伝子 ≠ 転写物

- ある状態のあるサンプル(例:目)の

ある遺伝子領域から転写(transcription)されている転写物(transcript)は、1種類とは限らない。例えば、遺伝子1の領域では、3種類の真の転写物が存在し、そのうち2種類は既知とする。

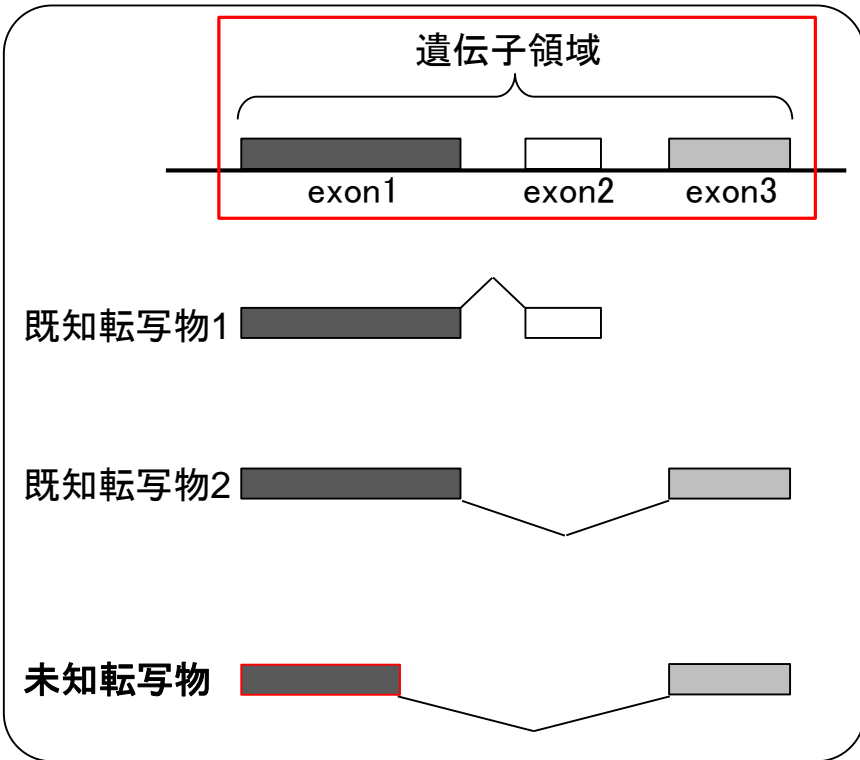
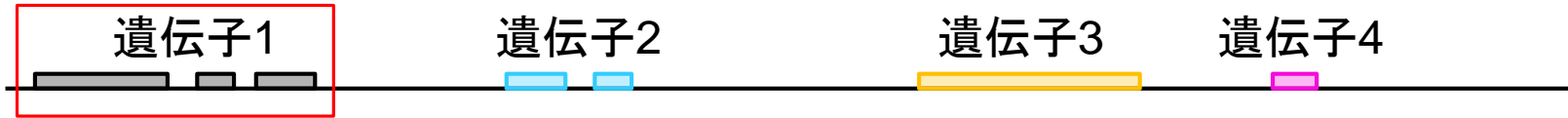


真の転写物情報

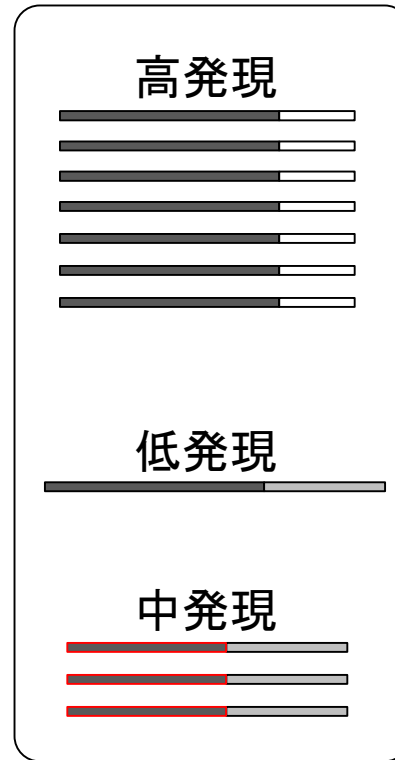
実際の細胞内での発現情報(働いている度合い)が①のような感じだったとする。

# 遺伝子 ≠ 転写物

- ある状態のあるサンプル(例:目)のあるゲノムの領域



真の転写物情報



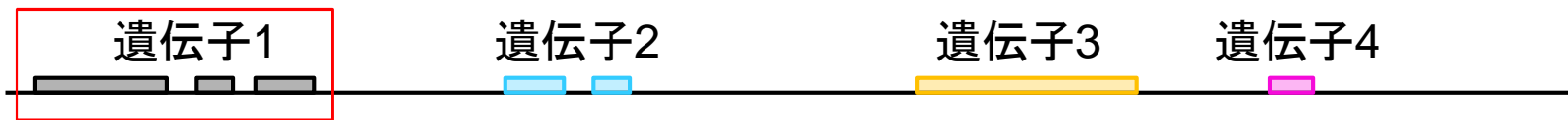
真の発現情報



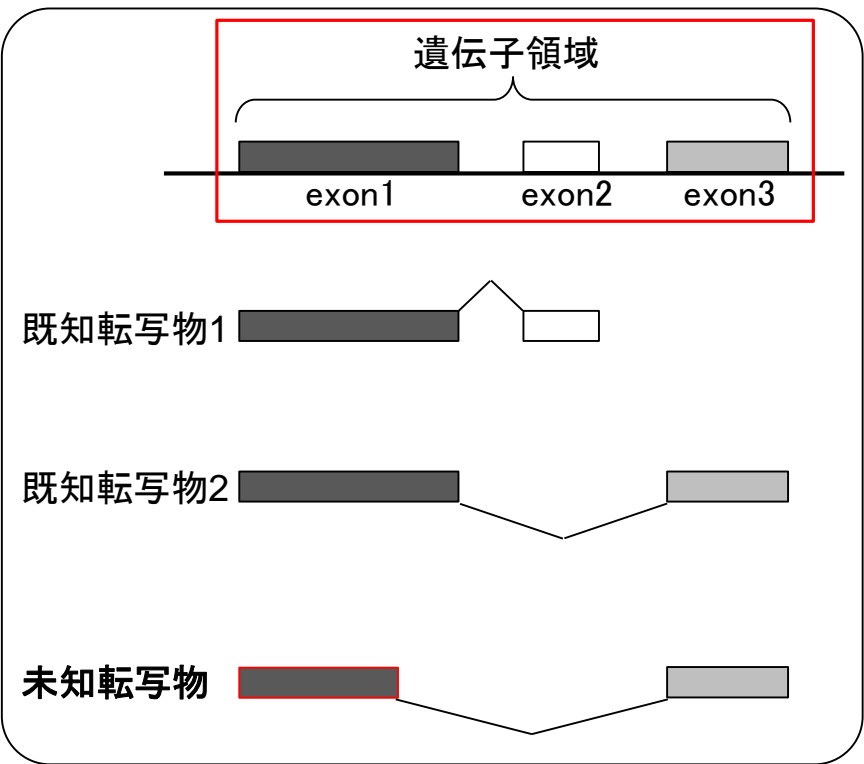


# 遺伝子 ≠ 転写物

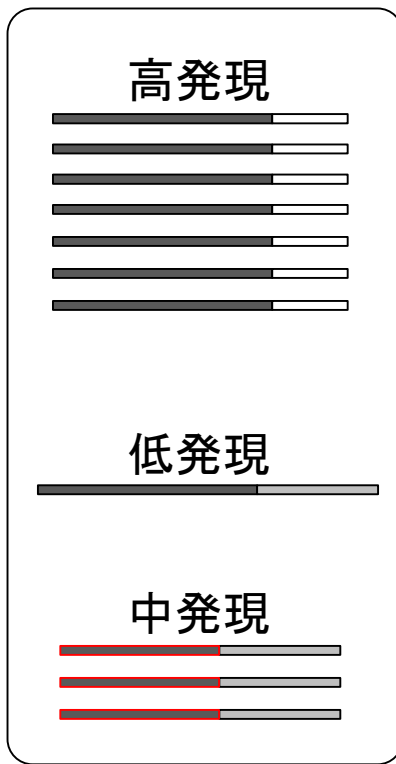
- ある状態のあるサンプル(例:目)のあるゲノムの領域



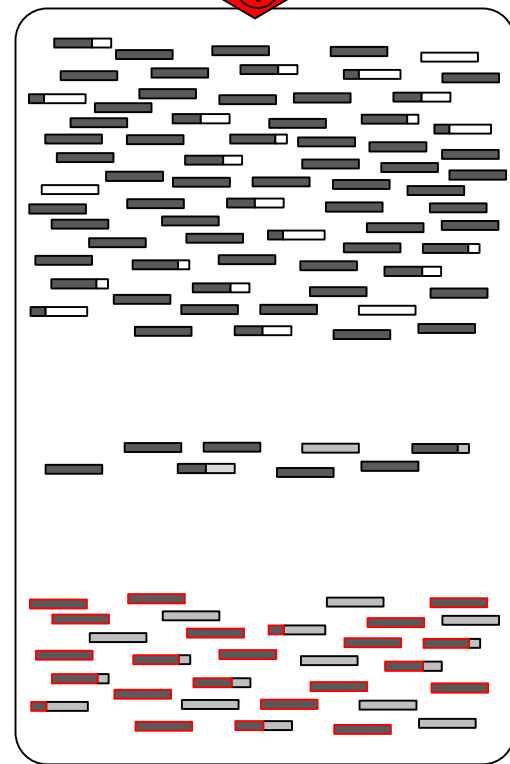
①



真の転写物情報



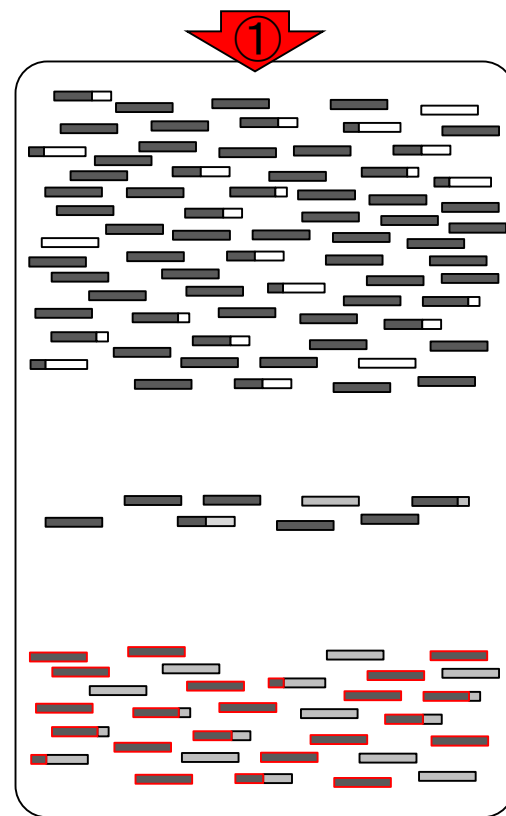
真の発現情報



RNA-seqで得られるリード情報 (色は不明)

# データ解析の出発点

トランスクリプトーム (RNA-seq) データ解析の出発点は、①RNA-seqデータファイル、



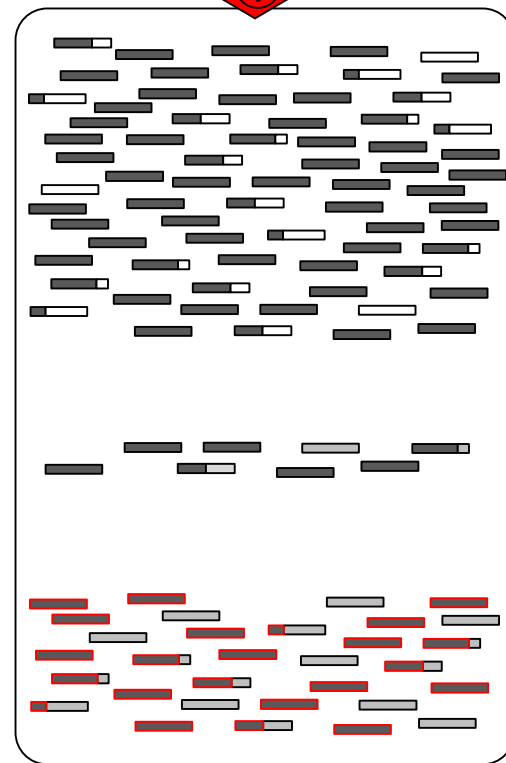
RNA-seqデータ

# データ解析の出発点

トランスクリプトーム (RNA-seq) データ解析の出発点は、①RNA-seqデータファイル、②ゲノム配列情報、

②

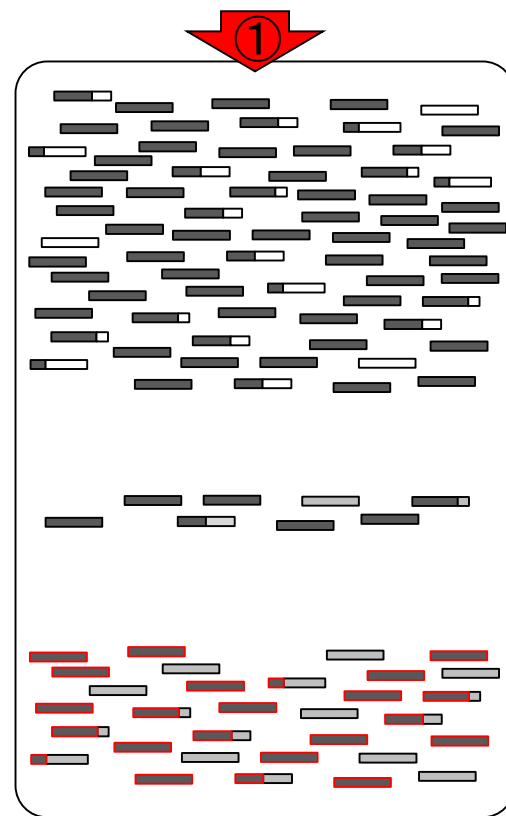
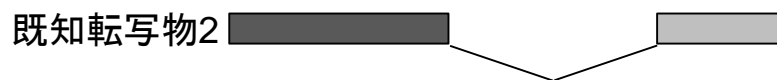
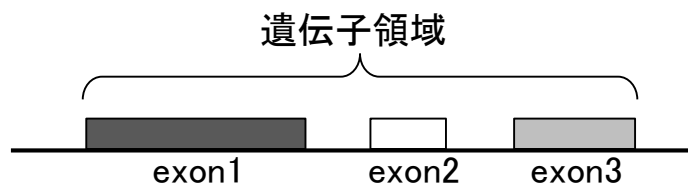
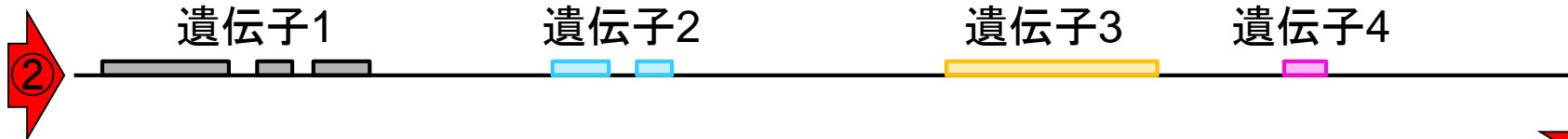
①



RNA-seqデータ

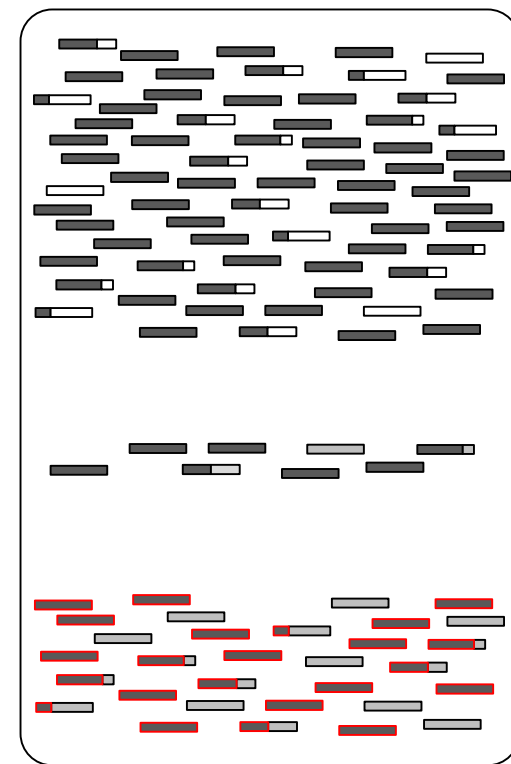
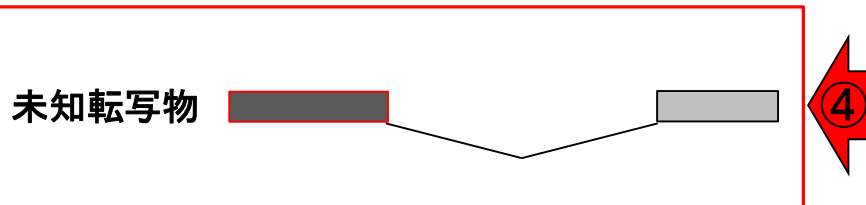
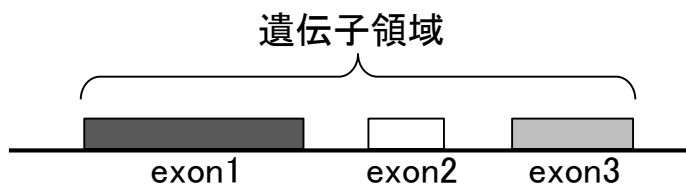
# データ解析の出発点

トランスクリプトーム (RNA-seq) データ解析の出発点は、①RNA-seqデータファイル、②ゲノム配列情報、③ゲノム上のどこにどんな遺伝子、exon、転写物が存在するかというアノテーション情報



# 解析結果のイメージ

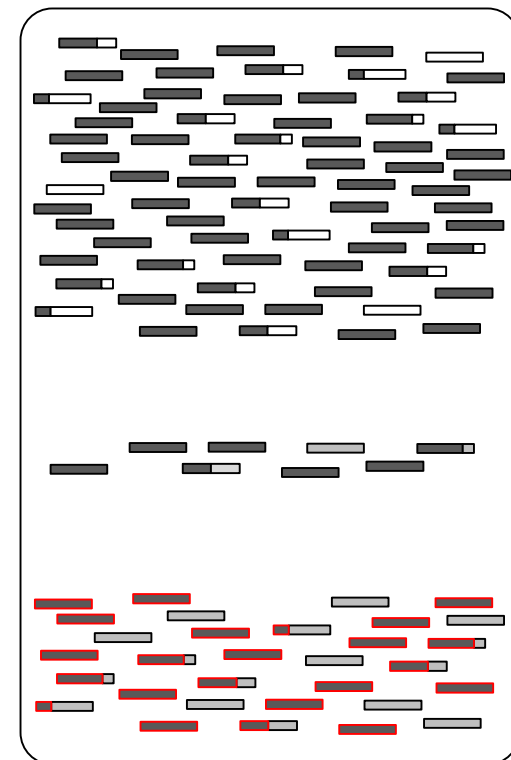
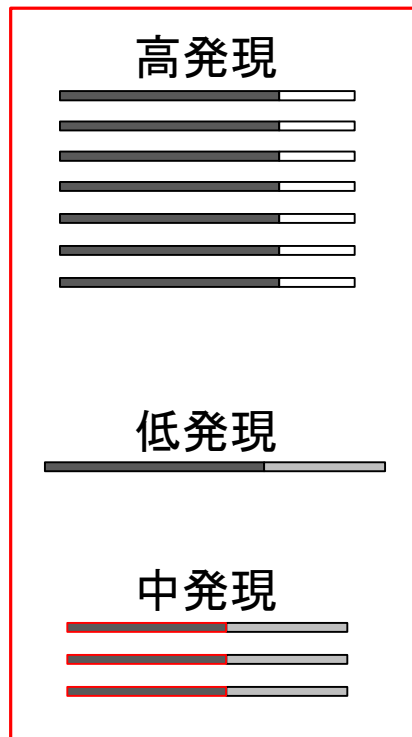
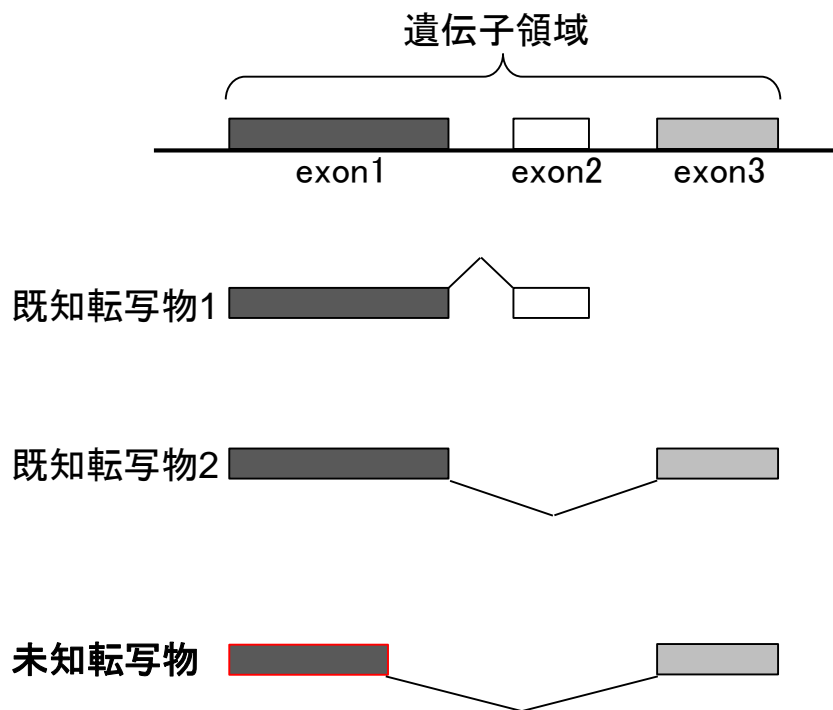
①RNA-seqデータ、②ゲノム配列情報、③アノテーション情報を利用して、④未知転写物(新規isoform)の同定ができる。



RNA-seqデータ

# 解析結果のイメージ

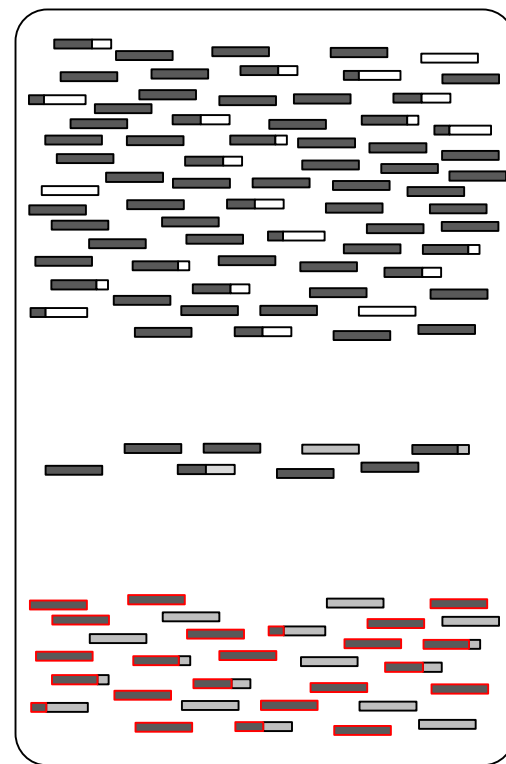
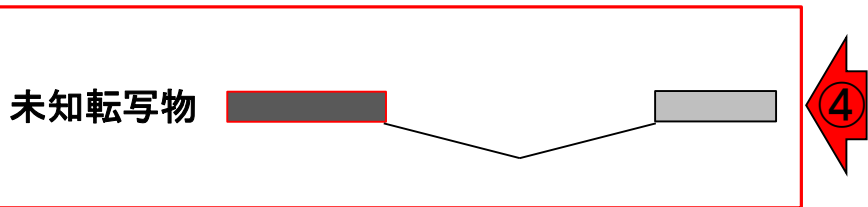
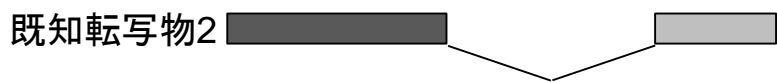
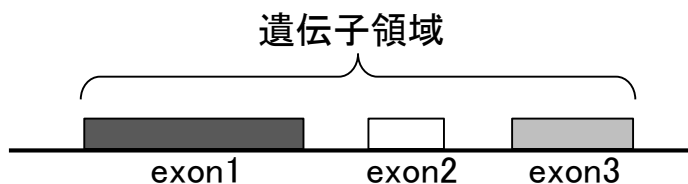
①RNA-seqデータ、②ゲノム配列情報、③アノテーション情報を利用して、④未知転写物(新規isoform)の同定ができる。⑤転写物の発現量(働いている度合い)推定も原理的に可能。



RNA-seqデータ

# 具体的な戦略は？

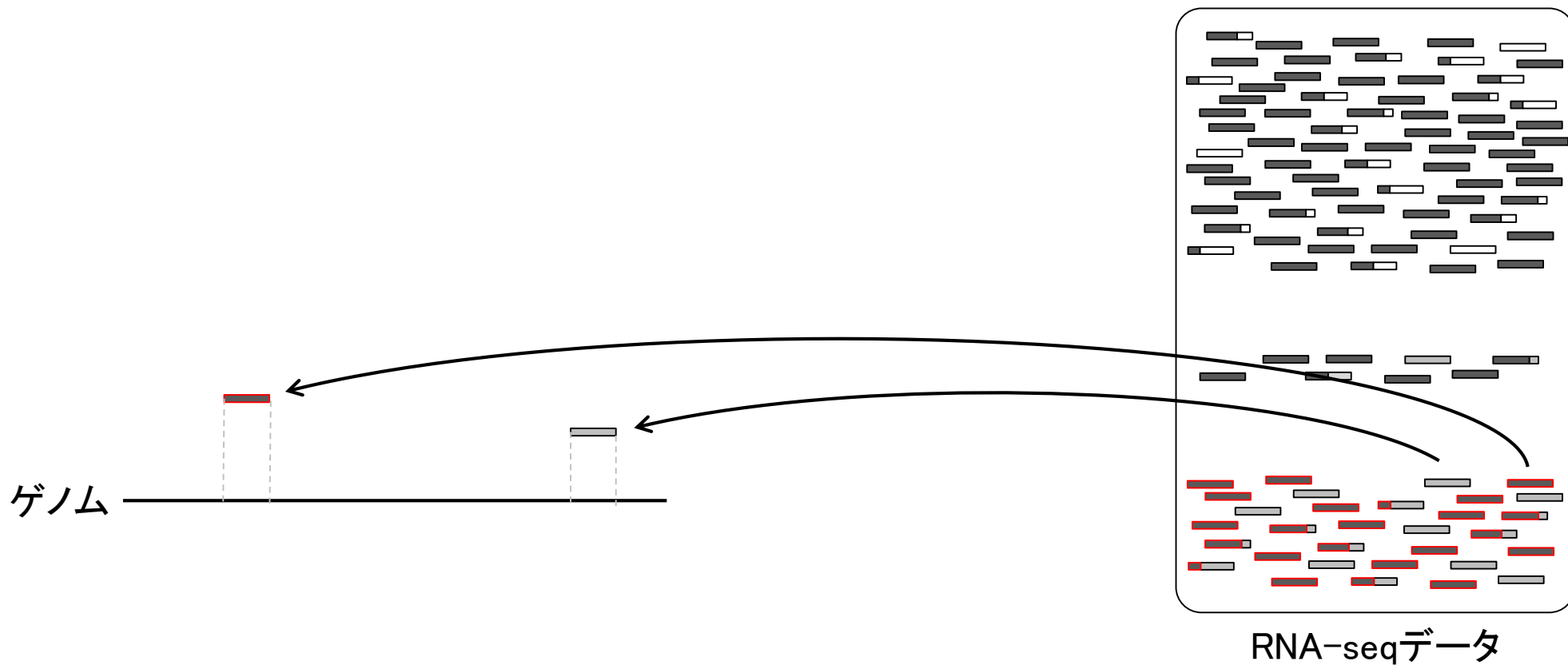
①RNA-seqデータ、②ゲノム配列情報、③アノテーション情報を利用して、④未知転写物(新規isoform)の同定ができる。



RNA-seqデータ

# 具体的な戦略

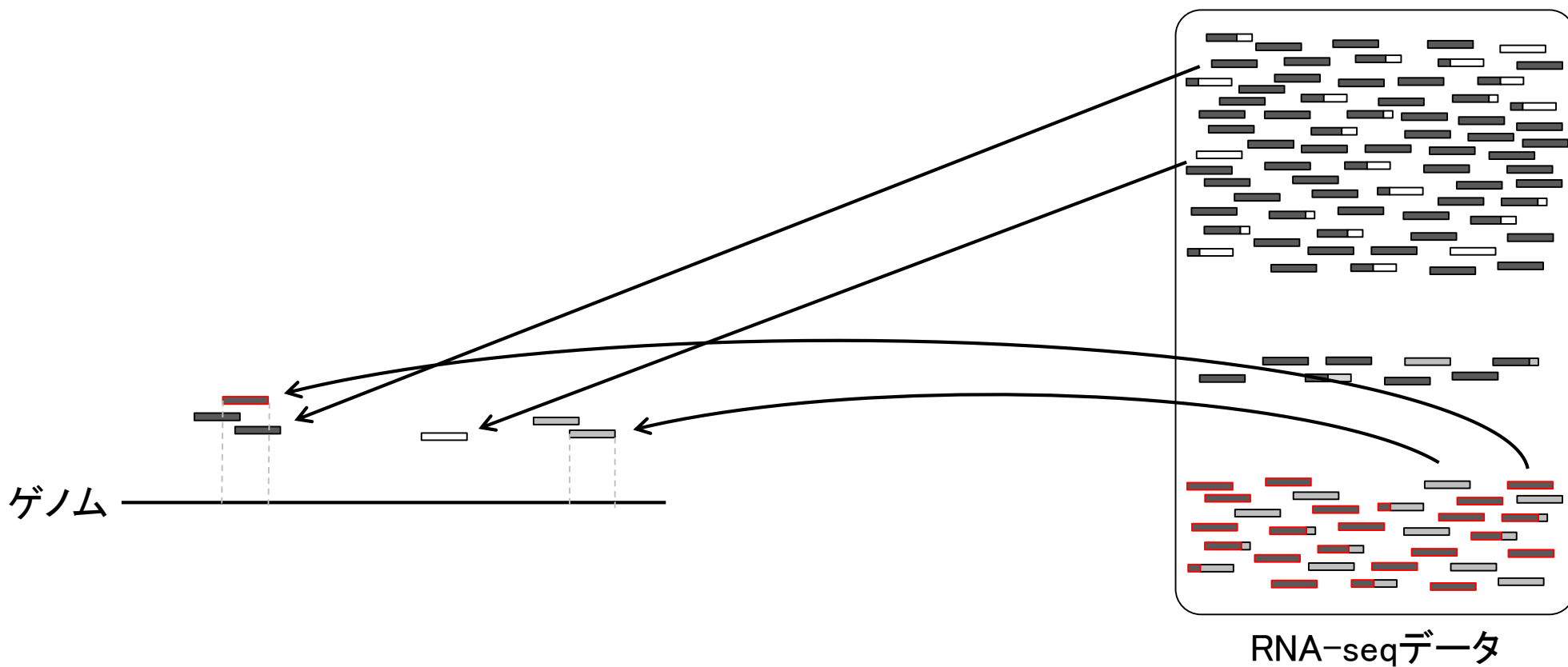
RNA-seqデータ中の1本1本のリード(横棒)がゲノム上のどの領域から転写されたのかを調べる。文字列検索と本質的に同じであり、これがマッピングという作業に相当する。





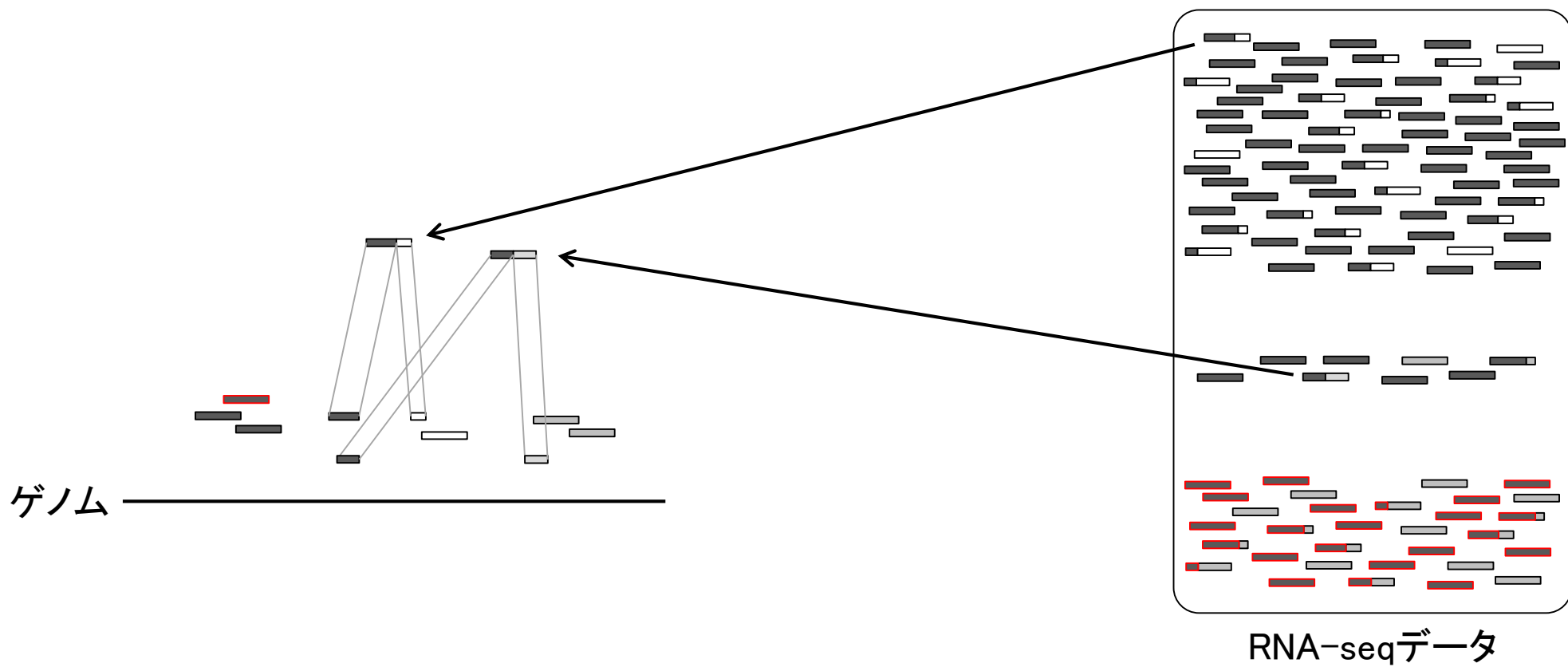
# 具体的な戦略

RNA-seqデータ中の1本1本のリード(横棒)がゲノム上のどの領域から転写されたのかを調べる。文字列検索と本質的に同じであり、これがマッピングという作業に相当する。



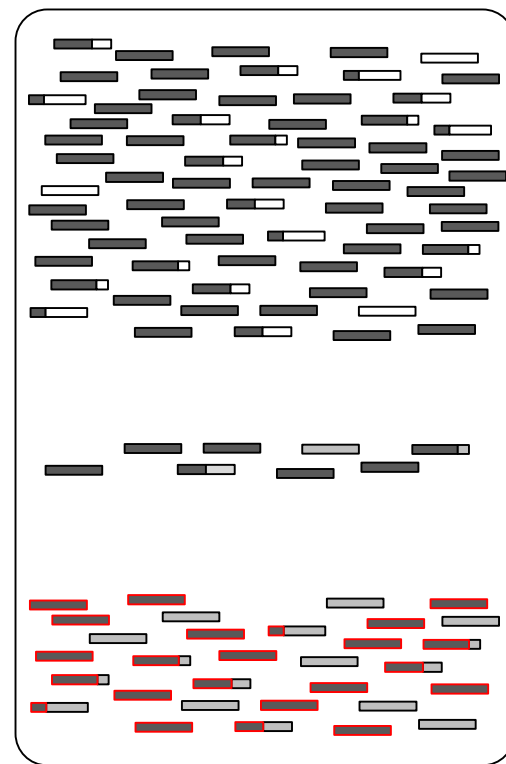
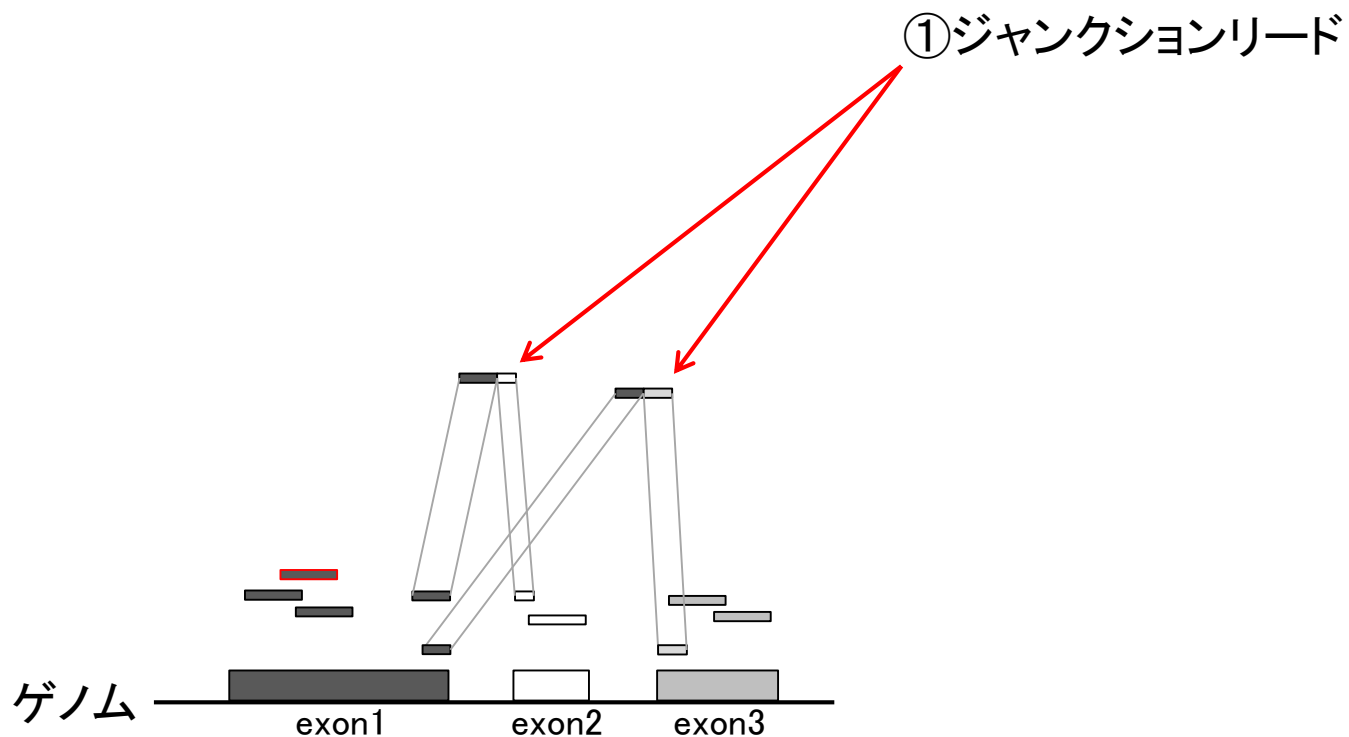
# 具体的な戦略

リードの長さが初期は35塩基程度だったが、現在は150塩基程度まで伸びている。そのおかげで、リードを分割してマッピングすることもできる。



# 具体的な戦略

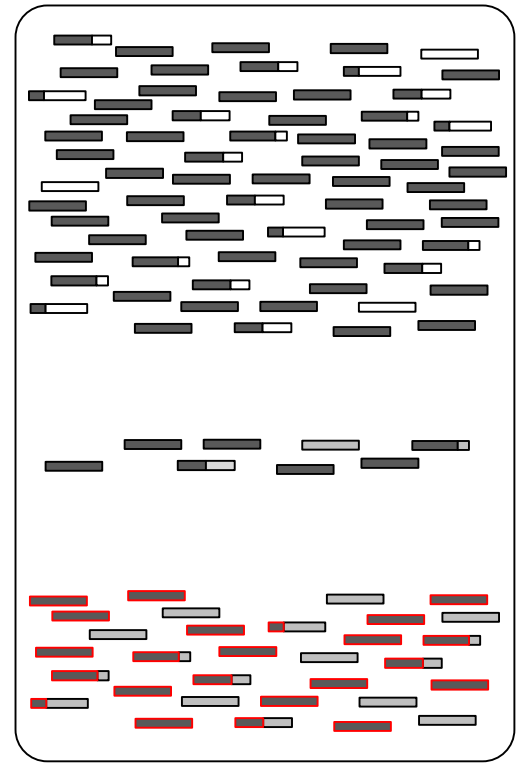
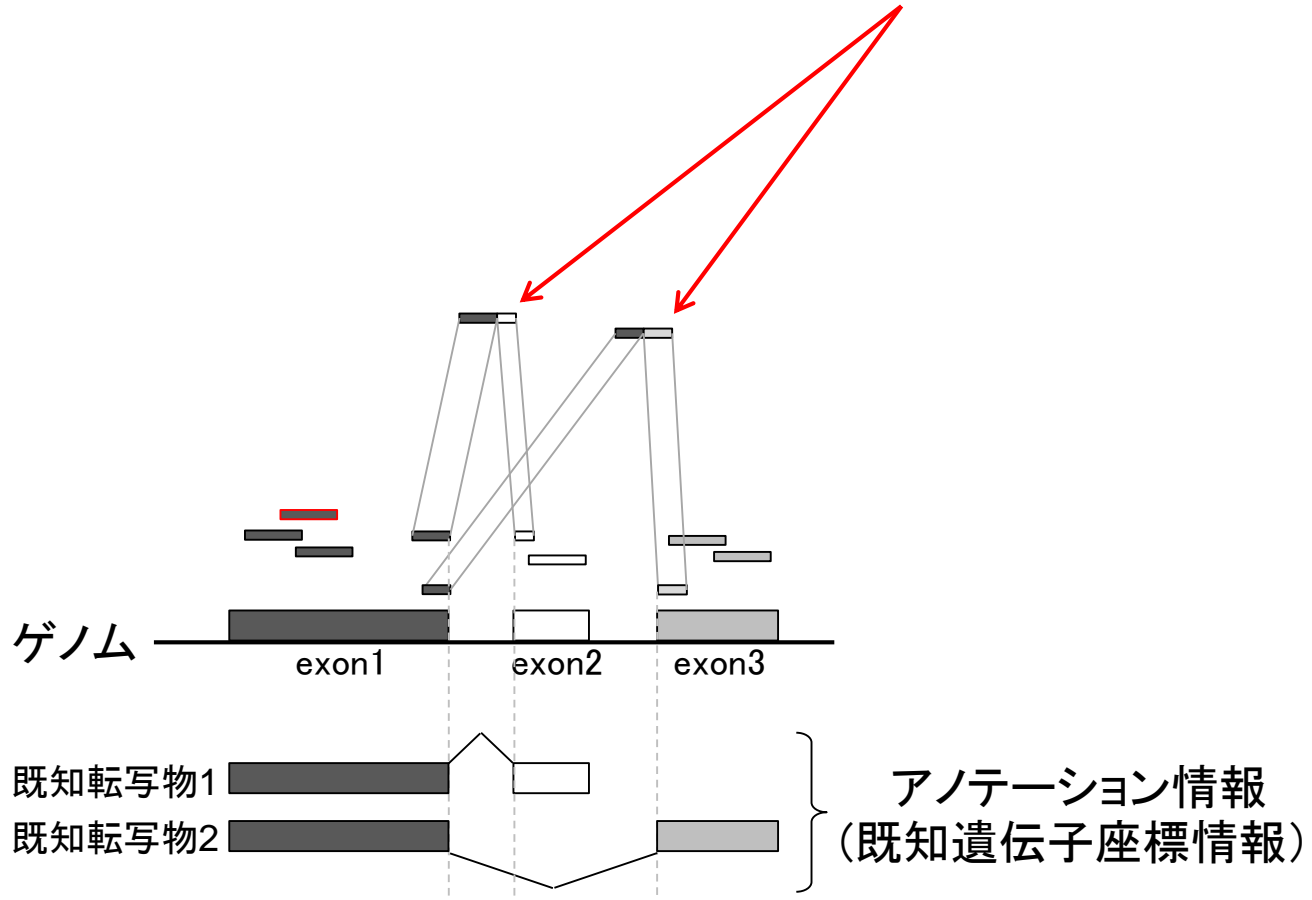
分割してマップされたリードは、大抵の場合複数のエクソン(exon)をまたぐリードであり、①ジャンクションリード(junction read)と呼ばれる。



RNA-seqデータ

# 具体的な戦略

①ジャンクションリード

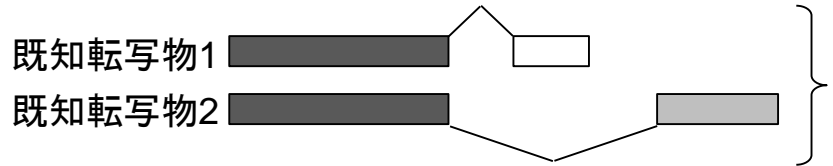
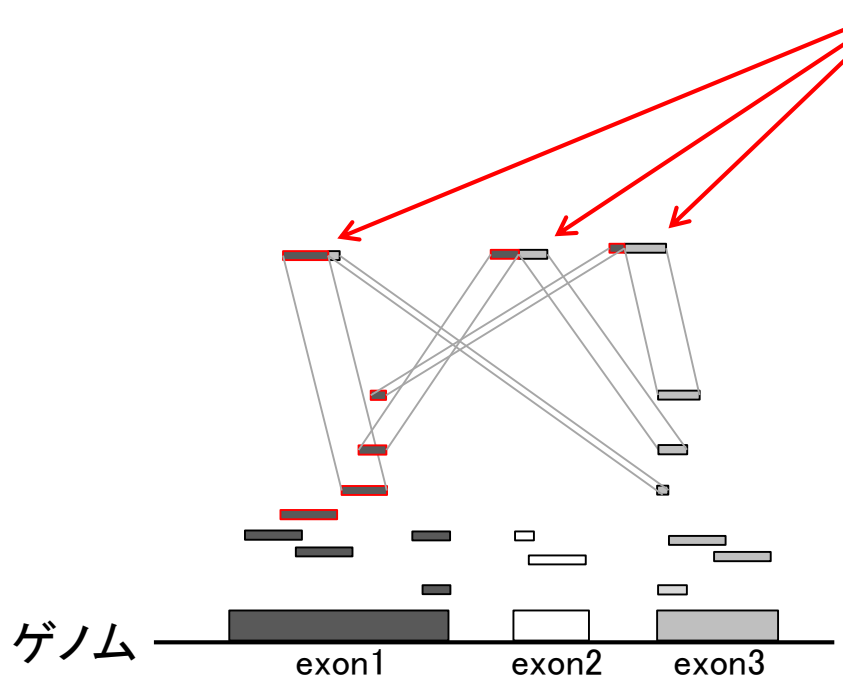


RNA-seqデータ

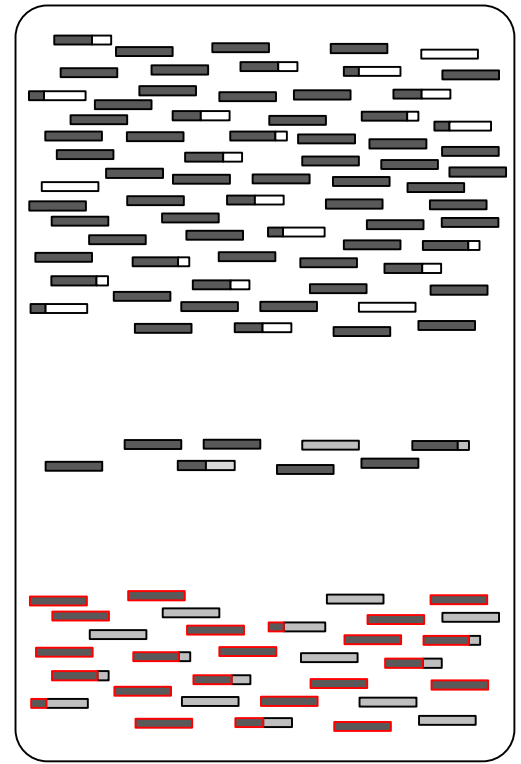
同様にして、他のジャンクションリードも既知転写物と比較することで…

# 具体的な戦略

①ジャンクションリード

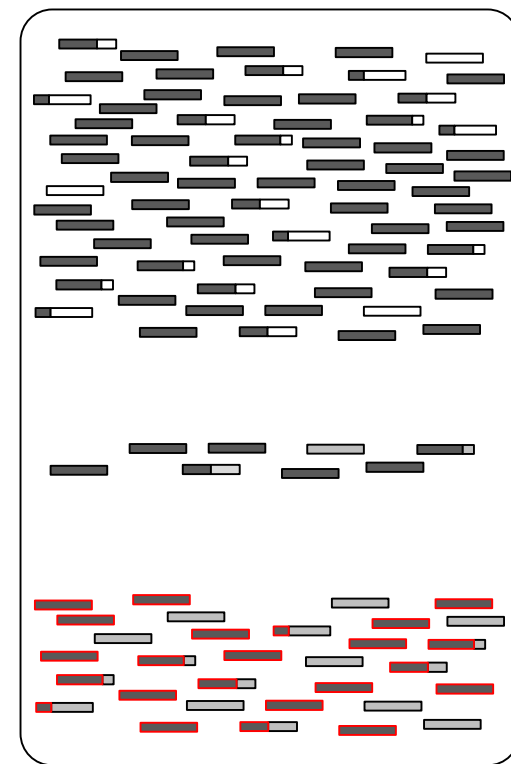
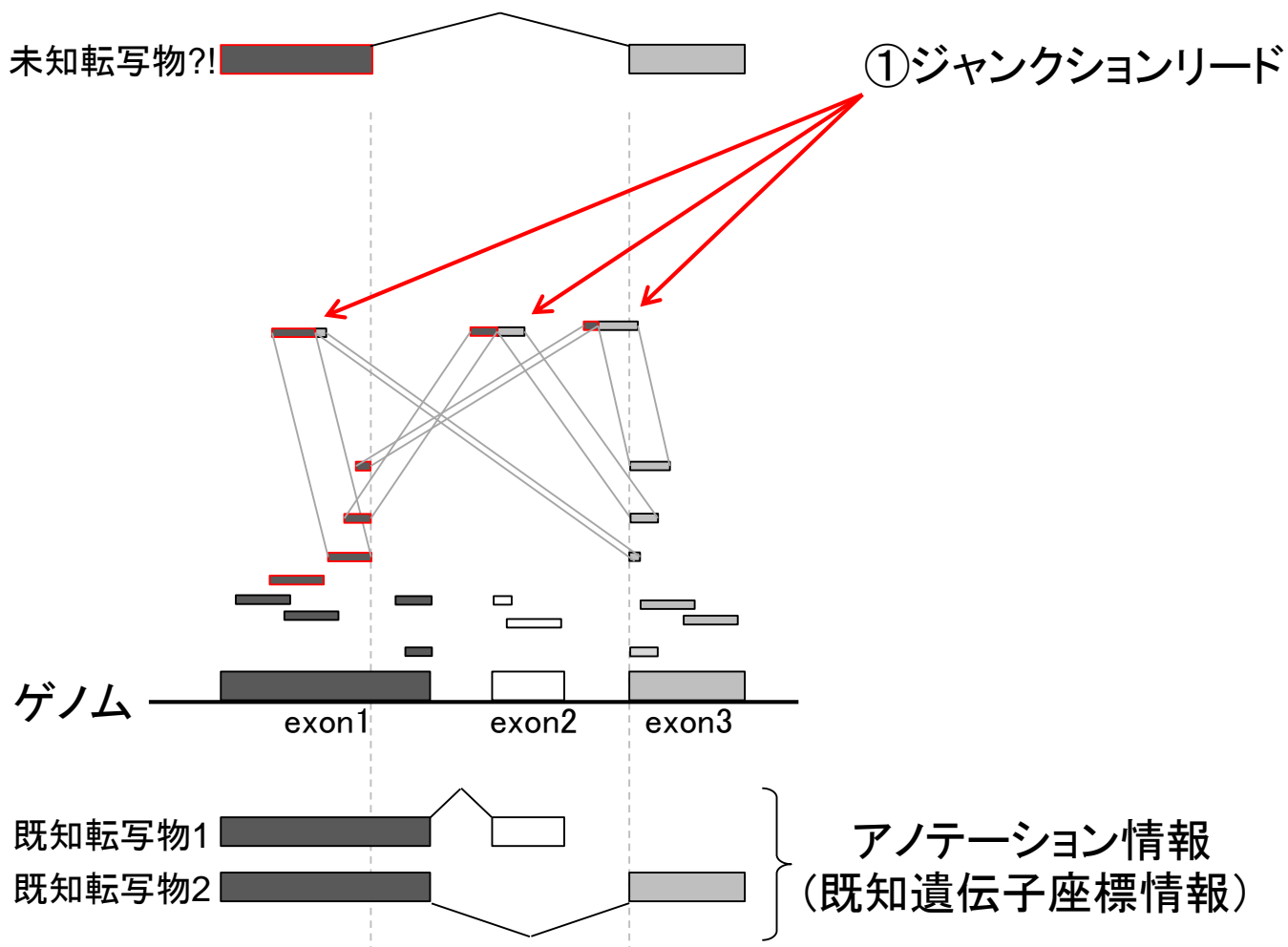


アノテーション情報  
(既知遺伝子座標情報)



RNA-seqデータ

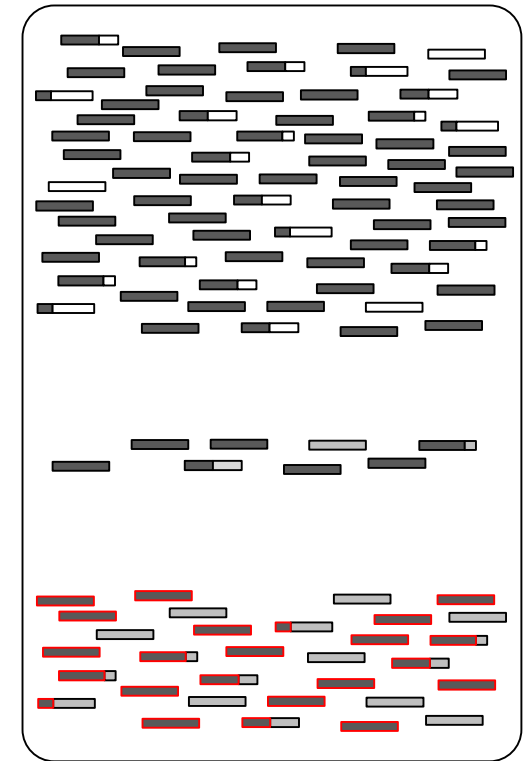
# 具体的な戦略



RNA-seqデータ

# RNA-seqリードの具体例

- Neyret-Kahn et al., *Genome Res.*, **23**: 1563-79, 2013
  - Proliferative and Ras-induced senescent human primary fibroblasts
  - Pro vs. **Ras**の2群間比較 (各群につき3反復)
  - (Pro1, Pro2, Pro3) vs. (**Ras1**, **Ras2**, **Ras3**)
  - 出力は計6ファイル (圧縮状態で約8GB)
  - ファイルあたり数千万リード



RNA-seqデータ

# RNA-seqリードの具体例

公共塩基配列データベース  
DDBJ上で、①リードごとの塩基  
配列情報を眺めることができる。

DRASearch  Send Feedback  Search Home  DRA Home

SRR616151 FASTQ SRA

## Run Detail

Alias	GSM1035443_r1
Instrument model	
Date of run	
Run center	
Number of spots	32,251,234
Number of bases	1,741,566,636

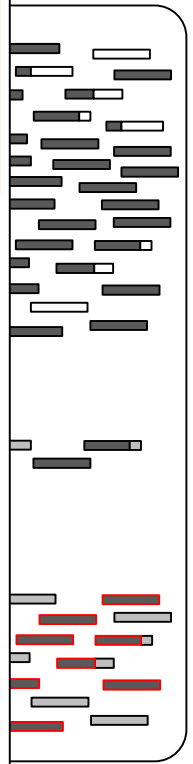
## Navigation

Submission	SRA061444	FTP
Study	SRP017142	
Experiment	SRX204181	FASTQ/S

## READS (joined)

quality  show 10 rows << < 1 / 3225124 Page > >>

```
>SRR616151.1
CCCAGCTCCAGCNCCTCTCGAGATCGACCGGTGCGACGCAGTTC CGCCTGTAGC
>SRR616151.2
TACCTGTACACCTCTTTGTATGCATGTATGACCTCCATGATGACAGGTGCAGTT
>SRR616151.3
CCGGCAGTAATTATGTTTGCCAGGCCAGTGCCCTGGGCACTGGGGGTCTGTGC
>SRR616151.4
CGCACAAATGTATNTTTCAAGGGTGAGCAAGTGTATTCTTAAATTTCTCAAAT
>SRR616151.5
GGCCAGTACCAAGGAGTGGCCCGACCTGTCTCCCGCATCTTCGGTCACTTGT
```



一タ



# RNA-seqリードの具体例

- Neyret-Kahn et al., *Genome Res.*, **23**: 1563-79, 2013
  - Proliferative and Ras-induced senescent human primary fibroblasts
  - Pro vs. **Ras**の2群間比較(各群につき3反復)
  - (Pro1, Pro2, Pro3) vs. (**Ras1**, **Ras2**, **Ras3**)
  - 出力は計6ファイル(圧縮状態で約8GB)
  - ファイルあたり数千万リード

```

>SRR616151.1
CCCAGCTCCAGCNCCTCTCGAGATCGACCGGTGCGACGCAGTTCCGCCTGTAGC

>SRR616151.2
TACCTGTACACCTCTTTGTATGCATGTATGACCTCCATGATGACAGGTGCAGTT

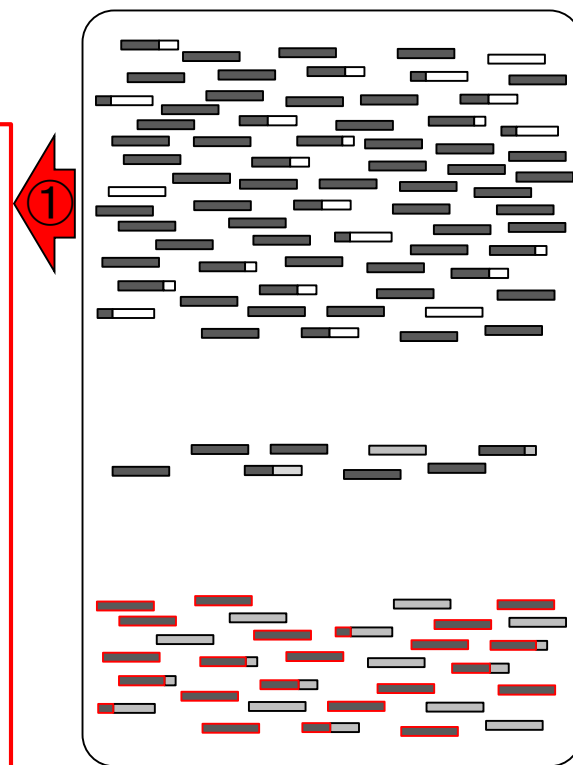
>SRR616151.3
CCGGCAGTAATTATGTTTGCCAGGCCAGTGCCCTGGGCACTGGGGGTCTGTGC

>SRR616151.4
CGCACAAATGTATNTTTC AAGGGTGAGCAAGTGTTATTTCTTAAATTTCTCAAAT

>SRR616151.5
GGCCAGTACCAGNGAGTGGCCCCACCTGTGTCCCCGATGCTTCCGTCACCTTGT

>SRR616151.6

```



RNA-seqデータ

# RNA-seqリードの具体例

- Neyret-Kahn et al., *Genome Res.*, **23**: 1563-79, 2013
  - Proliferative and Ras-induced senescent human primary fibroblasts
  - Pro vs. **Ras**の2群間比較(各群につき3反復)
  - (Pro1, Pro2, Pro3) vs. (**Ras1**, **Ras2**, **Ras3**)
  - 出力は計6ファイル(圧縮状態で約8GB)
  - ファイルあたり数千万リード

>SRR616151.1

CCCAGCTCCAGCNCCTCTCGAGATCGACCGGTGCGACGCAGTTCCGCCTGTAGC

>SRR616151.2

TACCTGTACACCTCTTTGTATGCATGTATGACCTCCATGATGACAGGTGCAGTT

>SRR616151.3

CCGGCAGTAATTATGTTTGCCAGGCCAGTGCCCTGGGCACTGGGGGTCTGTGC

>SRR616151.4

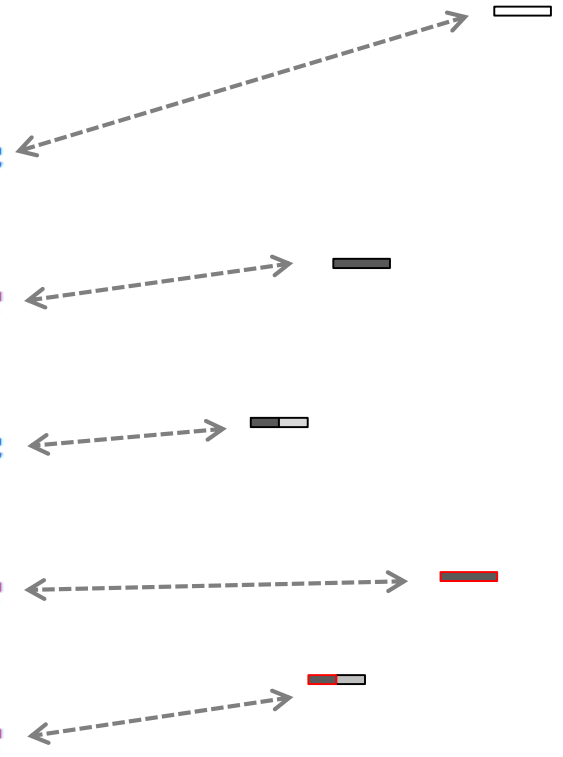
CGCACAAATGTATNTTTCAGGGTGAGCAAGTGTTATTTCTTAAATTTCTCAAAT

>SRR616151.5

GGCCAGTACCAGNGAGTGGCCCCACCTGTGTCCCCGATGCTTCCGTCACCTTGT

>SRR616151.6

CGCACAGCCAGCNCCTCTCGAGATCGACCGGTGCGACGCAGTTCCGCCTGTAGC



# RNA-seqリードの具体例

- Neyret-Kahn et al., *Genome Res.*, **23**: 1563-79, 2013
  - Proliferative and Ras-induced senescent human primary fibroblasts
  - Pro vs. **Ras**の2群間比較(各群につき3反復)
  - (Pro1, Pro2, Pro3) vs. (**Ras1, Ras2, Ras3**)
  - 出力は計6ファイル(圧縮状態で約8GB)
  - ファイルあたり数千万リード

>SRR616151.1

CCCAGCTCCAGCNCCTCTCGAGATCGACCGGTGCGACGCAGTTCCGCCTGTAGC

>SRR616151.2

TACCTGTACACCTCTTTGTATGCATGTATGACCTCCATGATGACAGGTGCAGTT

>SRR616151.3

CCGGCAGTAATTATGTTTGCCAGGCCAGTGCCCTGGGCACTGGGGGTCTGTGC

>SRR616151.4

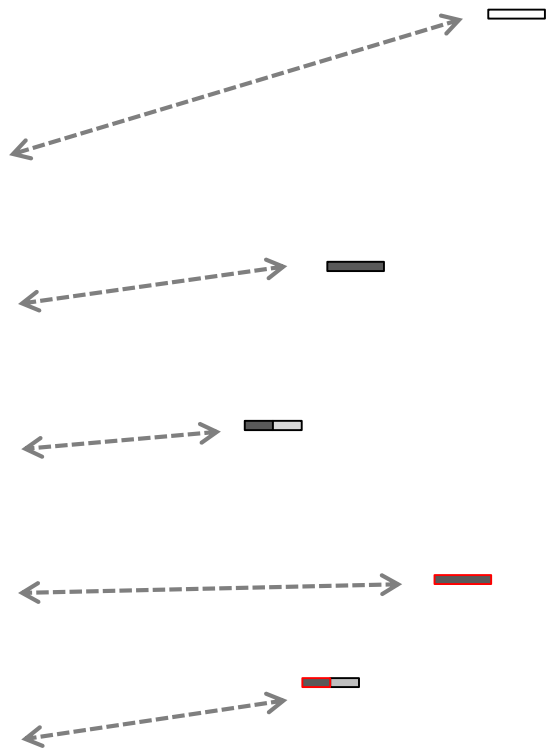
CGCACAAATGTATNTTTC AAGGGTGAGCAAGTGTTATTTCTTAAATTTCTCAAAT

>SRR616151.5

GGCCAGTACCAGNGAGTGGCCCCACCTGTGTCCCCGATGCTTCCGTCACCTTGT

>SRR616151.6

CGCAGACAGCCAGCNCCTCTCGAGATCGACCGGTGCGACGCAGTTCCGCCTGTAGC



# データ解析戦略1

- 解析受託企業に外注

-  ngs 受託解析

# データ解析戦略2

クラウド(ウェブツール)の多くは、Linuxコマンドなどを知らなくてもどうにか解析する方向性を志向。手元に高性能PCがなくてもNGS解析が可能。

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...

# データ解析戦略2

①DDBJ pipelineは、webインターフェースでアセンブルやマッピングを実行可能。内部的にスーパーコンピュータ(スパコン)で計算を実行している

## 解析受託企業に外注

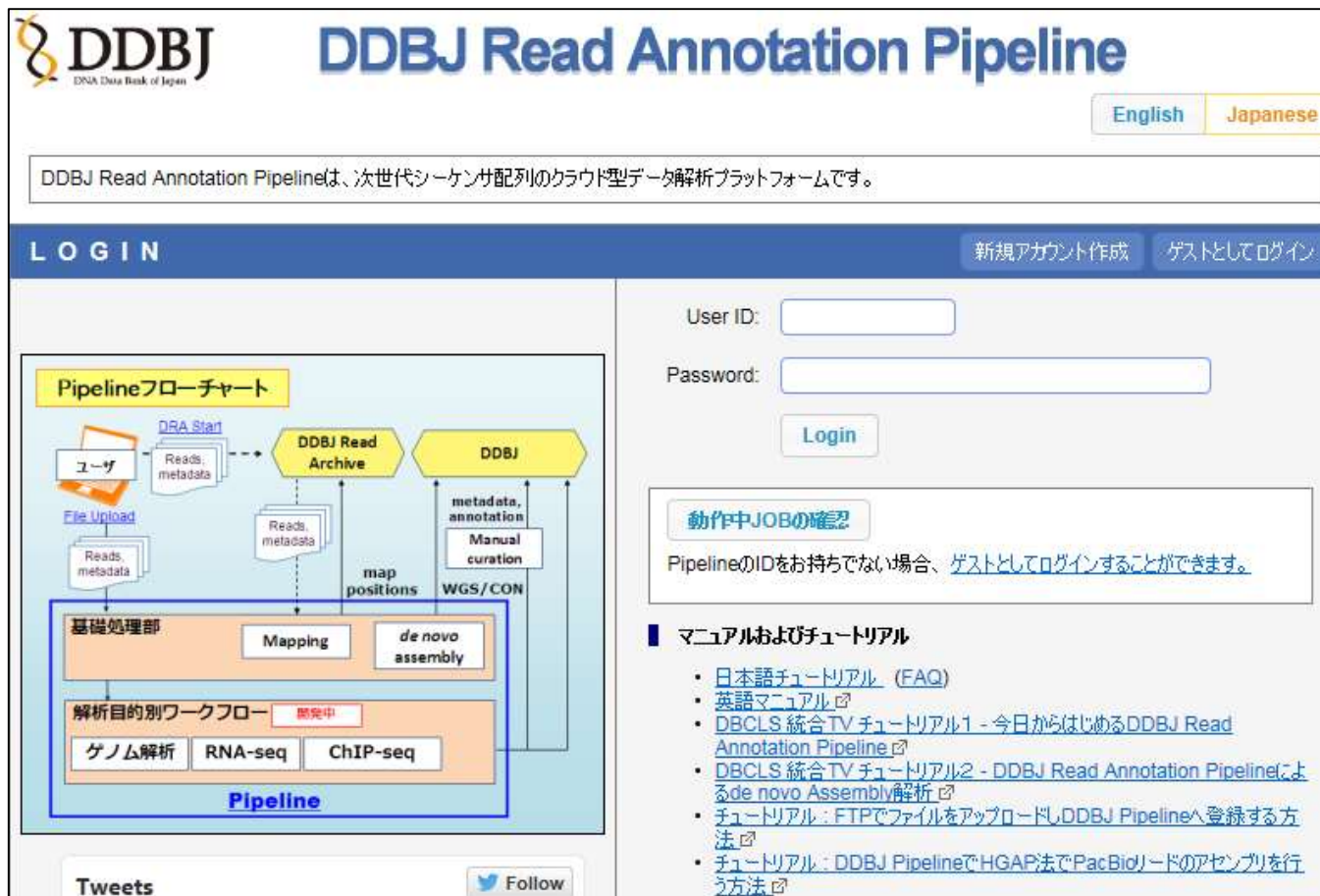
-  ngs 受託解析

## クラウド(ウェブツール)

- ① □ DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



Nagasaki et al., *DNA Res.*, 20:383-390, 2013



DDBJ Read Annotation Pipeline

English Japanese

DDBJ Read Annotation Pipelineは、次世代シーケンサ配列のクラウド型データ解析プラットフォームです。

LOGIN 新規アカウント作成 ゲストとしてログイン

User ID:

Password:

Login

動作中JOBの確認

PipelineのIDをお持ちでない場合、[ゲストとしてログインすることができます。](#)

マニュアルおよびチュートリアル

- [日本語チュートリアル \(FAQ\)](#)
- [英語マニュアル](#)
- [DBCLS 統合TV チュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline](#)
- [DBCLS 統合TV チュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析](#)
- [チュートリアル: FTPでファイルをアップロードしDDBJ Pipelineへ登録する方法](#)
- [チュートリアル: DDBJ PipelineでHGAP法でPacBioリードのアセンブリを行う方法](#)

Tweets Follow

# データ解析戦略2

NGSデータは公共データベースへの登録が義務付けられている。①DDBJ pipelineは、②その登録先(DDBJ)が運用しており、日本語でやり取りできるので便利。

## 解析受託企業に外注

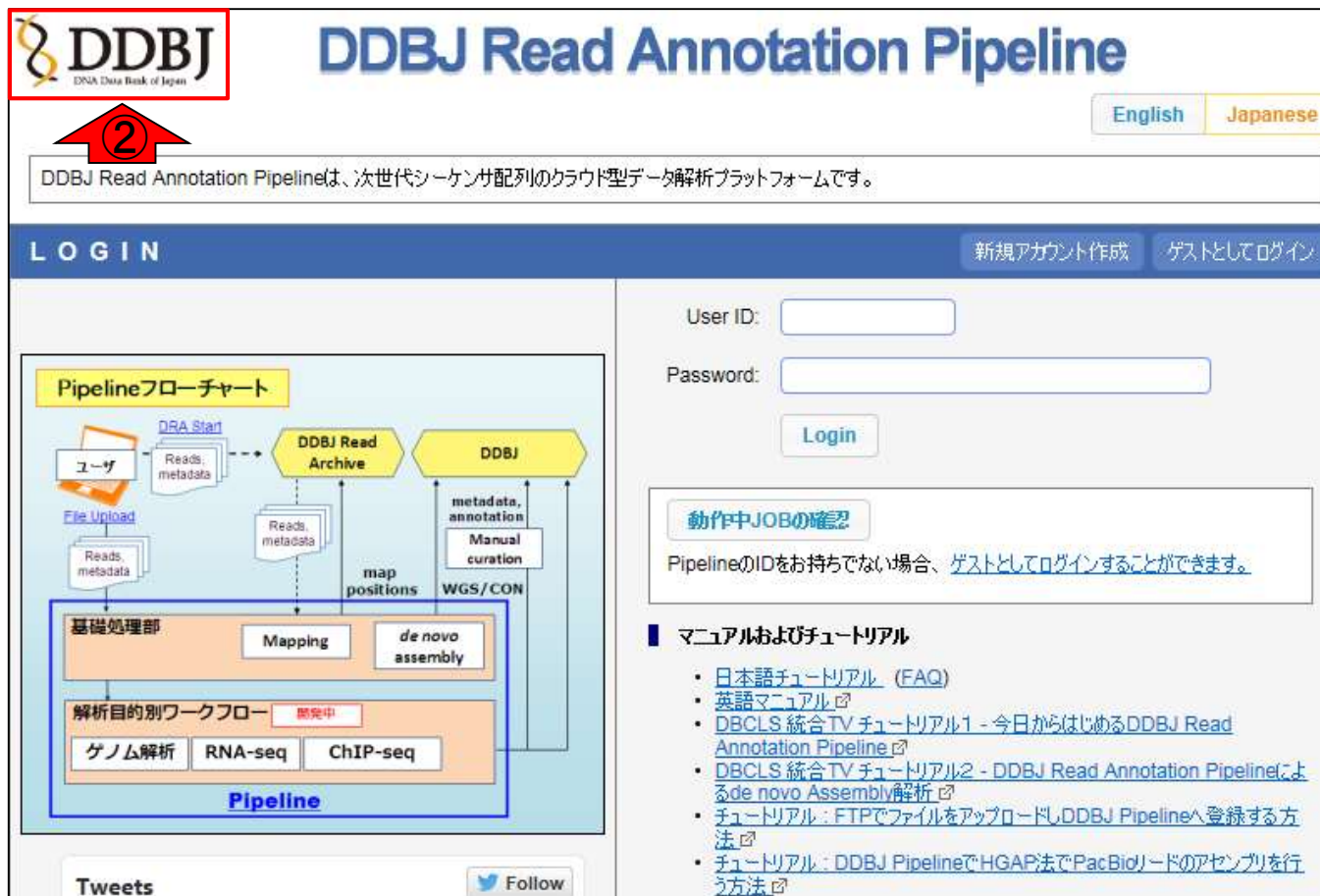
-  ngs 受託解析

## クラウド(ウェブツール)

- ① □ DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



Nagasaki et al., *DNA Res.*, 20:383–390, 2013



**DDBJ** DNA Data Bank of Japan

## DDBJ Read Annotation Pipeline

English Japanese

DDBJ Read Annotation Pipelineは、次世代シーケンサ配列のクラウド型データ解析プラットフォームです。

LOGIN 新規アカウント作成 ゲストとしてログイン

User ID:

Password:

Login

**動作中JOBの確認**

PipelineのIDをお持ちでない場合、[ゲストとしてログインすることができます。](#)

**マニュアルおよびチュートリアル**

- [日本語チュートリアル \(FAQ\)](#)
- [英語マニュアル](#)
- [DBCLS 統合TV チュートリアル1 - 今日からはじめるDDBJ Read Annotation Pipeline](#)
- [DBCLS 統合TV チュートリアル2 - DDBJ Read Annotation Pipelineによるde novo Assembly解析](#)
- [チュートリアル: FTPでファイルをアップロードしDDBJ Pipelineへ登録する方法](#)
- [チュートリアル: DDBJ PipelineでHGAP法でPacBioリードのアセンブリを行う方法](#)

Tweets Follow

# データ解析戦略2

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- ① □ Illumina BaseSpace
- Galaxy
- ...

①Illumina BaseSpaceは、名前の通りNGS機器シェアトップのIlluminaが提供するクラウド解析サービス。②ウェブツール系なのでマウスクリックでNGSデータ解析が可能。データ管理も可能。



illumina®

予算申請資料が充実済み お問い合わせ MyIllumina Tools

アプリケーション システム インフォマティクス 臨床研究 受託サービス サイエンス サポート カンパニー

ソフトウェア / 次世代シーケンサークラウドサービスBaseSpace

メールニュース登録 | Follow us

お問い合わせ

Welcome to push-button bioinformatics.

BaseSpace

Data storage, analysis, and collaboration made easy.

LOGIN OR REGISTER HERE

BaseSpaceは、次世代シーケンサーのデータ解析のためのクラウドサービスです。BaseSpaceにより、生物学者やインフォマティシャンが簡便かつ安全にゲノムデータの解析、保管と共有を行います。次世代シーケンサーのデータ解析も、メニュー選択ツールだけでシンプルに、そして迅速に行えます。面倒で時間がかかるデータ転送ステップを省略することができ、生産性が向上します。

BaseSpace の特長

データ解析と保管が容易に



# データ解析戦略2

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- ① □ Illumina BaseSpace
- Galaxy
- ...

Illuminaが提供するNGS機器と連動しており、シーケンス(塩基配列決定)した結果が直接BaseSpaceに吸い上げられるため、データ転送などの面倒な作業がないのが特徴。



illumina

アプリケーション システム インフォマティクス 臨床研究

ソフトウェア / 次世代シーケンサークラウドサービス BaseSpace

次世代シーケンサークラウドサービス: BaseSpace

概要  
特長  
利用方法  
データ管理  
システム統合  
エデュケーション  
BaseSpace Onsite システム

Interested in receiving newsletters, case studies, and information on new applications? Enter your email address below.

First  
Last  
Email

LOGIN OR REGISTER HERE

BaseSpaceの特長

データ解析と保管が容易に

# データ解析戦略2

Galaxyは、ウェブベースのデータ解析環境。おそらく世界的には、これを使うユーザは多い。

## 解析受託企業に外注

- Google ngs 受託解析

## クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- ① □ Galaxy
- ...

The screenshot displays the Galaxy web interface. The main content area features a large banner that reads "Try Galaxy on the Cloud" with a sub-message: "Now you can have a personal Galaxy within the infinite Universe." Below the banner, there are two tweets from the Galaxy Project. The left sidebar contains a search bar and a list of tool categories such as "Get Data", "Text Manipulation", "Filter and Sort", "Join, Subtract and Group", "NGS: QC and manipulation", "NGS: Mapping", "NGS: RNA-seq", "NGS: SAMtools", "NGS: BAM Tools", "NGS: Picard", "NGS: VCF Manipulation", "Extract Features", "Fetch Sequences", "Fetch Alignments", "Get Genomic Scores", "Operate on Genomic Intervals", "Statistics", "Graph/Display Data", "Phenotype Association", "snpEff", "BEDTools", "Genome Diversity", "EMBOSS", "Regional Variation", "FASTA manipulation", "Evolution", and "Multiple Alignments". The right sidebar shows a search bar for datasets and an "Unnamed history" section with a message: "ヒストリーは空です。 You can load your own data or get data from an external source".

# データ解析戦略2

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- ① □ Galaxy
- ...



The image shows a screenshot of the Pitagora Galaxy website. The header includes the logo and text: "ピタゴラ ギャラクシー Pitagora-Galaxy = VM + Cloud + Japan Community". A search bar is visible in the top right. The main content area is divided into a left sidebar and a main text area. The sidebar contains navigation links under "コンテンツ" (Content), "概要・メンバー" (Overview/Member), "Contents", and "リンク" (Link). The main text area has a "概要・メンバー" section, a "プロジェクトについて" (About the Project) section, and a "Galaxyとは?" (What is Galaxy?) section. Below the text is a diagram illustrating the Galaxy workflow: three boxes labeled "Data" on the left have arrows pointing to a central box labeled "GALAXY". Inside this box, several colorful boxes represent tools like BLAST, Bowtie, SAM, and GATK. Arrows from the "GALAXY" box point to three boxes labeled "Results" on the right.

# データ解析戦略3

最も大変ですが、通り一遍でない独自の解析、最新プログラムの利用などが可能であり、最も拡張性は高いと言えます。NGSデータ解析実務者(バイオインフォマティシャン)の多くはLinuxコマンドを自在に操ります。

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...


## ■ Linuxコマンドを駆使(旧来型)

- なるべく自力で解析
- LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
- NBDC/東大アグリバイオの「NGSハンズオン講習会」の方向性



どこを頑張ってどこで手を抜くかなどは、それぞれの与えられている環境やポリシーによっても異なる。どの選択肢も正解といえるのではないのでしょうか…

# データ解析戦略

- 解析受託企業に外注: Linuxコマンドを知らなくてもよい
  -  ngs 受託解析
- クラウド(ウェブツール): Linuxコマンドを知らなくてもよい
  - DDBJ Pipeline
  - Illumina BaseSpace
  - Galaxy
  - ...
- Linuxコマンドを駆使(旧来型)
  - なるべく自力で解析
  - LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
  - NBDC/東大アグリバイオの「NGSハンズオン講習会」の方向性



①所属先のアグリバイオインフォマティクス教育研究プログラム(東大アグリバイオ)は、バイオインフォマティクスの人材育成を行うのがミッション。

# データ解析戦略3

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...

## ■ Linuxコマンドを駆使(旧来型)

- なるべく自力で解析
- LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
- NBDC/東大アグリバイオの「NGSハンズオン講習会」の方向性

①

# データ解析戦略3

①所属先のアグリバイオインフォマティクス教育研究プログラム(東大アグリバイオ)は、バイオインフォマティクスの人材育成を行うのがミッション。②クラウド(ウェブツール)を教えない理由にはならないが、現実には本格的に教えられない理由がある。

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



## ■ Linuxコマンドを駆使(旧来型)

- なるべく自力で解析
- LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
- NBDC/東大アグリバイオの「NGSハンズオン講習会」の方向性



# アグリバイオ講義風景

教えられない理由は受講人数規模。講義室の後方はパイプ椅子。約130名。①私。





# データ解析戦略3

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



この受講人数規模でのウェブサイトへのアクセスは…

ニサイバーテロ!!

# データ解析戦略3

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



この受講人数規模でのウェブサイトへのアクセスは…

**ニサイバーテロ!!**①首謀者

# データ解析戦略3

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



この受講人数規模でのウェブサイトへのアクセスは…


**ニサイバーテロ!!** ①首謀者

# データ解析戦略3

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline 
- Illumina BaseSpace
- Galaxy
- ...



この受講人数規模でのウェブサイトへのアクセスは…

**ニサイバーテロ!!**①首謀者。


アク禁 and/or 始末書

# データ解析戦略3

## 解析受託企業に外注

-  ngs 受託解析

## クラウド(ウェブツール)

- DDBJ Pipeline 
- Illumina BaseSpace
- Galaxy
- ...



# データ解析戦略3

この受講人数規模での外部ウェブサイトへのアクセスは、サーバ側の処理能力や通信速度もボトルネック。クライアント側のネットワーク障害なども怖いため、なるべく自前サーバのみで完結したい。

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド (ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



スタッフ数が限界以下のため、ネットワークトラブル対応もままならない。

# 自己紹介 (アグリバイオインフォマティクス)

## 学歴および職歴

- 2002年3月 東京大学・大学院農学生命科学研究科 博士課程修了
- 2002年4月 産業技術総合研究所・生命情報科学研究センター
- 2003年11月 放射線医学総合研究所・先端遺伝子発現研究センター
- 2005年2月~ 東京大学・大学院農学生命科学研究科  
 アグリバイオインフォマティクス人材養成プログラム (科学技術振興調整費: 2004/10-2009/3)  
 アグリバイオインフォマティクス教育研究プログラム (特別教育研究経費: 2009/4~2014/3)

## アグリバイオインフォマティクス教育研究プログラム

- 他大学の学生や社会人も受講できる、希少なバイオインフォ教育プログラム

年度	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
修士課程	12	65	73	83	68	72	107	100	121	124	108	152
博士課程	3	7	11	13	6	8	12	21	16	19	24	24
社会人	5	3	8	4	1	0	11	19	32	26	55	32
<b>合計</b>	<b>20</b>	<b>75</b>	<b>92</b>	<b>100</b>	<b>75</b>	<b>80</b>	<b>130</b>	<b>140</b>	<b>169</b>	<b>169</b>	<b>187</b>	<b>208</b>
開講科目数	9	15	15	15	15	12	15	15	14	15	13	13
常勤教員数	6	6	7	7	7	3	4	4	3	2	2	2
ポスドク数	>2	>2	>2	>2	>2	1	1	1	1	1	1	0
門田担当コマ数	3	?	?	8	8	5	5	11	13	14	18	20



# データ解析戦略3

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...

## ■ Linuxコマンドを駆使(旧来型)

- なるべく自力で解析
- LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
- NBDC/東大アグリバイオの「NGS講習会」の方向性





科学技術振興機構(JST)傘下のバイオサイエンスデータベースセンター(NBDC)による全面的な協力のもと、Linuxコマンドを駆使してNGSデータ解析を行うための講習会をH26, 27年度と実施

# データ解析戦略3

## ■ 解析受託企業に外注

-  ngs 受託解析

## ■ クラウド(ウェブツール)

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...

## ■ Linuxコマンドを駆使(旧来型)

- なるべく自力で解析
- LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
- NBDC/東大アグリバイオの「NGS講習会」の方向性



# NGS用カリキュラム

H26年3月にNBDCによって、NGS解析に特化した「バイオインフォマティクス人材育成カリキュラム(次世代シーケンサ)」が策定された。

The screenshot shows the homepage of the National Bioscience Database Center (NBDC). The header includes the NBDC logo, the text "National Bioscience Database Center", and the Japanese name "バイオサイエンスデータベースセンター". There are navigation links for "English", "サイトマップ", and "サイト内検索". A search bar is also present. The main navigation menu includes "ホーム", "NBDCについて", "研究開発", "公募情報", "採用情報", "広報", "人材支援", "お問い合わせ", and "リンク". The breadcrumb trail is "Home > 人材支援 > 支援 > 講習会". The main content area is titled "バイオインフォマティクス人材育成のための講習会" and has a sub-section "NGS解析". The text describes the need for NGS analysis talent and mentions the curriculum developed by the NBDC Human Resource Development Subcommittee. A red arrow with the number "1" points to the link "バイオインフォマティクス人材育成カリキュラム (次世代シーケンサ)". Below this are links for "カリキュラムで習得できる技能" and "カリキュラム フロー図". At the bottom, there are two bullet points for seminars: "H26年度NGS速習コース講習会 (2014年9月1日~12日)" and "H27年度NGSハンズオン講習会 (2015年7月22日~8月6日)".

Home > 人材支援 > 支援 > 講習会

## バイオインフォマティクス人材育成のための講習会

### NGS解析

ライフサイエンス分野の研究現場においては、莫大・多様な研究データが産出され、取り扱うデータ量が飛躍的に増えています。一方で、それらのデータを整備・活用するための人材は不足している状況です。そのような研究現場の状況を踏まえ、NBDC運営委員会人材育成分科会において、研究データを整備・活用するバイオインフォマティクス人材を育成するためのカリキュラムに基づき実施した講習会です。本カリキュラムは、対応が急務であると思われる次世代シーケンサデータに焦点をあてた内容となっております。

- [人材育成分科会で策定したカリキュラム](#)

[バイオインフォマティクス人材育成カリキュラム \(次世代シーケンサ\)](#) ①

[カリキュラムで習得できる技能](#)

[カリキュラム フロー図](#)

- [H26年度NGS速習コース講習会 \(2014年9月1日~12日\)](#)
- [H27年度NGSハンズオン講習会 \(2015年7月22日~8月6日\)](#)

# NGS用カリキュラム

カリキュラムは、最低限必要とされる知識・技術を2週間程度で身につけることを想定した「速習」と、時間をかけて習得することを想定した「速習以外」に分かれている。

大項目	日数	No.	項目	習得技術	初級	中級	上級
1. コンピュータリテラシーとサーバー設計	4日	2日	1-1 OS, ハード構成	・コンピュータの基本の理解			
			1-2 ネットワーク基礎	・インターネット、セキュリティの基本の理解	初級	講義	
			1-3 UNIX I	・UNIXの基礎の理解 ・Linux導入	中級	実習	
	2日	1-4 スクリプト言語	・Perl ・シェルスクリプト	中級	実習		
2. 配列インフォマティクス	1日	2-1 配列解析基礎	・配列、ゲノムデータ記述のフォーマット、アラインメント(DP)、データベース検索(BLAST, BLAT)等の基礎的な配列比較解析の原理と実習	初級	実習		
		2-2 バイオ系データベース概論	・基本的な各種バイオ系データベースの理解、統合DBの利用法	初級	実習		
3. データ解析基礎	2日	3-1 R 基礎1	・R言語の基礎(インストールから利用まで)	初級	実習		
		3-2 R 基礎2	・ファイルの読み込み、行列演算の基本	初級	実習		
		3-3 R 各種パッケージ	・Rの各種パッケージのインストール法と代表的なパッケージの利用法	中級	実習		
		3-4 R bioconductor I	・bioconductorの利用法	中級	実習		
		3-5 R bioconductor II	・FASTA and FASTQ形式ファイルの読み込み。ファイル形式の変換(FASTQ → FASTA)、クオリティチェック、リード配列長分布、フィルタリングやトリミング、GC含量計算など	中級	実習		
4. 次世代シーケンサ	0.25日	4-1 次世代シーケンサ基礎I	・原理の理解	初級	講義		
		4-2 次世代シーケンサ基礎II	・応用分野とそのための計測技術の理解(RNA-seq, ChIP-seq, がんゲノム, 個人ゲノム, 環境ゲノム, Hi-C)	初級	講義		
	0.5日	4-3 次世代シーケンサ実習I	・ファイル形式、可視化、quality check、マッピング、アセンブル	初級	実習		
	1.5日	4-4 次世代シーケンサ実習II	・代表的なパイプラインについての実習(多型解析(IGV)、RNA-seq、ChIP-seq、及び統合解析、定番のツールを利用)	初級	実習		
5. ゲノム関連の倫理・法律	0.5日	5-1	ゲノム情報倫理概論	・ゲノム情報を扱う上で、プライバシー保護などの必要な倫理的問題、法的問題の国内外の状況を理解し、ゲノム情報を適切に利用できるようにする ・匿名化、暗号化、情報セキュリティ概要	初級	講義	
6. 分子生命科学	0.5日	6-1	分子生命科学概論	・複製、転写、翻訳、代謝、シグナル伝達などの基礎知識	初級	講義	
		6-2	オミクス概論	・ゲノム以外のオミクスデータの基礎知識	初級	講義	
		6-3	遺伝/進化概論	・ゲノムデータを扱う上での遺伝学、進化学の基礎知識	初級	講義	

平成26年9月に、終日2週間ぶつ  
通し(土日除く)で講習会を開催。

# NGS速習コース講習会



The screenshot shows the homepage of the National Bioscience Database Center (NBDC). The header includes the NBDC logo, the text "National Bioscience Database Center", and the Japanese text "バイオサイエンスデータベースセンター". There are navigation links for "English", "サイトマップ", and "サイト内検索". A search bar is present with the text "検索...". The main navigation menu includes "ホーム", "NBDCについて", "研究開発", "公募情報", "採用情報", "広報", "人材支援", "お問い合わせ", and "リンク". The breadcrumb trail is "Home > 人材支援 > 支援 > 講習会". The main content area is titled "バイオインフォマティクス人材育成のための講習会" and "NGS解析". The text describes the workshop's purpose: to train bioinformatics talent for NGS data analysis. It mentions the "NBDC運営委員会人材育成分科会" and the curriculum's focus on next-generation sequencing data. A list of workshops is provided, with the 2014 workshop (H26年度) highlighted by a red arrow and a circled number 1.

Home > 人材支援 > 支援 > 講習会

## バイオインフォマティクス人材育成のための講習会

### NGS解析

ライフサイエンス分野の研究現場においては、莫大・多様な研究データが産出され、取り扱うデータ量が飛躍的に増えています。一方で、それらのデータを整備・活用するための人材は不足している状況です。そのような研究現場の状況を踏まえ、NBDC運営委員会人材育成分科会において、研究データを整備・活用するバイオインフォマティクス人材を育成するためのカリキュラムに基づき実施した講習会です。本カリキュラムは、対応が急務であると思われる次世代シーケンサデータに焦点をあてた内容となっております。

- [人材育成分科会で策定したカリキュラム](#)

[バイオインフォマティクス人材育成カリキュラム \(次世代シーケンサ\)](#)

[カリキュラムで習得できる技能](#)

[カリキュラム フロー図](#)

- [H26年度NGS速習コース講習会 \(2014年9月1日~12日\)](#) ①
- [H27年度NGSハンズオン講習会 \(2015年7月22日~8月6日\)](#)

# NGS速習コース講習会

■平成26年度NGS速習コース講習会（2014年9月1日～12日）

「バイオインフォマティクス人材育成カリキュラム（次世代シーケンサ）」の速習部分について、講習会を開催しました。

H26年度概要

H26年度講義日程・参考資料

H26年度講習会時の情報については[こちら](#)もご覧ください。

H26年度講義資料・動画

●講習会実施報告書（PDFファイル：3.7MB）

●講師資料・動画



実施日	実施時間	大項目	項目番号および項目	レベル	形式	担当講師（敬称略）	講義資料・動画
9月1日	10:40-12:00	1.コンピュータリテラシーとサーバー設計	1-1. OS、ハード構成	初級	講義	中村 保一 (DDBJ)	<a href="#">統合TV</a>
	13:15-14:45		1-2. ネットワーク基礎				
	15:00-16:30		1-3. UNIX I	初級	実習	仲里 猛留 (DBCLS)	<a href="#">統合TV</a>
	16:45-18:15						
9月2日	10:30-12:00			中級	実習	仲里 猛留 (DBCLS)	<a href="#">統合TV</a>
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
9月3日	10:30-12:00		1-4. スクリプト言語	中級	実習	服部 恵美(アメリカ)	<a href="#">統合TV</a>

# NGS速習コース講習会

①報告書はこちら。実際に受けたヒトの感想や要望・不満をそのままコピペで掲載。

## ■平成26年度NGS速習コース講習会（2014年9月1日～12日）

「バイオインフォマティクス人材育成カリキュラム（次世代シーケンサ）」の速習部分について、講習会を開催しました。

### H26年度概要

### H26年度講義日程・参考資料

H26年度講習会時の情報については[こちら](#)もご覧ください。

### H26年度講義資料・動画

● [講習会実施報告書（PDFファイル：3.7MB）](#)

①

● [講師資料・動画](#)

実施日	実施時間	大項目	項目番号および項目	レベル	形式	担当講師（敬称略）	講義資料・動画
9月1日	10:40-12:00	1.コンピュータリテラシーとサーバー設計	1-1. OS、ハード構成	初級	講義	中村 保一 (DDBJ)	<a href="#">統合TV</a>
	13:15-14:45		1-2. ネットワーク基礎				
	15:00-16:30		1-3. UNIX I	初級	実習	仲里 猛留 (DBCLS)	<a href="#">統合TV</a>
	16:45-18:15						
9月2日	10:30-12:00			中級	実習	仲里 猛留 (DBCLS)	<a href="#">統合TV</a>
	13:15-14:45						
	15:00-16:30						
	16:45-18:15						
9月3日	10:30-12:00		1-4. スクリプト言語	中級	実習	服部 恵美(アメリカエフ)	<a href="#">統合TV</a>

講義の時間配分はほぼカリキュラム通り。  
 ①Linux (UNIX)の基礎は、約1.5日分。

# 講義日程、項目、内容

実施日	実施時間	項目番号および項目	レベル	形式	担当講師(所属)
9月1日	10:40-12:00	1-1. OS、ハード構成	初級	講義	中村保一(DDBJ)
	13:15-14:45	1-2. ネットワーク基礎	初級	講義	中村保一(DDBJ)
	15:00-18:15	1-3. UNIX I	初級	実習	仲里猛留(DBCLS)
9月2日	10:30-18:15	1-3. UNIX I	中級	実習	仲里猛留(DBCLS)
9月3日	10:30-18:15	1-4. スクリプト言語(シェルスクリプト)	中級	実習	服部恵美(Amelieff)
9月4日	10:30-18:15	1-4. スクリプト言語(Perl)	中級	実習	服部恵美(Amelieff)
9月5日	10:30-14:45	2-1. 配列解析基礎	初級	実習	坊農秀雅(DBCLS)
	15:00-18:15	2-2. バイオ系データベース概論	初級	実習	小野浩雅(DBCLS)
9月8日	10:30-12:00	3-1. R基礎1	初級	実習	門田幸二(東京大学)
	13:15-14:45	3-2. R基礎2	初級	実習	門田幸二(東京大学)
	15:00-18:15	3-3. R各種パッケージ	中級	実習	門田幸二(東京大学)
9月9日	10:30-14:45	3-4. R Bioconductor I	中級	実習	門田幸二(東京大学)
	15:00-18:15	3-5. R Bioconductor II	中級	実習	門田幸二(東京大学)
9月10日	10:30-12:00	4-1. 次世代シーケンサ基礎I	初級	講義	倉田哲也(NAIST)
	13:15-14:45	4-2. 次世代シーケンサ基礎II	初級	講義	倉田哲也(NAIST)
	15:00-18:15	4-3. 次世代シーケンサ実習I	初級	実習	山口昌雄(Amelieff)
9月11日	10:30-14:45	4-4. 次世代シーケンサ実習II (ChIP-seq)	初級	実習	河岡慎平(ATR)
	15:00-16:30	6. 分子生命科学	初級	講義	河岡慎平(ATR)
	16:45-18:15	5-1. ゲノム情報倫理概論	初級	講義	箕輪真理(NBDC) 川嶋実苗(NBDC)
9月12日	10:30-14:45	4-4. 次世代シーケンサ実習II (多型解析)	初級	実習	山口昌雄(Amelieff)
	15:00-18:15	4-4. 次世代シーケンサ実習II (RNA-seq)	初級	実習	山口昌雄(Amelieff)



全部ハンズオン講義(実習)形式にして!もちろん個別の講師への不満ではなく、カリキュラム改善要望

# 受講生の要望・不満

実施日	実施時間	項目番号および項目	レベル	形式	担当講師(所属)
9月1日	10:40-12:00	1-1. OS、ハード構成	初級	講義	中村保一(DDBJ)
	13:15-14:45	1-2. ネットワーク基礎	初級	講義	中村保一(DDBJ)
	15:00-18:15	1-3. UNIX I	初級	実習	仲里猛留(DBCLS)
9月2日	10:30-18:15	1-3. UNIX I	中級	実習	仲里猛留(DBCLS)
9月3日	10:30-18:15	1-4. スクリプト言語(シェルスクリプト)	中級	実習	服部恵美(Amelieff)
9月4日	10:30-18:15	1-4. スクリプト言語(Perl)	中級	実習	服部恵美(Amelieff)
9月5日	10:30-14:45	2-1. 配列解析基礎	初級	実習	坊農秀雅(DBCLS)
	15:00-18:15	2-2. バイオ系データベース概論	初級	実習	小野浩雅(DBCLS)
9月8日	10:30-12:00	3-1. R基礎1	初級	実習	門田幸二(東京大学)
	13:15-14:45	3-2. R基礎2	初級	実習	門田幸二(東京大学)
	15:00-18:15	3-3. R各種パッケージ	中級	実習	門田幸二(東京大学)
9月9日	10:30-14:45	3-4. R Bioconductor I	中級	実習	門田幸二(東京大学)
	15:00-18:15	3-5. R Bioconductor II	中級	実習	門田幸二(東京大学)
9月10日	10:30-12:00	4-1. 次世代シーケンサ基礎I	初級	講義	倉田哲也(NAIST)
	13:15-14:45	4-2. 次世代シーケンサ基礎II	初級	講義	倉田哲也(NAIST)
	15:00-18:15	4-3. 次世代シーケンサ実習I	初級	実習	山口昌雄(Amelieff)
9月11日	10:30-14:45	4-4. 次世代シーケンサ実習II (ChIP-seq)	初級	実習	河岡慎平(ATR)
	15:00-16:30	6. 分子生命科学	初級	講義	河岡慎平(ATR)
	16:45-18:15	5-1. ゲノム情報倫理概論	初級	講義	箕輪真理(NBDC) 川嶋実苗(NBDC)
9月12日	10:30-14:45	4-4. 次世代シーケンサ実習II (多型解析)	初級	実習	山口昌雄(Amelieff)
	15:00-18:15	4-4. 次世代シーケンサ実習II (RNA-seq)	初級	実習	山口昌雄(Amelieff)



計10日分を10人で担当。講師数が増えるほど全体の連携は低下傾向に...

# 受講生の要望・不満

実施日	実施時間	項目番号および項目	レベル	形式	担当講師(所属)
9月1日	10:40-12:00	1-1. OS、ハード構成	初級	講義	中村保一(DDBJ)
	13:15-14:45	1-2. ネットワーク基礎	初級	講義	中村保一(DDBJ)
	15:00-18:15	1-3. UNIX I	初級	実習	仲里猛留(DBCLS)
9月2日	10:30-18:15	1-3. UNIX I	中級	実習	仲里猛留(DBCLS)
9月3日	10:30-18:15	1-4. スクリプト言語(シェルスクリプト)	中級	実習	服部恵美(Amelieff)
9月4日	10:30-18:15	1-4. スクリプト言語(Perl)	中級	実習	服部恵美(Amelieff)
9月5日	10:30-14:45	2-1. 配列解析基礎	初級	実習	坊農秀雅(DBCLS)
	15:00-18:15	2-2. バイオ系データベース概論	初級	実習	小野浩雅(DBCLS)
9月8日	10:30-12:00	3-1. R基礎1	初級	実習	門田幸二(東京大学)
	13:15-14:45	3-2. R基礎2	初級	実習	門田幸二(東京大学)
	15:00-18:15	3-3. R各種パッケージ	中級	実習	門田幸二(東京大学)
9月9日	10:30-14:45	3-4. R Bioconductor I	中級	実習	門田幸二(東京大学)
	15:00-18:15	3-5. R Bioconductor II	中級	実習	門田幸二(東京大学)
9月10日	10:30-12:00	4-1. 次世代シーケンサ基礎I	初級	講義	倉田哲也(NAIST)
	13:15-14:45	4-2. 次世代シーケンサ基礎II	初級	講義	倉田哲也(NAIST)
	15:00-18:15	4-3. 次世代シーケンサ実習I	初級	実習	山口昌雄(Amelieff)
9月11日	10:30-14:45	4-4. 次世代シーケンサ実習II (ChIP-seq)	初級	実習	河岡慎平(ATR)
	15:00-16:30	6. 分子生命科学	初級	講義	河岡慎平(ATR)
	16:45-18:15	5-1. ゲノム情報倫理概論	初級	講義	箕輪真理(NBDC) 川嶋実苗(NBDC)
9月12日	10:30-14:45	4-4. 次世代シーケンサ実習II (多型解析)	初級	実習	山口昌雄(Amelieff)
	15:00-18:15	4-4. 次世代シーケンサ実習II (RNA-seq)	初級	実習	山口昌雄(Amelieff)

# NGSハンズオン講習会



The screenshot shows the homepage of the National Bioscience Database Center (NBDC). The main navigation bar includes links for Home, NBDCについて, 研究開発, 公募情報, 採用情報, 広報, 人材支援, お問い合わせ, and リンク. The breadcrumb trail is Home > 人材支援 > 支援 > 講習会. The main content area is titled "バイオインフォマティクス人材育成のための講習会" and features a section for "NGS解析". The text describes the workshop's purpose: to train bioinformatics talent by focusing on next-generation sequencing (NGS) data, which is becoming increasingly common in life science research. It mentions that the curriculum is based on the NBDC's talent development committee's recommendations. A list of workshops is provided, with the 2015 workshop (July 22 - August 6) highlighted by a red arrow and the number 1.

NBDC National Bioscience Database Center

独立研究開発法人 科学技術振興機構

文字サイズ変更 大 中 小

English サイトマップ サイト内検索 検索...

ホーム NBDCについて 研究開発 公募情報 採用情報 広報 人材支援 お問い合わせ リンク

Home > 人材支援 > 支援 > 講習会

## バイオインフォマティクス人材育成のための講習会

### NGS解析

ライフサイエンス分野の研究現場においては、莫大・多様な研究データが産出され、取り扱うデータ量が飛躍的に増えています。一方で、それらのデータを整備・活用するための人材は不足している状況です。そのような研究現場の状況を踏まえ、NBDC運営委員会人材育成分科会において、研究データを整備・活用するバイオインフォマティクス人材を育成するためのカリキュラムに基づき実施した講習会です。本カリキュラムは、対応が急務であると思われる次世代シーケンサデータに焦点をあてた内容となっております。

- [人材育成分科会で策定したカリキュラム](#)

[バイオインフォマティクス人材育成カリキュラム \(次世代シーケンサ\)](#)

[カリキュラムで習得できる技能](#)

[カリキュラム フロー図](#)

- [H26年度NGS速習コース講習会 \(2014年9月1日～12日\)](#)
- [H27年度NGSハンズオン講習会 \(2015年7月22日～8月6日\)](#) ①

# NGSハンズオン講習会

受講生の要望に応じて…①講師数を減らし項目間の連携強化(10 → 4人)。  
②ハンズオン講義(実習)のみ。

- 7月22日(水): Bio-Linux 8とRのインストール状況確認。基本自習(門田・寺田 透先生)
- 7月23日(木): Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田)
- 7月24日(金): スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月): スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火): スクリプト言語。**Python**(アメリエフ 服部恵美先生)
- 7月29日(水): データ解析環境R(門田)
- 7月30日(木): データ解析環境R(門田)
- 8月3日(月): NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月4日(火): NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月5日(水): NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口昌雄先生、後半:門田)
- 8月6日(木): NGS解析。ChIP-seq(理研 森岡勝樹先生)
- 8月26日(水): 予備日
- 8月27日(木): 予備日
- 8月28日(金): 予備日

# 申込者数(5/15締切時点)

最も受講希望が多かった8/3-5の項目(113名、110名、113名)を8/26-28の予備日に開催して分散。

- 7月22日(水):85名。Bio-Linux 8とRのインストール状況確認。基本自習(門田・寺田 透先生)
- 7月23日(木):93名。Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田)
- 7月24日(金):86名。スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月):94名。スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火):90名。スクリプト言語。Python(アメリエフ 服部恵美先生)
- 7月29日(水):96名。データ解析環境R(門田)
- 7月30日(木):98名。データ解析環境R(門田)
- 8月3日(月):113名。NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月4日(火):110名。NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月5日(水):113名。NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口先生、後半:門田)
- 8月6日(木):104名。NGS解析。ChIP-seq(理研 森岡勝樹先生)
- 8月26日(水):予備日
- 8月27日(木):予備日
- 8月28日(金):予備日



# 申込者数(5/15締切時点)

最も受講希望の多かった8/3-5の項目(113名、110名、113名)を8/26-28の予備日に開催して分散。

- 7月22日(水):85名。Bio-Linux 8とRのインストール状況確認。基本自習(門田・寺田 透先生)
- 7月23日(木):93名。Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田)
- 7月24日(金):86名。スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月):94名。スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火):90名。スクリプト言語。Python(アメリエフ 服部恵美先生)
- 7月29日(水):96名。データ解析環境R(門田)
- 7月30日(木):98名。データ解析環境R(門田)
- 8月3日(月):87名。NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月4日(火):84名。NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月5日(水):85名。NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口先生、後半:門田)
- 8月6日(木):104名。NGS解析。ChIP-seq(理研 森岡勝樹先生)
- 8月26日(水):26名。NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月27日(木):26名。NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月28日(金):28名。NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口先生、後半:門田)

# 申込者数(7/22開催時点)

- 7月22日(水):81名。Bio-Linux 8とRのインストール状況確認。基本自習(門田・寺田 透先生)
- 7月23日(木):86名。Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田)
- 7月24日(金):80名。スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月):84名。スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火):83名。スクリプト言語。Python(アメリエフ 服部恵美先生)
- 7月29日(水):80名。データ解析環境R(門田)
- 7月30日(木):85名。データ解析環境R(門田)
- 8月3日(月):82名。NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月4日(火):77名。NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月5日(水):78名。NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口先生、後半:門田)
- 8月6日(木):95名。NGS解析。ChIP-seq(理研 森岡勝樹先生)
- 8月26日(水):24名。NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月27日(木):25名。NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月28日(金):28名。NGS解析。RNA-seq、統計解析(前半:アメリエフ 山口先生、後半:門田)

# 申込者数→受講者数

## A日程

- 7月22日(水): 81→59名。Bio-Linux 8とRのインストール状況確認。基本自習(門田・寺田先生)
- 7月23日(木): 86→67名。Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田)
- 7月24日(金): 80→69名。スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月): 84→67名。スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火): 83→64名。スクリプト言語。Python(アメリエフ 服部恵美先生)
- 7月29日(水): 80→68名。データ解析環境R(門田)
- 7月30日(木): 85→68名。データ解析環境R(門田)
- 8月3日(月): 82→71名。NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月4日(火): 77→59名。NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月5日(水): 78→65名。NGS解析。RNA-seq、統計解析(アメリエフ 山口先生、門田)
- 8月6日(木): 95→71名。NGS解析。ChIP-seq(理研 森岡勝樹先生)
- 8月26日(水): 24→32名。NGS解析。基礎(アメリエフ 山口昌雄先生)
- 8月27日(木): 25→29名。NGS解析。ゲノムReseq、変異解析(アメリエフ 山口昌雄先生)
- 8月28日(金): 28→33名。NGS解析。RNA-seq、統計解析(アメリエフ 山口先生、門田)

## B日程

# 講習会では1日だけが...

受講希望者には、一週間程度はかかるかなり厳しい事前準備や予習を要求し、Linuxコマンドがある程度使えることを前提として①7/23のハンズオン講義を行った。

- 7月22日(水): 81→59名。Bio-Linux 8とRのインストール状況確認。
- 7月23日(木): 86→67名。Linux基礎。LinuxコマンドなどUNIXの基礎の理解(門田) ①
- 7月24日(金): 80→69名。スクリプト言語。シェルスクリプト(アメリエフ株式会社 服部恵美先生)
- 7月27日(月): 84→67名。スクリプト言語。Perl(アメリエフ 服部恵美先生)
- 7月28日(火): 83→64名。スクリプト言語。Python(アメリエフ 服部恵美先生) ①
- 7月29日(水): 80→68名。デ
- 7月30日(木): 85→68名。デ
- 8月3日(月): 82→71名。NGS
- 8月4日(火): 77→59名。NGS
- 8月5日(水): 78→65名。NGS
- 8月6日(木): 95→71名。NGS
- 8月26日(水): 24→32名。NG
- 8月27日(木): 25→29名。NG
- 8月28日(金): 28→33名。NG





自分の置かれている環境・予算・ポリシーによっても異なる。どの選択肢でも正解かも…。

# ビッグ (NGS) データ解析戦略

- 解析受託企業に外注: Linuxコマンドを知らなくてもよい

-  ngs 受託解析



- クラウド(ウェブツール): Linuxコマンドを知らなくてもよい

- DDBJ Pipeline
- Illumina BaseSpace
- Galaxy
- ...



- Linuxコマンドを駆使(旧来型)

- なるべく自力で解析
- LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
- NBDC/東大アグリバイオの「NGSハンズオン講習会」の方向性



# 統計は重要

ボタンを押せば、データ解析結果を簡単に得られる時代。重要なのは、結果を正しく解釈する能力や統計的なものの考え方を身につけること。

## (Rで)マイクロアレイデータ解析

(last modified 2015/01/16, since 2005)

### What's new?

- ・ 門田幸二 著 **シリーズ Useful R 第7巻** **トランスクリプトーム解析** 刊行(共立出版)。マイクロアレイ解析に関する最近の知見や、ROKU法 (Kadota et al., 2006)、WAD法 (Kadota et al., 2008) などについての解説も含まれています。書籍中のマイクロアレイ解析部分のRコードについては、このページの「書籍 | トランスクリプトーム解析」に掲載してあります。(2014/04/27)
- ・ お知らせは主に **(Rで)塩基配列解析** で行ってお資料なども **(Rで)塩基配列解析** 中の **参考資料**

- ・ [はじめに](#) (last modified 2014/05/14)
- ・ [過去のお知らせ](#) (last modified 2014/03/03)
- ・ [Rのインストールと起動](#) (last modified 2014/05/02)
- ・ [Rの昔のバージョンのインストール](#) (last modified 2014/05/02)
- ・ [使用例\(初心者向け\)](#) (last modified 2011/09/15)
- ・ [サンプルデータ](#) (last modified 2014/06/02)
- ・ [書籍 II について](#) (last modified 2014/05/12)
- ・ [書籍 | トランスクリプトーム解析 | 1.1はじめに](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.1 原理\(A\)](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.2 最近の](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.1 生データ](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.2 データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.3 データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.4 データの](#)

## (Rで)塩基配列解析

~NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル、バイオインフォマティクス~  
(last modified 2015/09/25, since 2011)

### What's new?

- ・ このウェブページは **インストール II について** の推奨手順 (**Windows2015.04.04版**と**Macintosh2015.04.03版**)に従ってフリーソフト R と必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は **基本的な利用法** (**Windows2015.04.03版**と**Macintosh2015.04.03版**)で自習してください。本ウェブページを体系的にまとめた **書籍** もあります。(2015/04/03)
- ・ **日本乳酸菌学会誌**のNGS関連連載の **第4回分ウェブ資料PDF**において、2015年9月18日に別のノートPCで、共有フォルダ関連の「`sudo apt-get install dkms`」(W4-5-2)実行時にエラーに遭遇したので、対処法を記載しました。(2015/09/18) **NEW**
- ・ **sample1.fastq**のようなコンティグ数が1つしかない場合に、`rowSums(x)`の計算時にエラーが出ることがわかったので、該当箇所を`apply(as.matrix(x), 1, sum)`のような感じに変更しました。(2015/09/12)
- ・ **NGSハンズオン講習会2015**のアメリカン様分(服部先生と山口先生)の講義資料を差し替えました。(2015/09/03)

- ・ [はじめに](#) (last modified 2015/03/31)
- ・ [参考資料\(講義、講習会、本など\)](#) (last modified 2015/09/09)
- ・ [過去のお知らせ](#) (last modified 2015/09/18) **NEW**
- ・ [インストール II について](#) (last modified 2015/04/04)
- ・ [インストール | R本体 | 最新版 | Win用](#) (last modified 2015/03/22) 推奨
- ・ [インストール | R本体 | 最新版 | Mac用](#) (last modified 2015/04/22) 推奨
- ・ [インストール | R本体 | 過去版 | Win用](#) (last modified 2015/03/22)
- ・ [インストール | R本体 | 過去版 | Mac用](#) (last modified 2015/03/22)
- ・ [インストール | Rパッケージ | まぼ全て\(20GB以上?\)](#) (last modified 2015/05/25)
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(数GB?\)](#) (last modified 2015/07/24) 推奨
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(アグリバイオ 居室のみ\)](#) (last modified 2015/06/16) [トップページへ](#)
- ・ [インストール | Rパッケージ | 必要最小限\(数GB?\)](#) (last modified 2015/05/25)

# 2つのウェブページ

私(門田)が提供している、よりよいデータ解析のための指南書、ノウハウ集、リンク集のようなもの。

## (Rで)マイクロアレイデータ解析

(last modified 2015/01/16, since 2005)

### What's new?

- ・門田幸二著シリーズ **Useful R 第7巻** トランスクリプトーム解析刊行(共立出版)。マイクロアレイ解析に関する最近の知見や、ROKU法(Kadota et al., 2006)、WAD法(Kadota et al., 2008)などについての解説も含んでいます。書籍中のマイクロアレイ解析部分のRコードについては、このページの「書籍|トランスクリプトーム解析」に掲載してあります。(2014/04/27)
- ・お知らせは主に(Rで)塩基配列解析で行って資料なども(Rで)塩基配列解析中の参考資料

- ・ [はじめに](#) (last modified 2014/05/14)
- ・ [過去のお知らせ](#) (last modified 2014/03/03)
- ・ [Rのインストールと起動](#) (last modified 2014/03/03)
- ・ [Rの昔のバージョンのインストール](#) (last modified 2014/03/03)
- ・ [使用例\(初心者向け\)](#) (last modified 2011/09/15)
- ・ [サンプルデータ](#) (last modified 2014/06/02)
- ・ [書籍 II について](#) (last modified 2014/05/12)
- ・ [書籍 | トランスクリプトーム解析 | 1.1はじめに](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.1 原理\(A\)](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.2 最近の](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.1 生データ](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.2 データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.3 データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.4 データの](#)

## (Rで)塩基配列解析

~NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル、バイオインフォーマティクス~  
(last modified 2015/09/25, since 2011)

### What's new?

- ・このウェブページは [インストール II について](#) の推奨手順 ([Windows2015.04.04版](#)と[Macintosh2015.04.03版](#))に従ってフリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は [基本的な利用法](#) ([Windows2015.04.03版](#)と[Macintosh2015.04.03版](#))で自習してください。本ウェブページを体系的にまとめた [書籍](#) もあります。(2015/04/03)
- ・ [日本乳酸菌学会誌](#)のNGS関連連載の [第4回分ウェブ資料PDF](#)において、2015年9月18日に別のノートPCで、共有フォルダ関連の「`sudo apt-get install dkms`」(W4-5-2)実行時にエラーに遭遇したので、対処法を記載しました。(2015/09/18) **NEW**
- ・ [sample1.fastq](#)のようなコンティグ数が1つしかない場合に、`rowSums(x)`の計算時にエラーが出るのがわかったので、該当箇所を`apply(as.matrix(x), 1, sum)`のような感じに変更しました。(2015/09/12)
- ・ [NGSハンズオン講習会](#)2015のアメリカ様分(服部先生と山口先生)の講義資料を差し替えました。(2015/09/03)

- ・ [はじめに](#) (last modified 2015/03/31)
- ・ [参考資料\(講義、講習会、本など\)](#) (last modified 2015/09/09)
- ・ [過去のお知らせ](#) (last modified 2015/09/18) **NEW**
- ・ [インストール II について](#) (last modified 2015/04/04)
- ・ [インストール | R本体 | 最新版 | Win用](#) (last modified 2015/03/22) 推奨
- ・ [インストール | R本体 | 最新版 | Mac用](#) (last modified 2015/04/22) 推奨
- ・ [インストール | R本体 | 過去版 | Win用](#) (last modified 2015/03/22)
- ・ [インストール | R本体 | 過去版 | Mac用](#) (last modified 2015/03/22)
- ・ [インストール | Rパッケージ | まぼ全て\(20GB以上?\)](#) (last modified 2015/05/25)
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(数GB?\)](#) (last modified 2015/07/24) 推奨
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(アグリバイオ居室のみ\)](#) (last modified 2015/06/16) [トップページへ](#)
- ・ [インストール | Rパッケージ | 必要最小限\(数GB?\)](#) (last modified 2015/05/25)

# 2つのウェブページ

私(門田)が提供している、よりよいデータ解析のための指南書、ノウハウ集、リンク集のようなもの。③講義や講演資料などの各種教材も含む。

## (Rで)マイクロアレイデータ解析

(last modified 2015/01/16, since 2005)

### What's new?

- ・門田幸二著シリーズ **Useful R 第7巻** **トランスクリプトーム解析** 刊行(共立出版)。マイクロアレイ解析に関する最近の知見や、ROKU法(Kadota et al., 2006)、WAD法(Kadota et al., 2008)などについての解説も含んでいます。書籍中のマイクロアレイ解析部分のRコードについては、このページの「書籍|トランスクリプトーム解析」に掲載してあります。(2014/04/27)
- ・お知らせは主に(Rで)塩基配列解析で行っており、資料なども(Rで)塩基配列解析中の参考資料

## (Rで)塩基配列解析

~NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル、バイオインフォマティクス~  
(last modified 2015/09/25, since 2011)

### What's new?

- ・このウェブページはインストール|についての推奨手順(Windows2015.04.04版とMacintosh2015.04.03版)に従ってフリーソフトRと必要なパッケージをインストール済みであるという前提で記述しています。初心者の方は基本的な利用法(Windows2015.04.03版とMacintosh2015.04.03版)で自習してください。本ウェブページを体系的にまとめた書籍もあります。(2015/04/03)
- ・日本乳酸菌学会誌のNGS関連連載の第4回分ウェブ資料PDFにおいて、2015年9月18日に別のノートPCで、共有フォルダ関連の「sudo apt-get install dkms」(W4-5-2)実行時にエラーに遭遇したので、対処法を記載しました。(2015/09/18) **NEW**
- ・sample1.fastqのようなコンティグ数が1つしかない場合に、rowSums(x)の計算時にエラーが出ることがわかったので、該当箇所をapply(as.matrix(x), 1, sum)のような感じに変更しました。(2015/09/12)
- ・NGSハンズオン講習会2015のアメリカ様分(服部先生と山口先生)の講義資料を差し替えました。(2015/09/03)

- ・はじめに (last modified 2015/03/31)
- ・参考資料(講義、講習会、本など) (last modified 2015/09/09)
- ・過去のお知らせ (last modified 2015/06/16) **NEW**
- ・インストール|について (last modified 2015/04/04)
- ・インストール|R本体|最新版|Win用 (last modified 2015/03/22) 推奨
- ・インストール|R本体|最新版|Mac用 (last modified 2015/04/22) 推奨
- ・インストール|R本体|過去版|Win用 (last modified 2015/03/22)
- ・インストール|R本体|過去版|Mac用 (last modified 2015/03/22)
- ・インストール|Rパッケージ|ほぼ全て(20GB以上?) (last modified 2015/05/25)
- ・インストール|Rパッケージ|必要最小限プラスアルファ(数GB?) (last modified 2015/07/24) 推奨
- ・インストール|Rパッケージ|必要最小限プラスアルファ(アグリバイオ居室のみ) (last modified 2015/06/16) [トップページへ](#)
- ・インストール|Rパッケージ|必要最小限(数GB?) (last modified 2015/05/25)

タイトルが「(Rで)...」となっているのは、フリーソフトウェアRを用いた統計解析手法周辺を記載しているため。

# 統計といえばR

## (Rで)マイクロアレイデータ解析

(last modified 2015/01/16, since 2005)

### What's new?

- ・ 門田幸二 著 **シリーズ Useful R 第7巻 トランスクリプトーム解析** 刊行(共立出版)。マイクロアレイ解析に関する最近の知見や、ROKU法 (Kadota et al., 2006)、WAD法 (Kadota et al., 2008) などについての解説も含まれています。書籍中のマイクロアレイ解析部分のRコードについては、このページの「書籍 | トランスクリプトーム解析」に掲載してあります。(2014/04/27)
- ・ お知らせは主に **(Rで)塩基配列解析** で行っており、資料なども **(Rで)塩基配列解析** 中の **参考資料** があります。

- ・ [はじめに](#) (last modified 2014/05/14)
- ・ [過去のお知らせ](#) (last modified 2014/03/03)
- ・ [Rのインストールと起動](#) (last modified 2014/05/05)
- ・ [Rの昔のバージョンのインストール](#) (last modified 2014/05/05)
- ・ [使用例\(初心者向け\)](#) (last modified 2011/09/15)
- ・ [サンプルデータ](#) (last modified 2014/06/02)
- ・ [書籍 | Rについて](#) (last modified 2014/05/12)
- ・ [書籍 | トランスクリプトーム解析 | 1.1はじめに](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.1 原理\(A\)](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.2 最近の\(A\)](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.1 生データ](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.2 データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.3 データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.4 データの](#)

## (Rで)塩基配列解析

~NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル  
(last modified 2015/09/25, since 2011)

### What's new?

- ・ このウェブページは **インストール | Rについて** の推奨手順 (**Windows2015.04.04版** とソフト R と必要なパッケージをインストール済みであるという前提で記述しています (**Windows2015.04.03版** と **Macintosh2015.04.03版**) で自習してください。本ウェブページ。(2015/04/03)
- ・ **日本乳酸菌学会誌** の NGS 関連連載の **第4回分ウェブ資料PDF** において、2015年9月関連の「`sudo apt-get install dkms`」(W4-5-2) 実行時にエラーに遭遇したので、対応として **sample1.fastq** のようなコンティグ数が1つしかない場合に、`rowSums(x)` の計算時間箇所を `apply(as.matrix(x), 1, sum)` のような感じに変更しました。(2015/09/12)
- ・ **NGSハンズオン講習会2015** のアメリカン様分(服部先生と山口先生)の講義資料

- ・ [はじめに](#) (last modified 2015/03/31)
- ・ [参考資料\(講義、講習会、本など\)](#) (last modified 2015/09/09)
- ・ [過去のお知らせ](#) (last modified 2015/09/18) **NEW**
- ・ [インストール | Rについて](#) (last modified 2015/04/04)
- ・ [インストール | R本体 | 最新版 | Win用](#) (last modified 2015/03/22) 推奨
- ・ [インストール | R本体 | 最新版 | Mac用](#) (last modified 2015/04/22) 推奨
- ・ [インストール | R本体 | 過去版 | Win用](#) (last modified 2015/03/22)
- ・ [インストール | R本体 | 過去版 | Mac用](#) (last modified 2015/03/22)
- ・ [インストール | Rパッケージ | まぼ全て\(20GB以上?\)](#) (last modified 2015/05/25)
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(数GB?\)](#) (last modified 2015/07/24) 推奨
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(アグリバイオ 居室のみ\)](#) (last modified 2015/06/16) [トップページへ](#)
- ・ [インストール | Rパッケージ | 必要最小限\(数GB?\)](#) (last modified 2015/05/25)



(半年かけて執筆した)教科書もあり。統計的なものの考え方や、数式の意味の理解を助ける内容(のつもり)。

# 統計といえばR

## (Rで)マイクロアレイデータ解析

(last modified 2015/01/16, since 2005)

### What's new?

- ・ 門田幸二 著 **シリーズ Useful R 第7巻** **トランスクリプトーム解析** 刊行(共立出版)。マイクロアレイ解析に関する最近の知見や、ROKUF法 (Kadota et al., 2006)、WAD法 (Kadota et al., 2008) などについての解説も含んでいます。書籍中のマイクロアレイ解析部分のRコードについては、このページの「書籍 | トランスクリプトーム解析」に掲載してあります。(2014/04/27)
- ・ お知らせは主に **(Rで)塩基配列解析** で行って資料なども **(Rで)塩基配列解析** 中の **参考資料**

- ・ [はじめに](#) (last modified 2014/05/14)
- ・ [過去のお知らせ](#) (last modified 2014/03/03)
- ・ [Rのインストールと起動](#) (last modified 2014/05/05)
- ・ [Rの昔のバージョンのインストール](#) (last modified 2014/05/05)
- ・ [使用例\(初心者向け\)](#) (last modified 2011/09/15)
- ・ [サンプルデータ](#) (last modified 2014/06/02)
- ・ [書籍 | 1](#) について (last modified 2014/05/12)
- ・ [書籍 | トランスクリプトーム解析 | 1.1](#) [はじめに](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.1](#) [原理\(A\)](#)
- ・ [書籍 | トランスクリプトーム解析 | 1.2.2](#) [最近の](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.1](#) [生データ](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.2](#) [データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.3](#) [データの](#)
- ・ [書籍 | トランスクリプトーム解析 | 2.2.4](#) [データの](#)

## (Rで)塩基配列解析

~NGS、RNA-seq、ゲノム、トランスクリプトーム、正規化、発現変動、統計、モデル  
(last modified 2015/09/25, since 2011)

### What's new?


- ・ このウェブページは [インストール | 1](#) についての推奨手順 ([Windows2015.04.04版](#) とソフト R と必要なパッケージをインストール済みであるという前提で記述しています ([Windows2015.04.03版](#) と [Macintosh2015.04.03版](#)) で自習してください。本ウェブページ。(2015/04/03)
- ・ [日本乳酸菌学会誌](#) の NGS 関連連載の [第4回分ウェブ資料PDF](#) において、2015年 9 月関連の「`sudo apt-get install dkms`」(W4-5-2) 実行時にエラーに遭遇したので、対応として `sample1.fastq` のようなコンティグ数が1つしかない場合に、`rowSums(x)` の計算時間箇所を `apply(as.matrix(x), 1, sum)` のような感じに変更しました。(2015/09/12)
- ・ [NGSハンズオン講習会2015](#) のアメリカ様分(服部先生と山口先生)の講義資料

- ・ [はじめに](#) (last modified 2015/03/31)
- ・ [参考資料\(講義、講習会、本など\)](#) (last modified 2015/09/09)
- ・ [過去のお知らせ](#) (last modified 2015/09/18) **NEW**
- ・ [インストール | 1](#) について (last modified 2015/04/04)
- ・ [インストール | R本体 | 最新版 | Win用](#) (last modified 2015/03/22) 推奨
- ・ [インストール | R本体 | 最新版 | Mac用](#) (last modified 2015/04/22) 推奨
- ・ [インストール | R本体 | 過去版 | Win用](#) (last modified 2015/03/22)
- ・ [インストール | R本体 | 過去版 | Mac用](#) (last modified 2015/03/22)
- ・ [インストール | Rパッケージ | まま全て\(20GB以上?\)](#) (last modified 2015/05/25)
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(数GB?\)](#) (last modified 2015/07/24) 推奨
- ・ [インストール | Rパッケージ | 必要最小限プラスアルファ\(アグリバイオ 居室のみ\)](#) (last modified 2015/06/16) [トップページへ](#)
- ・ [インストール | Rパッケージ | 必要最小限\(数GB?\)](#) (last modified 2015/05/25)



ビッグデータの扱いはいろいろ。自分の置かれている環境・予算・ポリシーによっても異なる。どの選択肢でも正解かも…。

# まとめ

- 解析受託企業に外注：Linuxコマンドを知らなくてもよい
  -  ngs 受託解析
- クラウド（ウェブツール）：Linuxコマンドを知らなくてもよい
  - DDBJ Pipeline
  - Illumina BaseSpace
  - Galaxy
  - …
- Linuxコマンドを駆使（旧来型）
  - なるべく自力で解析
  - LinuxコマンドやNGS解析用プログラムのインストールなどを練習し、スパコンを使いこなす
  - NBDC/東大アグリバイオの「NGSハンズオン講習会」の方向性

# 謝辞

- 科学技術振興機構バイオサイエンスデータベースセンター(JST-NBDC)  
高木 利久 先生、星 潤一 様、舘澤 博子 様、佐藤 恵子 様、白鳥 亜希子 様 他
- 関西学院大学  
藤 博幸 先生
- HPCI戦略プログラム分野1「予測する生命科学・医療および創薬基盤」  
人材養成プログラム(産業技術総合研究所 創薬基盤研究部門)  
杉原 稔 先生、坂井 寛子 様、寺田 朋子 様
- NGS講習会講師の先生方、TAのみなさま、および関係者の皆様
- 東京大学・大学院農学生命科学研究科アグリバイオインフォマティクス  
清水 謙多郎 先生、寺田 透 先生、三浦 文 様
- グラント
  - 基盤研究(C)(H24-26年度):「シーケンスに基づく比較トランスクリプトーム解析のためのガイドライン構築」(代表)
  - 新学術領域研究(研究領域提案型)(H22-26年度):「非モデル生物におけるゲノム解析法の確立」(分担;研究代表者:西山智明)
  - NBDCとの共同研究(H26-27年度)

