tactiq.io free youtube transcript
A community computational challenge to... - Gustavo Stolovitzky - Highlights - ISMB/ECCB 2015
https://www.youtube.com/watch/0SuHHQJlmrY

00:00:01.120 so hi everybody thank you for uh
00:00:03.919 coming to this talk i'm going to uh
00:00:06.399 report
00:00:06.960 on a challenge that we did with the
00:00:08.960 dream and ends
00:00:10.160 the nci on predicting drug synergy
00:00:15.120 um so first let's talk a little bit
00:00:17.440 about synergistic interactions between
00:00:19.359 drugs
00:00:20.320 um the reason why people and
00:00:23.920 doctors are in need for this kind of
00:00:27.119 cocktail therapy is because we need to
00:00:29.840 uh
00:00:30.800 reduce the resistance to treatment that
00:00:32.880 almost inevitably targeted therapy
00:00:35.040 eventually
00:00:35.680 uh creates the after time of
00:00:38.879 remittance uh the cancers tend to
00:00:41.360 relapse
00:00:42.320 improve the overall survival or decrease
00:00:44.399 the compound the independent compound
00:00:46.160 compound dose the mechanisms of actions
00:00:49.600 of drug combinations are
00:00:51.039 various and uh could be two drugs acting
00:00:54.800 on the same molecule like for example
00:00:56.719 trastuzumab which is herceptine and
00:00:58.320 lapatinib that both act on
00:01:00.160 her to for her two positive breast
00:01:02.239 cancer patients
00:01:03.680 uh acting on the same pathway for
00:01:05.519 example all acting on dna damage repair
00:01:08.479 or acting different pathways
00:01:10.000 synergistically for example
00:01:12.320 on dna damage and a cell cycle
00:01:17.360 so there is no one way of doing it there
00:01:19.280 are more ways of
00:01:20.560 interactions that we don't know and
00:01:24.000 most importantly in order to be
00:01:27.040 translationally
00:01:28.240 useful it is necessary to do
00:01:31.759 a screening of a lot of combinations
00:01:33.600 that probably are
00:01:34.799 synergistic but we don't know and in

00:01:36.960 order to do that you have to
00:01:38.720 combine pairs triplets
00:01:41.759 of a lot of drugs and if you have 1000
00:01:44.799 drugs and you
00:01:45.520 do n square you will have 1 million so
00:01:47.600 it doesn't make sense you have to get
00:01:49.119 some
00:01:50.159 way of prioritizing this by using
00:01:53.600 some in silico ways and so basically the
00:01:57.360 dream challenge was set up
00:01:58.719 in order to try to benchmark
00:02:02.000 methods to do these kinds of
00:02:06.399 triaging of of the possible n square
00:02:09.199 problem
00:02:10.000 and um um
00:02:13.040 in a way we are not trying to predict
00:02:14.720 compounds we are trying to find methods
00:02:16.319 that do predict compounds
00:02:17.680 that uh work synergistically so the
00:02:20.239 challenge was
00:02:21.360 um around 2012 actually the same year
00:02:24.319 that
00:02:24.800 the challenge that robert just mentioned
00:02:27.680 the
00:02:28.160 the price for life als challenge
00:02:31.920 this challenge um was more or less
00:02:35.200 organized as following we we had
00:02:37.440 14 drugs for which we had some data
00:02:40.800 and we made all the combinations of
00:02:43.120 these 14 drugs which are 91 combinations
00:02:45.519 and we asked patie
00:02:46.640 we asked participants can you rank
00:02:49.760 the combinations from the most
00:02:51.040 synergistic to the least synergistic and
00:02:53.680 eventually because we had the measure of
00:02:55.360 which ones were the most synergistic we
00:02:57.519 could
00:02:58.159 evaluate the predictions all these were
00:03:01.680 done on one cell line
00:03:03.440 this diffuse large b cell lymphoma cell
00:03:06.159 line called
00:03:07.519 ly3 so the data was pretty rich
00:03:11.519 very interesting data set
00:03:14.879 unique at the time that we was provided
00:03:17.519 we had
00:03:18.239 uh three time points for which we have
00:03:21.200 um
00:03:21.760 a gene expression arrays we had three

00:03:24.879 replicates per time points
00:03:26.799 and we had
00:03:29.920 two concentrations in which we uh gave
00:03:32.720 them
00:03:33.360 and they we gave the doc those response
00:03:36.720 for the individual monotherapies and we
00:03:38.959 gave
00:03:40.080 um genomic so by the by the way the gene
00:03:43.760 expression data was for
00:03:45.360 before treatment and after treatment and
00:03:48.239 we also had the copy number
00:03:49.840 uh variant variation
00:03:52.879 of the cell line as well as the single
00:03:55.519 nucleotide polymorphisms for that same
00:03:58.080 cell line so we gave all these data and
00:04:00.400 we asked
00:04:01.360 participants can you predict which are
00:04:03.920 the synergistic genes
00:04:06.560 how do we measure synergy this is a
00:04:08.560 little bit of a contentious
00:04:10.720 subject in the field there are many ways
00:04:12.640 of measuring synergy there is no
00:04:14.159 one best way it depends on what you want
00:04:16.079 to ask of the system
00:04:17.839 we use what is called excess over bliss
00:04:20.639 independence
00:04:22.079 if you call v the viability that means
00:04:23.840 the fraction of survive is
00:04:25.280 surviving cells after treating the cells
00:04:28.560 and i the inhibition which is basically
00:04:32.000 one minus the viability the fraction
00:04:33.680 that was inhibited
00:04:35.040 so bliss independence that means
00:04:36.800 assuming that the two drugs are not
00:04:38.240 synergies not synergistic
00:04:40.320 basically can be thought
00:04:42.000 probabilistically as the fraction the
00:04:44.240 probability that a cell survives
00:04:46.400 with two uh drugs is the probability
00:04:48.960 that it survives with one drug
00:04:51.120 times the probability that it survives
00:04:52.560 with the other
00:04:54.320 we are not saying that this is the way
00:04:56.400 uh it has to be computed
00:04:57.919 this is the baseline over which if we
00:05:01.120 have more or less viability we want to
00:05:02.960 call
00:05:03.840 uh the drug combinations synergistic or

00:05:06.479 antagonistic
00:05:07.520 another way of doing it and it's useful
00:05:09.280 for for the next slide so i want to say
00:05:11.520 is basically do one minus
00:05:14.320 the viability of the pair call it iab
00:05:17.600 and that can be interpreted as the
00:05:20.080 probability of being inhibition
00:05:21.680 inhibited by one drug
00:05:23.840 plus the probability that is inhibited
00:05:26.479 by the other given that it was not
00:05:28.240 inhibited by the first and so we will
00:05:29.759 discuss this in a second
00:05:32.240 basically what we are trying to do a
00:05:33.919 little bit more graphically
00:05:35.440 is if we have the drug response and we
00:05:37.680 have a given dose which is called the
00:05:40.080 ic20 concentration
00:05:42.240 the concentration at which each each of
00:05:44.160 the two drugs has 20
00:05:46.000 inhibition we estimate what is the
00:05:49.120 bliss additivity and if we have
00:05:52.320 more inhibition we say that the two
00:05:54.400 drugs are synergistic
00:05:56.400 and if we have less inhibition we say
00:05:58.960 that the two drugs are antagonistic one
00:06:01.039 goes against the other
00:06:03.039 so basically the excess of our bliss is
00:06:05.759 the inhibition that results from the
00:06:07.600 pair
00:06:08.080 minus the inhibition that you would
00:06:09.600 expect if they were independently acting
00:06:13.680 the measurements of this um
00:06:16.800 of this
00:06:21.199 accessor bliss is not one number is
00:06:24.479 several numbers if you do replicas you
00:06:26.240 will have an error so therefore we have
00:06:27.680 to take into account when we score
00:06:29.840 the fact that we don't have a unique
00:06:31.600 possible value
00:06:32.960 that if people rank things differently
00:06:35.440 it may be a probability that that
00:06:37.520 different ranking is probably going to
00:06:40.560 be measured if we redid the experiment
00:06:42.319 so we had into account that
00:06:43.840 and we define what is called the
00:06:45.199 probabilistic concordance index which is
00:06:47.120 like the concordance index
00:06:48.880 but under the fact that the gold

00:06:51.759 standard is noisy
00:06:52.960 the concordance index by the way is the
00:06:54.639 proportion of pairs of cell lines
00:06:57.039 so not of cell lines here of of drug
00:06:59.840 pairs this was for another
00:07:01.360 challenge whose excess over this order
00:07:03.759 was correctly predicted so if i say that
00:07:05.759 this
00:07:06.160 drug pair is more synergistic than this
00:07:08.880 other drug pair
00:07:10.560 uh and this is the true then i get a
00:07:13.680 brownie point if it is
00:07:15.120 false then i get subtracted uh value and
00:07:18.319 so the
00:07:19.120 the fraction of correct answers are the
00:07:21.280 ones that
00:07:22.479 contribute to the concordance index
00:07:25.919 so what was the result so if we assume
00:07:28.639 that the predictions were random
00:07:30.000 that basically that everybody was
00:07:31.680 sending the monkeys were predicting
00:07:33.520 basically and so the order of the
00:07:36.240 pairs were random you will have a null
00:07:39.599 distribution
00:07:40.639 and if you put on that null distribution
00:07:42.800 the
00:07:43.680 concordance index uh probabilistic
00:07:46.240 concordance index resulting from
00:07:48.560 the submissions we had kind of spanned
00:07:51.280 the support of this distribution
00:07:53.440 indeed some were kind of anti-correlated
00:07:56.000 but there were three
00:07:57.440 that seem to be uh in the tail
00:08:00.720 of the distribution and with a false
00:08:02.960 discovery rate less than 0.05
00:08:05.759 this is quite far from where we would
00:08:07.680 like to be this is the best possible
00:08:09.599 result that we could have gotten
00:08:11.039 remember that the gold standard is noisy
00:08:13.280 so we don't have one concordance index
00:08:15.039 but 0.9
00:08:16.800 so i like to think that either the the
00:08:19.680 glass is half
00:08:20.560 empty for these guys half empty for
00:08:23.360 these guys have full
00:08:25.120 and and i would like to say it's half
00:08:27.599 full
00:08:28.160 in the sense that this gives us the

00:08:29.919 sense that there is signal to be found
00:08:31.919 and um remember this and robert alluded
00:08:35.519 to this if
00:08:36.559 if we did it in for our algorithms most
00:08:39.760 likely we will find a way
00:08:41.839 to make it much better but this is blind
00:08:43.839 predictions
00:08:44.959 so this is the reality probably of what
00:08:47.440 the community can do
00:08:49.120 um today so
00:08:52.160 i wanted to discuss a little bit what
00:08:53.920 the um
00:08:57.200 the best performer did so what they were
00:08:59.279 thinking remember that we had said and
00:09:00.800 now i change notations to match the
00:09:02.720 notation of the best performers this
00:09:04.320 this group degree sorry the group was a
00:09:07.200 young cs lab from
00:09:09.040 ut southwestern but the algorithm they
00:09:11.120 call it degree
00:09:14.720 so they are thinking well if the
00:09:16.320 inhibition of the pair
00:09:17.920 is like the inhibition the probability
00:09:20.080 of being inhibited by one
00:09:22.160 plus the probability of being inhibited
00:09:24.160 by the other given that it wasn't
00:09:25.839 not inhibited by one they thought why
00:09:27.839 don't we think of this
00:09:29.200 as a sequential process in which we
00:09:31.279 start giving one drug
00:09:33.120 suppose you give a and you kill the
00:09:35.200 fraction fa of of cells
00:09:37.680 and then the ones that are not there you
00:09:40.000 add b
00:09:41.680 the other the other drug and you have to
00:09:44.000 think how b
00:09:44.880 is acting in the context of a having
00:09:47.600 been active
00:09:49.200 and so the idea is there you have the
00:09:51.279 gene expression profiles
00:09:53.920 you um you decide uh
00:09:57.279 on some prior knowledge what are the
00:09:59.760 genes that you are going to use and you
00:10:01.360 see what
00:10:03.040 some sort of similarity score between
00:10:05.440 the the
00:10:07.040 um the two drugs whether one created a
00:10:10.079 differential expression that was

00:10:11.360 consistent with the other drug or
00:10:13.680 anti-consistent with the other drug you
00:10:15.680 create how similar how
00:10:17.200 similar these two drugs are and this is
00:10:19.600 containing this similarity score
00:10:21.120 r you also have the drug response curves
00:10:24.000 and then you think
00:10:25.120 well i know f a right whether the
00:10:27.120 inhibition of that
00:10:28.959 because i have the drug response but
00:10:31.440 that drug
00:10:32.160 in the context of the other drug could
00:10:34.079 be fully aligned with the
00:10:36.079 with the with the other drug in its
00:10:38.320 action and therefore
00:10:39.920 it's as if i had a two times
00:10:42.959 the second drug or could not be aligned
00:10:45.279 in which case
00:10:46.079 in which case i was i will i would would
00:10:48.720 have
00:10:49.120 a fraction of the other drug um
00:10:51.600 contribution
00:10:52.720 if you think of this in that way then
00:10:54.959 the the
00:10:55.839 viability of uh the second drug in the
00:10:59.120 presence of the first drug could be
00:11:00.959 factored in this way this is a
00:11:02.320 mathematical model to do that so there
00:11:04.240 are many ways to do this this is how
00:11:05.600 they chose to do it
00:11:06.959 and they have a final inhibition that
00:11:08.720 they can subtract to the bliss
00:11:10.240 independence and get
00:11:11.680 a result and that's how they they mold
00:11:14.800 it
00:11:15.680 it's important to note that there was no
00:11:18.399 training set here this was pure
00:11:20.160 mathematical modeling so
00:11:21.680 we didn't give a training set in which
00:11:23.360 we said this is how two drugs actually
00:11:26.000 interact with each other they only had
00:11:28.320 the monotherapy
00:11:29.920 so another way of looking at the scoring
00:11:32.320 of this challenge is
00:11:33.440 take the most energistic the least
00:11:35.760 synergistic
00:11:36.560 or or antagonistic and forget about
00:11:38.800 these additive ones which are in the

00:11:40.560 middle
00:11:41.120 and how do people do when it comes to
00:11:44.320 um predicting the the
00:11:48.160 uh the synergistic on one side
00:11:50.480 synergistics are these ones
00:11:52.160 so this is what you expect by chance and
00:11:54.560 the antagonists which are these ones and
00:11:56.399 interesting
00:11:57.519 interestingly the the dots are
00:11:59.360 statistical significant with respect to
00:12:00.880 the dotted lines
00:12:02.800 everybody that was statistically
00:12:04.560 significant with respect to
00:12:06.240 either antagonism or synergism was not
00:12:09.200 significant with
00:12:10.000 respect to the other so for example the
00:12:12.079 best performers were
00:12:13.360 quite synergistic quite good at
00:12:15.440 predicting antagonism
00:12:16.880 but almost random at predicting
00:12:18.320 synergism and the ones that were good at
00:12:20.959 predicting synergism for example these
00:12:22.639 guys were almost random at predicting
00:12:24.320 antagonism
00:12:26.560 except for the wisdom of crowd solution
00:12:28.560 the aggregate of them all
00:12:29.760 that were significant in both counts
00:12:33.200 this this will give us something to
00:12:35.120 think in terms of
00:12:36.399 predicting synergism or antagonism
00:12:40.079 seems to be having different mechanisms
00:12:43.920 okay so let's move on to uh
00:12:46.959 some of the work that i am doing in my
00:12:48.639 lab that that came
00:12:50.079 as a result of what i learned from the
00:12:51.920 challenges um
00:12:53.279 which is the following what does
00:12:55.600 synergism mean
00:12:57.519 at the molecular level so at the
00:13:00.560 transcription level so
00:13:01.839 let's feel one blank that we had which
00:13:04.480 is that
00:13:05.760 we knew the gene expression after one
00:13:08.880 drug
00:13:09.920 after the other but we didn't know what
00:13:11.519 happens with this gene expression after
00:13:12.880 you put the two drugs we didn't have the
00:13:14.320 transcriptional profile in that case

00:13:16.480 by the way i use um three drugs tm
00:13:20.000 and the combination tm and mw is the
00:13:23.760 same
00:13:24.160 as this one and the combination and you
00:13:26.079 can see here the viability as a function
00:13:28.079 of time we measure the viability in time
00:13:30.399 was such that t and m didn't really
00:13:33.920 change the viability much
00:13:35.760 the excess of bliss predicted is this
00:13:38.639 but actually
00:13:39.680 the real viability is plummeting that
00:13:42.399 means these drugs are really synergistic
00:13:44.399 each one by themselves don't do anything
00:13:45.920 but the two together kill the cells
00:13:48.079 big time whereas these guys uh have the
00:13:51.519 m that that doesn't
00:13:52.800 do much w that does something and the
00:13:55.839 combination that that's almost the same
00:13:58.079 as w so let's see what what what do we
00:14:00.800 have at the level of gene expression
00:14:02.399 this is just a tip of the iceberg
00:14:04.959 analysis
00:14:06.720 if you look at tm that were the very
00:14:08.560 synergistic ones
00:14:10.560 t had some differential express genes
00:14:12.880 and some differential expressions for
00:14:14.639 m but when you put t and m together you
00:14:17.360 have a lot
00:14:18.800 new stuff that it seems to be very hard
00:14:21.120 to be predicted
00:14:22.320 by the independent action of each of the
00:14:24.480 drugs in principle but we are trying to
00:14:26.560 do it anyway
00:14:27.680 and this trend grows in time and you
00:14:30.639 know we have
00:14:31.839 i can see it here but you know 2 000
00:14:34.480 genes that seem to be differential
00:14:35.839 expression
00:14:36.560 here whereas we have hundreds here and
00:14:39.600 the same is true for m different is true
00:14:42.480 from mw which there seems
00:14:44.000 to be a clear dominance of w all the
00:14:47.279 differentially expressed genes
00:14:49.680 for uh mw are basically the same as w
00:14:54.240 so you see that there is a huge overlap
00:14:56.160 in the combination
00:14:57.360 with one of the drugs so it's
00:14:59.120 interesting that at the transcription

00:15:00.480 level
00:15:00.880 there is more that meets the eye than
00:15:02.800 what we did for the
00:15:04.000 uh thank you for the dream challenge so
00:15:06.399 basically what this is saying is
00:15:09.680 synergistic activity of a gene
00:15:12.720 might occur something new when you put
00:15:15.760 the two drugs
00:15:16.720 when you have for example a gene that is
00:15:18.639 downstream of two transcriptional
00:15:20.320 regulators and one transcriptional
00:15:21.760 regulators
00:15:22.959 is working only for one drug and the
00:15:25.440 other on the
00:15:26.000 other but this guy needs the two and if
00:15:29.199 that's the case for those genes
00:15:30.959 the combination will do something that
00:15:32.720 none of the individual ones
00:15:34.480 could uh could do and that's what we can
00:15:36.720 see when we do the
00:15:38.320 differential expressed in the combo
00:15:42.959 good so uh what's next
00:15:46.160 we have um as robert presented the new
00:15:49.600 ls to ls
00:15:50.880 als certification challenge we are going
00:15:53.680 to do
00:15:54.320 an astra seneca sanger recombination
00:15:57.279 prediction dream challenge which is
00:15:59.279 three years after the last one we are
00:16:02.480 increasing the scale going from 91
00:16:04.720 combinations to 13 000 provided by
00:16:07.920 astrosenica
00:16:08.959 and going from one cell line fro to um
00:16:12.800 79 cell lines uh
00:16:16.000 not just of um b cells but also of
00:16:19.600 breast
00:16:20.160 but but of breast lung and in the gi
00:16:22.560 tract
00:16:23.440 and what is different is that well
00:16:24.959 before we were doing gene expression
00:16:26.480 data
00:16:27.279 before and after treatment now we don't
00:16:29.440 do give gene expression data only basal
00:16:32.320 and we also give copy number variation
00:16:34.240 mutations and so on
00:16:35.600 we give a training set and the
00:16:37.040 leaderboard and we ask to predict which
00:16:39.360 before we didn't have neither training

00:16:40.880 set nor leaderboard
00:16:42.399 so we don't have the before and after
00:16:45.040 but we have other things and much more
00:16:47.360 scaled up and finally what are we asking
00:16:51.120 while in dream seven
00:16:52.240 we asked to rank synergy scores in dream
00:16:55.360 10
00:16:56.000 we are going to predict synergy scores
00:16:58.800 given a drag which added
00:17:00.399 agents lead to the highest synergy and
00:17:03.040 given the cell line which combination is
00:17:04.720 the most synergistic and if you are
00:17:06.000 interested to pre-register you can go to
00:17:08.199 dreamchallenges.org
00:17:09.439 upcoming challenges and that's where you
00:17:12.240 will be able to pre-register so
00:17:13.839 conclusions
00:17:15.039 for the dream 7 drug synergy challenge
00:17:17.679 we had
00:17:18.400 in three months about 90 researchers
00:17:20.799 from the 31 teams have participated
00:17:23.599 this would be about 23 percent years
00:17:26.880 that means we accelerated research by
00:17:28.640 that amount this is the way we can think
00:17:30.799 of doing challenges we are right not
00:17:32.880 just finding the best we are really
00:17:34.880 making a push
00:17:35.840 in the research in the field prediction
00:17:38.559 is possible without the training set
00:17:40.320 that that seems to be the
00:17:41.760 what our dream 7 challenge says
00:17:44.880 but methods that are good at predicting
00:17:46.640 synergy seem to be by predicting
00:17:48.240 antagonism and vice versa we may need to
00:17:50.640 model both independently transcriptomic
00:17:54.080 based energy
00:17:55.200 prediction needs a molecular
00:17:56.640 understanding of where
00:17:58.240 the novelty emerges at the single gene
00:18:00.400 level that's something that we couldn't
00:18:01.760 see before
00:18:02.640 and new experiments need to be done and
00:18:05.280 we have the new drag combo challenge
00:18:07.039 that uh
00:18:07.840 will open up probably later in the
00:18:10.160 summer
00:18:11.120 lots of people participate in dream
00:18:13.679 sagebio networks

00:18:15.520 has a very important participation my
00:18:18.160 folks at ibm
00:18:19.760 university of colorado costello oregon
00:18:22.080 health and sciences university joe gray
00:18:23.760 and laura heiser
00:18:24.960 nci columbia university where the data
00:18:27.760 from the synergy challenge
00:18:29.200 came astra seneca and the nci dream
00:18:32.799 community thank you
00:18:39.600 so are there any questions
00:18:49.760 at all well you know what we are seeing
00:18:53.200 is that what we call
00:18:54.240 targets the molecular target for example
00:18:56.320 a lot nib has egfr
00:18:58.160 or herceptin has her too
00:19:01.280 probably it's not the right way to think
00:19:02.960 in general when we do drug
00:19:04.480 combinations because there is a lot of
00:19:07.280 functional changes that occur after that
00:19:09.440 and there are molecular targets that
00:19:11.520 can be inferred from the transcriptional
00:19:13.440 changes and i think that that's
00:19:15.280 going to be more uh specific than the
00:19:18.559 target that sometimes it's not uh
00:19:20.400 unique or or you know it's a little bit
00:19:22.840 dirty
00:19:27.760 definitely there are algorithms to um to
00:19:30.080 dev uh
00:19:30.799 to infer new target new molecular
00:19:32.720 targets uh
00:19:34.480 that are working very well with
00:19:35.919 transcriptional data and that's that's
00:19:37.840 happening here