Poster #3176
Abstract #4678

Biomarker identification using Artificial Intelligence data analytics and xenograft mouse model based clinical trial simulation

Afshar M[1], Bichat F[2], Duchamp O[2], Etcheto A[1], France D[2], Kindermans M[1], Mignard C[2], Parmentier F[1], Ratisma H[2]

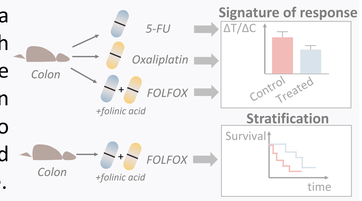
[1] Ariana Pharmaceuticals, Paris, France ; [2] Oncodesign, Dijon, France

Introduction

The growing number of anti-cancer drugs available at different stages of clinical development and generalized use of combination therapy further complexifies the early identification of companion markers, markers of synergy as well as novel indications for existing and new drug combinations.

Artificial Intelligence tools can integrate and analyze broad range of data generated by well characterized patient derived xenograft mouse models (PDX), PDX experiments provide an opportunity to simulate a clinical assessment using multiple mice.

In this study, we developed a PDX platform combined with the KEM® Artificial Intelligence data analytics, that is based on Formal Concept Analysis, to simulate a clinical trial and identify biomarkers of response.



The platform was tested on colon cancer patient derived PDX. mRECIST response was measured for 27 PDX exposed to either Oxaliplatin, 5-Fluorouracil (5-FU), or their combination in addition to folinic acid (FOLFOX). Survival was measured for 27 PDXs exposed to FOLFOX or Placebo (vehicle) simulating a clinical trial setting with 2 arms.

Methods

Data

- 27 PDX models were exposed to 5-Fluorouracil (5-FU), Oxaliplatin or FOLFOX. In a former study [1], tumor response was assessed using mRECIST for each drug, and survival was assessed for FOLFOX only, in comparison with a vehicle (control).
- PDX were previously [2] characterized with copy number (CGH array, Human Genome CGH Microarray-244A, Agilent Technologies, 25 869 genes) and transcriptomic (micro array, U133A GeneChip, Affymetrix, 12 112 genes) data for 26 and 21 PDX respectively.
- CGH data was limited to 409 genes relevant in oncology [3]. Copy numbers that covers the same PDX were clustered together, leading to 276 clusters of copy numbers
- Micro array data was analyzed using GSVA [4], limited to 2463 pathways (pathways with < 10 genes or > 500 genes were excluded) ; for each drug, top pathways were selected by computing moderated t-test of differential expression by empirical Bayes moderation from microarray linear model fitting [5]. Only genes from top pathways with p-value<0.01 were retained. Additional genes, not present in pathways, were also selected by the same method, thus leading to an overall number of 102 genes for 5-FU (74 genes in 4 pathways), 69 genes for Oxaliplatin (52 genes in 3 pathways), and 74 genes for FOLFOX (42 genes in 2 pathways)

Data handling

- Tumor response data and survival were discretized in 2 groups (‘*low*’, ‘*high*’) of 13 PDX separated by the median: 2-tiles discretization
- Gene expression levels were discretized in 3 groups (‘*low*’, ‘*medium*’, ‘*high*’) with 8 or 9 PDX in each groups: 3-tiles discretization
- Copy number was not modified as values are already discrete (‘*loss*’, ‘*gain*’, ‘*no change*’)

Artificial Intelligence

- Formal Concept Analysis as implemented in the KEM® platform generates all hypotheses consistent with the data in the form of association rules.

Example

If (Gene1Expression High) **Then** (TumorReduction High)

KEM® generates association rules $\text{Variable}_i \rightarrow \text{Endpoint}_j$ in an exhaustive manner. These rules are characterized by 4 metrics that help ranking them.

- KEM® (Knowledge Extraction Management) can combine multiple data sources and overcomes the over fitting challenge of analysis of biomarker data in small clinical studies [6]

Metrics

- **Support**: Number of times that the rule is checked in the dataset
- **Confidence**: Proportion of cases verifying Gene 1 = High and TumorReduction= High
- **Lift**: Ratio of the observed support to that expected if Gene 1 = High and TumorReduction= High were independent.
- **P-value**: Fisher’s exact test

KEM® Biomarker

- Identify variables alone and in combinations that best predict a binary outcome.
- Systematic exploration of combinations of variables.
- Predictive signatures derived from one or multiple rules.
- Performances of predictive signatures are assessed using metrics: sensitivity, specificity, efficiency, positive and negative predictive values.

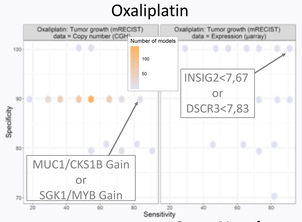
KEM® Clinical

- Systematic analysis to identify all patient characteristics at Baseline, or combination of characteristics, linked to outcomes, at multiple time points.
- Each interaction’s significance is statistically characterized.
- Each interaction’s amplitude is assessed using hazard ratio (HR) for continuous outcome, as well as odds-ratio (OR) for binary outcome.
- Odds-ratio represents the odds of outcome improvement during the whole trial period.
- Hazard-ratio represents the immediate chance of improvement at a given time point.

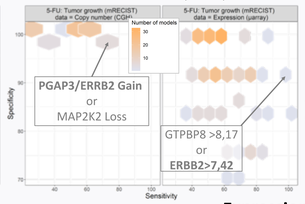
Results

4,792 biomarker signatures generated

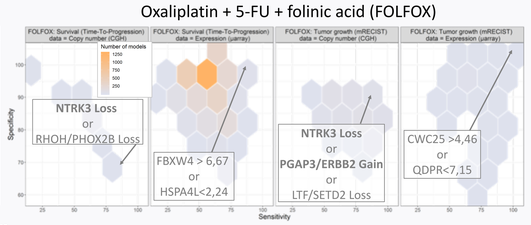
Oxaliplatin



5-Fluorouracil (5-FU)



Oxaliplatin + 5-FU + folinic acid (FOLFOX)



Copy Number

Gene	Variation	Start	Stop
ERBB2/PGAP3	Gain	17:37,831,500	17:38,068,895
NTRK3	Loss	15:87,614,479	15:88,696,754

Expression

Gene	Threshold	Value
ERBB2	High	>7.42

Odds-ratio (OR): cumulative risk, binary outcome: Survival > 38 days (treated) / 17 days (control)

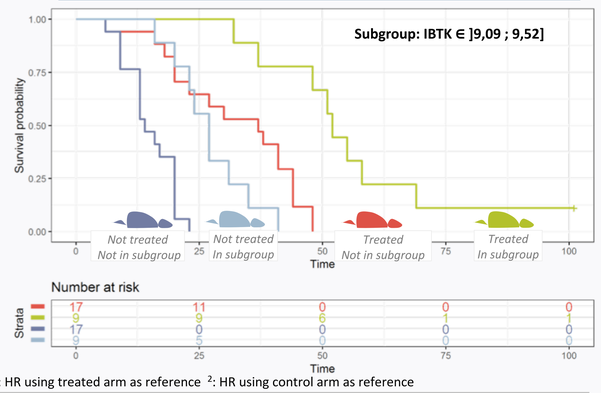
Gene	Feature	Value	Metric	Metric Value	p-value	Test
ERBB2/PGAP3	CopyNumberCluster 367	Gain	OR	6.25	0.027	likelihood ratio
PGAP3	CopyNumberCluster 368	Gain	OR	10	0.021	likelihood ratio
NTRK3	CopyNumberCluster 1186	Loss	OR	3	0.12	likelihood ratio
NTRK3	CopyNumberCluster 1229	Loss	OR	1.88	0.371	likelihood ratio
ERBB2	Expression	High	OR	1.67	0.544	likelihood ratio
NOTCH2	Expression	Low	OR	2.67	0.224	likelihood ratio
NOTCH2	Expression	Medium	OR	1.75	0.521	likelihood ratio
PGAP3	Expression	High	OR	2.78	0.227	likelihood ratio

Hazard ratio (HR): immediate risk, continuous outcome: survival

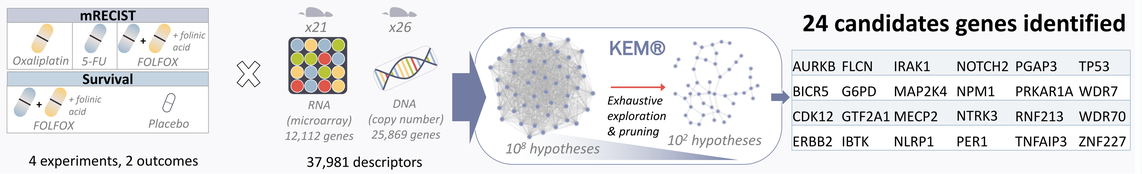
Gene	Feature	Value	Metric	Metric Value	p-value	Test
NOTCH2	CopyNumberCluster 1030	Gain	Log HR[1]	2.09	0.02	Wald
NOTCH2	CopyNumberCluster 1031	Gain	Log HR	2.03	0.14	Wald
NOTCH2	CopyNumberCluster 1032	Loss	Log HR	2.36	0.6	Wald
ERBB2	CopyNumberCluster 367	Gain	Log HR	2.2	0.28	Wald
PGAP3	CopyNumberCluster 367	Gain	Log HR	2.2	0.28	Wald
PGAP3	CopyNumberCluster 368	Gain	Log HR	2.59	0.13	Wald
ERBB2	Expression	Medium	Log HR	2.03	0.412	Wald
IBTK	Expression	Medium	Log HR	2.58	1.74E-06	Wald
NOTCH2	Expression	High	Log HR	2.33	0.619	Wald
NOTCH2	Expression	Low	Log HR	2.31	0.013	Wald
PGAP3	Expression	Low	Log HR	3.15	0.922	Wald
WDR70	Expression	Medium	Log HR	2.04	0.346	Wald
ZNF227	Expression	High	Log HR	2.16	6.70E-07	Wald
ZNF227	Expression	Medium	Log HR	2.05	0.491	Wald

2 subgroups identified

Data	Signature	Hazard Ratio[2]	[95% C.I.]	P-value (Wald)
CGH	NTRK3 ≠ Gain	0,10	0,05 – 0,26	5,1 10^{-6}
µarray	IBTK ∈]9,09 ; 9,52]	0,11	0,05 – 0,24	6,7 10^{-7}



[1]: HR using treated arm as reference [2]: HR using control arm as reference



24 candidates genes identified

AURKB	FLCN	IRAK1	NOTCH2	PGAP3	TP53
BICR5	G6PD	MAP2K4	NPM1	PRKAR1A	WDR7
CDK12	GTF2A1	MECP2	NTRK3	RNF213	WDR70
ERBB2	IBTK	NLRP1	PER1	TNFAIP3	ZNF227

Conclusion

This work demonstrates the ability of an Artificial Intelligence platform using PDX to simulate clinical trials and identify biomarkers of drug efficacy and synergy.
Candidate biomarkers were identified using the KEM® platform through automated workflows that can be easily repeated, deployed, and adapted to other omics data.

Systematic identification of both biomarker for tumor response and survival can be performed in parallel, thus enabling to extract knowledge that has an impact at the molecular level (tumor response) as well as at the clinical one (survival).

The platform can be used for drug repositioning or identification of innovative drug combinations, while maintaining a high level of robustness.

This study will be further extended to other indications (breast and lung), with the aim of validating the signatures obtained here in another cohort of PDX. Moreover, whole exome sequencing and RNA-seq data will be included.

We believe this work paves the way towards innovative Precision Medicine clinical trials, in which simulations performed in PDX and analyzed using Artificial Intelligence will deliver actionable hypothesis for patients inclusion and study extension designs.

References

[1] Mignard *et al*, Single Mouse Preclinical Trial (SMPT): a tool for translational research, AACR 2018, abstract #2170
[2] Julien *et al*, Characterization of a large panel of patient-derived tumor xenografts representing the clinical heterogeneity of human colorectal cancer, Clin Cancer Res., 2012
[3] Ion AmpliSeq™ Comprehensive Cancer Panel, www.thermofisher.com
[4] Hänzelmann *et al*, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinformatics, 2013
[5] Smyth GK, Linear models and empirical bayes methods for assessing differential expression in microarray experiments, Stat Appl Genet Mol Biol., 2004
[6] Afshar M, Lanoue A, and Sallantin J. Multiobjective/Multicriteria Optimization and Decision Support in Drug Discovery. Comprehensive Medicinal Chemistry, 2007

Ariana Pharma and Oncodesign are members of the IMODI consortium

www.arianapharma.com | www.oncodesign.com | m.afshar@arianapharma.com | contact@oncodesign.com