A Systematic Patient-Derived Xenograft Based Solution for Pre-Clinical Biomarker Discovery

Bruno Zeitouni[1], Anne-Lise Peille[1], Zakia Amalou[1], Thomas Metz[1], Heinz-Herbert Fiebig[1], Vincent Vuaroqueaux[1].

[1] Oncotest - Charles River, Am Flughafen 12-14, 79108 Freiburg, Germany.

1 INTRODUCTION

- There is an acute need for biomarkers at every phase of drug development from selecting preclinical models in pharmaco-genomic studies to enrollment of patients in clinical trials. However, their identification remains extremely challenging due to the limited availability of clinical samples. Use patient-derived xenografts (PDX) for testing anticancer agents is of increasing interest due to their closer similarity to patient tumors compared to cell lines.
- A collection of 400 PDX covering more than 30 different cancer types has been extensively characterized using the microarray or next-generation sequencing technologies for gene expression, copy number variations and whole-exome mutations.
- Molecular profiles of PDX in combination with drug response data from in vivo or in vitro 2D or 3D assays performed on large panels of 100-200 PDX significantly facilitate biomarker research.
- Here we present :
  - A fully integrated bioinformatics pipeline dedicated to biomarker discovery in which the complete molecular profiles of our PDX have been systematically tested for association with drug sensitivity.
  - A demonstration of the efficacy of our approach to retrieve biomarkers of known clinical utility, by using several datasets of PDX drug responses to chemotherapeutics and targeted therapies

2 MATERIALS & METHODS

- **Drug testing.** Compound testing in PDX was done either *in vitro* or *in vivo* following the Oncotest-CRL protocols for 3D tumor clonogenic assays or standard *in vivo* testing, respectively.
- **Molecular data.** Raw data (CEL or FASTQ files) were processed with bioinformatics analyses pipelines for gene expression, copy number variants and exome mutations. Gene x Sample matrices of expression values or alteration binaries were generated for drug correlation tests.
- **Statistics.** All statistical tests were done using R scripts. Drug response data were treated either as continuous variables using the Spearman or Wilcoxon tests, or as categorical variables (with two groups of responders and non-responders) using the limma, t-test or Fisher exact test. Significant genes were selected with p-values ≤ 0.05 for all tests (and absolute fold-changes > ±1 for expression with P1). Clustering and heatmaps were done with the EMA R package. The over-representation analysis of biological functions was done with the Enrichr web tool (http://amp.pharm.mssm.edu/Enrichr/).

3 RESULTS

1- Drug testing and molecular data for biomarker discovery

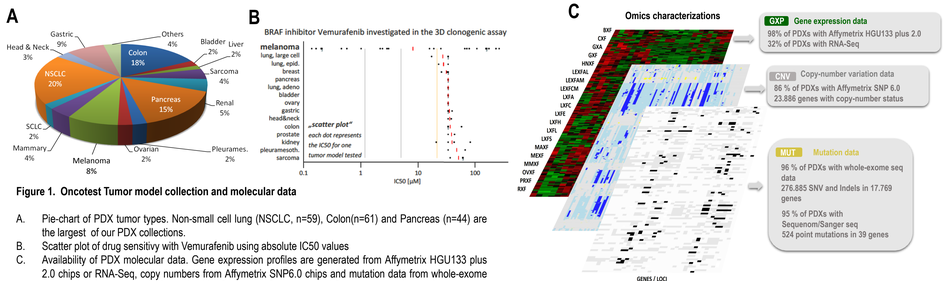

Figure 1. Oncotest Tumor model collection and molecular data

A. Pie-chart of PDX tumor types. Non-small cell lung (NSCLC, n=59), Colon(n=61) and Pancreas (n=44) are the largest of our PDX collections.
B. Scatter plot of drug sensitivy with Vemurafenib using absolute IC50 values
C. Availability of PDX molecular data. Gene expression profiles are generated from Affymetrix HGU133 plus 2.0 chips or RNA-Seq, copy numbers from Affymetrix SNP6.0 chips and mutation data from whole-exome sequencing data previously determined by Oncocarta Sequenom and Sanger sequencing analyses.

2- Pre-clinical study and biomarker discovery approaches

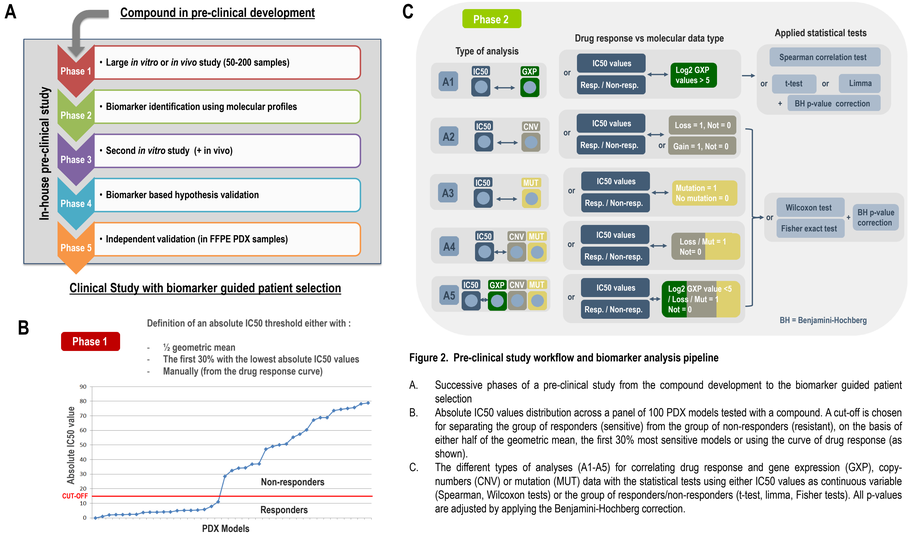

Figure 2. Pre-clinical study workflow and biomarker analysis pipeline

A. Successive phases of a pre-clinical study from the compound development to the biomarker guided patient selection
B. Absolute IC50 values distribution across a panel of 100 PDX models tested with a compound. A cut-off is chosen for separating the group of responders (sensitive) from the group of non-responders (resistant), on the basis of either half of the geometric mean, the first 30% most sensitive models or using the curve of drug response (as shown).
C. The different types of analyses (A1-A5) for correlating drug response and gene expression (GXP), copy-numbers (CNV) or mutation (MUT) data with the statistical tests using either IC50 values as continuous variable (Spearman, Wilcoxon tests) or the group of responders/non-responders (t-test, limma, Fisher tests). All p-values are adjusted by applying the Benjamini-Hochberg correction.

Compound	Cetuximab	Vemurafenib	PD0325901
Study	*in vivo*	*in vitro*	*in vitro*
Target	EGFR	BRAF	MEK
# Tissue Types	1	2	18
Types	Colon cancer (CXF)	Melanoma (MEXF), Colon cancer (CXF)	Colon cancer (CXF), Melanoma (MEXF), Pancreatic cancer (PAXF), …
Cut-off	30% most sensitive	Abs IC50 < 2 (curve)	30% most sensitive
# Models tested	54	92	162
# Responders (Sensitive)	15	15	48
# Non-responders (Resistant)	39	77	114
Gene expression availability	96%	98%	99%
Copy-numbers availability	93%	90%	95%
Exome mutation availability	96%	87%	99%

Table 1. Design of test biomarker studies

3 different biomarker studies were performed using different compounds, number of samples and size of datasets. Molecular data availability regarding gene expression, copy-numbers and exome mutation is provided.

3- Validation of biomarkers

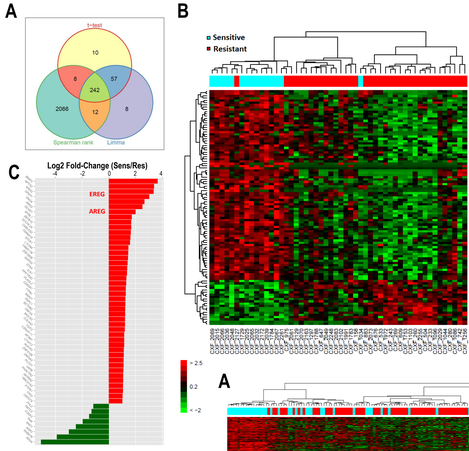

Figure 3. Cetuximab biomarker results using gene expression data identified with A1

A. Venn diagram of the significant probe-sets from the t-test, Spearman correlation and limma test (adjusted p-value ≤ 0.05 for all 3 tests).
B. Hierarchical heatmap clustering of the significant probe-sets with log fold change > ±1 (n=104). The blue columns indicate models that are in the responder group and the red columns indicate models that are in the non-responder group.
C. Distribution of significant genes with log fold change > ±1 (n=52). Most significant genes including AREG and EREG are found as expected.

	# Copy loss/ Sensitive	# Copy loss/ Resistant	# No copy loss/ Sensitive	# No copy loss/ Resistant	Fisher p-value	Wilcoxon p-value
OR2T11_chr1	3	1	11	68	5.35E-04	5.36E-02
PTEN_chr10	3	1	11	68	5.35E-04	7.85E-02
PPARA_chr22	2	0	12	69	8.99E-04	3.58E-02
CDKN2A_chr9	6	7	8	62	9.19E-04	6.75E-02
CDKN2B_chr9	5	7	9	62	6.71E-03	9.51E-02
OR2T10_chr1	2	1	12	68	1.65E-02	1.32E-01
PAPSS2_chr10	2	1	12	68	1.65E-02	1.82E-01
ATAD1_chr10	2	1	12	68	1.65E-02	1.82E-01
CFL1P1_chr10	2	1	12	68	1.65E-02	1.82E-01
DMD_chrX	2	1	12	68	1.65E-02	1.04E-01
MACROD2_chr20	0	15	14	54	2.05E-02	1.21E-02

Table 2. Significant gene copy losses correlated to the Vemurafenib drug response identified with A2 (# = number of models)

	# Mutated/ Sensitive	# Mutated/ Resistant	# Non mutated/ Sensitive	# Non mutated/ Resistant	# Low covered (gene deletion)/ Sensitive	# Low covered (gene deletion)/ Resistant	Fisher p-value	Wilcoxon p-value
BRAF_chr7	10	10	4	58	0	0	4.99E-05	8.78E-07
TP53_chr17	2	46	12	21	0	1	2.35E-04	3.02E-02
RGPD3_chr2	2	25	9	7	3	36	7.36E-04	1.29E-01
HLA-DRB1_chr6	6	59	8	9	0	0	1.01E-03	1.26E-01
NYAP2_chr2	3	2	0	34	11	32	1.09E-03	3.97E-03
KRAS_chr12	0	27	14	35	0	6	1.36E-03	2.90E-02
CDKN2A_chr9	2	0	0	28	12	40	2.30E-03	2.22E-02
SULF2_chr20	4	1	10	67	0	0	2.57E-03	1.14E-02
OTOG_chr11	0	27	14	41	0	0	3.46E-03	2.01E-02
HEXIM2_chr17	3	1	2	35	9	32	3.60E-03	4.70E-03

Table 3. The 10 most significant mutated genes correlated to the vemurafenib drug response identified with A3 (# = number of models)

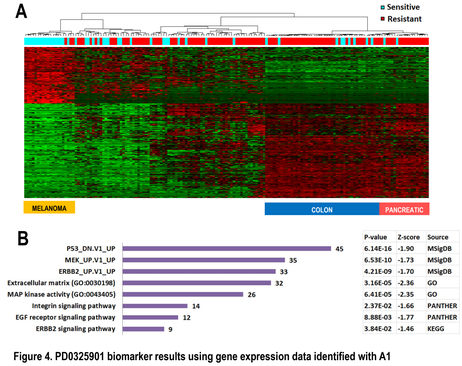

Figure 4. PD0325901 biomarker results using gene expression data identified with A1

A. Hierarchical heatmap clustering of the significant probe-sets with log fold change > ±1 (n=977). The blue columns indicate models that are in the responder group and the red columns indicate models that are in the non-responder group.
B. Over-representation analysis of biological functions/pathways from the significant differentially expressed genes (n=726). P-values, Z-scores and sources of annotation databases are shown.

4 CONCLUSION

The development of strategies for testing anticancer agents using PDX in large scale single mouse trials, or high throughput *in vitro* 2D, 3D screening approaches coupled to a more systematic biomarker research should significantly contribute to early biomarker identification and facilitate drug development.