



# ***From DNA Sequencing to Drug Discovery***

About Myself





[LinkedIn](#) - **Simon Prunean** **Currently Position** Bioinformatician - Ingenuity Systems [Qiagen] Cluj-Napoca, Romania **Past** Bioinformatician- Evaxion Biotech Copenhagen, Denmark **Studies** MSc - Bioinformatics, Computational Biomedicine SDU, Odense, Denmark

## Table of Content



- **Heterogeneity among us**
- **The Central Dogma**
- **Microarray analysis**
- **Experimental Procedure in Drug Combination**

## Heterogeneity among us. Part 1

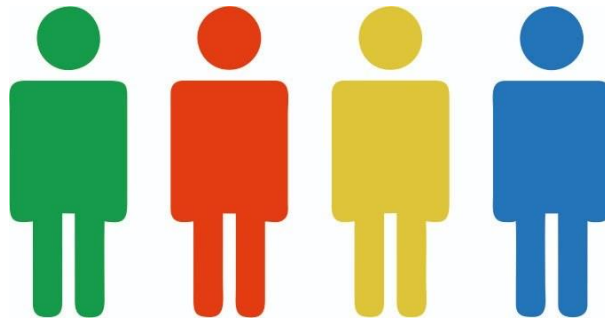




1 - Identical Twins



2 - Phenotype: 80 %

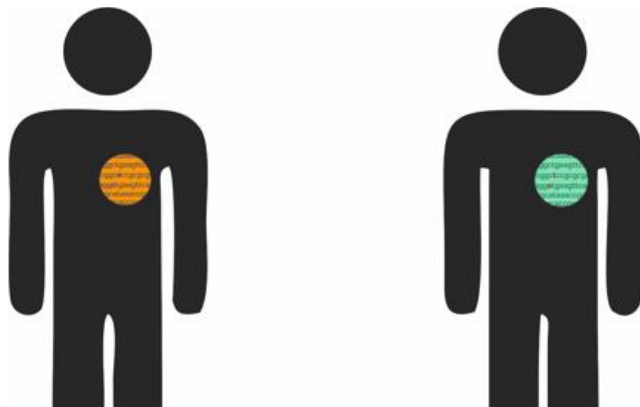


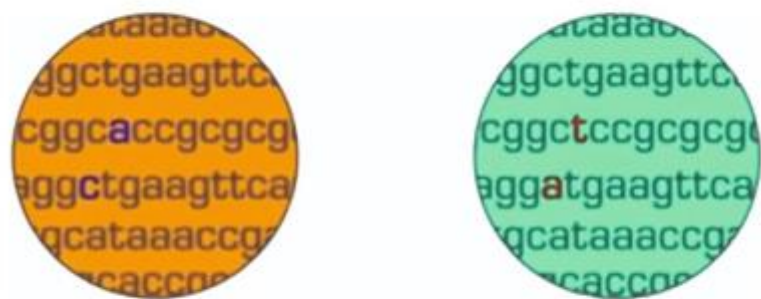
3 - Genotype: 99 %

## Heterogeneity among us. Part 2



- Same cell in 2 different humans
- 1 % dissimilarities make us unique humans
- develop personal immune system

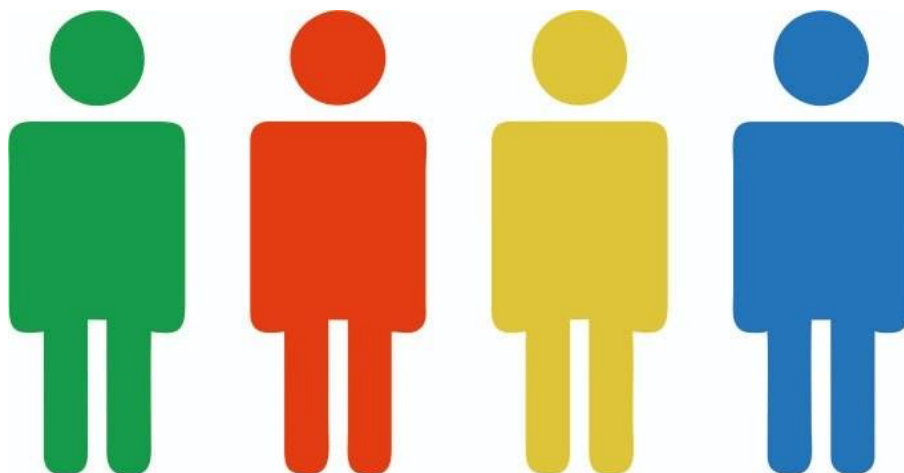




*4 - 0.5 % dissimilarities make us unique humans*

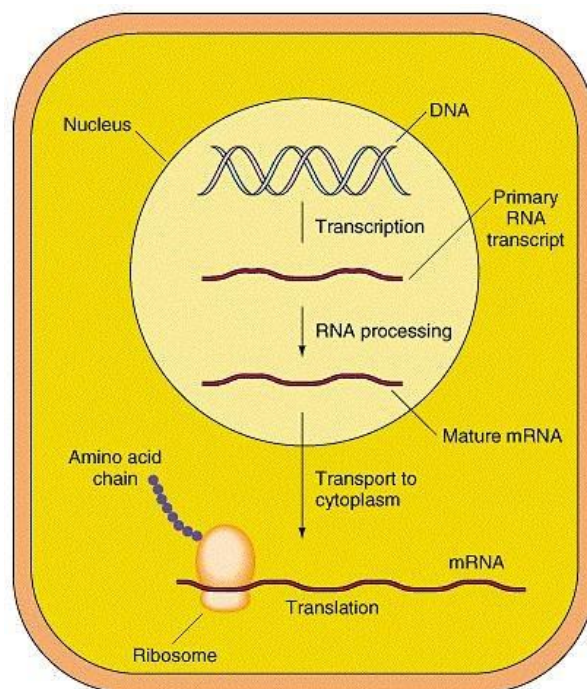


*5 - Flu Epidemic*

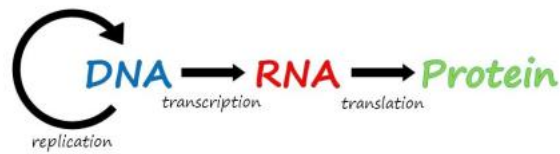




## The Central Dogma



## *The Central Dogma of Molecular Biology*



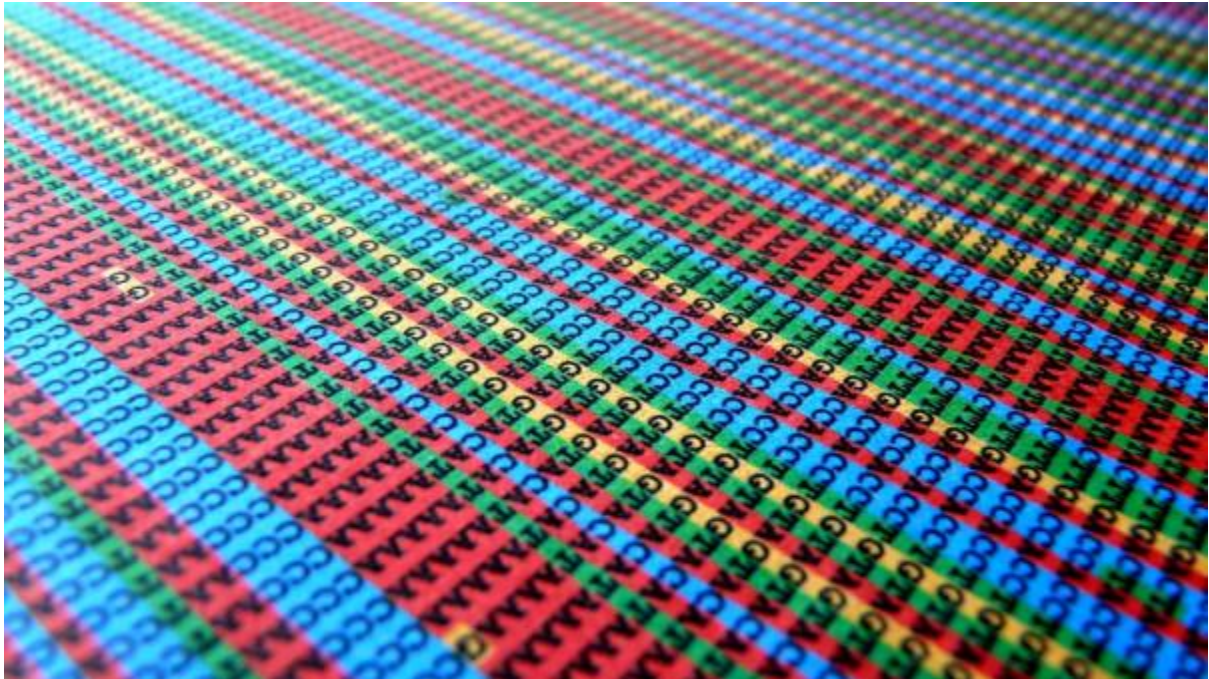
**Transcription** - the process within cell, where relevant genes for a target work are copied and compress into an RNA molecule. E.g. genes which produce immunoglobulin **Translation** - also called gene expression, is the process which the message (mRNA) is translated into proteins to perform different tasks. E.g. immunoglobulin produced in plasma cells to transport O<sub>2</sub> to all body tissues and transport CO<sub>2</sub> into lungs to be eliminated.

## *The Central Dogma of Molecular Biology*

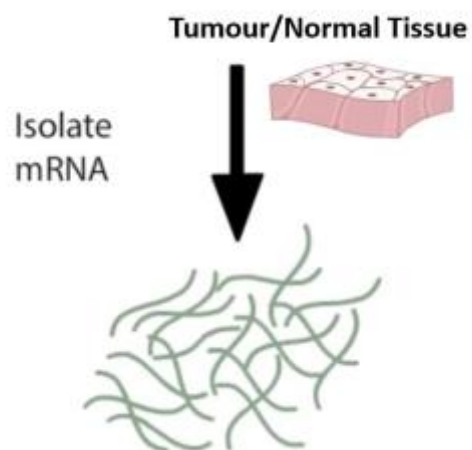




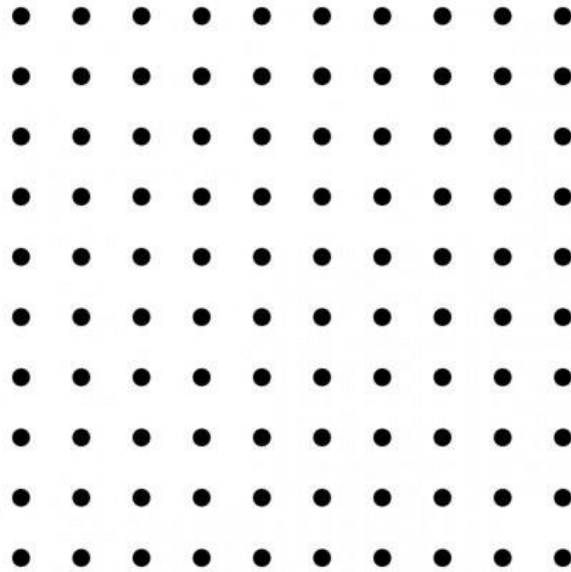
## DNA Sequencing Technologies



*6 - Lab safety!*



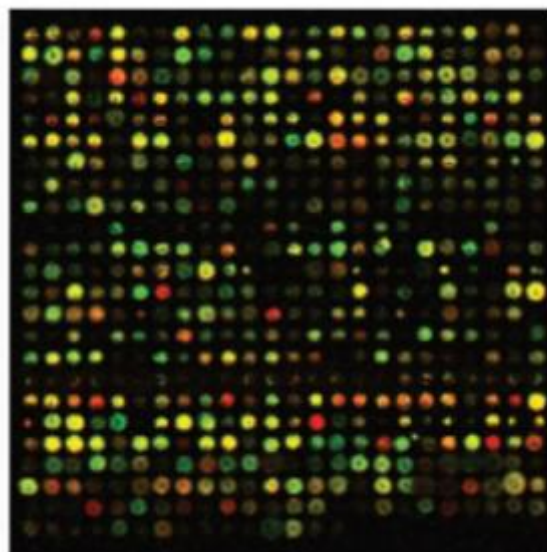
*7 - Isolate RNA and reverse transcribed to cDNA by enzyme*



8 - Empty Microarray Chip



9 - Genes - Base Pairs Spots



10 - Microarray - gene expression



11 - NGS fluorescent per nucleotide

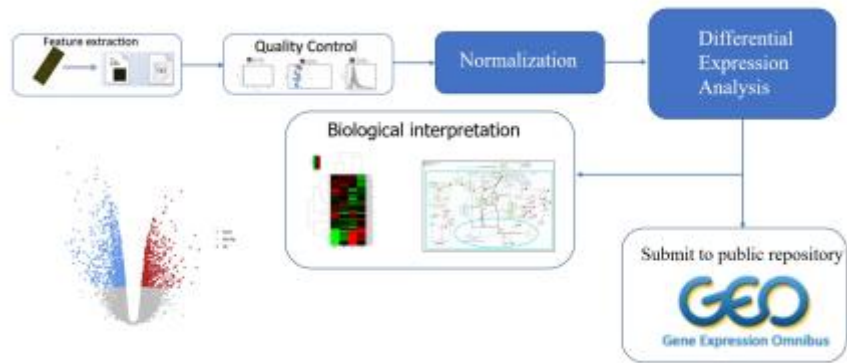
**DNA** - double stranded molecule | **RNA** - single strand molecule **Base Pairs** { A C G T }  
**Base Pairs**: { A C G U } **Reverse Transcription**: RNA --> cDNA (one strand molecule) **DNA Microarray**  
**Methodology**:

- fluorescent labelling (green - normal | red -tumour)
- > 6000 genes (probes)
- **microscope**: green/red laser store images for later analysis
- computational quantitative analysis

**NGS Technology**: PCR intervention

## Microarray Analysis





12 - Microarray analysis Pipeline

#### Data Repositories in Biomedicine



#### Experimental Procedure Part 1



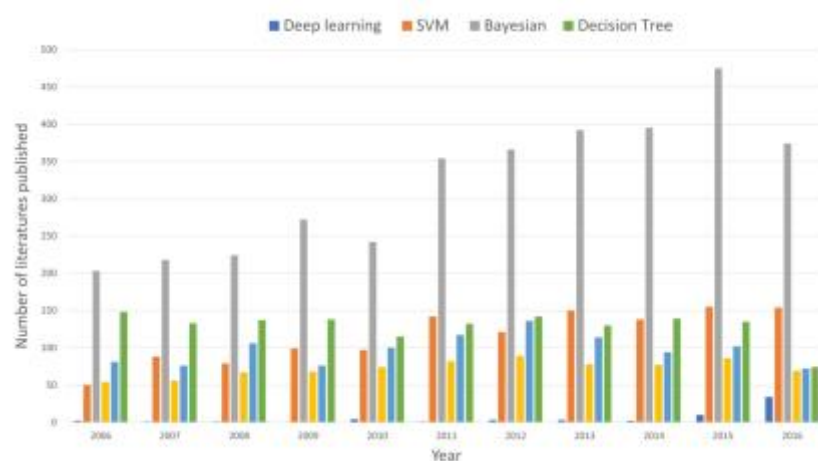


### Most targeted disease around the world

- Cancer across different tissue type
- Hepatitis C (viral) | approved cure - **Mavyret** [2019]
- HIV - Immunodeficiency virus (viral)

### How to find synergistic drug pairs? Machine Learning Methods !!!

**Loewe Score [antagonism | synergy]** Synergy : combine effect > separate effect



13 - Past 10 years performance



## DeepSynergy: predicting anti-cancer drug synergy with Deep Learning

Kristina Preuer<sup>1</sup>, Richard P. I. Lewis<sup>2</sup>, Sepp Hochreiter<sup>1</sup>, Andreas Bender<sup>2</sup>, Krishna C. Bulusu<sup>2,3</sup> and Günter Klambauer<sup>1,\*</sup>

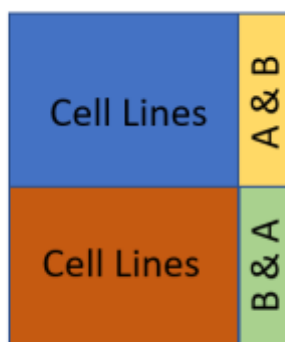
<sup>1</sup>Institute of Bioinformatics, Johannes Kepler University, 4040 Linz, Austria, <sup>2</sup>Department of Chemistry, Centre for Molecular Science Informatics, University of Cambridge, Cambridge CB2 1EW, UK and <sup>3</sup>Oncology Innovative Medicines and Early Development, AstraZeneca, Hodgkin Building, Chesterford Research Campus, Saffron Walden, Cambs CB10 1XL, UK

### 14 - Deesynergy Paper

**DeepSynergy** • Cell Gene Expression (38 cells) [[E-MTAB-3610](#)] • Synergy score: drug A & drug B (39 drugs) - [DrugCombo](#) • Artificial Neural Networks - TensorFlow • Normalization • Hidden/Output Units - ReLu • Dropouts - **Performance** - pretty well ... - neglects order of dosing - not so much data for training - [Tool](#)

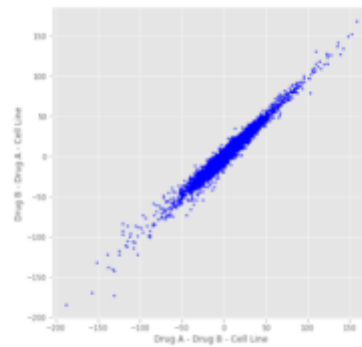
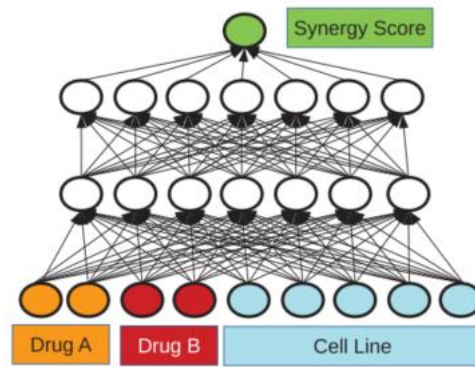
compound	target
ABT-888	PARP
AZD1775	Wee1
BEZ-235	Phosphatidylinositol-4,5-bisphosphate 3-kinase
DINACICLIB	Cyclin-dependent kinases (CDK)
GELDANAMYCIN	HSP90
L778123	Farnesyltransferase/ GGPTase-I (FTI/GGTI)
MK-2206	Protein kinase B (AKT)
MK-4541	Anti-androgen
MK-4827	PARP
MK-5108	Aurora kinase A
MK-8669	mTOR
MK-8776	Checkpoint kinase 1 (Chk1)
MRK-003	$\gamma$ -secretase

### 15 - Part of existing Drugs used



### 16 - Data Set

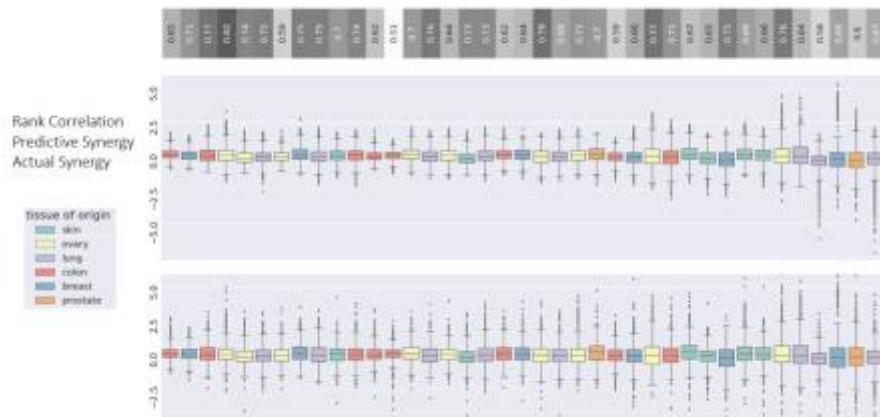




17 - Neglecting: Drug1\_Drug2 VS Drug2\_Drug1

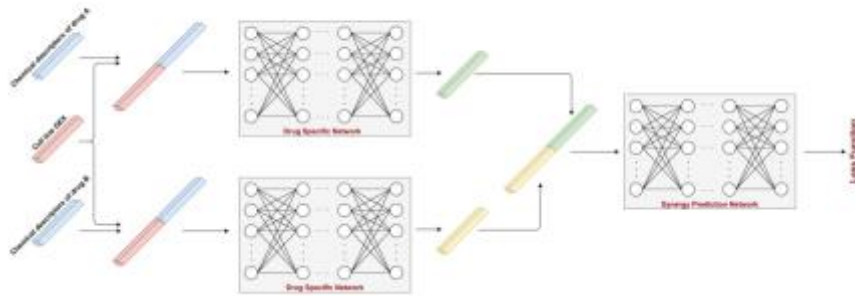
## Experimental Procedure Part 2



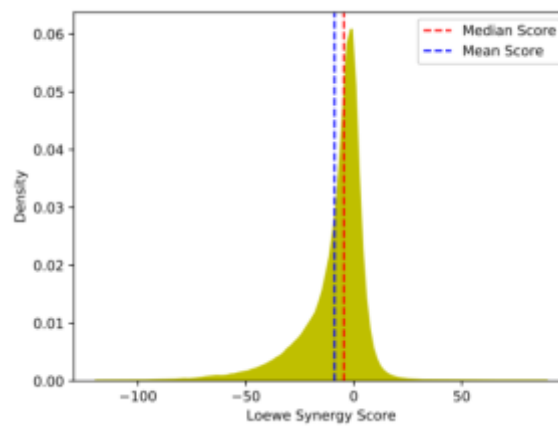


18 - Increasing Performance

### MatchMaker – schema (University, Ankara, Turkey [2020])



19 - MatchMaker: A Deep Learning Framework for Drug Synergy Prediction [2020] - Regression



20 - Loewe Distribution for current drug combinations

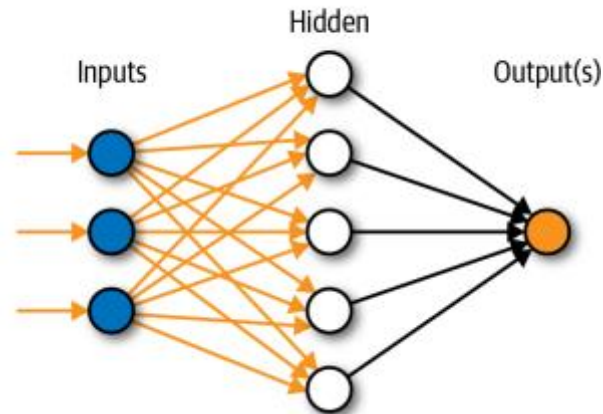
### Experiment- covers 3040 drugs | 81 cancer cell lines

- **untreated cancer cell lines** from > 10 different tissue types
- compound (physical and chemical features) | **chemopy**
- genomic information (gene expressions)
- **Loewe Synergy** - [DrugComb](#)

Dataset: [drug1\_drug2\_cell line]

**Evaluate with HTS** (High-throughput screening) - genes that modulate a particular biomolecular pathway- **83 %** two drug combinations - > 110k synergetic measurements [2020]

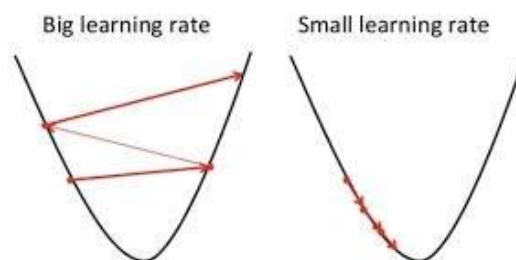
### Artificial Neural Network



21 - Neural Network Schema

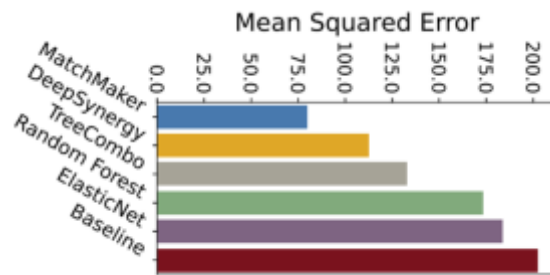


22 - Validation

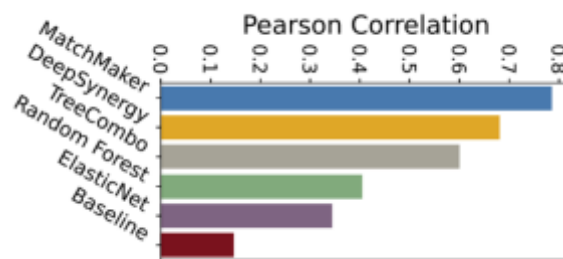


23 - Learning Rate

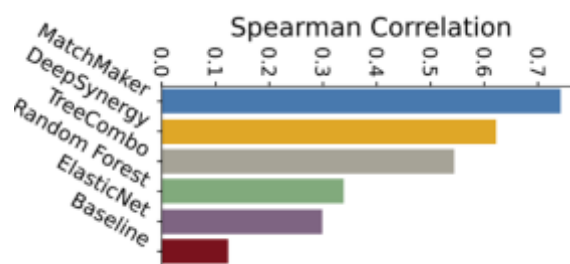
•**Normalization** • ANN – trio subnetworks • 2 subnetworks – representation of each drug conditioned • Same architecture: 3 fully connected layers (ReLU | Linear) • Last Net – continuity of firsts 2 nets • Same architecture (ReLU | Linear) • **Weights**: “Loewe Score Distance” – each point to min(Loewe)



24 - MSE



25 - Pearson



26 - Spearman

Model	MSE	Pearson	Spearman
DeepSynergy	0.112	0.66	0.66
TreeCombo	0.132	0.68	0.59
Bayesian	0.202	0.14	0.12
MatchMaker	0.079	0.79	0.74

27 - Results

**Classification label: -20 antagonism | 20 synergy Performance: 97% accuracy | 92% Future Perspectives - determine drug-disease relationships**

Model	AUC	Precision	Recall
MatchMaker	0.97	0.92	0.85
DeepSynergy	0.92	0.85	0.70
TreeCombo	0.89	0.79	0.61

The End



Simon Prunean

Bioinformatician - Qiagen

Q&A