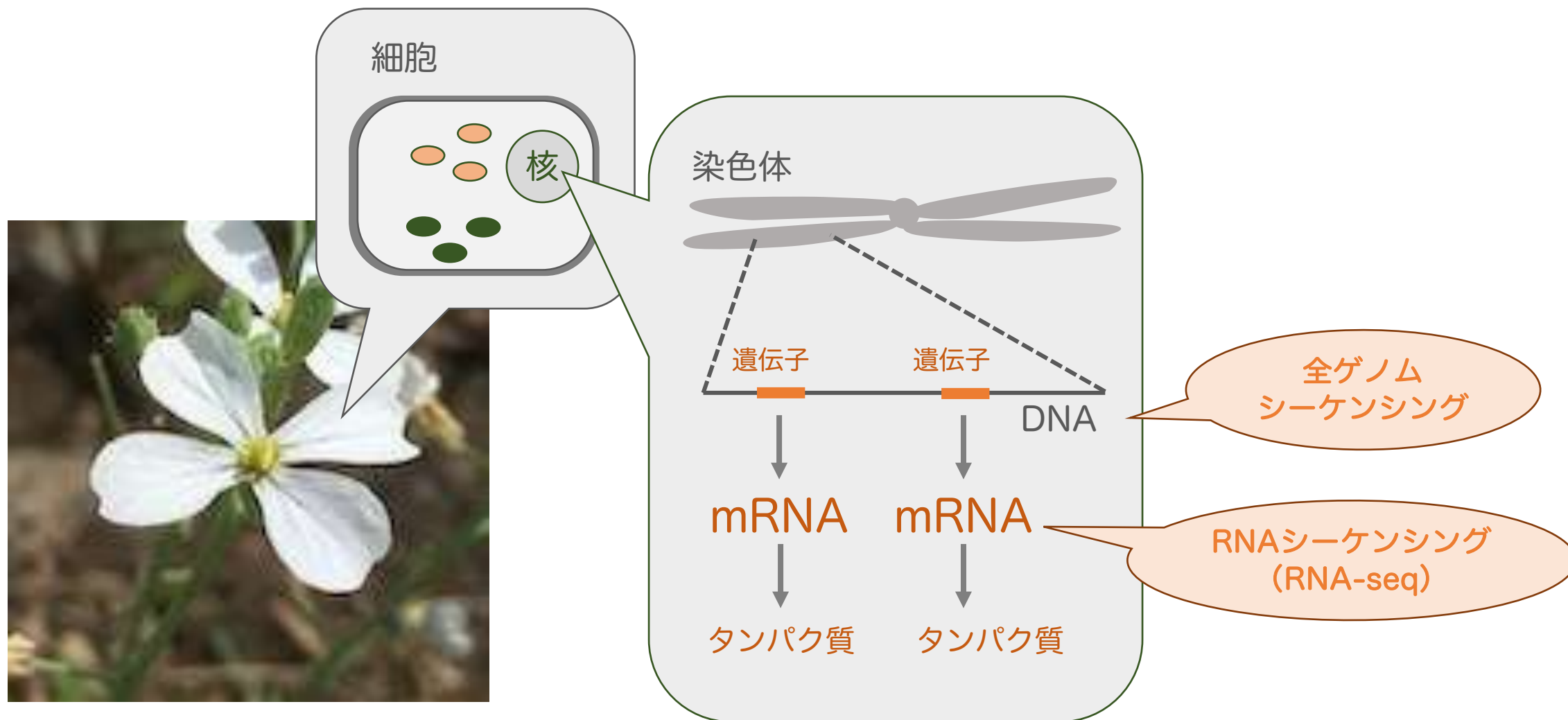


ILASセミナー 2022年6月

RNAシーケンシング (RNA-seq) 基礎

ゲノム情報はDNAに書き込まれている

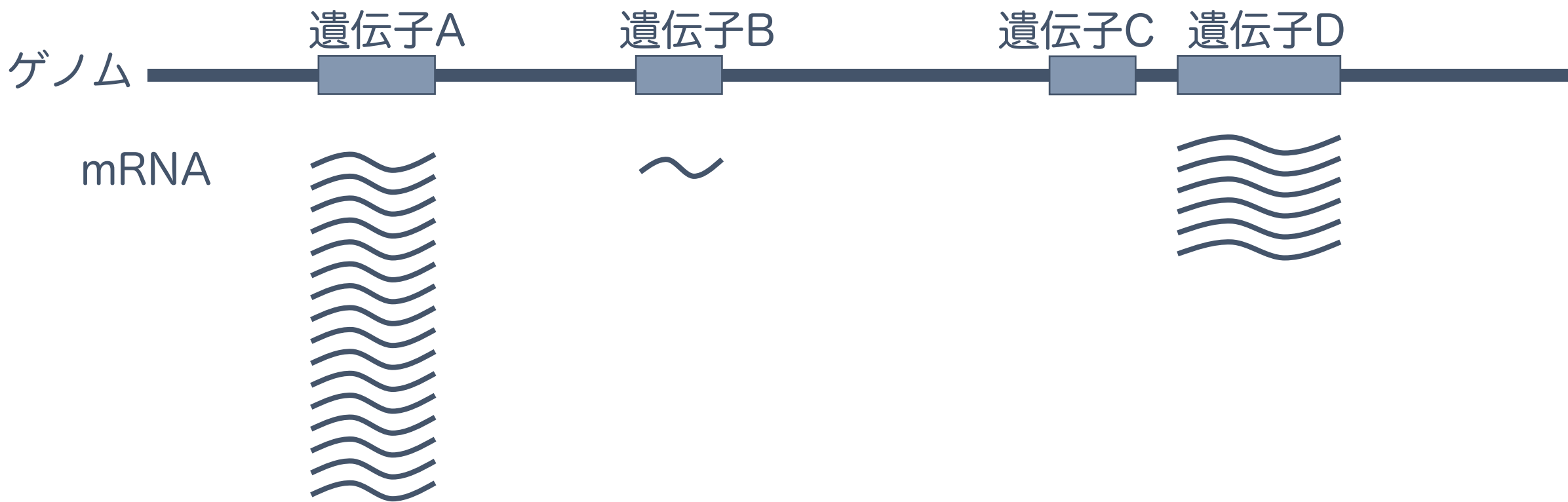


ゲノム上に遺伝子は多数ある



生物種	遺伝子数
ヒト	26,000 遺伝子
シロイヌナズナ	27,000 遺伝子
イネ	37,000 遺伝子
コムギ	120,000 遺伝子

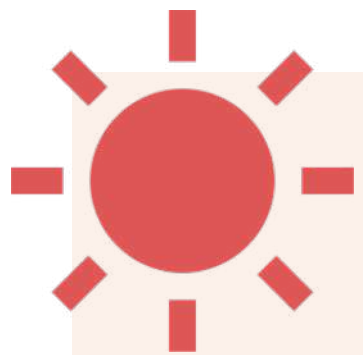
遺伝子発現量は遺伝子ごとに異なる



遺伝子発現量は個体ごとに異なる



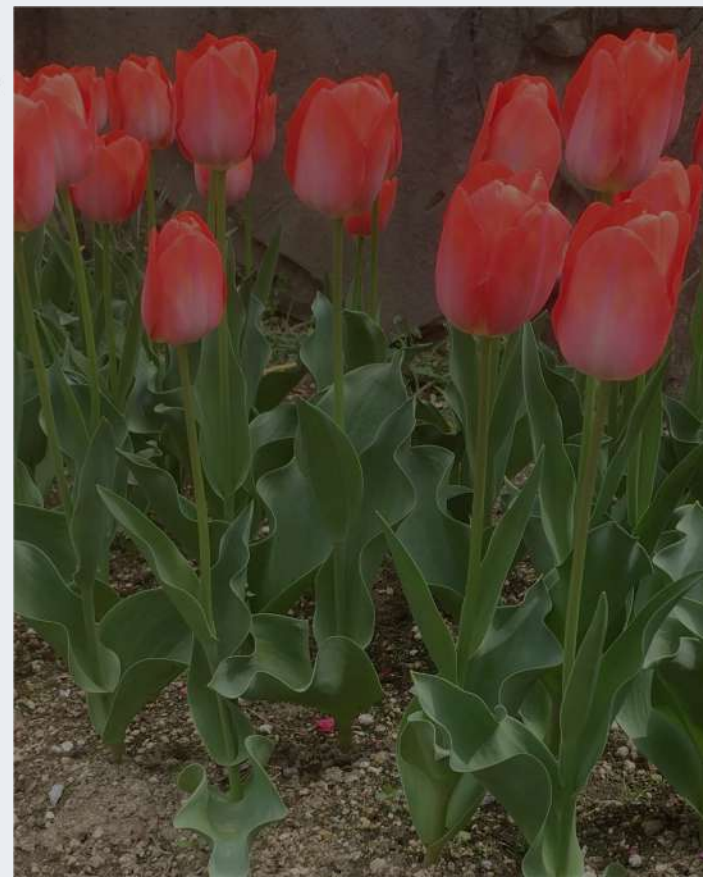
部位、時間でも発現量は異なる



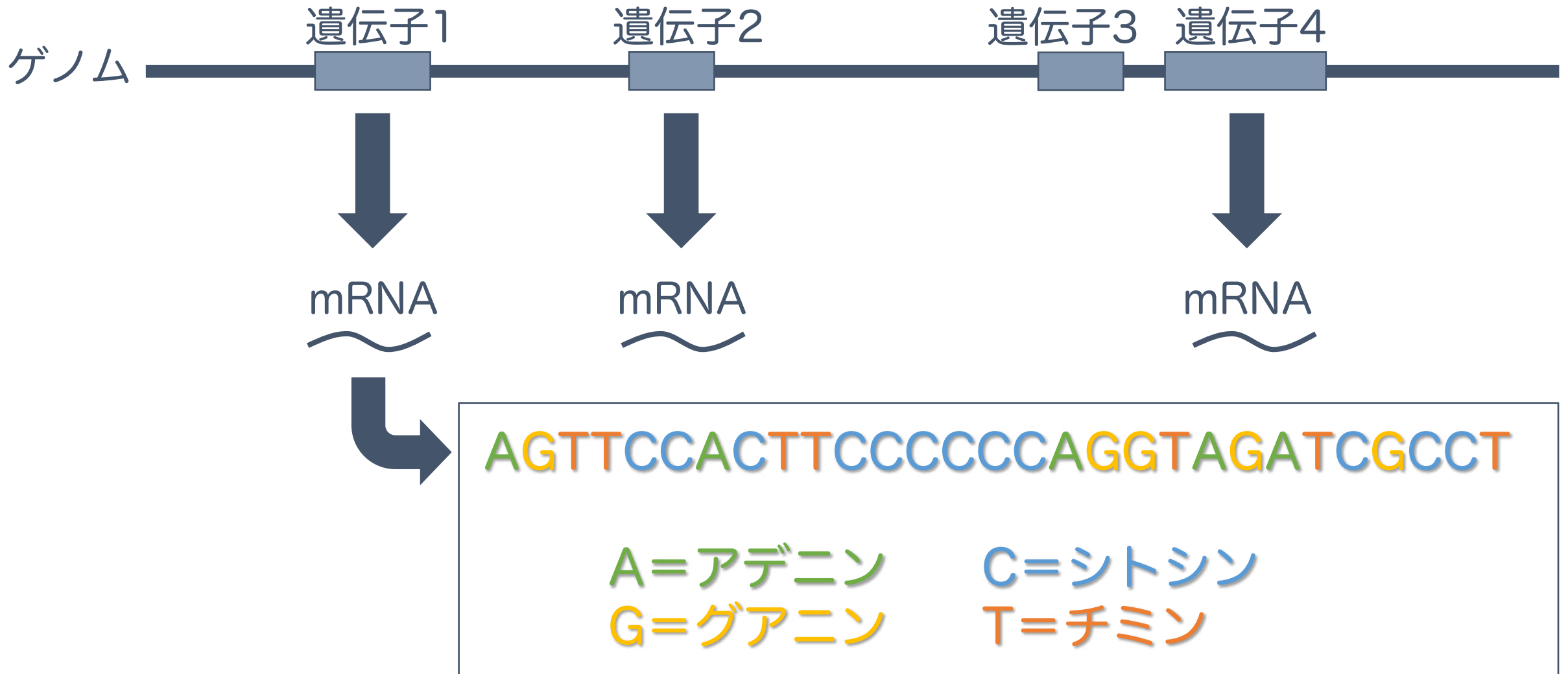
花



葉



RNA-seqは発現遺伝子の配列を得る方法



RNA-seqをすると何がうれしいのか？（１）

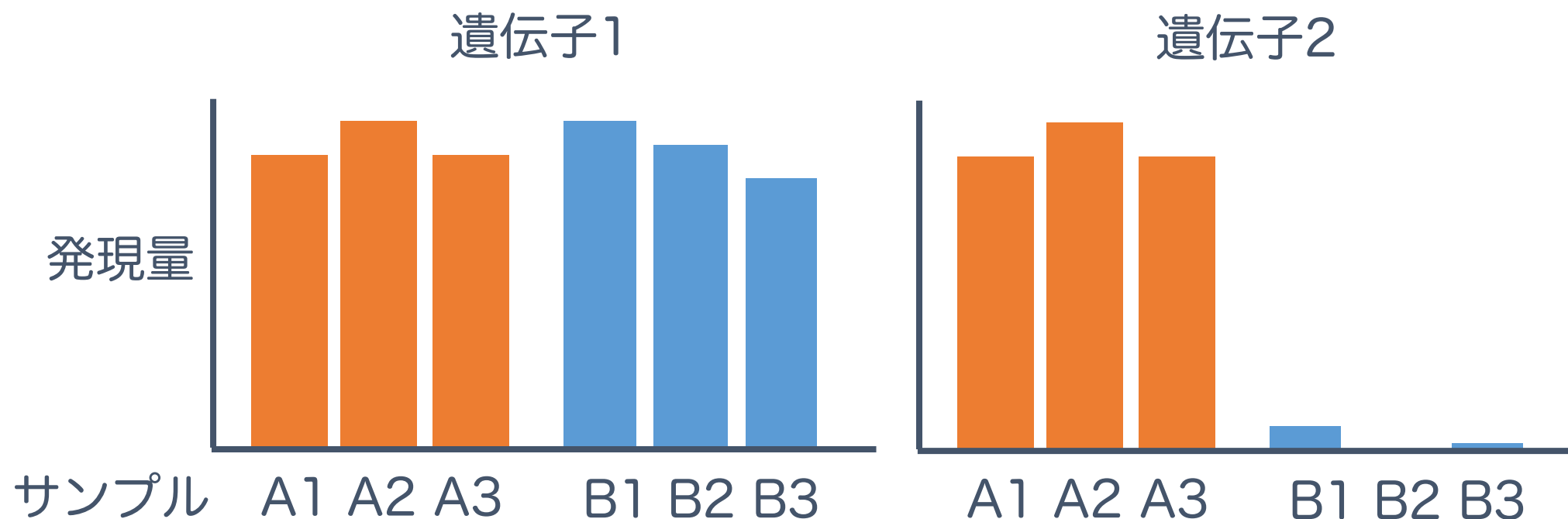
遺伝子配列情報から、
塩基の違いを見つけられる

遺伝子1の塩基配列

サンプルA	AGTTCCACTTCCC
サンプルB	AGTTC ^A ACTTCCC
サンプルC	AGTTCCACTT ^G CC

RNA-seqをすると何がうれしいのか？（2）

遺伝子発現量の違いを調べられる



RNA-seqで調べてみよう

ヤブガラシ

ねじれ部位 - ねじれなし部位

カイヅカイブキ

ネズ葉 - ヒノキ葉

水草

水上葉 - 水中葉

コムギ近縁野生種

芒あり穂 - 芒なし穂

どんな遺伝子に違いがあるのか？

形質の違いを決めている遺伝子は何か？

RNA-seqまで手順

サンプリング



RNA抽出 & RNA-seq前処理



RNA-seq

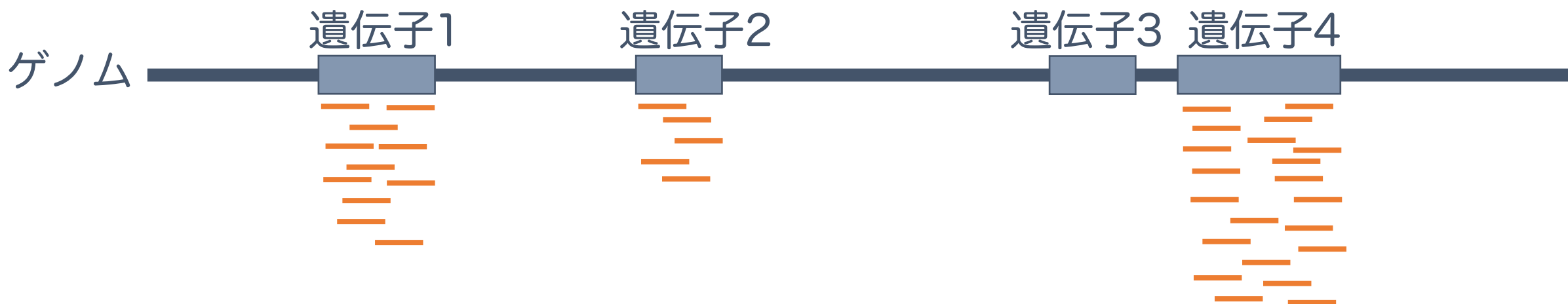


AGTTCCACTT
CTAGCTGGC
GGTACAAAT

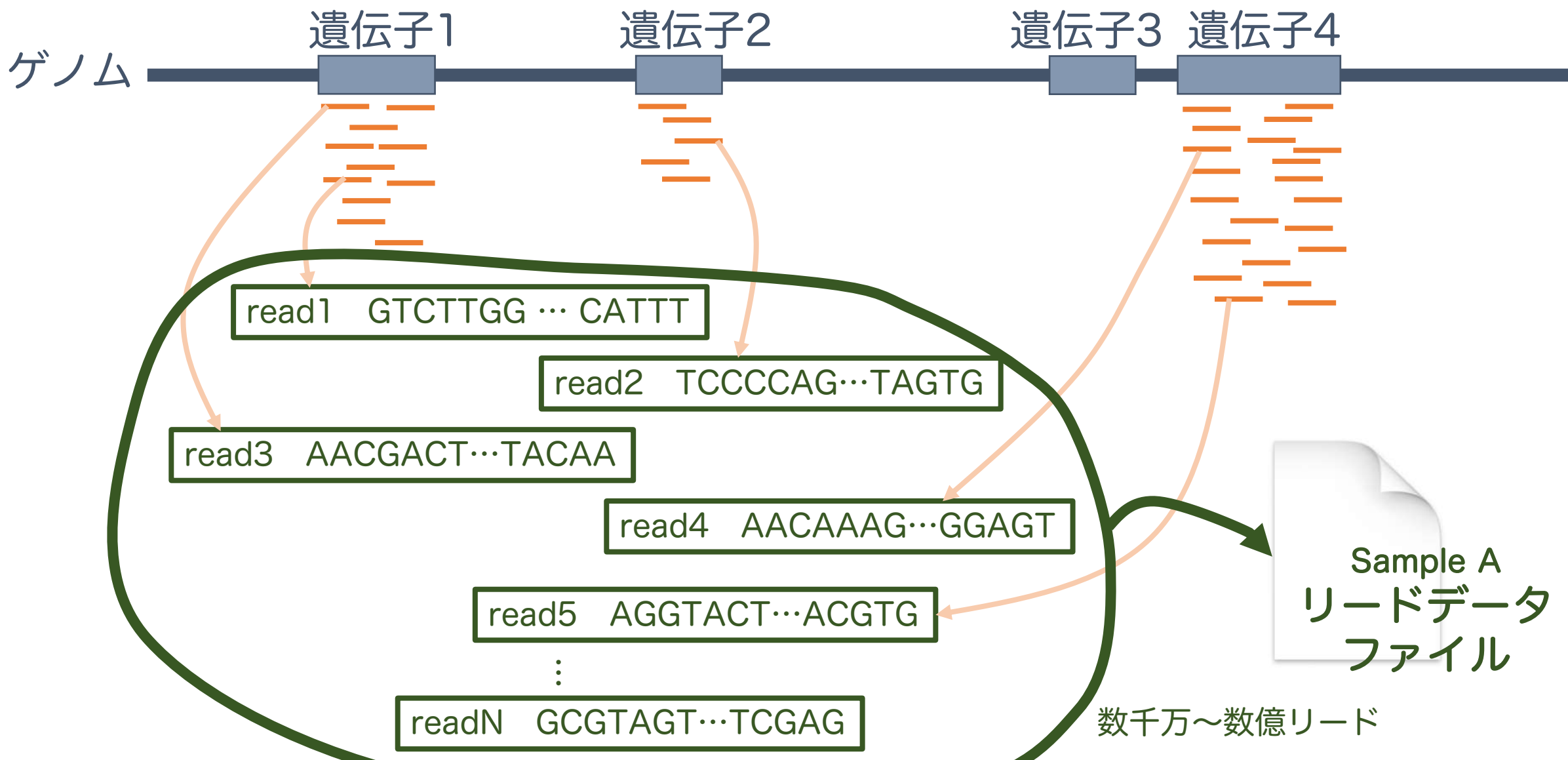
Illumina NextSeq 550

<https://www.illumina.com/>

発現遺伝子上の多数の短い塩基配列データが得られる




発現遺伝子上の多数の短い塩基配列データが得られる



RNA-seqデータ解析①

リードを参照配列にアライメント



リードデータ
ファイル

CTACCTAGCTCA
CTACCTAGCTCA
AGCTAGCTACCT
CTCCCAGCTAGC
CCGCCTCCCAGC
CCTCCCGCCTCC
TCCTCCTCCCGC
TCCTCCTCCCGCCTTCCAGCTAGCTACCTAGCTCA

TATTAATTCCGT
ACTCGTATTAA
AACTCGTATTAA
CCACACCAACTC
AGCCACACCAA
AGCCACACCAACTCGTATTGATTCCGT

参照配列（リファレンス）

参照配列が利用できない場合は？

リードデータ
ファイル

CTACCTAGCTCA
CTACCTAGCTCA
AGCTAGCTACCT
CTCCCAGCTAGC
CCGCCTCCCAGC
CCTCCCGCCTCC
TCCTCCTCCCGC
TCCTCCTCCCGCCTTCCAGCTAGCTACCTAGCTCA

TATTAAATTCCGT
ACTCGTATTAAAT
AACTCGTATTAA
CCACACCAACTC
AGCCACACCAA
AGCCACACCAACTCGTATTGATTCCGT

RNA-seqデータ解析①

リードから参照配列をつくる

リードデータ
ファイル

CTACCTAGCTCA
CTACCTAGCTCA
AGCTAGCTACCT
CTCCCAGCTAGC
CCGCCTCCCAGC
CCTCCCGCCTCC
TCCTCCTCCCGC



TCCTCCTCCCGCCTCCCAGCTAGCTACCTAGCTCA

TATTAATTCCGT
ACTCGTATTAAT
AACTCGTATTAA
CCACACCAACTC
AGCCACACCAA



AGCCACACCAACTCGTATTAATTCCGT

リードから作った参照配列

RNA-seqデータ解析のファーストステップ

① 参照配列をつくる

CTACCTAGCTCA
CTACCTAGCTCA
AGCTAGCTACCT
CTCCCAGCTAGC
CCGCCTCCCAGC
CCTCCCGCCTCC
TCCTCCTCCCGC



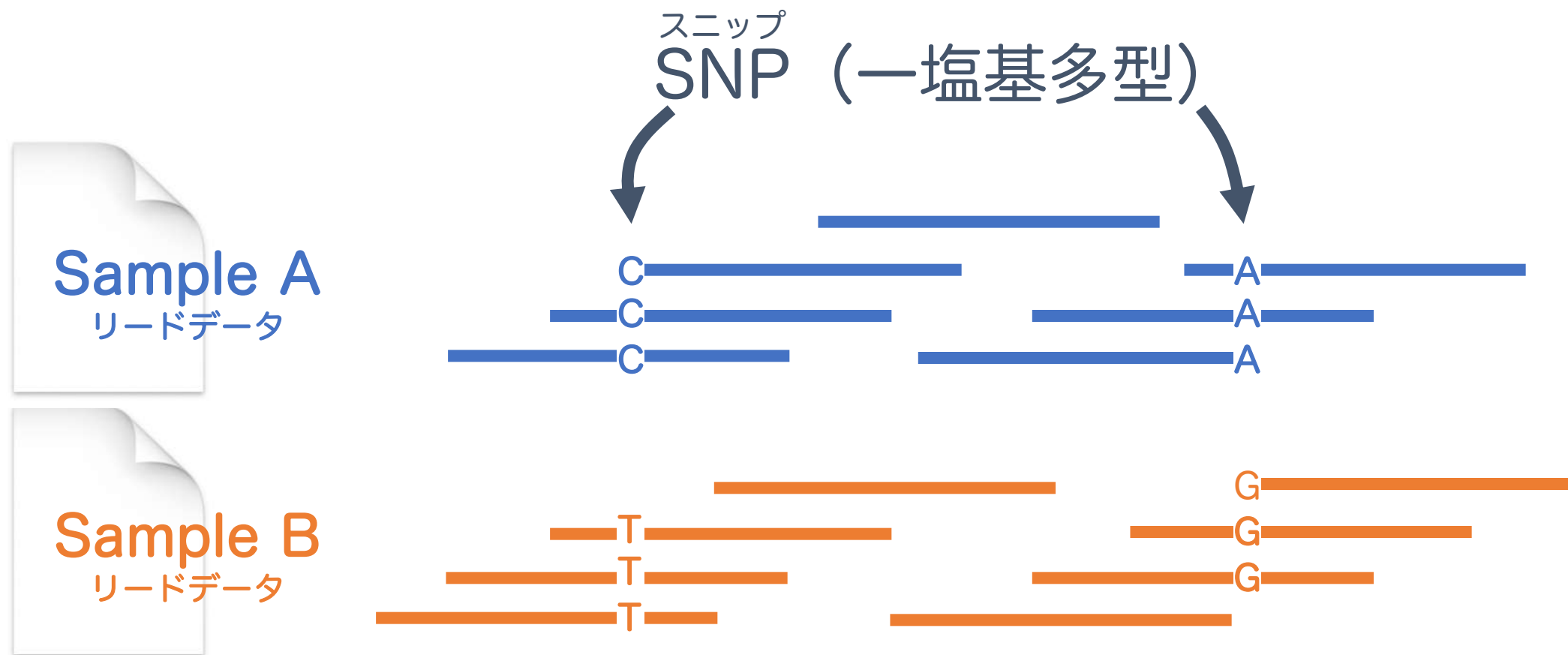
TCCTCCTCCCGCCTCCCAGCTAGCTACCTAGCTCA

② アライメント

CTACCTAGCTCA
CTACCTAGCTCA
AGCTAGCTACCT
CTCCCAGCTAGC
CCGCCTCCCAGC
CCTCCCGCCTCC
TCCTCCTCCCGC

TCCTCCTCCCGCCTCCCAGCTAGCTACCTAGCTCA

解析 1) 塩基配列を比べる→SNPを見つける



遺伝子1

TCCTCCTCCCGCCTCCCAGCTAGCTACCTAGCTCA

解析 2) リード数を数える→発現量を得る

遺伝子



発現量が多い



発現量が少ない

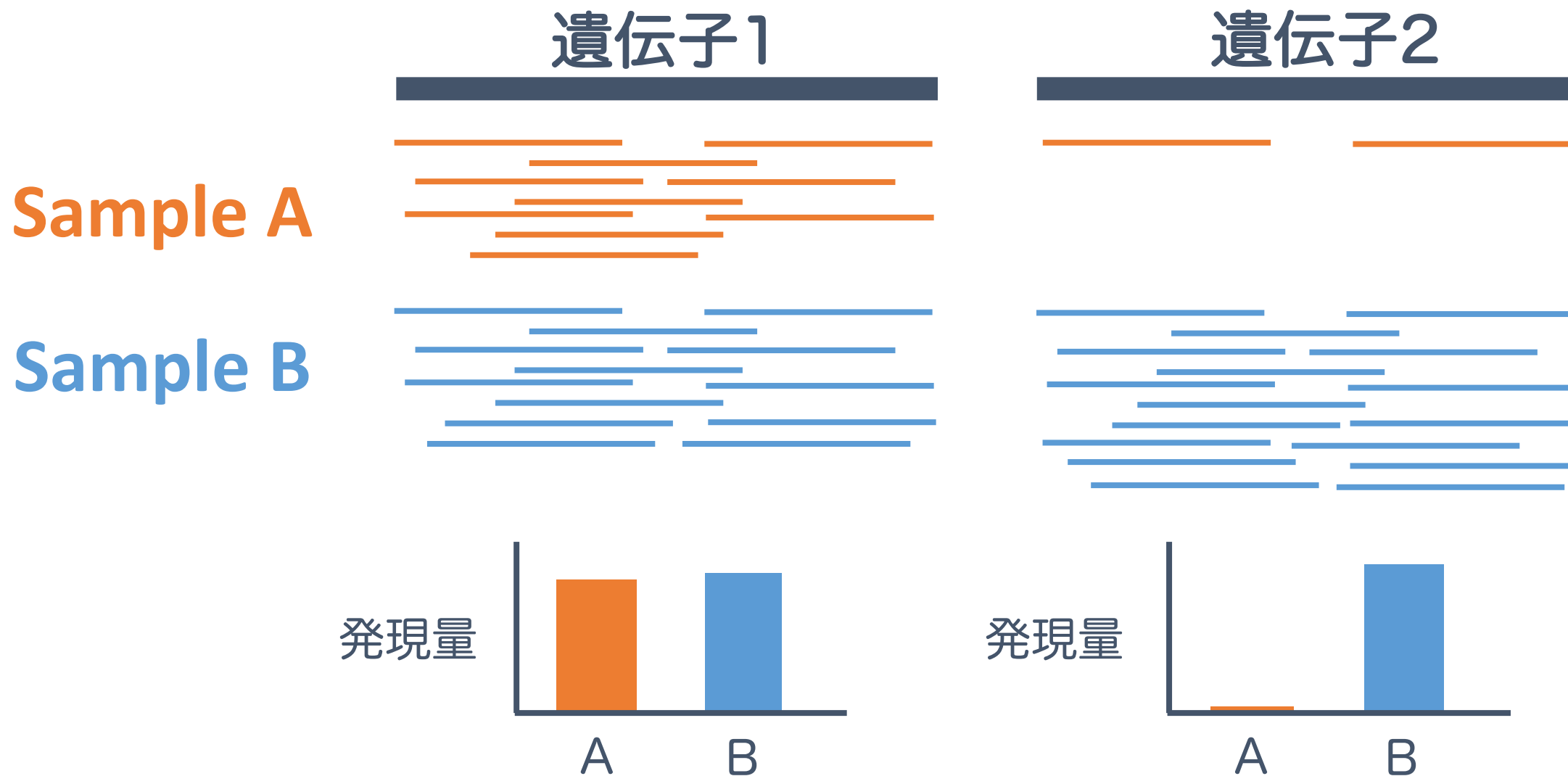
mRNA



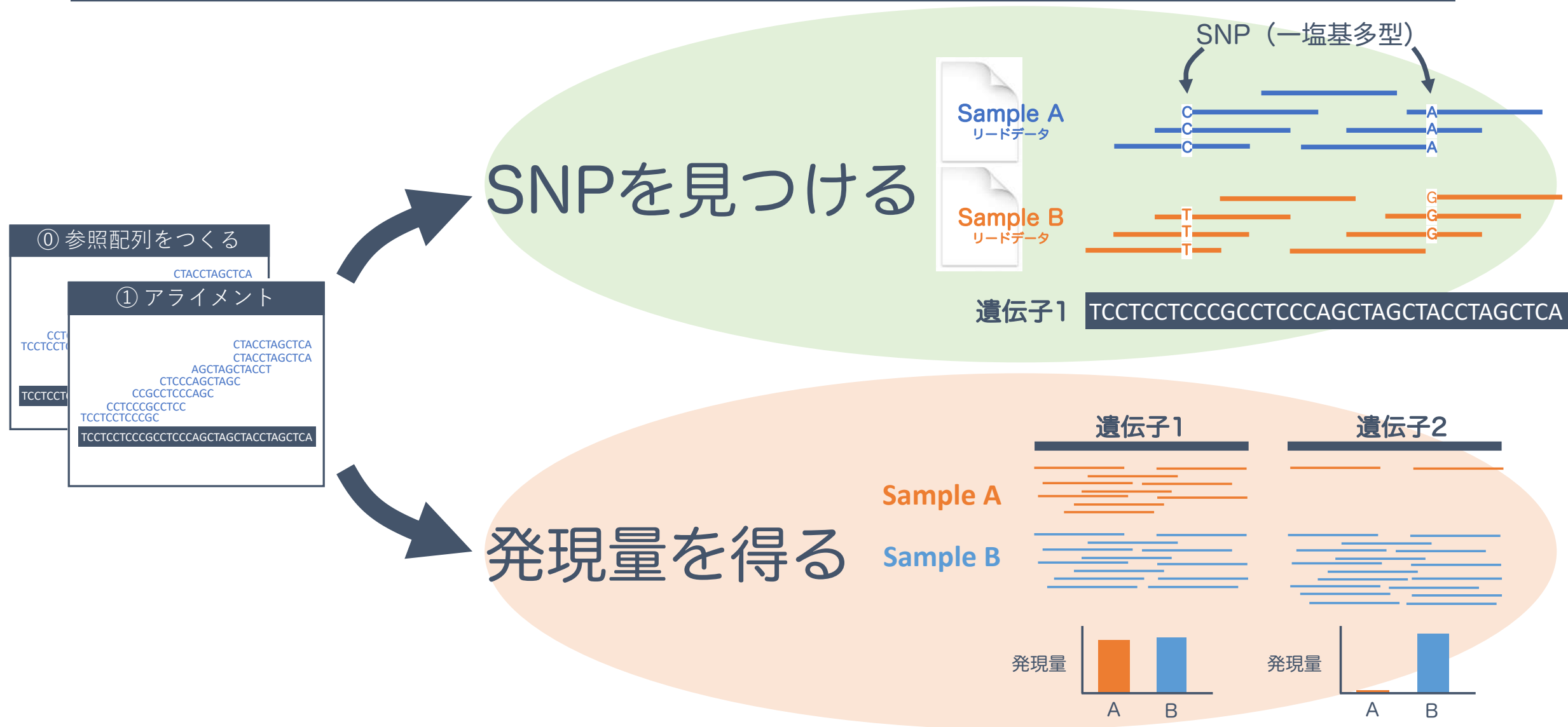
リード



解析2) リード数を数える→発現量を得る



RNA-seq解析で出来ること



差がある遺伝子を発見した後、出来ることは？



実習 1) データベースから類似遺伝子を探してみよう

相同性検索 (BLAST検索)

```
>gene1
TTCATCAGAGcagtgaaacacatgaacatactccctcgttccgaattactgtgcaggtatcaagatgtattttaattc
tagatacatccattttacgacgagtaatttgaacgaaggagtagacaagcTTACAAGTACAGATCAAGAACAGA
ACTGAACAAAGCTCAGCTCAGTTTTGGACGTGCCTTGCGTACAAAATTCATTTTCATTCTCATTG
GAGCTGCTAACAAAGTCCGGCCCGCGCTTATTTTGATATGAGCGATTGGAGGAATTACACATA
GCAATAATAGACATTTAacatagtagtagtagtagtaataaggcCTTGGGAGTTTTACGGGTGATC
ATGGAGGCGGTGGAGAACTGGGGACGAAGAAGGGATTACATTACTACATGCCGGTGGTCCGG
CCGGCTCGGAGTTCGTCGTCCATCTCTTACGCCGACGGACGGAACCCCGTGACCCGATC
GATCGTAGCCGGTCGCTGCAATTCAGCGCTCACTGACATGACATttagcttcttcttcaagtGGGT
GCGTGCGTGCGCCGGAGAGCCGGGGCGGCCGGGGTGTGTCTCTCCATGCTCACGAT
GTGTACTTACAGCGTTCAAACTCGCGCATCACGGCGCGGCCAGGCGCCACGCCGAGGTGG
CGATCAGCGGGCGCGTGTGCCATCCAGCGTCAGGCAGCCTTCTTCTCTCGACCTTGATCC
CGTCGCCCGCGCGGAAGAGCGCCAGCAGCGTGTCTGCCTGCTTGAGGCATTGGACCCAGG
TGGACGGTCTCGAACC CGCGTTGCCAGGCGGTTCCGCCACTGCCCGAGGGTCTCGTGCGC
CTCTGTGCGCTCCGCTCCCTCGCAGGCCACCACGTTGCAGATCTGCCGGCCGAGGTACACCTC
GGACATGACCTGCTGCGGCCGCGCGCAGGAGCGGCGAGCCGCCCGATGAGACTTCG
GATGGGCCGCCCGGAGCTGCCGCCCTCCAGAGAATCGAATCGGTGGAGTAGTAGTGCA
GGGACTCGGTGAAGCGGTCCAGGAATGTCCGGAGTTGTGGTTGGCCTCTGCTCCACCACG
GTGACGATCCTCGGCCGACGGCGCGCACGGTGCCAGGACCTTCTCCAGG
```

遺伝子配列

The screenshot shows the NCBI BLAST search interface. The 'blastx' tab is selected. The 'Enter Query Sequence' section has a text box containing the gene sequence, which is highlighted with an orange box. Below this, the 'Choose Search Set' section shows the 'Database' set to 'Non-redundant protein sequences (nr)'. The 'BLAST' button is highlighted with an orange box. The 'Algorithm parameters' section is partially visible at the bottom.

NCBIのタンパク質
データベース

実習2) どういう遺伝子かより詳しくを調べてみよう

関連論文を検索してみよう

Google Scholar

キーワード（遺伝子名や生物名など）を入力



☐ すべての言語 ☒ 英語 と 日本語のページを検索

巨人の肩の上に立つ

Rht-1 は半矮性遺伝子、草丈の低いコムギを作る

*Rht-1*をもたない
ヨーロッパのコムギ



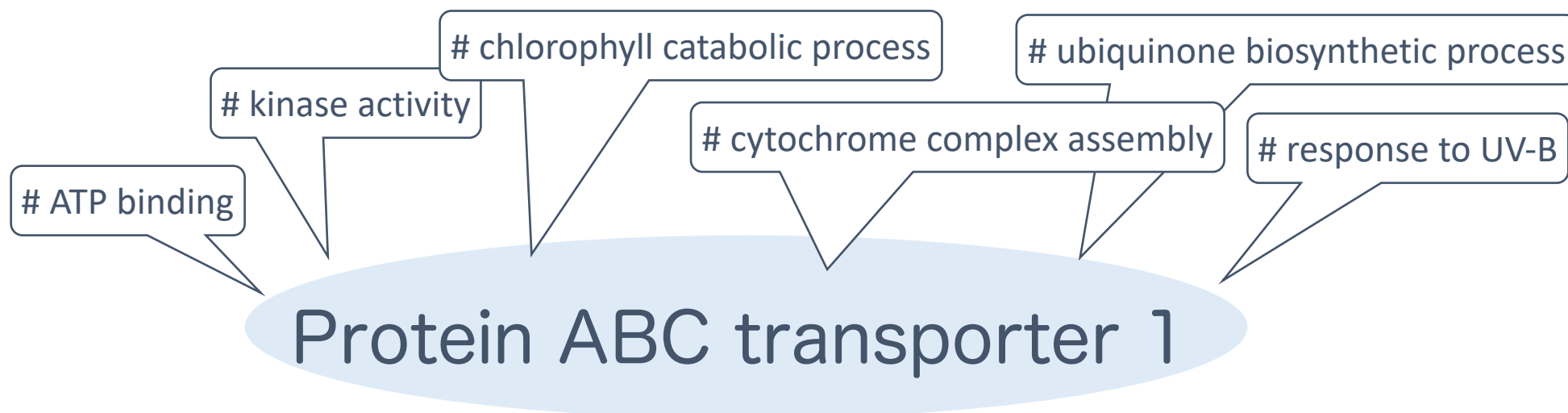
*Rht-1*をもつ
日本のコムギ

実習3) 遺伝子機能の情報も集めてみよう

データベース上の~~遺伝子~~

タンパク質には **GOterm** が付いている

Gene Ontology term
(遺伝子の機能に関するタグ)



実習 3) 遺伝子機能の情報も集めてみよう




The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.

UniProtKB

UniProt
Knowledgebase

Swiss-Prot
(567,483)

 Manually
annotated and
reviewed.

UniRef

Sequence clusters



UniParc

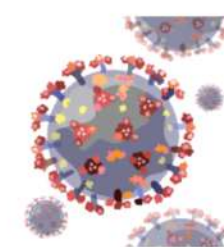
Sequence archive



Proteomes

Proteome sets





New UniProt portal for the latest SARS-CoV-2 coronavirus protein entries and receptors, updated independent of the general UniProt release cycle.

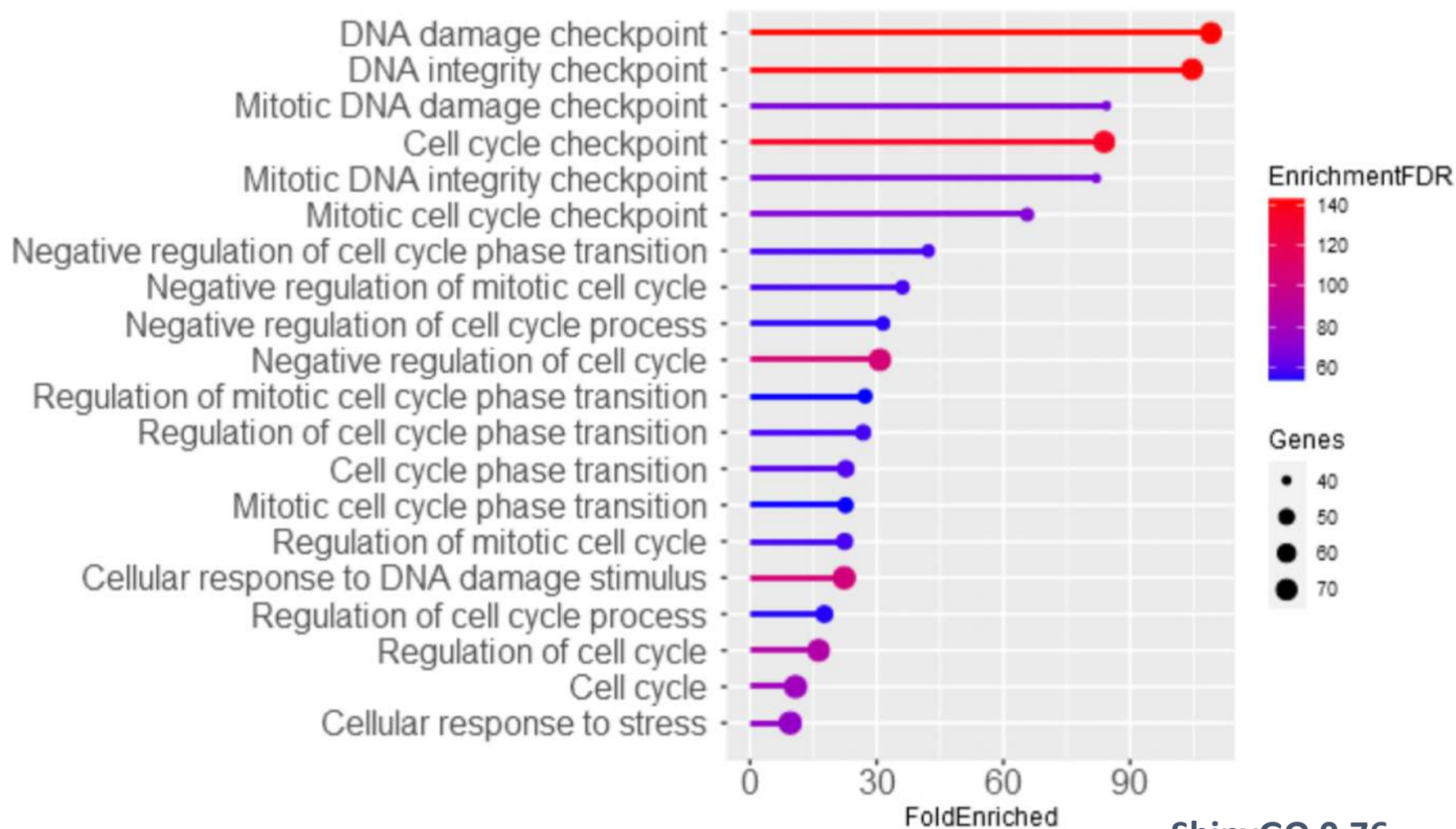
**View SARS-CoV-2
Proteins and Receptors**

News



Forthcoming changes

GOtermを調べると、頻出termがわかる



ShinyGO 0.76

<http://bioinformatics.sdstate.edu/go/>

実習 4) 100 遺伝子のBLAST検索 & GOterm検索

100遺伝子に対して、
実習 1 と 3 の作業をおこなってください。

実習 4) 100 遺伝子のBLAST検索 & GOterm検索

100遺伝子に対して、
実習 1 と 3 の作業をおこなってください。

単純な繰り返し作業は
コンピュータにやらせよう！