Deutsches Zentrum für Lungenforschung

# 3rd Munich Lung Conference

October 4-5, 2013


University of Sarajevo
Faculty of Electrical Engineering
Department of Computing and Informatics


# MODELING OF COMPUTER AIDED SIMULATOR IN CONTROL OF NSCLC TREATMENT BASED ON EGFR GENE MUTATIONS' ARTIFICIAL NEURAL NETWORK CLASSIFIER AND MICROARRAY EXPRESSION ANALYSIS



Prof. Zikrija Avdagic, DSc[1]
Artificial Intelligence and Bioinformatics


PhD Aida Saracevic[2]
Molecular Biology


MSc Dino Keco[1]
Computing in Bioinformatics


Ass.Prof. Dusanka Boskovic, DSc[1]
Biomedical Signals and Systems


MA Aja Avdagic[3]
Medicine


Ass.Prof. Samir Omanovic, DSc[1]
Pattern Recognition and Image Processing


Sen.Ass. Emir Buza, MSc[1]
Advanced Databases and Data Mining


Vedad Letic, MSc[4]
Calculus and Statistics


Ass. Tamer Bego, MSc[5]
Pharmacy

(1) Faculty of Electrical Engineering-University of Sarajevo, Bosnia and Herzegovina; (2) Center for Genetics, Medical Faculty-University of Sarajevo, Bosnia and Herzegovina;
(3) Faculty of Medicine-Ludwig Maximilian University of Munich, Germany; (4) Natural Sciences Faculty-University of Sarajevo, Bosnia and Herzegovina;
(5) Faculty of Pharmacy, University of Sarajevo, Bosnia and Herzegovina

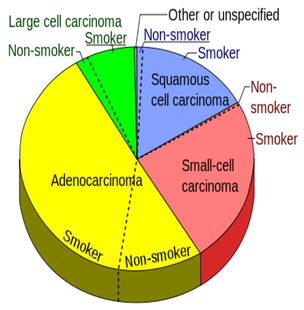

ID45

## LUNG CANCER

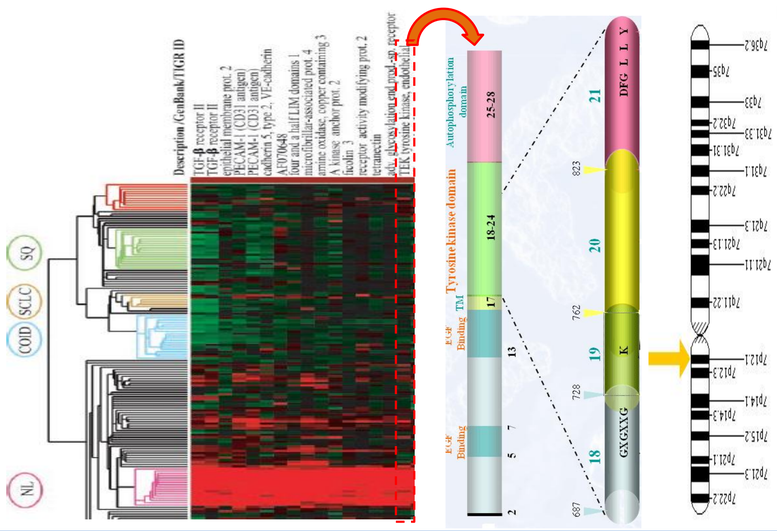
**This research has as its goal to improve diagnosis of Non-Small Cell Lung Cancer (NSCLC) based on sample patients' data with microdeletion mutations extracted from online EGFR mutation database, and samples with microdeletion mutations generated in own generator reqiured for simulation.**

### THE MAIN TYPES OF LUNG CANCER :

1. **Non-small-cell lung carcinoma (NSCLC)**
   - 1.1 Adenocarcinomas are often found in an outer area of the lung.
   - 1.2 Squamous cell carcinomas are usually found in the center of the lung next to an air tube (bronchus).
   - 1.3 Large cell carcinomas can occur in any part of the lung. They tend to grow and spread faster than the other two types.
2. **Small-cell lung carcinoma (SCLC),**
   - 2.1 Small cell carcinoma (oat cell cancer).
   - 2.2 Combined small cell carcinoma.




## GENE MICROARRAY




## EGFR EXONS



## EGFR GENE / NCBI / NG_007726

Different combinations of mutations (micro-deletions) exist within the EGFR kinase, and the most frequently observed mutations are on the exon 18, 9, 20 and 21.

**EXON 18: 159890-160012**

**MUTATIONS: 96-97 : Patient Class: 17**
CTTGTGGAGCCTCTTACACCCAGTGGAGAAGCTCCCAACCAAGCTCTCTTGAGGATCTTGAAGGAA
ACTGAATTCAAAAAGATCAAAGTGCTGGG--CCGGTGCGTTCGGCACGGTGTATAAG

**EXON 19: 160691-160789**

**MUTATIONS**: 50-64, **Patient class 1**
GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCA----------
CAACATCTCCGAAAGCCAACAAGGAAATCCTCGAT

**MUTATIONS**: 51-65, **Patient Class 2**
GGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCGTCGCTATCAAG----------
ACATCTCCGAAAGCCAACAAGGAAATCCTCGAT

**EXON 20: 167262-167447**

**MUTATIONS:25-26 : Patient Class: 15**
GAAGCCTACGTGATGGCCAGCGT---ACAACCCCCACGTGTGCCGCCTGCTGGGCATCTGCCTCA
CCTCCACCGTGCAGCTCATCACGCAGC TCATGCCCTTCGGCTGCCTCCTGGACTATGTCCG
GGAACACAAAGACAATATTGGCTCCCAGTACCTGCTCAACTGGTGTGTGCAGATCGCAAAG

**EXON 21: 177688-177843**

For example, exon 21 with convereted nucleotide G into A, and corresponding patient class 25

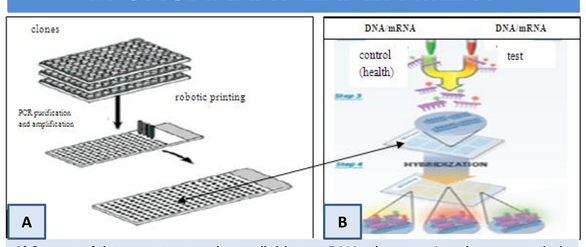
GGCATGAACTACTTGGAGGACCGTCGCTTGGTGCACCGCGACCTGGCAGCCAGGAACGTACTGGT
GAAAACACCGCAGCATGTCAAGATCACAGATTTTGGGCGGGCCAAACAGCTGGGTGCGGAAGAGA
AAGAATACCATGCAGAAGGAGGCAAA

## MICROARRAY GENE COMPARATOR

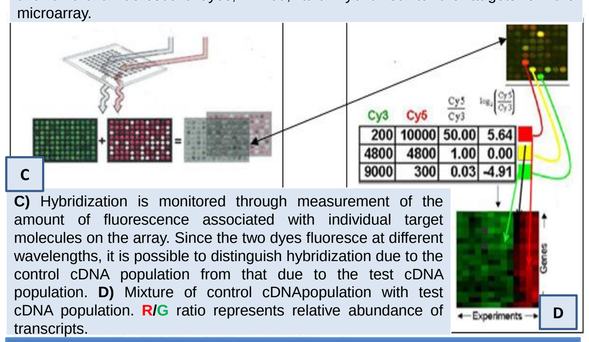
## INTELLIGENT COMPUTER AIDED SIMULATOR FOR NSCLC TREATMENT

## ARTIFICIAL NEURAL NETWORK CLASSIFIER

### MICROARRAY EXPERIMENT

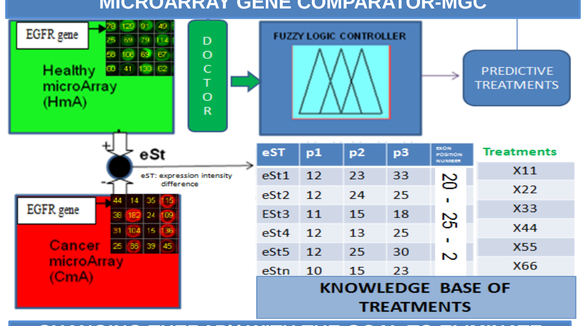



**A)**Genes of interest are made available as DNA clones. A polymerase chain reaction is used to amplify each gene to a sufficient amount to allow 'printing' onto array.**B)** RNA is isolated from reference (control) cells and test (experimental) cells. cDNA is synthesized from each RNA population by a reverse transcription process. The resulting two cDNA samples (control and experimentation) are labeled with two different fluorescent dyes, mixed, and hybridized to the targets on the microarray.

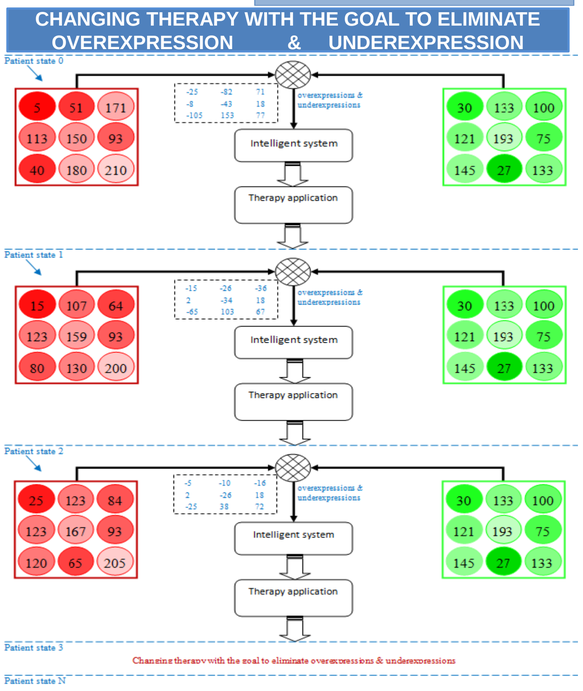


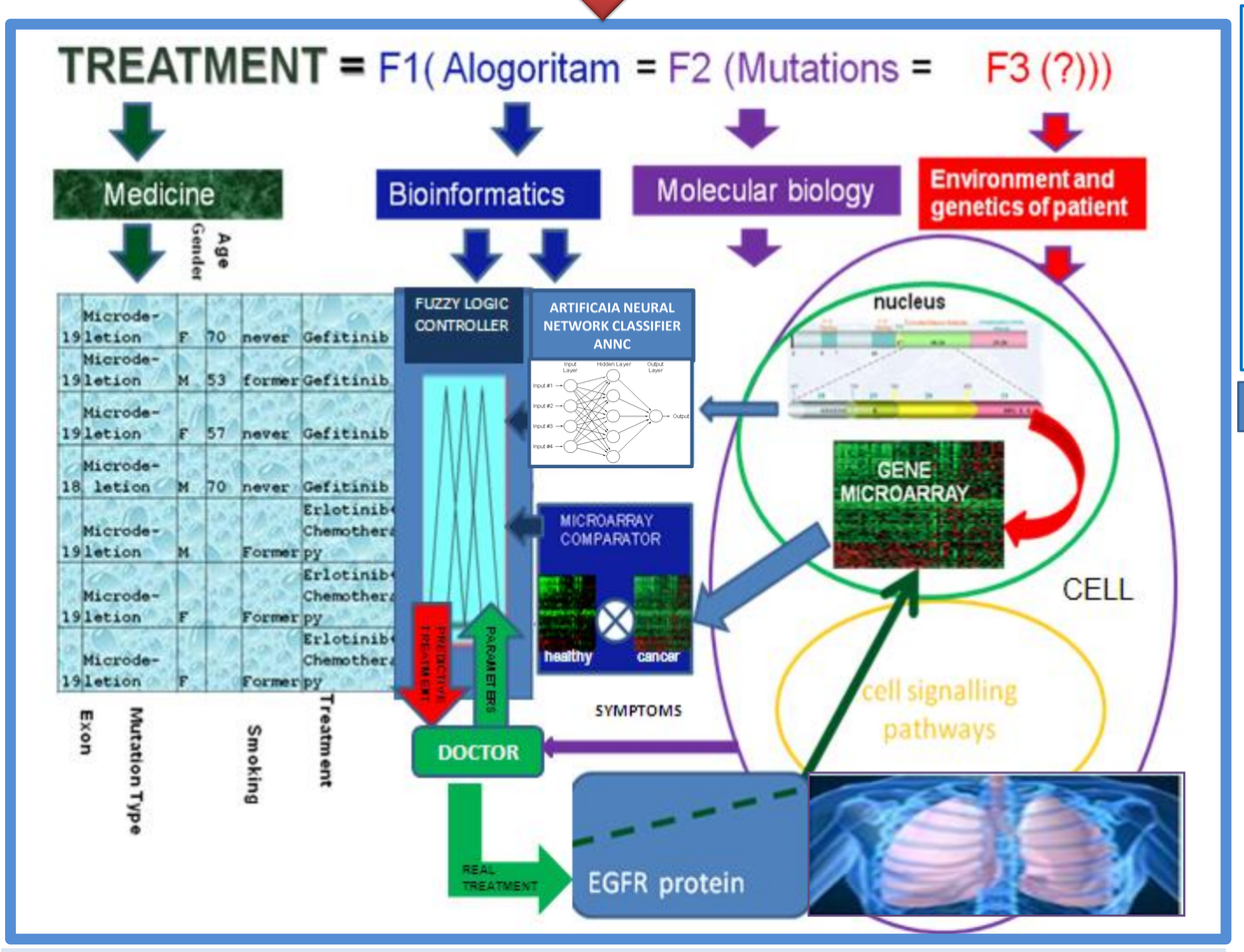

**C)** Hybridization is monitored through measurement of the amount of fluorescence associated with individual target molecules on the array. Since the two dyes fluoresce at different wavelengths, it is possible to distinguish hybridization due to the control cDNA population from that due to the test cDNA population. **D)** Mixture of control cDNApopulation with test cDNA population. **R/G** ratio represents relative abundance of transcripts.

### MICROARRAY GENE COMPARATOR-MGC




### CHANGING THERAPY WITH THE GOAL TO ELIMINATE OVEREXPRESSION & UNDEREXPRESSION




### STATISTICS FOR MUTATION GENERATOR

lung_cancer_mutation_statistics <26x8 cell>

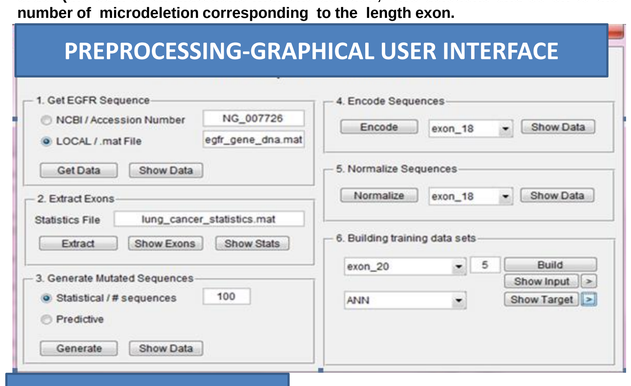
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 160691 | 160789 | 50 | 64 | 166 | 19 | '-' |
| 2 | 2 | 160691 | 160789 | 51 | 65 | 60 | 19 | '-' |
| 3 | 3 | 160691 | 160789 | 69 | 92 | 1 | 19 | '-' |
| 4 | 4 | 160691 | 160789 | 55 | 72 | 33 | 19 | '-' |
| 5 | 5 | 160691 | 160789 | 55 | 69 | 6 | 19 | '-' |
| 6 | 6 | 160691 | 160789 | 54 | 71 | 7 | 19 | '-' |
| 7 | 7 | 160691 | 160789 | 52 | 66 | 7 | 19 | '-' |
| 8 | 8 | 160691 | 160789 | 53 | 67 | 2 | 19 | '-' |
| 9 | 9 | 160691 | 160789 | 53 | 70 | 2 | 19 | '-' |
| 10 | 10 | 160691 | 160789 | 52 | 69 | 3 | 19 | '-' |
| 11 | 11 | 160691 | 160789 | 54 | 62 | 2 | 19 | '-' |
| 12 | 12 | 160691 | 160789 | 54 | 68 | 1 | 19 | '-' |
| 13 | 13 | 160691 | 160789 | 60 | 68 | 1 | 19 | '-' |
| 14 | 14 | 160691 | 160789 | 68 | 91 | 2 | 19 | '-' |
| 15 | 15 | 167262 | 167447 | 25 | 26 | 2 | 20 | '-' |
| 16 | 16 | 160691 | 160789 | 51 | 56 | 1 | 19 | '-' |
| 17 | 17 | 159890 | 160012 | 96 | 97 | 1 | 18 | '-' |
| 18 | 18 | 160691 | 160789 | 53 | 62 | 1 | 19 | '-' |
| 19 | 19 | 160691 | 160789 | 69 | 70 | 1 | 19 | '-' |
| 20 | 20 | 160691 | 160789 | 50 | 51 | 2 | 19 | '-' |
| 21 | 21 | 160691 | 160789 | 55 | 66 | 3 | 19 | '-' |
| 22 | 22 | 160691 | 160789 | 44 | 51 | 1 | 19 | '-' |
| 23 | 23 | 177688 | 177843 | 103 | 103 | 305 | 21 | 'G>T' |
| 24 | 24 | 177688 | 177843 | 112 | 112 | 7 | 21 | 'A>T' |
| 25 | 25 | 177688 | 177843 | 7 | 7 | 2 | 21 | 'G>A' |
| 26 | 26 | 177688 | 177843 | 103 | 104 | 2 | 21 | 'TG>GT' |

### MUTATION PREDICTION GENERATOR

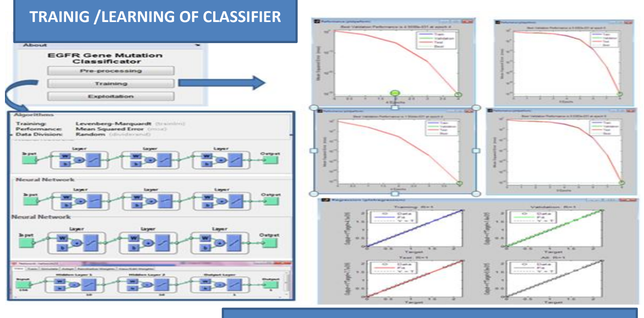
| | | |
|---|---|---|
| EXON 18 | 7627 | MUTATED NUCLEOTIDES |
| EXON 19 | 4951 | MUTATED NUCLEOTIDES |
| EXON 20 | 17392 | MUTATED NUCLEOTIDES |

**It is not enough to perform the classification of patients into one already well-known statistical group, but if there is a patient with a new mutated exons, our algorithm need to determine the position and number of mutated nucleotides. Because of that we have developed more powerfull mutation prediction generator for microdeletion mutations that take place over the nucleotides in consecutive order (the sum of all mutations with one deletion, two deletions and so on to the number of microdeletion corresponding to the length exon.**

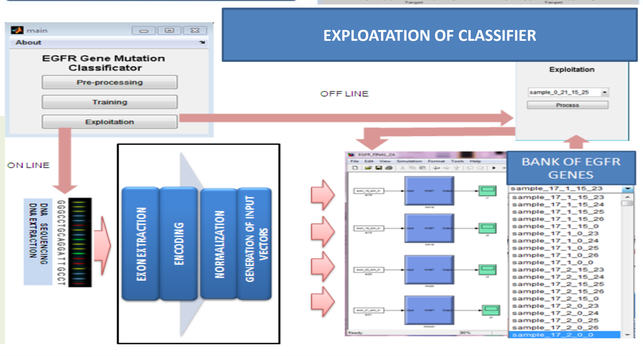
### PREPROCESSING-GRAPHICAL USER INTERFACE




### TRAINIG /LEARNING OF CLASSIFIER



### EXPLOATATION OF CLASSIFIER




The goal of this integration is to obtain a simulator that will process the appropriate inputs and generate predictive treatment. The inputs into Fuzzy Logic Controller-FLC are: (1) value of signal eSt, expression intensity difference between healthy microarray, and cancer microarray, (2)parameters from doctors (symptoms of patient), and (3) output from Artificial Neural Network Classifier-ANNC that means mutated exon, position in exon and number of mutated nucleotides. The outputs from FLC are: (1) type of therapy (for example, Gefitinib), (2)time duration of drug, (3) quantity of drug, (4) rules of behavior, (5) nutrition rules .

Knowing the principles and outputs of each module, we can now follow the functioning of our complete system. ANNC for input DNA sample reveals whether the observed gene is a healthy or a mutated one. Or, to be more precise, it identifies the exon of more of them on which mutations are present. According to mutations, the algorithm detects a certain treatment, or their combination, from the treatment knowledge base. A physician is free to accept that treatment as it is, or to adjust it to his previous

**New cancer treatments target and turn off EGFR signals. These therapies use antibodies that recognize only EGFR, stick to it, and block it from sending messages. By interrupting the signals, cancer cells are no longer told to overgrow, and eventually die. These treatments are called EGFR-targeted therapies.**

**•Gefitinib (Iressa) inhibits EGFR tyrosine kinase domain by binding to the adenosine triphosphate(ATP)-binding site of the enzyme. Thus the function of the EGFR tyrosine kinase in activating the anti-apoptotic RAS signal transduction cascade is inhibited, and malignant cells are inhibited.**

**•Erloinib (Tarceva) specifically targets the EGFR tyrosine kinase, which is highly expressed and occasionally mutated. It binds in a reversible fashion to the adenosine triphosphate (ATP) binding site of the receptor.**

## REFERENCES

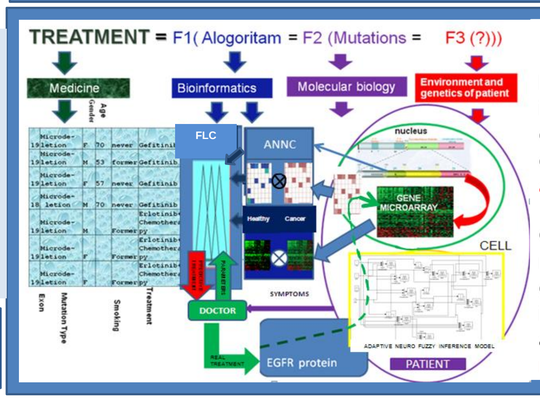

1. Melody K. Morris1,2, at all.Training Signaling Pathway Maps to Biochemical Data with Constrained Fuzzy Logic: Quantitative Analysis of Liver Cell Responses to Inflammatory Stimuli
2. E.Adetiba, J.C.Ekeh, at all., Estimating An Optimal Backpropagation Algorithm for Training An ANN with the EGFR Exon 19 Nucleotide Sequences: An Electronic Diagnostic Basis for Non-Small Cell Lung Cancer(NSCLC), JETEAS, 2011
3. G.J.Weiss, L.T.Bemis, E.Nakajima,at all.,EGFR regulation by microRNA in lung cancer: correlation with clinical and survival to gefitinib and EGFR expression in cell lines.Annals of Oncology, Februar 2008.
4. Jian Li, Vikash Pandey, Thomas Kessler, Hans Lehrach, Christoph Wierling, Modeling of miRNA and Drug Action in the EGFR Signaling Pathways, Plos ONE, January 2012
5. Bree B. Aldridge, Julio Saez-Rodriguez, Jeremy L.Muhlich at all., Fuzzy Logic Analysis of Kinase Pathway Crosstalk in TNF/EGF/Insulin-Induced Signaling, PLOS Computational Biology, April 2009.
6. Javed Khan, Jun S. Wei, Markus Ringner, at all., Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks, Nature Medicine, June 2001.
7. Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Che, Lang Cancer Cell Identification Based on Artificial Neural Networks Ensembles
8. P.K.Vaishali, Dr.A. Vinayababu, Application of Microarray Technology and Softcomputing in Cancer Biology : A Review,
9. Dennis YQ Wang, Luca Cardelli, Andrew Phillips, Nir Piterman and Jasmin Fisher, Computational modeling of the EGFR network elucidates control mechanisms regulating signal dynamics, BMC Systems Biology,2009.



**ACKNOWLEDGEMENTS**

We thank Federal Ministry of Education and Science – B&H, Federation of Bosnia and Herzegovina for support in 2012/2013


## PERSPECTIVES




**DEVELOPMENT**

**1**

**Microarray miRNA Comparator which will provide information on changes of miRNAs that target EGFR gene during the process of treatment. But there are open questions as how many targeted genes would be included and how many miRNAs we need to include?**

**2**

**Cell Signaling Pathways Module, which will enable us to monitor signals (downstream from EGFR protein to EGFR gene) through the regulatory miRNAs that target EGFR gene in continuous mode, not only in switching mode. But there are open question as how many reactions (interactions) we need to include?**

## FUTURE PLAN

**Due to combinatorial nature of all possible mutations calculated number of mutated combinations would be:**

| | | |
|---|---|---|
| EXON 18 | $2^{123}$ | $1.063 \times 10^{37}$ |
| EXON 19 | $2^{99}$ | $6.33 \times 10^{29}$ |
| EXON 20 | $2^{186}$ | $9.80 \times 10^{55}$ |
| EXON 21 | $2^{74} \cdot 3^{35}$ | $9.45 \times 10^{38}$ |

**To develop prediction mutation generator that cover all combinations of distributed nucleotides mutations related to positions in exons and number of mutated nucleotides would be very long process, and is not in line with human life. Because of that we look for a solution in new technologies (parallel multiprocessing cores), and the application of new models based on fusion of artificial intelligence methods.**