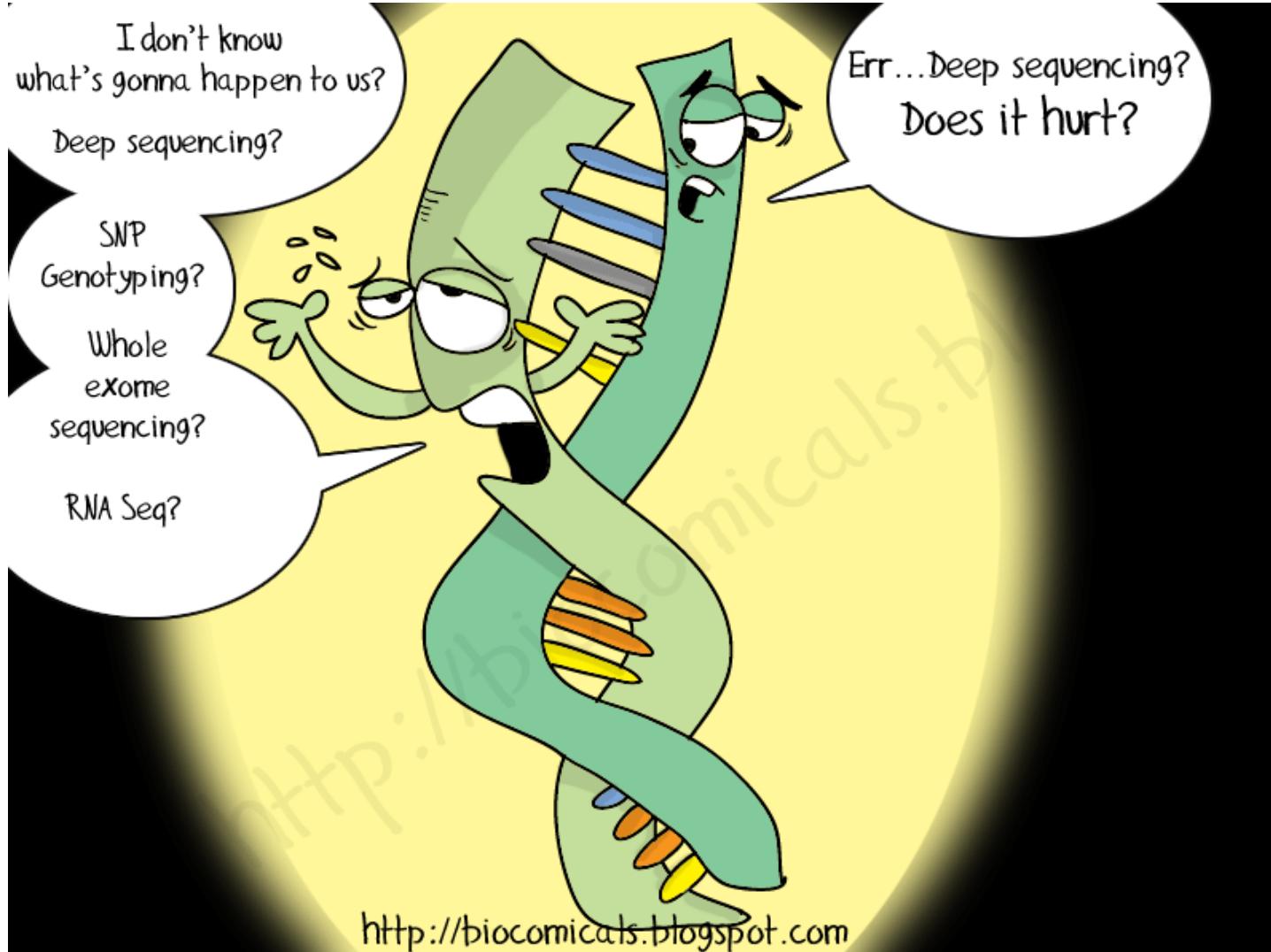


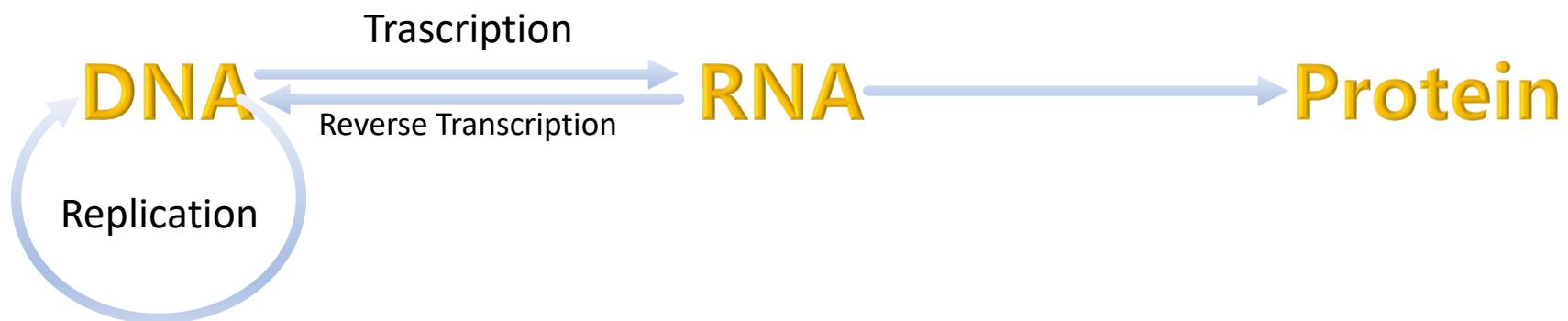
Omics Technologies

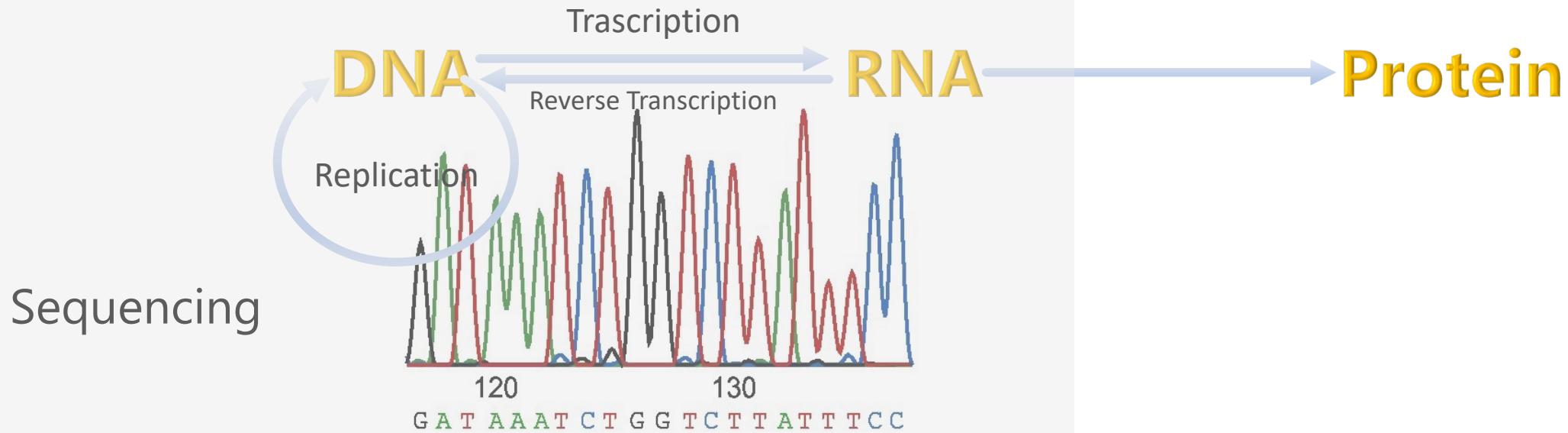


Omar S. Harb, PhD
University of Pennsylvania
 @omartheharb

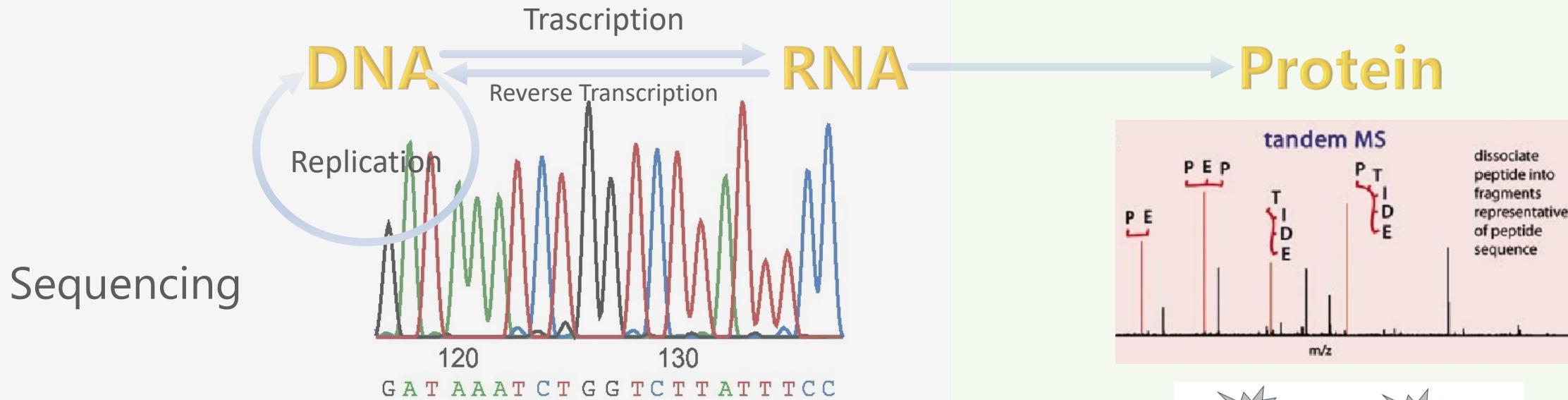
- ***Genome Assembly***
- ***Genome annotation***
- ***Transcriptomics***
- ***Proteomics***
- ***Metabolomics***
- ***Epigenomics***
- ***Evolutionary biology***

Central Dogma

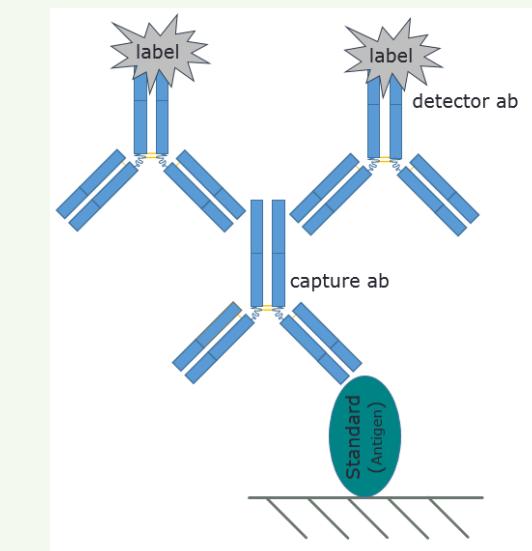




- Genome assembly
- Sequence variation
- Epigenomics
- RNA expression



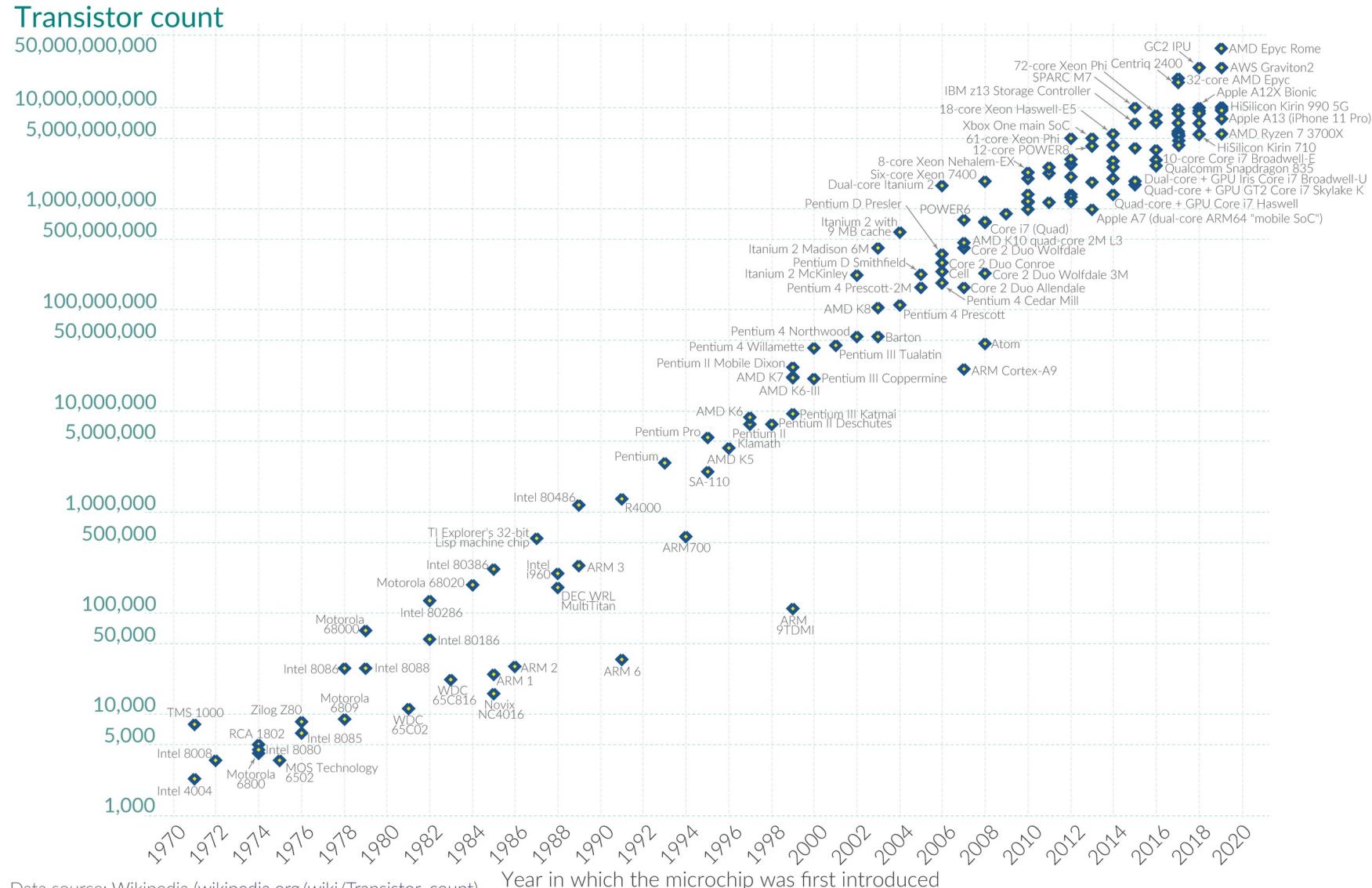
- Genome assembly
- Sequence variation
- Epigenomics
- RNA expression



Moore's Law: The number of transistors on microchips doubles every two years

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Our World in Data

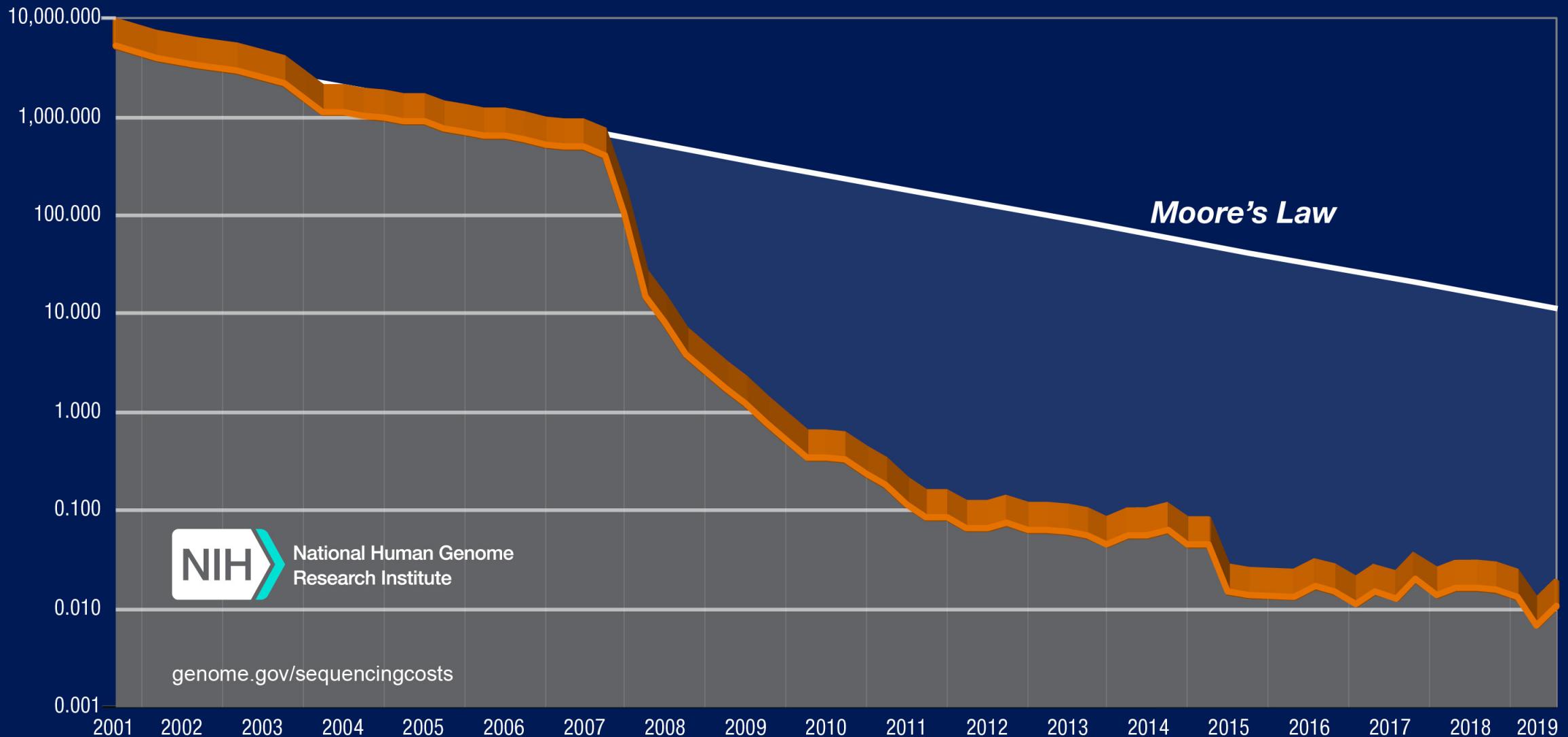


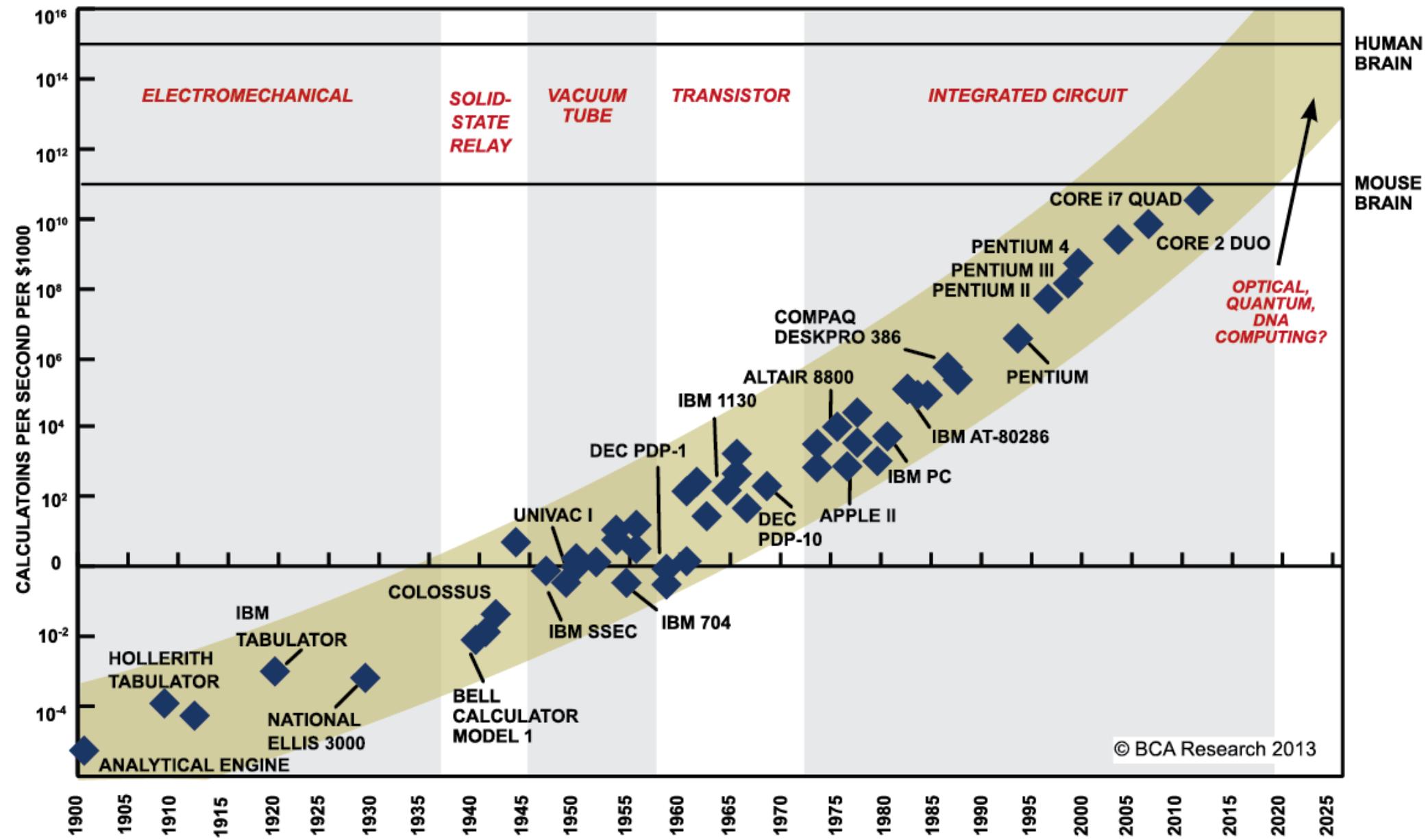
Data source: Wikipedia ([wikipedia.org/wiki/Transistor_count](https://en.wikipedia.org/w/index.php?title=Transistor_count&oldid=1000000000))

OurWorldinData.org – Research and data to make progress against the world's largest problems

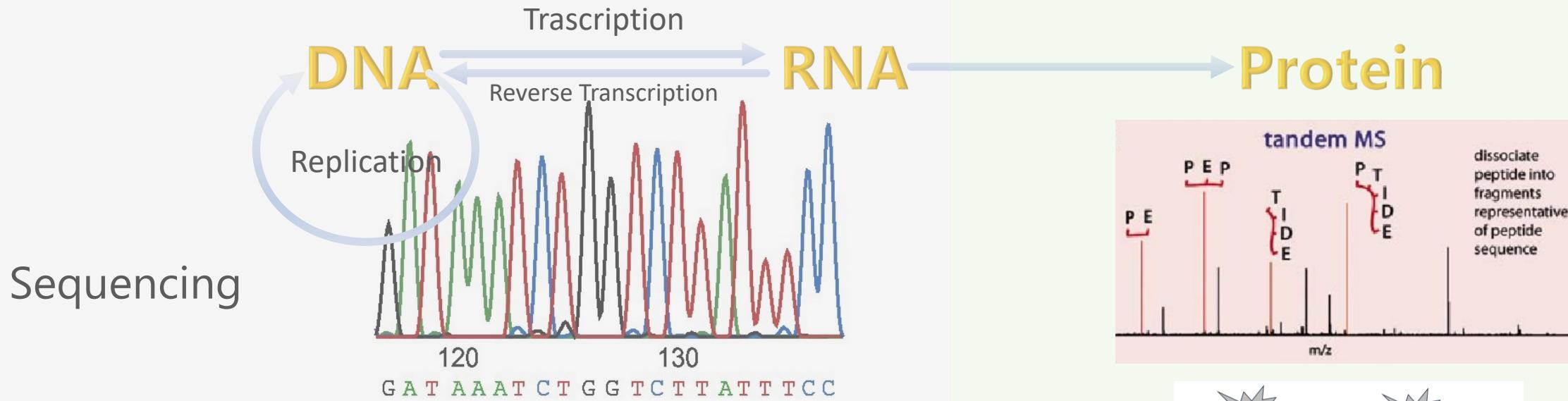
Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

Cost per Raw Megabase of DNA Sequence

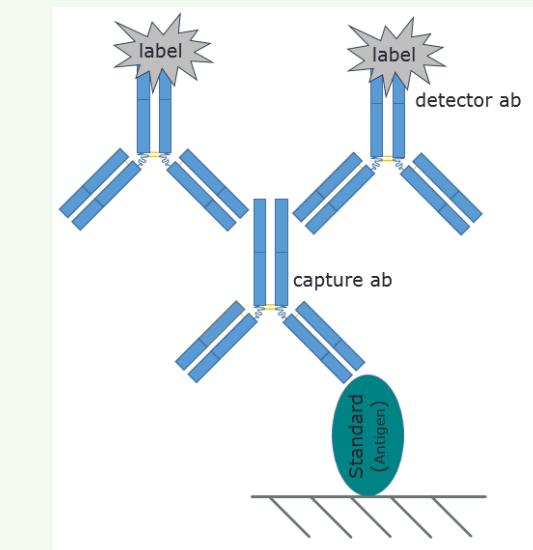




SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, THE VIKING PRESS, 2006. DATAPoints BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

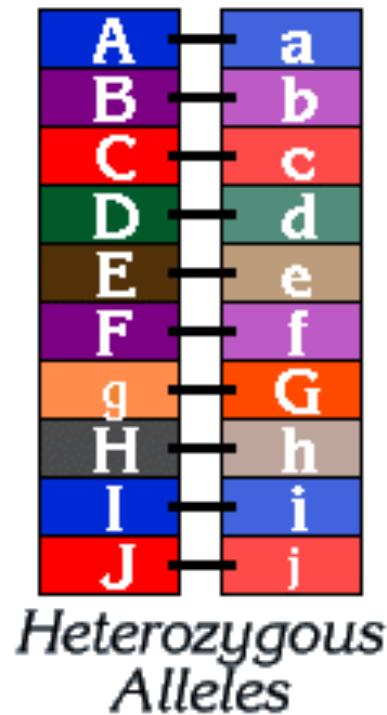
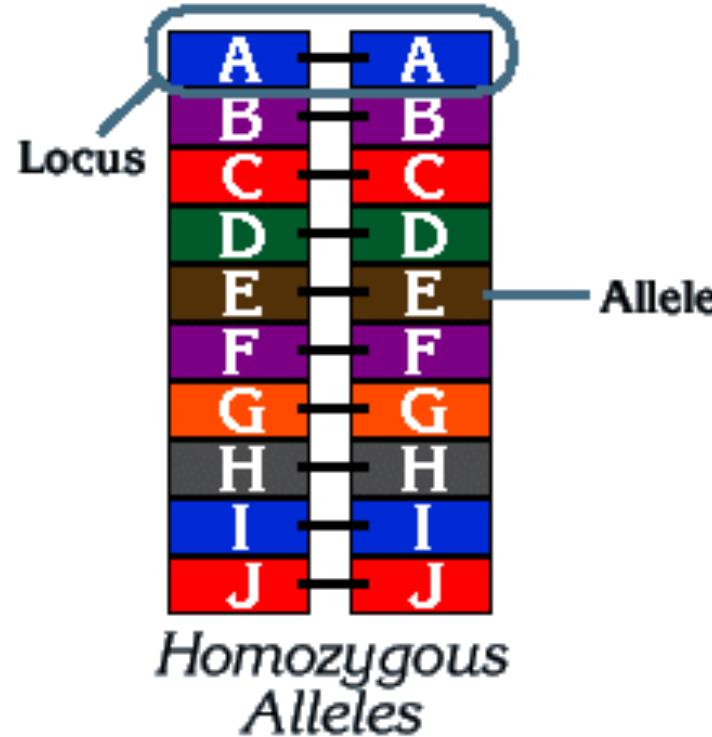


- Genome assembly
- Sequence variation
- Epigenomics
- RNA expression



Loci, alleles and SNPs in a population

Chromosome

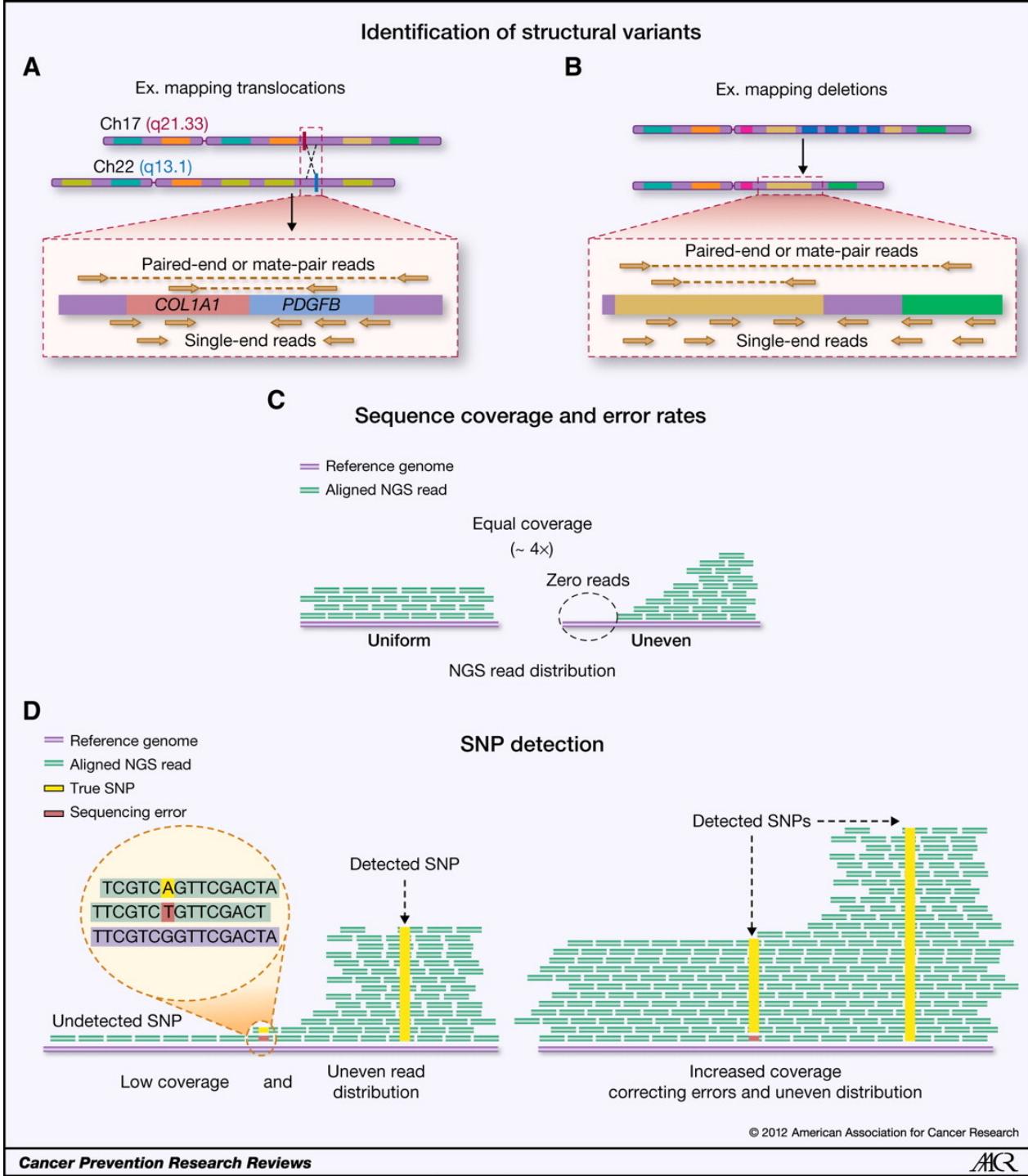


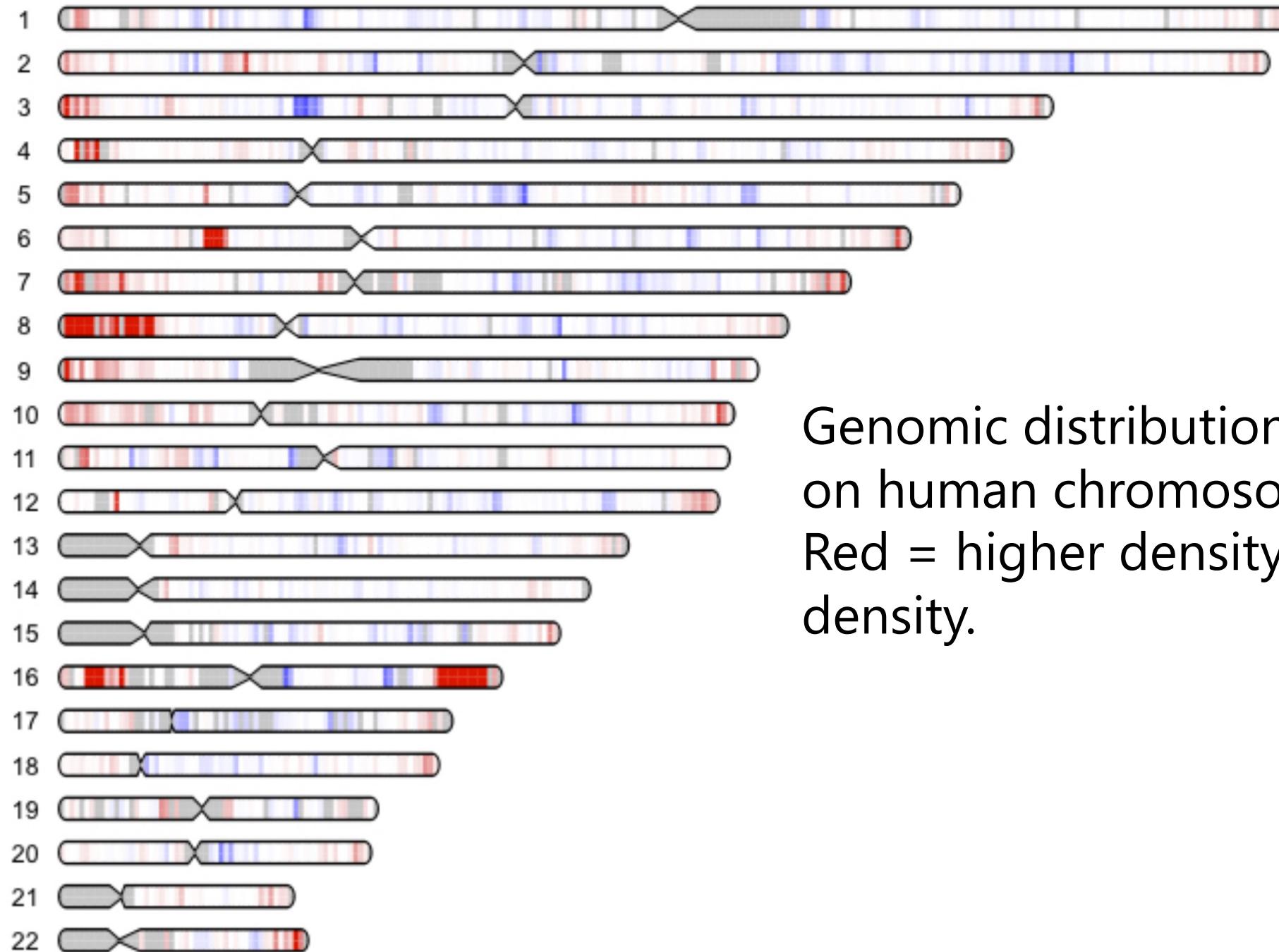
A AAGCCTCATC
a ACGCCTCATC

SNP =Single Nucleotide Polymorphism

Calling SNPs from sequence reads:

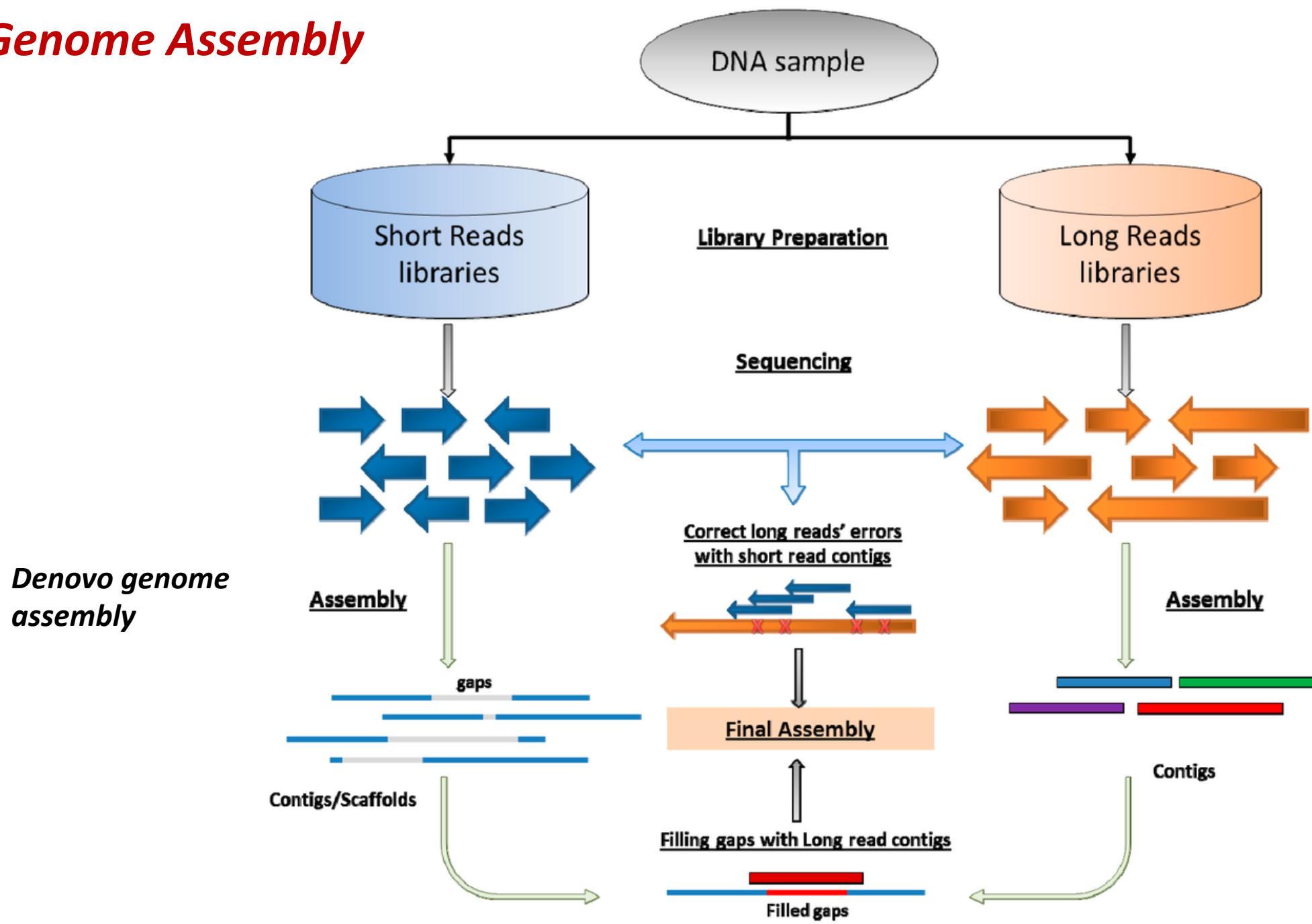
- Depth of coverage
- Allele frequency
- Copy number variation (CNV)



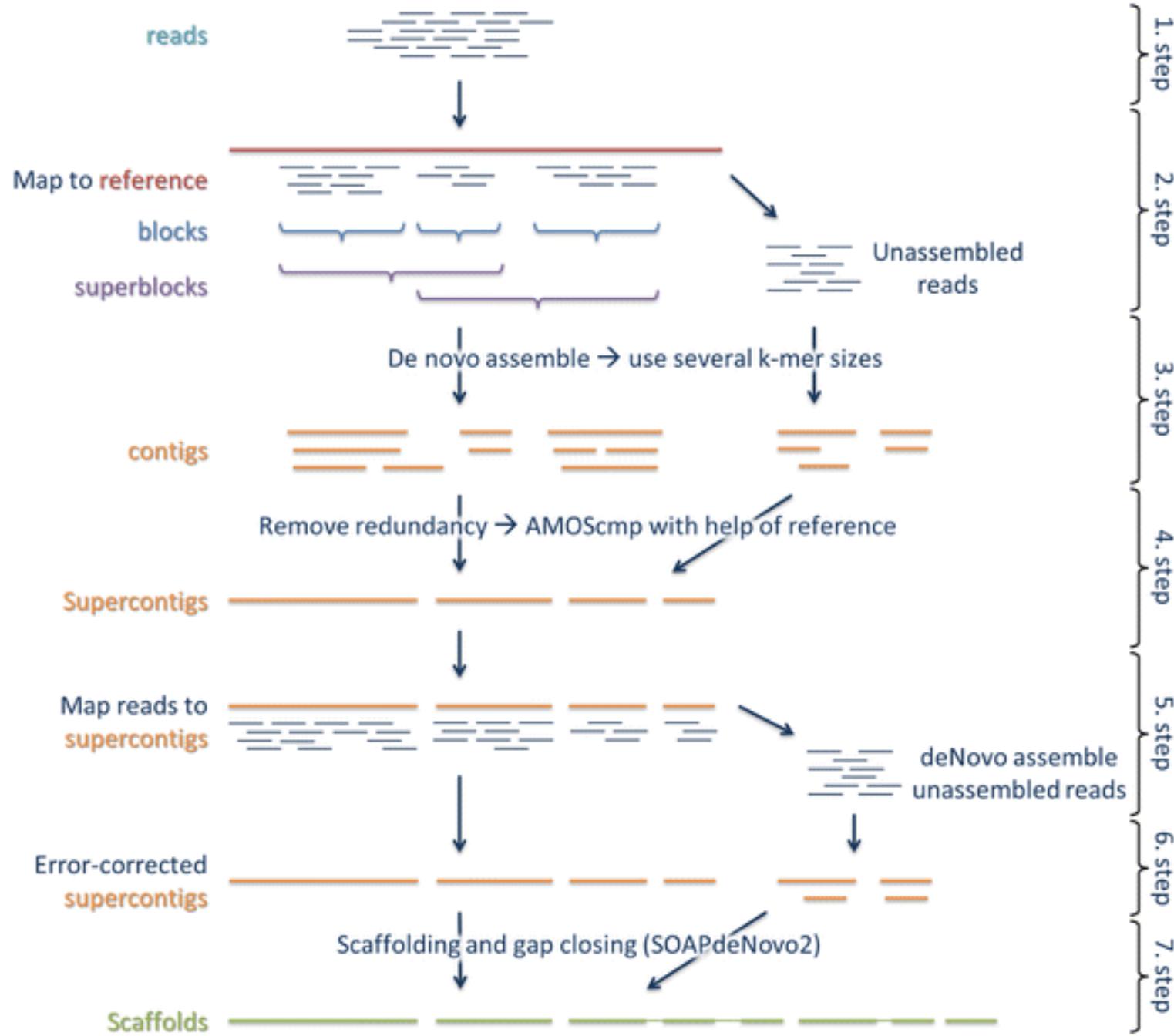


Genomic distribution of SNP density
on human chromosomes 1 to 22.
Red = higher density. Blue = lower
density.

Genome Assembly



Reference genome guides assembly



We have an assembly – now what?

AAGCTTCGCCAGGCTGTAAATCCCGTAGTCGTCCTCACAAATCATCAAGCAGGTGTCTCAGGGAGACTGCCTGACTGAGTTAGCTAATTCTTCTACTTTGGCGTGGTCACGTGT
AACCATAATCGGAATCATTCTCTAGCCCTACGAACAGGTAAAGAGCGTAGGGATGTCCTGGAGTAGTGTGCTTAATCGATAATATTCAAGTGGGACTACCAGCGAGGCCTCGCTTTG
CTCACGCAATGCCAGACAGTGAGAACATGAATGGTAACCGACAAACGCCTCATATGCCTTCAAACCTAGTAGACGCCTACTGTCTGAAACTGGCGGTACACGGACAGATAAC
GCCCTTGGCATCGGCATGTCGTACAGAGGTCCTATGAGTGCCACGACTCTAAATCCGGCAGACGGCTGGCTTTGCTTACACGTATTAGCCCGTGGCATTCTCGGAGC
GCACCTGTTCAACACTAGAAAACGGAGTTCTGATCGAGAACCCACCTTCCAGAAGTGTGAAACGCTAGCATGTCATTGATTCACCCCCCGTAGTTCTGTGTCAATTG
TTGTCGAGACAACACTGTCCCAGCCGGTGTCTCCATATGCCTGACTTCCCAGAATTTTCAAGACTTCAGGAAAGACAGGGCTCCGAACGATCTGTCCATGACTGGTAAATCC
ACGACACCGCAATGGCCCCAGCACCTCATCTCGTGCAGGGGACTAACGTTGATGCGTCTGCTTGCATTGCTTCAAAAAAGAGAGCCATCCGTTCCCCG
CACATTCAACGCCCGAGTGCCTTGTCTTTGAGTGGTAGGACGCTTCTAGCGCAGACTACGTGGACATTAAGTCCATTCTCTTTGACAGCACAAACCTTGCATT
AAACCCGCCCGAGATCCGATCTGCTGCTGTCAGTCCCAGTAGCGTCTGTCGGCCGCGCCTCTGTTGGTGGCAGCCGTAACACCTTGTATCTGACTGCCGTGCGGA
AAATGACGCCATTGGAAAATGGGAAACTCATTCTAAAGTATGGAGGTTCTCTTCTCTGTTCTGTTCTGGGTTGATAACCGTGGTCAAGC
TTTCCGTCTCCCTCGTCTTGACATCGAGACCAGGTGTGCAGATCCTCGCTTGTGATCGACGGAGACGCGTGTCTGAGAACCTTTCAACACACGGCAGTGCAGG
GCACTGCTCTAGTGCAACGGGAGGGGACTAACGCTTGTAGTGCCTTCTGCTACGGGGCTGTGCTGCTGGAGATGCAAGAACCCGGTGTGCTGGTGTGCGATG
ACCCCAAGAGGGGATCGGCATCAACAAACGCCCTCCGTTGGCCCACTTGACCACAGATTCAAACACTTTCTCGTGTGACAAAAACGACGCCGAAGAAGCCAGTGCCTGAACGG
GTGGCTCCAGGAAATTGCAAAGACGGCGACTCTGGACTTCCCTCCATCAGTGGCAAGAGATTCAACGCCGTTGTGATGGGAGGAAACCTGGGAAAGCATGCCCTGAAAGT
TTAGACCCCTCGTGGACAGATTGAAACATCGTGTCTCCCTCTCCCTGTGAGCACACACAGTAGTCGCCACACGCTGTTGAGACGTTGCAATCTCAAAGAGTGTGGACGCTGTCACG
TCTTCAAATTTCCAAACATCCGTCTAGTAGACACACCAACAAAAGCACCGCAATCTGCTCATCGGAGGGAGGAGCCGGGGGACACAAACTATCTCAACTCTGAACGA
ACATATCCGGGGCGCAAGACGTCAGTCTCAAATCCAACCCCGAACATCTGCACTCAAGTCACGATTGCGCCGGTACCTCATGTAAGCAGTTCATGAAACCTCC
GATATTACACAGACTGTTGATATGAAATTATGAGATGCAATATACTGAGACGCCGATGCAACTATAGGTTCTGCCCTCATGGATATTCGACCTTCATCACAATTGGT
TTGCCCGTACACCTCCGTACGCTTTCTGGCTTCTCGTCTGTTATCAGCAAAGAAGACATTGGCCGGAGAAGCCTCAAGCTGAAGGCCAGCGCGTCCGAG
TCTGTGCTTCACTCCCAGCAGCTCAGCCTCTGGAGGAAAGTAGAACGATTGTCGACCAGATTGGTGTGGGTATGTTGCTTCAAACCTTGGAACTCCATTGGTCAAG
AAACGTAACGAAACTGTATACTGATATACAGATGTATGGATAATCTAGAGAAAGATACAGGGAAAGACTGGCAAGGATGAAAGACATGCACTTAAACGAAGCAGAGGGCATTGGC
GAGAGGGACGCCGTTATGCTGTGATGTGGCTGTGAATCTACCTCGCCGTTGACTGTCAGCCTGGCAGCTTGTGTTCTCGTCTGTTGCTGTTCTACCTCCCCAACGCCCTC
TATTCCCTCACTCGCAAAGCGCCTCAGTGGCGTCACCGAACACCCCTGGTTCTTCGTTCTGCTAGCTGTTGCTTCTCGCCTGTTGCGTCTGGCTCGGCTCT
CTCTTCCGTGCTGCGTCCAGACTATGTCGCTGTTCCCCACCCCTCTGGCTTCTGCTTCAAGGAGGAGCGGACTGTACGAGGAGCGCTGTCCTGGCGTTGCCTC
GTACATCACCGTGTAGCCCGAGTTCCGTGCGACGTTCTCCCTCGTTCCCGAGATGACATTCTCAAACAAACTGTCGCGAGCTGCACTGCTCGAGCTCG
TGTGTTCCCTTGTCCGGAGCTCGGAAGAGAGAACATGAAGCGACGTATGACCCATCTCATTCCAAGACCTCTCAGACAACGGGGTACCCATGACTTTGTGTTCTC
GAGAAGAGAAGGAAGACTGACGACGAGCCACTGCCAACCGTAAGAGGAACCGAAGCGCTAGATAAGAAAACAACAAAGAGAAGGGTAAACACGAAGAGAAGGGAAATGCCGA
GAAACCGTGGATTACAAGAGATACTAACAGAGCAATGCTTGTGGAGATTGTTAATTCAAGTAGAGACACCCCGCGTGCAGGGTGTGAGAAATAACTGCAACCTGGAGACAGAGATG
CCCGCAGTACACCACTTGTGTTCTCTATGTCATGAGGGTGCTGAACGTCTATGCTACTTAATTGGAGGAGTCGCTCCGAAGCAGCTTGGCTGCCATCGTGTGTTGC
CTTGTGTTCCGAAAGCCAGAAGGCCGCTCCACAGTGAGGGGATATACAGGGACGCCACCGGAGCCGGTTCTGCTGACTCTGCAAGAGCAACGCAATGAGCTCCTGACGT
CCACGAGGGAGACAACCTCCCGTGCACGGGTTGCAAGGCTCTCTCGGCCGAGCCATTGCCCGGTTGGCTGGATGGAGAAGACCGGAAAAACCGCAGCAAAGGAACGT
ATTGGGCCGTTCCGCATGTTCACTTAGAGGCCATGAAGAATTCCAGTACCTTGATCTCATGCCACATTATAACAATGGATGACGAACGGTAACGGCAACTG
CGAGAAAAAGCCACACCGTTCTCGTGAATTGTCGCAAGCCCTCTTGCTTCTGCTTCCACACCGTCTTCTGCTATTCTCCGCCCTTCTGCTCCATGTTCAATTGTC
CTTCAGTCTTCCATCTTCCCTGTTACCTCTGTCATTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT
CAAGCAACTGTCATTGTCAGCGCCTCCCTCCACCGTGAATGGATTGTCGGTTGCCGGTCTCTGGGTCAAGAAAAGGCCGCGCAGTATTCTGAATAATACCCCGCCATTGAA
GAGGCAGAGGAACAAAGAGATAATTCCGGCGCATCTTGTCGGCGCGTCTCTGCTGCTTCAACCCGATGCCCTCTGCTGATGTCCTGCTCTGCTTCTGCTTCTGCT
TTAGGCGTTGGTGTCACTCCAAATCGGCTGCACTATGCGTACTCGCTGGATCAGGCCCTTCAACTTCTCACCACAAAGCGTGTGTTCTGAAAGGGTAAGGGCGTCTCAGTGA
TGCAATATTTGACTTCAGACATTCTAATGTTGACAACCAACGTACAATTGTTGTCGGTGTGCGTGTGACATGTCAGTATGTAAGAGTCGCTACTGAGACTAACGCAC
GAACAGATTGTTATCTGCATGCGTGTGCACCCGTTCTGAGTGTGGAGTTCCGCAACCTCCTTGAATTCTGGGTTGTTATGCGCGACTGGTTGATGTGGC
CTGAGAGAGCACAGATGAAGGTGGGGTATGTGGCGTGTGCAAGAGAAACTCCGGCAAGGGCAGACAGATAAGGAGAGTGGAAATCATTGAACAGTGTGGTGTGTTG
AGGGCCTCGAAGAGAGTGTGGTGTGGTCAATTGCGCGACACGAACGAAACCATCTTGAGAAGGGCGTGAAGGCAAGTACGTTGACCTCTGCTCTGCCAAGCTCAGATG
CTCCACGGCGTTGGTTCTTCTGTTCTGCTTCTGCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT

Bioinformatics uses algorithms

- Algorithms are sets of rules for solving problems or identifying patterns
- Algorithms can be general or case specific and often need to be trained
- Algorithms can have mistakes and bad assumptions which would lead to wrong results or interpretations.

What would you include in an algorithm to find genes?

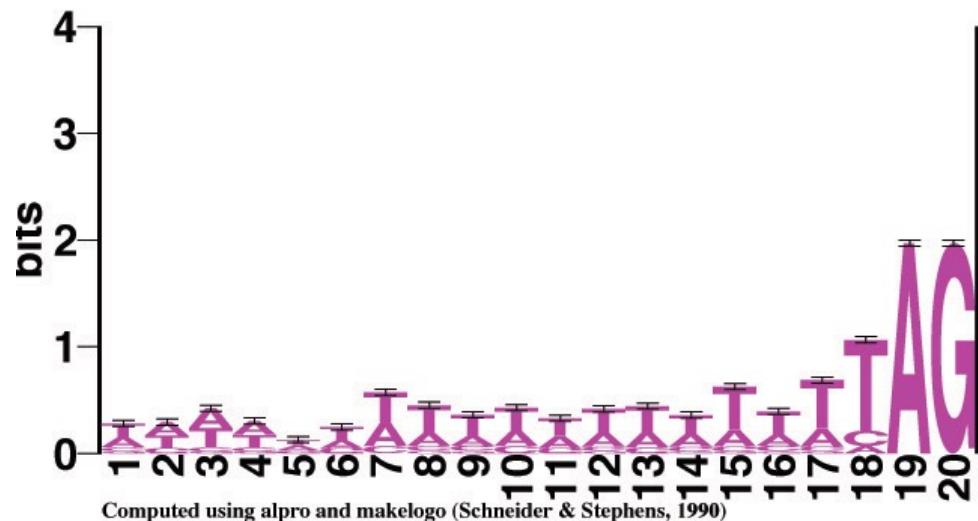
1. Promoter
2. Codon bias – reading frame
3. Start codon and stop codon
4. GC content
5. codons at 3'end
6. splice sites
- 7.
- 8.

What would go into and algorithm to find introns

- Usually begins with GT and end with AG
- Must be longer than 19 nucleotides
- Must contain a branchpoint “A”
- Donor GT often followed by a sequence pattern. This pattern is species-specific
- Acceptor AG often proceeded by pyrimidine stretch
- Has a mean length of “X” as is observed in this species

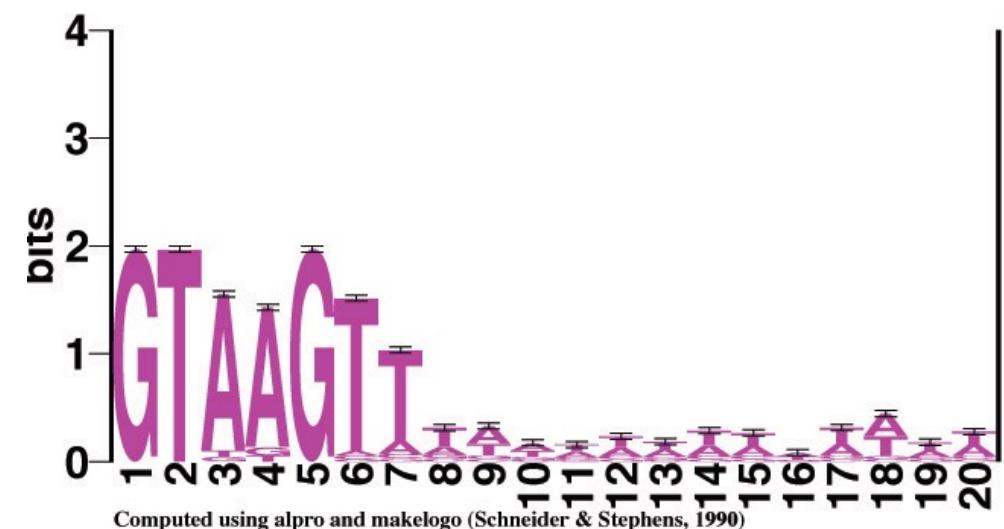
Acceptor site

Generated by <http://www.bio.cam.ac.uk/seqlogo/logo.cgi>



Donor Site

Generated by <http://www.bio.cam.ac.uk/seqlogo/logo.cgi>



Different prediction methods often generate different results



Six Frame Translation ORF-finding

<p>1/1</p> <pre> M Y A L L I L Y Y I I R H * S H H A C R G V Y Y I Y H V R F T D S I L Y Y Y * T L V T S C M * G G L L Y L A C T L Y * F Y I I L L L D T S H I M H V G G S T I S GCA TGT ACG CTT TAC TGA TTC TAT ATT ATA TTA TTA TTA GAC ACT AGT CAC ATC ATG CAT GTA GGG GGG TCT ACT ATA TCT CGT ACA TGC GAA ATG ACT AAG ATA TAA TAT AAT AAT AAT CTG TGA TCA GTG TAG TAC GTA CAT CCC CCC AGA TGA TAT AGA C T R K V S E I N Y * * * V S T V D H M Y P P R S Y R M Y A K S I R Y * I I I L C * D C * A H L P T * * I * H V S * Q N * I I N N N S V L * M M C T P P D V I D I </pre>	<p>31/11</p> <p>121/41</p> <p>151/51</p>	<p>61/21</p> <p>181/61</p>
---	--	----------------------------

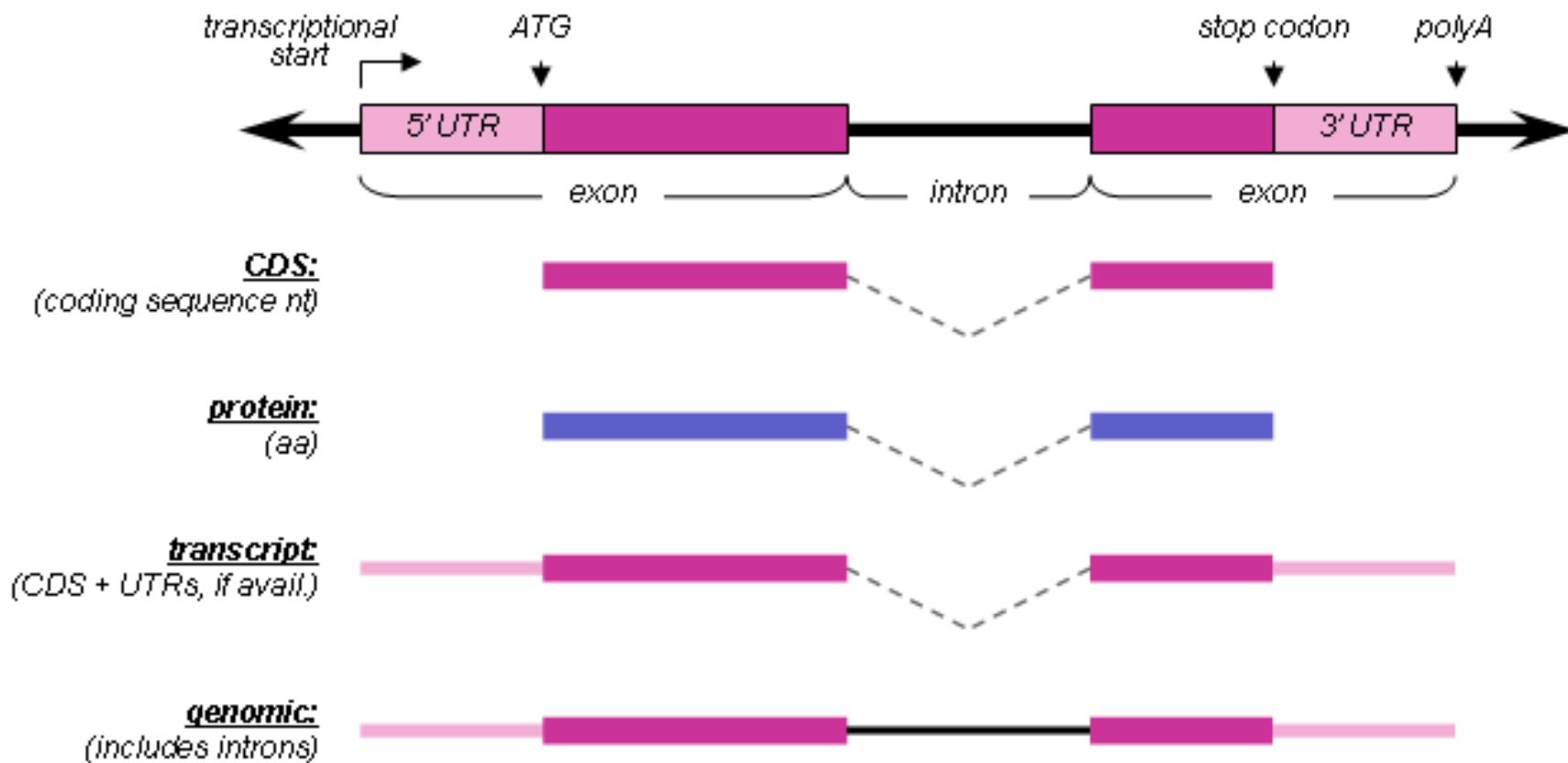
ORFs ≠ Genes

AAGCTTCGCCAGGCTGTAAATCCCGTGAAGTCGTCCTCACAAATCATCAAGCAGGTGTCCCTCAGGGAGACTGCCTGACTGAGTTATGCTAATTCTTCTACTTGGCGTG
GTCACGTGTAAACCATATCGAATCATTTCTAGCCCTACGAACAGGTAAGAGCGCTAGGGATGTCCGTGGAGTAGTGTGCTTACTCGATAATATTCAAGTGGGACTACC
AGCGAGGCGCTCGCTTGTCAAGCAATGCCTGAGACAGTTGAGAATGGTAACCGACAAACCGCTTACATGCGTTCAAACCTAGTAGCTGA
AACTGGCGGTACAGGCACCAGATAACGCCCTGGCATGGCATGTCTCGTACAGAGGTCCGTATGTAGTGCCTACGACTTCTAAATCCGGCAGACAGGCTGGTCTTTGTC
TTACACGTATTAGCCCGCGTGCATTTCTCGGAGCGCACCTGTTCAACACTAGAAAACGGAGTTCTGATCGAGAACGCCACCTTCCAGAAGTTGAACGCTAGCA
TGTCAATTGATTTCACCCCCCGCGTAGTTCTGTGTCATTGTTGAGACAACCTGTGCCCAGGCGTGTCCATATGCGTGAATTCCCAGAATTTC
AGACTTCAGGAAAGACAGGCTCCGGAACGATCTCGTCCATGACTGGTAATCCACGACACCGCAATGGCCCCAGCACCTATCTCTGTGCCAGGGACTAACGTTG
TATGCGTCTCGTCTTGTCTTTGCAATTGCTTCCAAAAAAGAGAGCCATCGTCTCCCGCACATTCAACGCCGAGTGCAGTTTGTCTTTGAGTGGTAGG
ACGCTTTATGCGGAACACTACGGACATTAAGTCCATTCTCTTTGACAGCACAAACCTGCAATTCAAACCCGCCGGAAGATCGATCTGCTGCTGTTCG
CAGTCCCAGTAGCGTCTGTGCGCCGCGTCTGTGGTGGCAGCGCTACACCTGTTATCTGACTGCCGTGCGGAAATGACGCCATTGGAAAATCGGG
AACTTCATTCTTAAAGATGCGGAGGTTCTTCTGTGTTCTTCTCGGGTTGATAACCGTGTGATGTAAGCACTTCCGTCTCCTCCGTGC
TTTGTGACATCGAGACAGGAGTCAGATCCTGCTGATCCGGAGACCGTGTCTGAGAACCTTCAACACAGGCACTGCCAGTGCAGGACTGCTCTG
AGTGCAGCAGGGACGGTGAAGTTCGCTTGTAGTGCCTCTGCTACGGGGCGTGTGCTGAGGAAAGATGCAAGAACCGGGTGTGCTGGCGATGAC
CCCCAAGAGGGGATCGGCATCAACAAACGGCCTCCGTGGCCCACTTGACCACAGATTCAAACACTTTCTGTGACAAAAACGACGCCGAAGAAGCCAGTCGC
TGAACGGGTGGCTTCCAGGAAATTGCAAAGACGGCGACTCTGACTTCCATCAGTCGGCAAGAGATTCAACGCCGTGTCATGGACGGAAAACCTGGGAA
AGCATGCCCTGAAAGTTAGACCCCTGTGGACAGATTGAACATCGTGTCTCCCTCTCCCTGTGAGCACACACAGTAGTCGCCACCGCTGTTGAGACGTGTCATCT
CCAAGAGTGTGGACGCTGTTCACTGCTTCAAATGTTCCAAACATCGTGTCTAGTAGACACACCAACAAAAGCACACGCCGAATCTGCTCATGGAGGGAGGCC
GGGGGGCACACAATCTCAACTCTCGAACGACATTCGGGGCGCGAAGACGCTCAGTCTCAAATCCAACCCCGGAACGCAAACATTCTGCAATCAAGTCACGA
TTGCGCCGGTACCTCCATGTGAAAGCAGTGTGTTCTCGATATTACACAGACTGTGGATATTGAGATATTGAGATCTGAGATCATATACTGAGACGCCGATGCA
ATAGGTTTCTGGCCCTCATGGATAATTGAGACCTTCTCACAATTGGTTGGCCCTACACCTCGTTACGCTTTTTTCTGGCTTCTTCTCGTCTCTGTTATC
AGCAAAGAAGAACATTGGCGCGAGAACGCTCAAGCTGAAGGCCAGCAGCGCTCCGAGCTGTGCTTCACTCCAGCAGCTCTGAGGAAAGAGTACAA
GGATTCTGCGACCAGATTGTGCGTGGTATGTTGCTAAACTCTGGAAACTCCATTCTGGTACGAAACTGTATACTGATATACTGAGATGTATG
GATAATATCTAGAGAACATAGGGAGACTGGCAAGGATGAAAAGACATGCACTTAAACGAGCTTACGAGGCTGTTCTCGTCTCTGAGGAGGGACGCCGTTATGCTGTGATGT
GCTGTGAATCTTACCTGCCGTTGACTGCTGAGCGCTTGTCCACTTGAACGTACTCTGTTCTACCTTCCCTGAGGCTGTTCTGAGGAGGGCTCTCTTCTGTT
CGCTCAGTGGCGTCACCGAACACCTTGGTTCTCGTCTGAGCTGTTCTCGTCTGTTCTGAGGAGGGCTGTTCTGAGGAGGGACGCCGTTATGCTGTGATGT
GTGCGTCCAGACTATGCGCTGTTCCCGACCCCTCTCGGCTTGTGCTTCAGGAGGAGGAGCTGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
ATCACCGTGTAGCCCGAGTTCCGTGCACTTCTCGTGTGTTCCCGAGATGACATTCTCAAACAAATCAACTGTCGCCAGGCTGAGCTGCCGA
GTCTGTGTTCTGTTCCCTTTGTCCGGAGCTCGGAAGAGAGAACATGAAGCGACGTATCGACCCATCTTCAGACAACGGGTTACCTAC
ACCTTGTGGTCTCGAGAACAGGAGACTGACGACGCCACTGCCAGGAGAACGGTAAGAGGCAACCGAAGCGCTGAGATAAGAAAAACACAAAGAGAAGGTGAAAC
ACGAAGAGAACGGGAAATGCCGAGAACCGTGGATTACAAAGATATCAAGAGCAATGCTTGAGGATTTTTTAATTCACTGAGGACACCCGCCGTGCGAGGTG
TAGAAATAACTGCAACCTGGAGACAGAGATGCCGAGTACACCACTGTGCTTTCTCCTATGTTGACGGGTGCTGAACGCTCTATGCTACTTAATTGGAGGAG
TCGTCTCCGAAGCAGCTTGGCTGCCATCCGTGTTGCTTGTGCTTGTGAGGAGAACGCCAGAACGGCGCTCCACAGTGAGGCGATAACGGGACGCCAACGGAGGCCGT
TTCTGCCTTGTGACTCTGCAAGCAACGCAATGAGCTCTGACGCCACAGGAGGAGAACACTCCGTGACGGGTGCTGCCCTCTCGGCCAGCCATTG
CCCCGGTGTGGCGTGGATGGACGAAGAACCGGAAAAACCGCAGGCAAAAGGAACGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
CTTGATCTCATGCCGACATTAAACATGGAGGACAATGGATGACCGAACGGTAACGGGAGACTGCGAGAAAAGCCACACCGTTCTCTGATTTCTGCTGTTACCTGCA
AGCCCTTTGCTTCCATCCACCCCTTGCTATTCTCCGCCCTCTTCTGCTCATGTTCAATTGCTCGTCTTCAGTCTTCCATCTCCCTGTTACCTCTG
TCATTGCTTCTTGCTCTATTAAACTGTTCTACTCACAGTCTGCACTGCCGATAGACGAGCTCCACGCTTGTGCTCGACAAGCAACTGTCATTGTAACGCG
CTCCCTCCACCGTGAATCGGATTGTCGGTTCGCCGTTCTGGGTCAAGAAAAGGCCCTGCCAGTATTCTGAATAATACCCCTGCCATTGAAAGAGGCGAAGGAACAA
AGAGATATTCCGGCGCATTTGTGCGCGCTTCTCGTCTCACACCGATGCCCTCTGCTGATGCTTCTGCTCTCGTCTCTGTTAGG
CGTTGGTGTGATCTCAAATTCCGGCTGCACTATGCGTACTCGCTGGATCAGGCCCTTCACTTCTCACCACAAACGCTGTTCTGAAAGGGTAAGGGCGTCTCAGT
GAATGCATATTGACTTCAGACATTCTAACTGTTGACAACCAACGTAACAAATTGTTGCTGCGTGTGACATGTCAGTGTGAAGAGTCGCTACTGT
AGACTAACGCAACGACAGATTGTTATCTGCATGCGTGTGACCCGTTCTGAGTGTCTGGAGGTTCCGCAACCTCCTTGAATTCTGGGTTGTTATGCA

AAGCTTCCGCCAGGCTGAAATCCCGTGAGTCGTCTCACAAATCATCAAGCAGGTGTCCTCAGGGAGACTGCTGACTGAGTTATGCTAATTCTTCTACTTTGGCGT
GTCACGTAAACCATACTCGAATCATTTCTAGCCCTACGAACAGGTAAGAGCGCTAGGGATGTCGTTGAGTAGTGCTTACTCGATAATATTCAAGTGGGACTAC
AGCGAGGCCTCGCTTGCACCGCAATGCCGAGACAGTGCAGAATGAATGGTAACCGACAAACCGCTCATATGCGTTCAAACCTAGTACGCGTACTGCTG
AACTGGCGGTACAGGCCAACAGATAACGCCCTGGCATCGCATGTCGTACAGAGGTCCGTATGTAGTGCACGACTTCTAAATCCGGCGACAGGCTGGCTTTGT
TTACCACGTTAGGCCGCGGATTCGAGCACCCTGTTCAACACTAGAAAACGGAGTTCTGATCGAGAAAGCCACACCTTCCAGAAGTTGAACGCTAGC
TGTCTTCAGGAAAGACAGGCTCCGGAACGATCTGTCATGACTGGTAATCCACGACACCGCAATGGCCCCCAGCACCTCTATCTCGTGCAGGGACTAACGTT
TATGCGTCTCGTCTGTCTTGTGATTGCTTCCAAAAAGAGAGGCCATCCGTTCCCCCGCACATTCAACGCCGAGTGCAGGGTTTGTCTTGTAGTGGTAG
ACGTTTCATGCGCAACTACGGACATTAGTCCTCTTCGACAGCACGAAACCTTGCATTCACAAACCCGCCGAGAAGATCGATCTTGTCTGTT
CAGTCCCAGTAGCGTCTGCGGCCGCGCTCTGTTGGGGCAGCGCTACACCTGTTATCTGACTGCGTGCAGGAAATGACGCCATTGGGGAAATCGGG
AACTTCATTCTTAAAGTATGCCGAGGTTCTCTCTGTTCTCGGGTTGATAACCGTGTGATGAAACGACTTCCGTCCTCCGTG
TTTGTGACATCGAGACCAGGTGTCAGATCTCGCTGCGATCCGGAGACCGGTGTCAGAACCTTCACTTACACACGGCAGTGCAGGAGCAGTGCCT
AGTGCAGCAGGGACGGGTGAAGTTGCTTAGTAGTGCCTCTGCTCACGGGCGTGTGCTGGAAAGATCAGAAACCGGTGTCGTCGCGATGA
CCCCAAGAGGGCAGCGCATCAACAACGCCCTCCCGTGGCCCACTTGACCAACAGATTCAACACACTTCTCGTGTGACAAAAACGACGCCAGTCG
TGAACGGGTGGCTTCCAGGAAATTGCAAAGACGGGCACTCTGAACTTCCCTCCATCGTGGCAAGAGATTCAACGCCGTTGTCATGGACGGAAACCTGGG
AGCATGCTCGAAAGTTAGACCCCTCGTGGACAGATTGAAACATCGTGTGTTCTCTCCCTGTGAGCACACACAGTAGTCGCCACACGCTGTTGAGACGTGTCATC
CCAAGAGTGTGGACGCTGTTCCACGTCATTCACATGTTCCCAACATCCGTCGTCTAGTAGACACACACCAACAAAAGCACACGCCGAATCTGTCATCGGAGGGAGGAGC
GGGGGGCACACAACATCTCAACTCTGAAACGACATATCCGGGGCCGCGAAGACGCTCAGTCTCTCAAATCCAACCCGGAACGCAAACATTCTGTCATCAAGTCACG
TTGCGCCGGTACCTCATGTAAGCAGTTCCATGAAACACTCCGATATTACACAGCAGTGTGAGATATGAAATTATGCAAGATGCAATATACTGAGACGCCGATGCAAC
ATAGGTTCTGGCCCTCCATGGATATTTCGACACCTTCTCACATTGGTTGCCGTAACCCCTCGTACGTTTTCTGCTTCTCGTCTGTT
AGCAAAGAAGAACGACATTGCCGGGGAGAACCTCAAGCTGAAGGCCAGCAGGCCGTCGGAGCTGTCCTCACTCCACGCTCTGCCCTCTGGAGGAAGACTACA
GGATTCTGCGACCGAGATTGTCGTTGGTATGTTGCTCTAAACTCTTGGAACTCCATTCTGGTCAGAAACGTACTGAAACTGATGATACAGATGTT
GATAATATCTAGAGAAGAACGAGACTGGCAAGGATGAAAGACATGCAAGCTTAAACGAGACGGGATTGGCAGAGGGCATTGGCAGAGGGACGCCGTTATGCTGTT
GCTGTAATCTTACCTCGCCGTTGACTTGTGTCAGCGCTTGTCCACATTGAAACGCTGACTCTTCTACCTTCCCAACGCCCTCTTACGCTGCAAAGC
CGCTCAGTGGGGCGTCACCGAACACCCCTTGGCTTCTTCCGTCAGCTGTTGCTCTTCTCCGCTGCTTCTGCGCTCTGCGCTTCTCCTCTTCTGTT
GTGCGTCCAGACTATGCGCTGTTCCCGTCTGGGATGGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCTT
ATCACCGCTGAGCTTCCGTCAGCTTCTCCCTGCGTTCCGGAGATGACATTCTCAAACAAACACTGCTGCGCAGGCTGCGCTCCTGCGCTTCTGCG
GTCTGTTGCTTCCCTTGTCCGGAGCTGGAGAGAACGAGAACGAGAACATGAAAGCAGCTATCGACCCCTCTTCACTTCCAGACACGGGTTACCCCTAC
ACTTTGTGGTTCTGAGAACGAGAACGACTGACGCCACTGCCAGCGAACCGGTAAAGGGCAACCGAACGCCGAGTAAAGAGAACGAGAACGAGGTT
ACGAAGAGAACGGAAATGCGGAGAACCGTGGATTACAAAGATACTCAAGAGCAATGCTTGTGGAGATTTTTAAATTCACTGAGAACACCCGCCGAGGGT
TAGAAATAACTGCGACCCCTGGAGAGACAGAGATGCCGAGTACACCACTGTCGTTTCTGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCG
TCGTCCTCGAAGCAGCTTGGCTGGCCATCCGTCAGTGTGTTGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCGCTTCTGCG
TTTCTGCCCTTGTGCACTCTGCAAGAGCAACCGAACGACTGCTCCITGACGTCCACGGGAGACAACCTCCGTCACGGCTTGCAGGCTCCCTCGGCCGAGCCATT
CCCCGGTGTGGCGTGGATGGACGAAGAACCGAAAAACCGCAGCAAAAGGAACGTTGATTGGCCGTTCCGCTGTTCTGCGCTTCTGCG
CTTGTCTCATTGCCGACATTATAACATGGAAGGACATGGATGACCGAACGGTAACGGCAGTCACTGCGAGAAAAAGCCACACGGTTTCTCCTGTAATTGTCG
AGCCCTCTTGTCTCATCCACCTTGTCTATTCTCCGCGCTTCTCTGCTCATGTCATTCCGCGATAGACGAGCTTCCACGCTTCTGCGCTCAGCAAGCAACTGTCATTGTCG
TCATTGTTCTGCTCTATTAACTGTTGCTACTCACAGTCGCTCATGTCATTCCGCGATAGACGAGCTTCCACGCTTCTGCGCTCAGCAAGCAACTGTCATTGTCG
CTCCCTCAGCGTGAATGGATTGTCGTTGCCGGTTCTGGGTCAAGAAAAGGCGTGCAGCAGTATTCTGAATAATACCCCTCGCCATTGTAAGAGGCGAAGGAACAA
AGAGATATTGCGCGATCTTTGTGCGGCCGTTCTCGTGTCTCACACCGATGCCCTCTGCGATGCTTCTGCTCTGCTCTCTCTCTTCTGTT
CGTTGGGTGTCATCTCAAATTGCCGCTGCACTATGCCGCTACTCGCTGGATCAGGCCCTTCCACTTCTCACACACAAAGCGTGTGTTCTGAAAGGTAAGGGCGTCTCG
GAATGCAATATTGACTTCAGACATTCTAAGTGTGACAAACCAACGTAACAAATTGTTGTCGTCGAGTGTGCAACTGTCAGTATGTAAGAGTCGCTACTG
AGACTAACGCAACGAAACGAGATTGTTTATGTCATGCCGCTGACGCCGTTCTGAGTGTGCTGGAGTTCCGCAACCTCCTTGAATTCTGGGTTGCTGTT
AGAGATATTGCGCGATCTTTGTGCGGCCGTTCTCGTGTCTCACACCGATGCCCTCTGCGATGCTTCTGCTCTGCTCTCTCTCTTCTGTT
CGTTGGGTGTCATCTCAAATTGCCGCTGCACTATGCCGCTACTCGCTGGATCAGGCCCTTCCACTTCTCACACACAAAGCGTGTGTTCTGAAAGGTAAGGGCGTCTCG
GAATGCAATATTGACTTCAGACATTCTAAGTGTGACAAACCAACGTAACAAATTGTTGTCGTCGAGTGTGCAACTGTCAGTATGTAAGAGTCGCTACTG

AAGCTTCGCCAGGCTGTAAATCCCGTGAAGTCGTCCTCACAAATCATCAAGCAGGTGTCCCTCAGGGAGACTGCCTGACTGAGTTATGCTAATTCTTTCTACTTTGGCGTG
GTCACGTGTAAACCATATCGAATCATTTCTCTAGCCCTACGAACAGGTAAGAGCGCTAGGGATGTCCGTGGAGTAGTGTGCTTACTCGATAATATTCAAGTGGGACTACC
ACCGAGGCGCTCGCTTGTCAAGCAATGCCAGACAGTGTGAGAATGAATGGTAACCGACAAACCGCTTATATGCCTTCAAACCTAGTAGCGTACTGTCTGA
AACTGGCGGTACAGGCACCAGATAACGCCCTGGCATGGCATGTCTGTACAGAGGTCCGTATGTAGTGCACAGACTTCTAATCAGGCGACAGGCTGGCTTTGTC
TTTACACGTATTAGCCCGCGTGCATTTCTCGAGCGCACCTGTTCAACACTAGAAAACGGAGTTTCTGATCGAGAACGCCACCTTCCAGAAGTTGAACGCTAGCA
TGTCAATTGATTTCAACCCCCCGCGTAGTTCTGTGTCATTGTTGAGACAACTCTGTCAGGACAACTGTGTCAGGCGTGTCCATATGCTGACTTTCCCAGAATTTC
AGACTTTCAAGAAAGACAGGCTCCGGAACGATCTCGCATGACTGGTAATCCACGACACCGCAATGGCCCCAGCACCTCTATCTCGTGCCAGGGACTAACGTTG
TATGCGTCTCGCTCTTGTCTTTGCAATTGCTTCCAAAAAAGAGAGCCATCCGTTCCCCGCACATTCAACGCCGAGTGCGGTTTTGTCTTTGAGTGGTAGG
ACGCTTTCACTGCGGAACACTACGGACATTAAGTCCATTCTCTTTGACAGCAGAACCTTGCATTCAAACCCGCCCGGAAGATCGATCTGCTGCTGTTCG
CAGTCCCAGTAGCGTCCGTGCGCCGCGTCTCTGGGGAGCCAGCGCTACACCTGTTATCTGACTGCCGTGCGGAAATGACGCCATTGGAAAATCGGG
AACTTCATTCTTAAAGATGCGGAGGTTCCCTTTCTGTGTTCTCTGGGTTGATAACCGTGTGATGTAAGCACTTCCGTCCTCCGTGC
TTTGTGACATCGAGACAGGTTGCAAGATCCTCGCTTGTGATCCGGAGACCGCTGCTGAGAACCTTTCATTTACACACGGCAGTGCGGAGCACTGCTG
AGTGCAGCAGGGACGGTGAAGTTCGCTTCTGAGTAGTGCCTCTGCTCACGGGGCGTGTGCTGCTGGGAAGATGCAAGAACCGGTTGCTGTCGCGATGAC
CCCAAGAGGGGATCGGCATCAACAACGGCCTCCGTTGGCCCACTTGACCCAGATTTCAACACACTTTCTGTCAGACAAAACGACGCCGAAGAAGCCAGTCGC
TGAACGGGTGGCTTCCCAGGAAATTGCAAAGACGGGCACTCTGACTTCCCTCTCCATCAGTCGGCAAGAGATTCAACGCCGTTGTCATGGGACGGAAAACCTGGGAA
AGCATGCCCTCGAAAGTTAGACCCCTCGTGGACAGATTGAAACATCGTGTCTTCCCTGTGAGCACACACAGTAGTCGCCACACCGTGTGTTGAGACGTGTCATCT
CCAAGAGTGGGACGCTTCCACGCTTCCAAAGTTCACACATCCGCTGCTAGTAGACACACCAACAAAAGCACACGGCGAATCTGTCATCGGAGGGAGGAGCC
GGGGGGCACACAACATCTCAACTCTCGAACGACATATCCGGGGCGCGAAGACGCTCCAGTCTCTCAAATCCACCCCGGAACGCAAACATTCTGCAAGTCACGA
TTGCGCCGGTACCTCCATGTGAAACGTTGAGAACCTCCGATATTACACACGACTGTTGAGATTGAAATTATGCAAGATCATATAACTGAGACGCCGATGCAACT
ATAGGTTTCTCGCCCTCCATGGATAATTTCAGACCTTCCCTCACATTGGTTGGCCCTACACCTCCGTTACGCTTTCTGCTTCTCTCTGTTTAC
AGCAAAGAAGAACATTGCGGCGAGAACGCTCAAGCTGAAGGCCAGCAGCGCTCCAGTGTGCTTCACTCCACGAGCTCTCAGGCTTCTGGAGGAAGAGTACAA
GGATTCTGCGACCAGATTGTCGTTGAGTGTGCTTAAACTCTGGAAACTCCATTGCTGAGAACCTGACTGAAACTGTATACATGATATAACAGATGTATG
GATAATATCTAGAGAACATAGGGAGACTGGCAAGGATGAAAAGACATGCACTTAAACGAAGACGAGGGCAATTGGCAGAGGGACGCCGTTATGCTGATGTG
GCTGTAATCTTACCTGCCGTTGACTTGCCTGCAAGCGTTGTCCACTTGAACGTACTCTTGTCTACCTTCCCAACGCCCTCTTCCACTGCGAAAGCG
CGCTCAGTGGCCGTACCGAACACCCCTGGTTCTCGTTCTGCTGAGTGTGCTTCTGCTGCTGCTGCTGCTCTCTTCCGTTG
GTGCGTCCAGACTATGCGCTGTTCCACCCCTCTCGGCTTGTGCTTCAAGAGGAGCGGACTGTACGAGGAGCGCTGCTGCTGCTGCTCCTCACCTGTAC
ATCACCGGTGAGCCCGGAGTTCCGTCGACGTTTCTCCCTCGCTTCCCGAGATGACATTCTTCAAAACAAATCAACTGCGCCAGGCTGCACTGCCGA
GTCTGTGTTCTCGAGAACAGGAAAGACTGACGACGCCACTGCCAACCGGTAAGAGGCAACCGAAGCCGCTAGATAAGAAAAACAAACAAAGAGAAGGTGAAAC
ACGAAGAGAACGGAAAGACTGCCAACCGGTTAACAGAGATCAAGAGCAATGCTTGTGGAGATTTTTTAATTCACTGAGACACCCGCCGCGAGGTG
TAGAAATAACTGCGACCCCTGGAGACAGAGATGCCGAGTACACCCTGTGCTTCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
TCGTCTCCGAAGCAGCTTGGCTGGCCATCCGTTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
TTTCTGCCCTTGTGACTCTTGCAGAGCAACGCAATGAGCTCTTGACGCTTCCACGAGGGAGACAACTCCGTCACGGGTTGCAAGGCTCTCTGCCGAGCCATTG
CCCCGGTGTGGCGTGGATGGACGAAGAACGCCGAAAAACCGCAGCAAGGAAACTGATTCGGGGCGTCCGCTGTTCACTTCTGCTGCTGCTGCTGCTGCTG
CTTGATCTCATTGCCGACATTATTAACAATGGAGACATGGATGACCGAACGGGTAACGGGCACTGCGAGAAAAGCCACACCGTTCTCTGCTGCTGCTGCTG
AGCCCTTTGCTTCTGCCCTTGTGCTTCTGCCCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
TCATTGCTTCTGCCCTTGTGCTTCTGCCCTTGTGCTTCTG
CTCCCTCCACCGTGAATCGGATTGTCGCTTCCGCTG
AGAGATAATTGCGGCATCTTGTGCGGCGCTTCTCGTGTGCTTCAACCGATGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
CGTTGGTGTGATCTCCAAATTGCGCTGCACTATGCGTACTCGCTGGATCAGGCCCTTCCACTTCTCACCACAAAGCGTGTGCTGCTGCTGCTGCTGCTG
GAATGCAATTGACTTCAGACATTCTAACTGTTGACAAACGACATGGAAATTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
AGACTAACGCACGAACGAGATTGTTATCTGCATGCGCTGCAACCGTTCTGAGTGTCTGGAGTTCCGCAACCTCCTTGAATTCTGGGTTGTTTATGC

Terminology



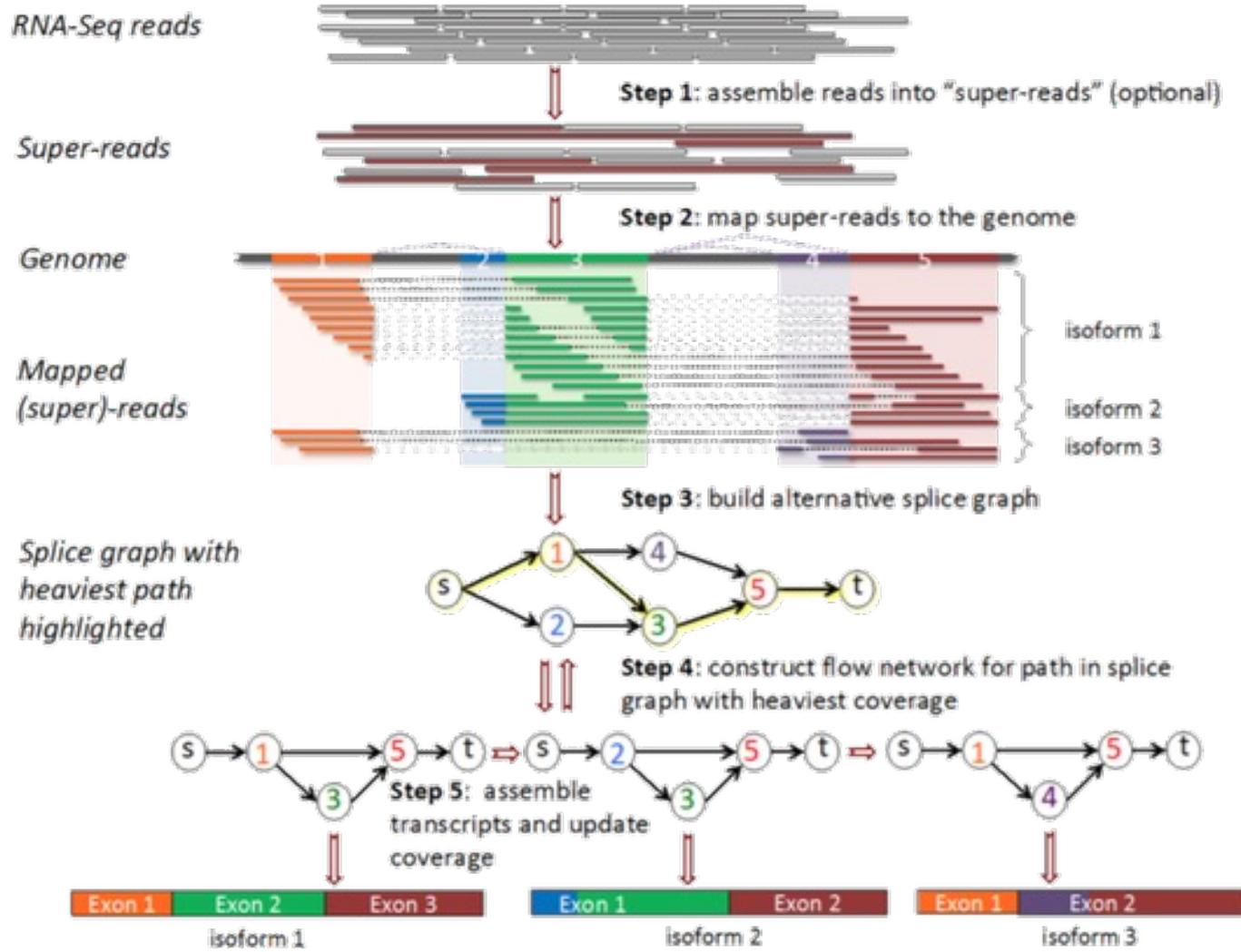
AAGCTTCGCCAGGCTGTAAATCCCGTGAAGTCGTCCTCACAAATCATCAAGCAGGTGTCCCTCAGGGAGACTGCCTGACTGAGTTATGCTAATTCTTTCTACTTTGGCGTG
GTCACGTGTAAACCATATCGAATCATTTCTCTAGCCCTACGAACAGGTAAGAGCGCTAGGGATGTCCGTGGAGTAGTGTGCTTACTCGATAATATTCAAGTGGGACTACC
ACCGAGGCGCTCGCTTGTCAAGCAATGCCAGACAGTGTGAGAATGAATGGTAACCGACAAACCGCTTATATGCCTTCAAACCTAGTAGCGTACTGTCTGA
AACTGGCGGTACAGGCACCAGATAACGCCCTGGCATGGCATGTCTGTACAGAGGTCCGTATGTAGTGCACAGACTTCTAATCAGGCGACAGGCTGGCTTTGTC
TTTACACGTATTAGCCCGCGTGCATTTCTCGAGCGCACCTGTTCAACACTAGAAAACGGAGTTCTGATCGAGAACGCCACCTTCCAGAAGTTGAACGCTAGCA
TGTCAATTGATTTCAACCCCCCGCGTAGTTCCTGTGTCAATTGTTGAGACAACTCTGTCCGGAGACAGGCTTCAATGCGTACTTCCAGAAGTTGAACGCTAGCA
AGACTTTCAAGAAAGACAGGCTCCGGAACGATCTCGTCCATGACTGGTAATCCACGACACCGCAATGGCCCCAGCACCTCTATCTCGTGCAGGGACTAACGTTG
TATGCGTCTCGTCTTGTCTTTGCAATTGCTTCCAAAAAAGAGAGCCATCGTCTCCCGCACATTCAACGCCGAGTGCGGTTTTGTCTTTGAGTGGTAGG
ACGCTTTCACTGCGGAACACTACGTGGACATTAAGTCCATTCTCTTTGACAGCAGAACCTTGCATTCAAACCCGCCGCGAAGATCGATCTGCTGCTGTTCG
CAGTCCCAGTAGCGTCCCTGCGCCGCGTCTCTGGGGAGCCGAGCGCTACACCTGTTATCTGACTGCCGTGCGGAAATGACGCCATTGGGAAATCGGGG
AACTTCATTCTTAAAGATAGCGGAGGTTCCCTTTCTGTGTTCTCTGGGTTGATAACCGTGTGATGTAAGCACTTCCGTCCTCCGTGC
TTTGTGACATCGAGACAGGTTGCAAGATCCTCGCTTGTGATCCGGAGACCGTGTCTGCTAGAACCTTTCATTTACACACGGCAGTGCGGAGCACTGCTG
AGTGCAGCAGGGACGGTGAAGTTCGCTTCTAGTGTGCTTCTGCTACGGGGCGTGTGCTGCTGGGAAGATGCAAGAACCGGTTGCTGTCGTCGCGATGAC
CCCAAGAGGGGATCGGCATCAACAACGGCCTCCGTGGCCCACTTGACCCAGATTTCAACACACTTTCTGTCGACAAAACGACGCCGAAGAAGCCAGTCGC
TGAACGGGTGGCTTCCAGGAAATTGCAAAGACGGGCACTCTGACTTCCCTCTCCATCAGTCGGAACAGAGATTCAACGCCGTTGTCATGGGACGGAAACCTGGGAA
AGCATGCCCTCGAAAGTTAGACCCCTCGTGGACAGATTGAAACATCGTGTCTCCCTCTCCCTGTGAGCACACACAGTAGTCGCCACACCGTGTGTTGAGACGTGTCATCT
CCAAGAGTGGGACGCTTCCACGCTTCCAAAGTTCACACATCGTGTCTAGTAGACACACCAACAAAAGCACACGGCGAATCTGTCATCGGAGGGAGGAGCC
GGGGGGCACACAACATCTCAACTCTCGAACGACATTCGGGGCGCGAAGACGTCAGTCTCTCAAATCCACCCCGGAACGCAAACATTCTGCAAGTCACGA
TTGCGCCGGTACCTCCATGTGAAACGTTGCAACGGTGGGATATTACACACGACTGTTGAGATTGAAATTATGCAAGATCATATAACTGAGACGCCGATGCAACT
ATAGGTTTCTCGCCCTCCATGGATAATTTCAGACCTCCCTCACATTGGTTGGCCCTACACCTCCGTTACGCTTTCTGCTTCTCTGCTCTGTTATC
AGCAAAGAAGAACATTGCGCGGAGAAGCCTCAAGCTGAAGGCCAGCAGCGCGTCCAGTGTGCTTCACTCCACAGCAGCTCTCAGGCTTCTGGAGGAAGAGTACAA
GGATTCTGCGACCAGATTGTCGTTGCTGGGATGTTGTCCTAAACCTCTGGAAACTCCATTGCTGCAAGAACCTGACTGAAACTGTATACATGATATAACAGATGTATG
GATAATATCTAGAGAACATAGGGAGACTGGCAAGGATGAAAAGACATGCACTTAAACGAAGACGAGGGCAITGGCAGAGGGACGCCGTTATGCTGTCGAGTGT
GCTGTAATCTTACCTCGCCGTTGACTTGTGCTGCAACGGCTTGTGCTTCACTTGAAACGTACTCTGTTCTACCTCCCAACGCCCTCTGCTTCTGCTCTTCTGTTG
GTGCGTCCAGACTATGCGCTGTTCCACCGAACACCCCTGGTTCTCGTGTGCTTCAAGAGGAGCGGACTGTACGAGGAGCGCTGCTCTGGCGTTGCTCTCACCTGTAC
ATCACCGGTGTAGCCCGCAGTTCCGTGCGACGTTTCTTCCCTCGTGTGTTCCACCGAGATGACATTCTTCAAAACAAATCAACTGCGCCAGGCTGCACTGCCGA
GTCTGTGTTCTCGAGAAGAGAAGAGACTGACGACGCCACTCGGAAACCGTAAGAGGCAACCGCGTAGATAAGAAAAACAAACAAAGAGAAGGTGAAAC
ACGAAGAGAACGGAAATGGGAGAACCGTGGATTACAAAGATATCAAGAGCAATGCTTGTGGAGATTCTTAAATTCACTGAGACACCCGCCGCGAGGTG
TAGAAATAACTGCGACCCCTGGAGACAGAGATGCCGCGAGTACACCACTGTGCTTCTCTCGTGTGCTTCTGCTGCTGCTGCTCTCGTCTCTGCTCTGCT
TCGGCTCCGAAGCAGCTTGGCTGGCCATCCGTGTGTTGCCCTGTTGAAAGGCCAGAAGGCCGCTCCACAGTGAGGCGATAACAGGGACGCCAACCGGAGCCCGT
TTTCTGCCCTTGTGCACTTGTGAGCAACGCAATGAGCTCTGACGCCACAGGGAGACAACTCCCGTGCACGGGTTGCAAGGCTCTCTGCCGAGCCATTG
CCCCGGTGTGGCGTGGATGGACGAAGAAGACCGGAAAAACCGCAGCAAGGAACCTGATTCGGGGCGTCCCGATGTTCACTTCTGAGGCGCATGAAAGAATTCCAGTAC
CTTGATCTCATTGCCGACATTATTAACAATGGAGGACAATGGATGACCGAACGGGTAAACGGCACTGCGAGAAAAGCCACACCGTTCTCTGCTGATTCTGCGCA
AGCCCTTTGCTTCACTCCACCCCTTGCTATTCTCCGCCCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT
TCATTGCTTCTGCTTCACTCCACCCCTTGCTATTCTCCGCCCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT
CTCCCTCCACCGTGAATCGGATTGTCGGTTGCCGGTCTGGTCAAGAAAGGCCCTGCCAGTATTCTGAATAACCCCTGCCATTGAAAGAGGCGAAGGAACAA
AGAGATAATTGCGCGCATTTGTGCGCGTGTGCTTCTCGTGTGCTTCAACCGATGCCCTCTGCTGATGCTTCTGCTTCTGCTTCTGCTTCTGCTTCTGCT
CGTTGGTGTCACTCCAAATTGGCTGCAACTATGCGTACTCGCTGGATCAGGCCCTTCCACTTCTCACCACAAAGCGTGTGTTCTGGAAAGGGTAAGGGCGTCTCAGT
GAATGCAATTGACTTCAGACATTCTAACTGTTGACAAACGACGTAACAAATTGTTGCTGCGTGTGCTGACATGTCAGTGTGAAAGAGTCGCTACTGT
AGACTAACGCACGAACAGATTGTTATCTGCATGCGTGTGCAACCGTTCTGAGTGTCTGGAGTTCCGCAACCTCCTTGAATTCTGGGTCGTTTATGC

Combine exons into a coding sequence and translate into a predicted amino acid sequence

ATGCAGAAACCGGTGTCTGGTCGCGATGACCCCCAAGAGGGCATCGGCATCAACAACGGCCTCCGTGGCCCC
ACTTGACCACAGATTCAAACACTTTCTCGTGTGACAAAAACGACGCCGAAGAACGCCAGTCGCCTGAACGGGTGGCT
TCCCAGGAAATTGCAAAGACGGCGACTCTGGACTTCCCTCCATCAGTCGGCAAGAGATTCAACGCCGTTGTATG
GGACGGAAAACCTGGGAAAGCATGCCTCGAAAGTTAGACCCCTCGTGGACAGATTGAACATCGTCGTTCTTCCC
TCAAAGAAGAAGACATTGCGCGGAGAAGCCTCAAGCTGAAGGCCAGCAGCGTCCGAGTCTGTGCTTCACTCCCAGC
AGCTCTCAGCCTCTGGAGGAAGAGTACAAGGATTCTGTCGACCAAGATTTGTCGAGGGAGGAGCGGGACTGTACGAG
GCAGCGCTGTCTGGCGTTGCCTCTCACCTGTACATCACCGTGTAGCCCGCAGTTCCGTGCGACGTTTCTTCC
CTGCGTTCCCCGGAGATGACATTCTCAAACAAATCAACTGCTGCGCAGGCTGCAGCTCCTGCCGAGTCTGTGTTCGT
TCCCTTTGTCGGAGCTCGGAAGAGAGAAGGACAATGAAGCGACGTATCGACCCATCTCATTCCAAGACCTTCTCA
GACAACGGGGTACCCCTACGACTTGTGGTTCTCGAGAAGAGAAGGAAAGACTGACGACGCAGCCACTGCGGAACCGAGCA
ACGCAATGAGCTCCTGACGTCCACGAGGGAGACAACACTCCCGTGCACGGGTGCAAGGCTCCTTCCGGCCGCAGCCAT
TGCCCCGGTGTGGCGTGGATGGACGAAGAAGACCGGAAAAAACCGCAGCAGCAGGAACTGATTGGCCGTTCCGCAT
GTTCACTTAGAGGCCATGAAGAATTCCAGTACCTGATCTCATTGCCGACATTATTAACAATGGAAGGACAATGGATG
ACCGAACGG

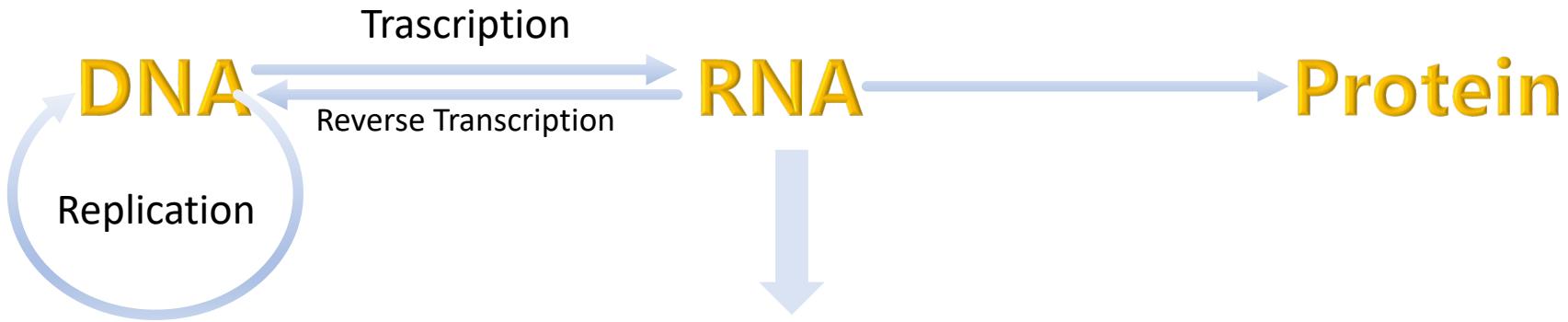
>Translation Frame 1

MQKPVCLVVAMTPKRGIGINNGLPWPHLTDFKHFSRVTKTPEEASRLN
GWLPRKFAKTGDSGLPSPSVGKRFNAVVMGRKTWE SMPRKFRPLVDRLN
VVSSLKEEDIAAEKPQAEGQQQRVRVCASLPAALSLLEEEYKDSVDQIFV
VGGAGLYEAALSLGVASHLYITRVAREFPCDVFFPAFPGDDILSNKSTA
QAAAPAESVFVPFCPELGREKDNEATYRPIFISKTFSNDNGVPYDFVVLEK
RRKTDDAATAEPSNAMSSLTSTRETPVHGLQAPSSAAAIAPVLAWMDEE
DRKKREQKELIRAVPHVHFRGHEEFQYLDLIADIINNGRTMDDRT



RNAseq data can help in predicting gene models

Supplementary Figure 12. The StringTie algorithm: RNA-seq reads are assembled into super-reads (Step 1) and then super-reads plus un-assembled reads are mapped to the genome (Step 2). In Step 3, mapped reads and super-reads are used to build an alternative splice graph. We use the path from source (s) to sink (t) with the heaviest coverage to build a flow network corresponding to the transcript represented by that path (Step 4). The maximum flow in this network represents the coverage of one assembled transcript, which is removed from the splice graph (Step 5). Steps 4 and 5 are repeated until no new transcripts are assembled. <http://yourgene.pixnet.net/album>

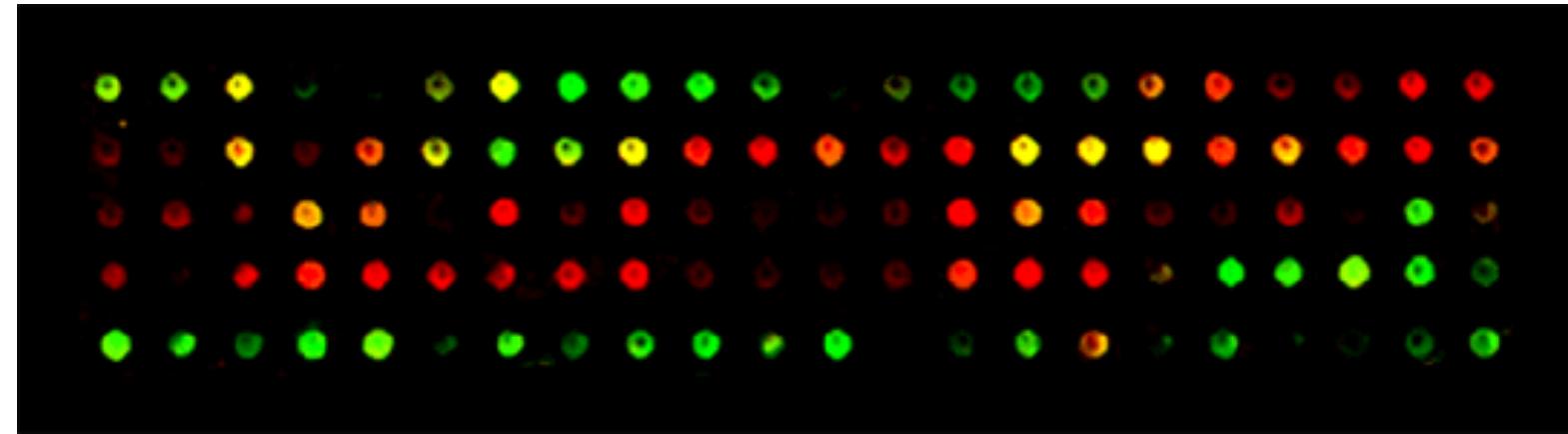


Expression Profiles

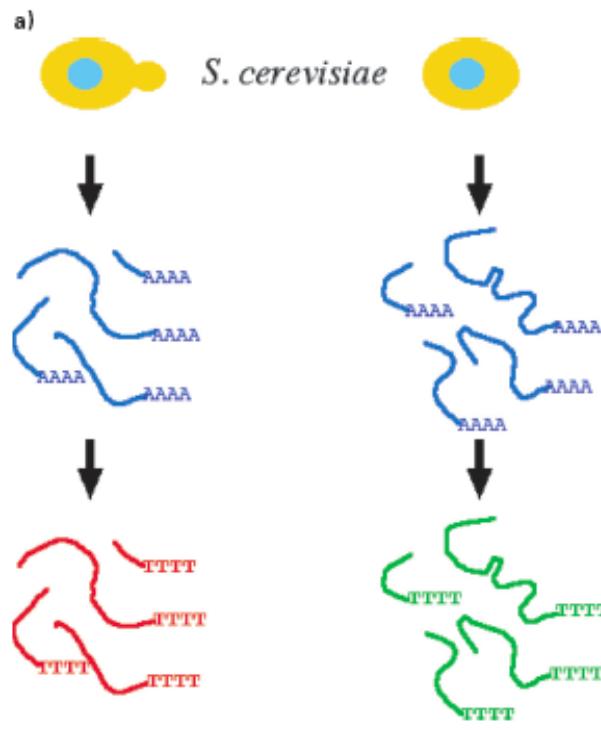
- The pattern of expression of one or more genes over time or a set of experimental conditions, e.g. during development or a drug treatment or in a genetic mutant such as a gene knock-out.
- Always... has a time and space component

Microarrays

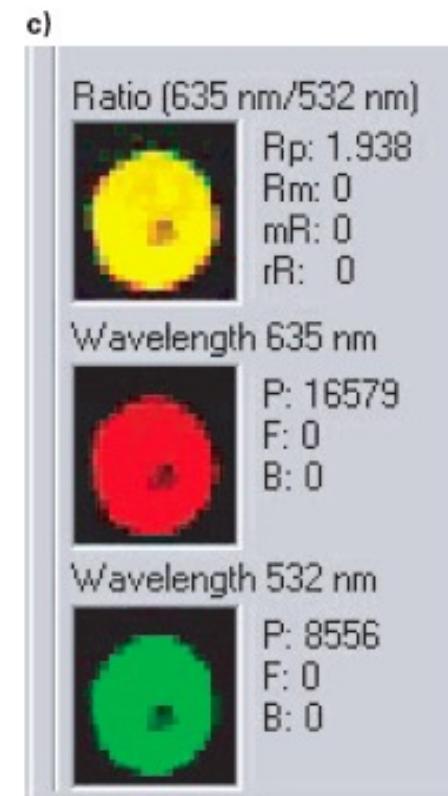
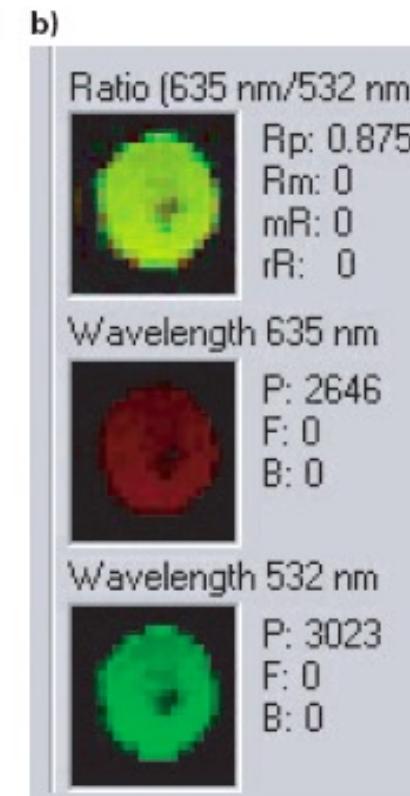
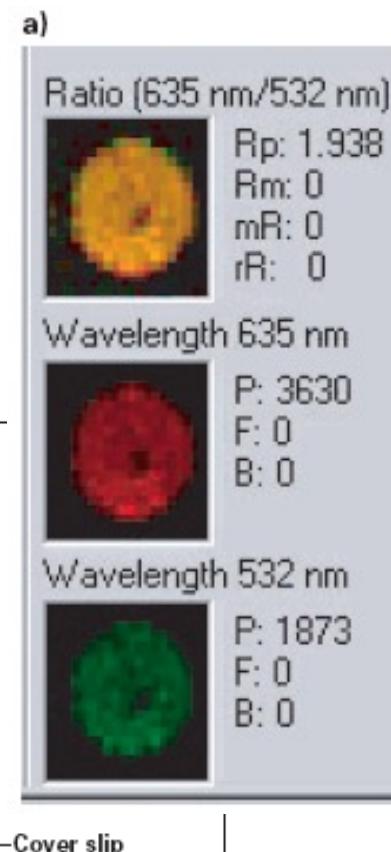
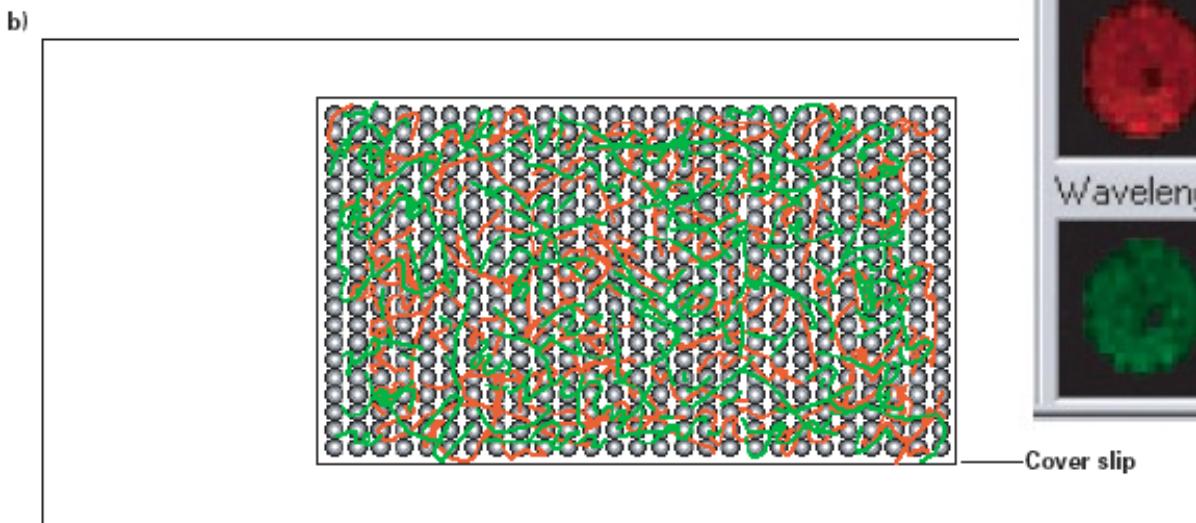
- cDNA microarrays
- “GeneChip” in situ synthesized oligonucleotide arrays
- Oligomer (~70mer) arrays



Experiments almost always a competitions between conditions or stages



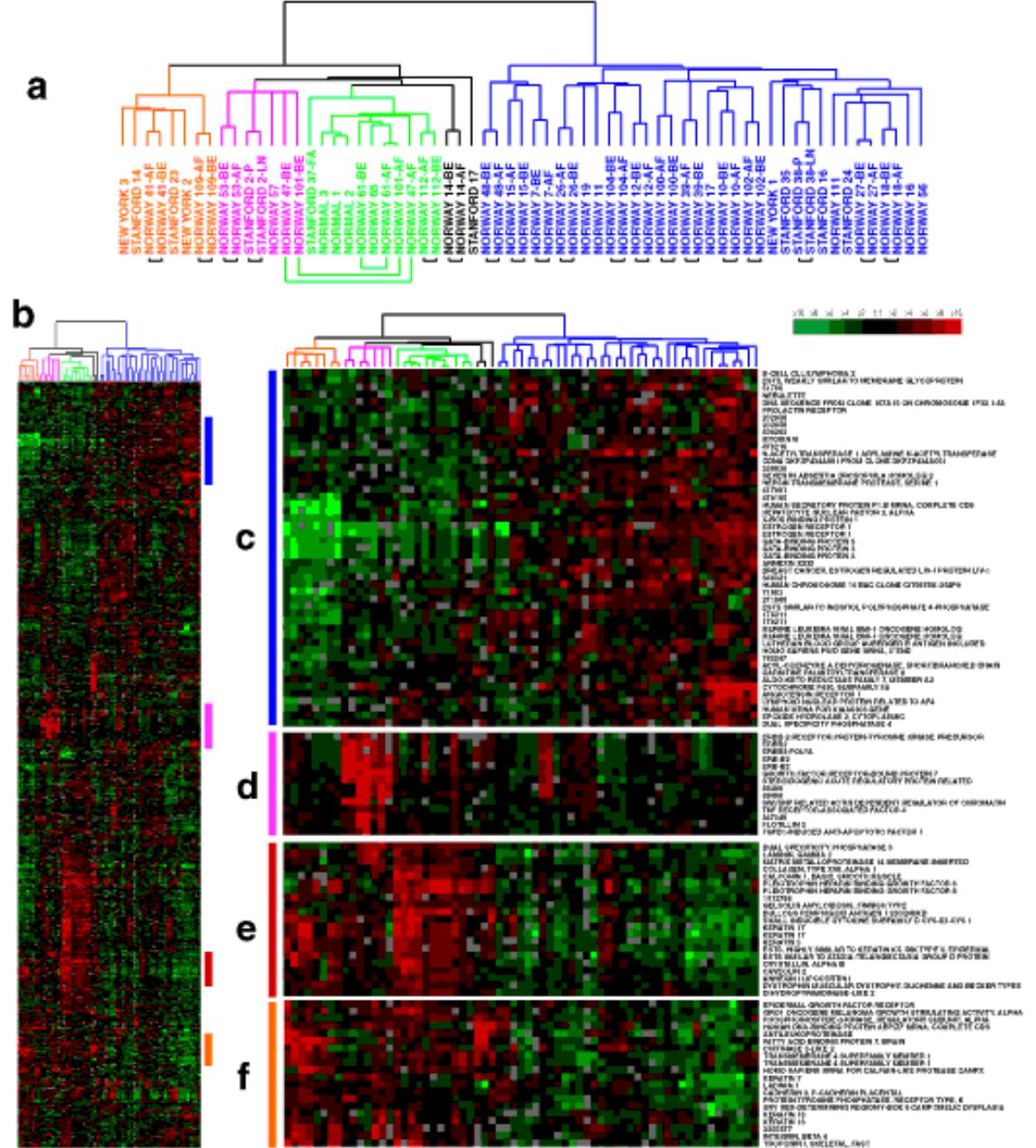
The RNA samples from the test and the control are labeled with different colors in a reverse-transcription reaction and then hybridized, together, competitively to a slide or chip containing gene sequences in multiple copies.



Ratios of experimental to control expression are often expressed as colors rather than numbers



**Clustered Microarray
Data:
Genes with similar
Expression profiles are
grouped together**



RNA sequencing



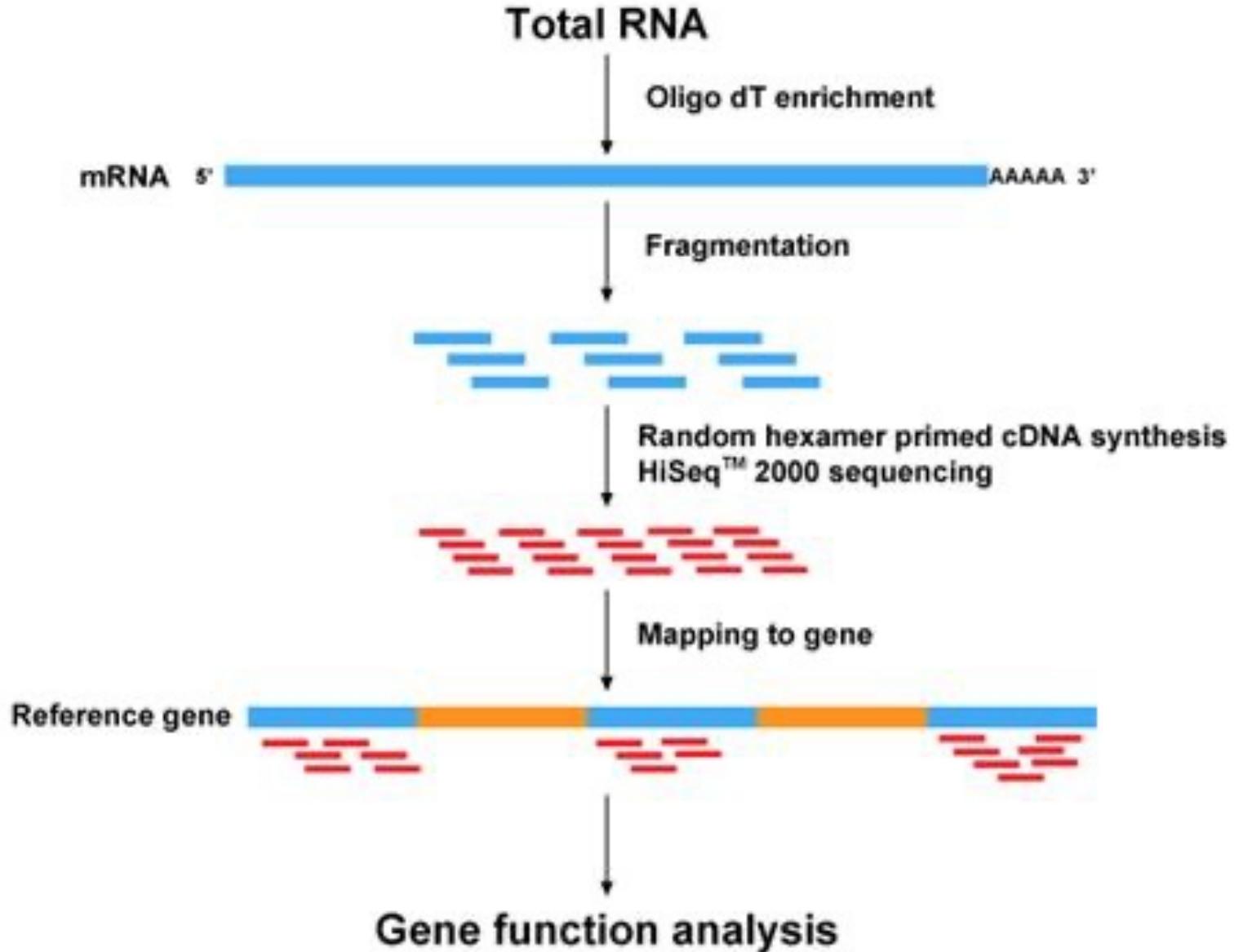
Illumina

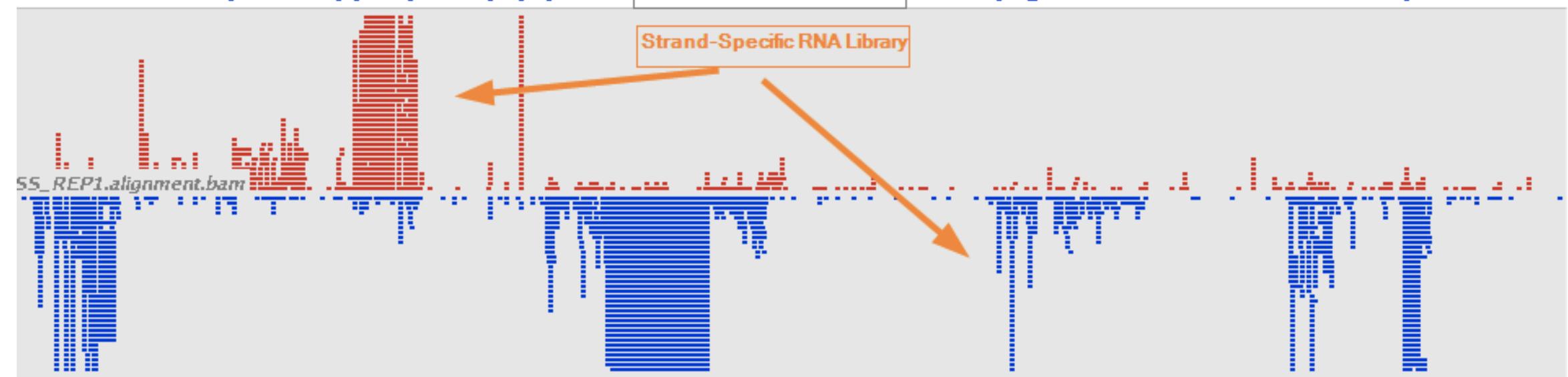
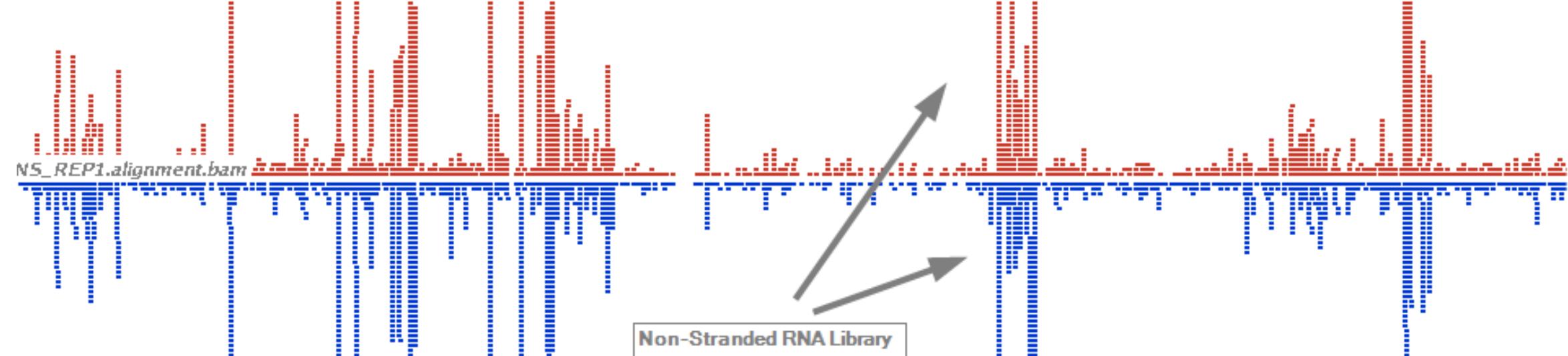
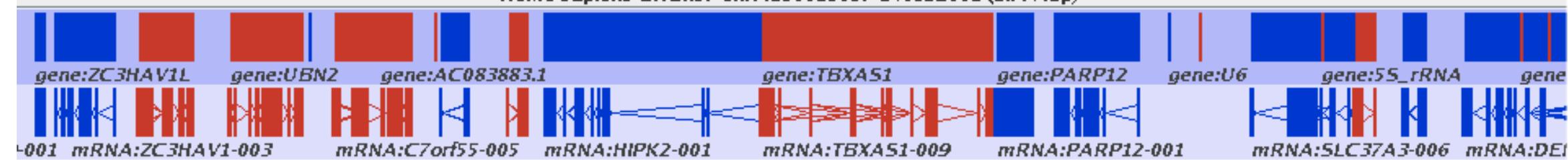


PacBio

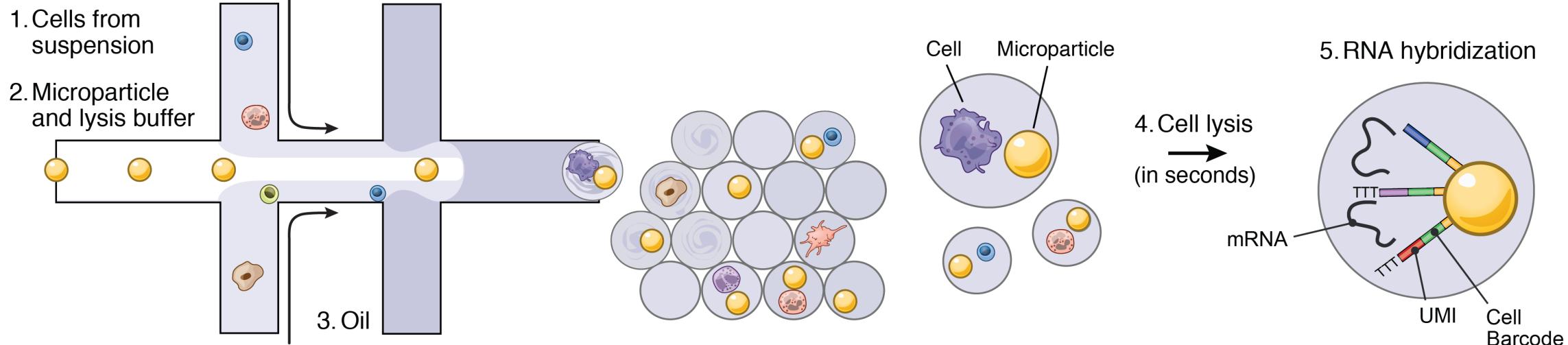


Nanopore

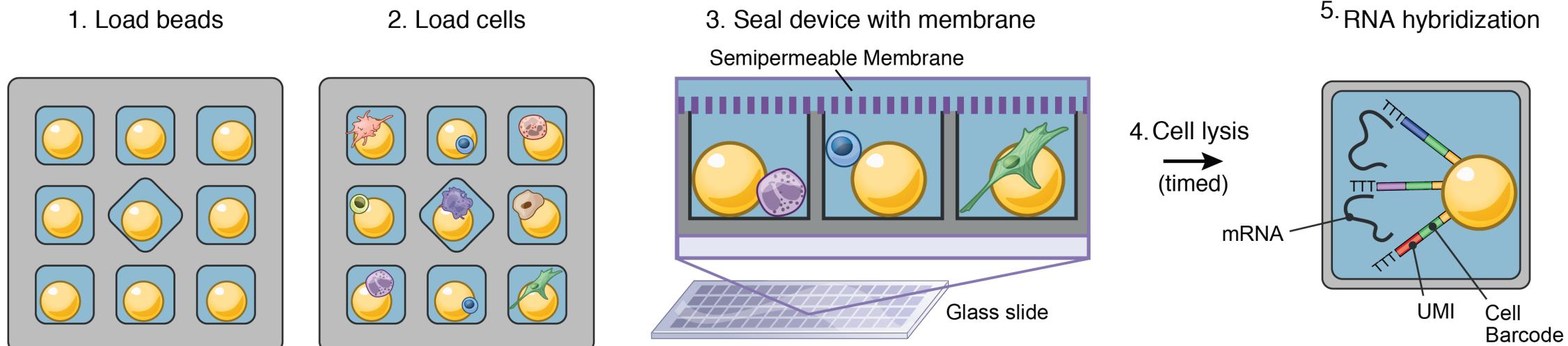




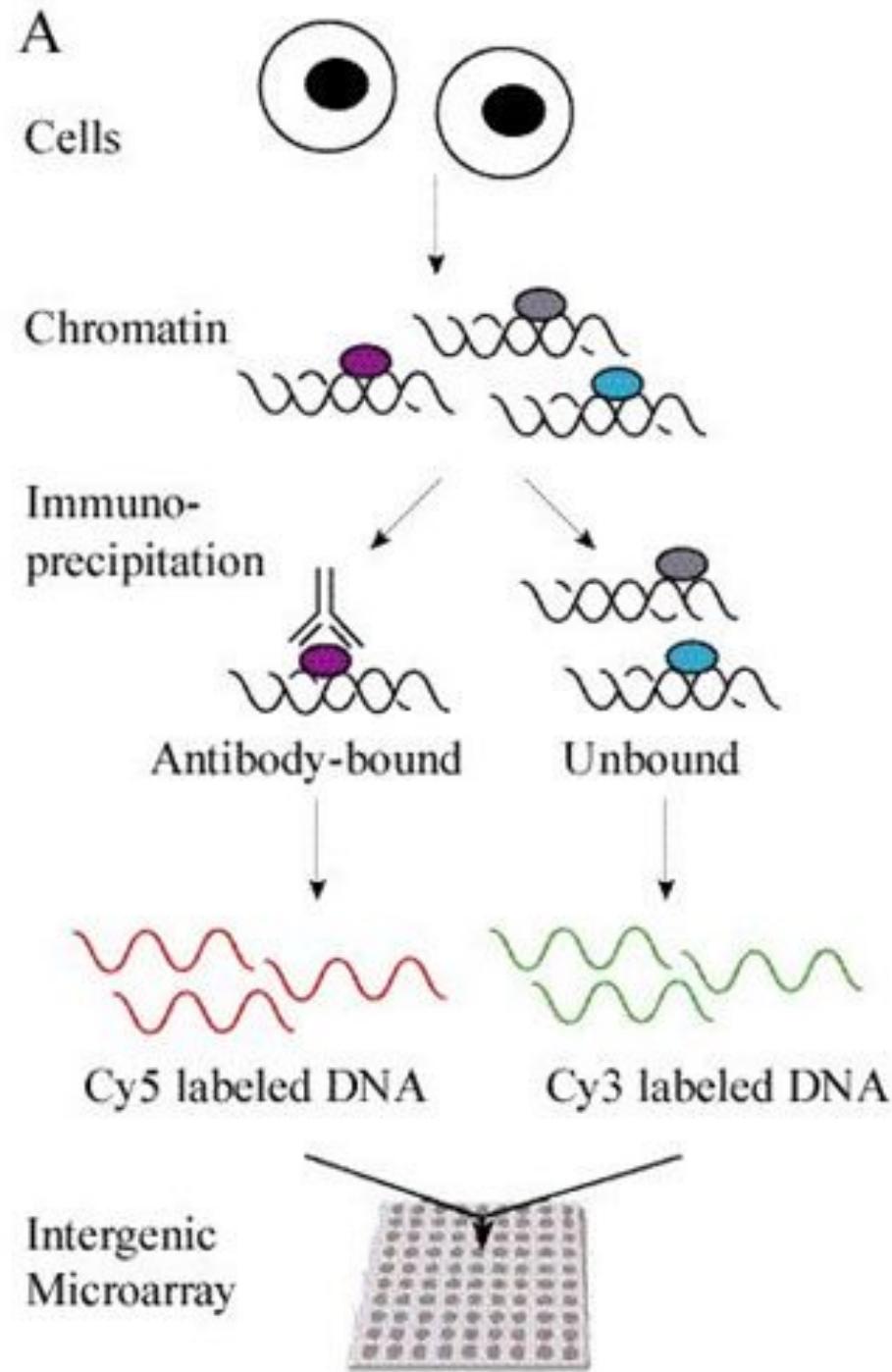
Drop-Seq



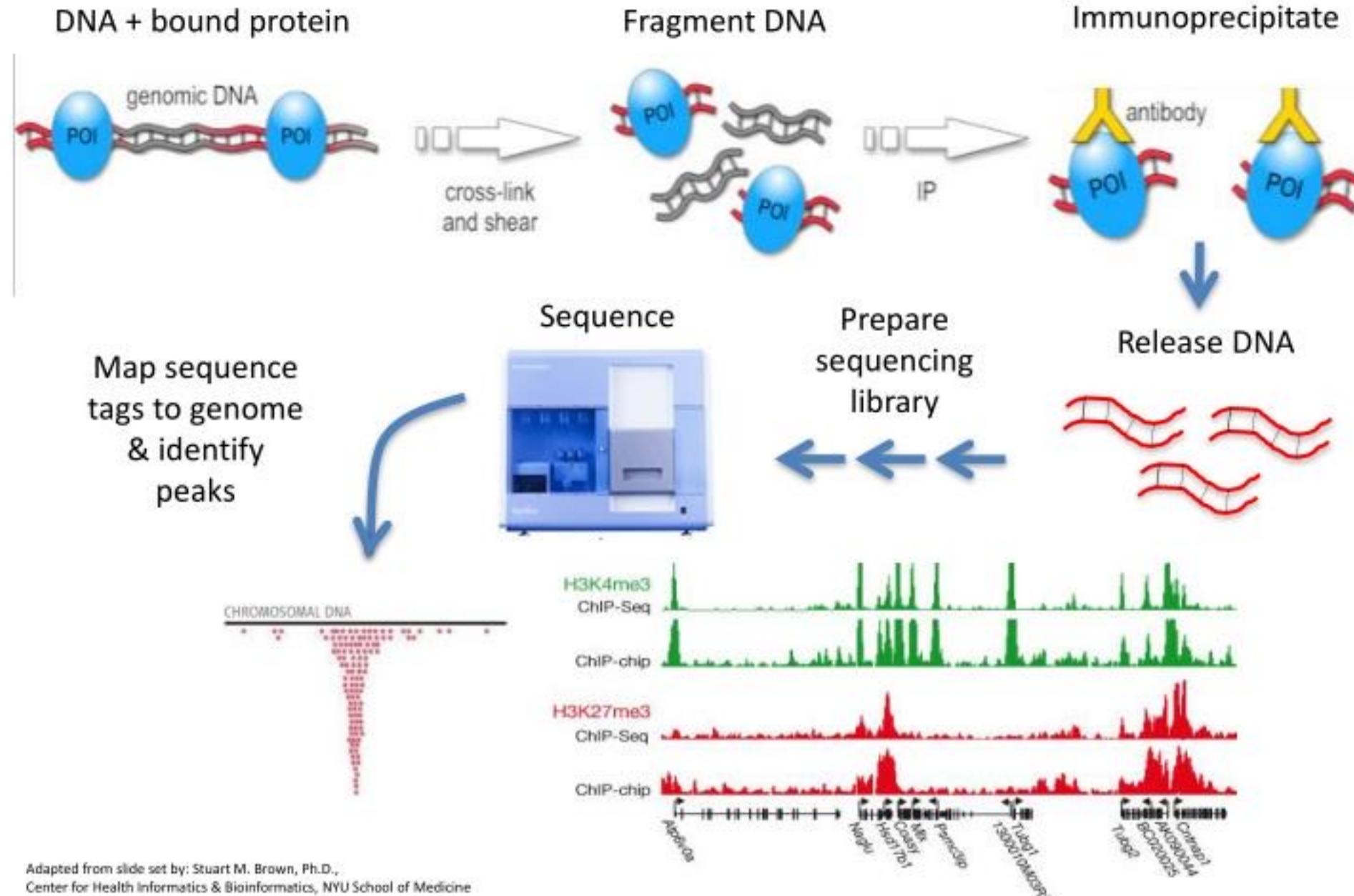
Seq-Well

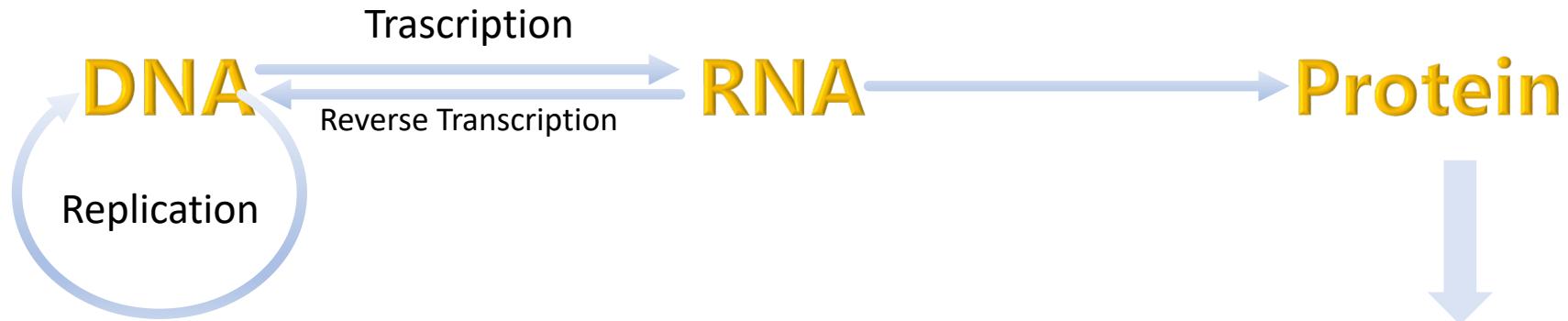


ChIP-chip

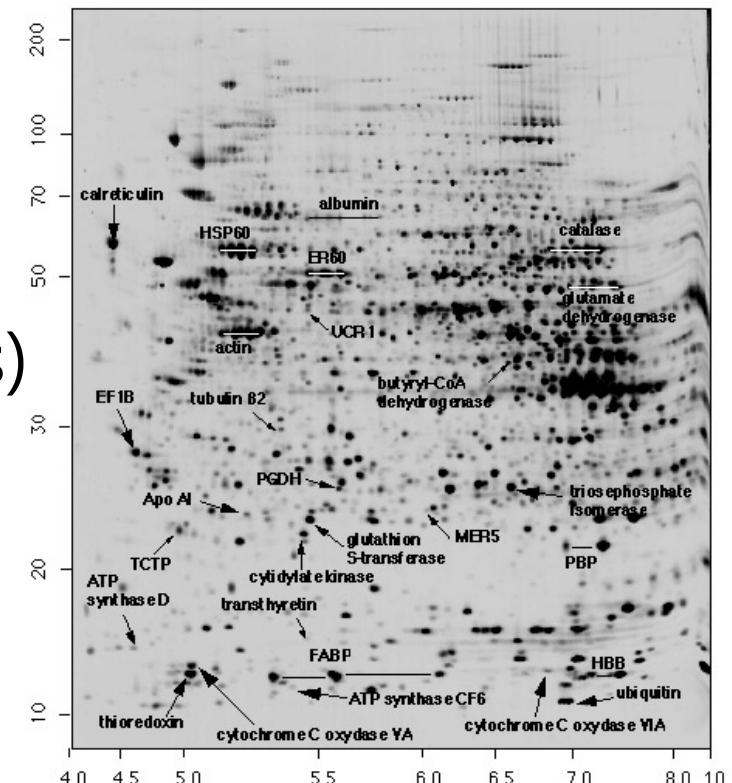


ChIP-seq overview





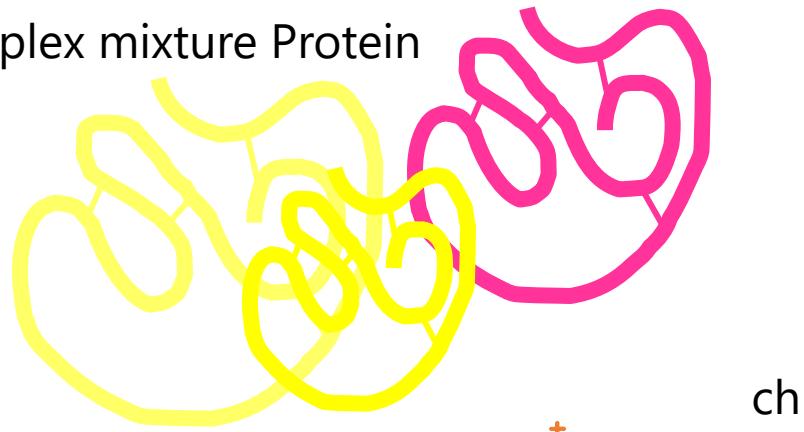
- Detect identity of peptides in a mixture.
- Quantify peptide levels (protein levels).
- Detect post-translational modifications (PTMs)



High throughput mass spectrometry

- Direct identification of proteins from biological sample.
- Capillary liquid chromatography apparatus (LC) coupled with...
- Electrospray tandem mass spectroscopy (MS/MS)
- “Sequest”, Mascot, or other software links mass spectra with genomic sequence database.

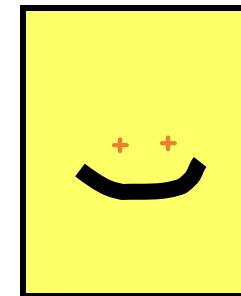
Complex mixture Protein



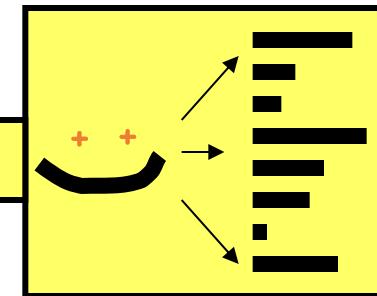
Tandem Mass Spec

Liquid chromatography

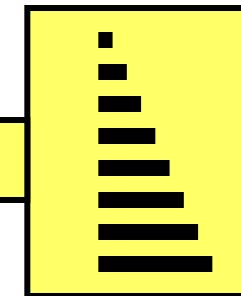
Collision Induced
Dissociation (CID)



Isolation

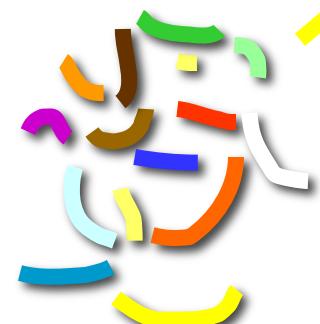


Fragmentation

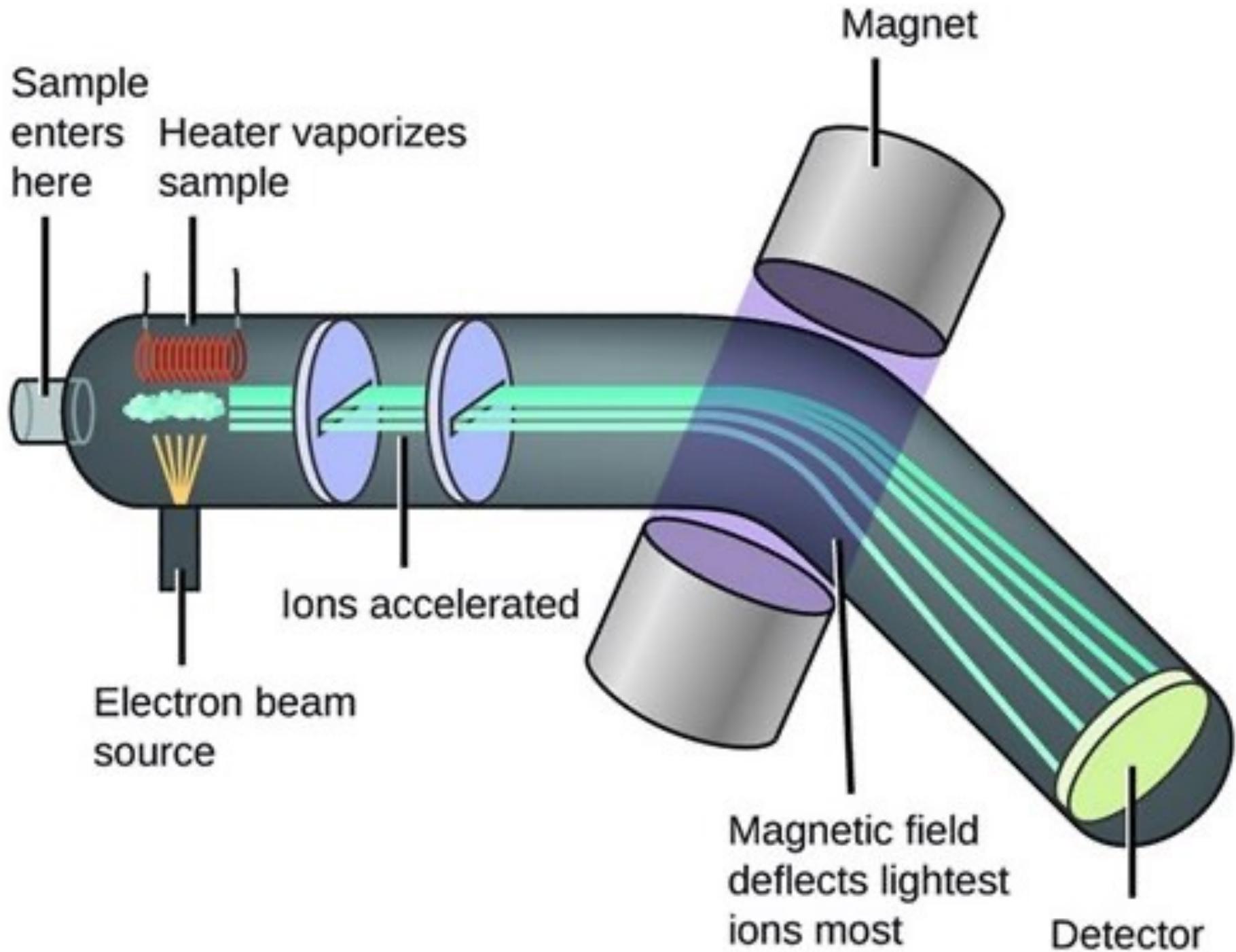


Measurement

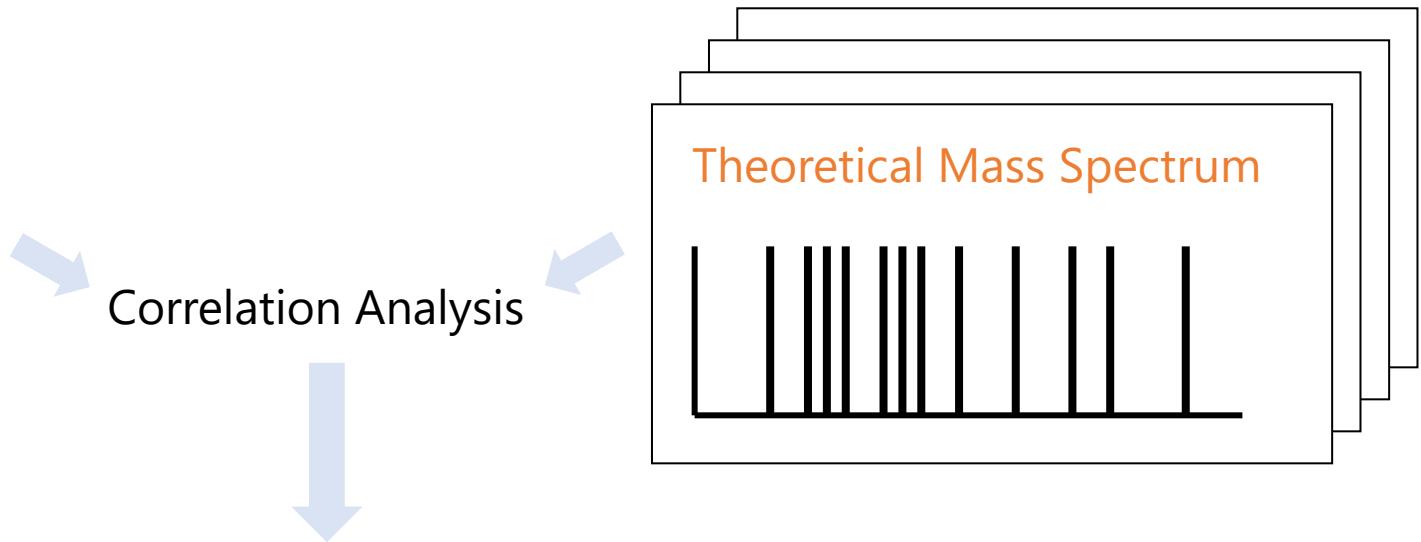
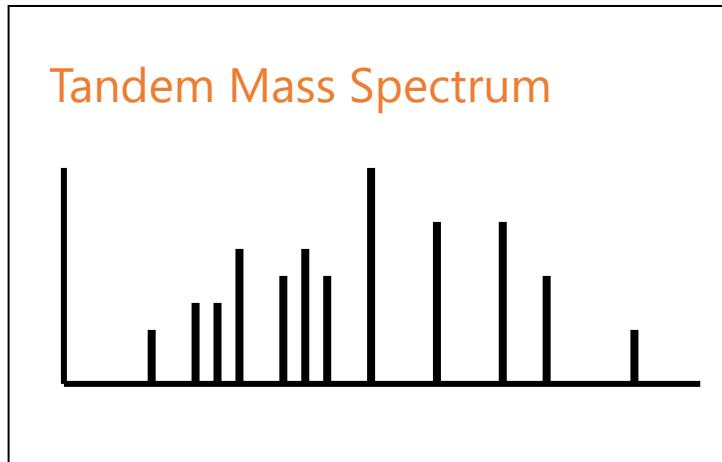
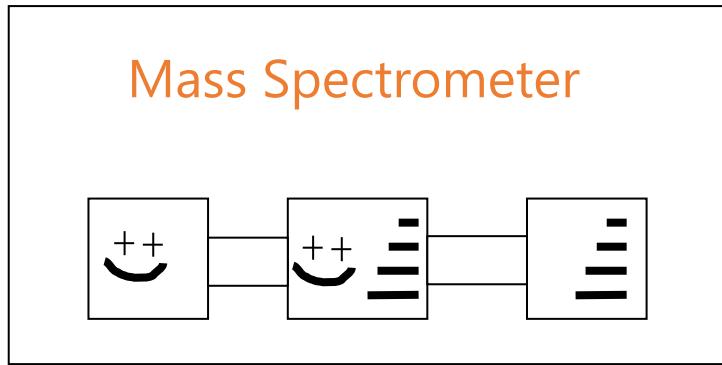
Ionized Peptides



Peptides

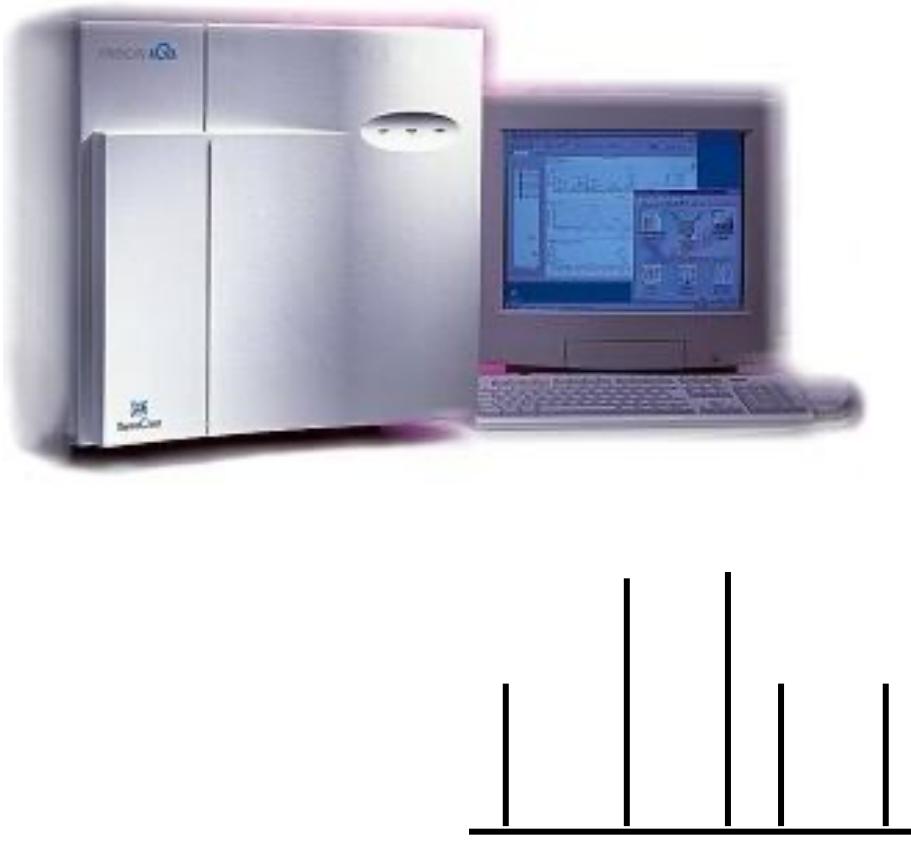


Database Search



Ranked Score of Matched Peptides

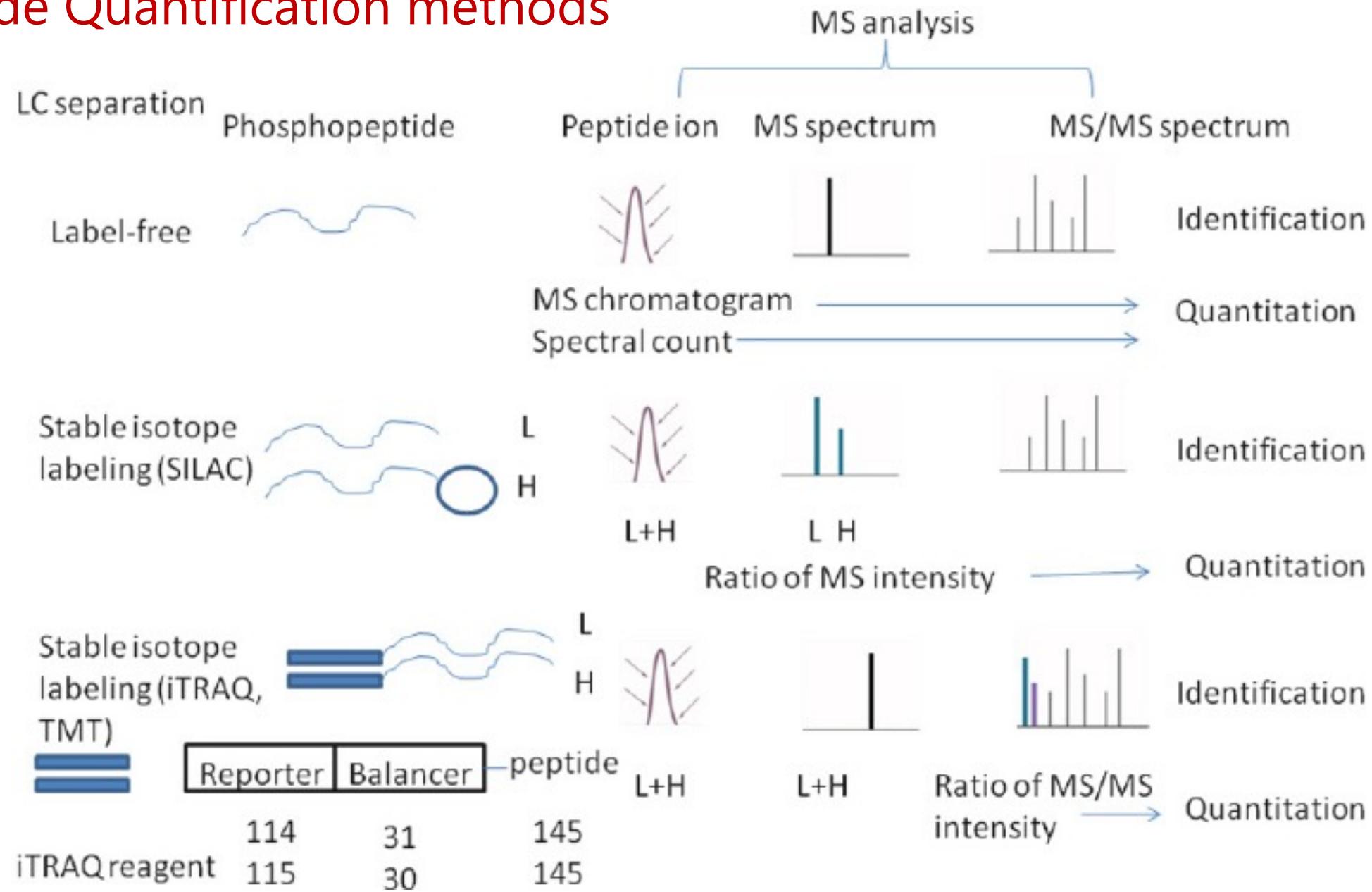
Peptide database



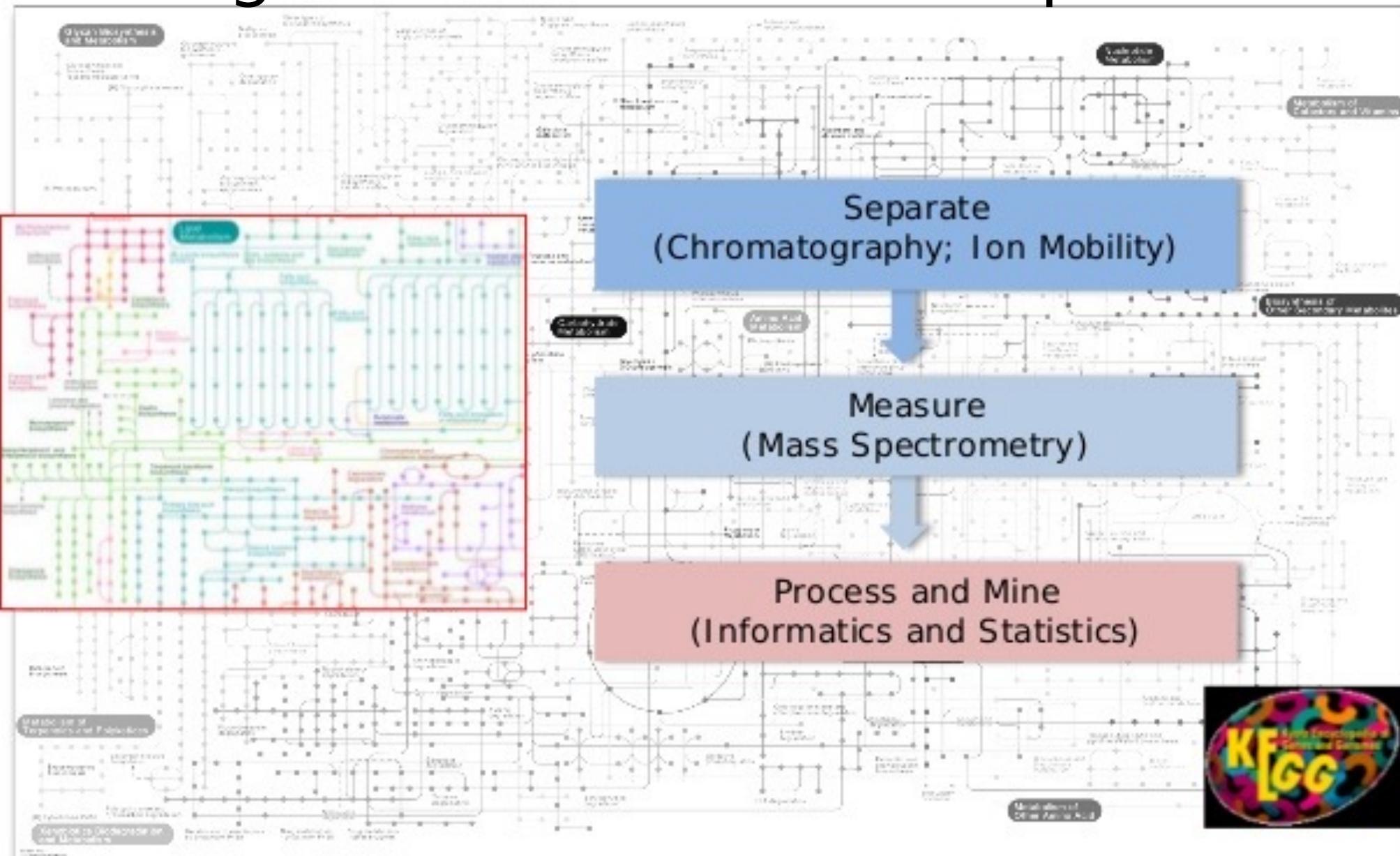
ENNPCKLQYDYNTNVTHGFGQEYPCETDIVERFSDEGAQCDDKKKIKDNSEGACAPYRRL
HVCVRNLENINDYSKINNKHNLLEVEVCLAALKYEGESITGRYPQHQETNPDTSQLCTVLA
RSFADIGDIIRGKDLYRGGNKEKKRKKLEENLKTIFGHYDELKNGKTNGEEELQKRY
RGDKDNDFYQLREDWWANRETVWKAITCNAGSYQYSQPTCGRGEIPYVTLSKCQCIAGE
VPTYFDYVPQYLRWFEEWAEDFCRKKKKIPNVKTNCRQVQRGKEKYCDRDGYNCDEGTIR
KQYIYRLDTDCTKCSLACKTFAEWIDNQKEQFDKQKQKYQNEISGGGRRQKRSTHSTKE
YEGYEKFNEELRNEGKDVRSFLQLLSKEKICKERIQVGEETANYGNFENESNTFSHTEY
CDRCPLCGVDCSSNDNCRKKPDKSCDEQITDKEYPPENTTKIPKLTAEKRKTGILKKYEKF
CKNSDGNGGQIKKWECHYEKNDKDDGNGDINNNCIQGDWKTSKNVYYPIYYSSFYGSII
DMLNESIEWRERLKSCINDAKLGCRKGCKNPCECYKRWVEKKDEWDKIKEFFRKQKDL
LKDIAGMDAGELLEFYLENIFLEDMKNANGDPKVIEKFKEILGKENEVQDPLKTTKTD
DFLEKELNEAKNCVEKNPDNECPKQKAPGDGAAPSDDPREDITHDGEHSSDEEEEEEE
EEQQPPAEGTEQGEEKSESKEVVEQQETPQKDTEKTVPTTPTVDVCDTVKTLADTGSL
NAACSLKYVTGNNNNNIAPIASGTTSGKDQCVPPRTCLYYLKLSDTTQKLLAE
FIKTAQAETYDQGKQDQGKQDQGKQDQGKQDQGKQDQGKQDQGKQDQGKQDQGKQD
LFLGGRYIGNDLDRVNNNNITAVTQDQGKQDQGKQDQGKQDQGKQDQGKQDQGKQD
GGKKTLTETYNNSNUVQGHTGTKLNEFASRPSFLRWMTIEWGDQFCRERITQLQILKER
CMVLTQINGDKGKDDKKEKCTEACTYYKEWLTNWQDNYKKQNQRYTEVKGTSPYKEDSDVK
ESKYAHGYLRKILKNIICTSGTDIAY**CNCMEGTST**TDSSNNNDNIPESLKYPPIEIEEGCT
CKDPSPGEVIPEKKVPEPKVLPKPKLPKRQPKERDFPTPALKNAMLSSTIMWSIGGFA
TFTYFYLKKTKSTIDLRLVINIPKSDYDIPPTKLSPNRYIPTYSGKYRGKRYIYLEGDSG
TDSGYTDHYSDITSSSESEYEELDINDIYAPRAPKYKTLIEVVLPSGNNTTASGNNTPS
DTQNDIQNDGIPSSKITDNEWNTLKDEFISQYLQSEQPNDVNDYSSGDIPLNTQPNTLY
FDNPDEKPFITSIHDRDLYSGEEYSYNVMVNTNNIDPISGKNGTYSGIDLINDSLSNN

Note: Typically an electronic database of translated genes is used but one could use translated ORFs

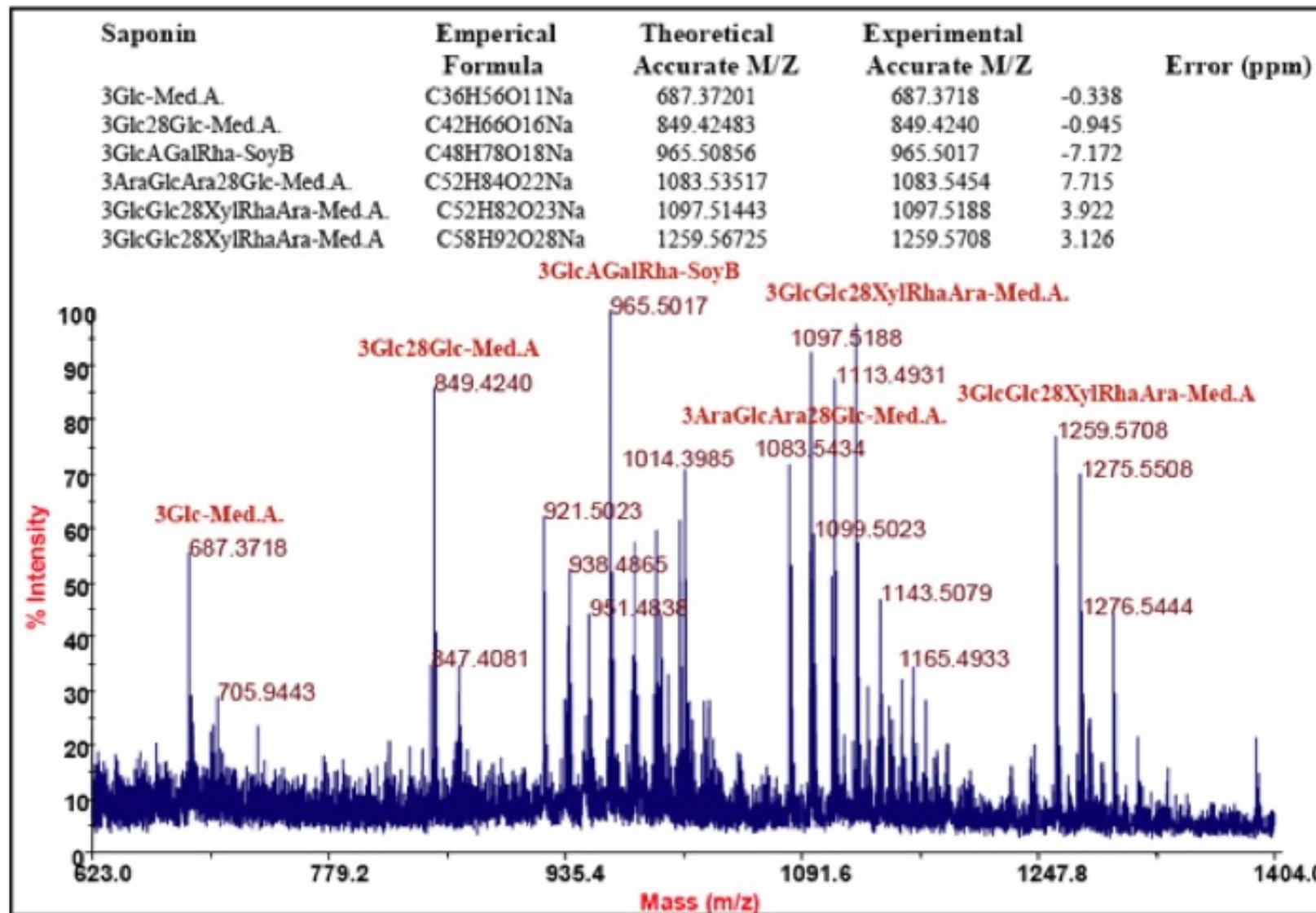
Peptide Quantification methods



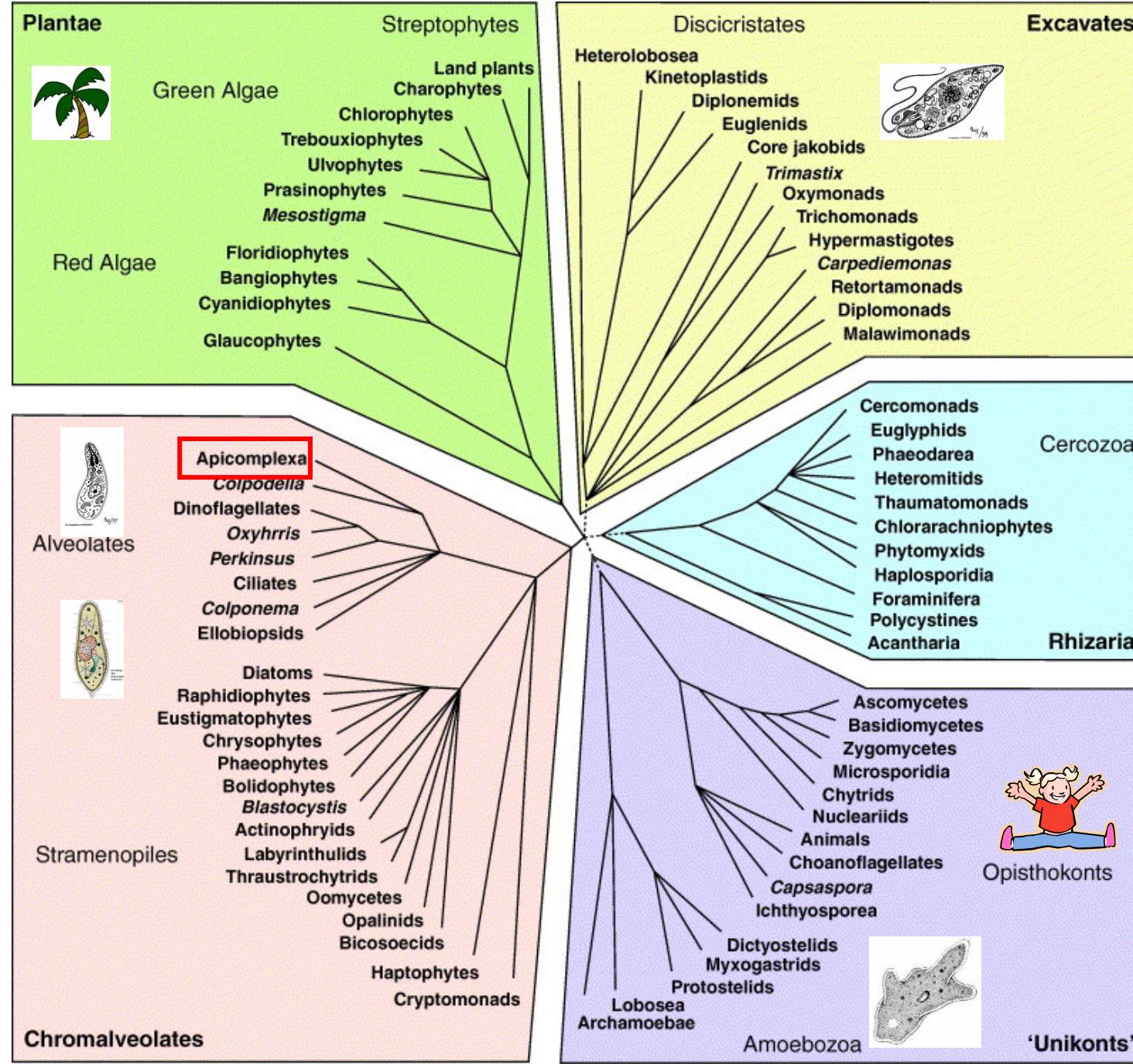
Untargeted metabolomics & lipidomics



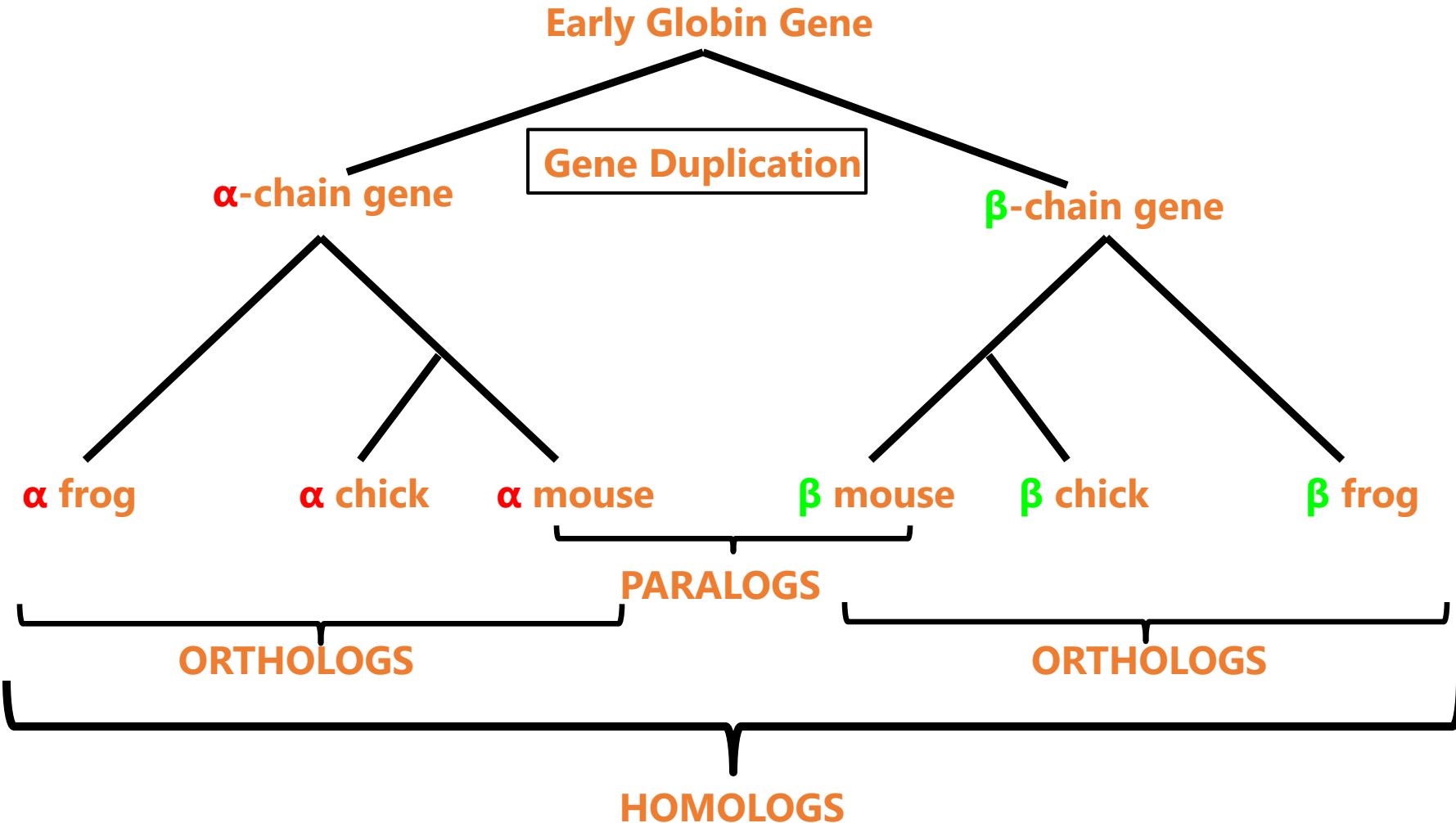
Complex mixtures can be analyzed and interpreted



Phylogeny



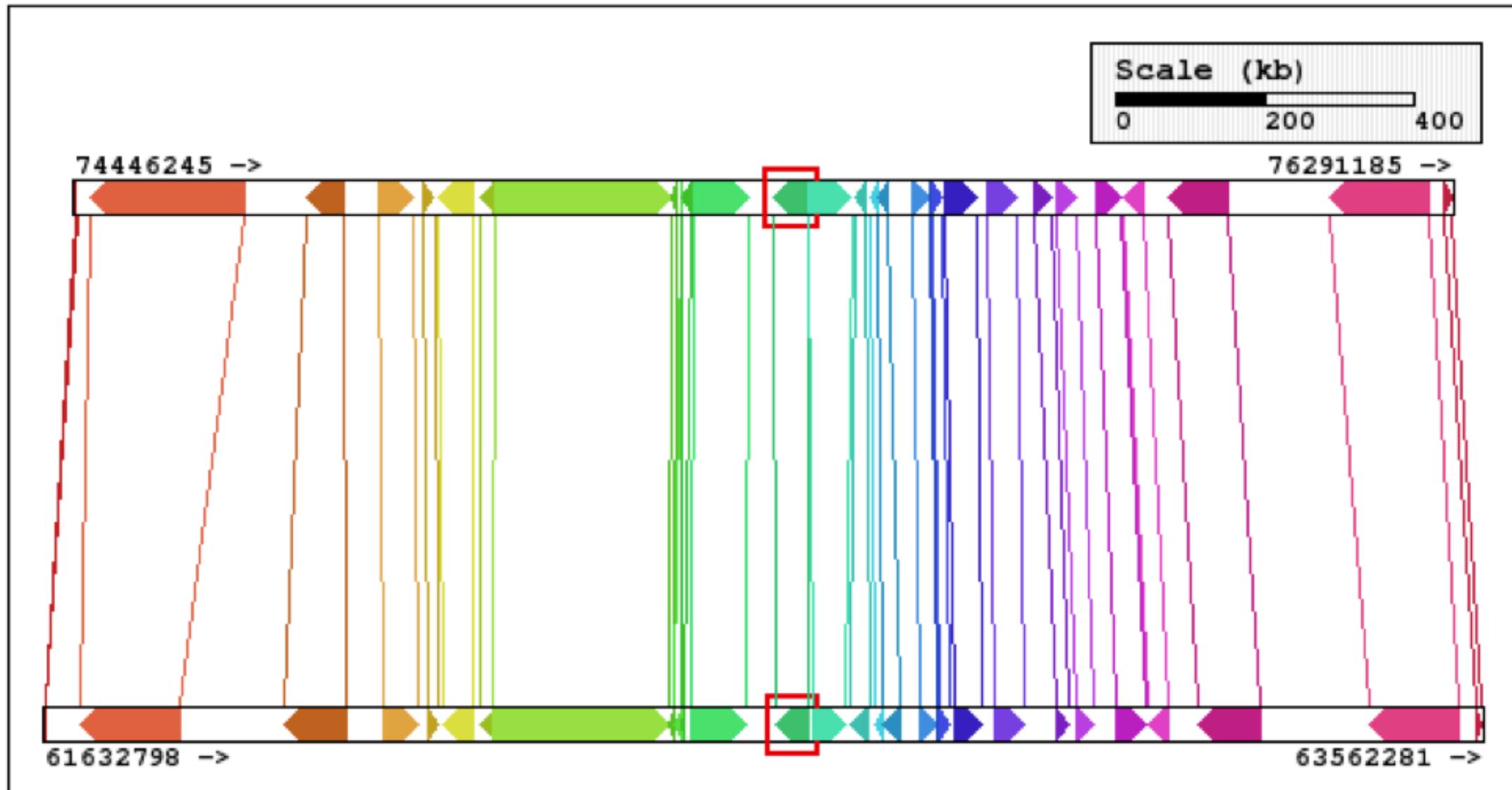
Homology



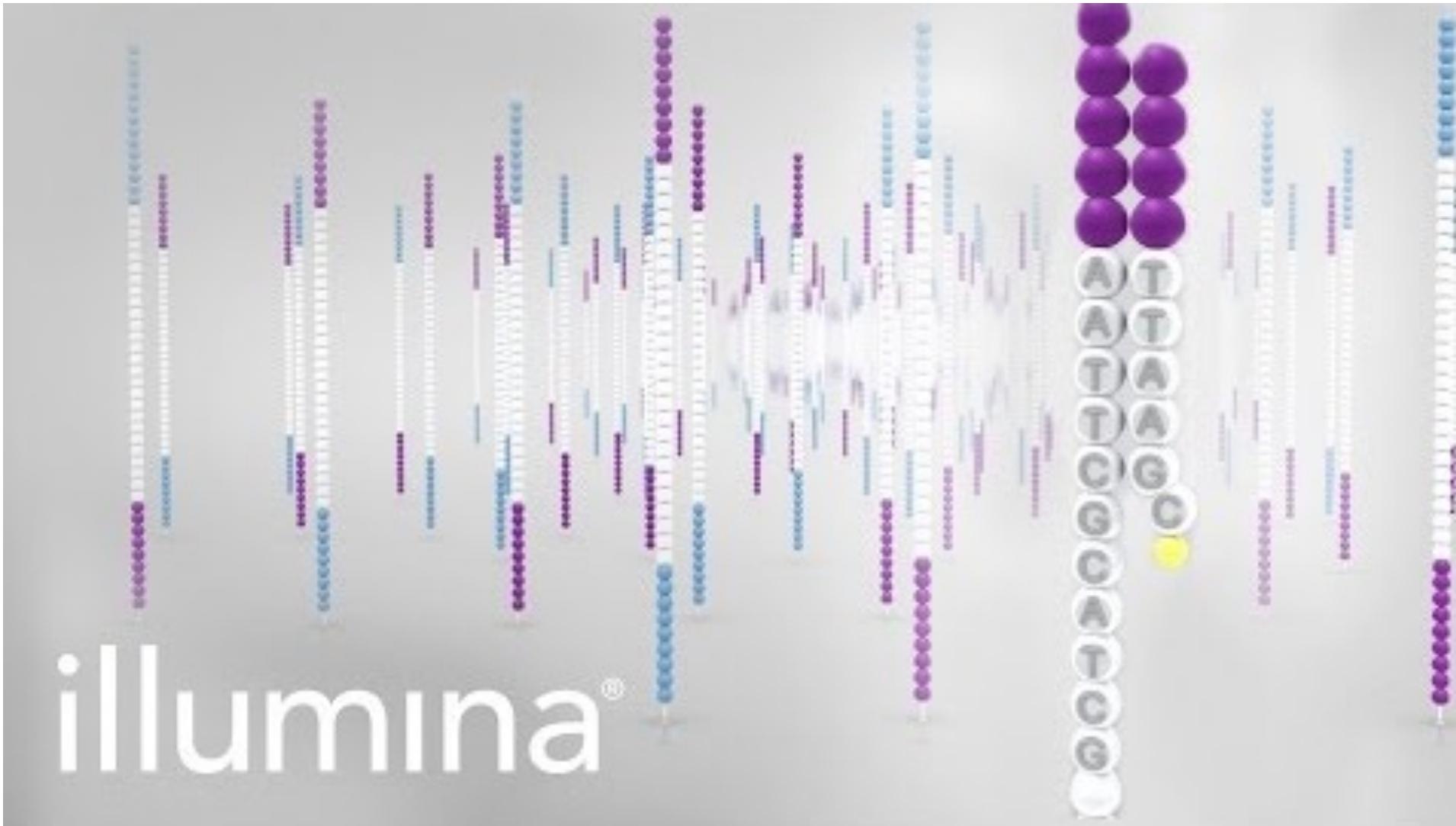
Evolutionary relationships

- Homology - related by evolutionary descent not equivalent to similarity
- Orthology - same gene in different organisms, e.g. alpha hemoglobin in humans and chimps
- Paralogy - genes within an organism related by gene duplication, e.g. alpha and beta hemoglobin in humans

Synteny (co-linearity) = large regions of chromosomes containing the same genes



Short Read Sequencing



<https://youtu.be/fCd6B5HRaZ8>

Long read sequencing

