

第6回(平成26年度第3回)CRCフォーラム(平成26年12月15日(月)開催)
「TDUで展開する先端微生物研究」

有用物質高生産菌株ゲノムのNGS解析

根本 航 助教

理工学部理工学科生命理工学系

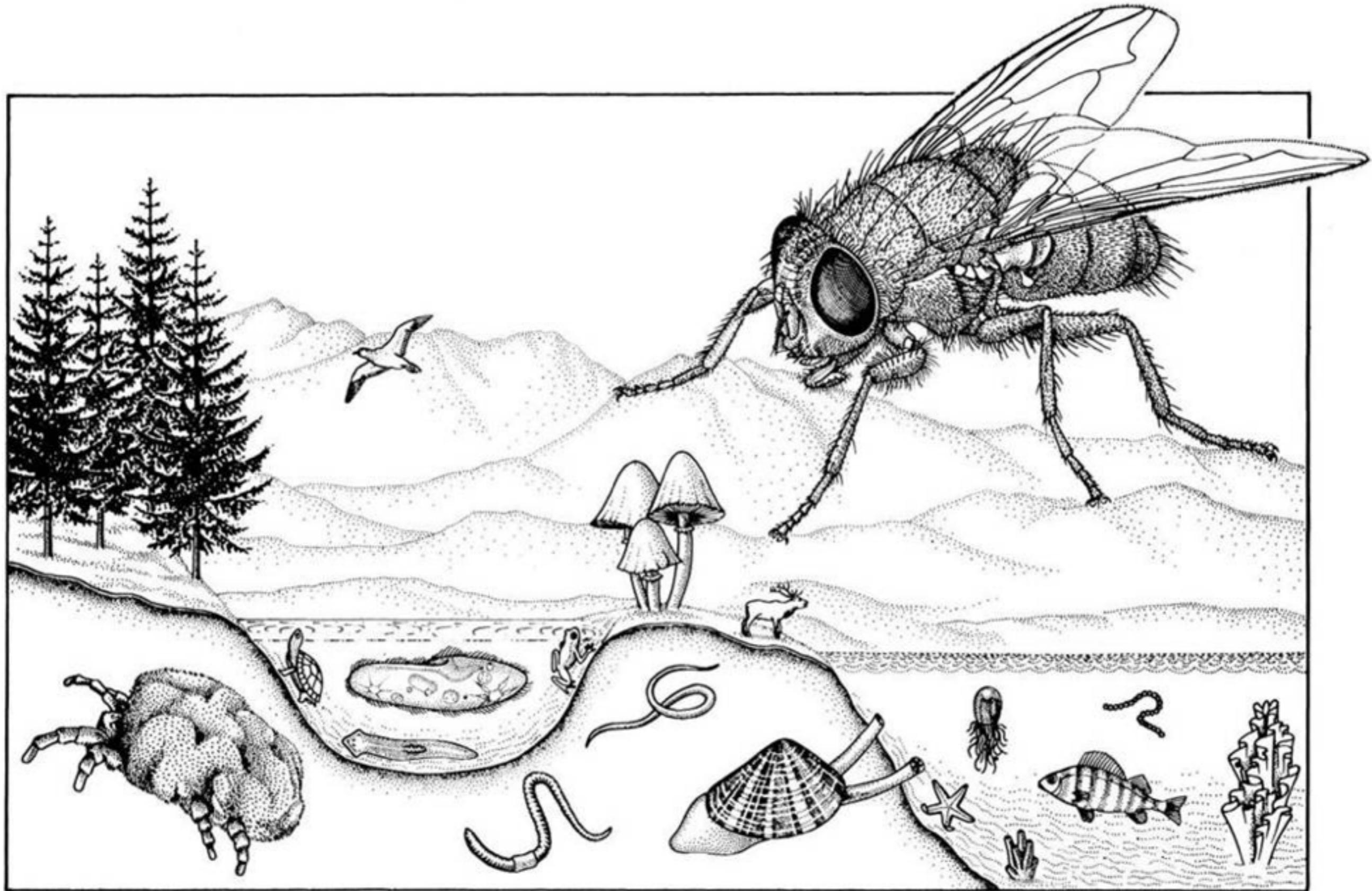
TDU
東京電機大学

有用物質高生産菌株ゲノムのNGS解析

東京電機大学 理工学部 生命理工学系

根本 航

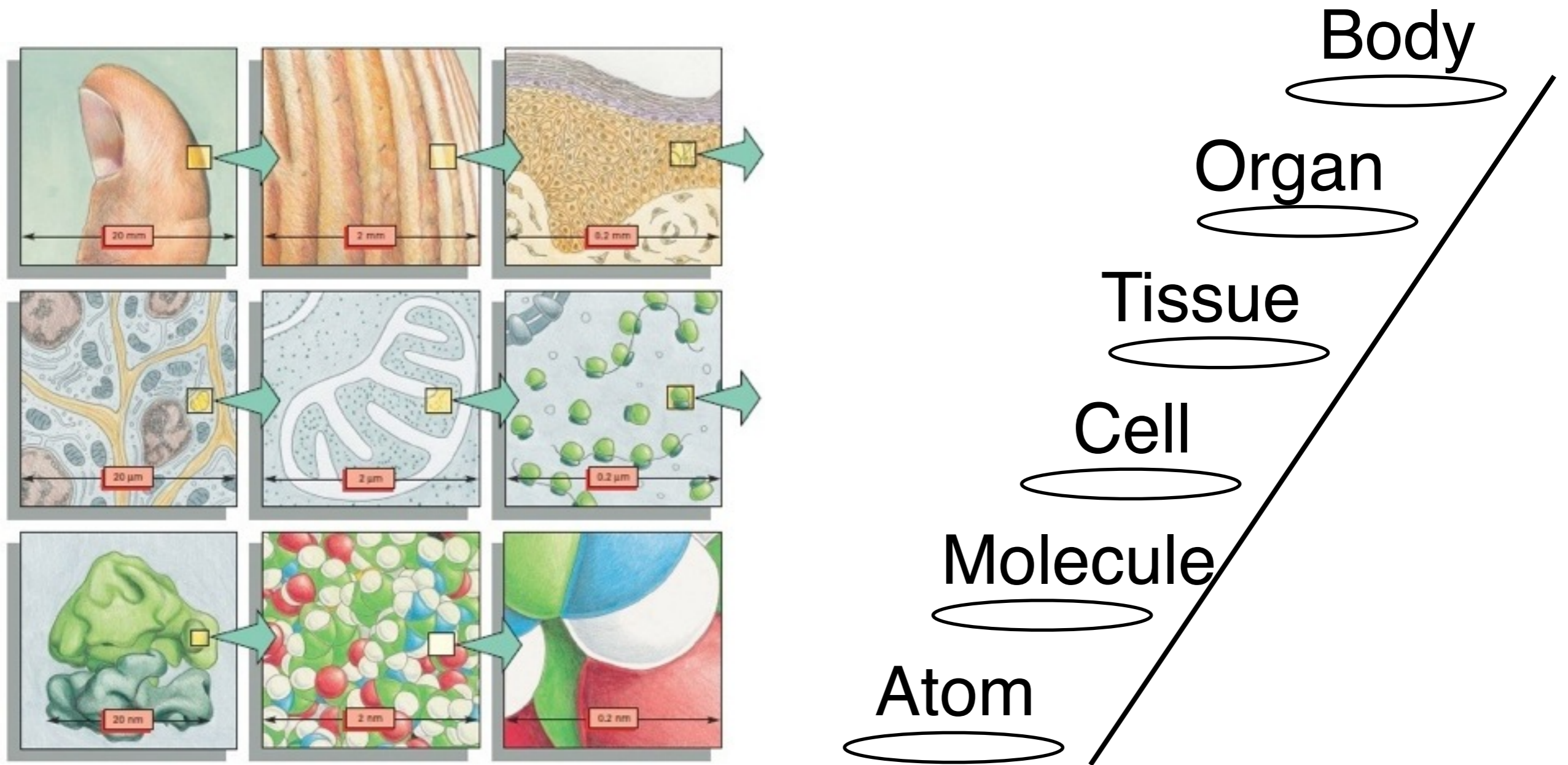
The Species-Scape Picture



本日の内容

- Next Generation Sequencer (NGS)の位置づけ
- 有用物質高生産菌株のNGSデータ解析
 - ゲノム配列比較
 - トランスクリプトーム解析

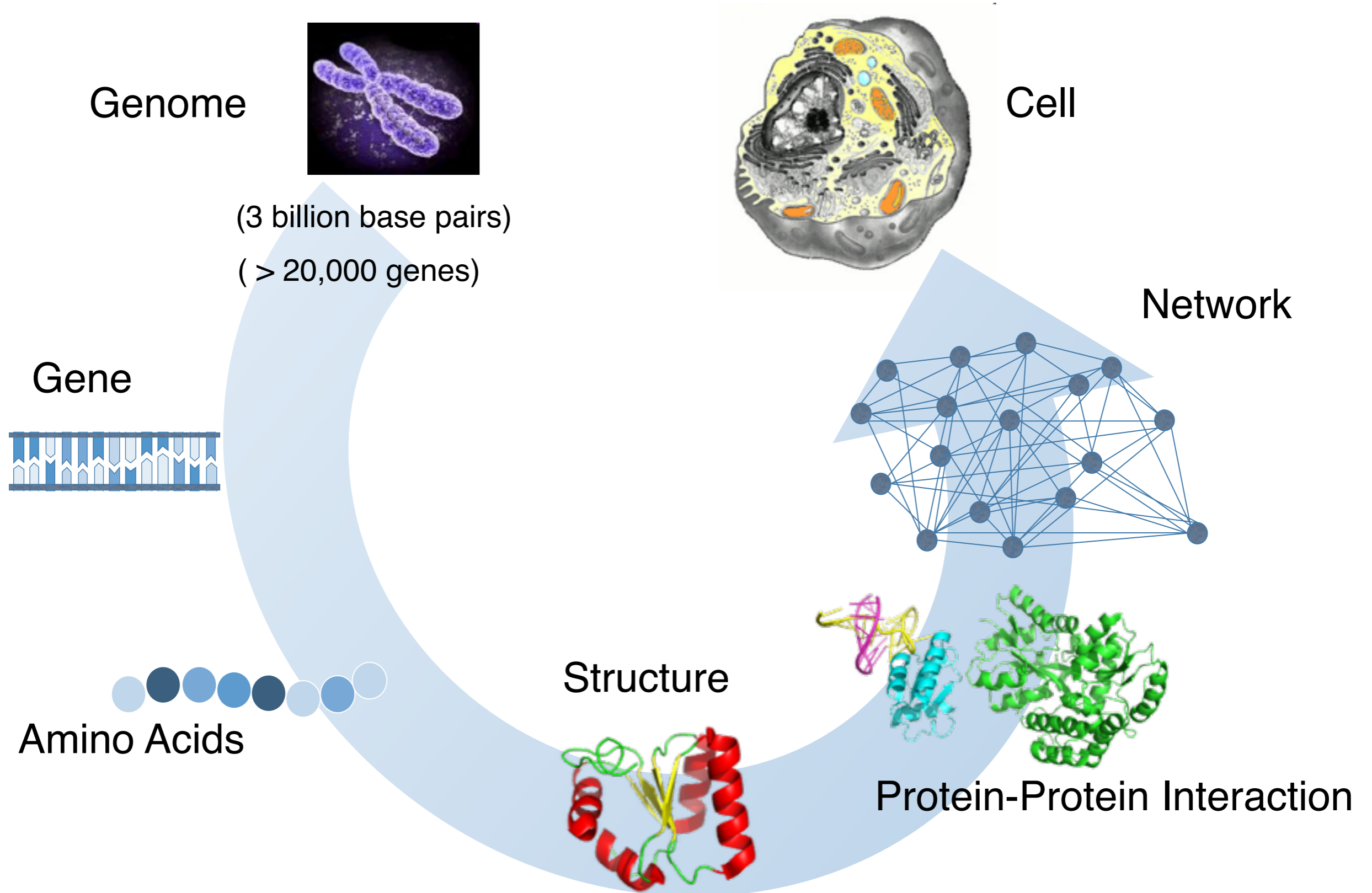
生命の階層構造



細胞の分子生物学第3版より

<http://www.garlandscience.com/textbooks/081533480X/pdf/ch01.pdf>

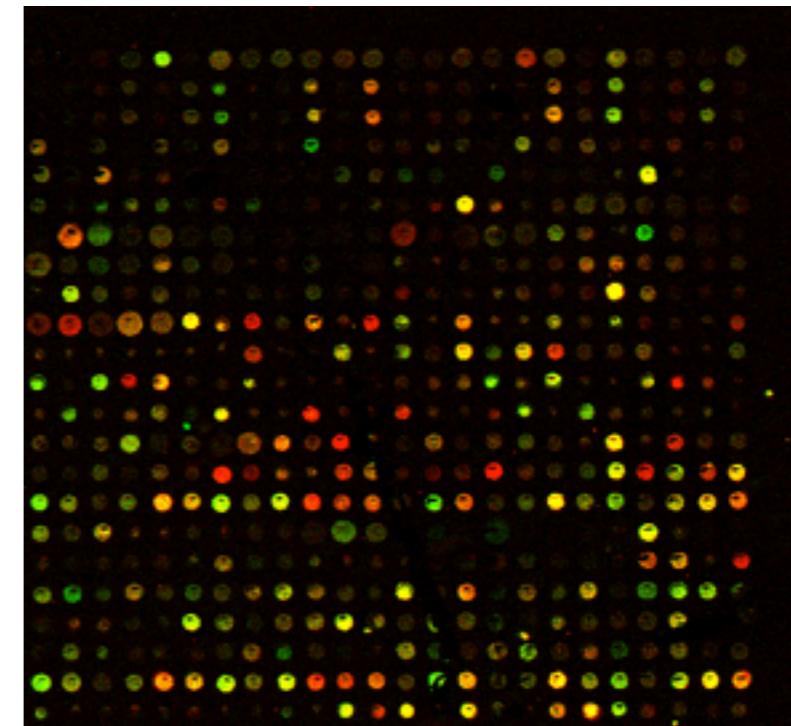
生命の階層構造



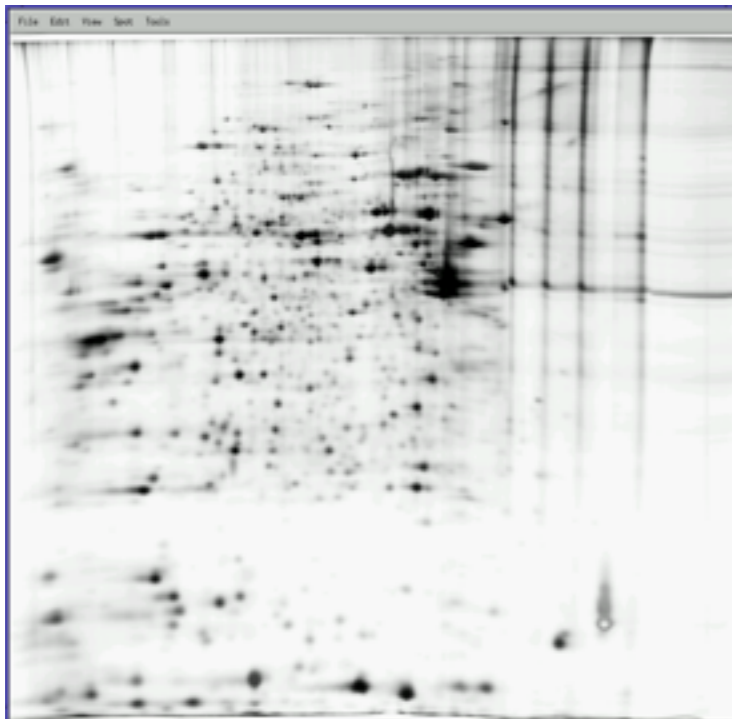
Ome研究の発達



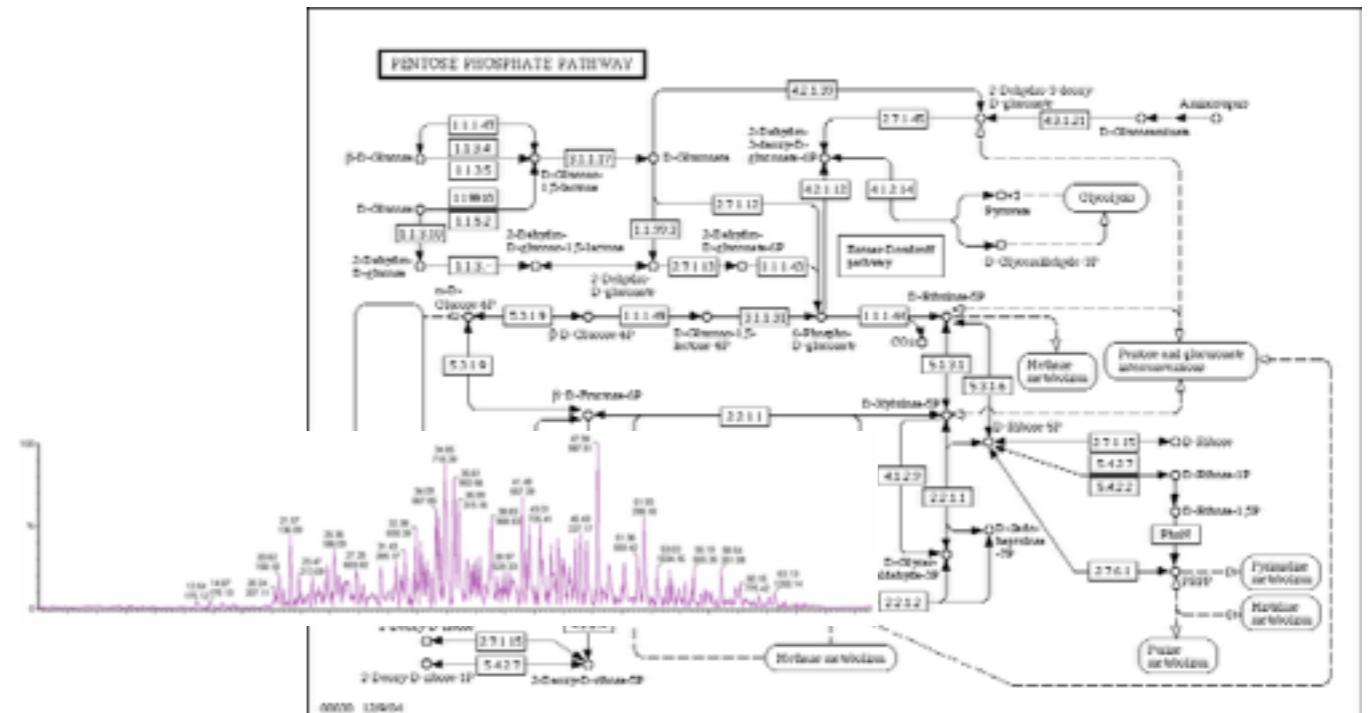
Genome: 設計情報



Transcriptome: DNA発現



Proteome: タンパク質翻訳・発現



Metabolome: 代謝物

ヒトゲノム配列解読終了宣言

2000年6月26日

『人類が作った最も重要な**地図**』

(X, Y)
モノが番地付け

↓

遺伝情報を文字情報/デジタルデータとして厳密に表現できるようになった → 0と1の組み合わせで記述可能

↓

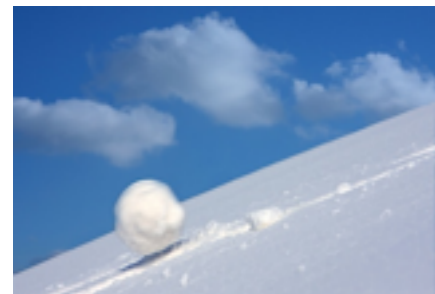
コンピュータ利用の促進

デジタルデータ・文字列処理が大得意

高速に、経済的に、安全に、

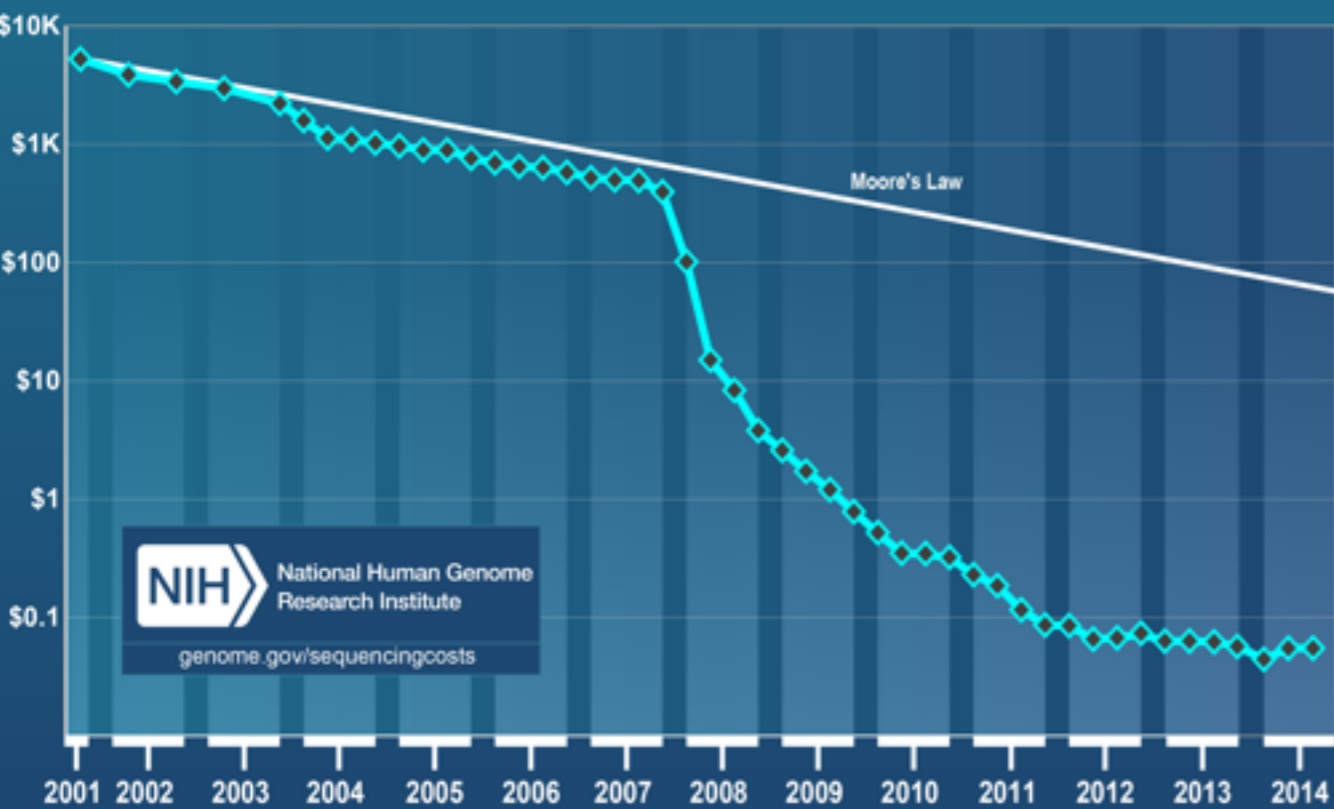
膨大な情報から、誰よりも早く、安全に、安く、飯の種を探す、

ゲノム情報の意味解釈(地図解読)競争が激化

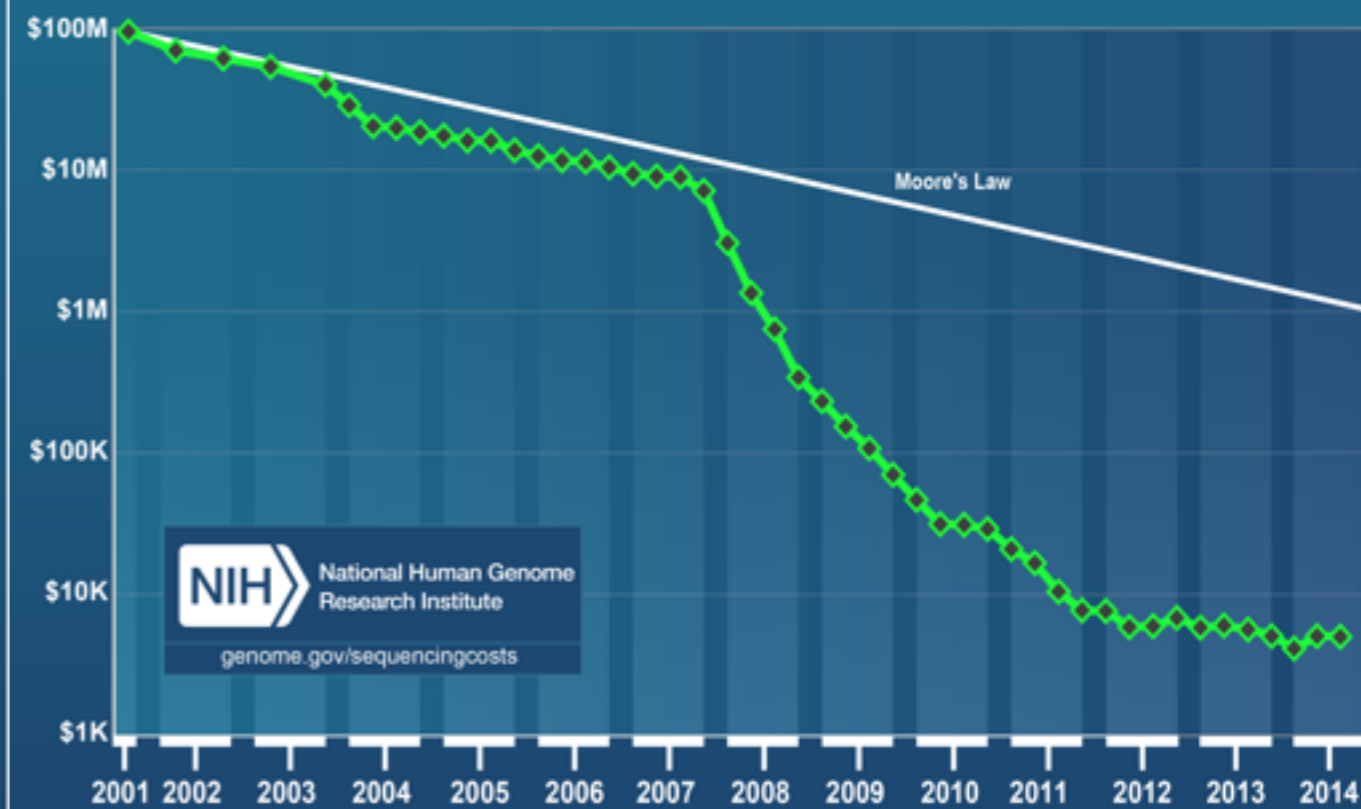


Sequencing Costs

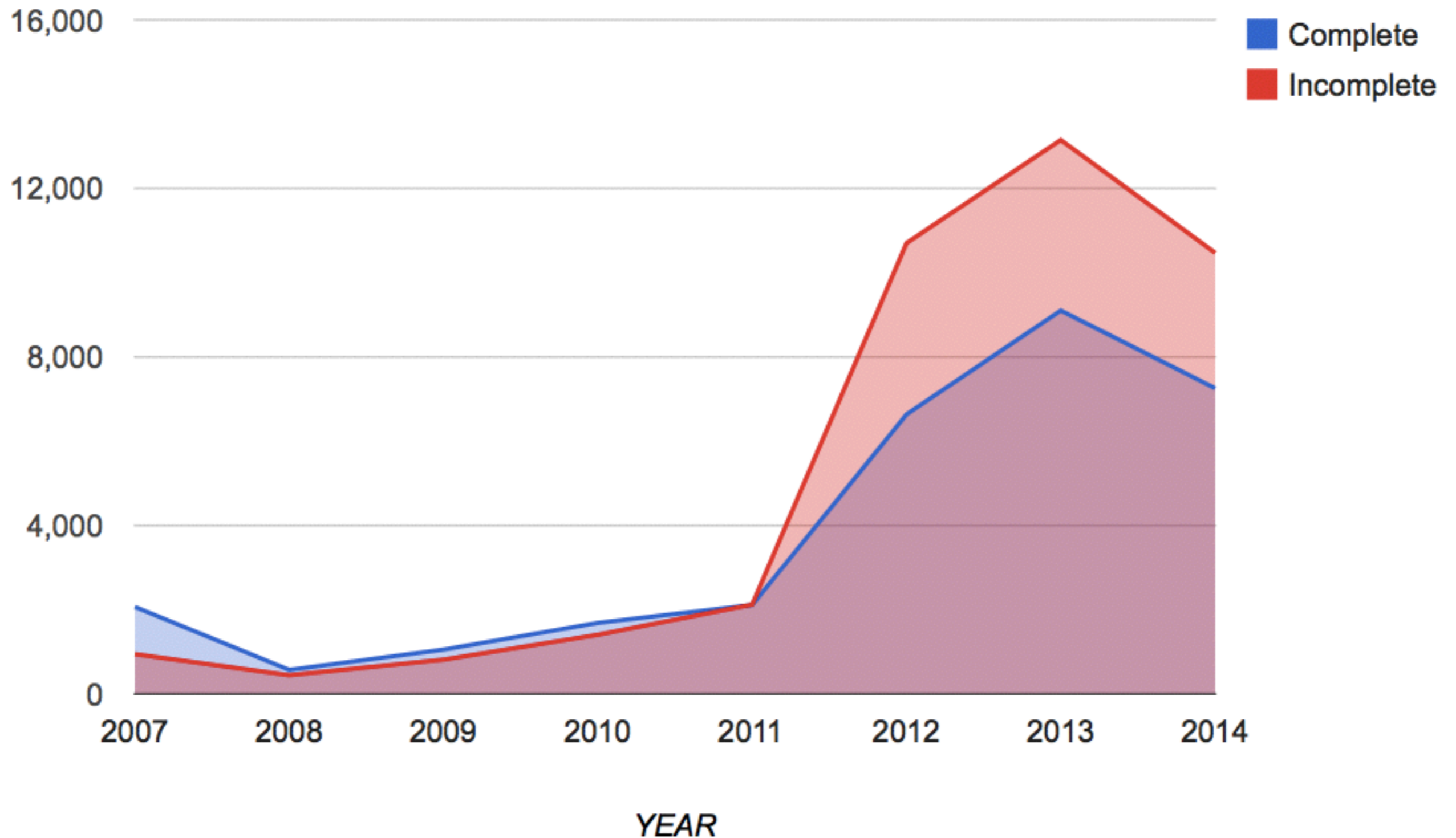
Cost per Raw Megabase of DNA Sequence



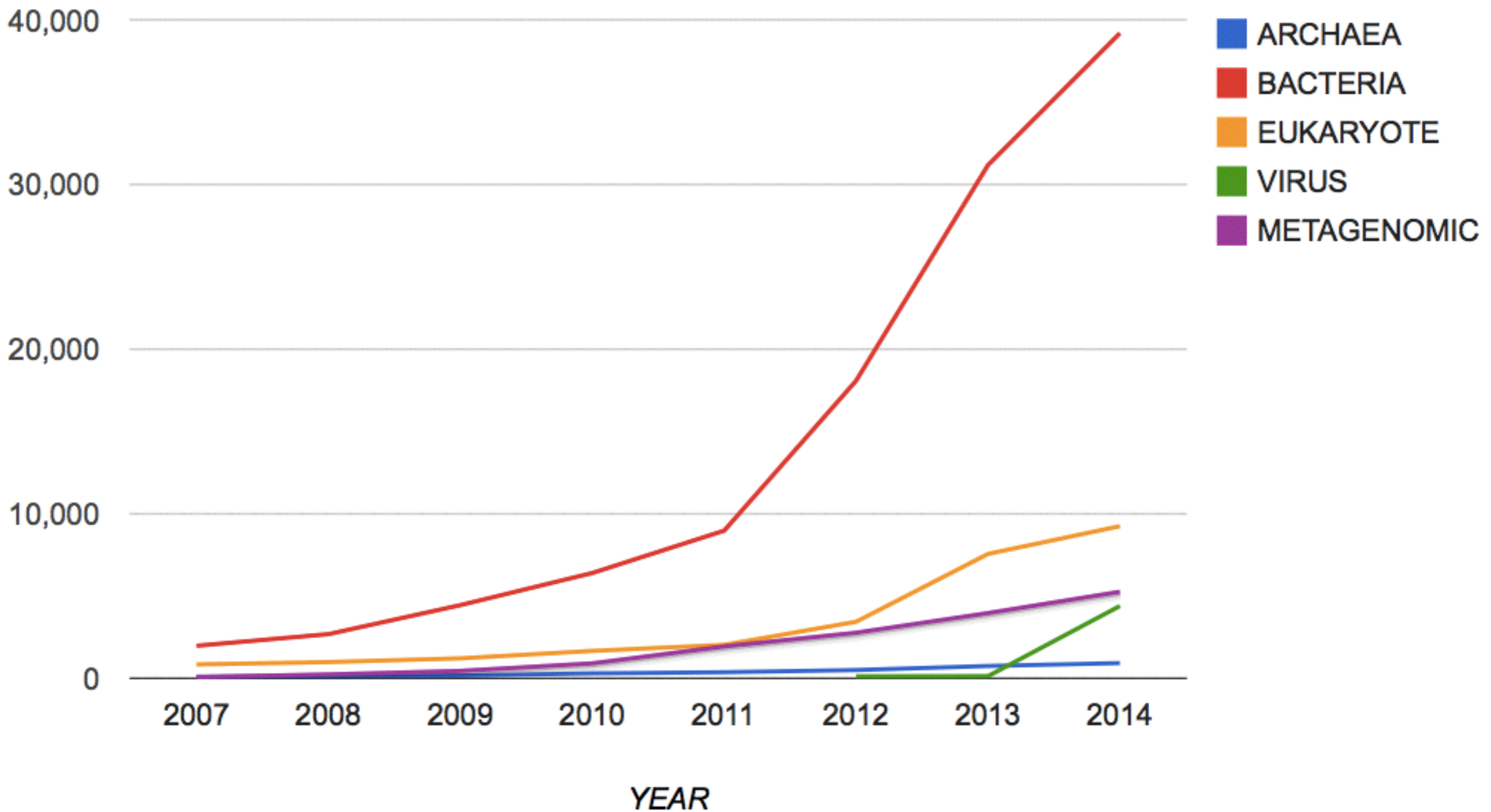
Cost per Genome



Genome Totals in GOLD



Project Total in GOLD

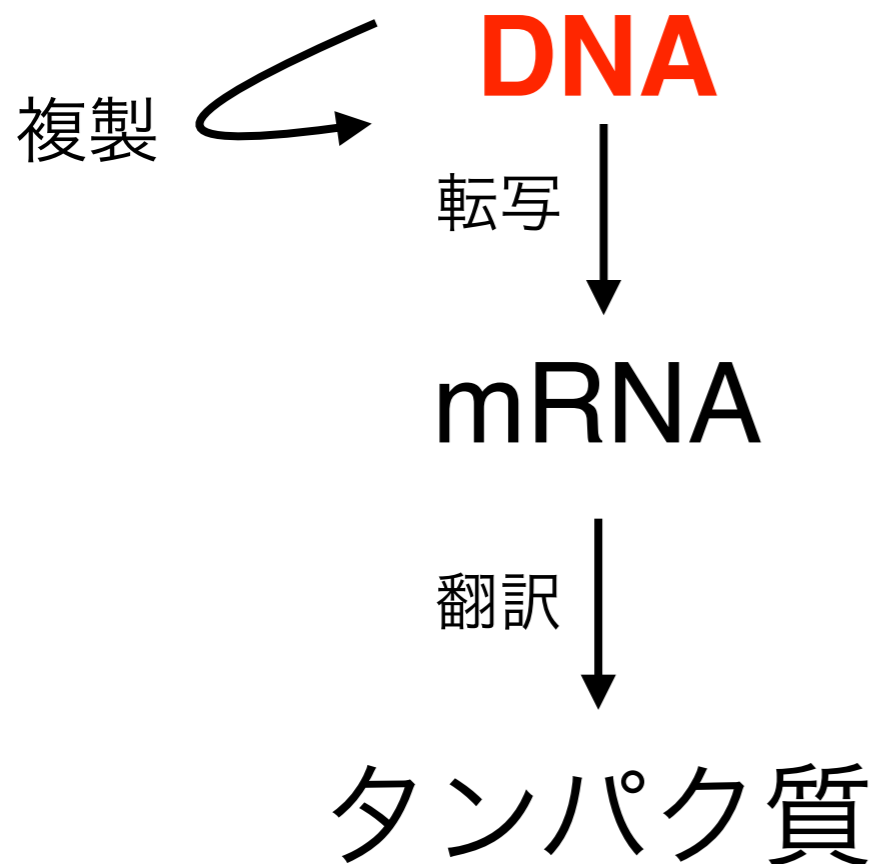


本日の内容

- Next Generation Sequencer (NGS)の位置づけ
- 有用物質高生産菌株のNGSデータ解析
 - ゲノム配列比較
 - トランスクリプトーム解析

各変異世代のゲノムデータを解析し、導入されている変異のうち、どれが高生産につながったのかを明らかにする

セントラルドグマ



オーム研究

ゲノム

トランスクリプトーム

プロテオーム

有用物質高生産菌株のゲノム配列比較

変異させる前の菌のゲノム
(のほんの一部)

高生産菌のゲノム
(のほんの一部)

GCCGATCTCTTCCTTGGCACGCTGGAGCGCCTCCCTCACCGTCAGATCCGACAAGTCTGTCCCGGCGAGC
TCAGTCGCTAGTCCACCGTCGTCGGCCAGCTTGCCGCCGGCCAGGCTCACTTCGTCGAACTGCGCTTGCA
GGGCTTCCGCCTTGGTGAGCGCCGTGGCCAGCTTCTTGCTGACCTCGTCGTAAGTCCGACGGATGTCCTT
CAGTTCCTGATCCTGGAGGCGGACGTAAGTCCCTTGGCCAGTCCGGCAGATCGTCGGCCTTCTGCAGGGCC
TGCCAGCGAGCCGCCCTCACCCAGCCGTGGACCGGTTCTCGGGAACCGAAGCGGCGCTCACCGAACCCA
GCCAGCGCCTCAGCTCATCGATCAGGAATGCGCCGCCCTCGTCGGGCATGCGGTGGGGAGGCCAGTACTT
GTGGTCGTACGGCTCCGAGTCTTTCCGAACTTCGGCCAGTACAGTCGCACAGCCCCCTGCCAGGCGGAC
AGCTGGGCGCCACCTCGTCGGTCAGGTTGCGGGCCGCCAGGTGAGTGGAGAGGTGCGCGACGTGGGCAA
GGCCGGCCAGCTCGCGTGACAGCTTCTGCGTGTCCACAAGAGGCCGACCGGTCGACGGATGGTTCGTGAC
CACAAACGACGGGCGAGGGCGCGACGGTGGTTCGGTTCAGCCACTCGACGAACCCCGGG
AGCATCCAGGGGGACGGCTTCGTCGGGTGTTCCGCATCGCCGACCACGAACCTCCC
CGATCGCCGGGGAGGAGAACCAGGAAGACGGGACGGCGCAGCTGGAACATGTCGCT
GATGCGTACGGTGGCGCACACCCCGTCGCTGCGGGAAGACAGGGCCAGACTCGTG
GGATCGTCCGCCGCGGG GCGTCGT
GGGTGGTCAACGTGTGT CAAGGAC
CGTTGGCCGGCCGACCC CGCCTCG
GGTTCGGTGGAGCAGGAGCTGGCATGCGTAAAGTTCGCGCATCTGGAACTTTTCGG
GCGATGCCGGAGCTGGGGCTTGTGGAGCACGTACCGGTGTGAAGTAGCAACGCCG
GTTCCAGCAGTGGGCGGAGGCGGGCGCCTCGGCCGACGAGTGCCTGCTACGAAC
CACGCCTTCGCCGTAAGTCTTCTCCATACGGGAGGGAGCAGAGTGGACGGCTGAG
CGGTTCTGGATGTACAGGTTCATCGACCCTCGTGACACACTGCGCCTGCTGTGCGCCATGATCGGCCAAAC
GTTTCGTAGGGGGAACAGTGCAGAACTCGGCCTGTCCGCGGTGATGATCTCCATCCTGGCTGTCACCGGG
TGCGGCGGAAGCCAGGACGGAGACGCCCGCCGCGGCAAGCCCGCAAGAGCAAGGGCGCGTCCGCATCCC
CGGCAACGCAGAAGAGCAAGGCCCTGAGCGGAACCTGCAGCAAGTGCAGCGGGACCTGAGGTTCCCGT
GGCGCCCCACAAGCCCCAGATGCAGCCCCCTACAGAGCGGATCCTGCATCGTGCACGGCGTGGCGCCCCACG
CGGAGGGTCTCAGCAAAGCCGAATTCGAGGCGATCGTCGGGCGCTTGCAGAAGCGTGGGTGGTTCGCCCC
TGGGTCCGCTCCGCTACTCCGACGATGAGACCTACGGGAAGATGAGTTACCTCAGACTCCAATCCGGCGA

数百万文字

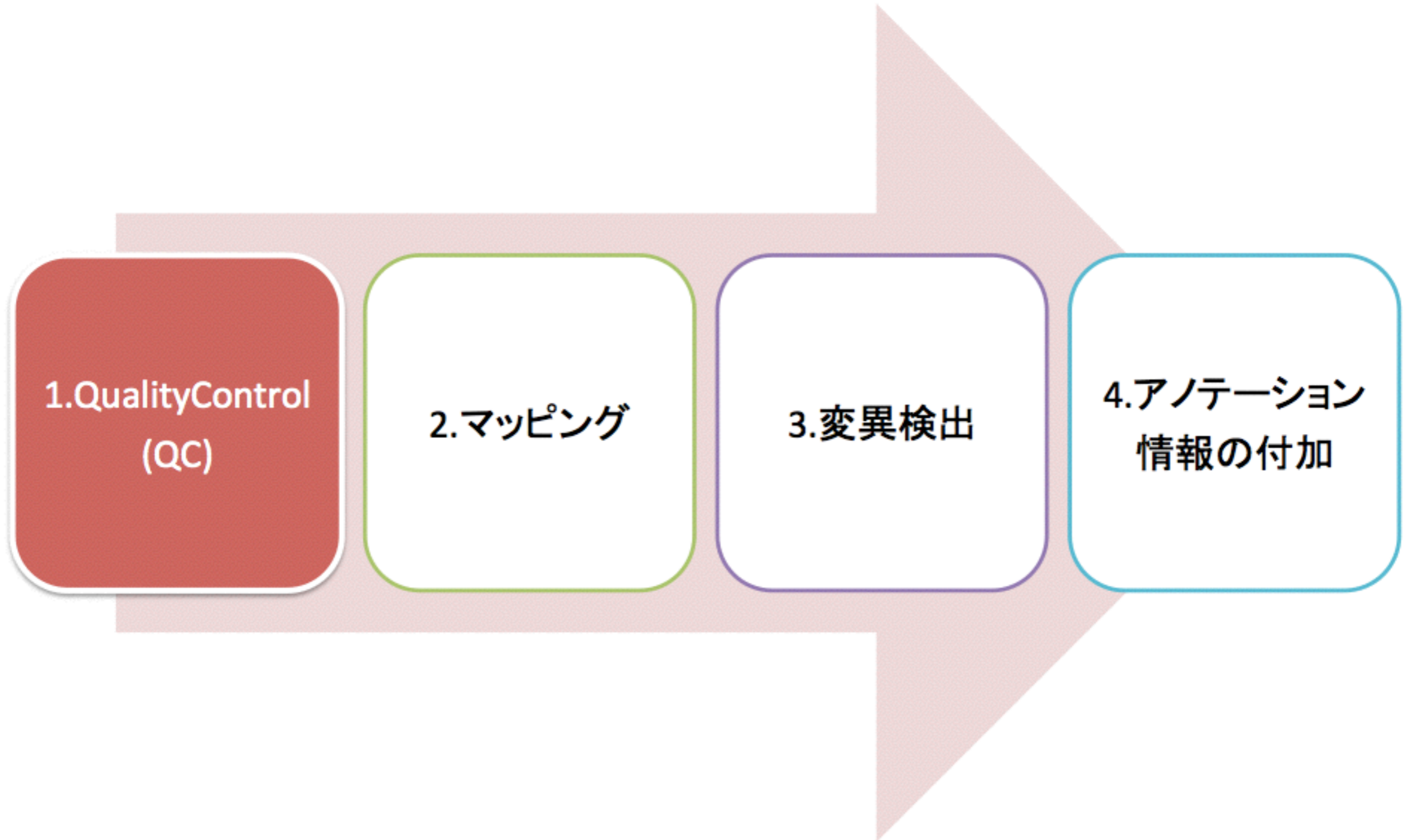


GCCGATCTCTTCCTTGGCACGCTGGAGCGCCTCCCTCACCGTCAGATCCGACAAGTCTGTCCCGGCGAGC
TCAGTCGCTAGTCCACCGTCGTCGGCCAGCTTGCCGCCGGCCAGGCTCACTTCGTCGAACTGCGCTTGCA
GGGCTTCCGCCTTGGTGAGCGCCGTGGCCAGCTTCTTGCTGACCTCGTCGTAAGTCCGACGGATGTCCTT
CAGTTCCTGATCCTGGAGGCGGACGTAAGTCCCTTGGCCAGTCCGGCAGATCGTCGGCCTTCTGCAGGGCC
TGCCAGCGAGCCGCCCTCACCCAGCCGTGGACCGGTTCTCGGGAACCGAAGCGGCGCTCACCGAACCCA
GCCAGCGCCTCAGCTCATCGATCAGGAATGCGCCGCCCTCGTCGGGCATGCGGTGGGGAGGCCAGTACTT
GTGGTCGTACGGCTCCGAGTCTTTCCGAACTTCGGCCAGTACAGTCGCACAGCCCCCTGCCAGGCGGAC
AGCTGGGCGCCACCTCGTCGGTCAGGTTGCGGGCCGCCAGGTGAGTGGAGAGGTGCGCGACGTGGGCAA
GGCCGGCCAGCTCGCGTGACAGCTTCTGCGTGTCCACAAGAGGCCGACCGGTCGACGGATGGTTCGTGAC
GGGCGCGACGGTGGTTCGGTTCAGCCACTCGACGAACCCCGGGATGTCACCTGCGGGT
CAGCTTCGTCGGGTGTTCCGCATCGCCGACCACGAACCTCCGCGAGAAGGGTGCAGG
GAGAACCAGGAAGACGGGACGGCGCAGCTGGAACATGTCGCTGTCTGCTGGAGACC
CGCACACCCCGTCGCTGCGGGAAGACAGGGCCAGACTCGTGGCCACCCGACGGTG
GGCTGGTA TCGGTACGTGGTGGT
TGTCCGCC CCGGACGGTACGAGACC
CCCACTCG GTCCGGCGTGGCCTGC
GAGTGGCA TCGGTAAAGTTCGCGCATCTGGAACTTTTCGACTCGCGGGGATCGA
GGGGCTTGTGGAGCACGTACCGGTGTGAAGTAGCAACCCGGTTCGGCAGGCGGGCTG
CGCGAGGCGGGCGCCTCGGCCGACGAGTGCCTGCTACGAACCCGCGGAACACCGTCC
ACTCCTTCTCCATACGGGAGGGAGCAGAGTGGACGGCTGAGGCGCCTACCGTCCGC
CGGTTCGGATGTACAGGTTCATCGACCCTCGTGACACACTGCGCCTGCTGTGCGCCATGATCGGCCAAAC
GTTTCGTAGGGGGAACAGTGCAGAACTCGGCCTGTCCGCGGTGATGATCTCCATCCTGGCTGTCACCGGG
TGCGGCGGAAGCCAGGACGGAGACGCCCGCCGCGGCAAGCCCGCAAGAGCAAGGGCGCGTCCGCATCCC
CGGCAACGCAGAAGAGCAAGGCCCTGAGCGGAACCTGCAGCAAGTGCAGCGGGACCTGAGGTTCCCGT
GGCGCCCCACAAGCCCCAGATGCAGCCCCCTACAGAGCGGATCCTGCATCGTGCACGGCGTGGCGCCCCACG
CGGAGGGTCTCAGCAAAGCCGAATTCGAGGCGATCGTCGGGCGCTTGCAGAAGCGTGGGTGGTTCGCCCC
TGGGTCCGCTCCGCTACTCCGACGATGAGACCTACGGGAAGATGAGTTACCTCAGACTCCAATCCGGCGA

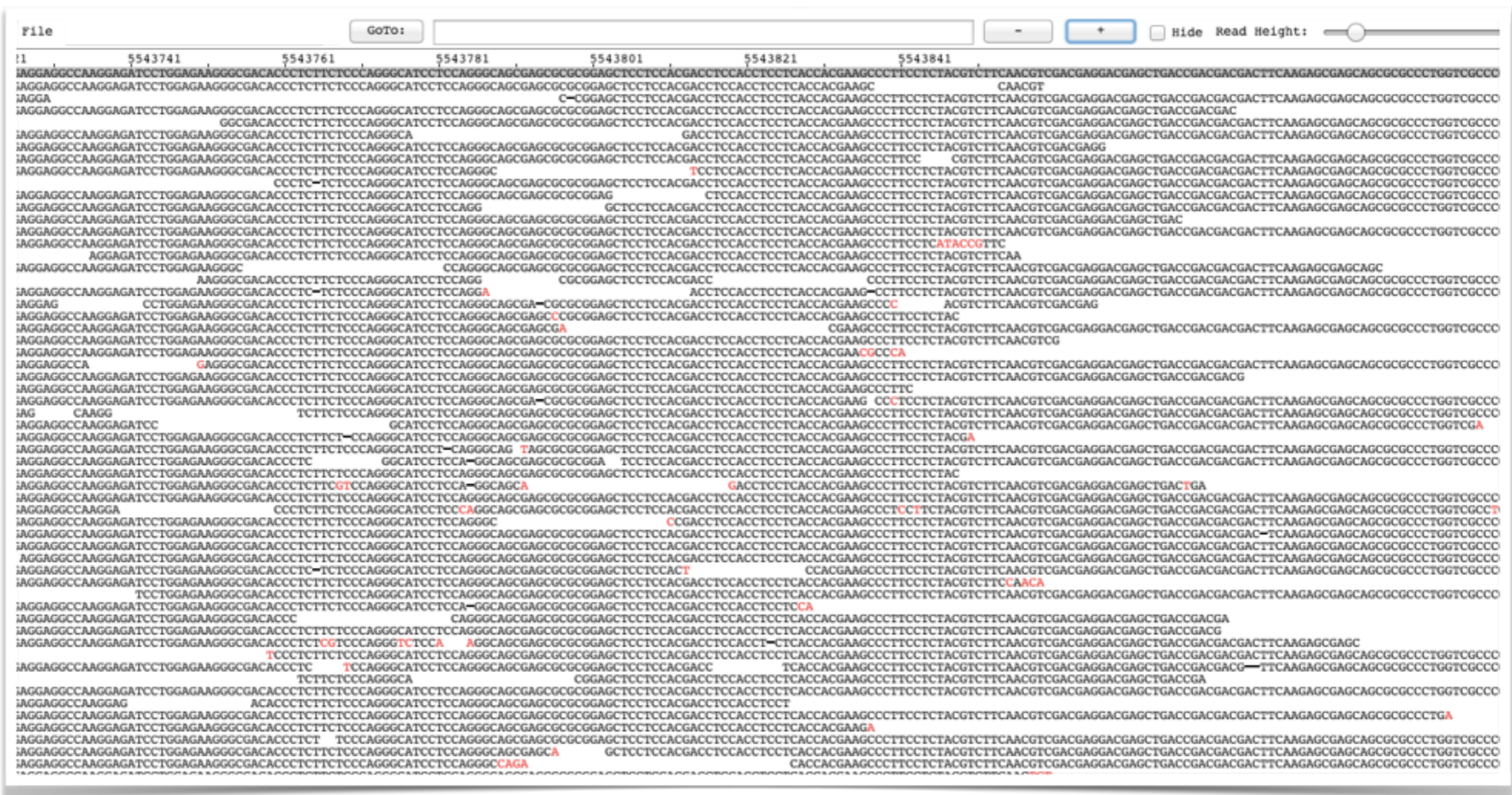
数百万文字

生産量の差を決定付ける違いはどこか？

データ解析の流れ



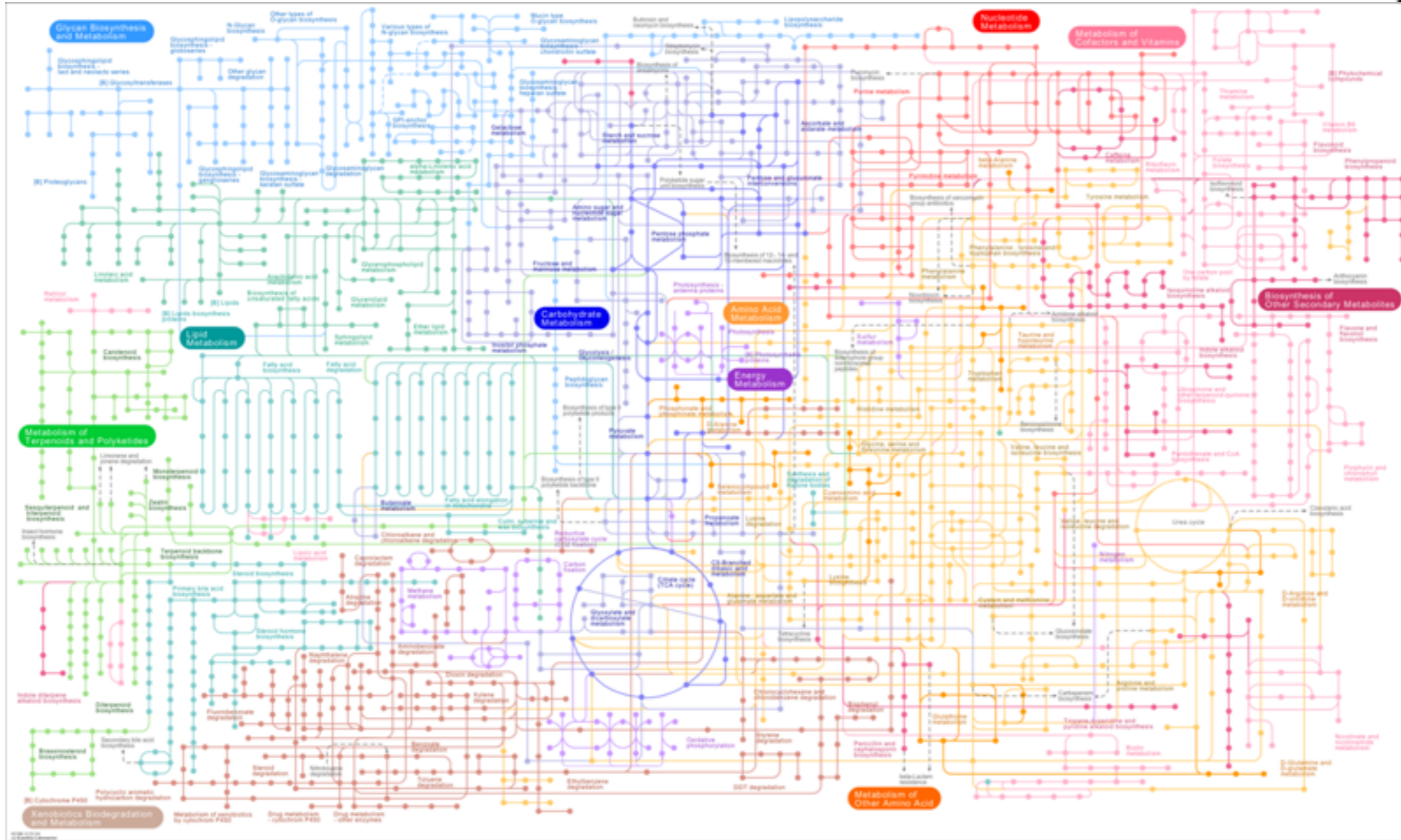
算出される生データの例



The image shows a screenshot of a sequence viewer interface. At the top, there is a 'File' menu on the left and a 'GoTo:' search bar in the center. To the right of the search bar are several control buttons: a minus sign, a plus sign (which is highlighted with a blue border), and a 'Hide Read Height:' checkbox. Below these controls is a horizontal scale bar with numerical markers at 5543741, 5543761, 5543781, 5543801, 5543821, and 5543841. The main area of the viewer displays a large block of text representing genomic sequence data. The text is organized into lines, with some lines starting with a line number (e.g., :1) and a column number (e.g., 5543741). The sequence consists of a mix of uppercase letters (A, C, G, T) and some lowercase letters, representing nucleotide bases. There are several instances of red text interspersed within the sequence, such as 'ATACGGTTC' and 'CGCCCA', which likely indicate specific features or mutations. The overall layout is clean and professional, typical of a bioinformatics software interface.

原因遺伝子の特定

細菌の中の配線図

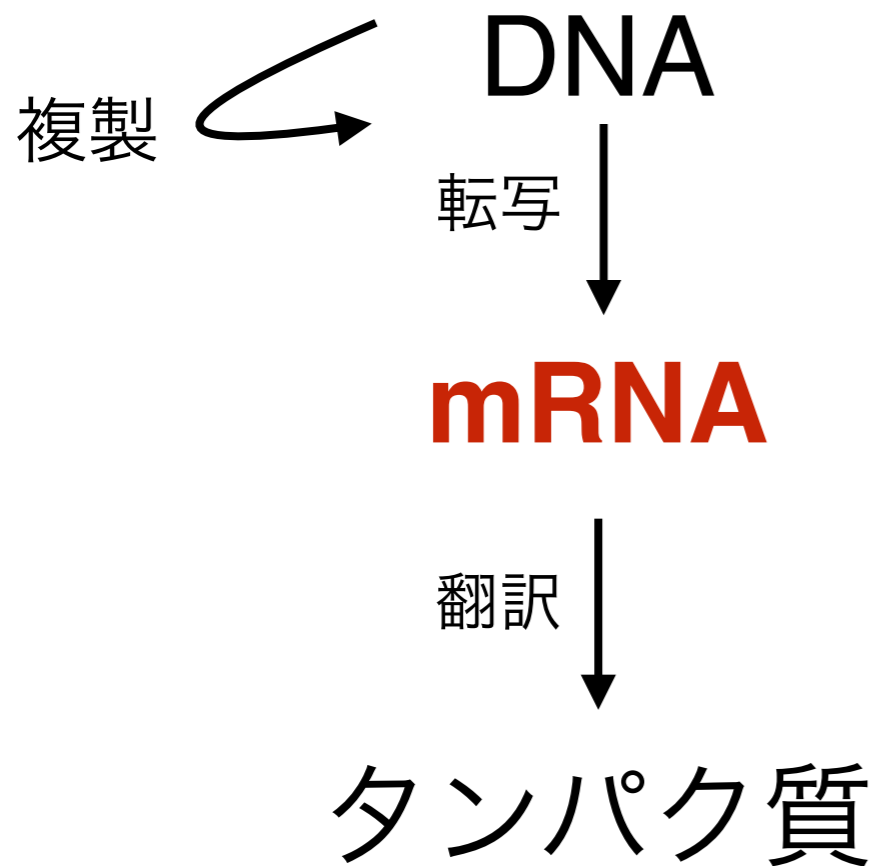


どこの遺伝子に変異が導入されているかを探す

有用物質高生産菌株のトランスクリプトーム解析

各変異世代において、どのような遺伝子の発現が上昇/低下したかを明らかにする

セントラルドグマ



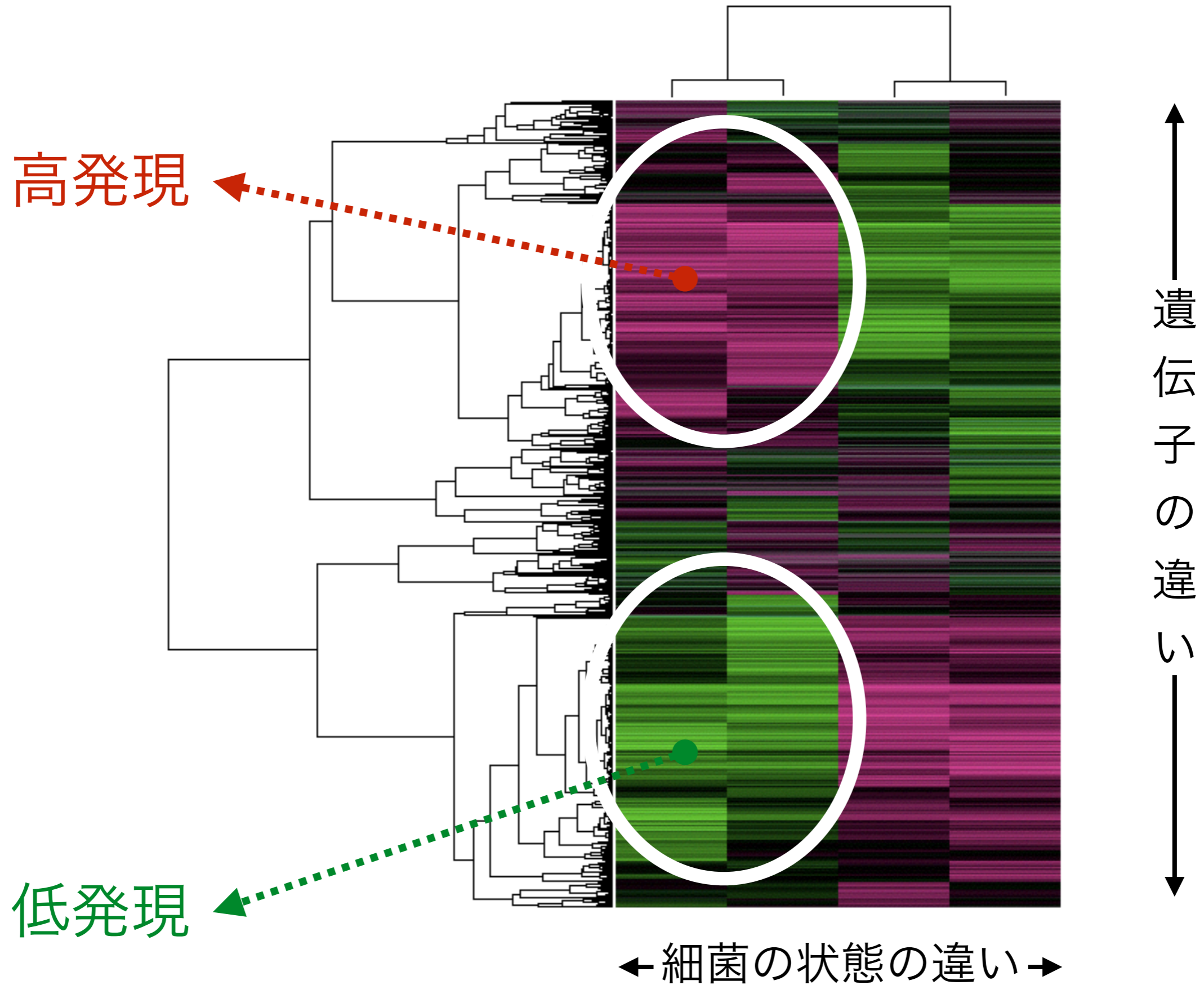
オーム研究

ゲノム

トランスクリプトーム

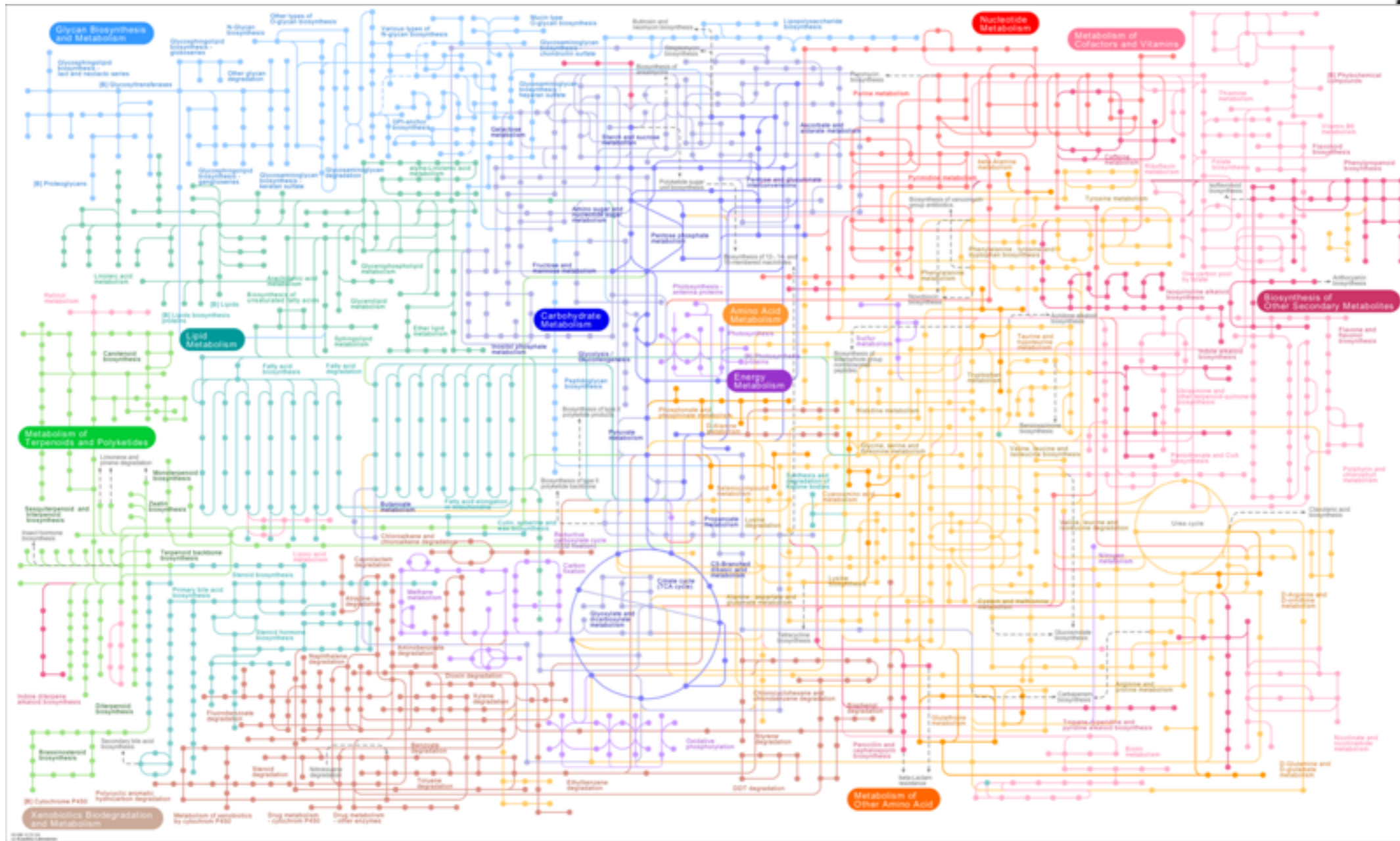
プロテオーム

有用物質高生産菌株のトランスクリプトーム解析



原因遺伝子の特定

細菌の中の配線図



どの遺伝子の発現量(増減)が変動しているかを探す