Decoding mechanism of action and susceptibility to drug candidates from integrated transcriptome and chromatin state

Caterina Carraro,[1] Lorenzo Bonaguro,[2,3] Jonas Schulte-Schrepping,[2,3] Arik Horne,[2,3] Marie Oestreich,[2] Stefanie Warnat-Herresthal,[2,3] Tim Helbing,[4] Michele De Franco,[1] Kristian Händler,[2,5,6] Sach Mukherjee,[7,8] Thomas Ulas,[2,3,5] Valentina Gandin,[1] Richard Göttlich,[4] Anna C. Aschenbrenner,[2,3,5,9] Joachim L. Schultze,[2,3,5,*] Barbara Gatto[1,*]

[1] Department of Pharmaceutical and Pharmacological Sciences, University of Padova, Padova, Italy

[2] Systems Medicine, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) e.V., Bonn, Germany

[3] Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany

[4] Institute of Organic Chemistry, Justus Liebig University Giessen, Giessen, Germany

[5] PRECISE Platform for Genomics and Epigenomics, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) e.V. and University of Bonn, Bonn, Germany

[6] Institute of Human Genetics, University of Lübeck, Lübeck, Germany

[7] Statistics and Machine Learning, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) e.V., Bonn, Germany

[8] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[9] Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands

* Corresponding author: j.schultze@uni-bonn.de, barbara.gatto@unipd.it

Abstract

Omics-based technologies are driving major advances in precision medicine but efforts are still required to consolidate their use in drug discovery. In this work, we exemplify the use of multi-omics to support the development of 3-chloropiperidines (3-CePs), a new class of candidate anticancer agents. Combined analyses of transcriptome and chromatin accessibility elucidated the mechanisms underlying sensitivity to test agents. Further, we implemented a new versatile strategy for the integration of RNA-seq and ATAC-seq data, able to accelerate and extend the standalone analyses of distinct omic layers. This platform guided the construction of a perturbation-informed basal signature able to predict cancer cell lines' sensitivity and to further direct compound development against specific tumor types. Overall, this approach offered a scalable pipeline to support the early phases of drug discovery, understanding of mechanism and potentially inform the positioning of therapeutics in the clinic.

Introduction

Omics technologies have revolutionized the classical *hypothesis-driven* paradigm of drug discovery, offering a new perspective for the systematic identification of targets and therapeutics.[1,2] An increasing number of examples are describing the use of these approaches to inspect the pharmacological profile of existing drugs, e.g. mechanism of action (MoA) and specific sensitivity biomarkers, as well as to assist their correct repositioning in the clinical practice.[3,4,5,6] Compared to traditional approaches, omics-based methods capture the complexity of biological systems and pathological processes in its entirety at increasingly affordable costs.[3] For this reason, refined strategies to handle the high-dimensional information of omics data are continuously investigated to expedite their routine use in drug development up to the clinics.[7,8–14]

Recent works from our group highlighted 3-chloropiperidines (3-CePs) as a novel class of candidate anticancer agents developed to improve the pharmacological profile of nitrogen mustard-based chemotherapeutics.[15–21] As intended, these agents were demonstrated to induce DNA lesions, a mechanism conceivably responsible for their cytotoxicity on tested cancer cell lines.[19–21] Interestingly, despite their expected broad-acting

MoA, a subset of derivatives showed a preferential activity against pancreatic adenocarcinoma BxPC-3 cells worth to be clinically translated, especially in light of the broad resistance of pancreatic tumors to most of the available treatments.[19–21]

The contribution of multi-omics to support early phases of drug discovery is growing exponentially in the era of precision medicine.[7] Omics technologies have the potential to address some of the intrinsic difficulties of the traditional drug discovery and development path, assisting it early from target prioritization and hit identification up to the evaluation of candidates' efficacy and safety.[4] Drug-perturbation experiments have been employed to inspect the functionality of target proteins[22] and the MoA of therapeutics, efficiently guiding the decision-making process in the development of lead compounds.[4] The massive accumulation of genomic and transcriptomic profiles offers a precious substrate for the optimization of strategies able to predict susceptibility to known therapeutics[23–26] refined by the continuous acquisition of data from high-throughput single-cell platforms.[10–12,27,28] Beyond the widely used transcriptome analysis, changes in gene regulation can be evaluated in terms of chromatin accessibility by ATAC-seq.[29–31] Examples of the joint use of these two omic techniques exist,[13,14,32,33] but their synergistic employment on compounds under early development is still underexplored.[3]

In this study, representative mono- (**M**) and bifunctional (**B**) 3-CePs bearing a single or double alkylating units (Fig. 1 A) were selected to exemplify the use of a multi-omic approach to investigate the molecular determinants of susceptibility to novel drug candidates and their MoA.[19–21] We analyzed transcriptional changes and chromatin status upon treatment in a high- (pancreatic adenocarcinoma BxPC-3) and low-sensitive (colorectal adenocarcinoma HCT-15) cancer cell lines by RNA-seq and ATAC-seq.[29–31] In addition, we implemented our multi-omics pipeline in drug discovery to derive perturbation-informed signatures predicting compound sensitivity. Overall, the proposed approach not only allowed to identify potentially more susceptible target tumor types for the further development of test compounds, but also offered a versatile predictive framework to support precision oncology in a clinical setting.

Results

Cancer tropism of 3-CePs is not explained by DNA damage

The mono- (**M**) and bifunctional (**B**) 3-chloropiperidines (3-CePs, Fig. 1 A), despite having different potencies, were shown to be particularly active against BxPC-3 pancreatic adenocarcinoma cells.[19,20] From this premise, the two compounds were selected along with the highly sensitive BxPC-3 cell line and the low-sensitive HCT-15 colorectal adenocarcinoma one to illustrate how integrative omics approaches unveil the molecular mechanisms responsible for the described cellular tropism.

First, to assess whether 3-CePs-induced DNA damage itself would differ in the two cell lines upon treatment, we measured the accumulation of DNA single-strand breaks after 6 h of treatment with both compounds at their cytotoxicity IC_{50}s in BxPC-3 and at a ten-times higher concentration (10 nM and 100 nM **M;** 200 nM and 2 µM **B**).[34] Surprisingly, the two cell lines showed very comparable DNA damage accumulation, in both cases higher after treatment with **M** compared to **B** (Fig. 1 B). These results clearly pointed towards differential responses in the two cell lines downstream of DNA damage.

Since alkylating agents are known to alter the progression of the cell cycle,[35–37] we next performed a cell cycle distribution analysis by flow cytometry after different times of treatment (6 h, 12 h, 72 h) with both compounds (Fig. 1 C). While **M** induced a persisting block in G1 throughout the observation time in BxPC-3 cells, this block was absent in HCT-15 cells. In contrast, **B** induced an early G2/S block in HCT-15 cells (6 h), which was not observed at later time points, while such a block was most obvious at 12 h for BxPC-3 cells. Despite similar DNA damage accumulation, these findings clearly indicated a different behavior for the two cancer cell lines in terms of cell cycle progression after treatment with the two 3-CePs.

To determine additional mechanisms explaining differential sensitivity to 3-CePs, we measured the activation of the DNA repair machinery as another key aspect in the cellular response to genotoxicants.[38] To verify the ability of the two cancer cell lines to detect double-strand breaks (DSBs), we assessed the phosphorylation of H2AX (γH2AX), an early

event of the DNA damage response (DDR),[39] by flow cytometry after 6 h and 12 h of treatment with both agents (Fig. 1 D). Interestingly, despite the comparable DNA damage accumulation in the two cell lines, only HCT-15 showed an increase in the γH2AX-positive population, suggesting a more efficient engagement of the DNA repair machinery.

Taken together, these results indicated that cell-specific mechanisms after the first event of DNA damage are responsible for the different susceptibilities to 3-CePs.

Treatments elicit cell-specific transcriptional changes

Different genetic and epigenetic factors define the responsiveness of tumor cells to chemotherapeutic agents.[40] To address these globally, we analyzed changes in the transcriptome of the high- and low-sensitive cell lines after treatment with the two 3-CePs (Fig. 1 E). RNA-seq was performed on total RNA of HCT-15 and BxPC-3 cells exposed to DMSO 0.5% (control) or treated with **M** (10 nM) or **B** (200 nM) for 6 h and 12 h (Fig. 2 A, S2 A) as in previous experiments.

Principal component analysis (PCA) of all transcripts separated samples within each cell line according to treatment and time-point (Fig. S2 B), suggesting a clear transcriptional reprogramming after treatment. In fact, differential expression (DE) analysis pointed out that the expression of a large number of genes changed significantly in both cell lines after exposure to 3-CePs (Fig. 2 B, S2 C), especially at 6 h in BxPC-3 cells and upon treatment with **B** in HCT-15 cells.

Gene Ontology (GO) enrichment was performed on the DE genes to determine signaling pathways and transcriptional programs explaining the observed differences. In a first explorative approach, we generated the union of DE genes per cell line irrespective of compound and time point, which allowed us also to distinguish between cell type-specific or shared DE genes (Fig. S2 D). The most representative biological processes identified by this analysis (Fig. S2 E, Supplementary data 1) are reported in Fig. 2 C (see *Methods* and Fig. S2 F for further details).

Unexpectedly, we identified a strong translational response in BxPC-3 cells after treatment, a process which is typically attenuated in stress conditions, as was the exposure to our DNA damaging agents, to allow

proper recovery of the protein quality control machinery.[41,42] In contrast, a strong regulation of genes mediating protein stability and catabolism was observed in the low-sensitive cell line. In addition, HCT-15 cells activated genes involved in the DDR, consistently with their higher ability to detect and respond to DSBs. Both these two mechanisms pointed towards the activation of an adaptive stress response in the low-sensitive cell line.

To further characterize these transcriptional changes over time in a cell type-specific context, we grouped the DE genes at 6 and 12 h in modules according to the similarity in their expression profiles and performed a functional enrichment on genes with similar expression patterns (Fig 2 D and S3 A, Supplementary data 2). Genes involved in ribosome biogenesis and DNA repair turned out to be upregulated particularly after 6 h of treatment in BxPC-3 cells (Clusters *2* and *3*, Fig. 2 D). Besides, silencing of pro-survival genes involved in microtubule organization and the JAK-STAT cascade (Cluster *1*, Fig 2 D) was detected at the same time point. Only after 12 h of treatment (Fig. S3 A), BxPC-3 cells boosted carbohydrate metabolism, most likely an attempt to recover *in extremis*.[43]

Also HCT-15 cells upregulated clusters of genes mediating DNA repair, protein stability and mitochondrial activity as early as 6 h of treatment, suggesting this time point as the most informative to describe the response to 3-CePs (Clusters *4* and *6*, Figure 2 D). In contrast to BxPC-3 cells, HCT-15 downregulated genes involved in translation and ribosome biogenesis from 6 h of exposure (Cluster *7*, Figure 2 D), while intensifying their response to oxidative stress after 12 h (Cluster *17*, Figure S3 A).

This exploratory analysis showed clearly different transcriptional responses and distinct time dynamics in BxPC-3 compared to HCT-15 cells, most likely responsible for their different susceptibility to 3-CePs. In particular, our findings pointed towards DNA repair and proteostasis as key mechanisms tuning sensitivity to the compounds, as further confirmed by inspecting the complete rank of DE genes *via* gene set enrichment analysis (GSEA, Fig. S3 B).[44]

DNA repair and proteostasis are key modulators of the response to 3-CePs

For their key role in the response to 3-CePs, DNA repair and protein homeostasis were further analyzed to clarify their contribution to BxPC-3 sensitivity.

Interestingly, DNA repair was activated in both cell lines early after 6 h of treatment but with a different modulation (Fig. 3 A). First, base-excision repair (BER) was suggested as the preferential pathway of BxPC-3 by GO enrichment while HCT-15 relied mostly on nucleotide-excision repair (NER), unleashing a generally stronger activation of the DDR. In detail, HCT-15 DE genes contributing to the response to the DNA damage stimulus were strongly upregulated already after 6 h especially in response to **B**, while activated only after 12 h in BxPC-3 (Fig. 3 B). In contrast, genes such as *PPP4R2* and *RAD51AP1*, both involved in the first phases of DSBs repair,[45,46] were even downregulated in BxPC-3 cells at 6h.

The more efficient activation of DNA repair in HCT-15 was further confirmed on the overall rank of genes by GSEA at 6 h of treatment (Fig. 3 C). As anticipated, most of the DE genes leading the enrichment in HCT-15 belonged to NER (e.g. *GTF2H3*, *RBX1*) and other recombinational pathways such as Homologous Repair (HR) (e.g. *MMS22L*, *BARD1*) and Fanconi Anemia (FA) (e.g. *BRIP1*, *FANCM*), all better suited for the efficient repair of bulky lesions and highly toxic DSBs and crosslinks.[47–52] On the other hand, DE genes in BxPC-3 cells were mostly related to BER (e.g. *APEX1*, *UNG*) and MMR (Mismatch Repair) (e.g. *MSH6*, *EXO1*), which contribute to the repair of smaller lesions and mismatches.[53,54]

In the analysis, proteostasis was identified as a second key biological process strictly related to genotoxic stress.[55,56] HCT-15 cells engaged the protein folding and catabolism apparatus in response to 3-CePs, especially to **B** already at the early time point (Fig. 3 D). As observed for DNA repair, DE genes contributing to protein catabolism were upregulated as early as 6 h of exposure in HCT-15 cells, while even downregulated at the same time point in BxPC-3 and only upregulated after 12 h (Fig. 3 E). This response involved chaperones and co-chaperones (e.g. *HSPA8*, *HSPA1B*, *BAG2*, *BAG5*), other genes mediating protein catabolism (e.g. *LAMP2*, *CUL3*) and ER morphogenesis (e.g. *RTN4*).[57–60] Interestingly, a transcriptional pattern revealed by GSEA at 6 h of treatment highlighted an intense positive modulation of the PERK-mediated branch of the

unfolded protein response (UPR) specifically in BxPC-3 (Fig. 3 F). Even more enlightening were the DE genes leading the enrichment: *ATF4*, *DDIT3* (CHOP) and *PPP1R15A* (GADD34) were significantly upregulated after 6 h of exposure only in this cell line (Fig. 3 G, H). These genes participate in the PERK-mediated UPR triggering cell death after prolonged ER stress through the aberrant recovery of translation, which induces proteotoxicity.[61,62] This mechanism would reasonably explain the ribosome biogenesis signature observed in BxPC-3 cells. Consistently, recent work reported a particular susceptibility for pancreatic cancer adenocarcinoma to ER stress and protein dyshomeostasis.[63]

Furthermore, the ability of HCT-15 cells to control proteostasis may also depend on the activation of lipid and cholesterol biosynthesis in response to the compounds (Fig. S4 A). In fact, among other known pro-survival functions, these pathways contribute to resolving ER stress through pathways involving e.g. the Stearoyl-CoA Desaturase (*SCD*) enzyme, for which we detected a significant upregulation of the respective transcript in HCT-15 (Fig. S4 B).[64,65]

Overall, the transcriptome analysis of this *in vitro* perturbation experiment allowed us to dissect the different responses to 3-CePs in our model cell lines, pointing towards protein homeostasis and DDR imbalances as mechanisms responsible for the high susceptibility of BxPC-3 cells.

The response to 3-CePs is further regulated at the chromatin level

The transcriptome analysis unveiled a defined framework of responses tuning the sensitivity to 3-CePs. To further characterize them at the epigenetic level, we examined chromatin accessibility in nuclei of BxPC-3 and HCT-15 cells treated with **M** and **B** for 6 h and 12 h (Fig. 4 A, Fig. S5 A) by ATAC-seq.

3-CePs induced evident epigenetic changes in both cell lines, as suggested by PCA (Fig. S5 B) and confirmed by the number of differentially accessible regions (DARs) identified especially in BxPC-3 cells (Fig. 4 B, Fig. S5 C). For further downstream analyses we focused on DARs mapping to promoters, whose specific condensation or compaction contribute to modulation of transcription of associated genes (Fig. 4 B).

Also in this case, to better describe the timing of chromatin remodeling, cell-specific promoter-associated DARs elicited after 6 h and 12 h of treatment were grouped in clusters sharing a similar pattern of regulation and functional enrichment was performed on the associated genes (Fig. 4 C, S5 D, Supplementary data 3).

In BxPC-3 cells, we observed condensation of promoters involved in carbohydrate metabolism and others mediating protein folding and UPR after 6 h of exposure (Cluster *1*, Fig. 4 C), most likely contributing to the transcriptional downregulation of such processes observed at the same time point.[57,66] On the contrary, relaxation of peaks involved in tRNA metabolism and mRNA splicing were detected, in line with the upregulation of translation and RNA processing evidenced by RNA-seq. In HCT-15 cells, relaxation of promoters involved in the DDR, lipid metabolism (Cluster *6*, Fig. 4 C) as well as protein catabolism (Cluster *7*, Fig. 4 C) was observed, again in line with our observations on transcriptome level. Altogether, these results attested that the regulation of elicited transcriptional pathways was accommodated by changes at the chromatin level, adding new information on the possible mechanisms determining the cellular responses to 3-CePs.

A critical step in the analysis of multi-omic datasets is the integration of information obtained from the different layers. Though valuable strategies have been developed in recent years to integrate RNA-seq and ATAC-seq data, alternatives are still required to optimize and enlarge the functional information obtained from the combination of these powerful techniques.[67–69] In this study, we approached data integration through two alternative strategies, that we called *pairwise* and *crosswise.*

As a first level of integration, we identified genes with concordant regulation in RNA-seq and ATAC-seq upon treatment. In this *pairwise* integration, we compared the direction of transcriptional regulation of genes to the accessibility of their promoters, as specified in the *Methods* section and shown in Fig. 4 D. Given the biological delay that could exist between chromatin remodeling and a detectable variation in transcript level, pairwise comparisons were also considered between chromatin changes after 6 h and transcriptional responses after 12 h of treatment.

Among genes with coherent regulation in BxPC-3 cells we found the tumor suppressors *ADRA2A*, *NME1,* and *KLF6* to be upregulated, elicited after treatment with both agents, and *LATS2* and *NME2* specific for **M** and **B**, respectively.[70–73] Besides, other genes were involved in translation and RNA processing such as *RNPS1* and *LARP4B*,[74,75] and apoptosis such as *AEN*, *PAWR,* and *CASP7*.[76–78] Interestingly, BxPC-3 also negatively regulated *TXNIP*, an inhibitor of the oxidative stress regulator thioredoxin, after treatment with **B**.[79] Conversely, among downregulated hits we found apoptosis inhibitors such as *WRAP53* and *TRADD*, as well as *HSPB8*, *CALCOCO1* and *SELENOH*, all involved in the resolution of ER and oxidative stress.[78,80,81] In HCT-15 cells, among identified positively regulated genes some were involved in DNA repair such as *MRE11*, *MDM4*, *RNF138*,[82,83] others were oncogenes such as *VIM* and *ARID4B* or apoptosis inhibitors like *TRADD*.[84,85] Notably, some genes involved in the modulation of the redox balance (*GLRX3*, *SELENOF*) showed double regulation after treatment with **B** as well as others active in proteostasis (*PSMA5* after exposure to **M**, *UBE2N* to **B**).[81,86] Among the downregulated genes, some were associated to cell adhesion (*PLEKHO1*, *ITGB3*, *ICAM1*) and translation (*RPL19*, *RPL13*).[87]

Collectively, *pairwise* integration of RNA-seq and ATAC-seq shed light on genes with robust regulation at the transcriptional and chromatin level, adding further details to the previously identified response pathways.

Crosswise integration expedites the comprehension of multi-omic data

Through the *pairwise* approach, we identified genes with both transcriptional and chromatin regulation which significantly contributed to the observed cellular response. We further evaluated the crosstalk between RNA-seq and ATAC-seq at a different level by focusing on groups of genes co-regulated in the two omic layers. The identification of genes sharing similar regulation across conditions either at the transcriptional or chromatin level would maximize the detection of interacting pathways and regulatory processes, e.g. as a result of chromatin changes in promoters tuning the transcription of a certain gene set. This approach, which we termed *crosswise* integration, was achieved by vertical Construction of Co-expression network analysis (vCoCena).

vCoCena is designed to define modules of genes and/or genomic markers such as DARs with a similar pattern of regulation across conditions in multiple omic datasets. As a first step, we created separate co-expression networks for the RNA-seq and ATAC-seq layers (Fig. 5 A, S6 A). To prevent the construction of a network mostly describing the difference between the two cell lines, we first calculated separate networks for BxPC-3 and HCT-15 cells which were then integrated horizontally (hCoCena).[88] The union of all DE and promoter DAR-associated genes detected in treated conditions was selected as input for constructing all networks. Clustering of the resulting RNA-seq and ATAC-seq networks identified a relevant number of gene modules with highly specific regulatory patterns (Fig. S6 B and C). At this point, the vertical, inter-omic integration (vCoCena) was applied to construct the final network consolidating the information from transcriptome and chromatin accessibility (Fig. S6 D, see *Methods* for details). The new network was then reclustered resulting in integrated modules of co-regulation including nodes originally derived from the two separate layers in different ratios, as shown in Fig. 5 B.

The approach combined genes sharing similar regulation in the respective omic dataset, as approximated by the GFC pattern, with the postulate that genes grouped together cooperate in specific cellular processes. To define the underlying mechanisms, GO enrichment was performed on genes included in each of the modules and representative biological terms for the most relevant clusters were reported in Fig. 5 C (Supplementary data 4). Some modules validated the information obtained through previous analyses (Fig. S6 E): both the *maroon* and *darkorange* clusters suggested macroautophagy as a putative pathway accounting for the enhanced catabolism observed in HCT-15 cells.[89] Consistently, the former RNA-seq-based module was downregulated at 6 h in BxPC-3 but upregulated already after 6 h with **B** in HCT-15, while the latter ATAC-seq-based module included peaks condensing after 6 h only in BxPC-3,confirming the latter cell line as refractory to a rapid engagement of its protein catabolism apparatus. Another mostly RNA-seq-based module validating our previous approach was the *orchid* module, upregulated after 6 h in BxPC-3, containing genes involved in ribosome biogenesis. The *darkgrey* cluster instead, more balanced in terms of contribution from the two omic layers, showed positive regulation only in HCT-15 cells and included hits involved in DDR.

However, the *crosswise* integration also identified additional regulation, exemplified by the *pink* module. As approximated by the associated GFCs pattern, its 163 genes were positively modulated only in HCT-15 cells especially after 6 h of treatment with **B** (Fig. 5 D). Interestingly, functional enrichment identified hits both from RNA-seq and ATAC-seq involved in actin remodeling (Fig. 5 E), a mechanism affecting morphology and function of cancer cells (e.g. *FHOD1*, Fig. 5 F).[90,91,92] Other module genes, such as *FURIN*, positively regulated at the chromatin level (Fig. 5 F), belonged to TGFβ signaling (Fig. 5 E), an emerging player in cancer drug resistance.[93,94] In addition, the module included genes of lipid metabolism and DNA repair belonging to both omic layers, which was in line with our initial findings (Fig. 5 E and F).

Overall, the *crosswise* integration of RNA-seq and ATAC-seq data allowed an efficient combination of the functional information from the two omics layers. Clearly, this approach added further biology to what we had identified when analyzing transcriptional and chromatin landscape regulation individually.

Perturbation-informed basal signatures efficiently predict sensitivity to our candidate drugs

The information derived from the *crosswise* integration was employed to construct a signature of sensitivity to 3-CePs. Being more potent, **M** was selected as reference to describe a sensitivity prediction framework based on the use of a perturbation-informed omic signature (Fig. 6 A, S7 A, Methods).

First, we selected vCoCena clusters with a marked difference in regulation between the two cell lines after treatment with **M,** considering only the most informative time point of 6 h (selected modules: *cyan*, *darkgreen*, *darkgrey*, *darkorange*, *gold*, *indianred*, *khaki*, *lightgreen*, *steelblue*, *orchid*; module selection criteria are described in detail in the *Methods* section). According to our analysis, genes that belong to these modules, coming both from RNA-seq and ATAC-seq analyses, are expected to be the major determinants of the differential susceptibility in the two cell lines.

Importantly, we postulated that features accounting for sensitivity should be intrinsic for the two cell lines, thus explained already by significant differences in their basal status. For this reason, we performed DE

analysis between untreated BxPC-3 and HCT-15 control groups, identifying genes up- and downregulated at the transcriptional level in the high-sensitive cell line, and sorted out only those belonging to previously selected modules. This approach resulted in a subgroup of genes with different basal expression in BxPC-3 cells as well as a sufficiently compound- and cell line-specific regulation upon perturbation. This perturbation-informed signature was composed of 294 genes upregulated (signature *up*) and 170 genes downregulated (signature *down*) in the high-sensitive BxPC-3 cells (Fig. 6 B, gene list available in Supplementary data 5). GO enrichment on these genes identified protein synthesis, folding and catabolism, as well as cell adhesion, matrix organization and actin remodeling among the most significant biological functions (Fig. 6 C). Some interesting genes in the *up* signature were *BNIP3* and *FADD*, both proapoptotic, as well as *TXNIP*, already identified as a thioredoxin inhibitor. Among those composing the *down* signature, we identified *YOD1*, *HERPUD1* and *HSPA5*, involved in protein homeostasis and ER stress, but also *ERCC6* and *AUNIP* of the DDR (Fig. 6 B).[95–98]

To determine the robustness of the obtained signature and its ability to predict sensitivity to **M**, we next performed a gene set variation analysis (GSVA) on publicly available transcriptomes of common cell lines,[99] testing for both the *up* and *down* signatures (Fig. S7 B). A sensitivity score was calculated for each cell line as the difference between the enrichment scores (ES) of the *up* and the *down* signatures. The predicted rank was validated experimentally on representative cell lines (A-431, A549, HEK-293, NTERA-2, PC-3, SH-SY5Y) demonstrating the strong predictive capacity of our perturbation-informed signature (Pearson's R=0.84, p=0.038, Fig. 6 D). This signature outperformed a random one containing the same number of genes (R=-0.48, p=0.34, Fig. S7 C and D) and also a signature of equal size composed by the top up- and downregulated genes between the two cell lines (R=0.34, p=0.51, Fig. S7 E and F, gene list available in Supplementary data 6). Collectively, our crosswise integration approach resulted in a perturbation-informed signature capable of predicting drug sensitivity in a wide range of untreated tumor cell lines commonly used in cancer research.

Encouraged by these results, we adapted our strategy to mimic a clinical setting utilizing the primary tumor samples of the Cancer Genome Atlas

TCGA database (Fig. S7 G). By applying GSVA, we examined the relative distribution of samples from different tumor types based on the calculated sensitivity score, unveiling which cancer types were predicted as generally more susceptible (i.e. kidney renal clear cell carcinoma KIRC, pancreatic adenocarcinoma PAAD, kidney renal papillary cell carcinoma KIRP, glioblastoma multiforme GBM) or less sensitive (i.e. tenosynovial giant cell tumor TCGT, cervical squamous cell carcinoma and endocervical adenocarcinoma CESC, rectum adenocarcinoma READ, colon adenocarcinoma COAD) to **M**, providing a framework for further *in vivo* development of this compound (Fig. 6 E).

Interestingly, the predicted tumor types with the highest and lowest sensitivity turned out to be KIRC and TCGT, respectively, demonstrating that the designed signature was not driven by the original cell type of the cell lines used for its extrapolation and could go beyond the original cancer type. At the same time, PAAD and COAD (pancreatic and colorectal adenocarcinoma, as BxPC-3 and HCT-15 cells) were still among the most and least sensitive, confirming that cell type intrinsic determinants of susceptibility exist and are represented in our signature. Interestingly, intra-tumor variability resulted in a continuous distribution of samples scores within each cancer group, confirming the importance of clinically translating such predictions beyond the tumor type to better address patient-specific therapeutic needs.

To enlarge the accessibility and clinical translatability of our framework, we finally introduced a LASSO regression model to predict the sensitivity of tumor samples in the external reference dataset (Fig. 6 F). We trained a regression model using TCGA basal transcriptomic profiles labelled with the previously predicted sensitivity scores in order to create a self-supervised system able to emulate the prediction irrespective of the context dataset, detaching the predictive tool from the data space. From a clinical perspective, this further step would permit to collect a patient basal transcriptome and feed it to the model, not only improving the performance of the prediction but also avoiding any issue related to data sharing since the model itself does not contain any patients' sensitive data.

In detail, TCGA samples were labeled according to the calculated continuous sensitivity scores. Next, the model was trained on 80% of the

data and tested on the remaining 20%, which efficiently predicted drug sensitivity within the test samples (R = 0.97, RMSE = 0.063) (Fig. 6 G). Notably, such predictive capacity was maintained even when excluding from the transcriptomes all the signature genes used to define the sensitivity score label of the samples, suggesting the biological robustness of the predictive system (R = 0.97, RMSE = 0.066, Fig. S7 H). In fact, while the signature itself was good enough to rank samples based on experimental biological evidence, the model showed to go beyond the initial signature relying on additional predictive features previously not identified.

Overall, we demonstrated how to further employ the integrated RNA-seq and ATAC-seq information to assemble an accurate and clinically-accessible predictive strategy able both to orient drug development and to support the medical practice in the context of precision oncology.

Discussion

Despite the advances of the last decades, efforts are continuously required to expedite routine use of omic-scale approaches in clinical and pre-clinical settings. Recent work illustrated the potential for omics technologies to accelerate the process of drug discovery from the initial identification of candidate lead compounds up to their pre-clinical and clinical development.[8–14] Further, improvements in computational approaches for omics data analyses[4,6,7] and an ever-increasing availability of public reference datasets[100] make it now possible to develop completely new pipelines to address the pharmacological profile of any given drug, from its MoA to sensitivity biomarkers.[1–3]

Here, we combined transcriptome and chromatin accessibility analyses within perturbation experiments to investigate the specific activity profile of 3-CePs, a new class of potential anticancer agents acting as DNA alkylators.[15–21] Our combined analysis unveiled the basis of the preferential activity of 3-CePs against the pancreatic cancer cell line BxPC-3, which was demonstrated to be unable to properly control proteostasis and DDR under stress conditions upon exposure to the alkylating agents. On the contrary, the low-sensitive colorectal adenocarcinoma cell line HCT-15 potentiated protein folding and catabolism all together activating a more efficient DNA repair after treatment. Due to unresolved genotoxic stress and proteostasis dysregulation, widely described as crosstalking events,[55,56] BxPC-3 cells activated the apoptotic branch of the PERK-mediated UPR via CHOP and GADD34, both upregulated after treatment.[61,62] Accordingly, such behavior is in line with the described susceptibility of pancreatic cancer adenocarcinoma to ER stress and protein dyshomeostasis.[63]

Beyond validating the described results, the analysis of chromatin accessibility was first employed to identify genes with concordant transcriptional and epigenetic regulation, a step we called *pairwise* integration. Among these genes, we found apoptotic mediators and tumor suppressors upregulated in BxPC-3 and downregulated in HCT-15, as well as redox balance and proteostasis hits upregulated in HCT-15 and downregulated in BxPC-3.

To further evaluate the interaction between transcriptional and chromatin accessibility responses, we proposed here a new versatile approach for the *crosswise* integration of RNA-seq and ATAC-seq, based on vCoCena (vertical Construction of Co-expression network analysis). This approach identified modules of genes co-regulated in the two omic layers across the analyzed experimental conditions. With this standalone method, we not only recapitulated the result of the independent transcriptomic and epigenomic analysis, but we also discovered additional pathways, e.g. actin and TGF β signaling, which modulate the response to the compounds. In detail, actin dynamics were recognized to potentially assist DSBs repair[91] and a protumorigenic role was established for TGF β in mediating epithelial-mesenchymal transition, both processes that could additionally explain the more efficient response of HCT-15 cells to 3-CePs.[93] Efficient and versatile, this approach demonstrated to represent a valid option to integrate the information from multi-omic studies substituting the separate examination of each omic dataset.

To further assist the development of 3-CePs, we set up a pilot sensitivity prediction framework readily transferable from the bench to the clinics. We designed a perturbation-informed signature derived from the integrated omic layers filtering the differentially expressed genes between the two cell lines at a steady state for those specifically involved in the cellular response to the treatment. Though based on a limited number of perturbed profiles, this gene signature predicted with high precision the sensitivity to 3-CePs only relying on the untreated transcriptome of test cell lines. The possibility to improve predictions from basal transcriptomes sounds attractive from a clinical perspective since it overcomes the need to screen for thousands of drugs and collect the same amount of profiles from limitedly-available patient samples, such as biopsies.[101] Applied to TCGA tumor samples, this approach provided a list of susceptible cancer types, e.g. KIRC and PAAD, to support the further development of our drug candidate, and, once transferred on an ML platform, could offer a versatile predictive strategy translatable to the clinics.[6,102]

In this study, we combined transcriptomic and epigenetic data to guide our exemplary analysis. Nevertheless, the modularity of our framework allows, with only minimal adjustment, its application to other omic technologies or experimental designs. Indeed, the vCoCena integration,

565 which is instrumental for both the biological interpretation of the data and
566 the definition of the perturbation-informed signature, is agnostic of the type
567 of data used as soon as this is reduced to a network of co-regulation.

568 In conclusion, we present a complete end-to-end workflow to implement
569 the use of multi-omics in drug development, providing a human-readable
570 toolbox to interrogate pharmacological questions in both pre-clinical and
571 clinical settings. We applied this framework to understand the MoA of 3-
572 CePs revealing the cellular determinants of sensitivity to this novel class
573 of drugs and providing precious information for their clinical development
574 as anticancer candidates. Given its versatility, we envision our workflow
575 to be a broadly applicable resource to assist researchers in different steps
576 of the drug discovery and development process.

577

Methods

Cell lines culturing

Colon (HCT-15), pancreatic (BxPC-3), lung (A549) carcinoma cell lines and human embryonic kidney (HEK-293) cells were purchased from ATCC (American Type Culture Collection) while prostate (PC-3) and testis (NTERA-2) carcinoma cell lines were kindly provided by Prof. W. Kolanus (LIMES institute; University of Bonn), neuroblastoma (SH-SY5Y) by Prof. D. Schmucker (LIMES institute; University of Bonn) and epidermoid (A-431) carcinoma by Prof. G. Zunino (Istituto Nazionale dei Tumori di Milano). Cell lines were maintained in logarithmic phase at 37 °C in a 5% carbon dioxide atmosphere using RPMI-1640 (for BxPC-3, HCT-15, PC-3), DMEM (for A-431, HEK-293, NTERA-2, SH-SY5Y) or Ham's F-12K (for A549) media (by Gibco or Euroclone) containing 10% fetal calf serum, antibiotics (50 units/mL penicillin and 50 µg/mL streptomycin) and 2 mM L-glutamine (Euroclone).

Direct detection and quantification of early DNA damage

The extent of early DNA damage induced by 3-CePs in treated cells was assessed by the Fast Micromethod single-strand-break assay. This approach can detect both single and double-strand breaks, as well as alkali-labile adduct sites in the DNA of treated cells. 5,000 cells/well were seeded in 96-well microplates and treated next day for 6 h with **M** (10 nM and 100 nM), **B** (200 nM and 2 µM) or DMSO 0.5%. After treatment, we measured the effect of double and single-strand breaks on the rate of unwinding of cellular DNA in denaturing alkaline conditions by monitoring the fluorescence of a dye that preferentially binds to dsDNA up to 20 min (Pico488 dsDNA quantification reagent, Lumiprobe). The assay was performed following the protocol of Schröder et al.[34] Two experimental replicates were performed, each one including three technical repeats. Fluorescence signal was acquired by the FLUOstar Omega microplate reader using Omega 5.11 software (BMG LABTECH). The resulting curves based on mean normalized fluorescence values obtained for each treatment and the control (DMSO 0.5%) are reported in Fig. 1 B.

Cell cycle and flow cytometric H2AX phosphorylation analyses

Possible effects of 3-CePs treatments on the cell cycle distribution of both cell lines were analyzed by FACS, staining cellular DNA with the PI (propidium iodide) dye. In addition, we monitored by antibody staining the phosphorylation of histone H2AX, upstream event of the DDR cascade, after 6 h and 12 h of treatment in order to investigate the ability of BxPC-3 and HCT-15 cells to detect DSBs. 200,000 cells/well were seeded in 12-well plates and treated next day for 6 h, 12 h or 72 h with **M** (10 nM), **B** (200 nM) or DMSO 0.5%. Cells were harvested, washed with PBS, fixed and permeabilized with the Foxp3 Transcription Factor Staining Buffer Set (eBioscience, cat. #00-5523-00). In detail, cell suspensions were fixed for 1 h at room temperature with FixBuffer, washed twice with PermBuffer and stained with anti-human γH2AX AlexaFluor 488 (Biolegend, clone 2F3, cat. #613405) for 1 h at 4 °C. After the first staining, cells were washed first with PermBuffer, then with PBS and stained secondly with PI (30 min, dark). Samples were acquired on a BD Symphony instrument equipped with 5 lasers (UV, violet, blue, yellow-green, red), the spectral overlap between the channels were determined with single stained samples using FACSDiva (v 9.1.2). Samples were analyzed in FlowJo (BD, v 10.7.1). Events were gated first according to FSC-A and SSC-A and cleaned from cell doubles with 3 consecutive gates (FSC-A vs. FSC-H; SSC-A vs. SSC-H and PI-A vs. PI-H). The frequency of cells within each phase of the cell cycle was calculated using the PI-A signal with the FlowJo built-in algorithm (Watson model with constrained G2 peak). Three biological replicates were obtained per condition and unpaired two-tailed Student's *t*-test was performed to assess statistical significance ($p < 0.05$).

RNA-seq and ATAC-seq experiments

For both RNA-seq and ATAC-seq analyses, 300,000 cells/well were seeded in 6-well plates and treated next day for 6 h, 12 h or 72 h with **M** (10 nM), **B** (200 nM) or DMSO 0.5%. Both for RNA-seq and ATAC-seq samples, three experimental replicates were obtained for each condition.

RNA-seq: at the end of the treatment, cells were washed, resuspended in 1 mL QIAzol reagent (Qiagen) and stored at -80 °C.

ATAC-seq: at the end of the treatment, cells were washed, harvested, resuspended in PBS with EDTA, stained with the LIVE/DEAD Near-IR fixable dye (Invitrogen, cat. #10119) for 10 min at 4 °C, centrifuged and

suspended in PBS with EDTA. 20,000 living cells/sample were sorted by FACS and further processed for nuclei isolation and transposition reaction following the protocol of Buenrostro et al.[29]

We extracted the RNA using the miRNeasy mini kit (Qiagen) and checked the RNA integrity and quantity using the tapestation RNA assay on a tapestation4200 instrument (Agilent). We used 750ng total RNA to generate NGS libraries using the TruSeq stranded total RNA kit (Illumina) following manufacturer's instructions and generated ATAC-libraries from tagmented cells following the protocol of Buenrostro et al. In both cases we checked library size distribution via tapestation using D1000 (RNA) and D5000 assays (ATAC) respectively on a Tapestation4200 instrument (Agilent) and quantified the libraries via Qubit HS dsDNA assay (Invitrogen). We clustered the libraries at 250pM final clustering concentration on a NovaSeq6000 instrument using SP and S2 v1 chemistry (Illumina) and sequenced paired-end 2*50 cycles before demultiplexing using bcl2fastq2 v2.20.

RNA-seq data analysis

Reads were aligned and quantified with STAR (v 2.5.2a)[103] using standard parameters and mapped against the GRCh38p13 human reference genome (Genome Reference Consortium). Raw counts were imported, pre-filtered to exclude low-count genes (<100 reads, 17.693 mapped transcripts), normalized and VST-transformed (variance stabilizing transformation) following the DESeq2 (Bioconductor, v 1.26.0) pipeline using default parameters.[104,105] SVA (surrogate variable analysis) was applied to identify latent variables responsible for batch effects and four of them were included in the DESeq2 model.[106] All present transcripts were used as input for principal component analysis (PCA). The call for differentially expressed genes was performed for all treated vs control comparisons (separate cell lines) using an adjusted p-value threshold equal to 0.05, where IHW (IHW: independent hypothesis weighting) was adopted for multiple testing. Only protein-coding hits were considered for further functional analyses on DE genes. GSEA (gene set enrichment analysis) based on the GO (gene ontology) *biological process* database was employed for functional enrichments, both based on DE genes (Supplementary data 1, 2) or $\log_2$FC-based ranks. All enrichment dotplots report the Count and p-value associated with each term, when $p < 0.05$.

Representative enrichment terms in Fig. 2 C were selected manually from enrichment maps obtained for each group of genes depicted in the dotplot (Supplementary data 7): to remove semantic redundancy, only the most significant nodes among those converging into the same hub were reported (higher Count and lower p-value, example in Fig. S2 F). SVA batch-corrected normalized vst-transformed counts were used as input for boxplots, heatmaps and log_2FC-based GSEA. Hierarchical clustering was applied to identify blocks of DE genes with similar regulations across conditions as reported in the presented heatmaps (Fig. 2 D, S3 A). In the same heatmaps, row-scaled expression levels of cell-specific DE genes elicited at 6 h and 12 h were reported separately for each of the analyzed conditions.

ATAC-seq data analysis

After adapter trimming using Trimmomatic v 0.36[107], the sequencing reads were aligned bowtie2 v 2.3.5 against the GRCh38p13 human reference genome.[108] Subsequently, duplicated reads were removed using Picard *dedup* function and the transposase-induced offset was corrected using the deeptools v 3.1.3 *alignmentSieve* function.[109] After sorting and indexing bam files with samtools v 1.9.,[110] peak calling was performed using MACS2 v 2.1.2.[111] Peak regions from sample-specific peak calling results were unified in R v 3.6.2 using the *reduce* function implemented in the GenomicRanges package v 1.38.0.[112] prior to quantification of sequencing reads in these unified peak regions using the *summarizeOverlaps* function implemented in the GenomicAlignments package v1.22.1.[112] Raw counts were pre-filtered to exclude low-count peaks (<20 reads, 63.434 mapped peaks), normalized and VST-transformed following the DESeq2 (Bioconductor, v 1.26.0) pipeline using default parameters.[104,105] Peak regions were annotated using *ChIPseeker* v1.22.1. All present peaks were used as input for principal component analysis (PCA). The call for differentially accessible regions (DARs) was performed for all treated vs control comparisons (separate cell lines) considering a $p < 0.05$ threshold. Only peaks mapping in promoters of protein-coding regions were considered for further functional analyses. GSEA (gene set enrichment analysis) based on the GO (gene ontology) *biological process* database was employed for functional enrichments based on DAR-associated genes (Supplementary data 3). Normalized

and vst-transformed counts were used as input for heatmaps and boxplots. Hierarchical clustering was applied to identify blocks of DAR-associated genes with similar regulations across conditions as reported in the presented heatmaps (Fig. 4 C, S5 D). In the same heatmaps, row-scaled accessibility levels of cell-specific DARs at 6 h and 12 h were reported separately for each of the analyzed conditions. For the *pairwise* integration between transcriptional and chromatin accessibility data, we identified hits having the same sign of regulation in RNA-seq and ATAC-seq which were DE (protein-coding) and/or DAR-associated (protein-coding mapping in promoters). Since a delay could exist between a prior chromatin remodeling and a detectable variation in the respective transcript level, pairwise comparisons were considered not only at the same time point in both omic layers but also between chromatin changes at 6 h and transcriptional responses at 12 h. We reported in Fig. 4 D only hits with ($log_2FC_{RNA\text{-}seq}$ + $log_2FC_{ATAC\text{-}seq}$) > 1 or < -1. Interesting gene names for each of the considered comparisons were also reported.

Crosswise integration of RNA-seq and ATAC-seq data

The *crosswise* integration of transcriptomic and chromatin accessibility data was achieved through an adaptation of the CoCena (construction of co-expression network analysis - automated) tool, which can identify modules of genes showing similar regulation across conditions of interest. The core principles driving both network construction and gene modules detection by CoCena have been described previously.[88] In this analysis, we first optimized the design of separate co-expression networks for the RNA-seq and ATAC-seq layers. To avoid the creation of networks mostly describing cell type differences, we calculated separate networks for BxPC-3 and HCT-15 cells which were then integrated horizontally through hCoCena[2].[88] The union of all DE and promoter DAR-associated genes detected in treated conditions was selected as input for constructing all networks. For the construction of cell-specific networks, the specified Pearson correlation cutoffs, edges and nodes for RNA-seq (BxPC-3: cutoff=0.801, edges=356851, nodes=4266; HCT-15: cutoff=0.772, edges=154497, nodes=4321) and ATAC-seq (BxPC-3: cutoff=0.702, edges=48280, nodes=3479; HCT-15: cutoff=0.733, edges=13336, nodes=3350) were used. The horizontally integrated networks contained the union of all nodes and edges coming from parent networks, where

edges between nodes connected in both parent layers were recalculated as a mean of their original weights. Clustering of the resulting RNA-seq and ATAC-seq networks was performed based on the *infomap* algorithm, where a threshold of minimum of 15 nodes per cluster was applied (Fig. S6 B, C).[113]

Subsequently, inter-omic integration by vCoCena was applied to construct the final network. In this case, the correlation between the mean group-fold change (GFC) pattern of modules belonging to the two layers was calculated to identify clusters of genes with similar regulation, suitable for *crosswise* integration. Edges from the two separate networks were selected for contributing to the integrated one based on a minimum cross-layer correlation which could guarantee the maximum mixture between layers in identified module pairs (minimum correlation cutoff=0.73, edges=628783, nodes=8067). The new network was reclustered exploiting again the *infomap* algorithm, applying a higher threshold of a minimum of 30 nodes per cluster, and mean GFCs were recalculated: the resulting integrated modules included nodes originally derived from the two separate layers in different ratios, as shown in the relative heatmap (Fig. 5 B). GO-based GSEA was performed on detected modules of genes (Supplementary data 4) and the most significant terms ($p < 0.05$) were reported.

Sensitivity signature construction and prediction pipeline

For the signature of sensitivity to **M**, relevant modules from the crosswise vCoCena integration were selected as follows (Fig. S7 A): for each module, in both cell lines separately, we calculated the difference between the GFC (group fold-change) of the control and the **M** 6 h treated groups (ΔGFC(cell line) = GFC(**M**6h) - GFC(ctrl)). The early time point was selected to guide the signature construction since from upstream analyses it turned out to be the most informative of cell responses to 3-CePs. The threshold score was then calculated as the difference between the previously obtained ΔGFCs for the two cell lines (thr_{score} = ΔGFC(BxPC-3) - ΔGFC(HCT-15)). Modules with thr_{score} above q50, thus modules where

the regulation was sufficiently different in the two cell lines after treatment with **M**, were selected (*cyan*, *darkgreen*, *darkgrey*, *darkorange*, *gold*, *indianred*, *khaki*, *lightgreen*, *steelblue*, *orchid*). Genes from the identified modules were grouped together and further considered to drive the definition of our signature of interest.

Further on, DE analysis was performed between BxPC-3 and HCT-15 untreated control groups to identify baseline DE genes up- and downregulated in the high-sensitive cell line (log_2FC threshold equal to 1, padj < 0.01). In fact, given the much higher availability and clinical spendability of RNA-seq compared to ATAC-seq profiles, the signature was finally constructed only from basal transcriptomes. In particular, we further selected among the identified module genes only those that were also DE between the two untreated controls, ending up with a restricted group of genes showing compound- and cell line-specific regulation upon perturbation but, meanwhile, a significantly different basal expression in BxPC-3 cells. This perturbation-informed signature was composed by 294 genes upregulated (signature *up*) and 170 genes downregulated (signature *down*) in the high-sensitive BxPC-3 cells (listed in Supplementary data 5).

To validate the predictive performance of the obtained signature, GSVA (gene set variation analysis) was performed both with *up* and *down* signatures on the basal RNA-seq profiles of cancer cell lines included in the HPA (Human Protein Atlas).[99] A sensitivity score was calculated for each cell line as the difference between the ESs (enrichment scores) of the *up* and the *down* signatures. The predicted rank was validated on selected cell lines (A-431, A549, HEK-293, NTERA-2, PC-3, SH-SY5Y) as described in the next paragraph and Pearson correlation between predicted sensitivity scores and viability decrease in cells treated with **M** 10 nM for 72 h was calculated. Two control signatures of the same size were also tested: 1) a random genes signature (composed by random genes among those annotated in the RNA-seq profile of HPA cell lines) 2) a control signature composed by the top up- and down- log_2FC DE genes between the two cell lines (listed in Supplementary data 6).

GSVA (v 1.38.2) was applied also on basal transcriptomes of samples from the Cancer Genome Atlas TCGA database and their sensitivity score was calculated as previously indicated. The relative distribution of

822 samples from different tumor types in terms of calculated sensitivity score
823 was plotted in Fig. 6 E, together with the indicated median value for each
824 group, to identify possibly more susceptible tumor types.

825 Finally, the signature-based prediction was used to train a LASSO-based
826 classifier (*cv.glmnet* function in *glmnet* package v 4.1 to assess lambda
827 penalty, *predict* function in *stats* package v 4.0.3 for actual prediction).
828 Briefly, TCGA samples were assigned to a continuous label based on the
829 previously inferred sensitivity scores. We next trained the classifier with
830 80% of these profiles and tested it on the remaining 20%: Pearson
831 correlation and RMSE were calculated to evaluate the predictive
832 performance of the classifier. To assess the biological robustness of our
833 signature and of the obtained model, the classifier was trained and tested
834 also using transcriptomes cleaned up from genes belonging to our
835 signature.

836 **Validation of 3-CePs sensitivity prediction on cancer cell lines**

837 The rank of sensitivity to **M** obtained from the newly constructed signature
838 was validated on a subset of available cell lines included in the Human
839 Cell Atlas. The selected cell lines spanned quite well between the max
840 and min detected susceptibility scores. Here are the screened cell lines
841 from the one predicted as most sensitive: PC-3, A549, A-431, SH-SY5Y,
842 NTERA-2, HEK-293. 5,000 cells/well were seeded in 96-well microplates
843 and after 24 h treated with **M** 10 nM for 72 h. Cell viability was assessed
844 at the end of the treatment by MTT, following previously adopted
845 protocols.[19] Mean values of residual viability and standard deviations
846 obtained from two independent experiments in duplicated microplates,
847 each one containing three technical replicates, are reported in Table S1.
848 Pearson correlation between mean residual viability and predicted
849 susceptibility score in considered cell lines was calculated and reported in
850 Fig. 6.

851 **Statistics and reproducibility**

852 Sample size was defined empirically to ensure robust statistical analysis.
853 Unpaired two-tailed Student's *t*-test was performed to assess statistically
854 significant differences ($p < 0.05$) in cell cycle and H2AX phosphorylation
855 analyses between treated and control conditions ($n=3$). All correlation
856 coefficients were calculated with a Pearson's test. The adopted statistical

tests, the considered significance levels and the number of biological replicates are also reported in figure legends. Box plots are in the style of Tukey, where the center of the box represents the median of values, hinges represent the 25th and 75th percentile and the whiskers are extended not further than the 1.5 * IQR (inter quartile range). The analysis was performed on R (v. 3.6.2 or 4.0.3): the specific packages used for the analysis, their version and relevant parameters used are explained in the *Methods* sections. All plots were generated with *ggplot* (v. 3.3.2) except for the heatmaps which were generated with the R package *complexheatmap* (v. 2.2.0). To ensure the reproducibility of the manuscript results, all the analyses were conducted within a containerized environment (Docker). RNA-seq and ATAC-seq analyses were performed with the docker image jsschrepping/r_docker:jss_R362 (https://hub.docker.com/r/jsschrepping/r_docker). The rest of the analysis was conducted with the image lorenzobonaguro/cocena:v3 (https://hub.docker.com/r/lorenzobonaguro/cocena) for compatibility with the CoCena pipeline.

Data Availability

All raw data included in this study are available at gene expression omnibus (GEO). Raw RNA-seq data and count matrix under the GEO accession number GSE179057. Raw ATAC-seq data and peak matrix are available under the accession number GSE179059. Both datasets are collected in a GEO SuperSeries (GSE179064).

During the review process reviewer can access the private dataset at the link:

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179064

using the provided access token.

The cell line expression data employed in the prediction pipeline were downloaded from https://www.proteinatlas.org/about/download. The file *RNA HPA cell line gene data* contains transcript expression levels summarized per gene in 69 cell lines and is based on the Human Protein Atlas version 20.0 and Ensembl version 92.38.

Similarly, the TCGA expression data from cancer cell samples (the Cancer Genome Atlas) were downloaded from the same web page of the Human Cell Atlas (*Transcript expression levels summarized per gene in 7932 samples from 17 different cancer types*). Data are based on The Human Protein Atlas version 20.0 and Ensembl version 92.38.

Code availability

The code to reproduce both pre-processing and downstream analyses reported in this manuscript will be made publicly available on GitHub upon acceptance. The CoCena script is accessible at https://github.com/Ulas-lab/CoCena2.

Author contributions

Conceptualization was by C.C., L.B, S.M., V.G., R.G., A.C.A., J.L.S and B.G. The methodology was devised by C.C., L.B., J.S.-S., M.O, S.W.-H., T.U., while C.C, L.B., A.H., T.H., M.D.F., K.H. performed formal experiments and C.C., L.B., J.S.-S. the data analyses. C.C, L.B. carried out the investigations. The manuscript was written by C.C, L.B., J.L.S and B.G. while all authors reviewed and edited it. The project was supervised by C.C., L.B, J.L.S and B.G. Resources were provided by V.G., J.L.S., B.G.

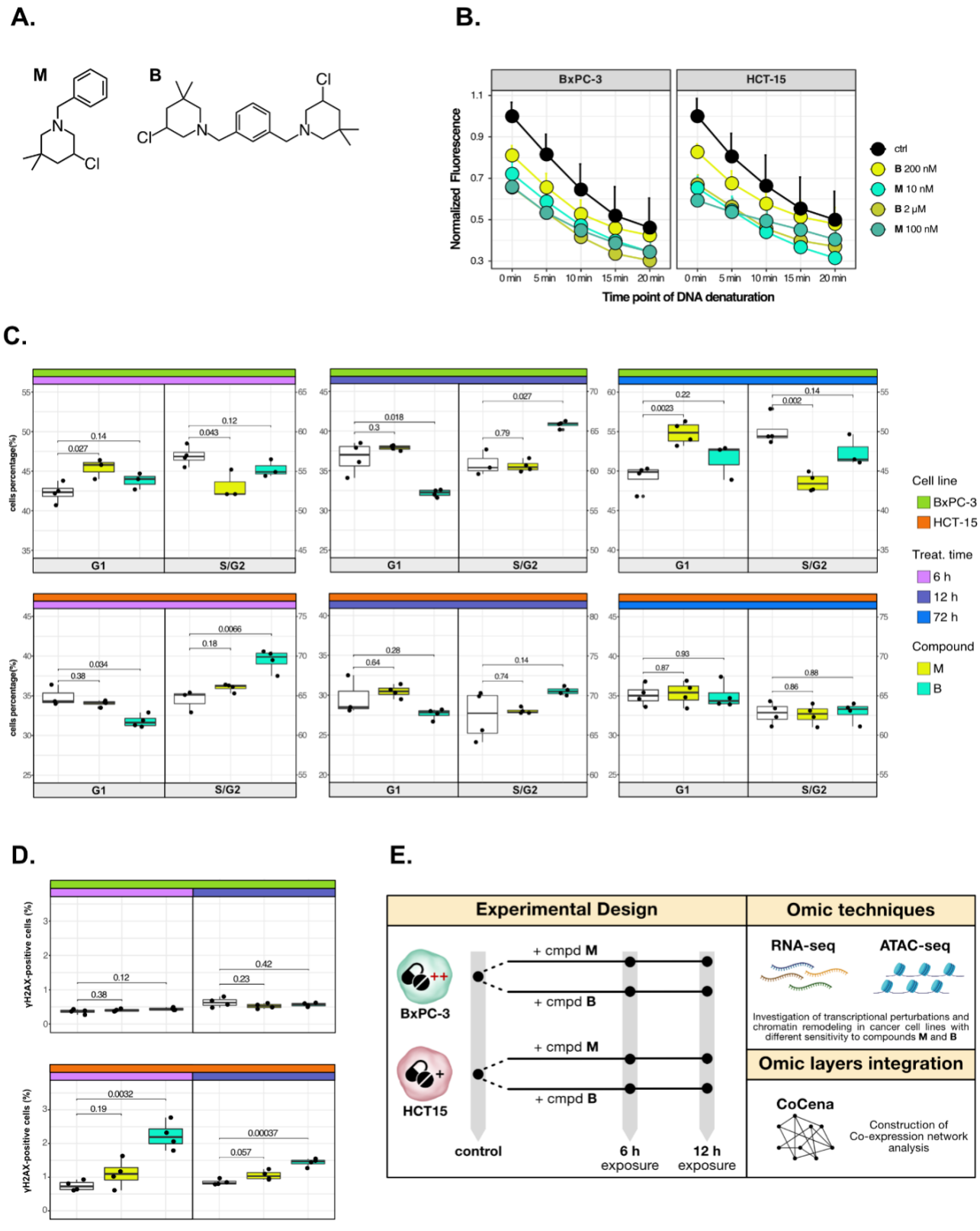
Correspondance

Correspondence to J. L. Schultze (j.schultze@uni-bonn.de) and B. Gatto (barbara.gatto@unipd.it).

Competing interests

The authors declare no competing interests.

Figure 1



913

Figure 1. Cancer tropism of 3-CePs is not explained by DNA damage

A Chemical structure of the analyzed 3-CePs (M = monofunctional, B = bifunctional). **B** Quantification of genomic DNA damage in BxPC-3 and HCT-15 cells treated with M (10 nM and 100 nM), B (200 nM and 2 µM) or DMSO 0.5% (ctrl) for 6 h and analyzed by the Fast Micromethod single-strand-break assay: alkaline denaturation of DNA is followed in time up to 20 min by monitoring the fluorescence of the dsDNA-specific PicoGreen dye. **C** Cell cycle distribution (accumulation in G1 vs G2/S phases) of BxPC-3 and HCT-15 cells treated with M (10 nM), B (200 nM) or DMSO 0.5% for 6 h, 12 h and 72 h analyzed by FACS. Three biological replicates were obtained per condition and unpaired two-tailed Student's *t*-test was performed to assess statistical significance ($p < 0.05$). **D** Analysis of H2AX phosphorylation in BxPC-3 and HCT-15 cells treated with M (10 nM), B (200 nM) or DMSO 0.5% for 6 h and 12 h analyzed by FACS. Three biological replicates were obtained per condition and unpaired two-tailed Student's *t*-test was performed to assess statistical significance ($p < 0.05$). **E** Schematic representation of the adopted omic-based approach.

Figure 2

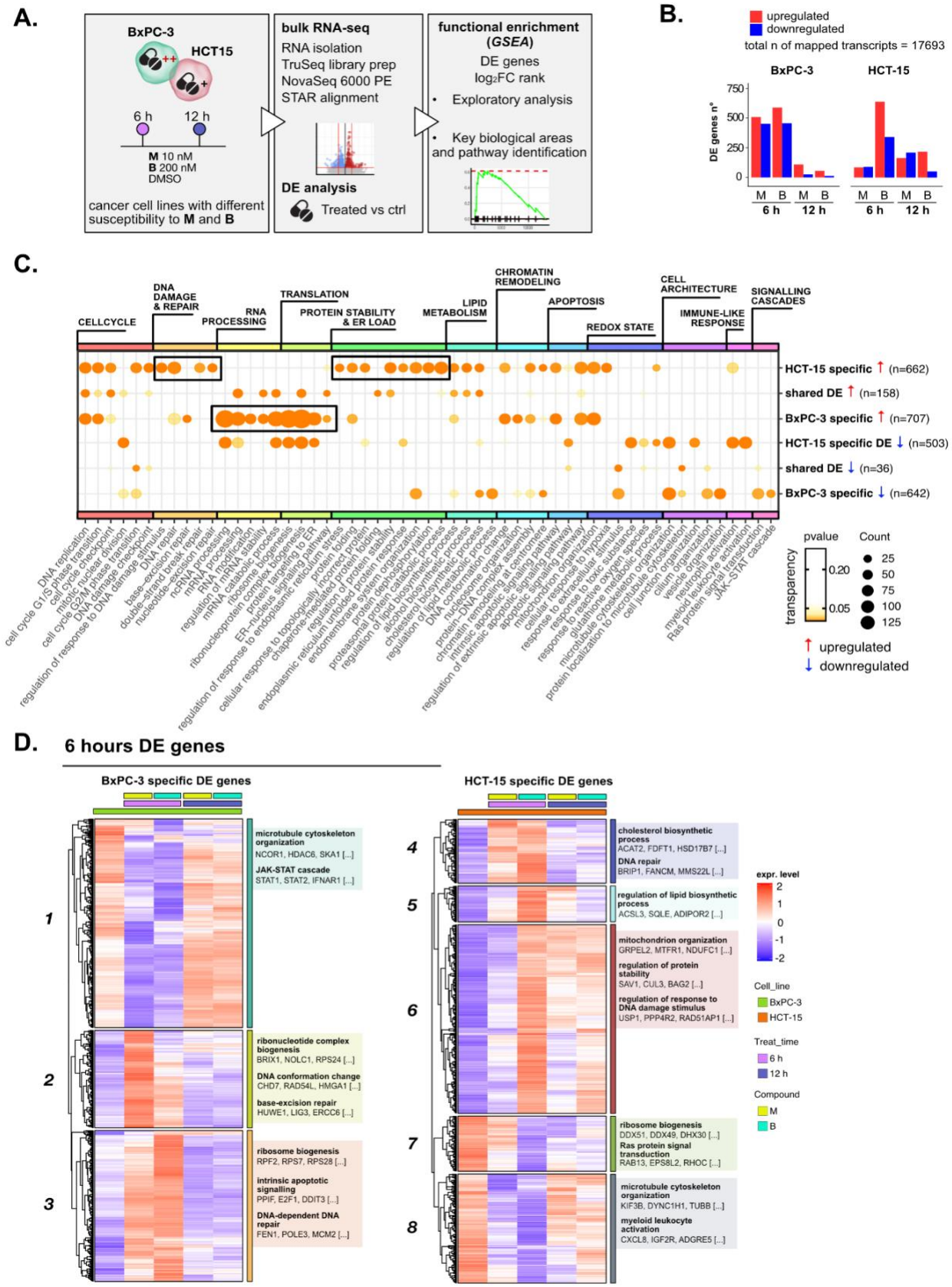
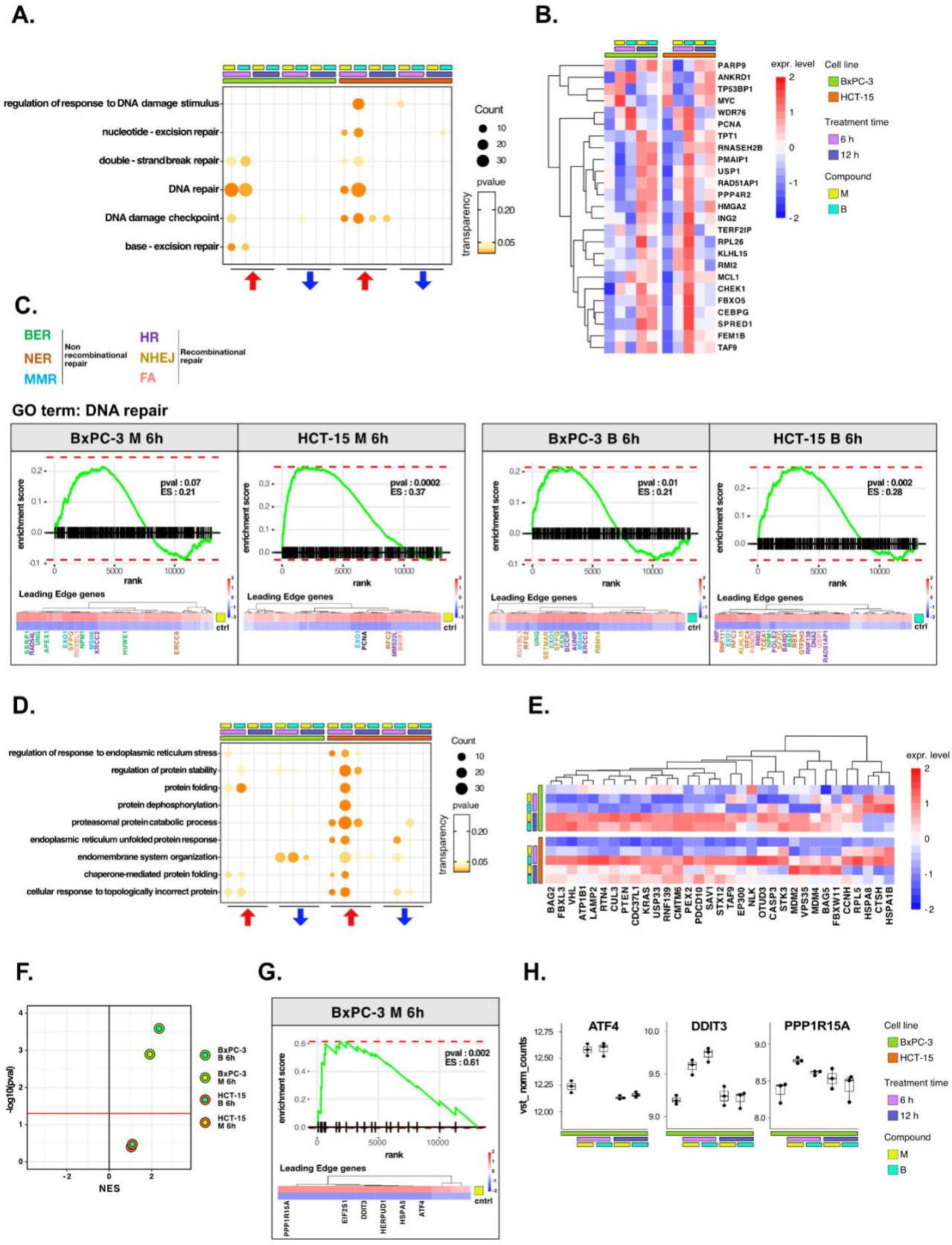


Figure 2. Treatments elicit cell-specific transcriptional changes

A Overview of the applied workflow for the RNA-seq analysis. **B** Number of up- (red) and downregulated (blue) DE genes in BxPC-3 and HCT-15 cells after treatment with M (10 nM), B (200 nM) or DMSO 0.5% (ctrl) for 6 h and 12 h (adjusted p threshold = 0.05, shrinkage = TRUE, multiple testing method = IHW). **C** GO database functional enrichment (GSEA) on cell-specific and shared up- and downregulated DE genes. For each identified biological process, enrichments in terms of Count and p-value of representative terms are reported ($p < 0.05$). **D** Expression level of cell-specific 6 h DE genes across test conditions. GSEA was performed on modules with similar regulation identified by hierarchical clustering: for each cluster, representative GO terms and genes of the associated load are reported.

Figure 3

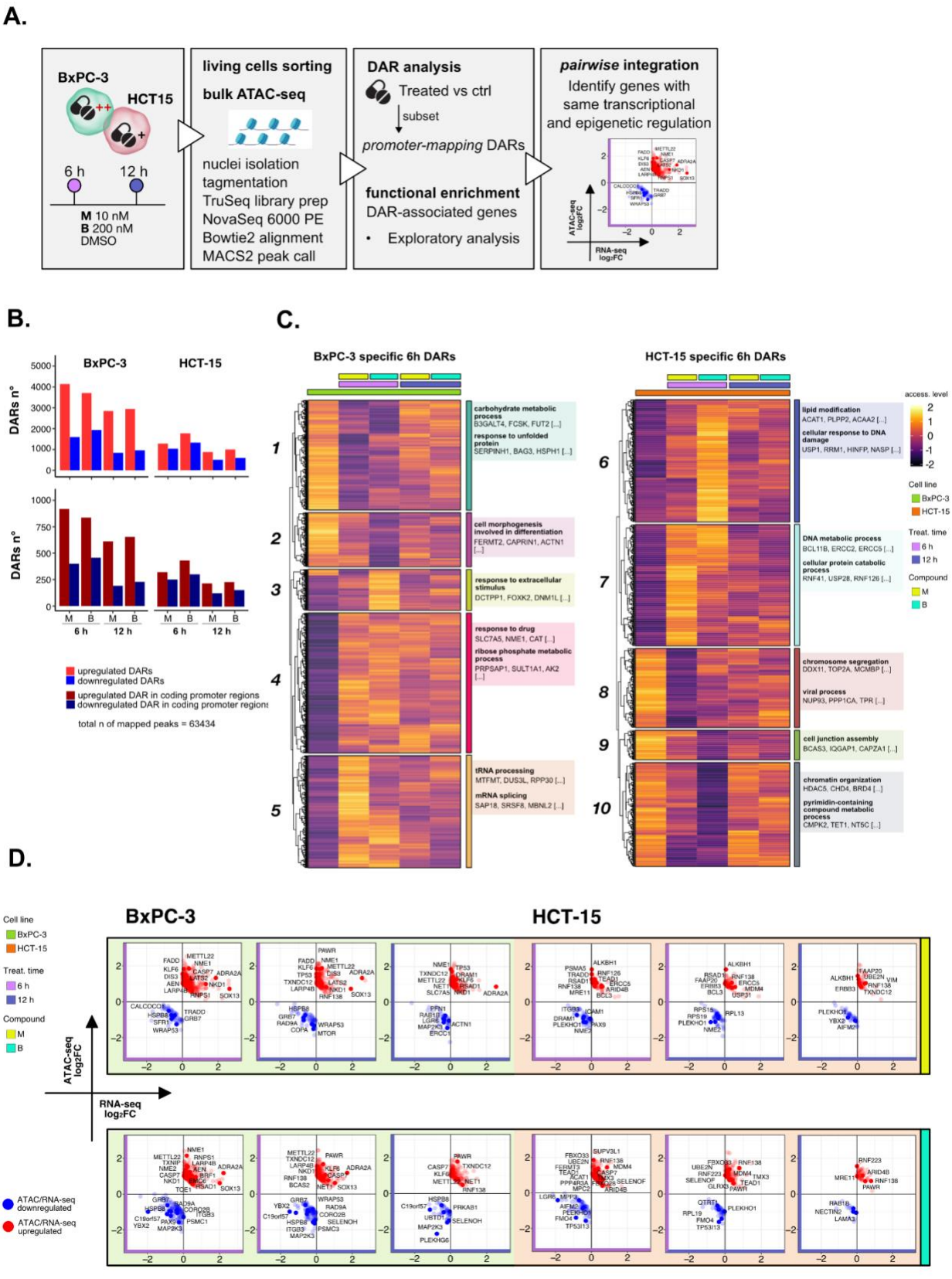


941

Figure 3. DNA repair and proteostasis are key modulators of the response to 3-CePs

A GSEA for terms related to *DNA damage and repair* performed on DE genes detected in each of the considered treated vs control comparisons. For each GO term ($p < 0.05$), enrichments in terms of Count and p-value are reported. **B** Expression level of DE genes included in the load of the GO term *regulation of response to DNA damage stimulus* (HCT-15 cells, $p < 0.05$) in BxPC-3 and HCT-15 cells. **C** GSEA enrichment plots for the *DNA repair* pathway obtained from log_2FC ranks for each of the considered treated vs control comparisons. The expression of leading edge genes is also shown, where key DE genes are reported with the same color of their associated DNA repair pathways (BER=base excision repair, NER=nucleotide-excision repair, MMR=mismatch repair, HR=homologous recombination, NHEJ=non-homologous end joining, FA=Fanconi anemia pathway).[47–54,82,95,96,114–133] **D** GSEA for terms related to *protein stability and ER load* performed on DE genes detected in each comparison. For each GO term, enrichments in terms of Count and p-value are reported. **E** Expression level of DE genes included in the load of the GO term *proteasomal protein catabolic process* (HCT-15 cells, $p < 0.05$) in BxPC-3 and HCT-15 cells. **F** NES (normalized enrichment score) and - log_{10}pval for the log_2FC rank-based GSEA enrichment of the GO term *PERK-mediated UPR* in treated vs control comparisons. **G** GSEA enrichment plot for the *PERK-mediated UPR* pathway obtained from log_2FC rank in the M 6 h vs control comparison in BxPC-3 cells. The expression of leading edge genes is also shown, where key DE genes of the mentioned pathway are reported. **H** Boxplots showing the expression level of ATF4, DDIT3 and PPP1R15A (vst-transformed normalized counts) in BxPC-3 cells.

Figure 4

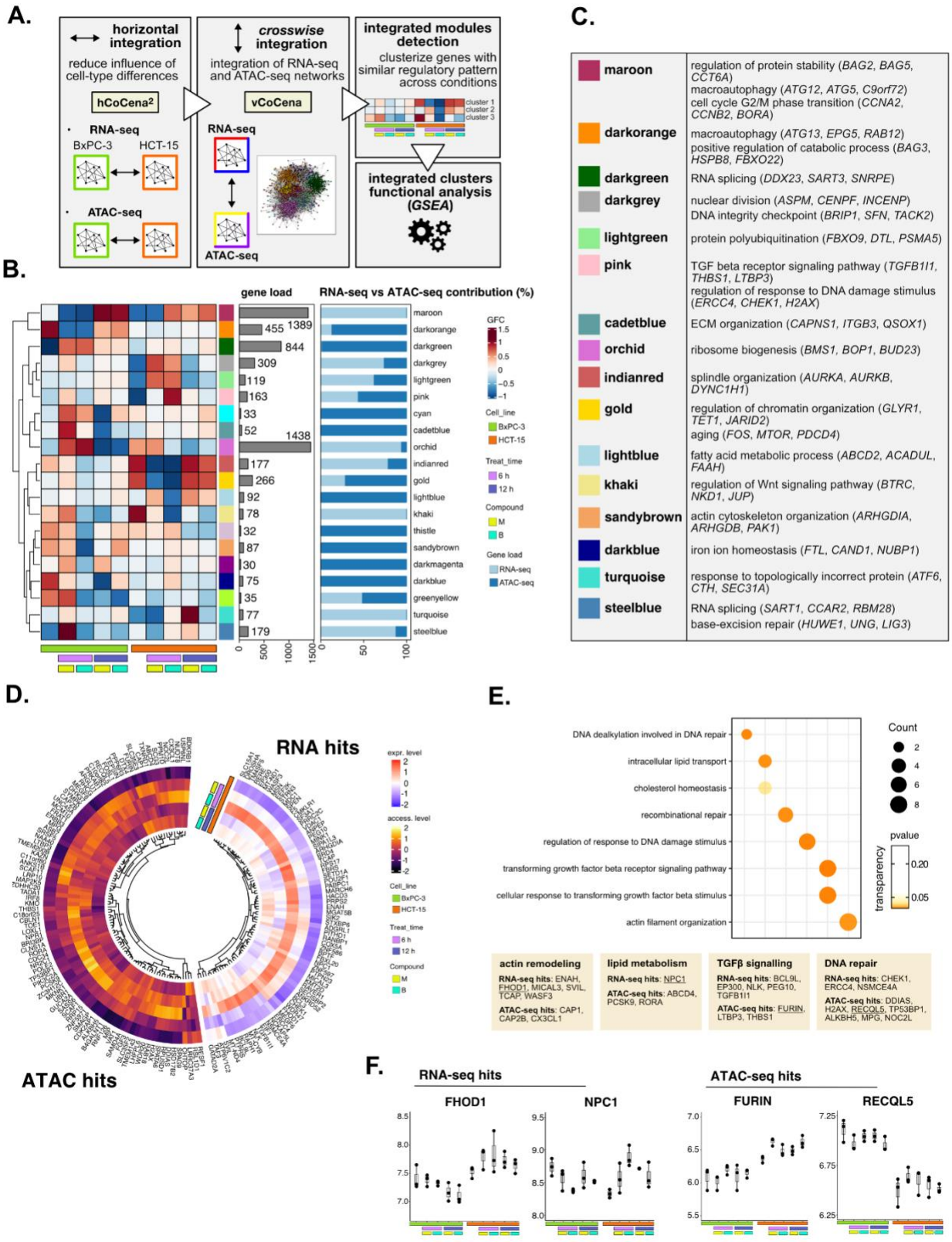


966

Figure 4. The response to 3-CePs is further regulated at the chromatin level

A Overview of the applied workflow for the ATAC-seq analysis. **B** Number of up- (red) and downregulated (blue) DARs in BxPC-3 and HCT-15 cells after treatment with M (10 nM), B (200 nM) or DMSO 0.5% (ctrl) for 6 h and 12 h (p-value threshold = 0.05, shrinkage = TRUE). Light blue/red = all detected DARs, dark blue/red = protein coding DARs mapping in promoter regions. **C** Accessibility level of cell-specific 6 h DARs across test conditions. GSEA was performed on genes associated with DARs with similar regulation, grouped in modules identified by hierarchical clustering: for each cluster, representative GO terms and genes of the associated load are reported. **D** Pairwise integration: ratio-ratio plots report the RNA-seq and ATAC-seq log_2FCs of genes showing the same direction of transcriptional and chromatin accessibility regulation. Integration was performed not only at the same time point in both omic layers, but also between chromatin changes at 6 h and transcriptional responses at 12 h.

Figure 5

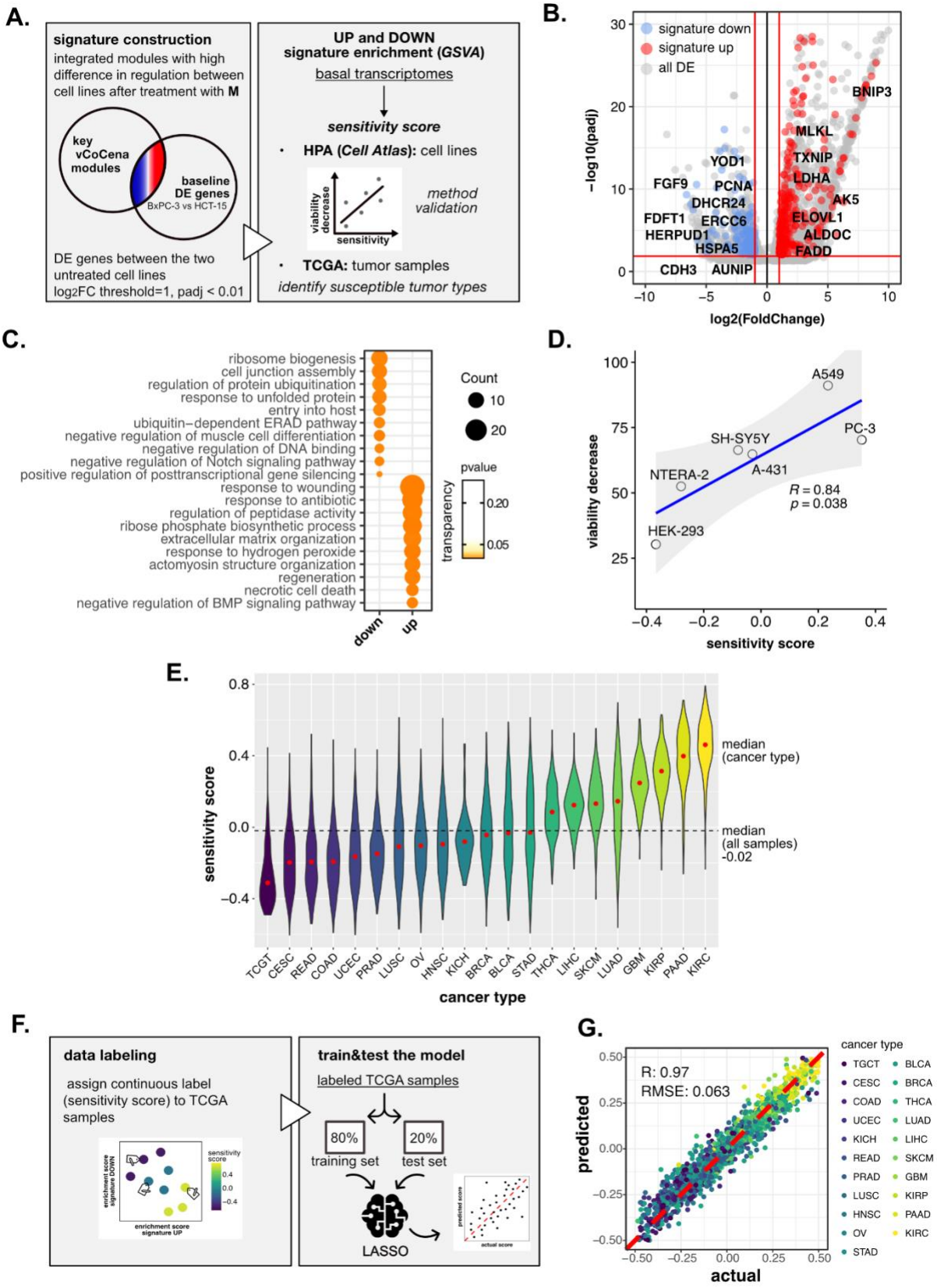


981

Figure 5. Crosswise integration expedites the comprehension of multi-omic data

A Overview of the applied workflow for the *crosswise* integration analysis. **B** Integrated modules of genes from the RNA-seq and ATAC-seq layers obtained with vCoCena and associated GFC (group fold change) pattern of regulation across conditions. The relative contribution of hits from the RNA-seq or ATAC-seq layers is also reported for each module. **C** Representative GO terms ($p < 0.05$) for the most relevant modules of genes, identified by GSEA. Enrichments in terms of Count and p-value are reported. **D** Expression and chromatin accessibility levels in HCT-15 cells of genes included in the *pink* module (nodes can come from the RNA-seq or ATAC-seq layer). **E** Most representative GO terms from GSEA on genes of the *pink* module (key areas: actin remodeling, lipid metabolism, TGF β signaling, DNA repair). For each GO term ($p < 0.05$), enrichments in terms of Count and p-value are reported. **F** Boxplots showing the expression level of FHOD1, NPC1, FURIN and RECQL5 (vst-transformed normalized counts) in BxPC-3 and HCT-15 cells.

Figure 6



996

Figure 6. Perturbation-informed basal signatures efficiently predict sensitivity to our candidate drugs

A Overview of the applied workflow for the sensitivity signature construction and associated drug susceptibility prediction. **B** M sensitivity signature genes (red = signature up, blue = signature down) pinpointed from all DE genes in the BxPC-3 vs HCT-15 baseline comparison. **C** Representative GO terms ($p < 0.05$) for genes of the M sensitivity signature (up and down), identified by GSEA. Enrichments in terms of Count and p-value are reported. **D** Pearson correlation between predicted sensitivity score and viability decrease in a subset of HPA (*Cell Atlas*) cell lines (validation set). **E** Sensitivity scores predicted from GSVA enrichment of our up and down signatures in RNA-seq profiles of TCGA tumor samples. Median values for all sample scores and within each tumor type are reported. **F** Overview of the applied workflow for the LASSO-based ML setup. **G** Predictive outcome of the trained model (Pearson correlation R and RMSE are reported).

Bibliography

1. Paananen, J. & Fortino, V. An omics perspective on drug target discovery platforms. *Brief. Bioinformatics* **21**, 1937–1953 (2020).
2. Dugger, S. A., Platt, A. & Goldstein, D. B. Drug development in the era of precision medicine. *Nat. Rev. Drug Discov.* **17**, 183–196 (2018).
3. Matthews, H., Hanison, J. & Nirmalan, N. "Omics"-Informed Drug and Biomarker Discovery: Opportunities, Challenges and Future Perspectives. *Proteomes* **4**, (2016).
4. Mun, J., Choi, G. & Lim, B. A guide for bioinformaticians: 'omics-based drug discovery for precision oncology. *Drug Discov. Today* (2020) doi:10.1016/j.drudis.2020.08.004.
5. Tsimberidou, A. M., Fountzilas, E., Nikanjam, M. & Kurzrock, R. Review of precision cancer medicine: Evolution of the treatment paradigm. *Cancer Treat. Rev.* **86**, 102019 (2020).
6. Koromina, M., Pandi, M.-T. & Patrinos, G. P. Rethinking Drug Repositioning and Development with Artificial Intelligence, Machine Learning, and Omics. *OMICS* **23**, 539–548 (2019).
7. Li, Y., Ma, L., Wu, D. & Chen, G. Advances in bulk and single-cell multi-omics approaches for systems biology and precision medicine. *Brief. Bioinformatics* (2021) doi:10.1093/bib/bbab024.
8. Shaheen, S., Fawaz, F., Shah, S. & Büsselberg, D. Differential Expression and Pathway Analysis in Drug-Resistant Triple-Negative Breast Cancer Cell Lines Using RNASeq Analysis. *Int. J. Mol. Sci.* **19**, (2018).
9. Li, X.-X. *et al.* RNA-seq reveals determinants for irinotecan sensitivity/resistance in colorectal cancer cell lines. *Int. J. Clin. Exp. Pathol.* **7**, 2729–2736 (2014).
10. Ye, C. *et al.* DRUG-seq for miniaturized high-throughput transcriptome profiling in drug discovery. *Nat. Commun.* **9**, 4307 (2018).
11. Srivatsan, S. R. *et al.* Massively multiplex chemical transcriptomics at single-cell resolution. *Science* **367**, 45–51 (2020).
12. McFarland, J. M. *et al.* Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. *Nat. Commun.* **11**, 4296 (2020).
13. Rendeiro, A. F. *et al.* Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nat. Commun.* **11**, 577 (2020).
14. Kagohara, L. T. *et al.* Integrated single-cell and bulk gene expression and ATAC-seq reveals heterogeneity and early changes in pathways associated with resistance to cetuximab in HNSCC-sensitive cell lines. *Br. J. Cancer* **123**, 101–113 (2020).
15. Zuravka, I. *et al.* Synthesis and DNA cleavage activity of Bis-3-chloropiperidines as alkylating agents. *ChemMedChem* **9**, 2178–2185 (2014).
16. Zuravka, I., Roesmann, R., Sosic, A., Göttlich, R. & Gatto, B. Bis-3-

chloropiperidines containing bridging lysine linkers: Influence of side chain structure on DNA alkylating activity. *Bioorg. Med. Chem.* **23**, 1241–1250 (2015).

17. Zuravka, I., Sosic, A., Gatto, B. & Göttlich, R. Synthesis and evaluation of a bis-3-chloropiperidine derivative incorporating an anthraquinone pharmacophore. *Bioorg. Med. Chem. Lett.* **25**, 4606–4609 (2015).
18. Sosic, A. *et al.* Direct and Topoisomerase II Mediated DNA Damage by Bis-3-chloropiperidines: The Importance of Being an Earnest G. *ChemMedChem* **12**, 1471–1479 (2017).
19. Carraro, C. *et al.* Behind the mirror: chirality tunes the reactivity and cytotoxicity of chloropiperidines as potential anticancer agents. *ACS Med. Chem. Lett.* **10**, 552–557 (2019).
20. Helbing, T. *et al.* Aromatic Linkers Unleash the Antiproliferative Potential of 3-Chloropiperidines Against Pancreatic Cancer Cells. *ChemMedChem* **15**, 2040–2051 (2020).
21. Carraro, C. *et al.* Appended Aromatic Moieties in Flexible Bis-3-chloropiperidines Confer Tropism against Pancreatic Cancer Cells. *ChemMedChem* **16**, 860–868 (2021).
22. Faivre, E. J. *et al.* Selective inhibition of the BD2 bromodomain of BET proteins in prostate cancer. *Nature* **578**, 306–310 (2020).
23. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
24. Subramanian, A. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437-1452.e17 (2017).
25. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357**, (2017).
26. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
27. Bush, E. C. *et al.* PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nat. Commun.* **8**, 105 (2017).
28. Corsello, S. M. *et al.* Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. *Nat. Cancer* **1**, 235–248 (2020).
29. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
30. Schmidl, C. *et al.* Combined chemosensitivity and chromatin profiling prioritizes drug combinations in CLL. *Nat. Chem. Biol.* **15**, 232–240 (2019).
31. Granja, J. M. *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
32. Tung, K.-L. *et al.* Integrated chromatin and transcriptomic profiling of patient-derived colon cancer organoids identifies personalized drug targets to overcome oxaliplatin resistance. *Genes Dis.* **8**, 203–214 (2021).
33. Suzuki, A. *et al.* Characterization of cancer omics and drug perturbations in panels of lung cancer cells. *Sci. Rep.* **9**, 19529 (2019).

34. Schröder, H. C., Batel, R., Schwertner, H., Boreiko, O. & Müller, W. E. G. Fast micromethod DNA single-strand-break assay. *Methods Mol. Biol.* **314**, 287–305 (2006).
35. Meyn, R. E. & Murray, D. Cell cycle effects of alkylating agents. *Pharmacol. Ther.* **24**, 147–163 (1984).
36. Kaufmann, W. K. & Paules, R. S. DNA damage and cell cycle checkpoints. *FASEB J.* **10**, 238–247 (1996).
37. Tu, L. C., Melendy, T. & Beerman, T. A. DNA damage responses triggered by a highly cytotoxic monofunctional DNA alkylator, hedamycin, a pluramycin antitumor antibiotic. *Mol. Cancer Ther.* **3**, 577–585 (2004).
38. Li, L.-Y., Guan, Y., Chen, X.-S., Yang, J.-M. & Cheng, Y. DNA repair pathways in cancer therapy and resistance. *Front. Pharmacol.* **11**, 629266 (2020).
39. Sharma, A., Singh, K. & Almasan, A. Histone H2AX phosphorylation: a marker for DNA damage. *Methods Mol. Biol.* **920**, 613–626 (2012).
40. Vasan, N., Baselga, J. & Hyman, D. M. A view on drug resistance in cancer. *Nature* **575**, 299–309 (2019).
41. Shenton, D. *et al.* Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *J. Biol. Chem.* **281**, 29011–29021 (2006).
42. Burger, K. *et al.* Chemotherapeutic drugs inhibit ribosome biogenesis at various levels. *J. Biol. Chem.* **285**, 12416–12425 (2010).
43. Lin, X., Xiao, Z., Chen, T., Liang, S. H. & Guo, H. Glucose metabolism on tumor plasticity, diagnosis, and treatment. *Front. Oncol.* **10**, 317 (2020).
44. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550 (2005).
45. Chowdhury, D. *et al.* A PP4-phosphatase complex dephosphorylates gamma-H2AX generated during DNA replication. *Mol. Cell* **31**, 33–46 (2008).
46. Modesti, M. *et al.* RAD51AP1 is a structure-specific DNA binding protein that stimulates joint molecule formation during RAD51-mediated homologous recombination. *Mol. Cell* **28**, 468–481 (2007).
47. Satoh, M. S. & Hanawalt, P. C. TFIIH-mediated nucleotide excision repair and initiation of mRNA transcription in an optimized cell-free DNA repair and RNA transcription assay. *Nucleic Acids Res.* **24**, 3576–3582 (1996).
48. Li, J.-M. & Jin, J. CRL ubiquitin ligases and DNA damage response. *Front. Oncol.* **2**, 29 (2012).
49. Piwko, W. *et al.* The MMS22L-TONSL heterodimer directly promotes RAD51-dependent recombination upon replication stress. *EMBO J.* **35**, 2584–2601 (2016).
50. Westermark, U. K. *et al.* BARD1 participates with BRCA1 in homology-directed repair of chromosome breaks. *Mol. Cell. Biol.* **23**, 7926–7936 (2003).
51. Cantor, S. B. *et al.* BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105**, 149–160 (2001).
52. Niraj, J., Färkkilä, A. & D'Andrea, A. D. The fanconi anemia pathway in cancer. *Annu. Rev. Cancer Biol.* **3**, 457–478 (2019).

53. Fortini, P. *et al.* The base excision repair: mechanisms and its relevance for cancer susceptibility. *Biochimie* **85**, 1053–1071 (2003).
54. Kunkel, T. A. & Erie, D. A. DNA mismatch repair. *Annu. Rev. Biochem.* **74**, 681–710 (2005).
55. Hutt, D. & Balch, W. E. Cell Biology. The proteome in balance. *Science* **329**, 766–767 (2010).
56. González-Quiroz, M. *et al.* When endoplasmic reticulum proteostasis meets the DNA damage response. *Trends Cell Biol.* **30**, 881–891 (2020).
57. Ran, R., Lu, A., Xu, H., Tang, Y. & Sharp, F. R. Heat-shock protein regulation of protein folding, protein degradation, protein function, and apoptosis. in *Handbook of neurochemistry and molecular neurobiology* (eds. Lajtha, A. & Chan, P. H.) 89–107 (Springer US, 2007). doi:10.1007/978-0-387-30383-3_6.
58. Eskelinen, E.-L. Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy. *Mol. Aspects Med.* **27**, 495–502 (2006).
59. Scott, D. C. *et al.* Two distinct types of E3 ligases work in unison to regulate substrate ubiquitylation. *Cell* **166**, 1198-1214.e24 (2016).
60. Jozsef, L. *et al.* Reticulon 4 is necessary for endoplasmic reticulum tubulation, STIM1-Orai1 coupling, and store-operated calcium entry. *J. Biol. Chem.* **289**, 9380–9395 (2014).
61. Han, J. *et al.* ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. *Nat. Cell Biol.* **15**, 481–490 (2013).
62. Urra, H., Dufey, E., Lisbona, F., Rojas-Rivera, D. & Hetz, C. When ER stress reaches a dead end. *Biochim. Biophys. Acta* **1833**, 3507–3517 (2013).
63. Garcia-Carbonero, N., Li, W., Cabeza-Morales, M., Martinez-Useros, J. & Garcia-Foncillas, J. New hope for pancreatic ductal adenocarcinoma treatment targeting endoplasmic reticulum stress response: A systematic review. *Int. J. Mol. Sci.* **19**, (2018).
64. Tadros, S. *et al.* De Novo Lipid Synthesis Facilitates Gemcitabine Resistance through Endoplasmic Reticulum Stress in Pancreatic Cancer. *Cancer Res.* **77**, 5503–5517 (2017).
65. Romero, F. *et al.* Lipid synthesis is required to resolve endoplasmic reticulum stress and limit fibrotic responses in the lung. *Am. J. Respir. Cell Mol. Biol.* **59**, 225–236 (2018).
66. Zhou, W. *et al.* Clinical significance and biological function of fucosyltransferase 2 in lung adenocarcinoma. *Oncotarget* **8**, 97246–97259 (2017).
67. Höllbacher, B., Balázs, K., Heinig, M. & Uhlenhaut, N. H. Seq-ing answers: Current data integration approaches to uncover mechanisms of transcriptional regulation. *Comput. Struct. Biotechnol. J.* **18**, 1330–1341 (2020).
68. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* **21**, 22 (2020).
69. Ackermann, A. M., Wang, Z., Schug, J., Naji, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016).
70. Wang, W., Guo, X. & Dan, H. α2A-Adrenergic Receptor Inhibits the Progression

of Cervical Cancer Through Blocking PI3K/AKT/mTOR Pathway. *Onco Targets Ther* **13**, 10535–10546 (2020).

71. Pamidimukkala, N. *et al.* Nme1 and Nme2 genes exert metastasis-suppressor activities in a genetically engineered mouse model of UV-induced melanoma. *Br. J. Cancer* **124**, 161–165 (2021).
72. Reeves, H. L. *et al.* Kruppel-like factor 6 (KLF6) is a tumor-suppressor gene frequently inactivated in colorectal cancer. *Gastroenterology* **126**, 1090–1103 (2004).
73. Furth, N. & Aylon, Y. The LATS1 and LATS2 tumor suppressors: beyond the Hippo pathway. *Cell Death Differ.* **24**, 1488–1501 (2017).
74. Mayeda, A. *et al.* Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *EMBO J.* **18**, 4560–4570 (1999).
75. Schäffler, K. *et al.* A stimulatory role for the La-related protein 4B in translation. *RNA* **16**, 1488–1499 (2010).
76. Hebbar, N., Wang, C. & Rangnekar, V. M. Mechanisms of apoptosis by the tumor suppressor Par-4. *J. Cell. Physiol.* **227**, 3715–3721 (2012).
77. Kawase, T. *et al.* p53 target gene AEN is a nuclear exonuclease required for p53-dependent apoptosis. *Oncogene* **27**, 3797–3810 (2008).
78. Lim, Y., Dorstyn, L. & Kumar, S. The p53-caspase-2 axis in the cell cycle and DNA damage response. *Exp. Mol. Med.* **53**, 517–527 (2021).
79. Nishiyama, A., Masutani, H., Nakamura, H., Nishinaka, Y. & Yodoi, J. Redox regulation by thioredoxin and thioredoxin-binding proteins. *IUBMB Life* **52**, 29–33 (2001).
80. Nthiga, T. M. *et al.* CALCOCO1 acts with VAMP-associated proteins to mediate ER-phagy. *EMBO J.* **39**, e103649 (2020).
81. Hawkes, W. C. & Alkan, Z. Regulation of redox signaling by selenoproteins. *Biol. Trace Elem. Res.* **134**, 235–251 (2010).
82. Yard, B. D., Reilly, N. M., Bedenbaugh, M. K. & Pittman, D. L. RNF138 interacts with RAD51D and is required for DNA interstrand crosslink repair and maintaining chromosome integrity. *DNA Repair (Amst)* **42**, 82–93 (2016).
83. Eischen, C. M. Role of Mdm2 and Mdmx in DNA repair. *J. Mol. Cell Biol.* **9**, 69–73 (2017).
84. Zhu, Q. S. *et al.* Vimentin is a novel AKT1 target mediating motility and invasion. *Oncogene* **30**, 457–470 (2011).
85. Winter, S. F., Lukes, L., Walker, R. C., Welch, D. R. & Hunter, K. W. Allelic variation and differential expression of the mSIN3A histone deacetylase complex gene Arid4b promote mammary tumor growth and metastasis. *PLoS Genet.* **8**, e1002735 (2012).
86. Ogata, F. T., Branco, V., Vale, F. F. & Coppo, L. Glutaredoxin: Discovery, redox defense and much more. *Redox Biol.* **43**, 101975 (2021).
87. Janiszewska, M., Primi, M. C. & Izard, T. Cell adhesion in cancer: Beyond the migration of single cells. *J. Biol. Chem.* **295**, 2495–2505 (2020).
88. Aschenbrenner, A. C. *et al.* Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med.* **13**, 7 (2021).

89. Chang, H. & Zou, Z. Targeting autophagy to overcome drug resistance: further developments. *J. Hematol. Oncol.* **13**, 159 (2020).
90. Tafazzoli-Shadpour, M., Mohammadi, E. & Torkashvand, E. Mechanics of actin filaments in cancer onset and progress. *Int. Rev. Cell Mol. Biol.* **355**, 205–243 (2020).
91. Caridi, C. P., Plessner, M., Grosse, R. & Chiolo, I. Nuclear actin filaments in DNA repair dynamics. *Nat. Cell Biol.* **21**, 1068–1077 (2019).
92. Gasteier, J. E. *et al.* Activation of the Rac-binding partner FHOD1 induces actin stress fibers via a ROCK-dependent mechanism. *J. Biol. Chem.* **278**, 38902–38912 (2003).
93. Brunen, D. *et al.* TGF-β: an emerging player in drug resistance. *Cell Cycle* **12**, 2960–2968 (2013).
94. Dubois, C. M. *et al.* Evidence that furin is an authentic transforming growth factor-beta1-converting enzyme. *Am. J. Pathol.* **158**, 305–316 (2001).
95. Lou, J. *et al.* AUNIP/C1orf135 directs DNA double-strand breaks towards the homologous recombination repair pathway. *Nat. Commun.* **8**, 985 (2017).
96. Pascucci, B. *et al.* CSA and CSB play a role in the response to DNA breaks. *Oncotarget* **9**, 11581–11591 (2018).
97. Schulze, A. *et al.* The ubiquitin-domain protein HERP forms a complex with components of the endoplasmic reticulum associated degradation pathway. *J. Mol. Biol.* **354**, 1021–1027 (2005).
98. Ernst, R., Mueller, B., Ploegh, H. L. & Schlieker, C. The otubain YOD1 is a deubiquitinating enzyme that associates with p97 to facilitate protein dislocation from the ER. *Mol. Cell* **36**, 28–38 (2009).
99. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
100. Perez-Riverol, Y. *et al.* Quantifying the impact of public omics data. *Nat. Commun.* **10**, 3512 (2019).
101. Adam, G. *et al.* Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis. Oncol.* **4**, 19 (2020).
102. Warnat-Herresthal, S. *et al.* Swarm Learning for decentralized and confidential clinical machine learning. *Nature* **594**, 265–270 (2021).
103. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
104. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
105. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
106. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
107. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
108. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*

Methods **9**, 357–359 (2012).
109. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187-91 (2014).
110. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
111. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
112. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
113. Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *Eur. Phys. J. Spec. Top.* **178**, 13–23 (2009).
114. Charles Richard, J. L. *et al.* FACT assists base excision repair by boosting the remodeling activity of RSC. *PLoS Genet.* **12**, e1006221 (2016).
115. Poletto, M., Lirussi, L., Wilson, D. M. & Tell, G. Nucleophosmin modulates stability, activity, and nucleolar accumulation of base excision repair proteins. *Mol. Biol. Cell* **25**, 1641–1652 (2014).
116. Parsons, J. L. *et al.* Ubiquitin ligase ARF-BP1/Mule modulates base excision repair. *EMBO J.* **28**, 3207–3215 (2009).
117. Prasad, R., Dianov, G. L., Bohr, V. A. & Wilson, S. H. FEN1 stimulation of DNA polymerase beta mediates an excision step in mammalian long patch base excision repair. *J. Biol. Chem.* **275**, 4460–4466 (2000).
118. Smirnova, E., Toueille, M., Markkanen, E. & Hübscher, U. The human checkpoint sensor and alternative DNA clamp Rad9-Rad1-Hus1 modulates the activity of DNA ligase I, a component of the long-patch base excision repair machinery. *Biochem. J.* **389**, 13–17 (2005).
119. Poulsen, S. L. *et al.* RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response. *J. Cell Biol.* **201**, 797–807 (2013).
120. Overmeer, R. M. *et al.* Replication factor C recruits DNA polymerase delta to sites of nucleotide excision repair but is not required for PCNA recruitment. *Mol. Cell. Biol.* **30**, 4828–4839 (2010).
121. Fousteri, M. & Mullenders, L. H. F. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* **18**, 73–84 (2008).
122. Mazin, A. V., Mazina, O. M., Bugreev, D. V. & Rossi, M. J. Rad54, the motor of homologous recombination. *DNA Repair (Amst)* **9**, 286–302 (2010).
123. Johnson, R. D., Liu, N. & Jasin, M. Mammalian XRCC2 promotes the repair of DNA double-strand breaks by homologous recombination. *Nature* **401**, 397–399 (1999).
124. Lu, H., Yue, J., Meng, X., Nickoloff, J. A. & Shen, Z. BCCIP regulates homologous recombination by distinct domains and suppresses spontaneous DNA damage. *Nucleic Acids Res.* **35**, 7160–7170 (2007).
125. Li, Y. *et al.* HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J. Biol. Chem.* **284**, 23525–23531 (2009).

126. Xu, D. *et al.* RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability. *Genes Dev.* **22**, 2843–2855 (2008).
127. McVey, M., Khodaverdian, V. Y., Meyer, D., Cerqueira, P. G. & Heyer, W.-D. Eukaryotic DNA polymerases in homologous recombination. *Annu. Rev. Genet.* **50**, 393–421 (2016).
128. Liu, S. & Kong, D. End resection: a key step in homologous recombination and DNA double-strand break repair. *GENOME INSTAB. DIS.* **2**, 39–50 (2021).
129. Jaafar, L., Li, Z., Li, S. & Dynan, W. S. SFPQ•NONO and XLF function separately and together to promote DNA double-strand break repair via canonical nonhomologous end joining. *Nucleic Acids Res.* **45**, 1848–1859 (2017).
130. Tellier, M. & Chalmers, R. The roles of the human SETMAR (Metnase) protein in illegitimate DNA recombination and non-homologous end joining repair. *DNA Repair (Amst)* **80**, 26–35 (2019).
131. Ferretti, L. P. *et al.* Cullin3-KLHL15 ubiquitin ligase mediates CtIP protein turnover to fine-tune DNA-end resection. *Nat. Commun.* **7**, 12628 (2016).
132. Rajendra, E., Garaycoechea, J. I., Patel, K. J. & Passmore, L. A. Abundance of the Fanconi anaemia core complex is regulated by the RuvBL1 and RuvBL2 AAA+ ATPases. *Nucleic Acids Res.* **42**, 13736–13748 (2014).
133. Nijman, S. M. B. *et al.* The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol. Cell* **17**, 331–339 (2005).

Supplementary Information

Decoding mechanism of action and susceptibility to anticancer candidates from integrated transcriptome and chromatin accessibility state

Caterina Carraro,[1] Lorenzo Bonaguro,[2,3] Jonas Schulte-Schrepping,[2,3] Arik Horne,[2,3] Marie Oestreich,[2] Stefanie Warnat-Herresthal,[2,3] Tim Helbing,[4] Michele De Franco,[1] Kristian Händler,[2,5,6] Sach Mukherjee,[7,8] Thomas Ulas,[2,3,5] Valentina Gandin,[1] Richard Göttlich,[4] Anna C. Aschenbrenner,[2,3,5,9] Joachim L. Schultze,[2,3,5,*] Barbara Gatto[1,*]

[1] Department of Pharmaceutical and Pharmacological Sciences, University of Padova, Padova, Italy

[2] Systems Medicine, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) e.V., Bonn, Germany

[3] Genomics and Immunoregulation, Life & Medical Sciences (LIMES) Institute, University of Bonn, Bonn, Germany

[4] Institute of Organic Chemistry, Justus Liebig University Giessen, Giessen, Germany

[5] PRECISE Platform for Genomics and Epigenomics, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) e.V. and University of Bonn, Bonn, Germany

[6] Institute of Human Genetics, University of Lübeck, Lübeck, Germany

[7] Statistics and Machine Learning, Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) e.V., Bonn, Germany

[8] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[9] Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands

* Corresponding author

Figure S1

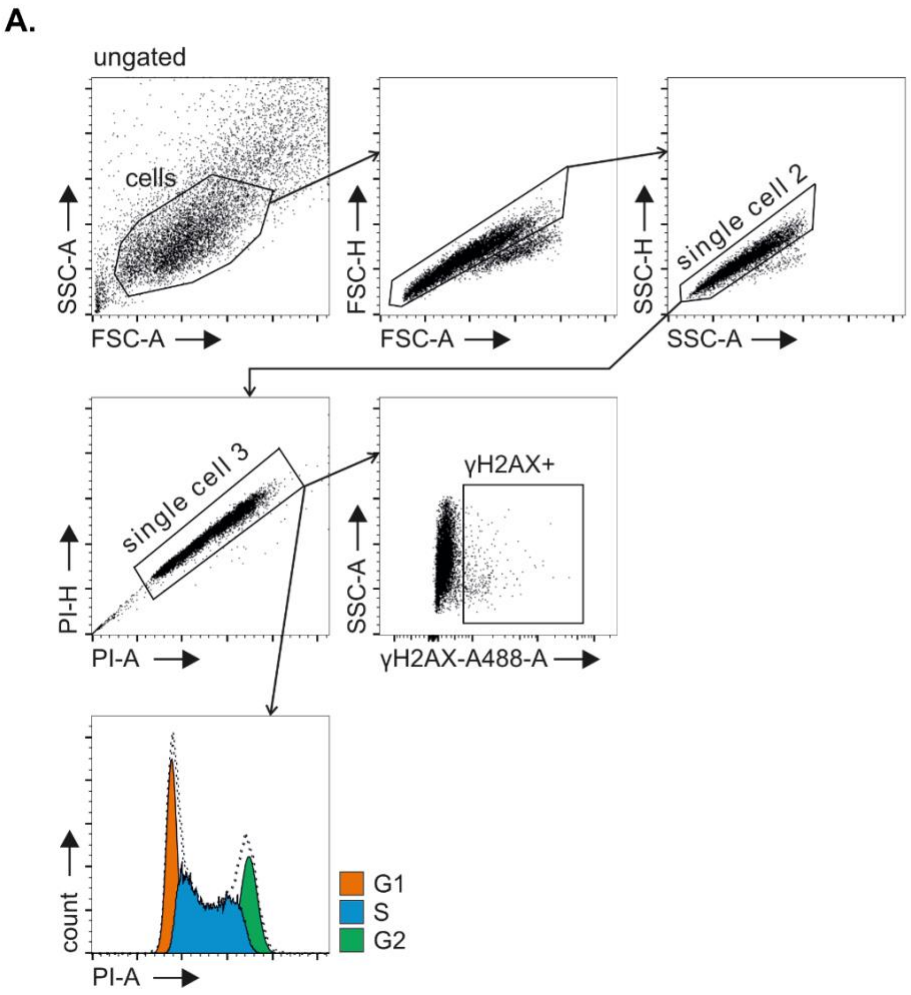


Figure S1. A Flow cytometry gating strategy for the cell cycle analysis and γH2AX induction reported in Fig.1 C and 1 D.

Figure S2

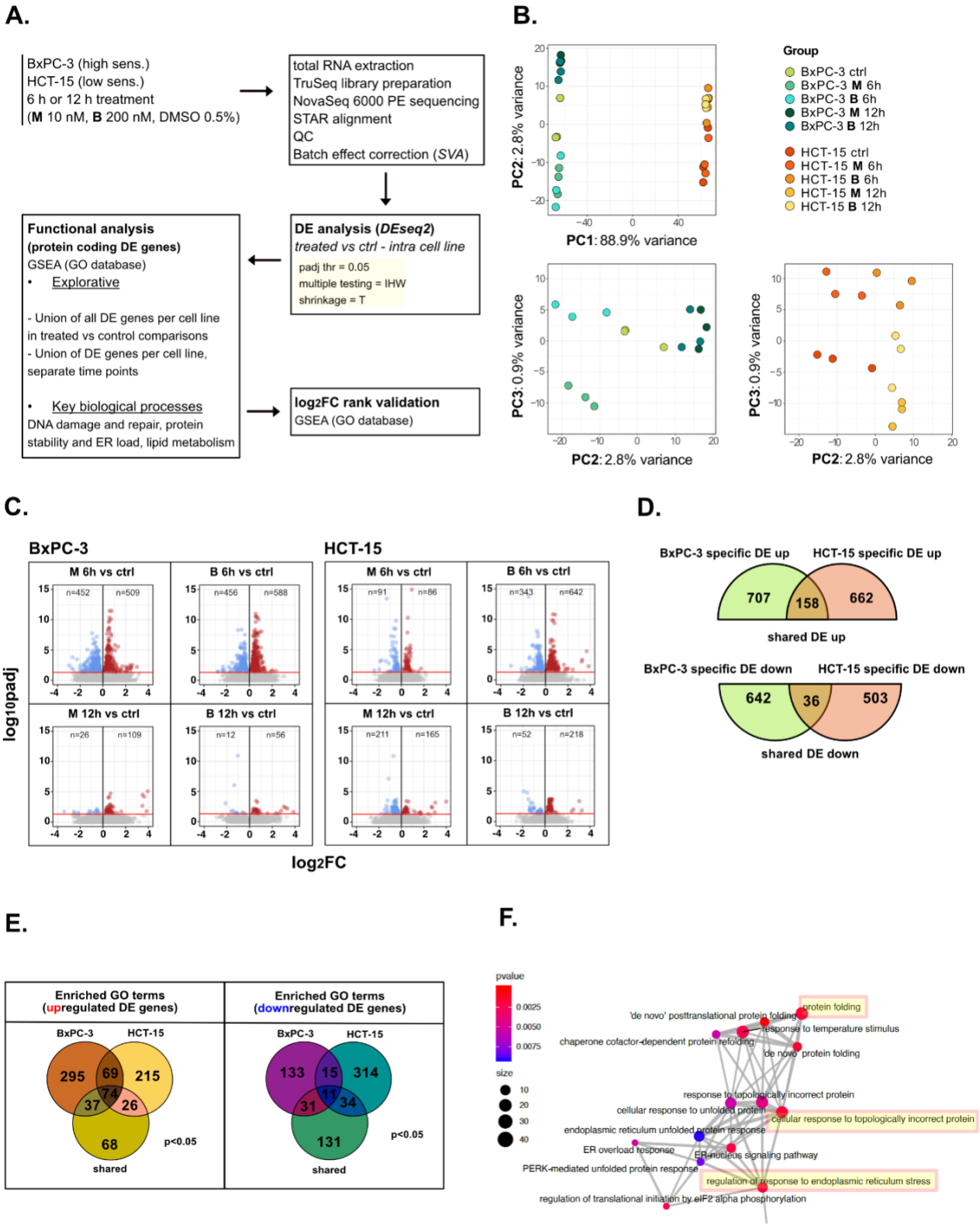


Figure S2. A Scheme of the applied workflow for the RNA-seq analyses. **B** Principal component analysis (PCA) post SVA batch correction of RNA-seq data: PC1 vs PC2 showed sample separation by cell line, PC2 vs PC3 (cell lines depicted separately) showed treatment and time point separation. **C** Volcano plots reporting up- and downregulated DE genes in all treated vs control comparisons (adjusted p-value < 0.05). **D** Venn plot reporting the number of specific and shared up- and downregulated DE genes between BxPC-3 and HCT-15 cells (union of DE genes in all treated vs control comparisons). **E** Enriched GO terms ($p < 0.05$) derived from GSEA on BxPC-3 and HCT-15 DE genes (union of DE genes in all treated vs control comparisons) and on shared DE genes (up- and downregulated separately). **F** Schematic representation of enrichment map-based selection of representative GO terms to be reported in Fig. 2 C.

Figure S3

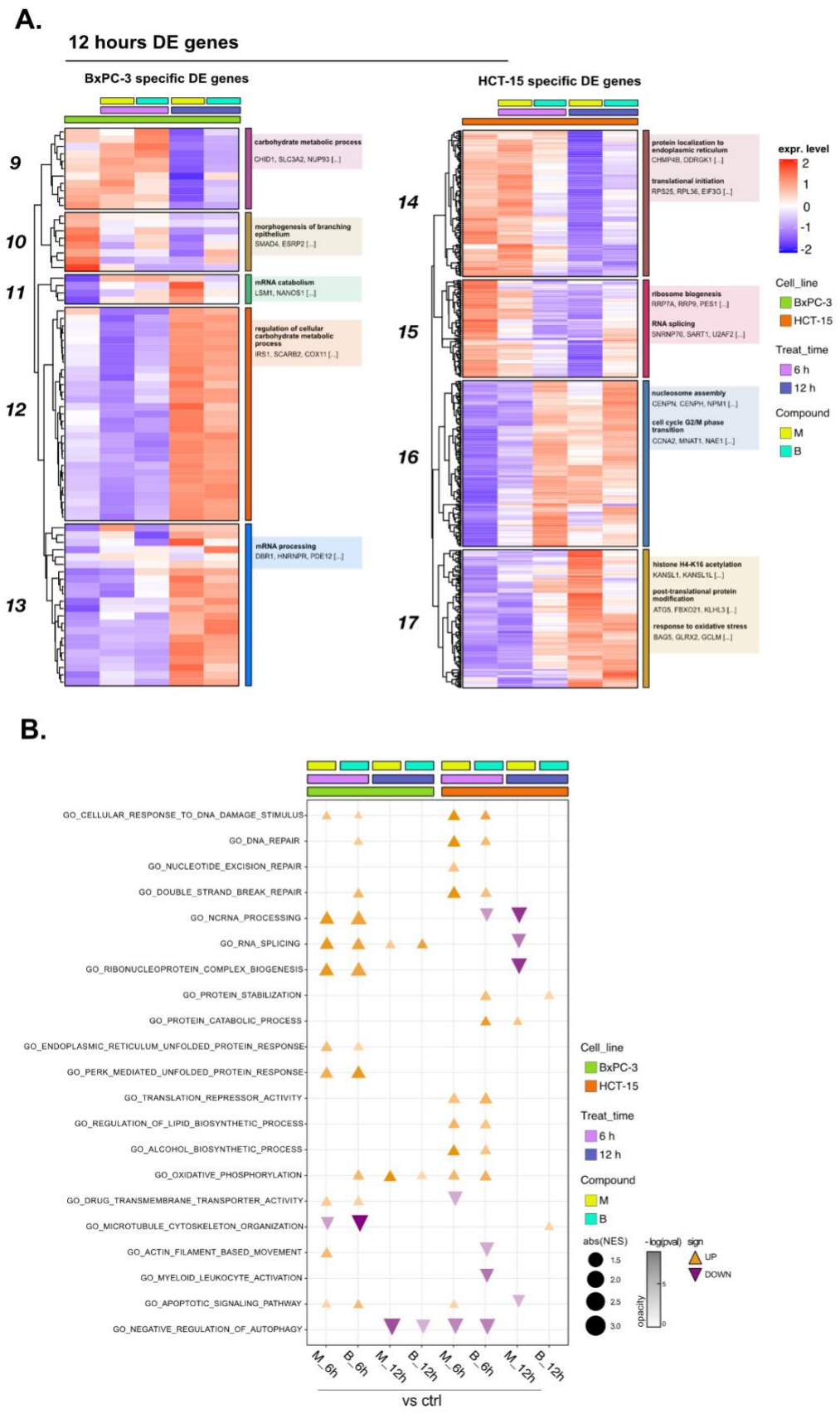
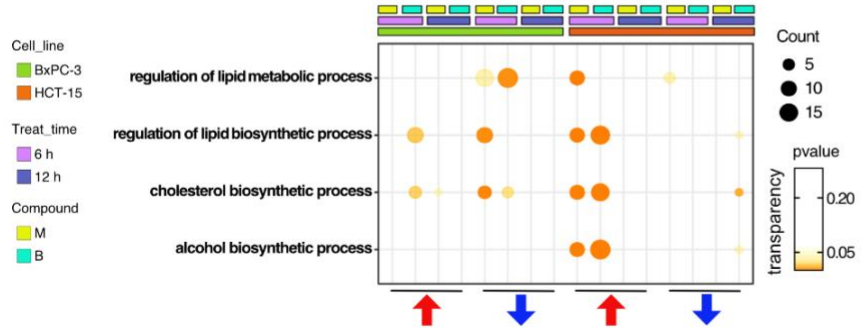


Figure S3. A Expression level of cell-specific 12 h DE genes across test conditions. GSEA was performed on modules with similar regulation identified by hierarchical clustering: for each cluster, representative GO terms and genes of the associated load are reported. **B** GO database functional enrichment (GSEA) obtained from log_2FC ranks in all treated vs control comparison both in BxPC-3 and HCT-15 cells. For each identified biological process, enrichments in terms of absolute normalized enrichment score (abs(NES)) and -log(p) of representative terms are reported ($p < 0.05$).

Figure S4

A.



B.

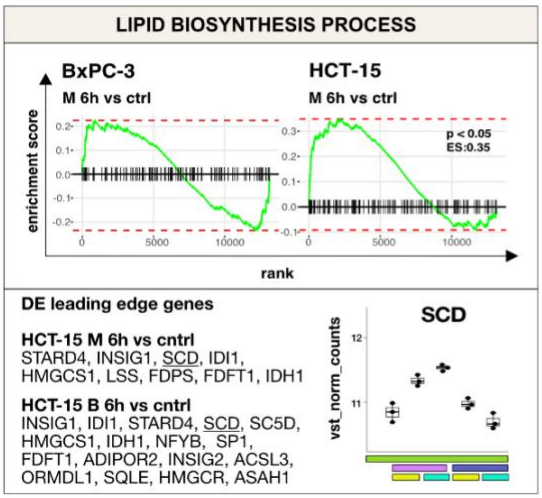


Figure S4. A GSEA for terms related to *lipid metabolism* performed on DE genes detected in each of the considered treated vs control comparisons. For each GO term ($p < 0.05$), enrichments in terms of Count and p-value are reported. **B** GSEA enrichment plots for the *lipid biosynthesis process* pathway obtained from log_2FC ranks for each of the considered treated vs control comparisons. DE leading edge genes are also reported, together with boxplots showing the expression level of SCD (vst-transformed normalized counts) in BxPC-3 cells.

Figure S5

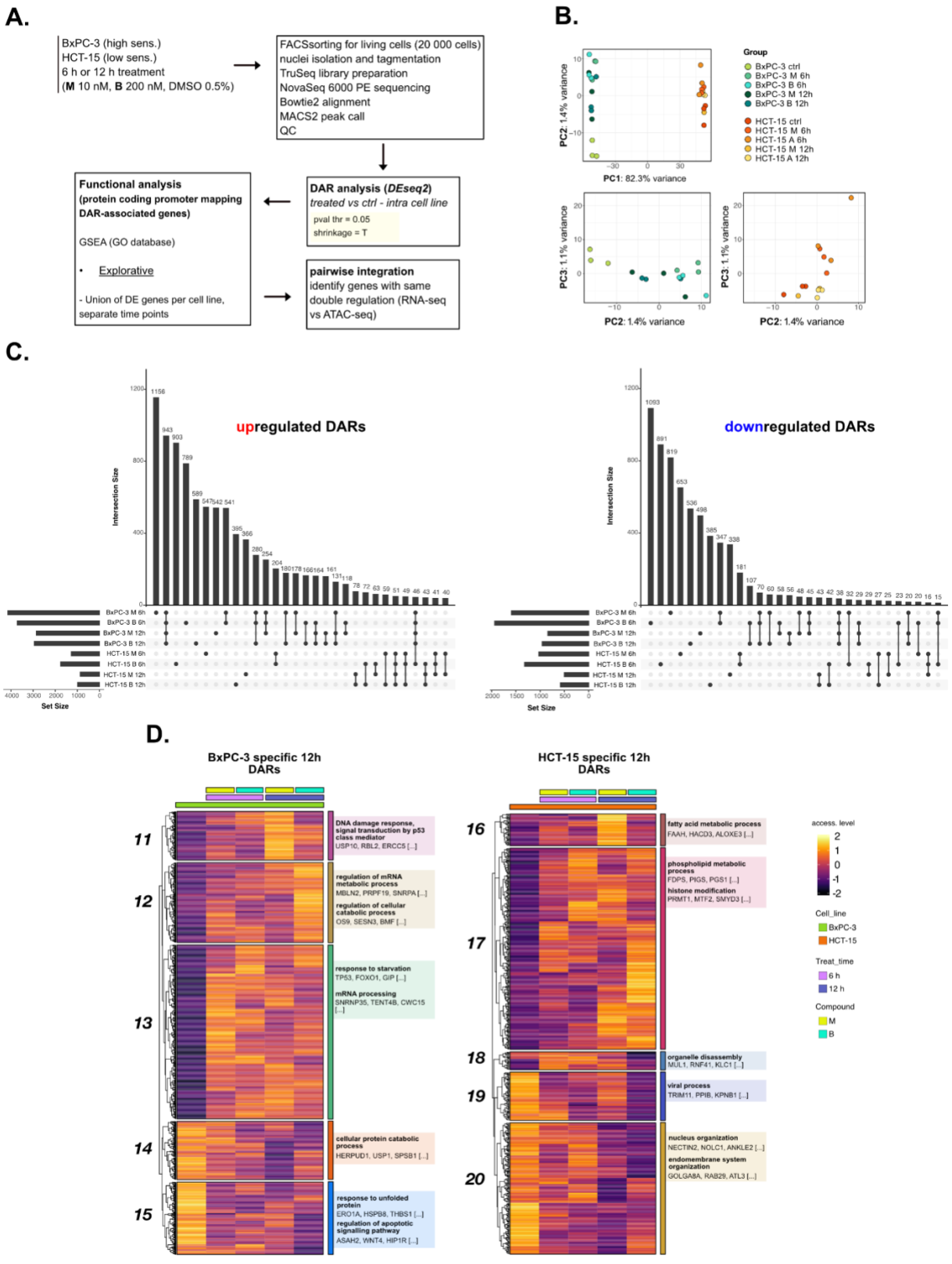


Figure S5. A Scheme of the applied workflow for the ATAC-seq analyses. **B** Principal component analysis (PCA) of ATAC-seq data: PC1 vs PC2 showed samples separation by cell line, PC2 vs PC3 (cell lines depicted separately) showed treatment and time point separation. **C** Upset plots reporting up- and downregulated DARs ($p < 0.05$) and their overlap between all treated vs control comparisons in both cell lines. **D** Accessibility level of cell-specific 12 h DARs across test conditions. GSEA was performed on genes associated with DARs with similar regulation, grouped in modules identified by hierarchical clustering: for each cluster, representative GO terms and genes of the associated load are reported.

Figure S6

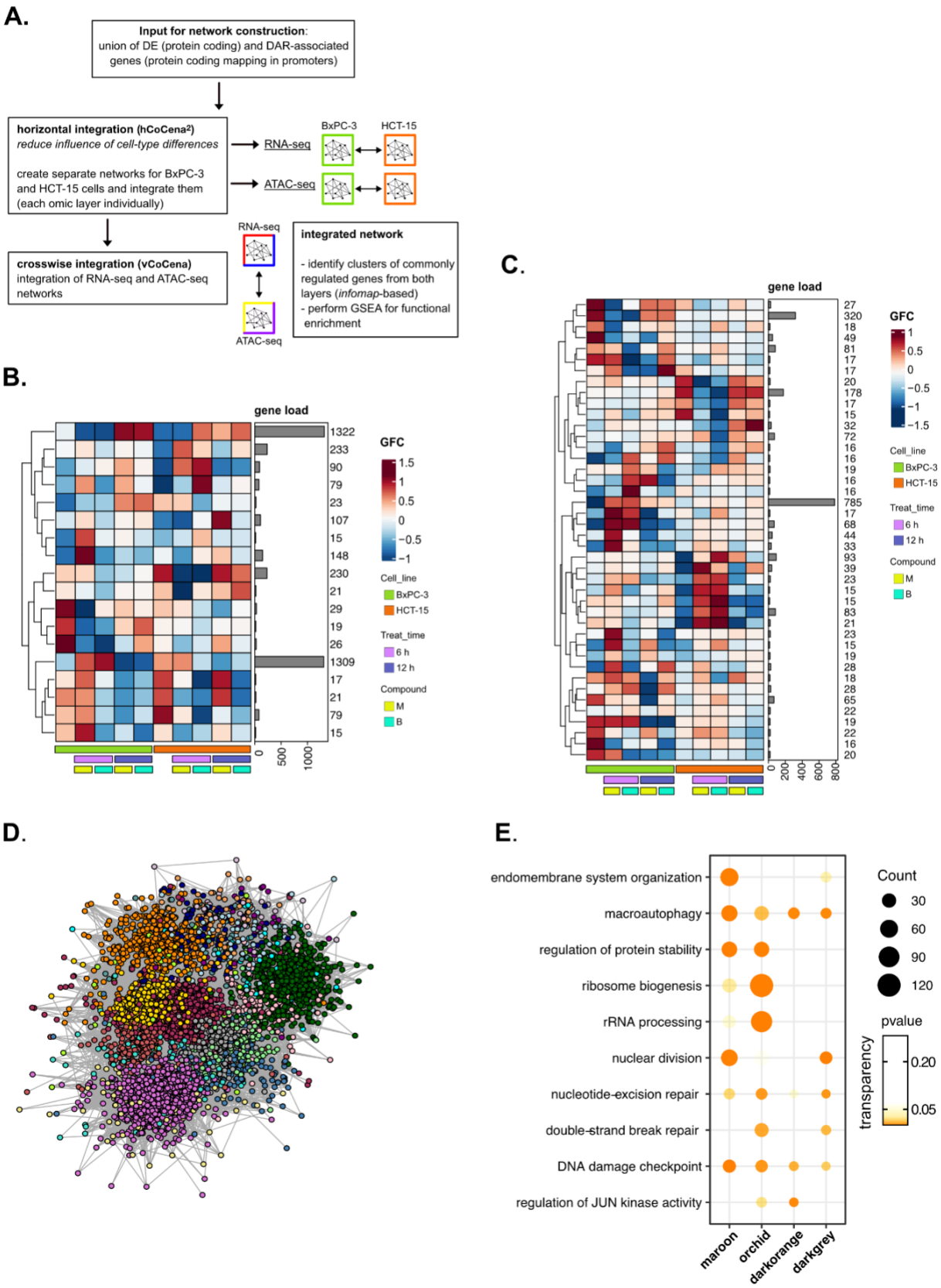


Figure S6. A Scheme of the applied workflow for the *crosswise* integration analysis. **B** Horizontally integrated modules of genes from the RNA-seq layer and associated GFC (group fold change) pattern of regulation across conditions. **C** Horizontally integrated modules of genes from the ATAC-seq layer and associated GFC (group fold change) pattern of regulation across conditions. **D** *Crosswise* integrated vCoCena network. **E** Most representative GO terms from GSEA on genes of the *maroon, orchid, darkorange, darkgrey* modules. For each GO term ($p < 0.05$), enrichments in terms of Count and p-value are reported.

Figure S7

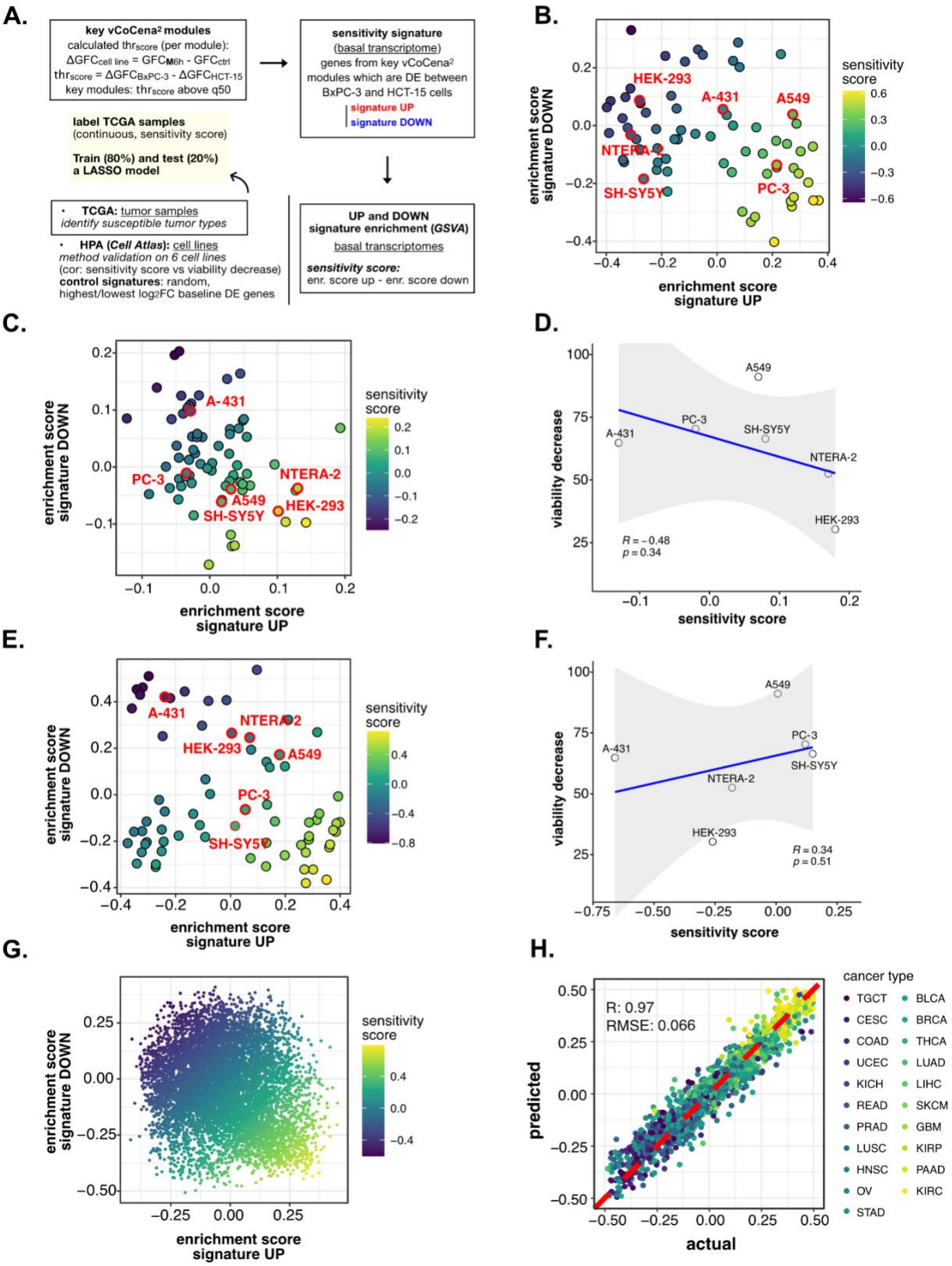


Figure S7. A Scheme of the applied workflow for the sensitivity signature construction and associated drug susceptibility prediction. **B** HPA (*Cell Atlas*) cell lines separation based on GSVA enrichment of our newly constructed up vs down signatures of sensitivity to M. Color scale reflects samples predicted sensitivity score (up signature enrichment - down signature enrichment). **C** HPA (*Cell Atlas*) cell lines separation based on GSVA enrichment of random up vs down signatures. Color scale reflects samples predicted sensitivity score (up signature enrichment - down signature enrichment). **D** Pearson correlation between predicted sensitivity score and viability decrease in a subset of HPA (*Cell Atlas*) cell lines (validation set) using a random signature. **E** HPA (*Cell Atlas*) cell lines separation based on GSVA enrichment of control up vs down signatures. GSVA was performed using a control signature composed by DE genes with top up and down log_2FC. Color scale reflects samples predicted sensitivity score (up signature enrichment - down signature enrichment). **F** Pearson correlation between predicted sensitivity score and viability decrease in a subset of HPA (*Cell Atlas*) cell lines (validation set) using a control signature composed by DE genes with top up and down log_2FC. **G** TCGA tumor samples separation based on GSVA enrichment of our newly constructed up vs down signatures of sensitivity to M. Color scale reflects samples predicted sensitivity score (up signature enrichment - down signature enrichment). **H** Predictive performance after the exclusion of genes belonging to our signature from training and test set transcriptomes (Pearson correlation R and RMSE are reported).

Table S1.

Cell line	Viability decrease	SD	signature SS	random SS	topFC SS
HEK-293	30,3	15,1	-0,37	0,18	-0,26
NTERA-2	52,5	6,14	-0,28	0,17	-0,18
SH-SY5Y	66,4	4,61	-0,08	0,08	0,15
A-431	64,8	7,54	-0,03	-0,13	-0,66
A549	91,1	2,85	0,23	0,07	0,007
PC-3	70,3	0,14	0,35	-0,02	0,12

SD: standard deviation
signature SS: perturbation-informed signature sensitivity score
random SS: random signature sensitivity score
topFC SS: top up and down log_2FC control signature sensitivity score

Table S1. Average viability decrease in cell lines treated with **M** 10 nM for 72 h with associated standard deviation (SD). For each cell line, predicted sensitivity scores based on our perturbation-informed signature (signature SS), a random one (random SS), a control one based on top up and down log_2FC DE genes between BxPC-3 and HCT-15 (topFC SS) were also reported.