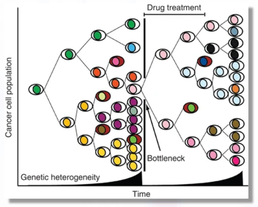
Inferring the mechanism of action of new drugs through the analysis of the predetermined heterogeneous response to treatment of different subpopulations of cancer cells

Houssein Chhouri[1], Nina Verstraete[2], Jiyoung Lee[1], David Alexandre[1], Vera Pancaldi[2] and Luca Grumolato[1]

[1] Université de Rouen Normandie, INSERM, NorDiC, Rouen, France

[2] Centre de Recherches en Cancérologie de Toulouse, INSERM, CNRS, Université Paul Sabatier, Université de Toulouse, Toulouse, France

Introduction



Tumors are composed of dynamic and **heterogeneous cell populations** characterized by distinct genetic and epigenetic profile. This intratumor heterogeneity has dramatic consequences, not only for cancer **progression** and **metastatic** spread, but also for **resistance to therapy**.

During therapeutic intervention, small populations of **resistant or tolerant clones are selected** thanks to their higher-fitness, resulting in **tumor relapse**. Lung cancer is the most common cause of cancer-related mortality worldwide. **Non-small cell lung cancer** (NSCLC), which represents 85% of the cases, can be driven by the **activation of certain oncogenes**, such as the epidermal growth factor receptor (**EGFR**). This receptor tyrosine kinase, mutated in about 15% of NSCLCs, plays a major role in the control of cell survival and proliferation, and it constitutes a **target of choice for therapy**.

We used a **lentiviral genetic barcode strategy** to label individual clones within a mass population of PC9 cells, a cell line derived from human NSCLC. We then assessed the **barcode profile** of cells treated with different **compounds capable of inhibiting cell growth**.

We found that some subpopulations display a **specific and predetermined response to the treatment**, probably reflecting particular epigenetic profiles of the cells. We extended these findings and showed that each drug exerts a **characteristic effect on the clonal architecture** of the cell population, resulting in a **specific barcode pattern** that can be used as a **signature** to **compare different compounds** and investigate their **mechanism of action**.

Methods

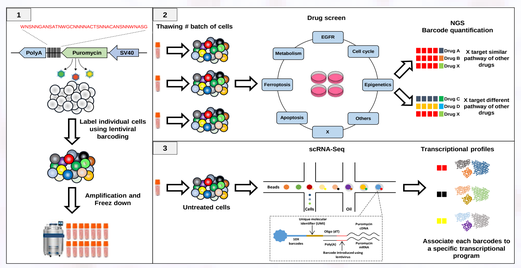
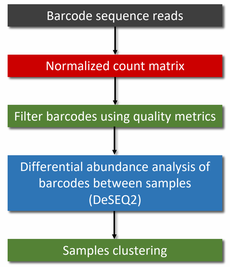


Fig. 1. Diagram illustrating the Barcode-Signature strategy.

We generated a high-complexity lentiviral barcode library containing semi-random 38 base-pair long DNA barcodes. PC9 NSCLC cells were transduced with the library at low multiplicity of infection to **label each cell with one unique barcode**. After barcoding, 25'000 cells isolated, amplified and frozen down to be stored in liquid nitrogen.

To perform our screen, different vials of PC9 barcoded cells were thawed and screened against 87 compounds with well-defined mechanisms of action. After 9 days of treatment, the cells were harvested and the genomic DNA was extracted. The barcode-containing sequence were PCR amplified from genomic DNA and sequenced to identify which clones had survived the drug treatment and in which numbers.

The DNA barcodes are located in the 3'-UTR region of the puromycin resistance gene, which makes it possible to **combine lineage tracing with gene expression single cell analysis**.



Barcoding computational analysis

Our dataset contains in total 520 samples including 40 control samples and 12 time zeros samples representing 11 experiments spanning over 11 runs. Barcode sequences associated with each sample were recorded in a barcode-count matrix, normalized by sample to one million (i.e. such that the sum over all barcodes per sample equaled one million), reaching a total of 3'120'235 unique barcodes. To filter out non-specific PCR amplification, barcodes detected in less than 5 control samples and 5 time zeros over all the samples were discarded. Filtered data composed of 12'305 barcodes was then analyzed using the statistical computing language R, specifically the package DeSEQ2 to evaluate log2 fold change (FC) between condition and control samples within each experiment to the drug-specific effects on the clonal architecture in terms of barcode profile. For determining the significance of the differences of logFC between two conditions, the Wald test was performed on log-transformed values, and p-values <0.05 were considered statistically significant.

Drugs used in the study

EGFR	Chemotherapy	Multi-kinase	Epigenetics	Antibiotics	RAF	Autophagy	JAK/STAT	Ferroptosis	MEK
Osimertinib WZ4002 Rociletinib Lazertinib Mavelertinib Gefitinib Gefitinib-PROTAC	Pemetrexed Methotrexate Pralatrexate 5-Fluorouracil Carboplatin Oxaliplatin Cisplatin Paclitaxel Cisplatin Doxorubicin	Sorafenib Regorafenib Sunitinib Lenvatinib Cabozantinib	SAHA (Vorinostat) Trichostatin A Sodium Butyrate Sodium phenylbutyrate Tazemetostat 5-Azacytidine	Bafilomycin A1 Trimethoprim Ngercin Mitomycin C	LY3009120 Vemurafenib RAF265 CCT196969	Chloroquine Spautin-1 SBI-0206965	Napabucasin Stattic-V AZD1480	Erastin Ferrostatin-1 Deferoxamine	U0126 Trametinib
AMPK	**ER stress**	**Apoptosis**	**Proteasome**	**Gap junction**	**CDK**	**MNK**	**Wnt**	**ROCK**	**ALK**
Ibudilast Phenformin HCL Dorsomorphin	CCT020312 Tunicamycin Thapsigargin	S63845 Navitoclax	MG-132 Bortezomib	Mefloquine Probenecid	TH21 Palbociclib	eFT-508 CGP57380	XAV-939 LGK-974	LIMKi 3 Y-27632	Crizotinib TAE684
mTORC	**Mitochondrial biogenesis**	**CaM-Kki**	**SHP2**	**GPX4**	**Notch**	**Adrenergic receptor**	**FGFR**	**ERK1/RasGAP PI#3K**	**PI3K**
Temsirolimus	Levofloxacin	STO-609	SHP099 HCL	ML-210	IMR-1	L-755,507	Infigratinib	Pluripotin	AZD8186
BRCA	**AKT**	**RAS**	**Antioxidant**	**MDM2 antagonist**	**Src**	**AURKA**	**IGFR**	**NFKB**	**Unkwon**
Olaparib	Capivasertib	Sotorasib	N-Acetyl-L-cysteine	Nutlin-3	Saracatinib	Alisertib	BMS-536924	ISH-23	X-13271

Results

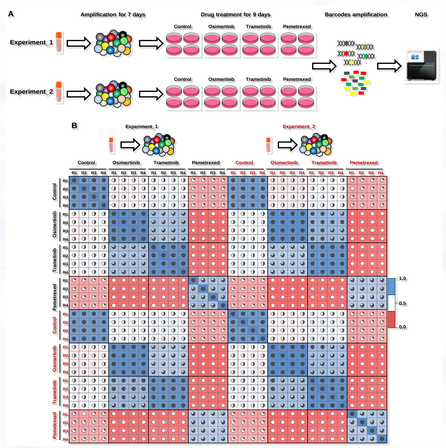


Fig. 2. Reproducibility of Barcode-Signature approach.

(A) Two different vials of PC9 cells containing the lenti-barcode library were thawed, amplified for one week, and treated with or without osimertinib (100 nM; EGFR inhibitor), trametinib (30 nM; MEK inhibitor) or pemetrexed (100 nM; chemotherapeutic agent) for 9 days. (B) Spearman's pairwise correlation analysis of positively selected barcodes revealed high correlation between samples treated with the same drug across different experiments.

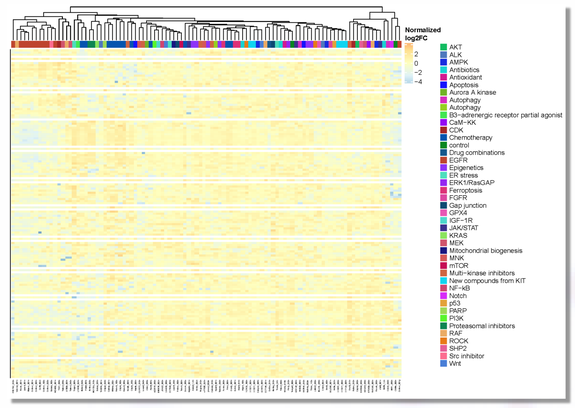


Fig.3 Annotated drug clustering based on barcodes signatures.

Each column represents a condition and rows represent barcodes. The heatmap shows normalized log2FC values obtained after barcodes sequencing data analysis with the DeSEQ2 R package. Annotation of the drugs shows that **all EGF inhibitors cluster together** (Gefitinib, Lazertinib, Mavelertinib, Osimertinib, PROTAC, Rociletinib, WZ4002). A considerable **clustering is also observed among chemotherapy agents** (Carboplatin, Mitomycin C, Pemetrexed, Cisplatin and to a lower extent with Methotrexate and Pralatrexate). CPG57380 (MN kinase inhibitor) and Lenvatinib (Multi-kinase inhibitor) cluster together. Na-Ibutyrate / Trichostatin on one hand and Azacytidine / Vorinostat, Phenylbutyrate and Tazemetostat on the other hand impact epigenetics processes and cluster together as well.

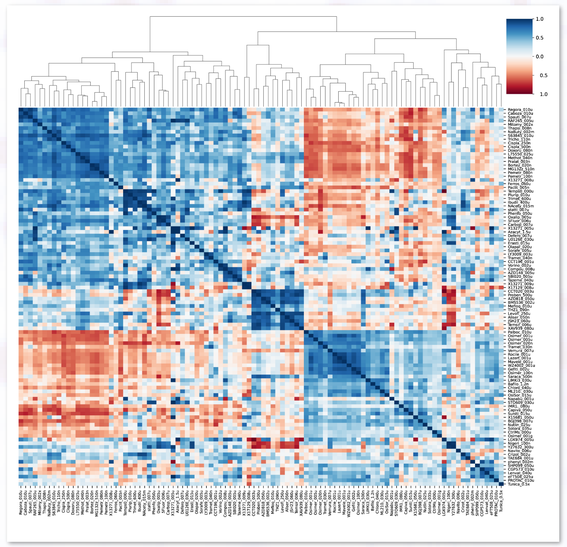


Fig.4 Drug-drug correlation clustered heatmap.

Each column in the heatmap represents one condition and shows the Pearson correlation to all other conditions (including itself), with blue for correlation = 1, red for correlation = -1 and white for 0 or insignificant correlation, based on the logFC between the condition and the control, obtained through DeSEQ2 analysis.

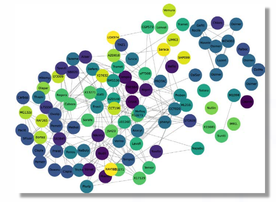


Fig. 5 Drug network.

The network is drawn from the correlation matrix in shown Fig. 4 with a Spring layout using the Fruchterman-Reingold force-directed algorithm. Edges are displayed between nodes if the two nodes' Pearson correlation coefficient is over 0.8. The color of the nodes is based on their annotation.

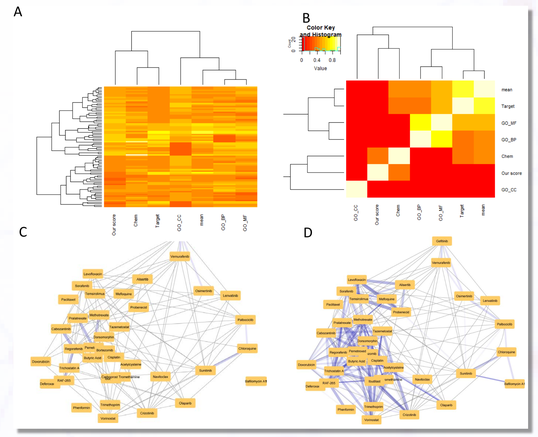


Fig. 6 Comparison with other drug similarity scores.

(A) Clustered heatmap of estimated similarity for 82 different drug pairs. We filtered our correlation matrix for drug pairs, resulting in all significant interactions between 93 drugs. Out of the possible interactions between them, 82 pairs are annotated as interacting in DrugSimDB, which collects drug-drug similarity based on different criteria. The heatmap shows the similarity based on our correlation score and the different DrugSimDB scores (similarity based on Chemical structure, Targets, Gene Ontology Cellular component, Molecular Function and Biological process of induced pathways, mean of all scores). Columns show similarity scores, rows are drug pairs. **(B)** Correlation matrix of similarity measures based on (A) calculated over the 82 drug pairs. Our similarity correlates best with chemical similarity of drugs (Pearson R=0.35, p=0.001), and similarity based on shared targets (Pearson R= 0.23, p=0.04). **(C)** Network of 27 drugs in common between DrugSimDB and our own set (converted to DrugBank ID) showing DrugSimDB edges (grey) and edges in common between DrugSimDB and our own network filtered at score > 0.6. **D)** Union of the DrugSimDB network and our own filtered at score > 0.6 showing in blue our score (thicker and less transparent lines show higher similarity score).

Conclusion

- We derived barcode signatures from about 90 different drugs and generated a drug similarity score displaying good **correlation** with other **drug similarity scores for drugs that were in common**.
- **Importantly, our approach does not require any previous knowledge on the target of the compounds or their functional characterisation.**
- Through comparison with the profile of known drugs, this **strategy could be used to identify the mechanism of action of new compounds**.
- **scRNA-Seq analysis of untreated cells** is ongoing, to investigate potential mechanisms underlying the **predetermined drug response** of certain cell subpopulations.