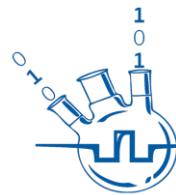




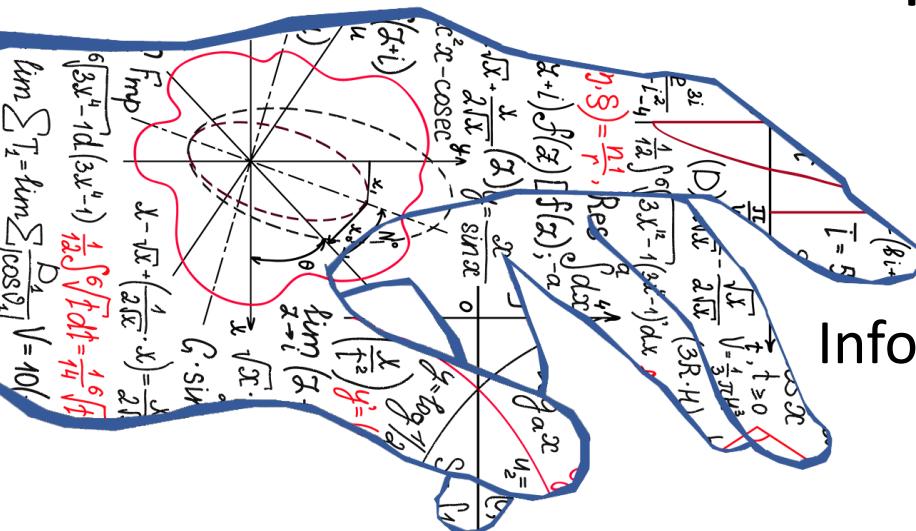
ITMO UNIVERSITY



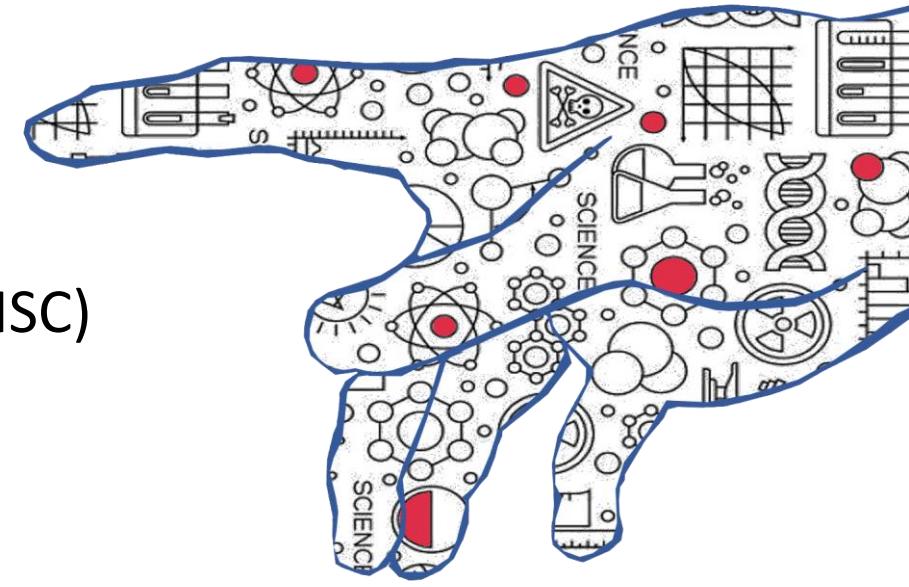
INFOCHEMISTRY SCIENTIFIC CENTER

Cheminformatics and synthetic biology: synthetic data analysis and design

Prof. Sergey Shityakov

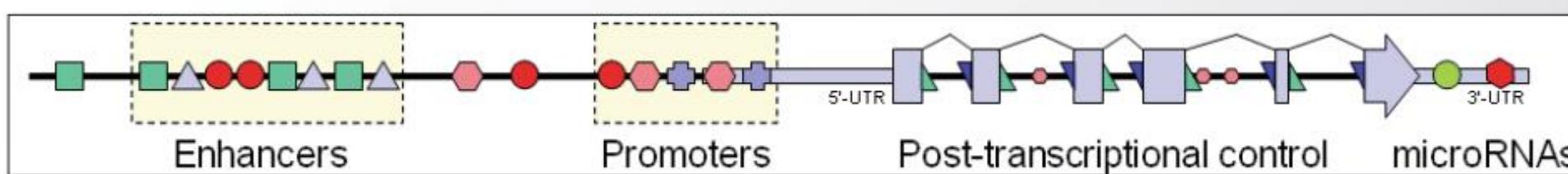
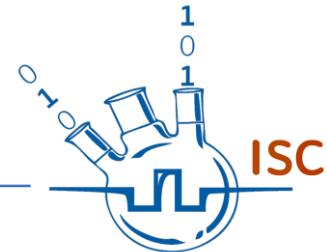


Infochemistry Scientific Center (ISC)
ITMO University
Saint-Petersburg, 2024





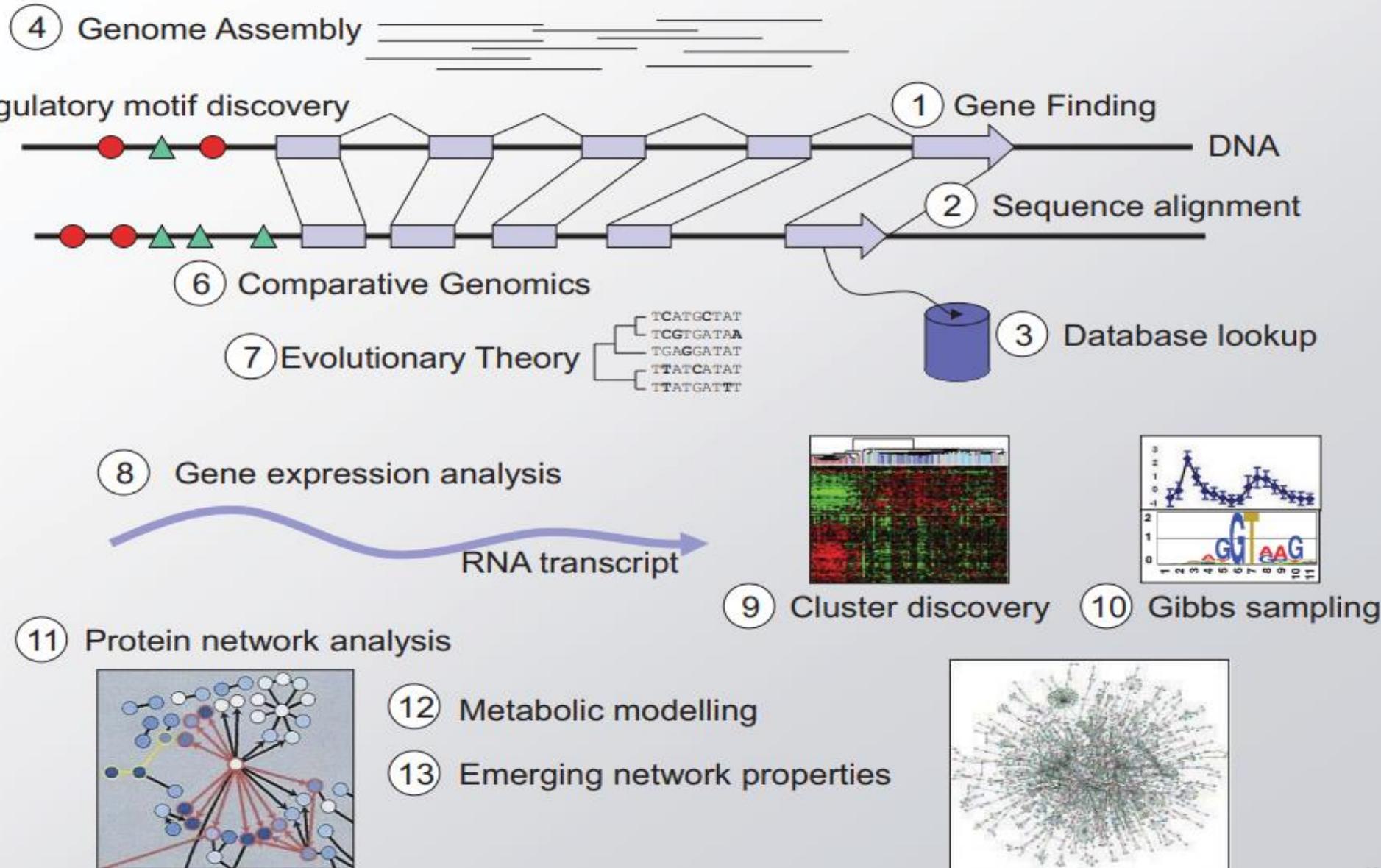
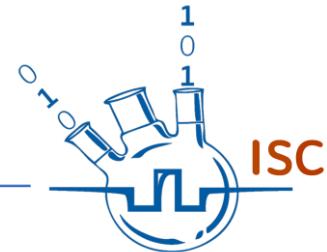
Components of genes and genomes



Goal: A systems-level understanding of genomes and gene regulation:

- The genome: Map reads, align genes/genomes, assembly strategies
- The genes: Protein-coding exons, introns, non-coding RNA, RNA folding
- The control regions: Promoters, enhancers, insulators, chromatin states
- The actual words: Regulatory motifs, high-resolution accessibility maps
- The regulators: Transcription factors, chromatin modifiers, nucleosomes
- The dynamics: Changing maps between cell types, across development
- The networks: regulator → enhancer → target, ChIP-seq, correlated activity
- The grammars: TF/motif/mark combinations, predictive models
- Human variation: Human diversity, population genomics, linkage maps
- Evolution: Phylogenetics, phylogenomics, coalescent, human ancestry
- GWAS/QTLs: Genome variation ↔ organismal/molecular phenotypes

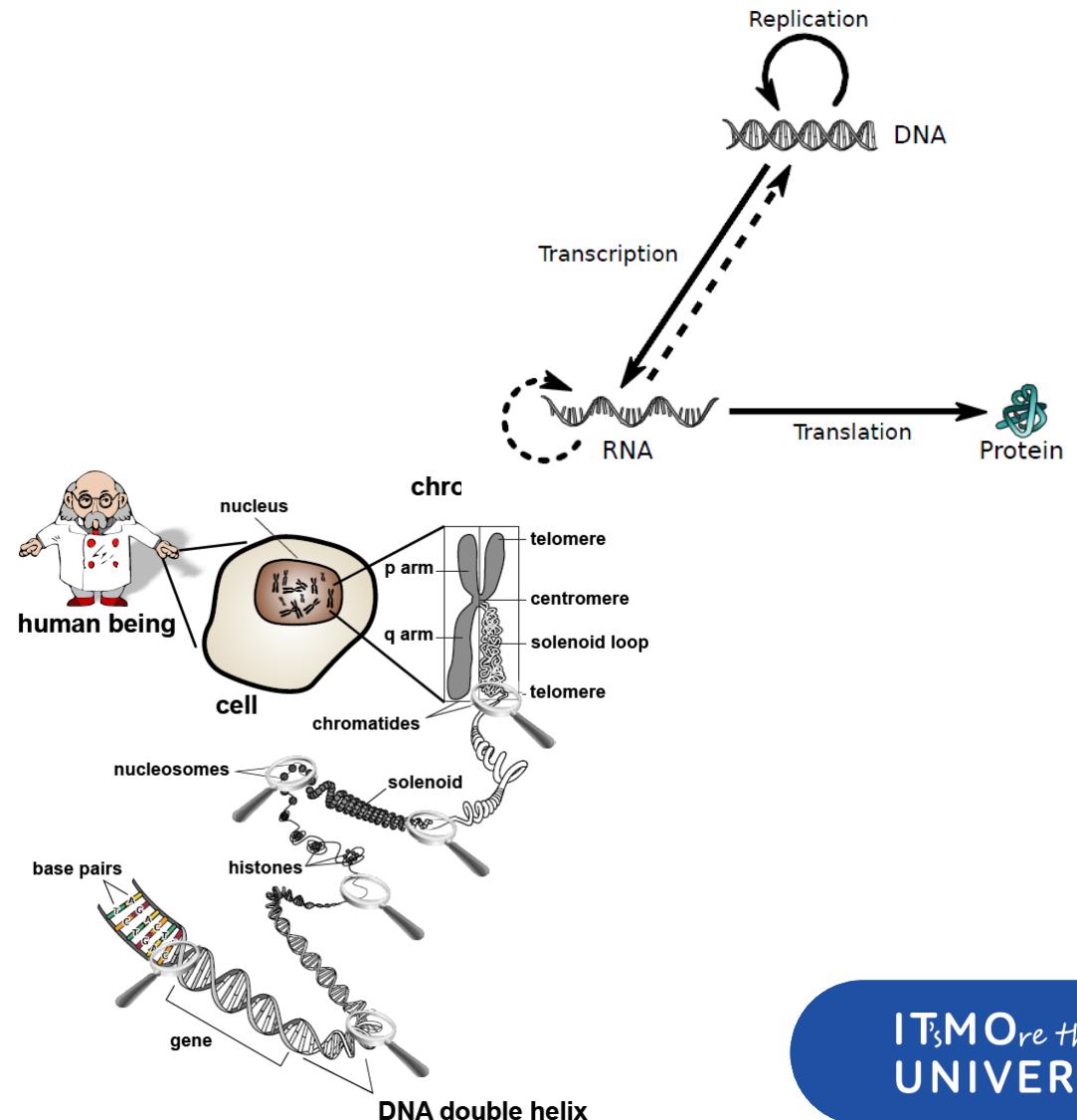
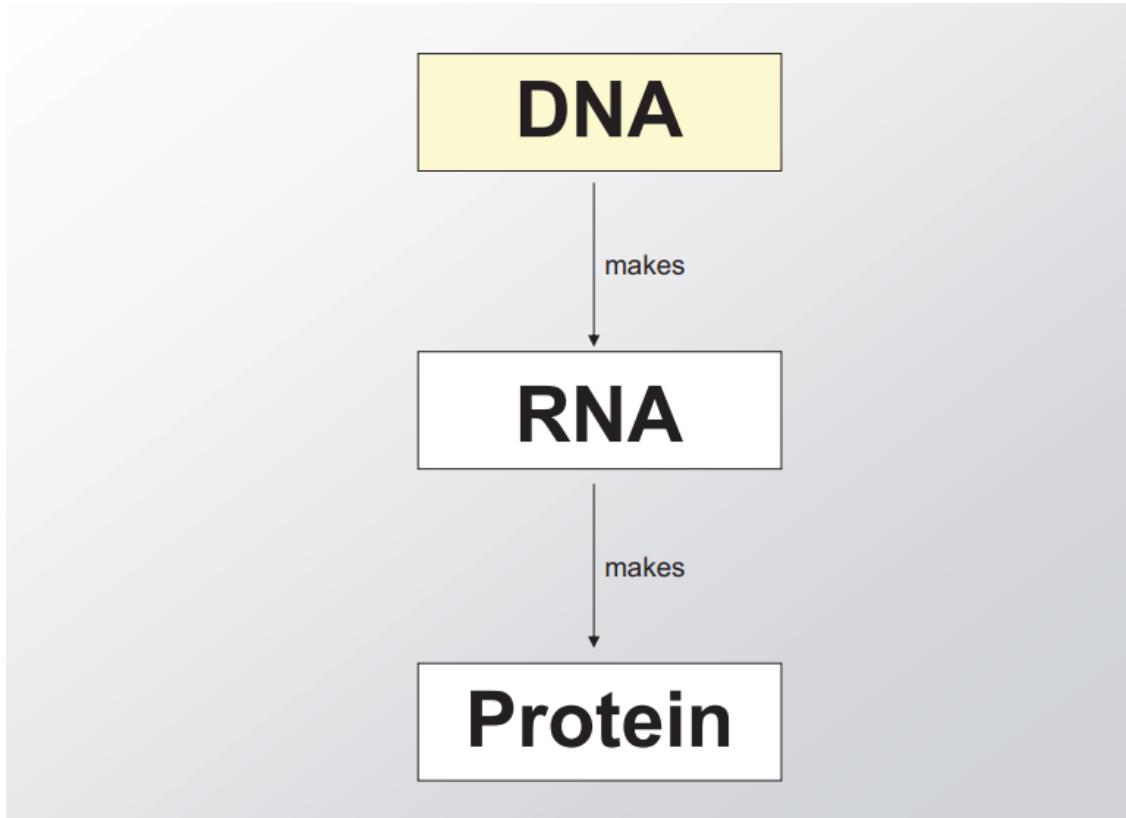
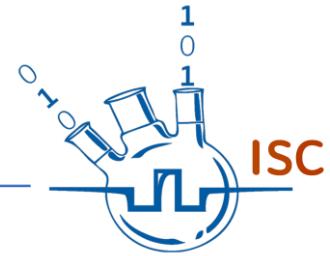
Challenges in Synthetic Biology



ITMO more than a
UNIVERSITY



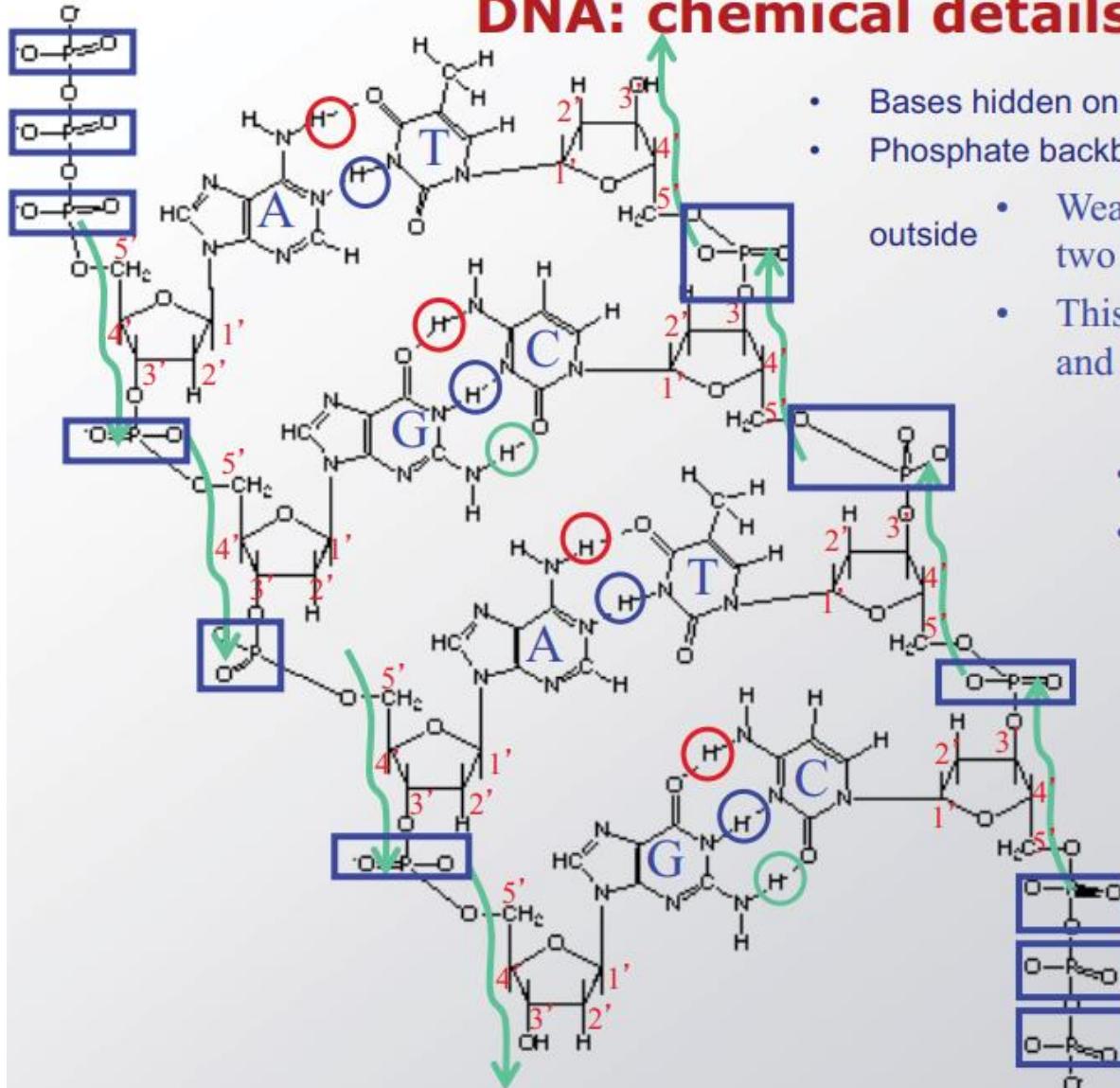
Central dogma in synthetic biology



IT'S MORE than a
UNIVERSITY



DNA: chemical details



- Bases hidden on the inside
- Phosphate backbone outside
- Weak hydrogen bonds hold the two strands together
- This allows low-energy opening and re-closing of two strands
- Anti-parallel strands
- Extension $5' \rightarrow 3'$ triphosphate coming from newly added nucleotide

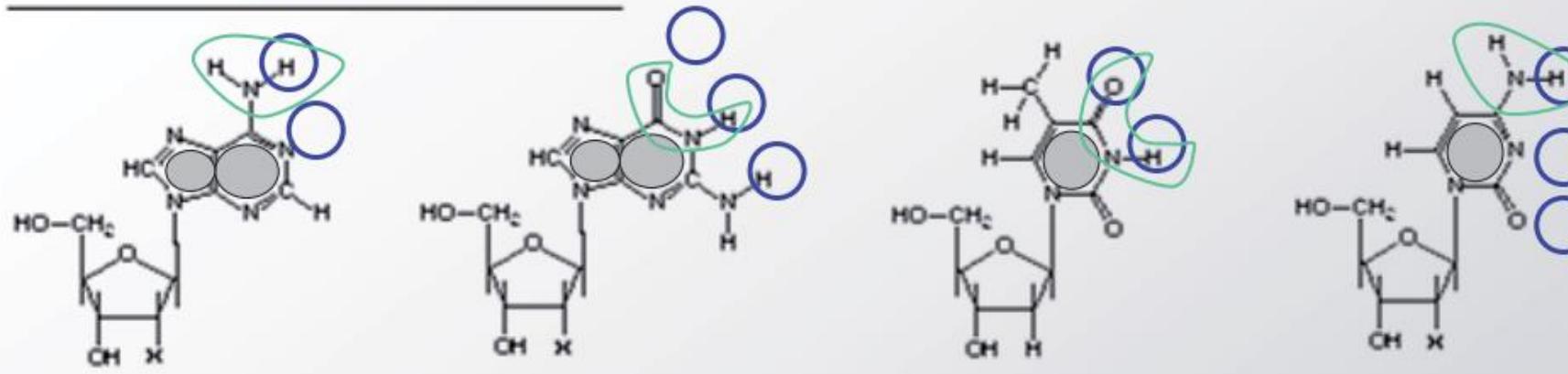
The only pairings are:

- A with T
- C with G



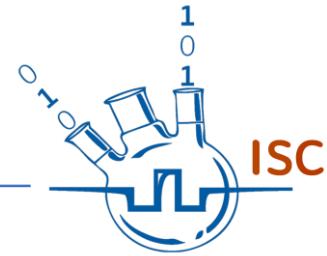
The Nucleotides of DNA

DNA: the four bases



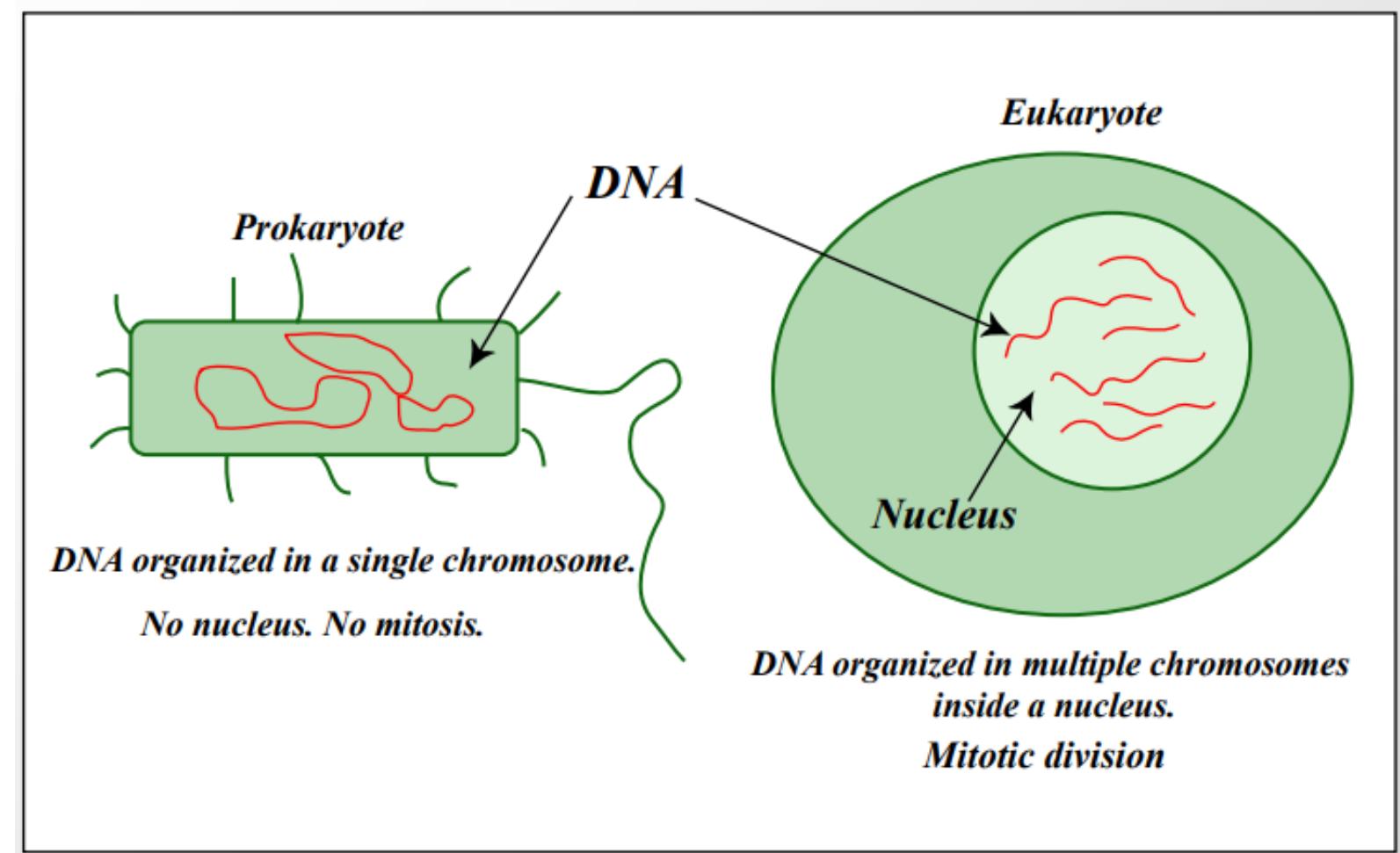
| Adenine | Guanosine | Thymine | Cytosine |
|---------|-----------|------------|------------|
| Purine | Purine | | |
| | | Pyrimidine | Pyrimidine |
| Weak | | Weak | |
| | Strong | | Strong |
| Amino | | | Amino |
| | Keto | Keto | |

Chromosomes inside the cell

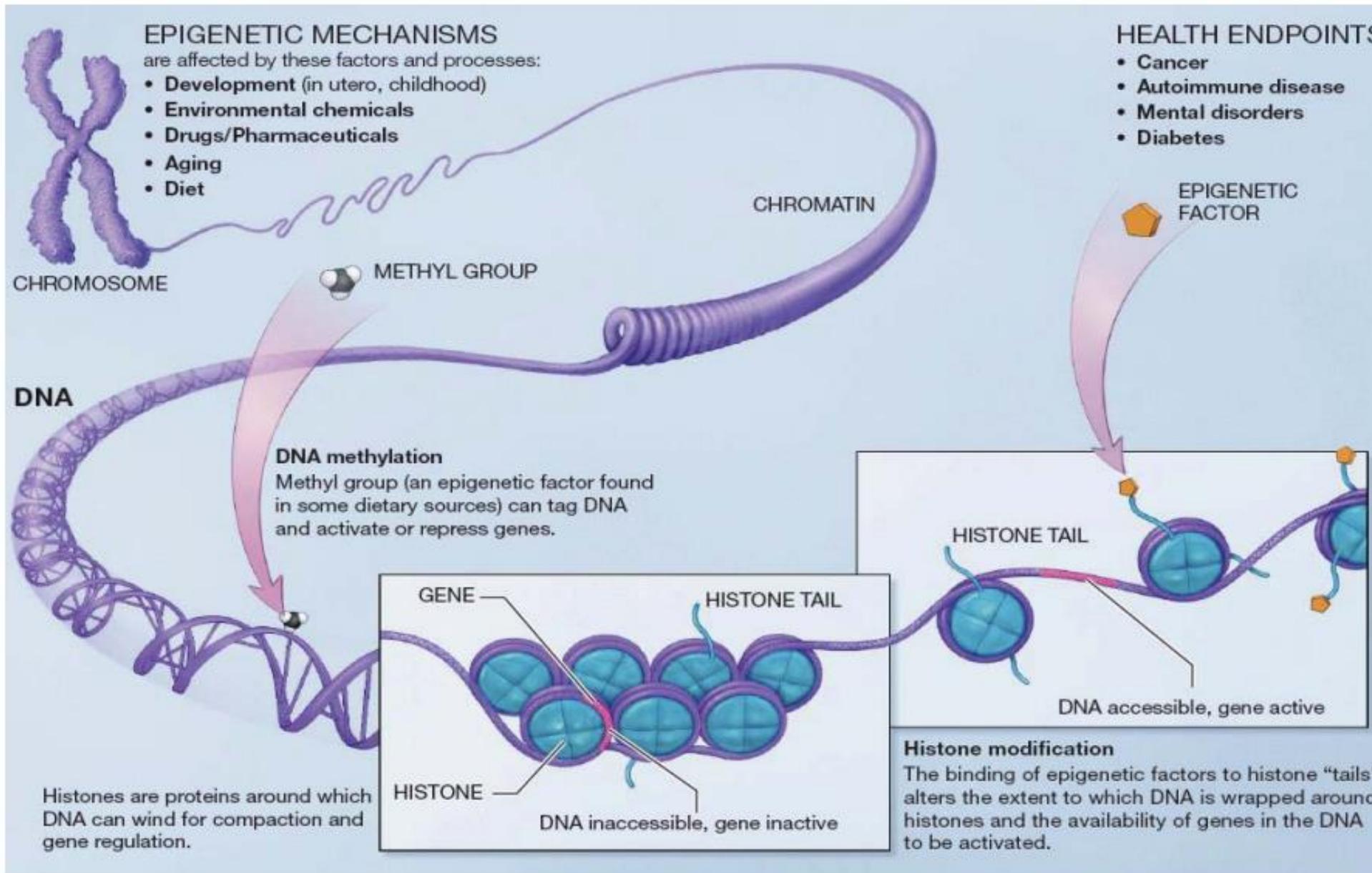


DNA packaging

- Why packaging
 - DNA is very long
 - Cell is very small
- Compression
 - Chromosome is 50,000 times shorter than extended DNA
- Using the DNA
 - Before a piece of DNA is used for anything, this compact structure must open locally
- Now emerging:
 - Role of accessibility
 - State in chromatin itself
 - Role of 3D interactions



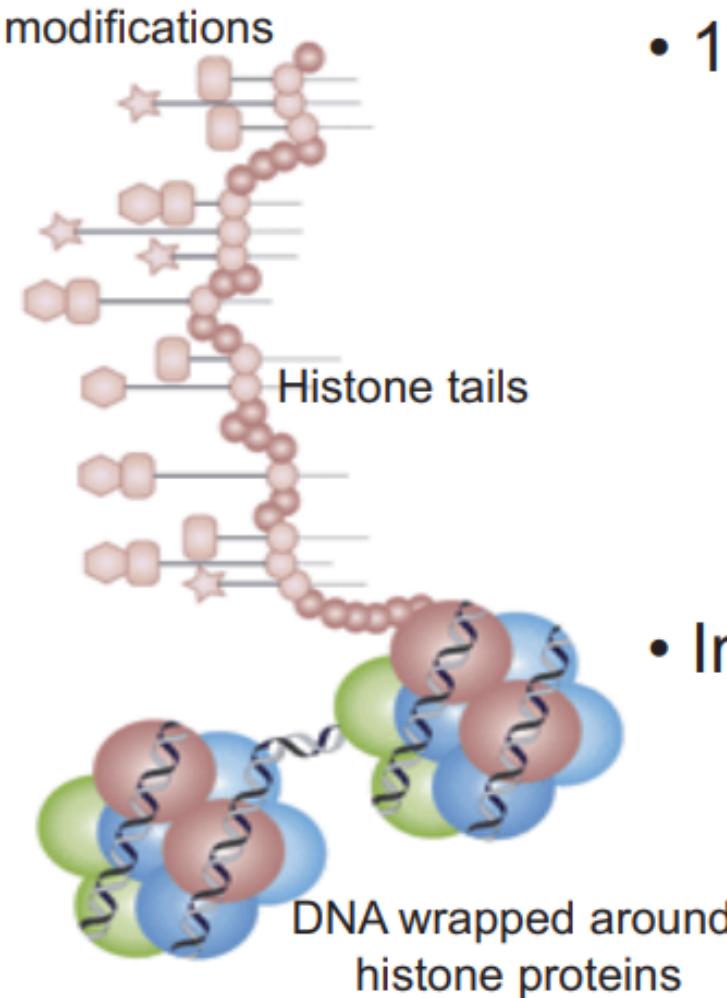
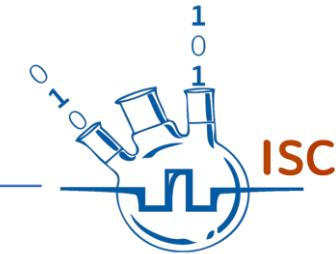
Diverse epigenetic modifications



IT'S MORE THAN A
UNIVERSITY



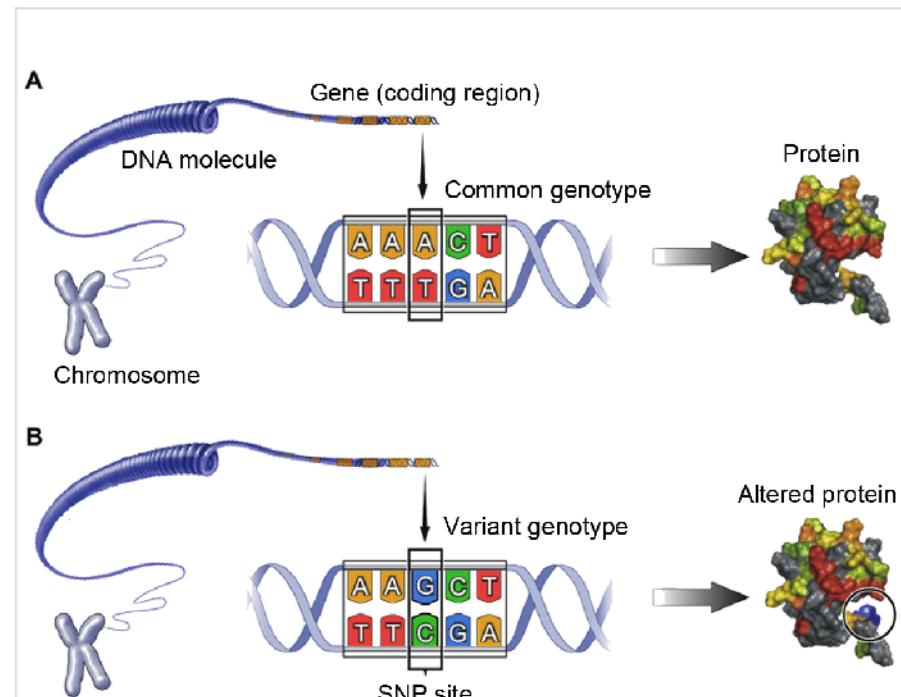
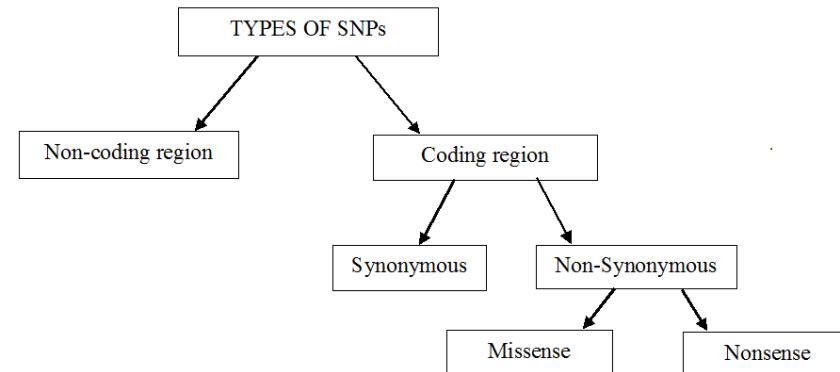
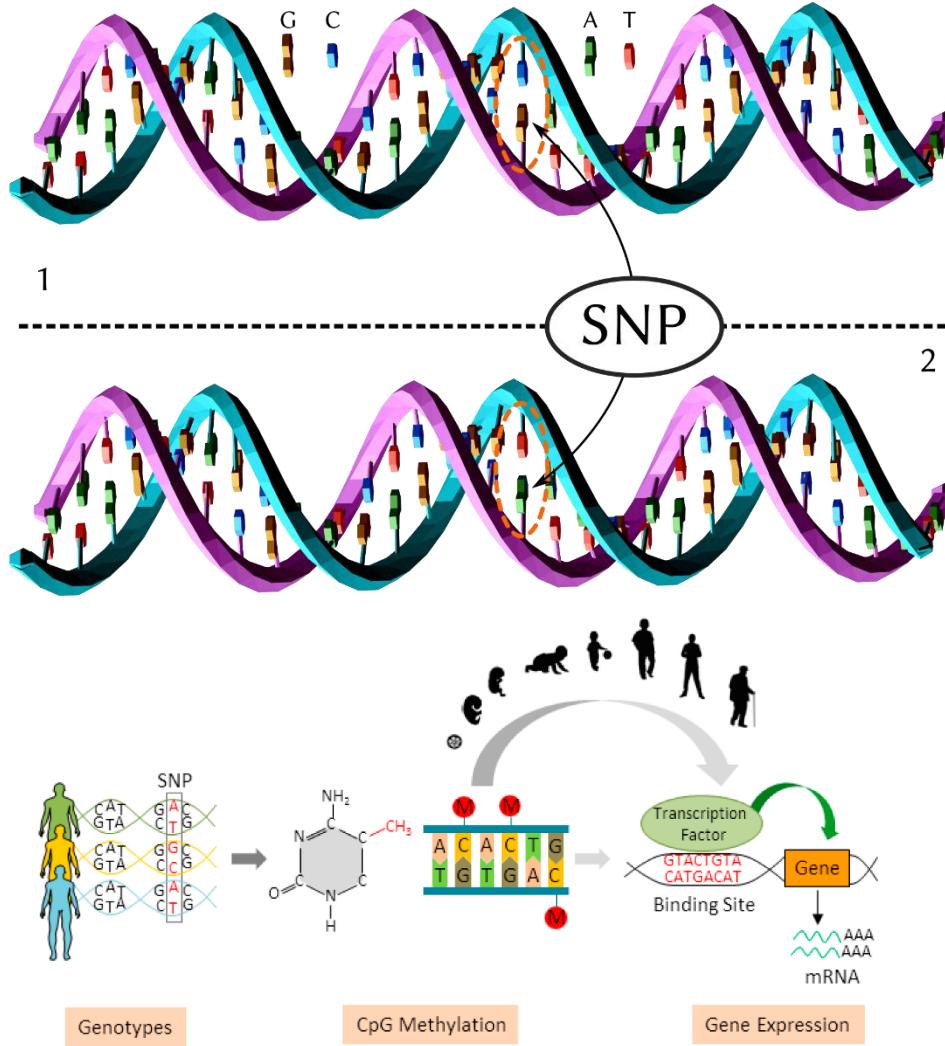
Diverse epigenetic modifications



- 100+ different histone modifications
 - Histone protein → H3/H4/H2A/H2B
 - AA residue → Lysine4(K4)/K36...
 - Chemical modification → Met/Pho/Ubi
 - Number → Me-Me-Me(me3)
 - Shorthand: H3K4me3, H2BK5ac
- In addition:
 - DNA modifications
 - Methyl-C in CpG / Methyl-Adenosine
 - Nucleosome positioning



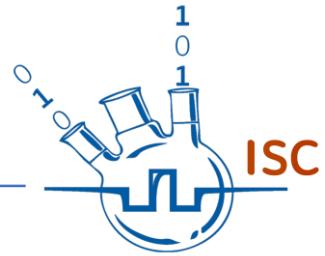
SNPs



IT'S MORE than a
UNIVERSITY

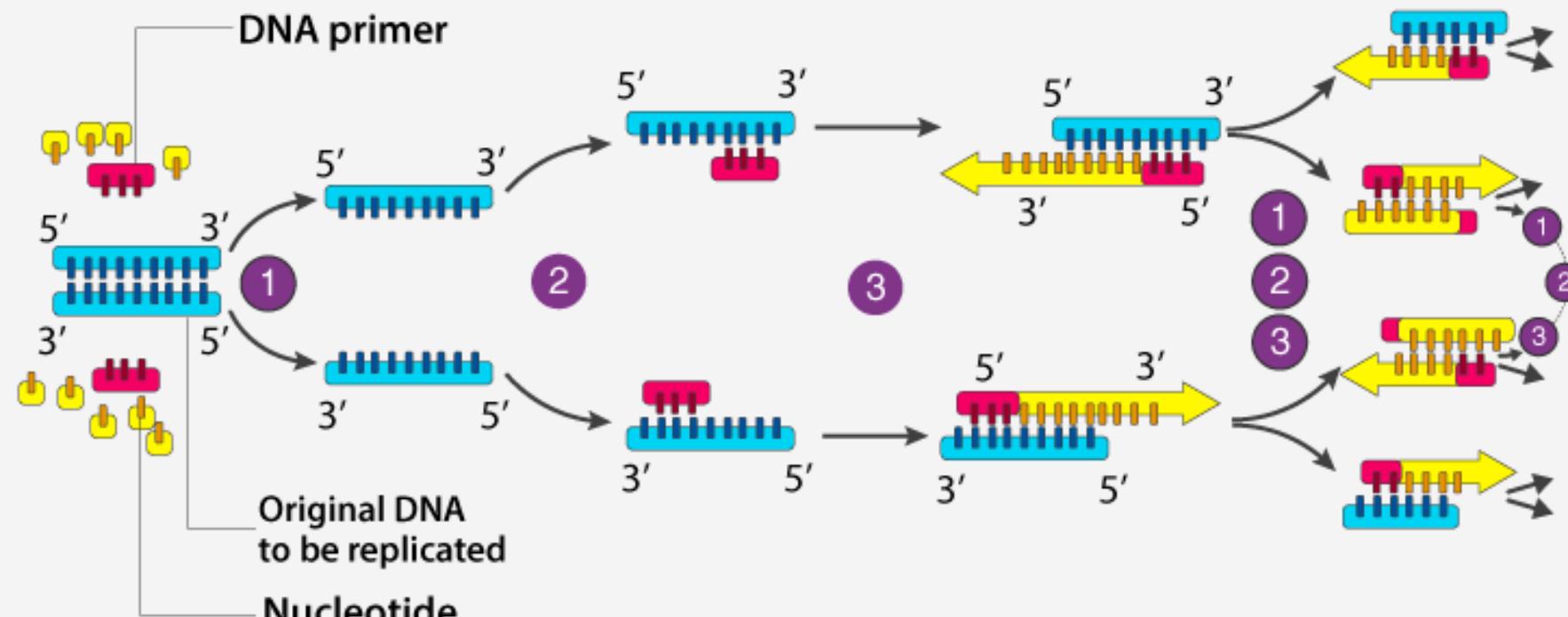


PCR



POLYMERASE CHAIN REACTION (PCR)

BYJU'S
The Learning App



1 Denaturation

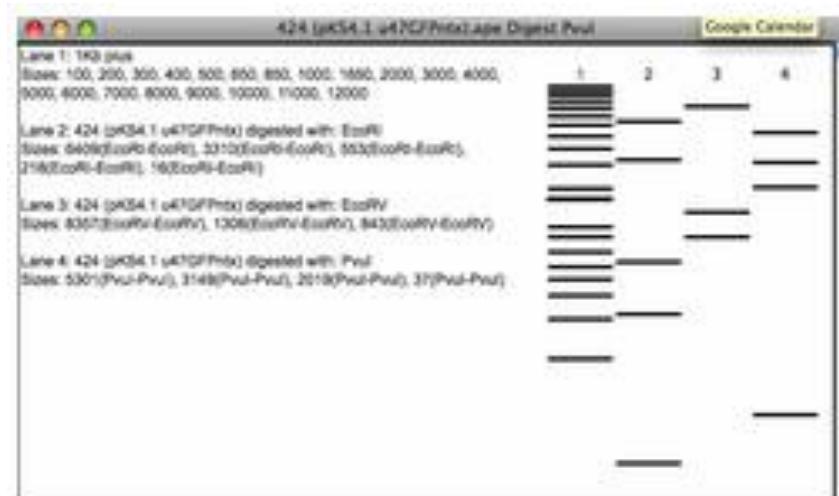
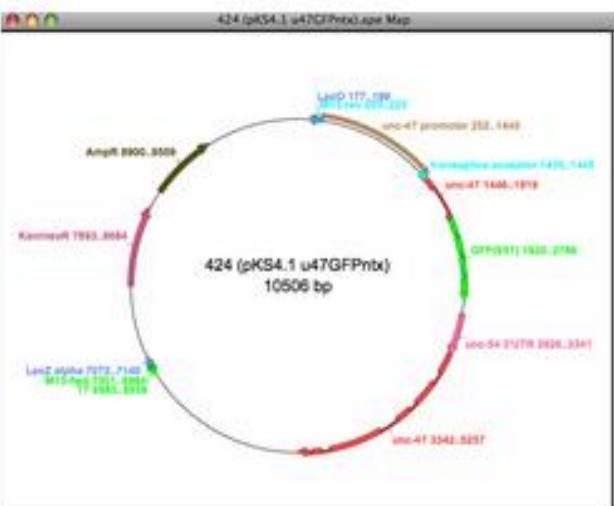
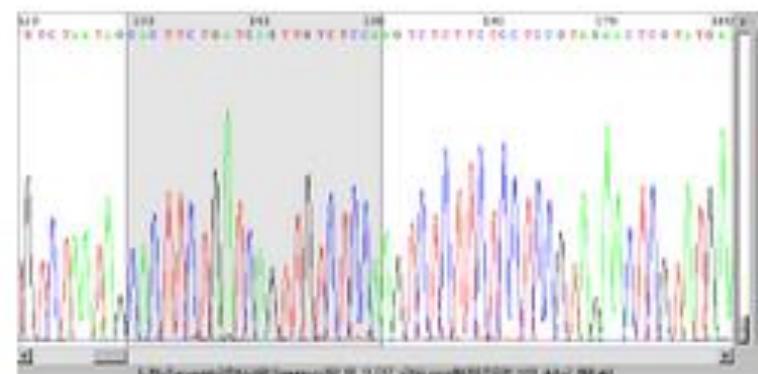
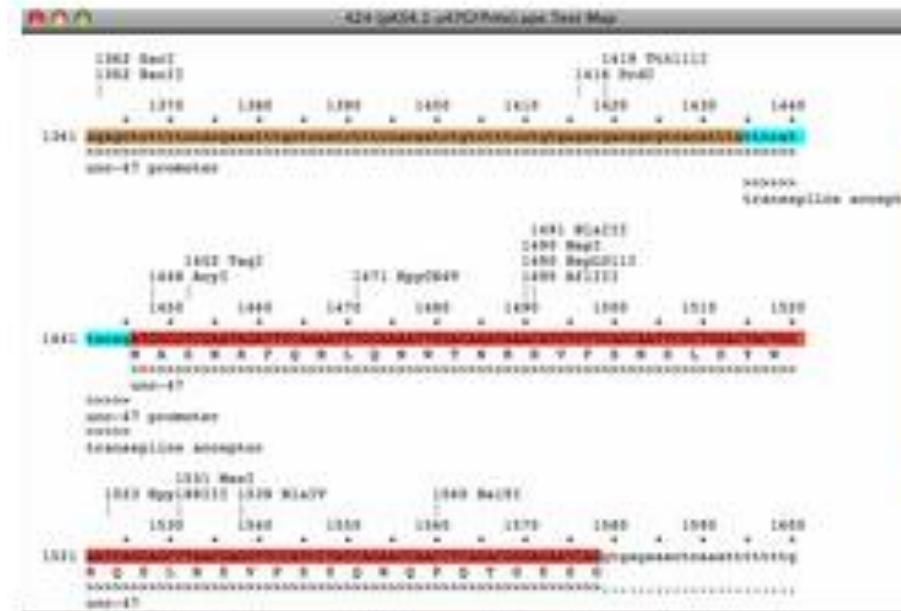
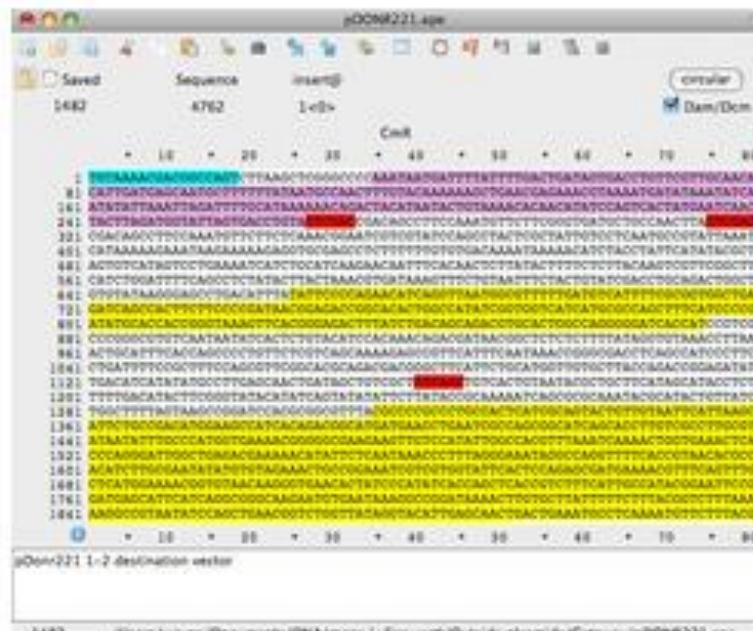
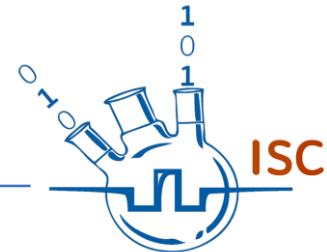
2 Annealing

3 Elongation

IT'S MORE than a
UNIVERSITY



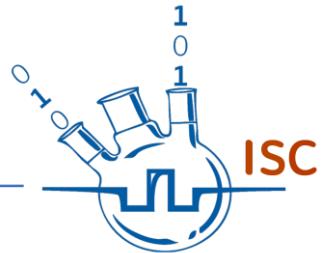
ApE software



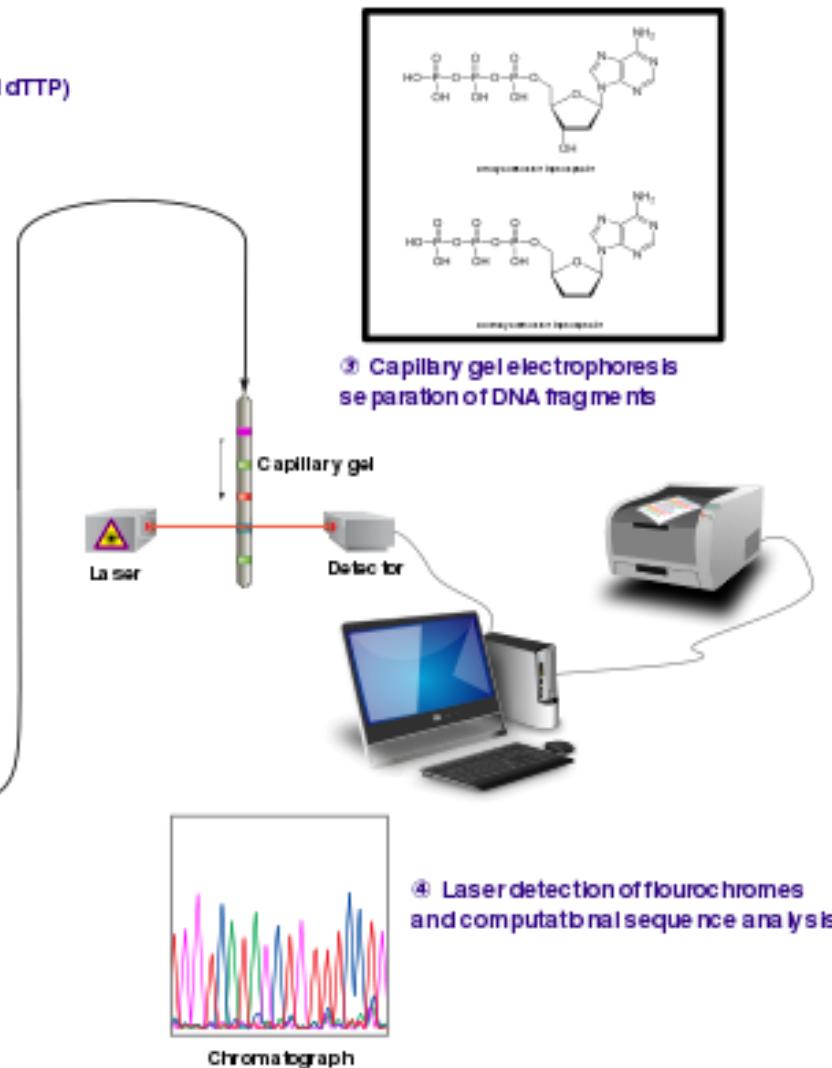
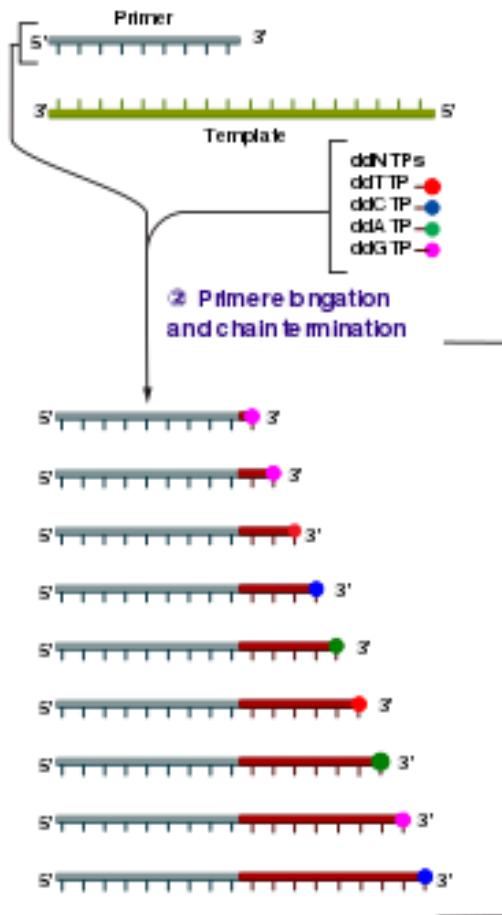
**IT'S MORE THAN A
UNIVERSITY**



Chain-termination sequencing

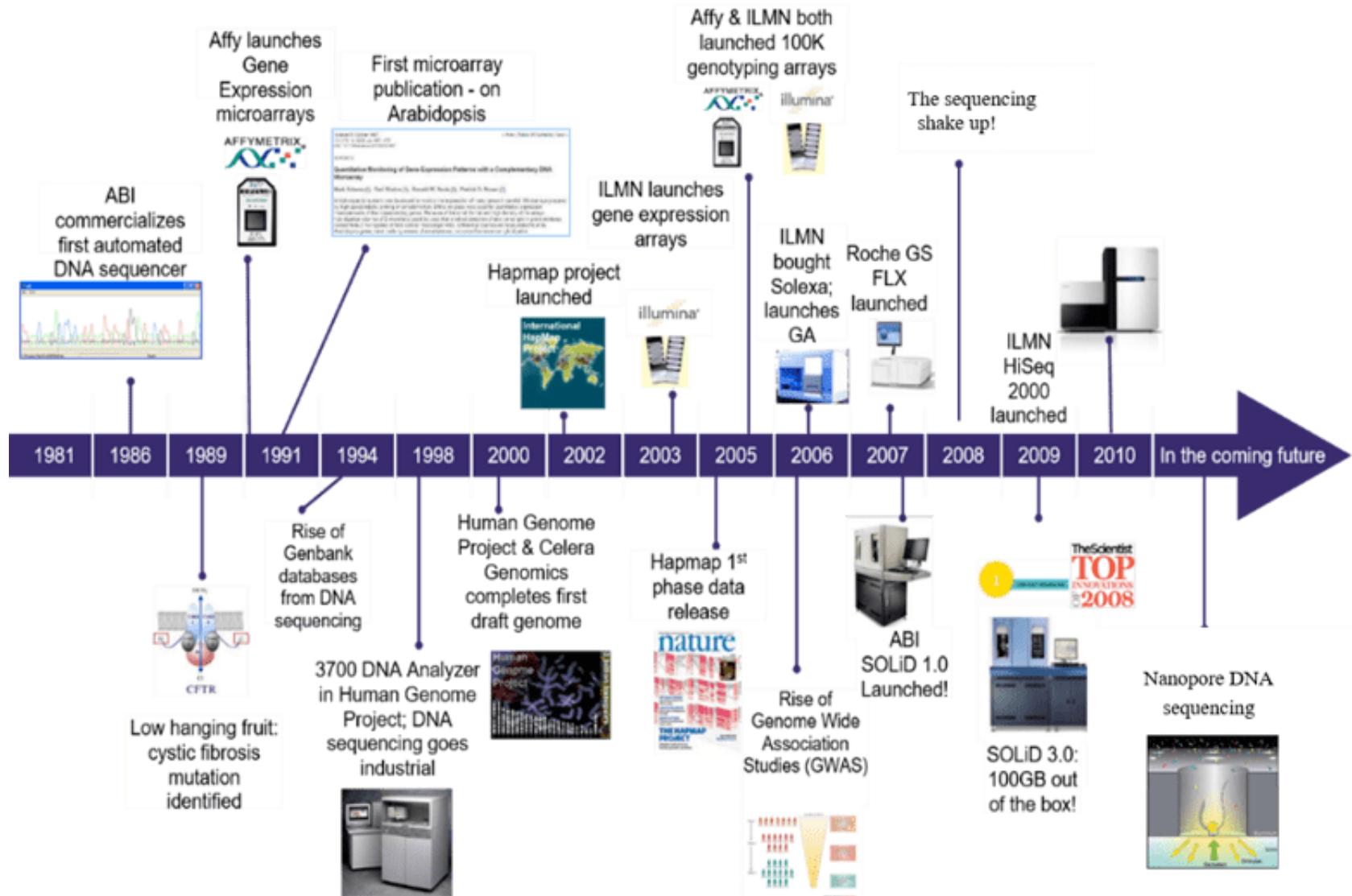


- ① Reaction mixture
 - Primer and DNA template
 - ddNTPs with fluorochromes
 - DNA polymerase
 - dNTPs (dATP, dCTP, dGTP, and dTTP)





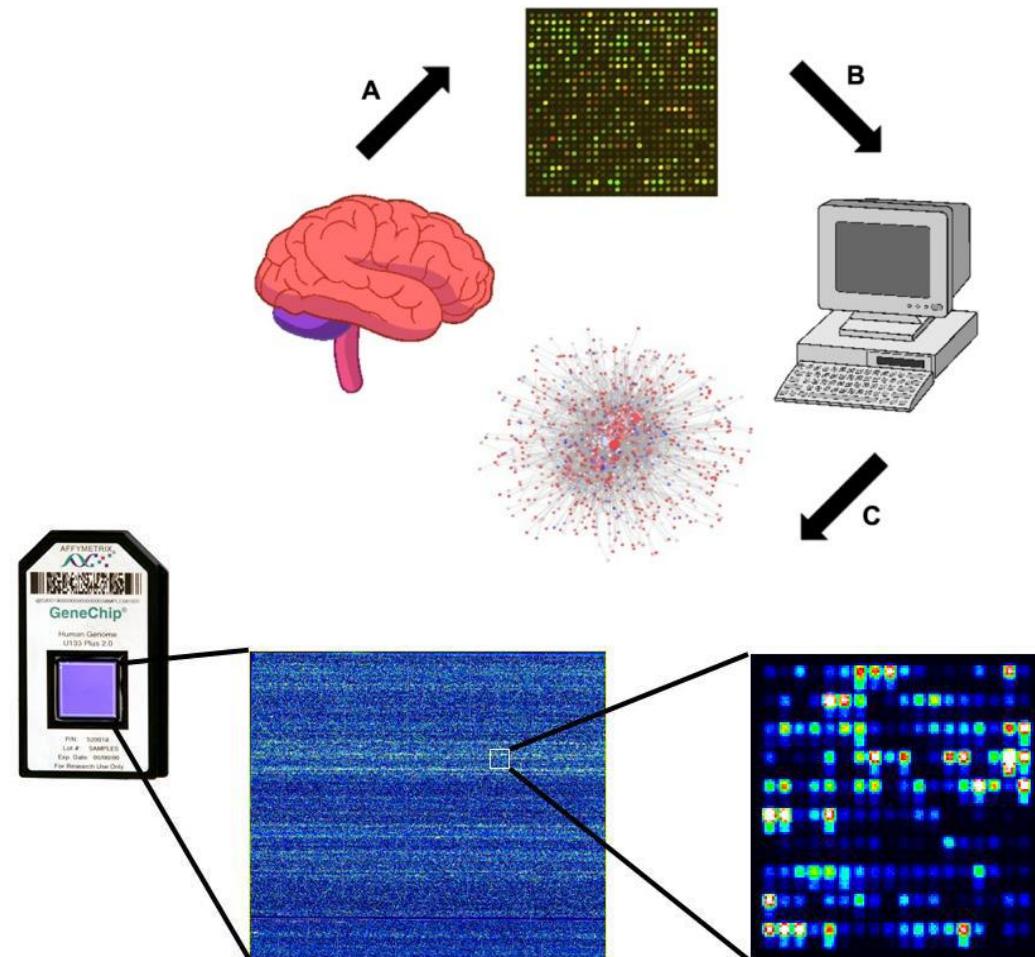
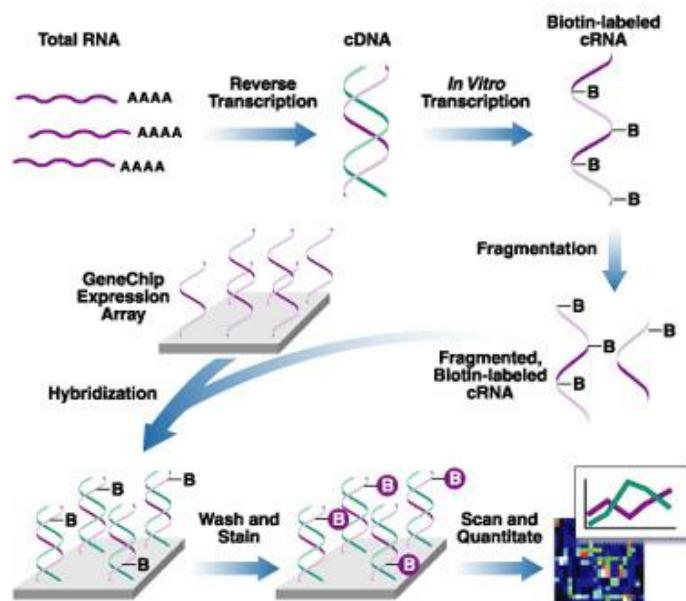
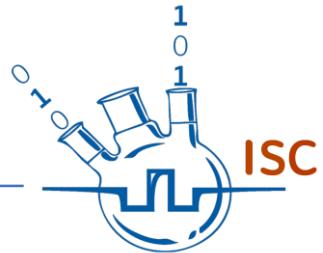
NGS technology



IT'S MORE than a
UNIVERSITY



Affymetrix microarray

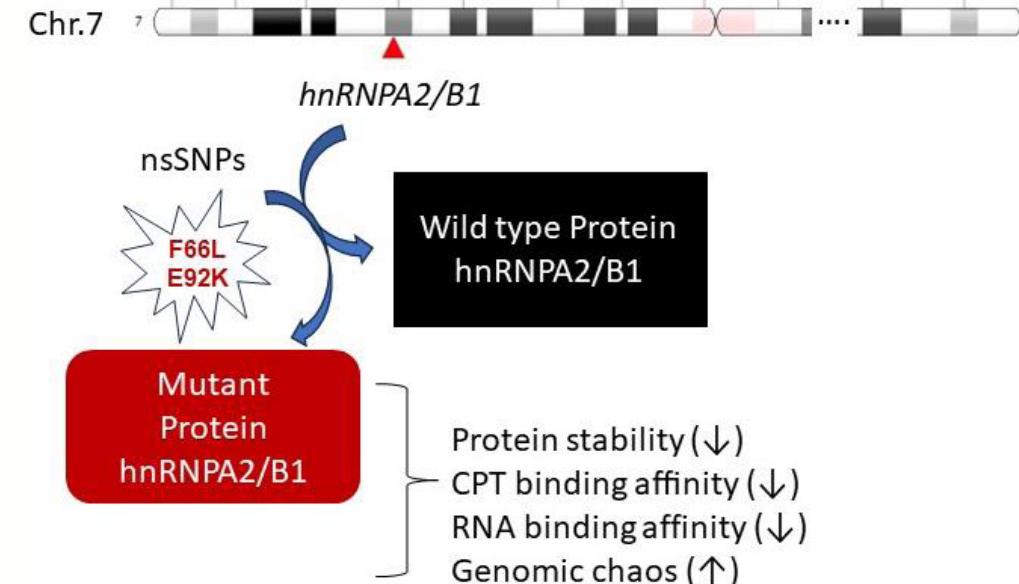
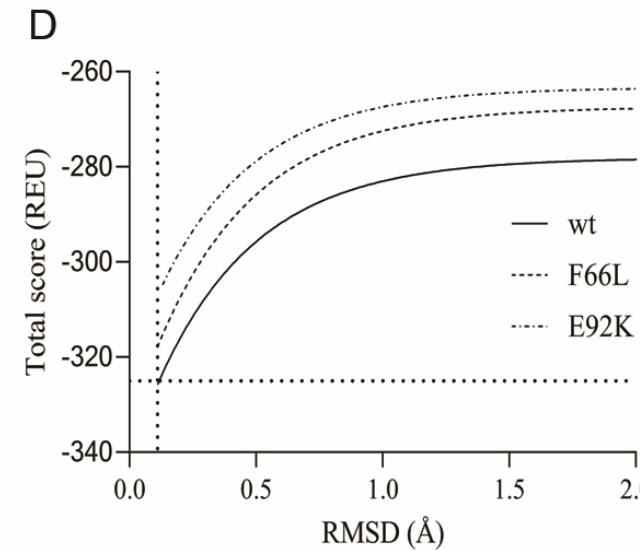
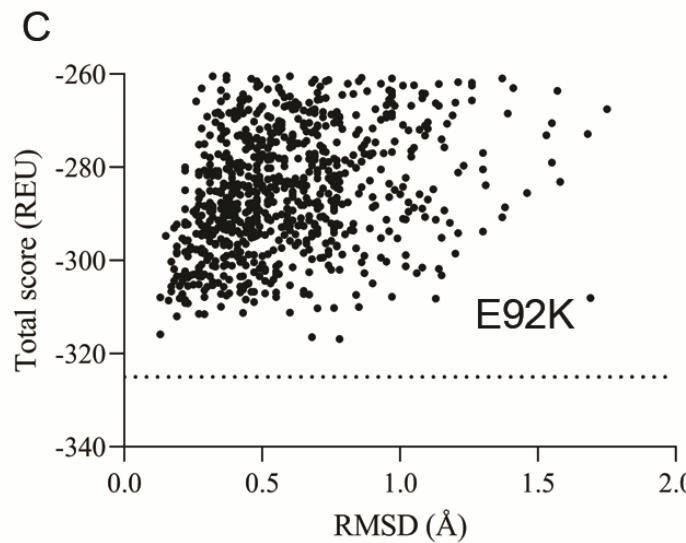
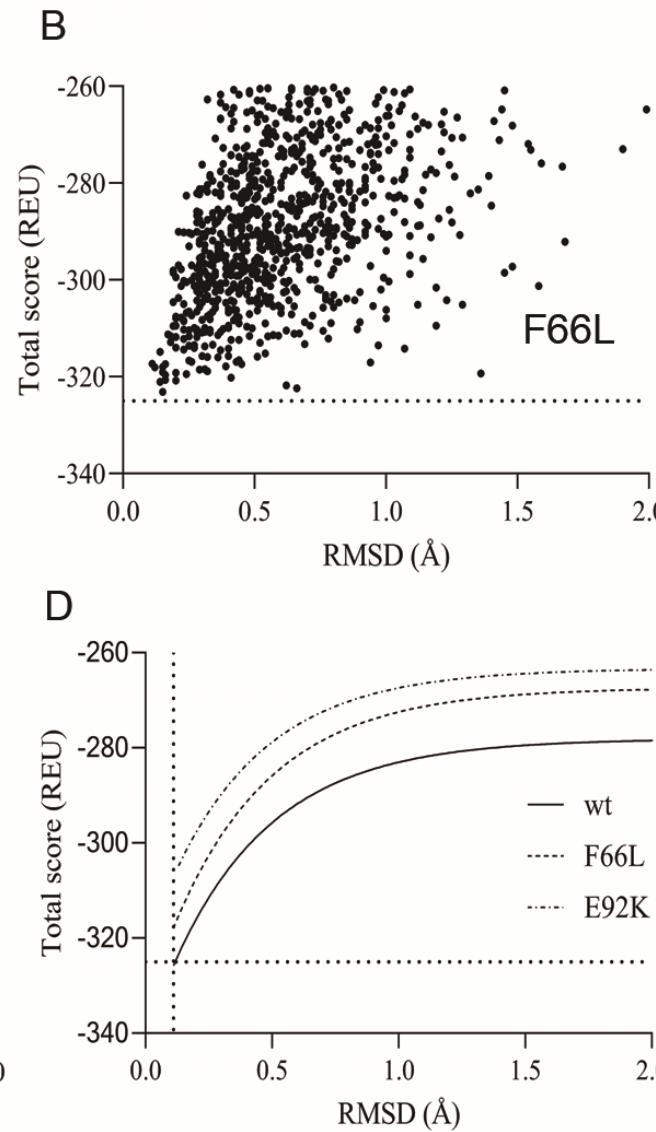
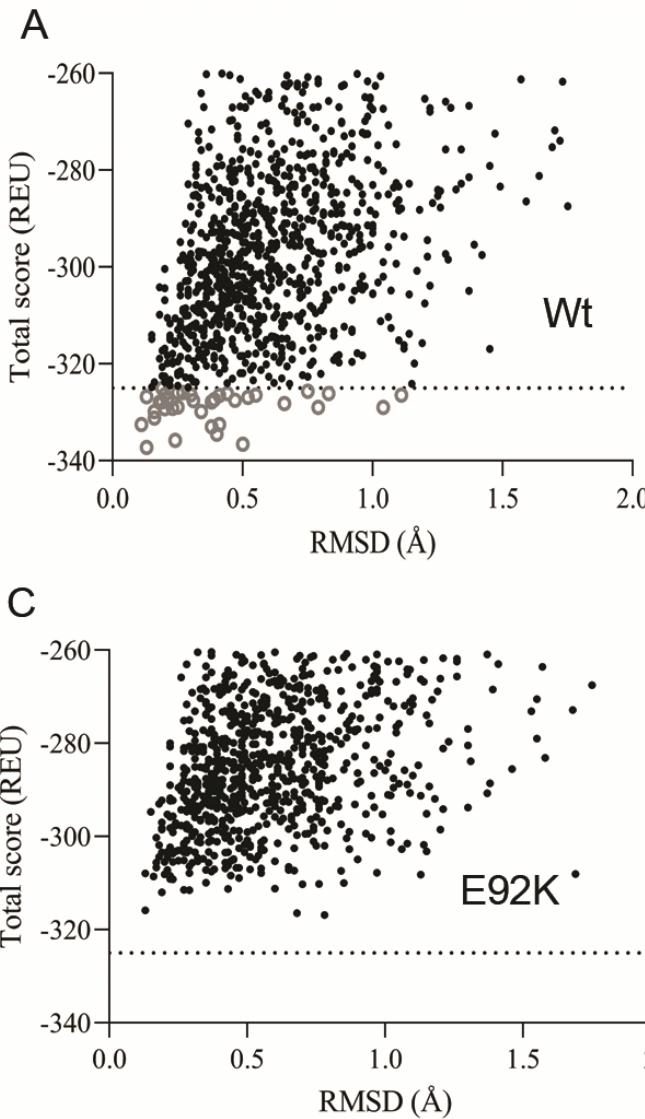
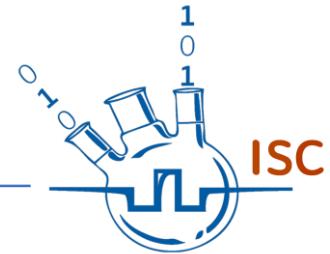


Shityakov et al., 2015

IT₃MOre than a
UNIVERSITY



SNP in hnRNPA2/B1





SNP in hnRNPA2/B1

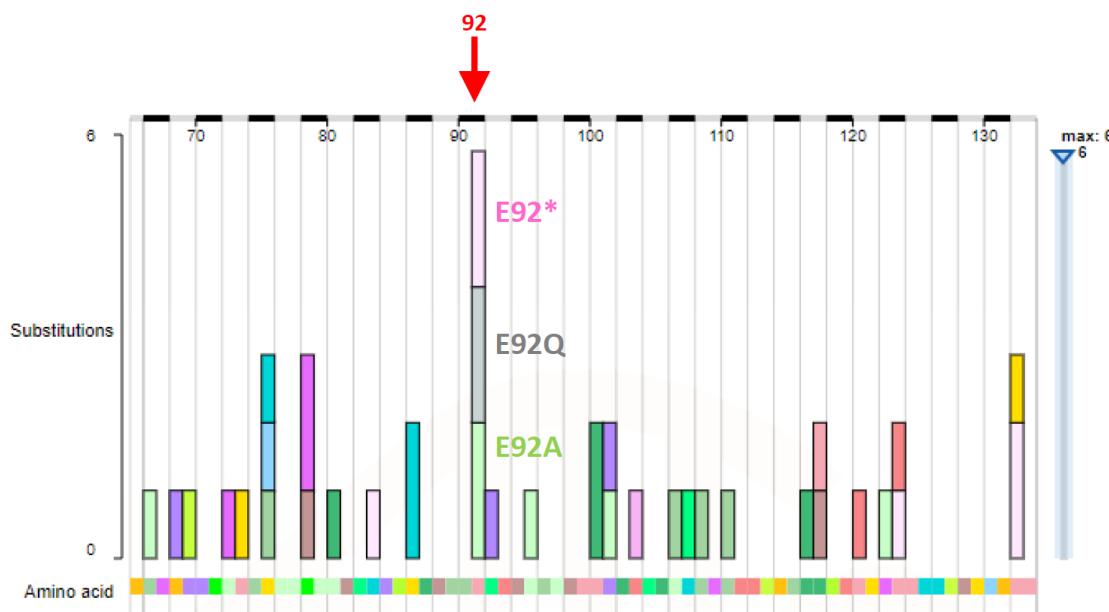
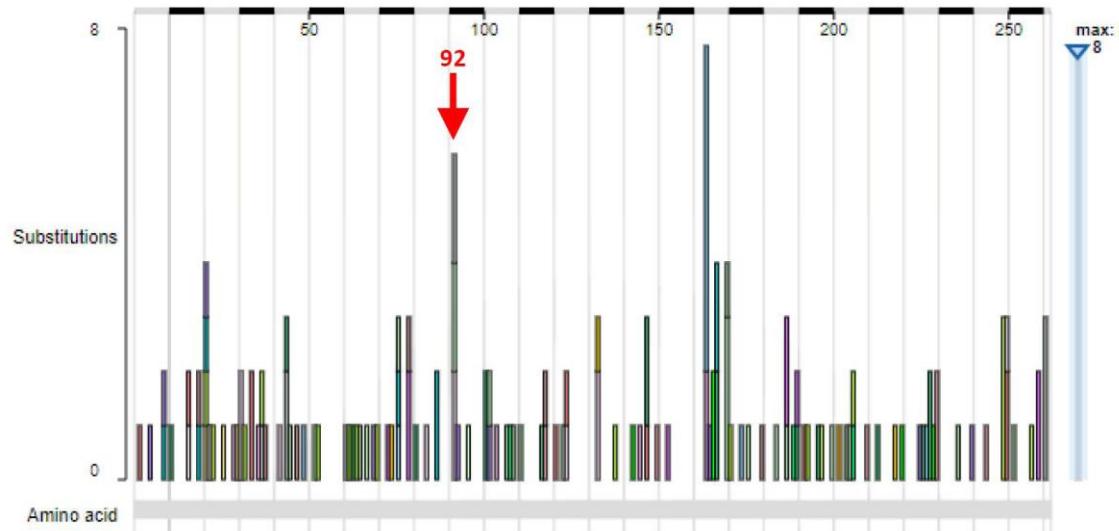
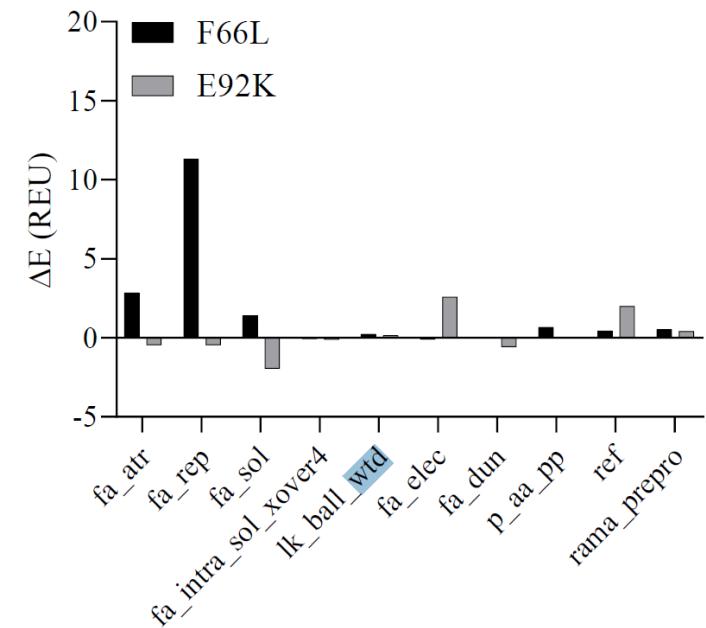


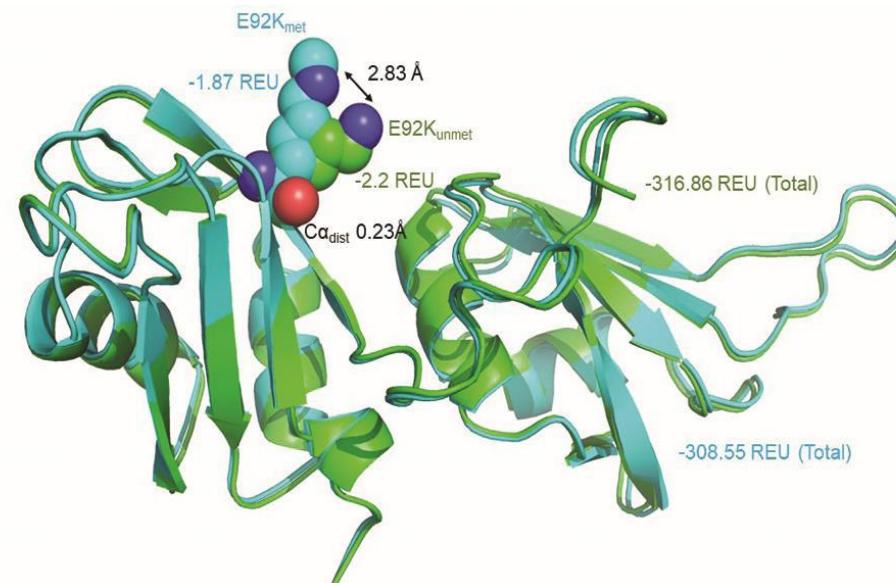
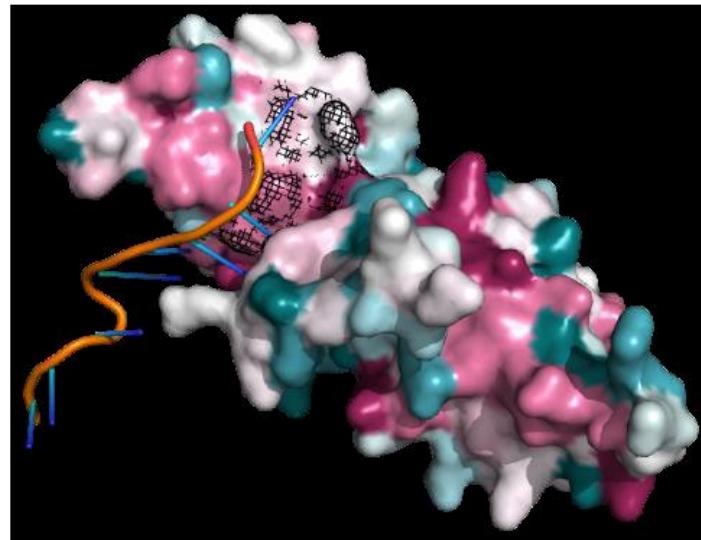
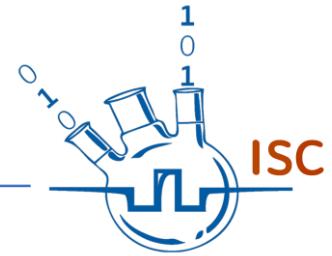
Table 1. Molsoft protein stability prediction

| Protein | Chain | Residue | Wild-type | Mutant | ΔE_{mol} |
|---------|-------|---------|-----------|--------|-------------------------|
| Wt | A | 66 | Phe | Leu | 1.33 |
| Wt | A | 92 | Glu | Lys | 0.18 |





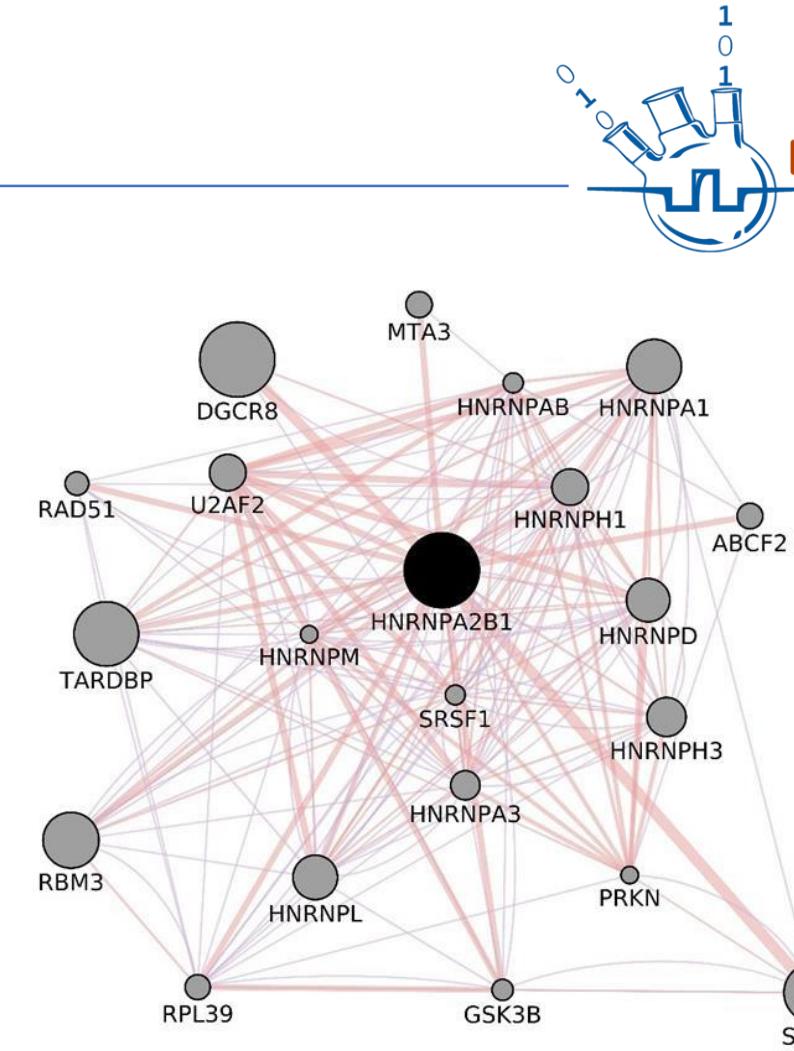
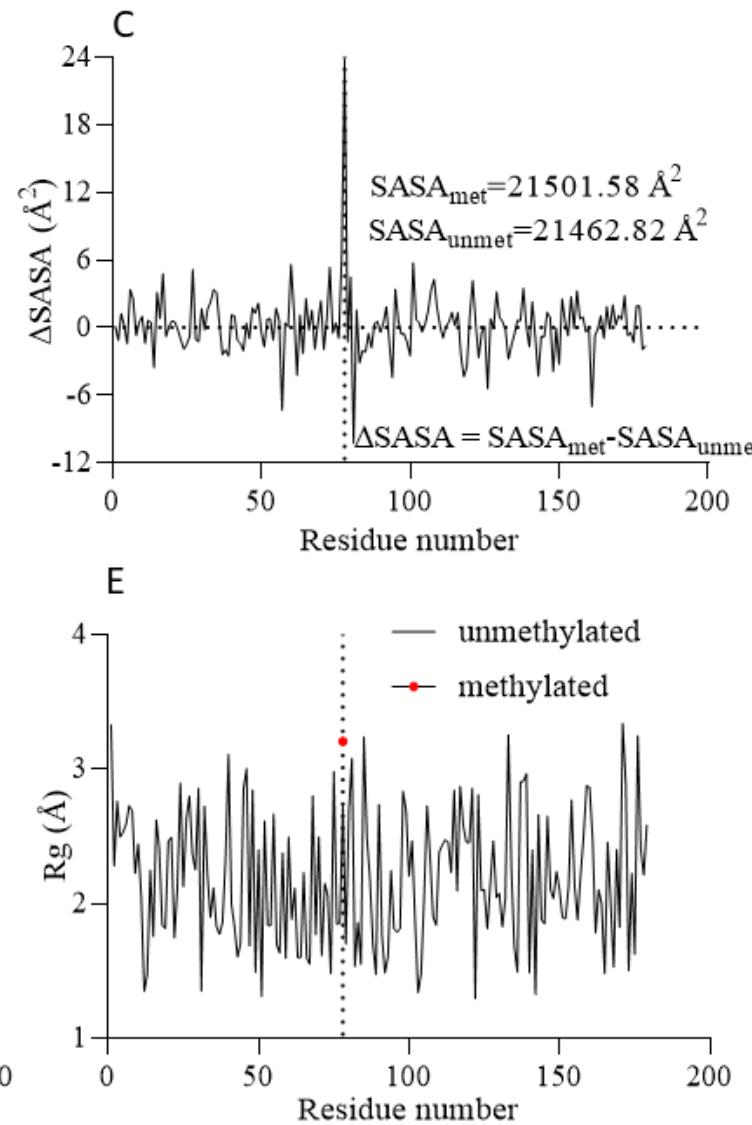
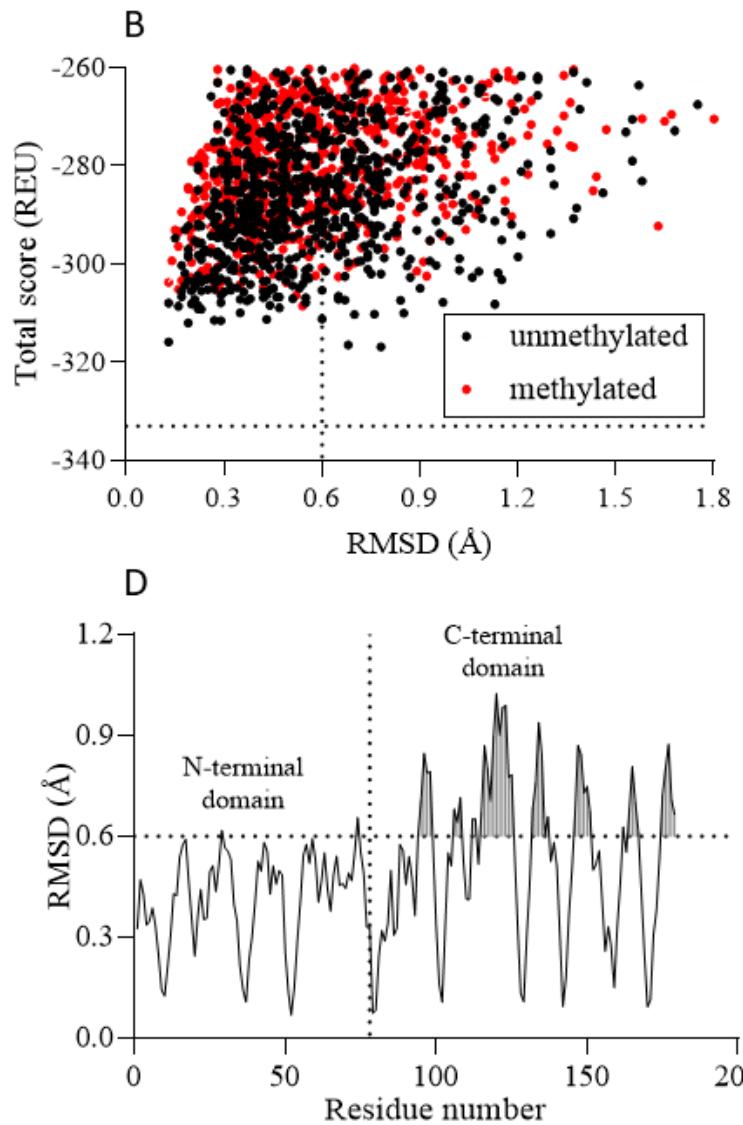
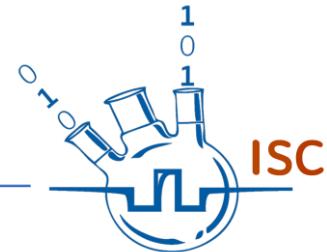
SNP in hnRNP A2/B1



IT'sMOre than a
UNIVERSITY



SNP in hnRNPA2/B1





SNP in hnRNPA2/B1

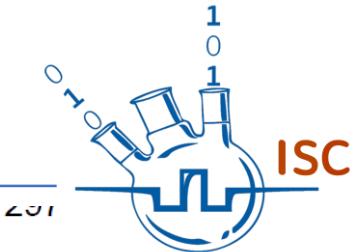


Table 2. Predicted posttranslational modifications and pathogenicity of the wild-type and substitution mutants of hnRNPA2/B1

| Position | PTM scores | SNP type | MutPred2 score | SNP effect | PTMs |
|----------|------------|----------|----------------|------------|-------------|
| F66 (wt) | - | - | - | - | - |
| E92 (wt) | - | - | - | - | - |
| F66L | - | NS | 0.877 | Pathogenic | - |
| E92K | 0.677 | NS | 0.779 | Pathogenic | Methylation |

NS, non-synonymous.

Table 3. Dihedral angle analysis of substitution
mutant E92K_{met} and E92K_{unmet}

| Angle | Unmethylated | Methylated |
|-------|--------------|-------------------------|
| φ | -131.98° | -131.95° ³⁰⁰ |
| ψ | 87.08° | 161.87° ³⁰¹ |
| ω | -176.31 | 87.13° ³⁰² |
| X1 | -175.51 | -175.51 |
| X2 | 177.99 | 177.98° ³⁰³ |
| X3 | 178.49 | 178.49° ³⁰⁴ |
| X4 | -177.33 | -177.33° ³⁰⁵ |



SNP in hnRNPA2/B1

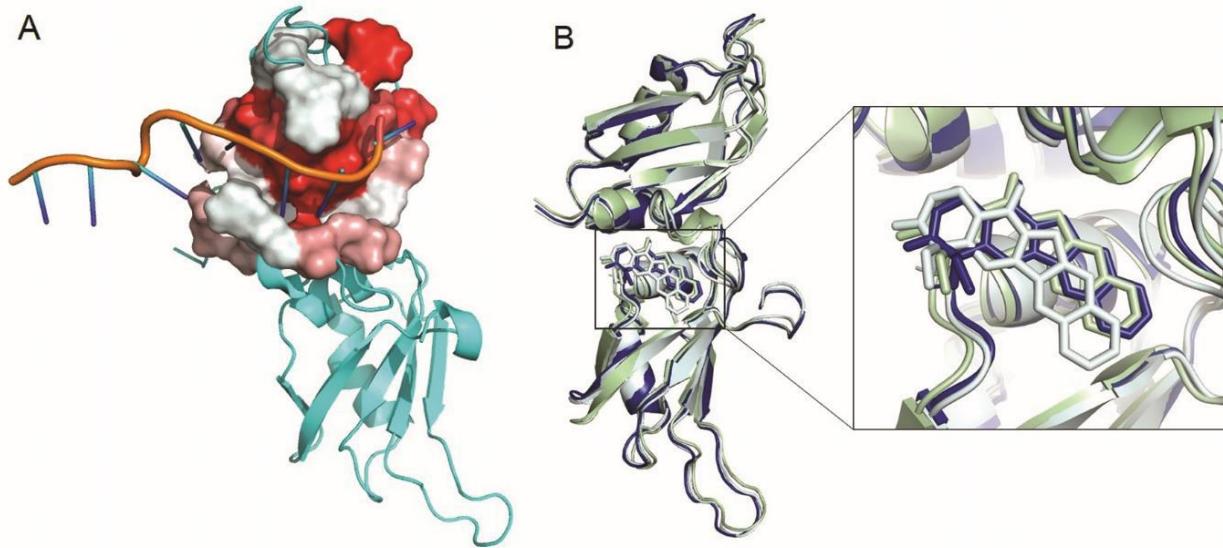
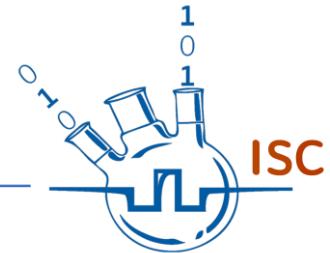
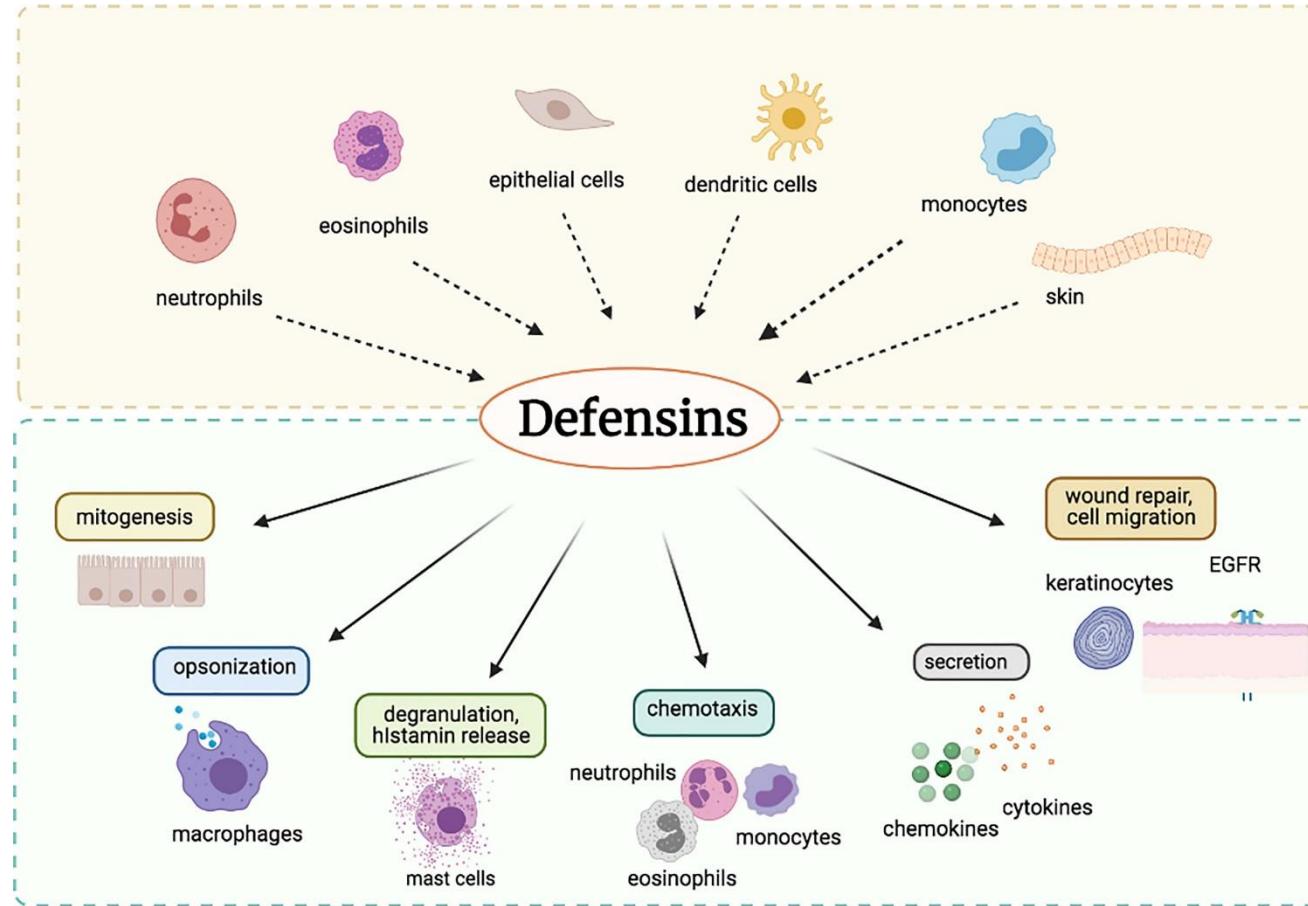
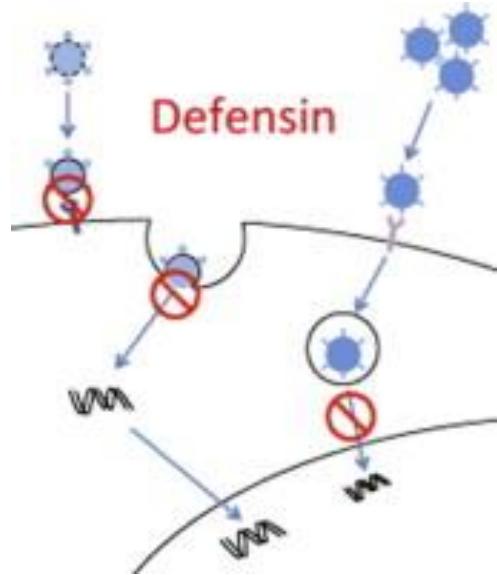
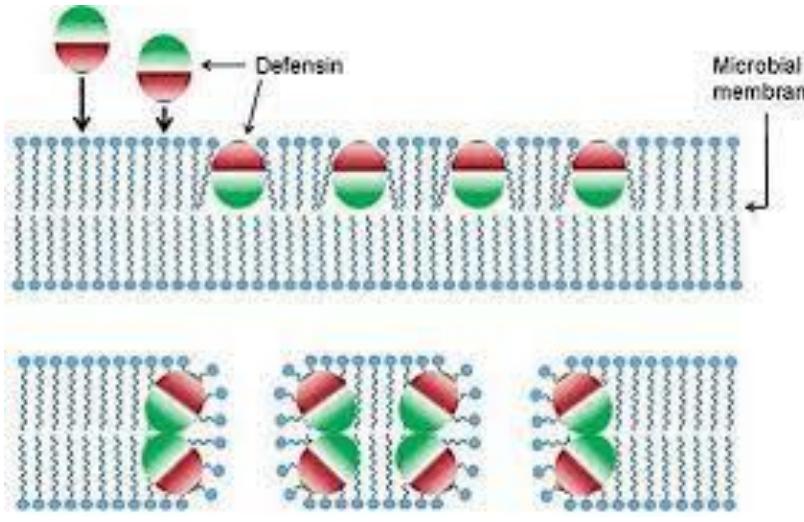
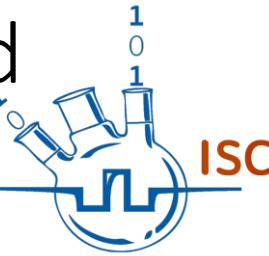


Table 4. Summary of molecular docking and structure-truncated MM/PB(GB)SA rescoring

| Proteins | Total energy (kcal/mol) | | | |
|----------|---------------------------|---------------------|-------------------|--------------------|
| | FoldX protein-RNA docking | Protein-CPT docking | fastDRH* | |
| | | | GB | PB |
| Wt | -18.52 | -9.55 | -30.83 ± 4.54 | -14.21 ± 11.50 |
| F66L | -16.40 | -9.41 | -30.72 ± 3.77 | -15.23 ± 12.17 |
| E92K | -15.42 | -8.23 | -27.13 ± 4.35 | -11.07 ± 11.18 |

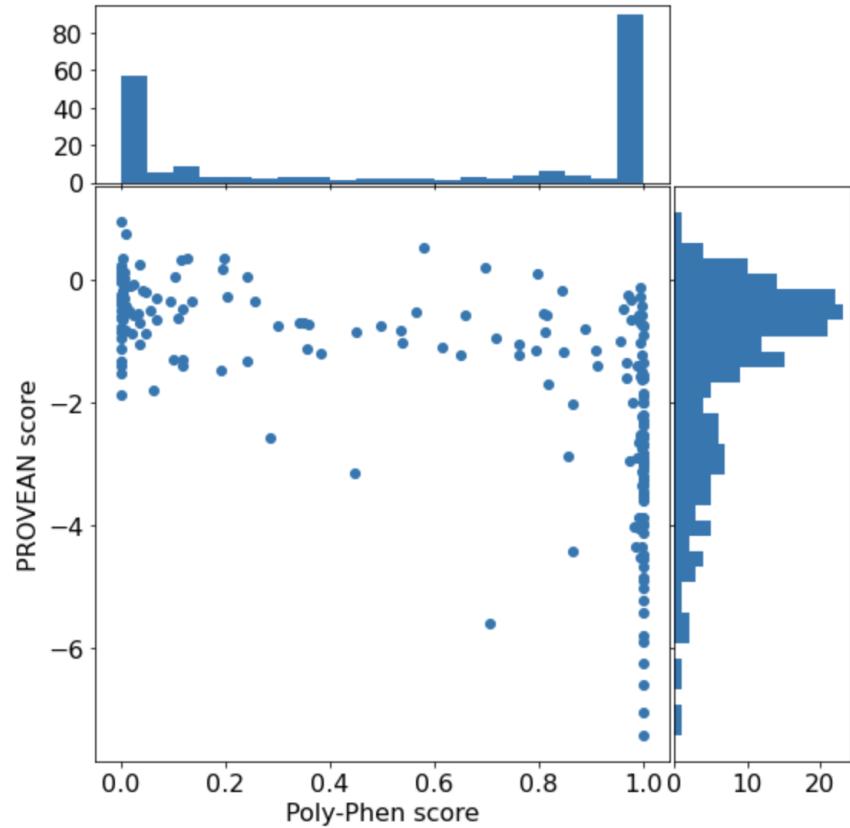
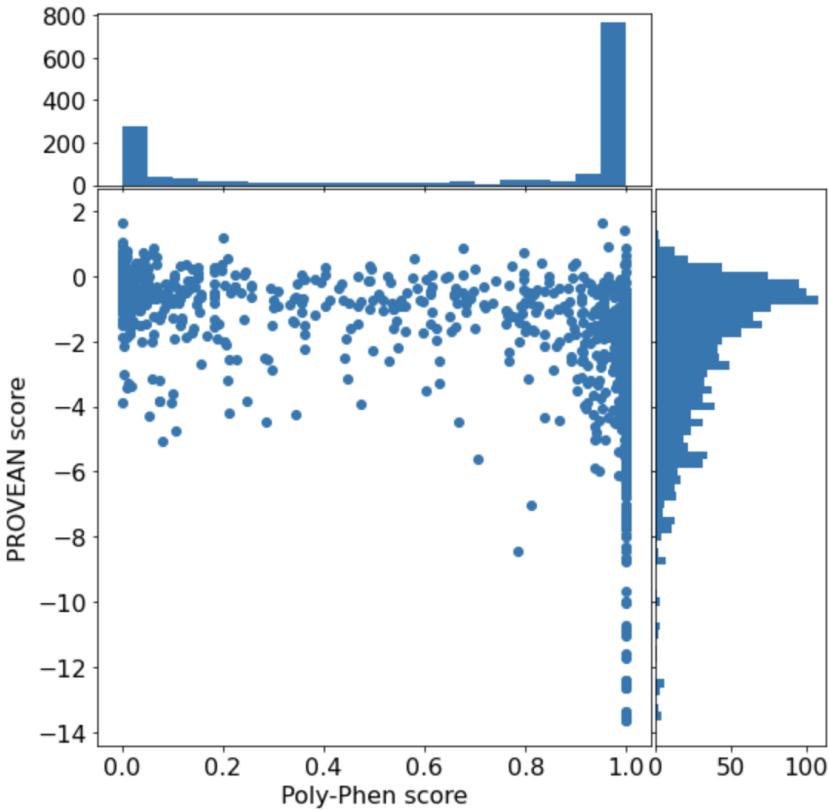
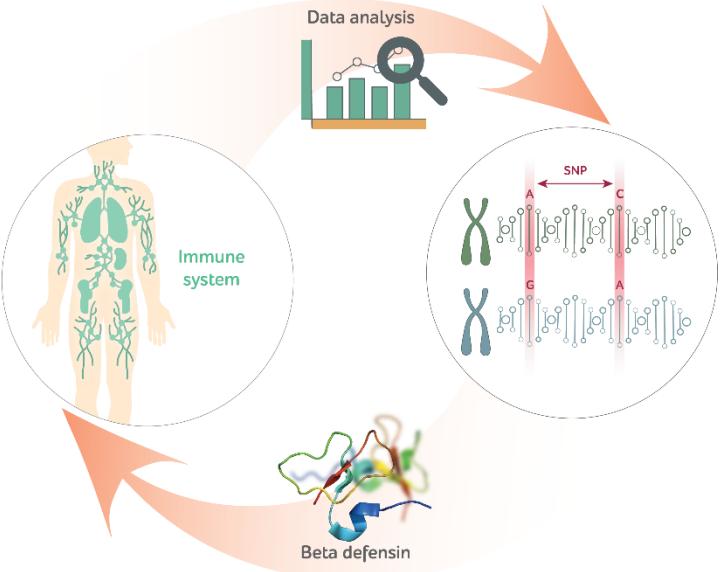


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites



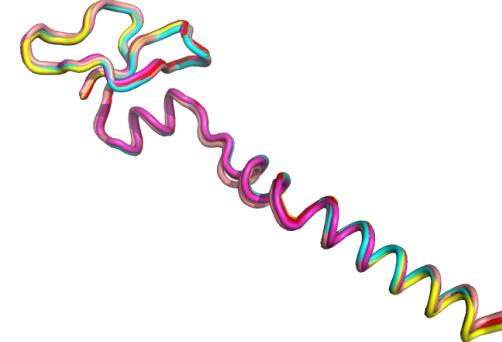
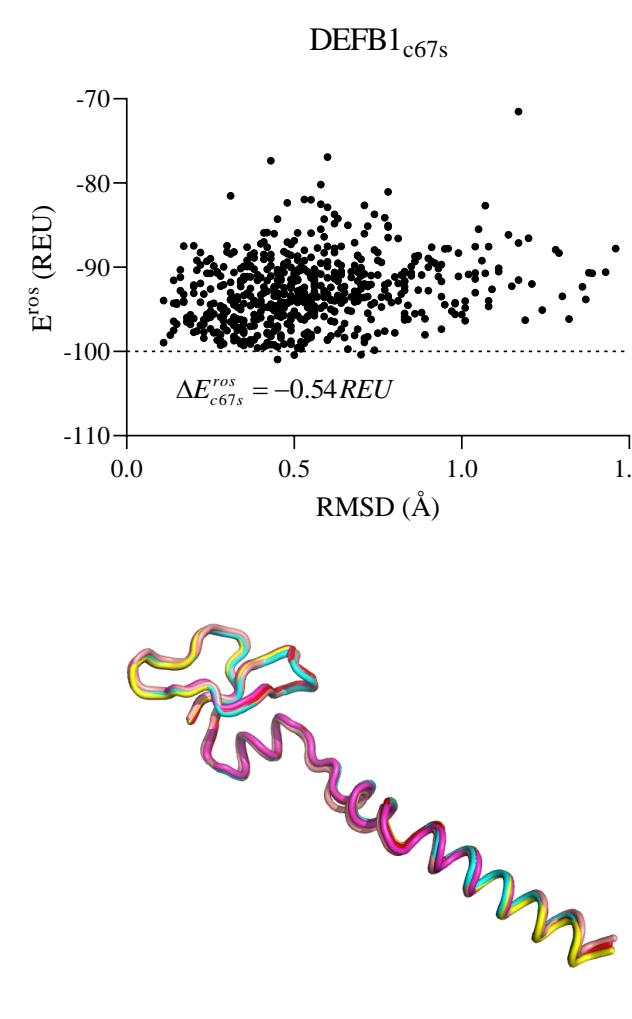
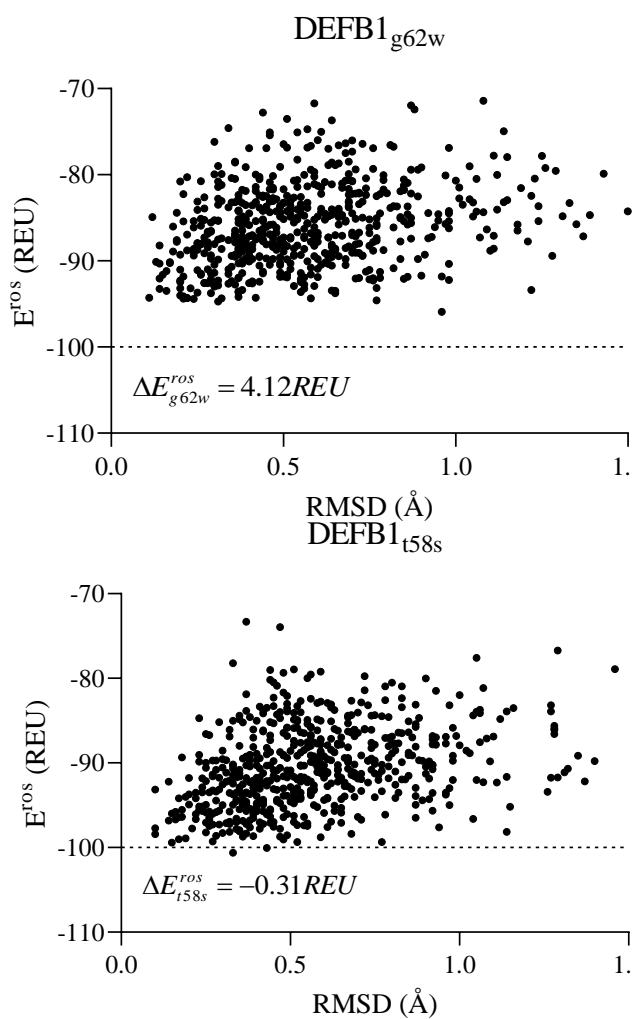
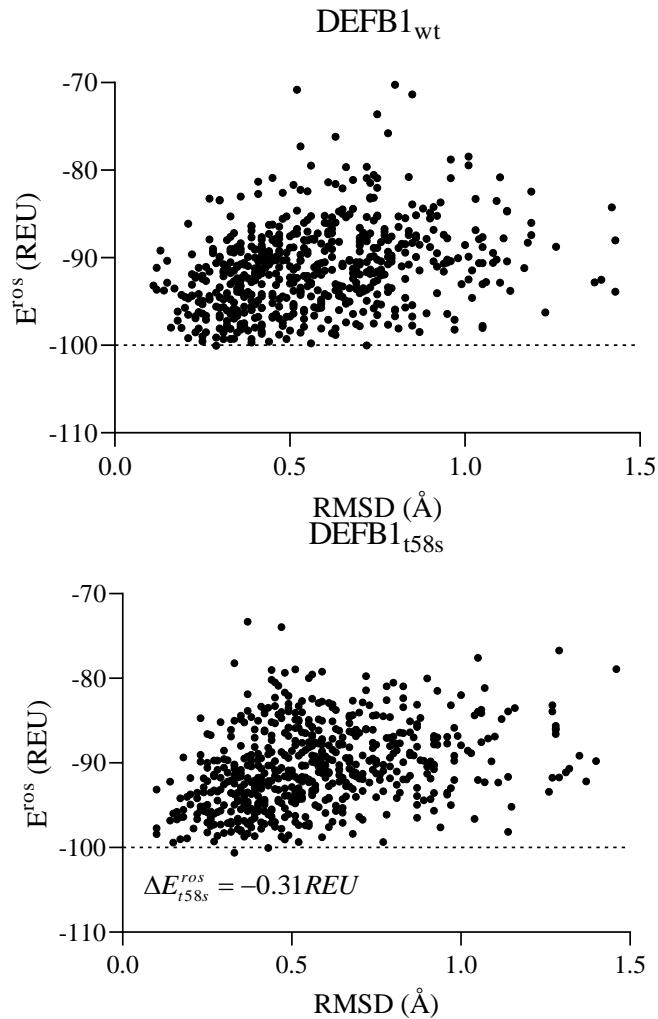
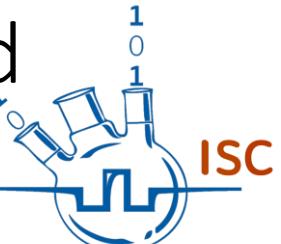


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites



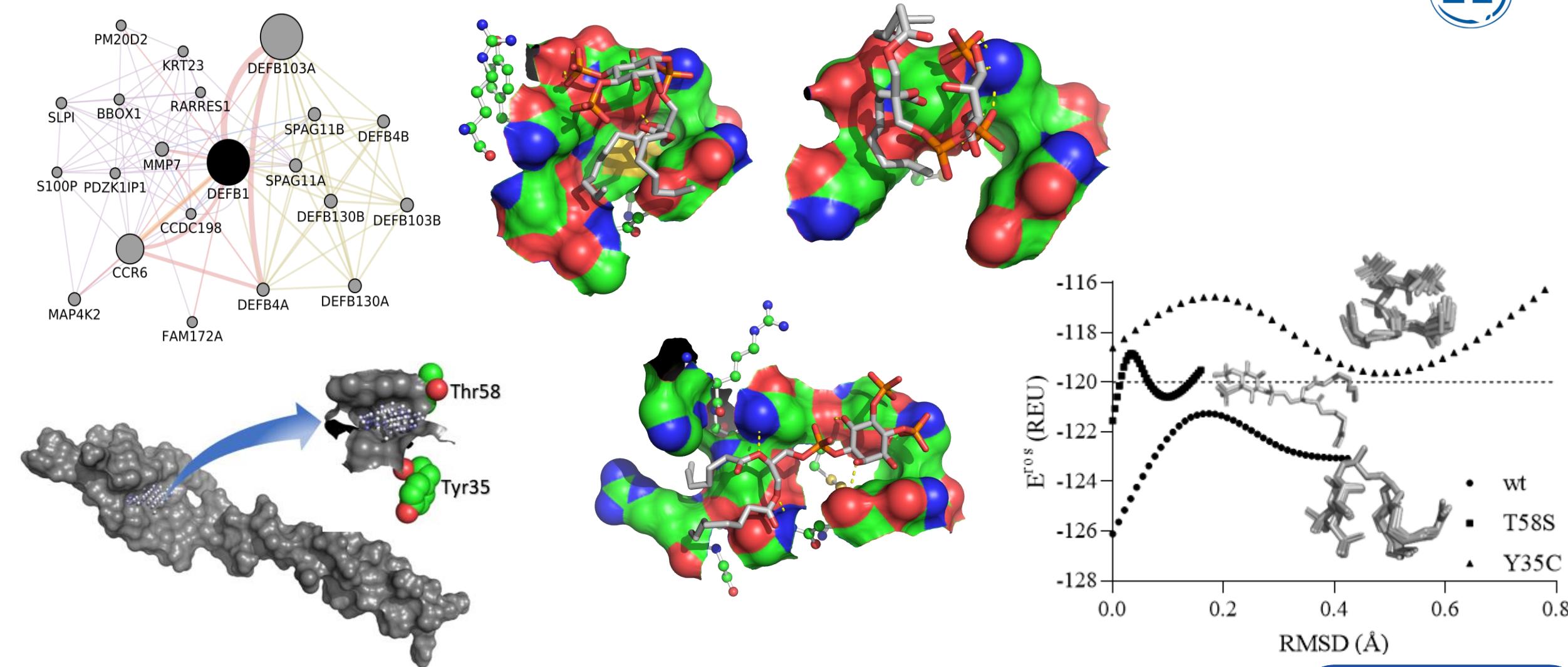
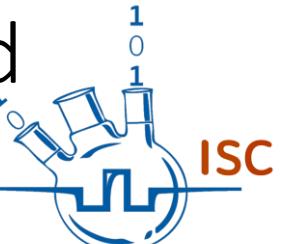


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites

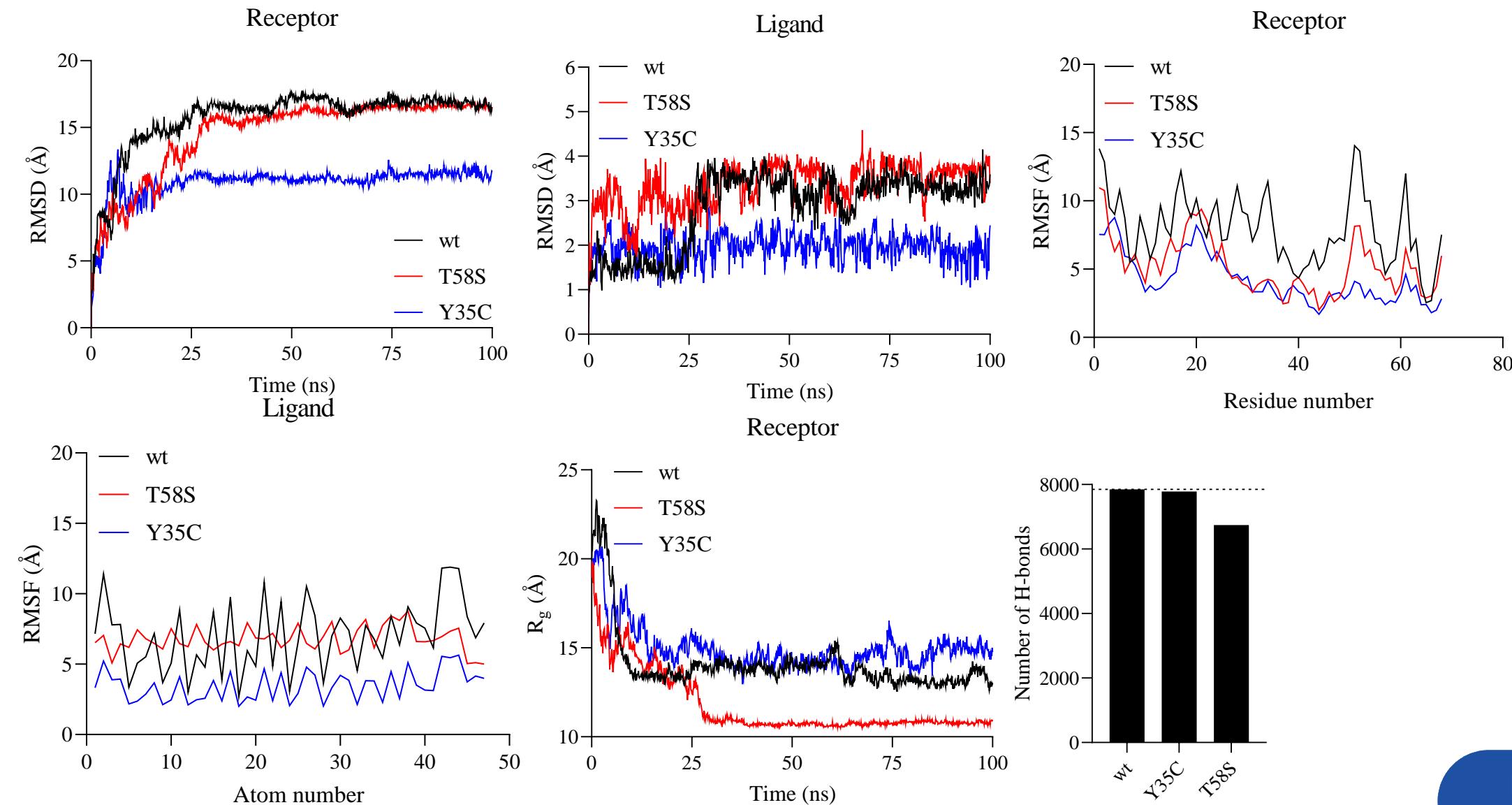




nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites

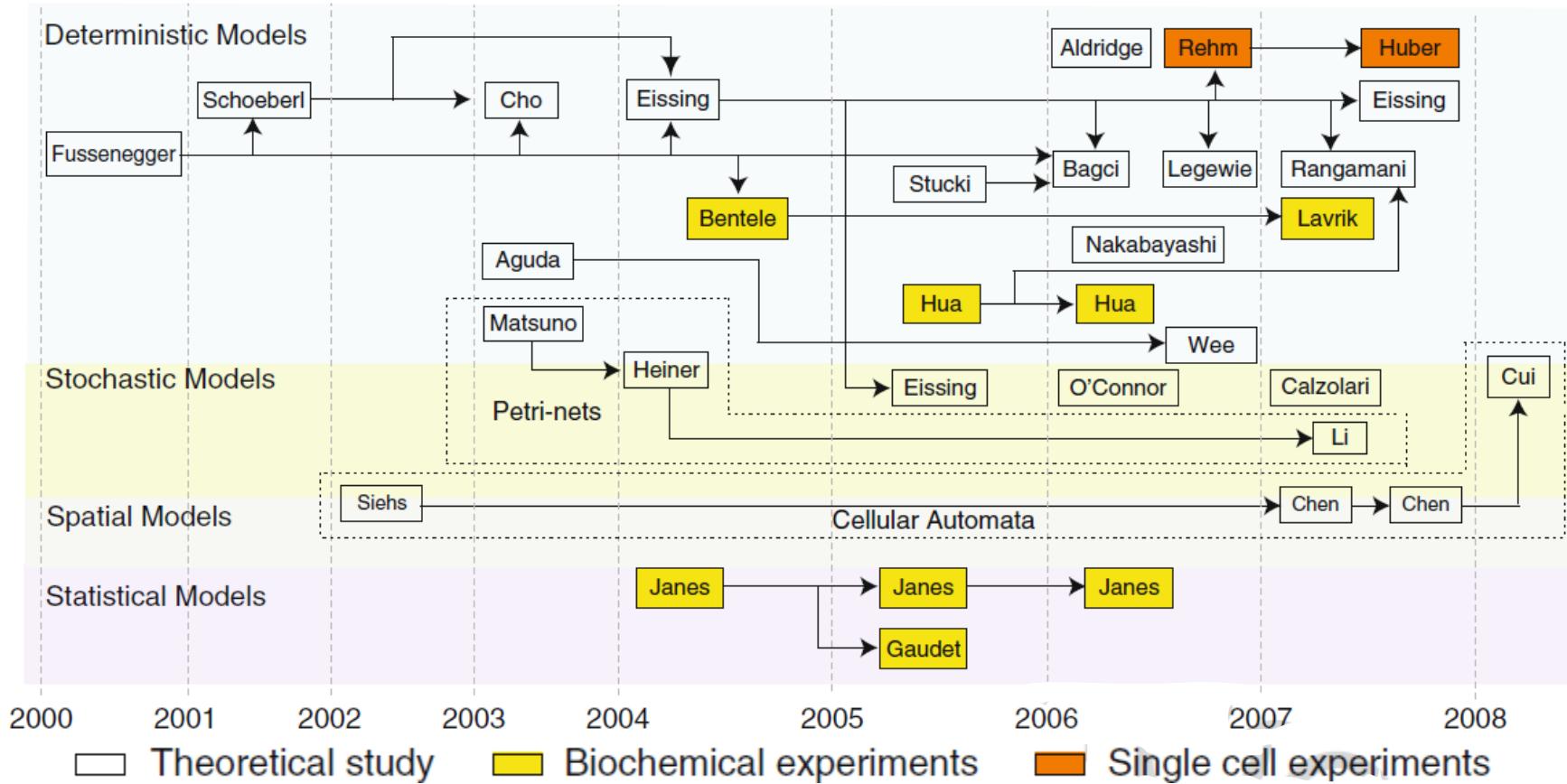


nsSNPs in DEFB1 gene reveal impact on protein-ligand binding sites





Cell death modelling



(From Huber, Bullinger and Rehm, Systems Biology Approaches to the Study of Apoptosis 2009)

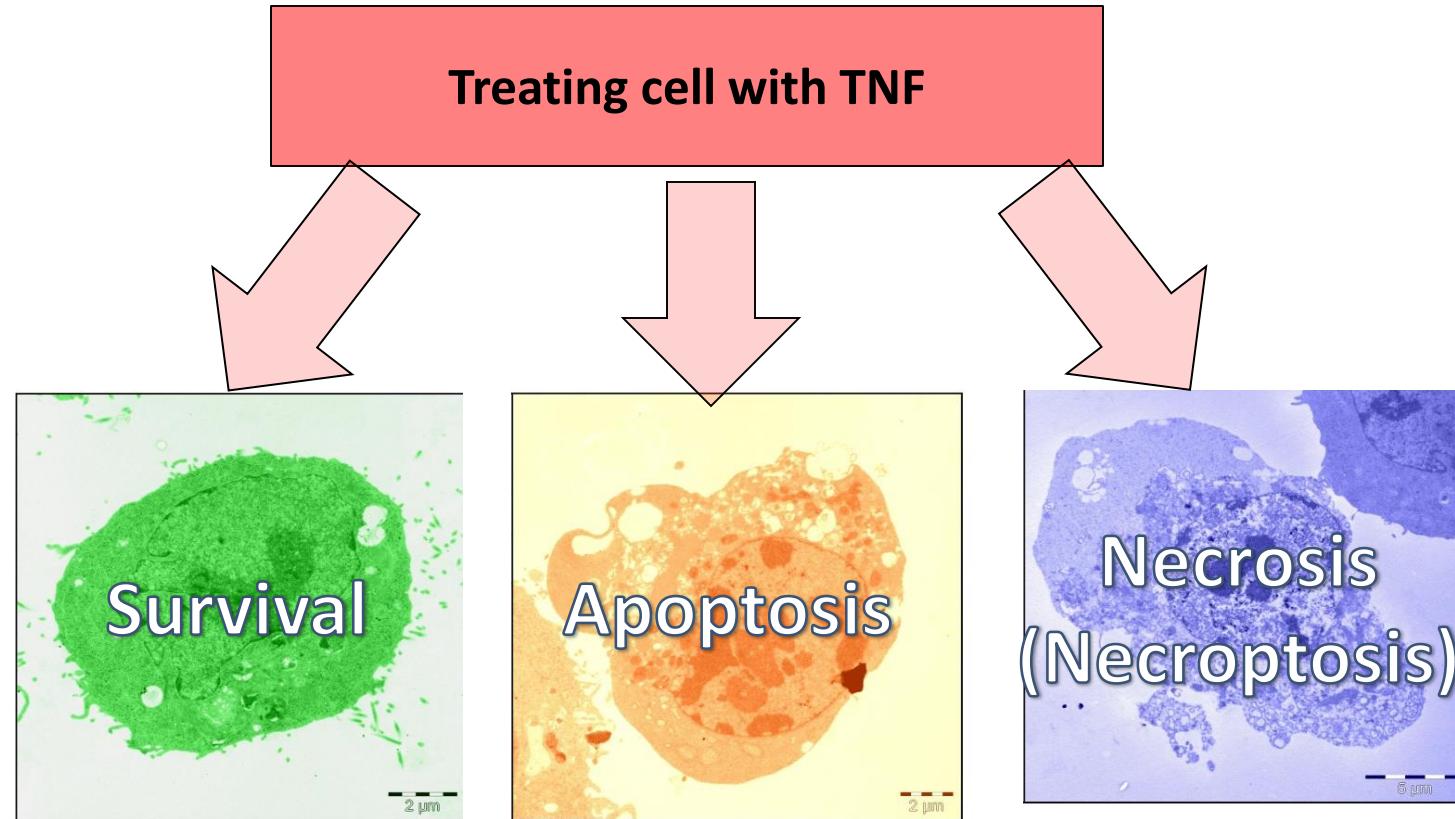
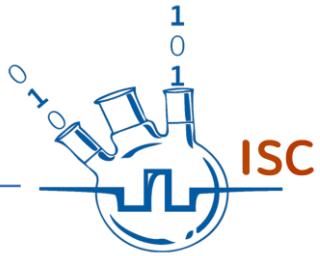
Cell Death Modalities

(From Galuzzi et al, Cell Death and Diff, 2007)





Apoptosis vs. Necrosis vs. Survival



OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Mathematical Modelling of Cell-Fate Decision in Response to Death Receptor Engagement

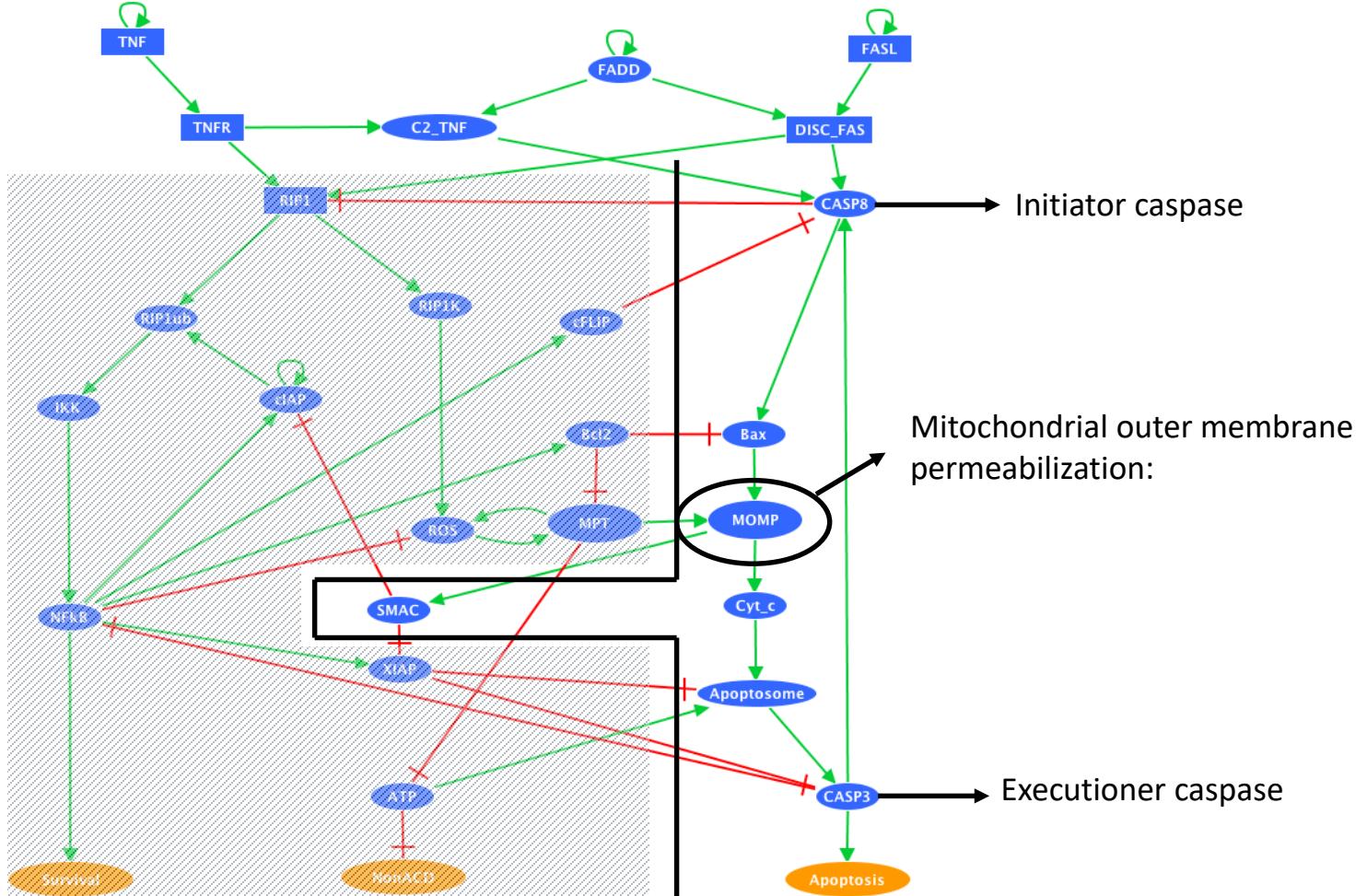
Laurence Calzone^{1,2,3*}, Laurent Tournier^{1,2,3}, Simon Fourquet^{1,2,3}, Denis Thieffry^{4,5}, Boris Zhivotovsky⁶, Emmanuel Barillot^{1,2,3†}, Andrei Zinov'yev^{1,2,3‡}

¹ Institut Curie, Paris, France, ² Ecole des Mines ParisTech, Paris, France, ³ INSERM U900, Paris, France, ⁴ TAGC – INSERM U928 & Université de la Méditerranée, Marseille, France, ⁵ CONTRAINTES Project, INRIA Paris-Rocquencourt, France, ⁶ Karolinska Institutet, Stockholm, Sweden

EMOre than a
NIVERSITY

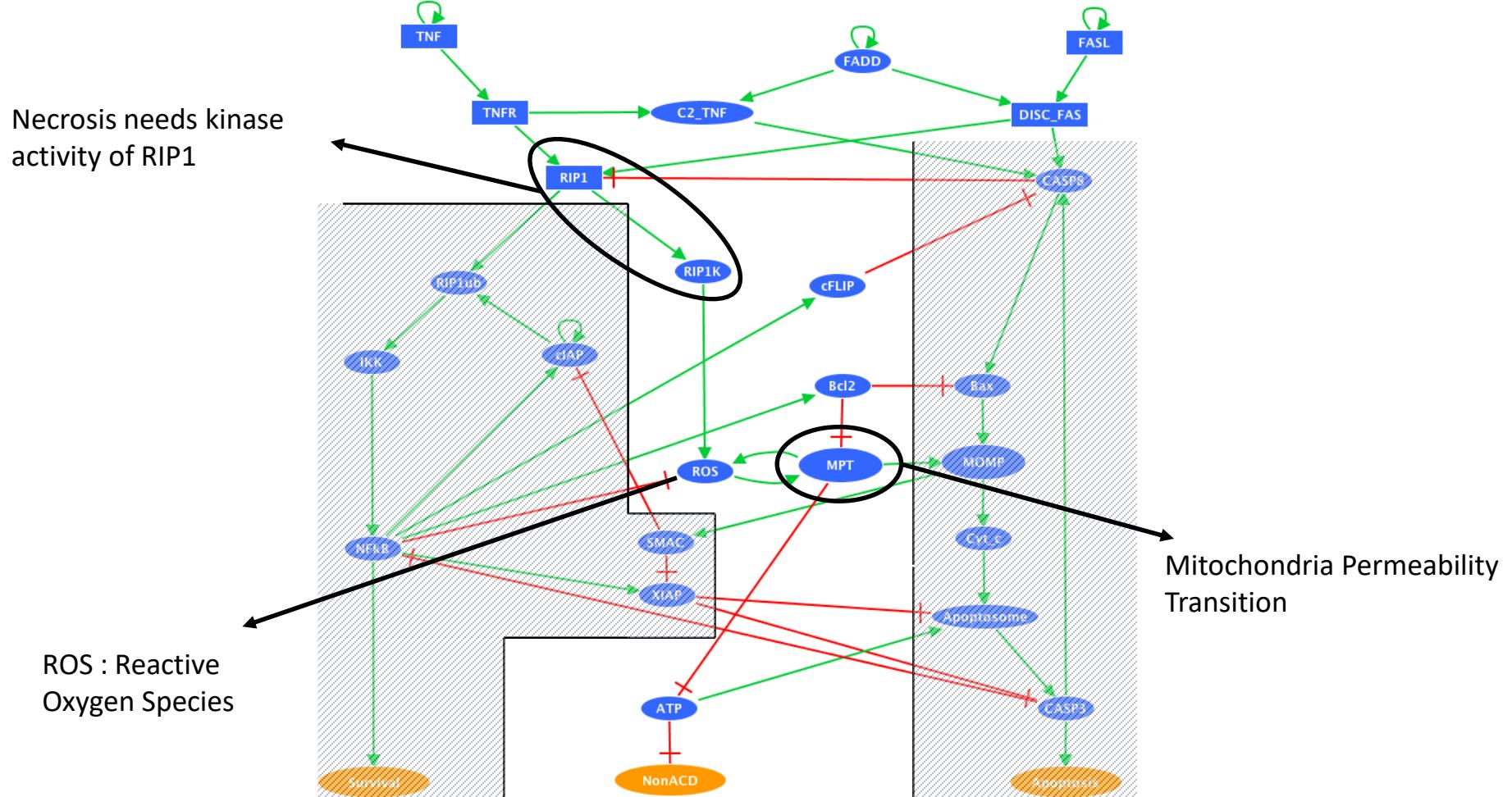
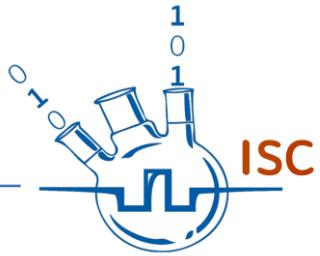


Apoptosis



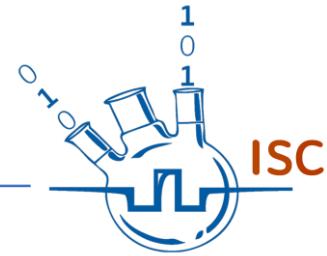


Necrosis





Boolean modeling



Назначить булеву функцію на узел

Example of CASP8

CASP8 = 0 when

DISC-Fas=0 and DISC-TNF=0 and CASP3=0

(equivalent to no external signals from death receptors
and no intracellular problems)

cFLIP=1

(equivalent to inhibition by the NFkB pathway)

CASP8 = 1 when

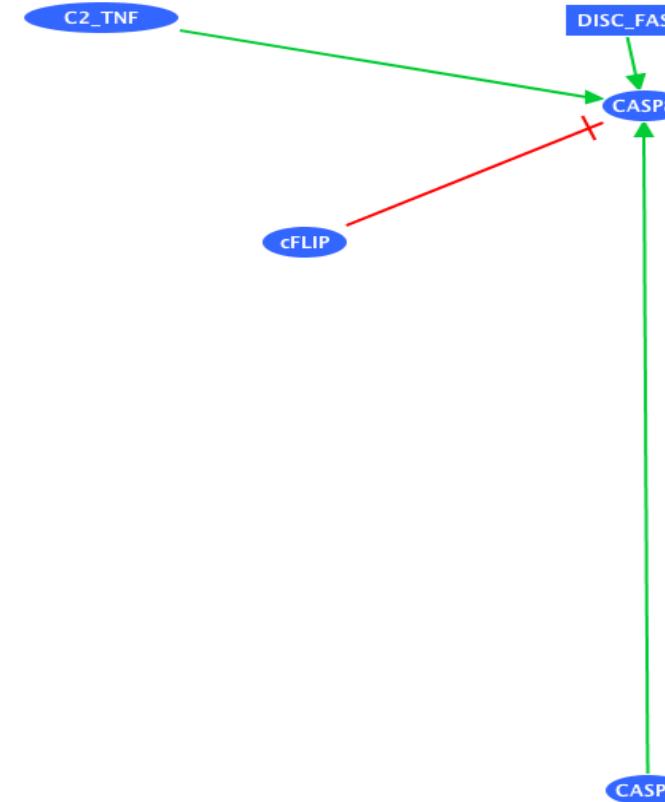
DISC-Fas=1 or/and DISC-TNF=1

(equivalent to signal from death receptors)

CASP3=1

(amplification signal, feedback activation)

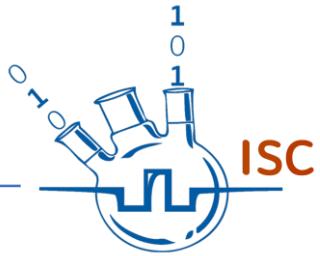
AND no cFLIP



One node = one species

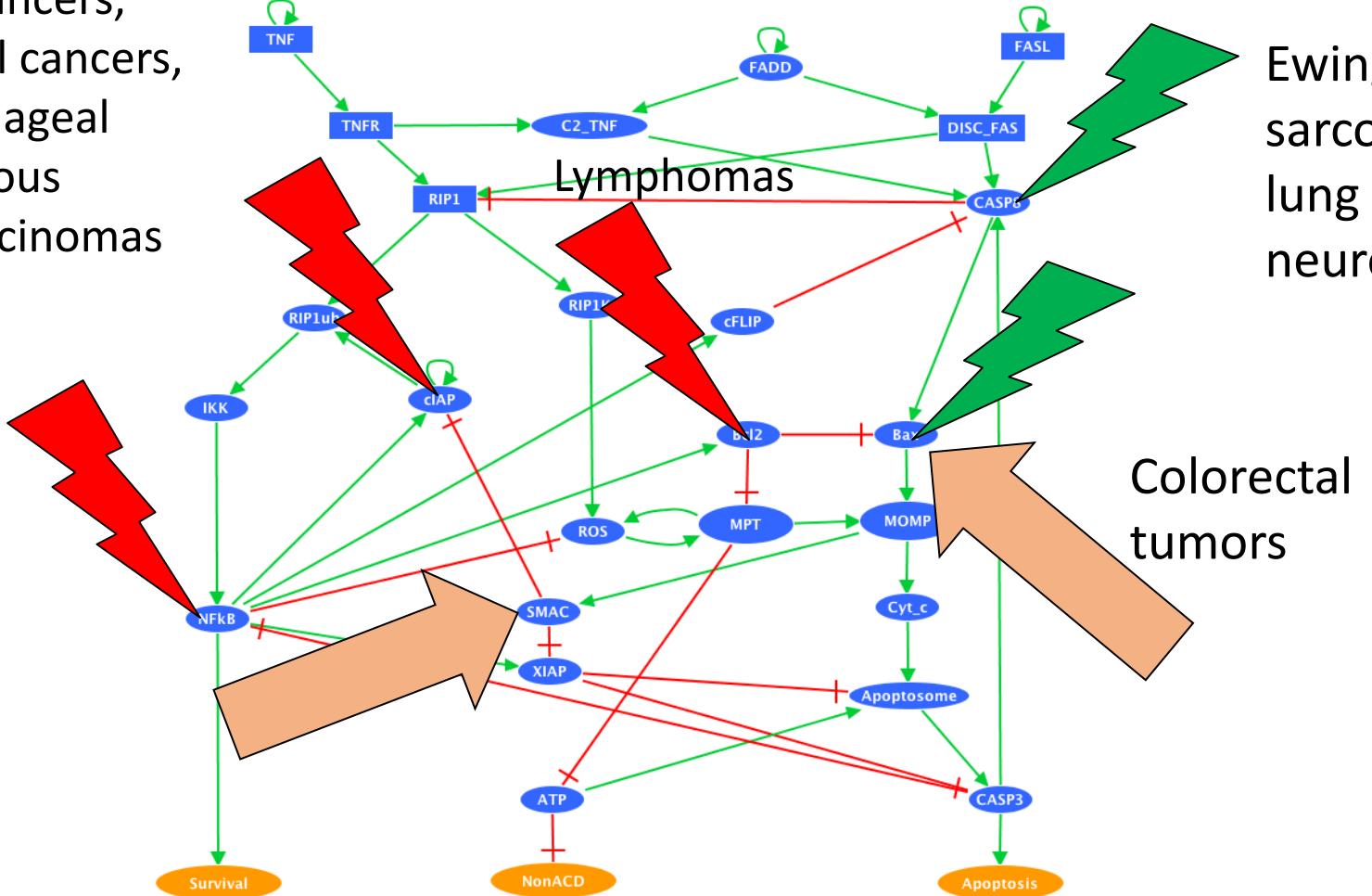


Target molecules in various tumors



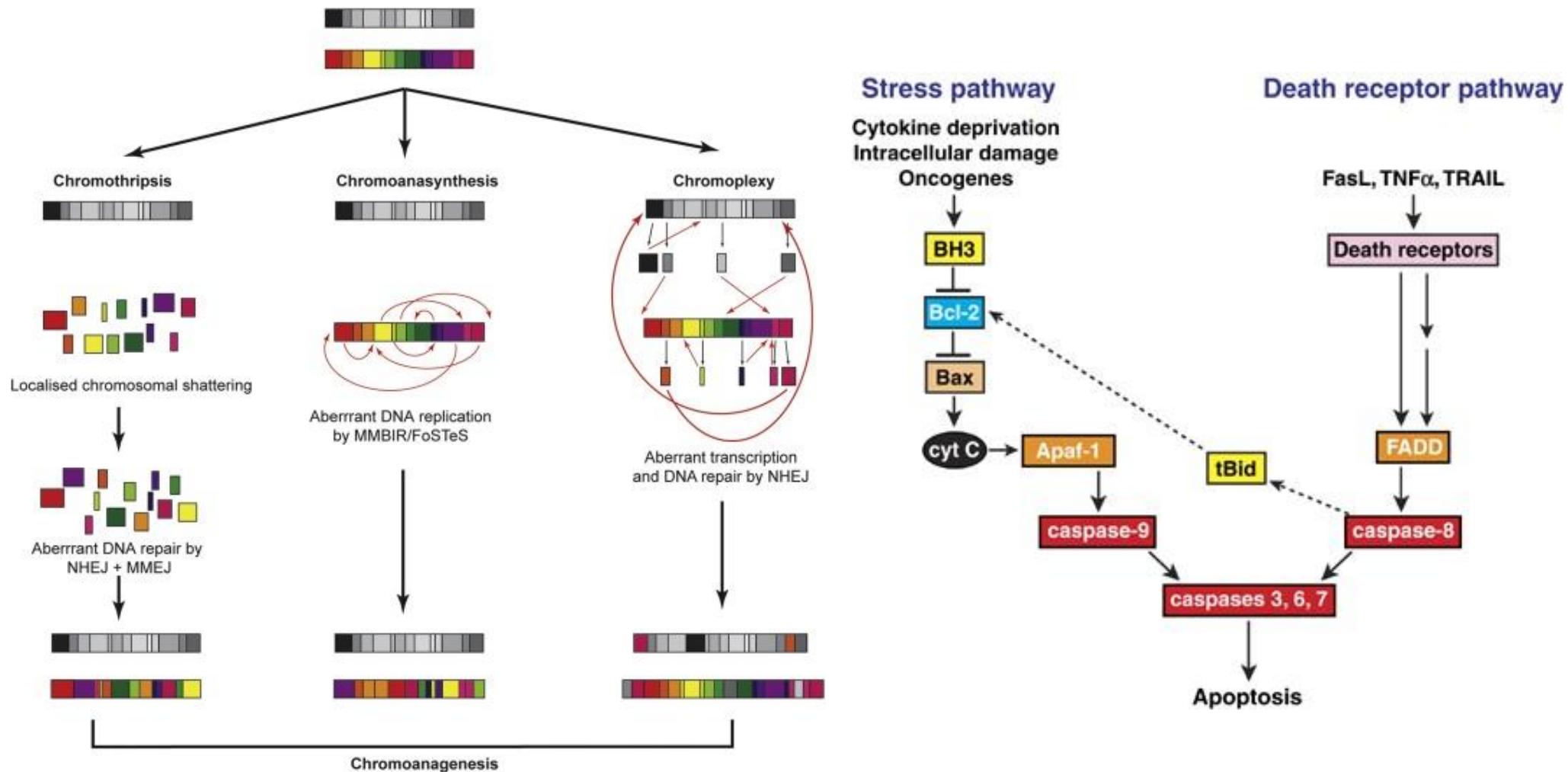
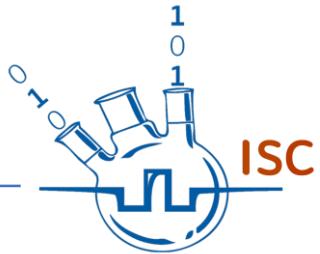
Lung cancers,
cervical cancers,
oesophageal
squamous
cell carcinomas

Lymphomas,
breast cancer





Genome chaos





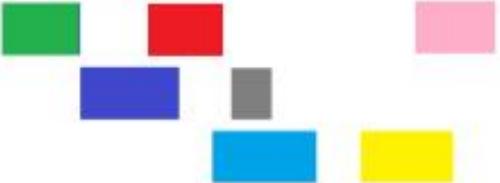
Genome chaos



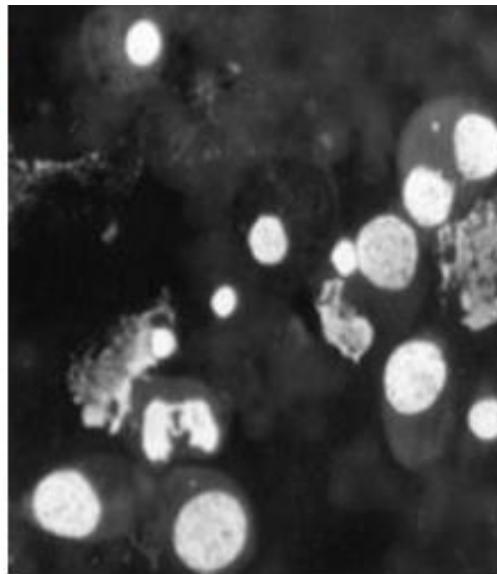
Normal chromosome



Chromothripsis

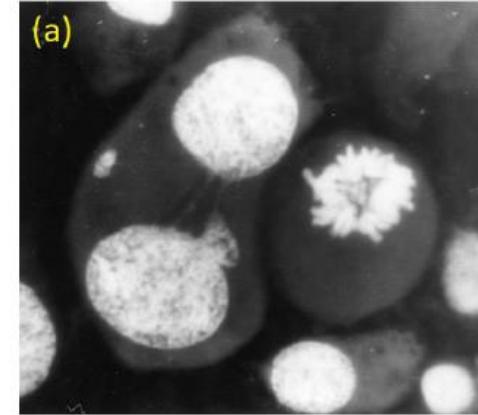


Altered chromosome



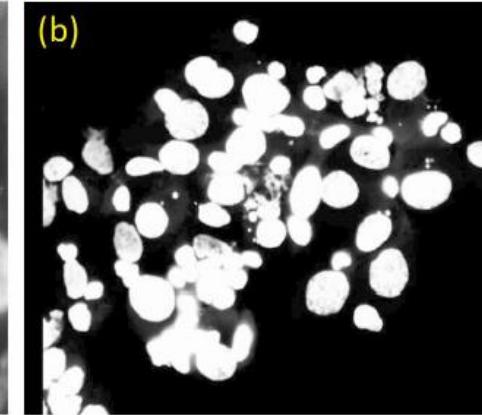
(a)

(c)



(b)

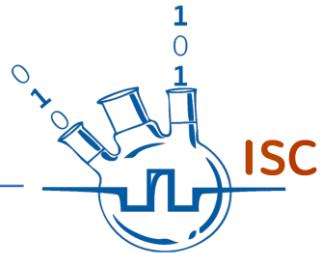
(d)



(d)



Genome chaos algorithm



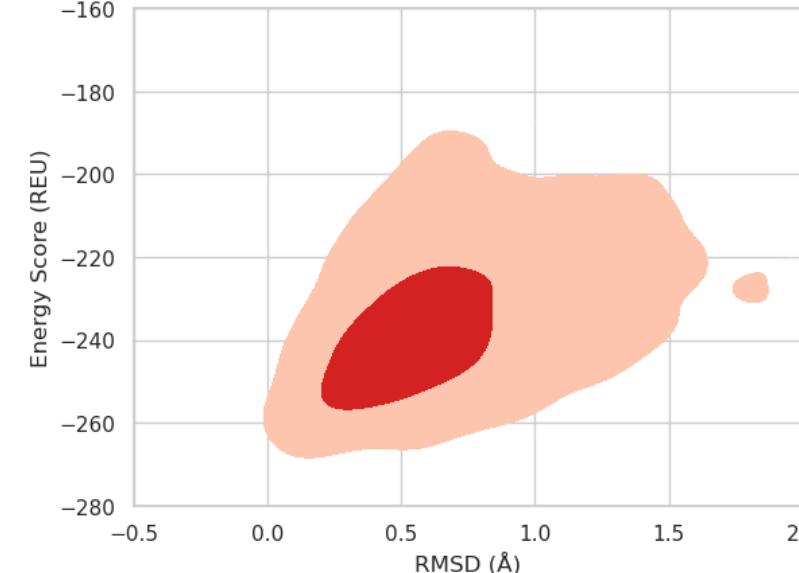
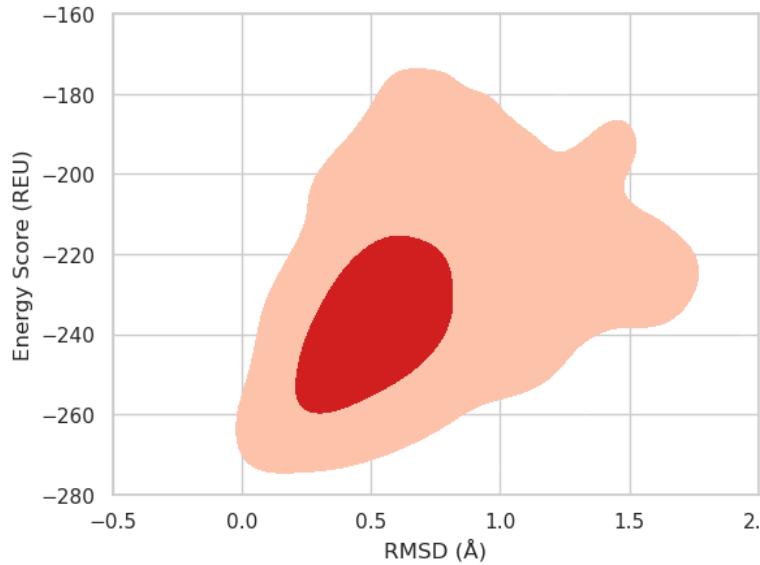
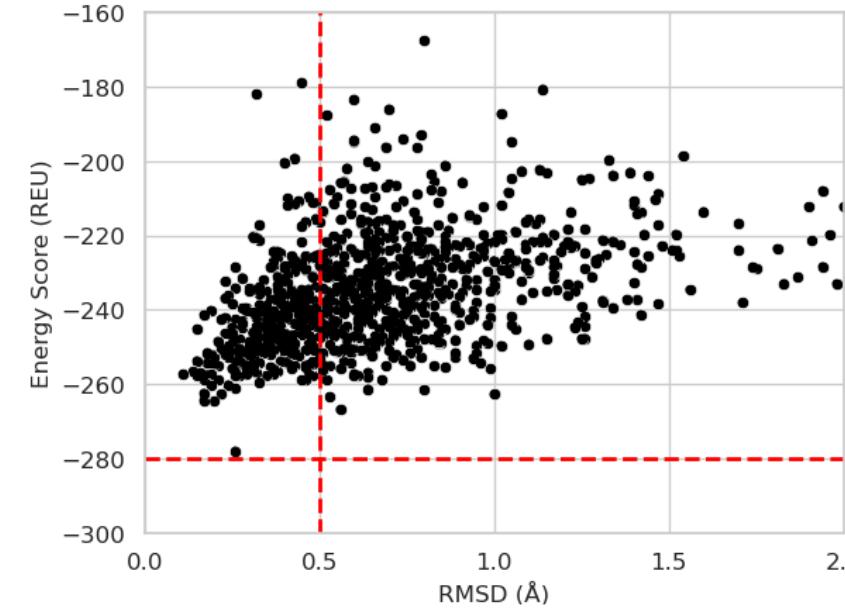
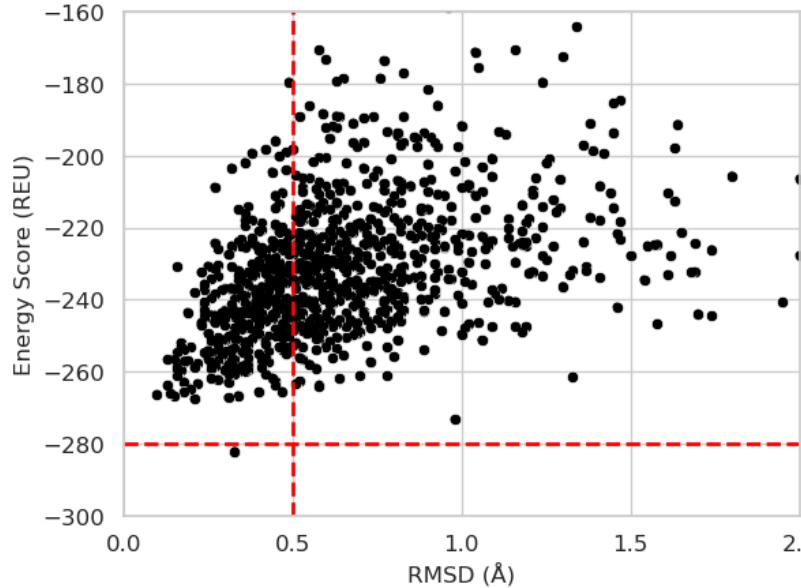
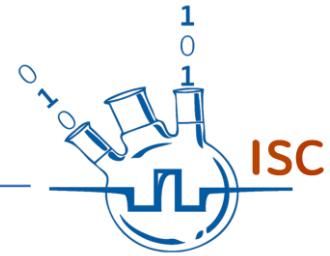
```
gene_sequence = list("ATGGCGCACGCTGGGAGAACAGGGTACGATAA  
# Calculate the number of nucleotides  
num_nucleotides = len(gene_sequence)  
  
print("Number of Nucleotides:", num_nucleotides)  
  
import random  
  
# Mutation rate  
mutation_rate = 1.0e-4  
  
# Number of cell divisions  
num_divisions = 60  
  
# Initialize the gene sequence (as a list of characters, for example)  
gene_sequence = list("ATGGCGCACGCTGGGAGAACAGGGTACGATAACCAGGAGATAGTGATGAAGTACA  
  
# Function to apply mutations  
def apply_mutation(gene_sequence, mutation_rate):  
    mutated_sequence = []  
    for base in gene_sequence:  
        if random.random() < mutation_rate:  
            # Mutate the base (for simplicity, just change to a random base)  
            mutated_base = random.choice("ATCG")  
            mutated_sequence.append(mutated_base)  
        else:  
            mutated_sequence.append(base)  
    return mutated_sequence  
  
# Simulate cell divisions and mutations  
for _ in range(num_divisions):  
    gene_sequence = apply_mutation(gene_sequence, mutation_rate)  
  
# Print the mutated gene sequence  
mutated_gene = ''.join(gene_sequence)  
print("Mutated Gene Sequence:", mutated_gene)
```

```
# Genetic code dictionary  
genetic_code = {  
    "TTT": "F", "TTC": "F", "TTA": "L", "TTG": "L",  
    "CTT": "L", "CTC": "L", "CTA": "L", "CTG": "L",  
    "ATT": "I", "ATC": "I", "ATA": "I", "ATG": "M",  
    "GTT": "V", "GTC": "V", "GTA": "V", "GTG": "V",  
    "TCT": "S", "TCC": "S", "TCA": "S", "TCG": "S",  
    "CCT": "P", "CCC": "P", "CCA": "P", "CCG": "P",  
    "ACT": "T", "ACC": "T", "ACA": "T", "ACG": "T",  
    "GCT": "A", "GCC": "A", "GCA": "A", "GCG": "A",  
    "TAT": "Y", "TAC": "Y", "TAA": "*", "TAG": "*"  
    "CAT": "H", "CAC": "H", "CAA": "Q", "CAG": "Q",  
    "AAT": "N", "AAC": "N", "AAA": "K", "AAG": "K",  
    "GAT": "D", "GAC": "D", "GAA": "E", "GAG": "E",  
    "TGT": "C", "TGC": "C", "TGA": "*", "TGG": "W",  
    "CGT": "R", "CGC": "R", "CGA": "R", "CGG": "R",  
    "AGT": "S", "AGC": "S", "AGA": "R", "AGG": "R",  
    "GGT": "G", "GGC": "G", "GGA": "G", "GGG": "G",  
}  
  
# Function to translate a DNA sequence to a protein sequence  
def translate_dna_to_protein(dna_sequence):  
    protein_sequence = []  
    for i in range(0, len(dna_sequence), 3):  
        codon = dna_sequence[i:i + 3]  
        amino_acid = genetic_code.get(codon, 'X') # 'X' for unknown codons  
        protein_sequence.append(amino_acid)  
    return ''.join(protein_sequence)  
  
# Translate the reference and mutated sequences  
translated_reference_sequence = translate_dna_to_protein(reference_sequence)  
translated_mutated_sequence = translate_dna_to_protein(mutated_gene)  
  
print("Translated Ref Gene Sequence:", translated_reference_sequence)  
print("Translated Mut Gene Sequence:", translated_mutated_sequence)
```

Protein Sequence 1: MAHAGRTGYDNREIVMKYIHYI
LTLRQAGDDFSRRYRRDFAEMSQLHLTPFTARGRFATVVEI
LFDFSWLSLKTLSSLALVGACITLGAYLGHK*
Protein Sequence 2: MAHAWRTGYDNREIVMKYIHYI
LTLRQAGDDFSRRYRRDFAEMSQLHLTPFTARGRFATVVEI
LFDFSWLSLKTLSSLALVGACITLGAYLGHK*
Sequence Identity: 99.58%

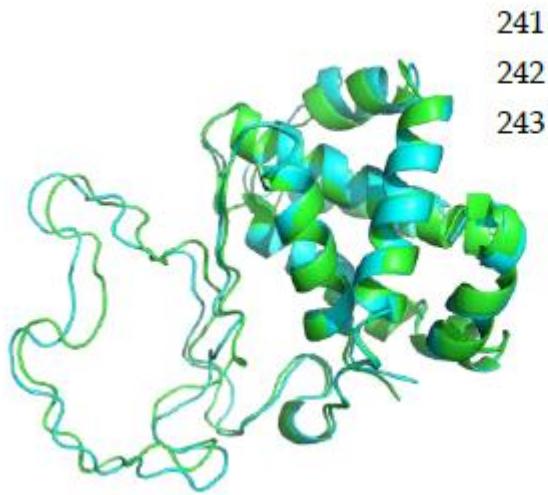
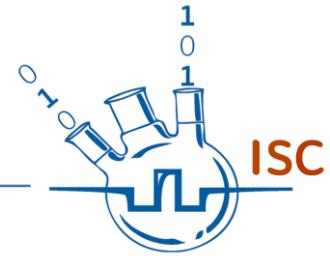


Genome chaos

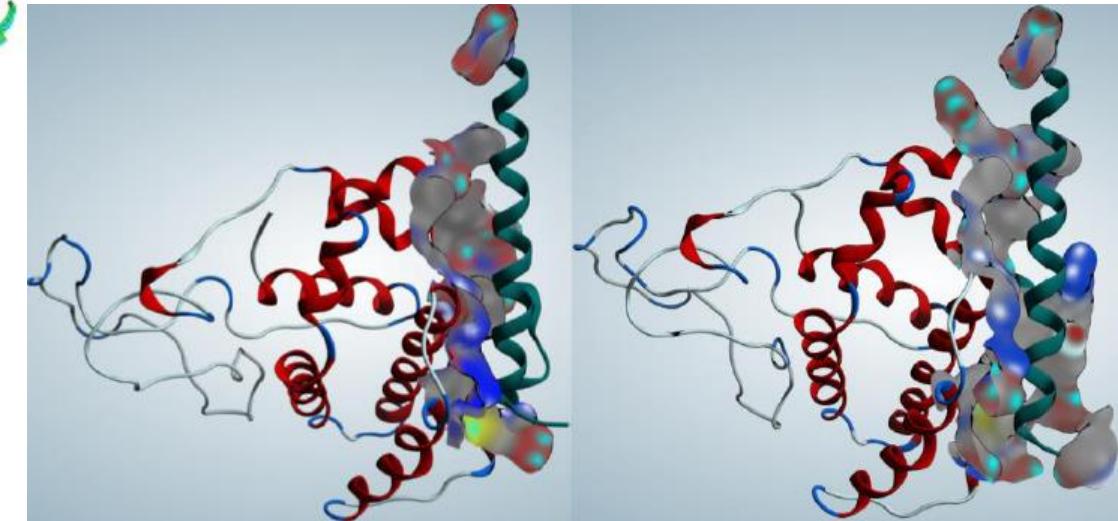
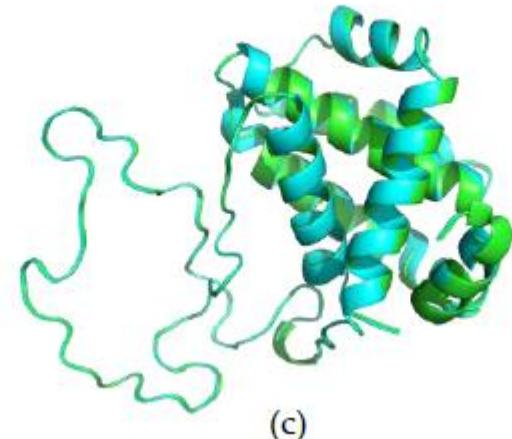
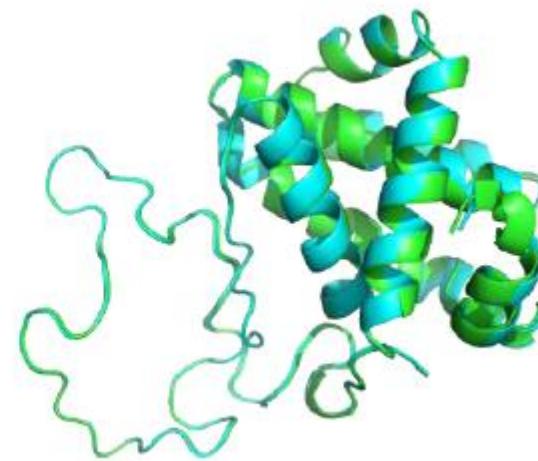




Genome chaos (BCL-2/BH3 interaction)



(a)





Genome chaos

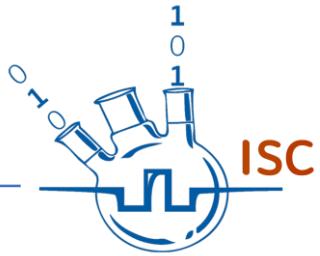
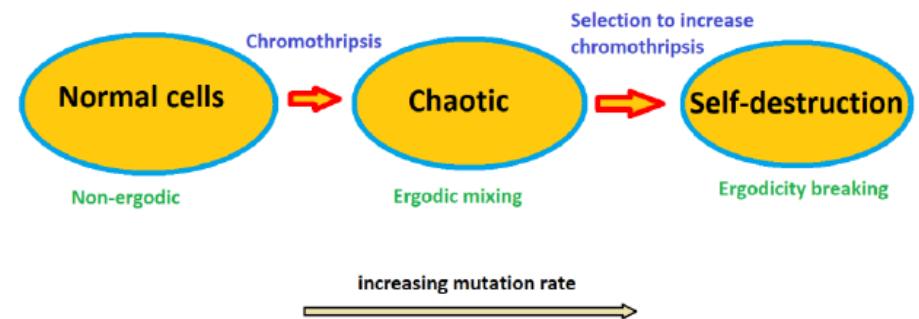
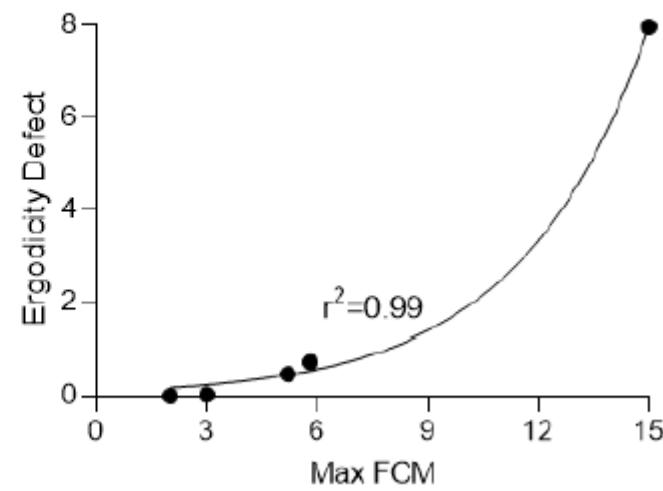
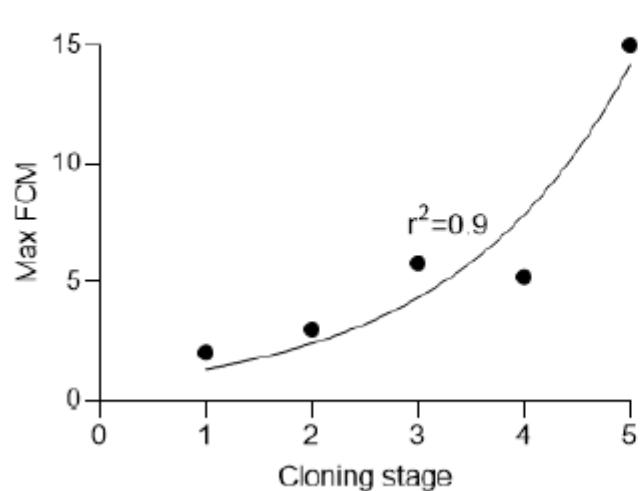
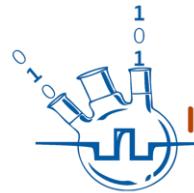


Table 1. Changes of observed minimum, average, and maximum FCM and calculated effective mutation rate and ergodicity defect at different stages of the selection for increasing FCM.

| Stage | Clones | Min FCM | Average FCM | Max FCM | Effective mutation rate, $\times 10^{-4}$ | Ergodicity defect |
|-------|--------|---------|-------------|---------|---|-------------------|
| 0 | 48 | 0.0 | 0.6 | 2.0 | 6.99 | 0 |
| 1 | 50 | 0.0 | 0.8 | 3.0 | 8.99 | 0.03 |
| 2 | 52 | 0.0 | 1.7 | 5.8 | 17.98 | 0.74 |
| 3 | 48 | 0.0 | 1.6 | 5.2 | 16.98 | 0.48 |
| 4 | 47 | 1.0 | 4.7 | 15.0 | 47.95 | 7.96 |





Thank you for your attention

