Genome Medicine

METHOD   **Open Access**

Beyondcell: targeting cancer therapeutic heterogeneity in single-cell RNA-seq data

Coral Fustero-Torre[1], María José Jiménez-Santos[1], Santiago García-Martín[1], Carlos Carretero-Puche[2,3], Luis García-Jimeno[1], Vadym Ivanchuk[1], Tomás Di Domenico[1], Gonzalo Gómez-López[1] and Fátima Al-Shahrour[1*]

Abstract

We present Beyondcell, a computational methodology for identifying tumour cell subpopulations with distinct drug responses in single-cell RNA-seq data and proposing cancer-specific treatments. Our method calculates an enrichment score in a collection of drug signatures, delineating therapeutic clusters (TCs) within cellular populations. Additionally, Beyondcell determines the therapeutic differences among cell populations and generates a prioritised sensitivity-based ranking in order to guide drug selection. We performed Beyondcell analysis in five single-cell datasets and demonstrated that TCs can be exploited to target malignant cells both in cancer cell lines and tumour patients. Beyondcell is available at: https://gitlab.com/bu_cnio/beyondcell.

Keywords: Single-cell RNA-seq, Intratumoural heterogeneity, Drug repositioning, Therapeutic clusters, Personalised therapy

Background

Tumour heterogeneity (TH) refers to genetic and epigenetic differences of the same tumour type between patients, between tumours in a single patient, and within the cells of a tumour. It contributes to the medical complexity of cancer treatment and can lead to drug resistance, therapeutic failure and higher lethal outcome [1]. TH is closely related to tumour evolution which has been described as branching clonal evolution due to the accumulation of somatic mutations [2]. The heterogeneity of cancer cells, tumour evolution and clonal dynamics introduce significant challenges in designing effective treatment strategies. Unfortunately, these issues are not satisfactorily addressed in routine clinical practice as systematic efforts to consider inter-patient TH when choosing therapies are still limited, while they are extremely rare for intra-TH [3, 4].

Single-cell RNA sequencing (scRNA-seq) has become an established technology to dissect TH at the transcriptional level, revealing high-resolution cellular composition and dynamics and offering an unprecedented opportunity to address TH therapeutically [5–7]. For instance, scRNA-seq studies have been successfully employed to identify novel cancer cell subpopulations [8], tumour biomarkers [9], drug resistance pathways and therapeutic targets [10]. In addition, large-scale projects focused on obtaining a comprehensive molecular and pharmacological characterisation of the cancer cell lines have provided valuable datasets relating gene expression signatures with drug response and treatment sensitivity. Some examples include the Cancer Cell Line Encyclopedia (CCLE) [11], the Genomics of Drugs Sensitivity in Cancer (GDSC) [12], the Cancer Therapeutic Response Portal (CTRP) [13] and the Drug Repurposing Hub/LINCS [14].

In this situation, it is reasonable to hypothesise that drugs (or drug combinations) capable of targeting TH at single-cell resolution can be identified by integrating drug response profiles and scRNA-seq data in order to

* Correspondence: falshahrour@cnio.es
[1]Bioinformatics Unit, Spanish National Cancer Research Centre (CNIO), Calle Melchor Fernandez Almagro, 3, 28029 Madrid, Spain
Full list of author information is available at the end of the article

Fustero-Torre *et al. Genome Medicine*   (2021) 13:187

Page 2 of 15

discover therapeutic clusters (TCs), which we have defined as groups of cells with a similar drug response. This would help to address the tumour therapeutic complexity, revealing the impact of TH in response to drugs and the scope of tumour cells whose therapeutic approach could be managed with approved drugs, clinical trials or drug repositioning strategies. However, there are no current computational methodologies capable of relating single-cell gene expression and high-throughput drug screening datasets to suggest knowledge-driven treatments. To address these challenges, we developed Beyondcell (https://gitlab.com/bu_cnio/beyondcell) [15], a novel method for detecting tumour cell subpopulations with distinct drug response in order to estimate the tumour therapeutic complexity and propose differential drugs to target cell subpopulations, experimental conditions or phenotypes in scRNA-seq experiments. Our analysis covers the detection of TCs as well as the visualisation of results for biological and clinical interpretation. We have applied Beyondcell to five different studies focused on cancer therapeutic heterogeneity, spanning cancer cell lines [16–18], primary cancer samples [19] and mouse-derived xenografts [5]. Our results demonstrated that Beyondcell can characterise single-cell variability in drug response and recapitulate the biology of the datasets analysed. Beyondcell is able to reveal sensitive, innate and acquired drug-resistant cell subpopulations in cell lines and propose drugs to target them. Beyondcell also allowed us to explore tumour heterogeneity in patients relating it to clinical drug response data predicting responders and non-responders to immunotherapy in melanoma and suggesting drugs to overcome drug resistance in lung cancer patients.

Methods

Dataset processing
A re-analysis of the samples was applied using the bollito pipeline (https://gitlab.com/bu_cnio/bollito) [20], which enables to perform automated cell filtering, data normalisation and integration, cell cycle regression, expression cluster detection and differential expression via Seurat v3 [21].
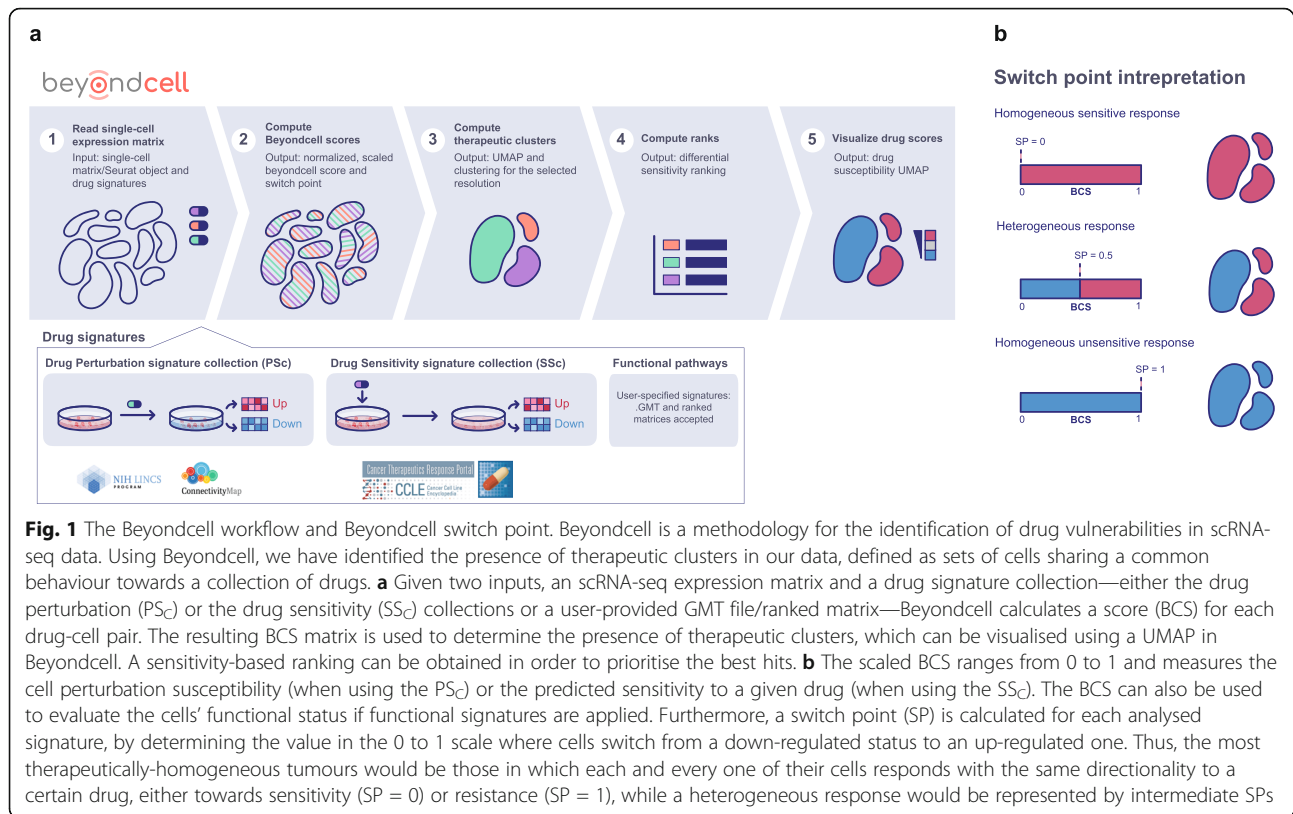
Beyondcell workflow and therapeutic clusters
An analysis with Beyondcell (https://gitlab.com/bu_cnio/beyondcell) [15] starts with a single-cell expression matrix and a collection of drug signatures: the drug perturbation (PS_C) and the drug sensitivity (SS_C) collections containing 4690 and 819 signatures respectively [11–14]. PS_C captures the transcriptional changes induced by a drug, while SS_C contains signatures reflecting the transcriptional status of sensitivity or resistance prior to drug treatment (Fig. 1a).

Our method calculates the Beyondcell Score (BCS) which estimates, for each cell in the preprocessed single-cell expression matrix, the enrichment in every drug signature in the specified collections. The BCS ranges from 0 to 1 and measures the cell perturbation susceptibility (using the PS_C) or the predicted sensitivity to a given drug (using the SS_C). The BCS can also evaluate the cells' functional status using functional gene sets such as molecular pathways, or cancer hallmarks. The calculated BCS matrix allows the determination of therapeutic clusters (TCs) within cellular populations defined as 'a set of cells that share a common response to a set of drugs'. In order to find potential TCs, a clustering analysis is applied to the BCS matrix where cells are grouped by their differential response to the selected drugs. For each drug, Beyondcell also calculates the switch point (SP), which reflects the homogeneity of the drug response throughout the single-cell dataset (Fig. 1b). Thus, the most therapeutically homogeneous tumours would be those in which each and every one of their cells responds in the same way to a certain drug, either with a sensitivity (SP = 0) or resistance (SP = 1) response, while a heterogeneous response would be represented by intermediate SPs. Then, the BCS matrix and the SP are used to generate a Uniform Manifold Approximation and Projection (UMAP), enabling the visualisation of the TCs reflecting the homogeneity in the drug responses within cellular populations, and to highlight in TC drug response, experimental conditions, cell subtypes, biomarker expression and cellular functional activity. Additionally, Beyondcell determines therapeutic differences among cell populations and generates a prioritised ranking of the differential sensitivity drugs between chosen conditions to guide drug selection (Additional file 1: Supplementary Methods and Figures).

Data description
The Library of Integrated Network-based Cellular Signatures (LINCS, http://www.lincsproject.org) [14] is a catalogue of gene expression data associated with cell lines exposed to a variety of perturbing agents, such as small molecules (about 5500), FDA drugs (~ 1300) and shRNA silencing (22,119 genetic perturbagens). LINCS, based on L1000 high throughput technology, is an extension of the Connectivity Map, which has been successfully used for drug repositioning [22]. From the ~ 20 K small molecules tested in the LINCS L1000 dataset, only those with an identifiable common drug name were selected. This reduced the number of drugs to 4690. In the CCLE [11], 28 drugs were tested, and the mRNA expression of 1037 cell lines was profiled using Affymetrix U133Plus2 arrays. The gene-centric RMA-normalised mRNA expression data, the cell line information and the drug response data from Cancer Cell Line Encyclopedia were

Fig. 1 The Beyondcell workflow and Beyondcell switch point. Beyondcell is a methodology for the identification of drug vulnerabilities in scRNA-seq data. Using Beyondcell, we have identified the presence of therapeutic clusters in our data, defined as sets of cells sharing a common behaviour towards a collection of drugs. **a** Given two inputs, an scRNA-seq expression matrix and a drug signature collection—either the drug perturbation (PS$_C$) or the drug sensitivity (SS$_C$) collections or a user-provided GMT file/ranked matrix—Beyondcell calculates a score (BCS) for each drug-cell pair. The resulting BCS matrix is used to determine the presence of therapeutic clusters, which can be visualised using a UMAP in Beyondcell. A sensitivity-based ranking can be obtained in order to prioritise the best hits. **b** The scaled BCS ranges from 0 to 1 and measures the cell perturbation susceptibility (when using the PS$_C$) or the predicted sensitivity to a given drug (when using the SS$_C$). The BCS can also be used to evaluate the cells' functional status if functional signatures are applied. Furthermore, a switch point (SP) is calculated for each analysed signature, by determining the value in the 0 to 1 scale where cells switch from a down-regulated status to an up-regulated one. Thus, the most therapeutically-homogeneous tumours would be those in which each and every one of their cells responds with the same directionality to a certain drug, either towards sensitivity (SP = 0) or resistance (SP = 1), while a heterogeneous response would be represented by intermediate SPs

downloaded from the CCLE data portal (https://portals.broadinstitute.org/ccle/home). In the CTRP [13], 545 drugs (single or combination) were tested against 887 cell lines. We downloaded the drug response data of the CRTP 2.1 project from the CTD2 Data Portal (https://ocg.cancer.gov/programs/ctd2/data-portal). The gene expression data used in order to obtain the CTRP expression signatures were those of the CCLE. Of the 887 cell lines, 805 were present in the CCLE dataset. In the GDSC data [12], 265 compounds corresponding to 250 different drugs were tested against 1074 cell lines. Cell lines were profiled for gene expression using the Affymetrix Human Genome U219 array. Expression data were normalised using RMA [23]. Drug response data and preprocessed mRNA gene expression data were downloaded from the GDSC web portal (http://www.cancerrxgene.org/downloads). The release 21Q1 of the Achilles dataset was downloaded from the DepMap portal [24].

Signature generation

A gene expression signature is a general model for the representation of the transcriptional changes associated to a given biological phenotype or perturbation. In this study, we have considered two expression signature collections: the PS$_C$ captures the transcriptional changes induced by a drug; while the SS$_C$ captures the sensitivity to the drug

effect. In both cases, these expression signatures are obtained from a differential expression analysis, although using different designs. Despite its different biological interpretations, both types of signature were represented as the two gene sets formed by the N most upregulated and downregulated genes (the *UP* and *DN* genesets). Several N were tested (50, 100, 250 and 500), and no significant disagreement was found between the top synergistic and antagonistic interactions (data not shown). Thus, following Iorio et al. rationale [25], N was taken as 250.

Drug perturbation signature collection (PS$_C$)

Drug-induced expression signatures were obtained from experiments in which the transcriptional state of the cell is measured before and after treatment with the drug. This makes it possible to study the transcriptional effect of the drug, that is, it informs on whether the transcriptional state of each cell is more similar to that of untreated or treated cells. But it also allows us to predict sensitivity under the *signature reversion* principle, which aims at the identification of drugs inducing a transcriptional response complementary to that of the disease. In order to obtain consensus expression signatures for each drug, a differential expression analysis was performed on control vs treated cells using *limma* [26]. Full details of PS$_C$ signature collection are available at Additional file 1: Supplementary Methods and Figures.

Fustero-Torre *et al. Genome Medicine*   (2021) 13:187

Page 4 of 15

Drug sensitivity signature collection (SS_C)

Drug-induced expression signatures were obtained from pharmacogenomics experiments in which the pharmacological response to a drug and the transcriptional state before treatment were considered. In order to obtain the expression signature, we considered the area under the curve (AUC) as a measure of drug response. For each drug, we performed a differential expression analysis between AUC high cell lines and AUC low cell lines by considering the AUC as a continuous variable using the *limma* R package. The area under the curve (AUC) was used as the measure of drug response; as contrary to IC_{50}, it can always be estimated without extrapolation from the dose-response curve and also because it has shown more accuracy in the prediction of drug response [27]. In all the projects considered, the tumour origin of the cell lines were considered as confounding variables. In the CCLE, the covariate was created by combining the *Site.Primary* and the *Histology* information of the cell line. In the GDSC 2.0, the variable *Site* and in the CTRP, the '*CCLE primary Site*' was used.

Further details of SS_C signature collection are available at Additional file 1: Supplementary Methods and Figures.

Functional pathways

The Beyondcell package [15] also includes a small collection of curated functional pathways obtained from the Molecular Signatures Database (MSigDB) [28]. These are meant to give the user some insight about the cells' viability and status. These pathways are related to the regulation of the epithelial-mesenchymal transition (EMT), cell cycle, proliferation, senescence and apoptosis. Furthermore, the package is able to accept external signatures in GMT format or as ranked matrices

Beyondcell score calculation

The Beyondcell score evaluates the activity of a signature of interest in a single-cell RNA-seq experiment. The transcriptomics data needs to be pre-processed, meaning that proper cell-based quality control filters, as well as normalisation, scaling and clustering of the data, should be applied prior to the analysis with Beyondcell. When analysing a bidirectional gene signature (a signature with separate sets of upregulated and downregulated genes), the Beyondcell score is independently obtained for each signature mode. The individual sum of the expression is calculated and divided by the number of genes in the given signature that are present in the single-cell expression matrix. The obtained raw scores are normalised in order to penalise cells with a great number of zeroes and/or with outliers. The individual up and down normalised scores are summed and rescaled between 0 and 1 by calculating their *switch point* (SP). In cases where

the gene signature is unidirectional, all steps remain the same, although the rescaling will only be applied to one set of normalised scores (see full details in Additional file 1: Supplementary Methods and Figures).

Drug-background selection

In order to obtain TCs, the Beyondcell methodology can optionally generate a background score matrix. The background score matrix allows the user to reduce the computation time. It aims to characterise the heterogeneity of the drug responses in the whole dataset, by using a reduced collection of drugs that is able to capture the main differences between individual cells. To generate this background selection, the drug specificity score (DSS) of the whole PS_C collection has been calculated and the first and last decile have been selected. Here, the rationale is the following: for each drug, the DSS score quantifies the similarity between the induced expression patterns across cell types. And while some drugs induce similar patterns across distinct cell types, the majority of them have different effects across all of them [29].

Regression of unwanted sources of variation

After obtaining the normalised and scaled Beyondcell scores, we observed that the normalisation step was not sufficient to avoid the correlation between sample-level metrics (number of genes, number of UMIs or even cell cycle status) and the calculated scores. To correct for this, we have implemented a regression function. *bcRegressOut* removes the effect of all unwanted sources of variation from the scoring matrix in two steps: first, it imputes missing data (if necessary) using a k-nearest neighbours (KNN) algorithm; second, it obtains the residuals derived from the regression model via QR decomposition. The obtained residuals can then be used for the downstream dimensional reduction steps.

Therapeutic clusters

The BCS matrix, including the computed scores for all drugs of interest, can be used as an input for a dimensionality reduction and clustering analysis. In this step, cells are grouped by their differential response to the analysed drugs. A Uniform Manifold Approximation and Projection (UMAP) allows the visualisation of the identified TCs. With this analysis, cells can be classified into distinct TCs, which represent sets of cells sharing a common response to a particular drug exposure.

Drug prioritisation

The *bcRank* function computes the BCS matrix statistics. In particular, it calculates the switch point, mean, median, standard deviation, variance, minimum, maximum, proportion of NaN and residuals' mean of each signature. We recommend prioritising drugs taking into

account both the switch point and the residuals. The Beyondcell package includes the function *bc4Squares*, which helps to visualise this prioritisation. A 4 squares plot consists in a scatter plot of the residuals' means (*x*-axis) vs the switch points (*y*-axis) of a specified cluster (either a TC or a group defined by experimental condition or phenotype). The 4 quadrants are highlighted: the top left and bottom right corners contain the drugs to which these cells are least/most sensitive, respectively. The centre quadrants show the drugs to which the selected cells are differentially insensitive or sensitive when compared to the other clusters.

Summary of datasets analysed to validate Beyondcell

We have applied Beyondcell to 5 studies featuring scRNASeq: (i) the Ben-David et al. study [16] focused on dissecting the therapeutic heterogeneity of 27 strains from the MCF7 cancer cell line after exposure to 321 anti-cancer compounds; (ii) Ho. and colleagues dataset [17], which identified drug-resistant cellular populations in melanoma cell lines, after exposure to BRAF inhibitors; (iii) the Pan-Cancer scRNA-seq dataset consisting on 198 cancer cell lines analysed by Kinker et. al. [18]; (iv) Jerby-Arnon et al. [19] studying malignant cells from melanoma patients resistant to immunotherapy and (v) Stewart et al. [5], focused on circulating tumour cells derived from mouse xenografts of small cell lung cancer patients undergoing chemotherapy, Further details about the preprocessing steps and Beyondcell analysis [15] are available in Additional file 1: Supplementary Methods and Figures.

Results

Reliability of the Beyondcell score and its application in cancer cell lines under drug exposure

We applied Beyondcell to the Ben-David et al. dataset [16] to demonstrate the reliability of the BCS and to validate its usefulness for identifying drug-response cell subpopulations. This study dissects the genetic and transcriptional heterogeneity within cancer cell lines, providing an scRNA-seq dataset that includes 7101 cells obtained from one single-cell-derived clone from the MCF7 cell line (MCF7-AA) exposed to bortezomib. Cells were collected before and after 12, 48 and 96 h of drug exposure (t0, t12, t48 and t96) to study its antiproliferative effects. Also, a drug screening of 321 anticancer compounds was performed to study drug response heterogeneity across 27 strains of the MCF7 breast cancer cell line.
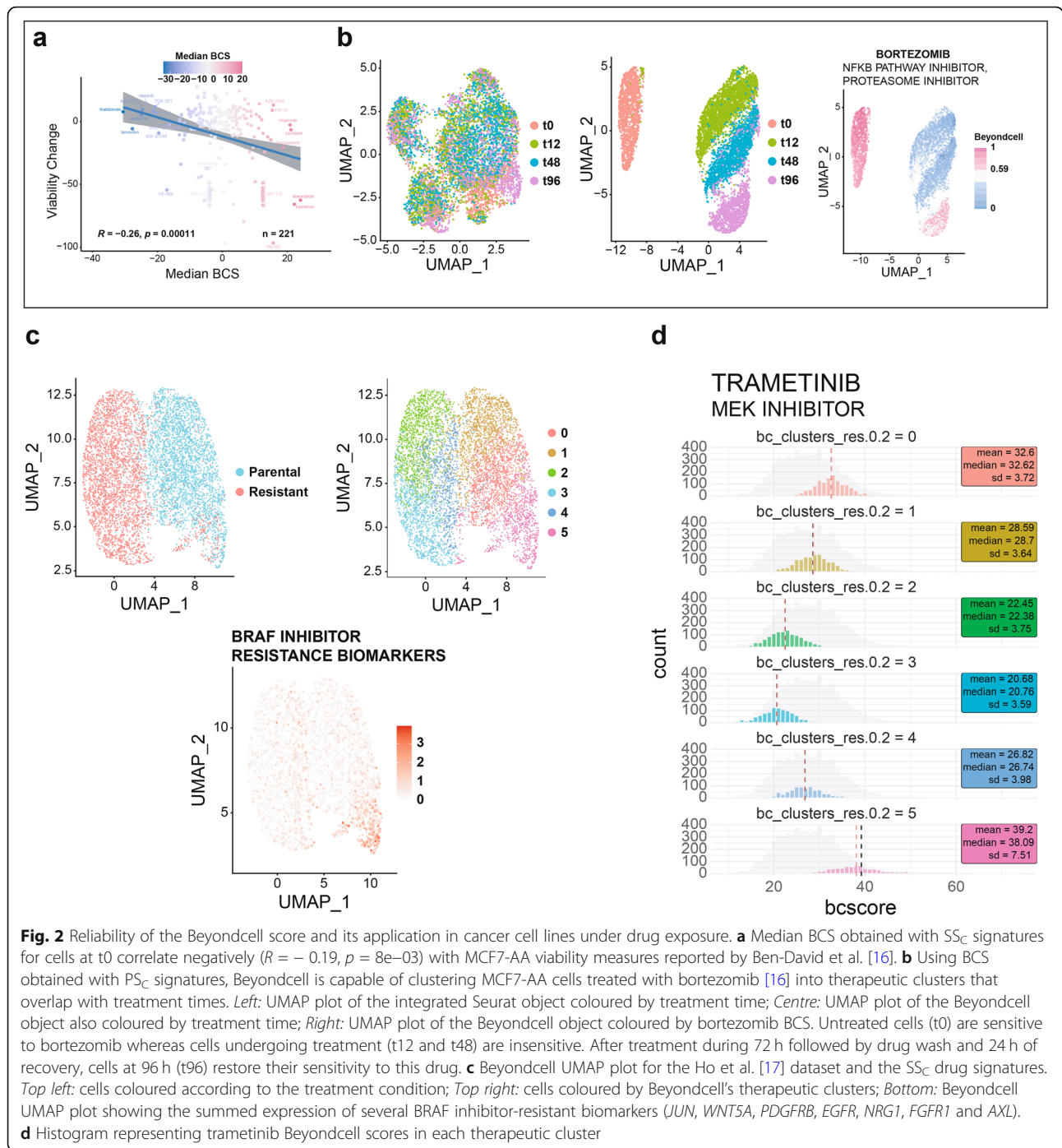
First, to demonstrate the reliability of the BCS, we computed BCS for each compound using SS_C drug signatures to the collected MCF7-AA cells at t0. We found that median BCS obtained with SS_C signatures for cells at t0 correlate significantly ($R = -0.19$, $p = 8e{-}03$) with MCF7-AA

cell viability measures reported by Ben-David et al. [16] after treatment, demonstrating that BCS reflects drug sensitivity (Fig. 2a). When we employed PS_C signatures, drugs were then classified in three groups: chemotherapy, targeted therapy and others (including immunotherapy, hormone therapy and photodynamic therapy). The BCS for targeted therapies showed a significant correlation with MCF7-AA cell viability ($R = -0.24$, $p = 8.4e{-}03$) (Additional file 1: Fig. S1A) while the rest of the therapies were not found significant. This could suggest that PS_C signatures for targeted therapies reflect more accurately which cells are more likely to respond than chemotherapy signatures.

Next, we were interested in evaluating Beyondcell's ability to identify distinct drug-response cell subpopulations before and after bortezomib exposure. Beyondcell's analysis with the PS_C collection clearly separates the cells into discrete clusters, in contrast to the mixed cell groups found using transcriptional profiles. The resulting therapeutic clusters not only separated bortezomib-treated and untreated cells but also retrieved drug exposure time-points. By focusing on a PS_C bortezomib expression signature, t12 and t48 cells reflected the perturbation status induced by bortezomib in contrast to t0 cells. Interestingly, t96 cells reverted to a pre-perturbation status for the bortezomib signature, highlighting the reversible inhibitory capacity of this proteasome inhibitor [30] (Fig. 2b). In particular, cells were more sensitive to bortezomib when combined with SNX-2112, an *HSP90* inhibitor, than in treatment with bortezomib alone, suggesting its role as a sensitiser to bortezomib treatment (Additional file 1: Fig. S1B). Finally, we applied Beyondcell removing the expression signatures of bortezomib from the PS_C collection in order to demonstrate that the therapeutic clusters are not solely driven by the effect of bortezomib signatures and that the remaining signatures are able to rescue the biology of untreated and treated cells (Additional file 1: Fig. S1C). In summary, in this study, we validate the BCS as a measure of drug sensitivity, determining its significant correlation with drug screening experimental data. In addition, we demonstrate Beyondcell's utility to recapitulate drug response in the MCF7 cancer cell line and propose single and drug combinations to sensitise the cells.

Analysis of BRAF inhibitor sensitive and resistant subpopulations in melanoma cells

Next, using data from a 451HLu human melanoma cell line treated with BRAF inhibitors (BRAFi), we evaluated Beyondcell's ability to identify drug-resistant cellular populations and to propose drug treatments [17]. Beyondcell analysis with SS_C drug signatures recapitulated 'parental' and 'BRAFi-resistant' cell populations and identified 6 different therapeutic clusters shown in a Beyondcell UMAP plot where cells are coloured according to the treatment condition or the Beyondcell

Fig. 2 Reliability of the Beyondcell score and its application in cancer cell lines under drug exposure. **a** Median BCS obtained with SS_C signatures for cells at t0 correlate negatively ($R = -0.19$, $p = 8e-03$) with MCF7-AA viability measures reported by Ben-David et al. [16]. **b** Using BCS obtained with PS_C signatures, Beyondcell is capable of clustering MCF7-AA cells treated with bortezomib [16] into therapeutic clusters that overlap with treatment times. *Left:* UMAP plot of the integrated Seurat object coloured by treatment time; *Centre:* UMAP plot of the Beyondcell object also coloured by treatment time; *Right:* UMAP plot of the Beyondcell object coloured by bortezomib BCS. Untreated cells (t0) are sensitive to bortezomib whereas cells undergoing treatment (t12 and t48) are insensitive. After treatment during 72 h followed by drug wash and 24 h of recovery, cells at 96 h (t96) restore their sensitivity to this drug. **c** Beyondcell UMAP plot for the Ho et al. [17] dataset and the SS_C drug signatures. *Top left:* cells coloured according to the treatment condition; *Top right:* cells coloured by Beyondcell's therapeutic clusters; *Bottom:* Beyondcell UMAP plot showing the summed expression of several BRAF inhibitor-resistant biomarkers (*JUN, WNT5A, PDGFRB, EGFR, NRG1, FGFR1* and *AXL*). **d** Histogram representing trametinib Beyondcell scores in each therapeutic cluster

therapeutic cluster. Therapeutic cluster 5 (TC5) included both parental and resistant cells. A small fraction of the parental cells in TC5 (15%) expressed BRAFi-resistance markers like *AXL, NRG1, DCT* and *FGFR1* that are also expressed in BRAFi-resistant cells, showing that they were clonally selected from the parental population contributing to the resistance (Fig. 2c; Additional file 1: Fig. S2).

Additionally, Beyondcell identified specific drugs to target TC5, BRAFi-resistant and parental cells. For

instance, cells in cluster 5 are differentially sensitive to dasatinib (SRCi) and unresponsive to dinaciclib (CDKi) and gemcitabine. On the other hand, cells in the resistant condition are differentially sensitive to bardoxolone methyl (NF-kBi) and unresponsive to AZD6482 (PIK3i) (Additional file 1: Fig. S3; Additional file 2: Table S1). MEK inhibitors (MEKi) are shown to target 451HLu cells including pre-resistant clones (TC5). For instance, trametinib had positive Beyondcell scores in each

Fustero-Torre *et al. Genome Medicine*   (2021) 13:187

Page 7 of 15

therapeutic cluster obtained for SS_C drug signatures, showing a higher BCS score (higher sensitivity) in TC5, and therefore, it could be proposed to target BRAFi-unresponsive cells. In fact, MEKi is a standard treatment in advanced melanoma patients in combination with BRAFi [31] (Fig. 2d; Additional file 1: Fig. S4A). We observed that TC5 cells were also grouped in expression UMAPs overlapping with the scRNA-seq expression cluster 2 (EC2), suggesting a direct relationship between the drug response profiles and gene expression patterns (Additional file 1: Fig. S4B). In order to identify novel genes involved in BRAFi-resistance, we performed a differential gene expression analysis for TC5, revealing 572 significantly overexpressed genes ($|\log_2(FC)| > 2$, FDR < 0.05) (Additional file 3: Table S2), including some known BRAFi-resistance biomarkers (e.g. *AXL* and *NRG1*). In addition, vemurafenib and dabrafenib BCSs obtained using SS_C showed that TC5 had lower sensitivity to RAF inhibitors than the rest of the TCs, confirming that TC5 cells are pre-resistant to BRAFi (Additional file 1: Fig. S4C). This validation demonstrates that Beyondcell is able to identify therapeutic clusters that recapitulate drug effects on cells, propose drugs to target sensitive and resistant cells and identify drug-response biomarkers.

Beyondcell characterises single-cell variability in drug response in pan-cancer cell line data

We also applied Beyondcell to describe the therapeutic heterogeneity in 198 cell lines from 22 cancer types [18]. Using SS_C drug signatures, we found 5 TCs where 12 of 22 cancer types were overrepresented in at least one of these TCs. TC0 was mostly constituted by cells from skin cancer (melanoma) and endometrial/uterine cancer cell lines, while TC1 included cells from bladder, gallbladder and pancreatic cancer cell lines. TC3 was enriched in breast and colon/colorectal cancer, while gliomas were exclusively located in TC2 together with kidney and thyroid cancer (Additional file 1: Fig. S5; Additional file 4: Table S3). TC4 was mainly constituted by two cell lines: the osteosarcoma (HOS) and sarcoma (A204) cell lines. Interestingly, 11 out of 12 brain cancer cell lines clustered together in TC2 independently of whether the lineage is astrocytoma or glioblastoma, with the exception of the single medulloblastoma cell line that is located in TC0. These results suggest that the cell lines from these 12 cancer types that tend to cluster in the same TCs have a common drug response.

In contrast, cancer types such as lung, gastric and ovarian among others showed high cellular therapeutic heterogeneity. For instance, the four bile duct cancer cell lines are each grouped in a different TC showing different drug response profiles despite belonging to the same cancer type (Additional file 1: Fig. S6A). Interestingly,

lung cancer cell lines are distributed between the TCs regardless of whether they are squamous or adenocarcinoma subtype showing diverse drug response (Additional file 1: Fig. S6B). We also tested whether these lung cancer cell lines expressing this distinct drug response pattern exhibit a unique pattern of mutations and genetic vulnerabilities. For this, we used the CCLE mutation dataset [32] as well as the Achilles dataset [24] of genome-wide CRISPR knockout screens to map known driver mutations in lung cancer (e.g. *KRAS*, *EGFR*, *PI3KCA*) and the genes identified as essential for proliferation. These analyses showed that lung cancer cell lines did not cluster together; therefore, TCs detected by Beyondcell are not driven by mutational and essentiality events (Additional file 1: Fig. S6C).

We hypothesised that the heterogeneous therapeutic landscape observed in cancer cell lines could also be an opportunity to infer drug repositioning strategies to target cell lines with different tissue origin or genetic background but clustered together in Beyondcell by similar drug responses [33]. For instance, most of the colon/colorectal cancer cell lines were clustered in TC3 except cells from the NCIH747 colorectal cell line, which are mostly concentrated in TC1 where bladder, gallbladder and pancreatic cancer cell lines are located (Additional file 1: Fig. S7). This suggests that the NCIH747 cell line could respond to tyrosine kinase inhibitors (TKIs) prescribed by Beyondcell for these tumour types such as EGFRi and also to inhibitors to target members from MAPK pathway (MEK and SRC). Interestingly, NCIH747 cell line has shown experimental sensitive response to selumetinib, a MEKi, this drug being the most differential sensitive drug for TC1 and for bladder and pancreatic cancer in our results. These findings highlight Beyondcell's ability to propose drugs for repurposing, providing additional drug response information using transcriptional data that could complement diagnostic information (i.e. tissue origin, stage, mutational status, etc.) that commonly guide cancer treatment.

Overall, global expression profiles clustered cells by cancer type; however, the TCs did not show such separation, suggesting less variability in the drug response. More specifically, 172 of 198 cell lines were overrepresented in the same TC, indicating that these cell lines had a lower cellular therapeutic heterogeneity than the rest of the cell lines ($n = 26$), which were spread across the different TCs (Additional file 4: Table S3). The high-grade serous ovarian cancer cell lines are a clear example of high therapeutic heterogeneity, since cells from these cell lines are mostly distributed between TCs (Additional file 1: Fig. S8). This observed differential drug sensitivity led by varying patterns of gene expression could result from clonal dynamics and continuous genetic instability that translates into heterogeneity in cancer cell lines [16].

Beyondcell results for SS_C revealed 136 differential sensitivity drugs for TCs and 183 drugs in cancer type comparison. In general, TC1 and TC4 showed higher sensitivity to EGFR and topoisomerase inhibitors respectively while TC0 and TC3 both showed higher sensitivity to PLK and CDK inhibitors (Fig. 3a; Additional file 1: Fig. S9). Using PS_C 174 drugs showed differential sensitivity for TCs and 569 drugs in cancer type comparison (Additional file 5: Table S4). However, TCs did not form discrete clusters so we expect that changes in drug responses are subtle, with a lot of commonalities. Beyondcell was also used to compute drug-response similarity correlation modules using BCS matrix (Additional file 1: Fig. S10; Additional file 6: Table S5). These correlation modules could be used to infer therapeutic mechanisms of action (MoA) for those drugs with unknown targets.

We validated Beyondcell results using a recently generated dataset of clinical compounds screened across 578 cell lines [33]. Differential drug vulnerability analysis showed that TC1 and TC2 are the most relevant clusters in terms of drug sensitivity, with TC1 featuring sensitivity to multiple drug families (EGFR, MEK, AKT and Aurora kinase inhibitors) compared to TC2 with lower sensitivity (Additional file 7: Table S6). TC1 and TC3 are enriched in cell lines with higher sensitivity to EGFR inhibitors, while TC2 showed decreased sensitivity, confirming Beyondcell results. Cell lines from TC1 (but not TC3) are differentially sensitive to MEK inhibitors compared to those from TC2. Interestingly, Beyondcell predicts not only that EGFRi is the most enriched drug for TC1, but also for bladder and gallbladder cancer (Additional file 1: Fig. S11; Additional file 8: Table S7). Moreover, *EGFR* is overexpressed in up to 74% of bladder cancer tissue specimens but has a relatively low expression in normal urothelium suggesting that it could be a potential therapeutic target. In addition, *EGFR* is an independent predictor of decreased survival and stage progression in bladder cancer [34].

Beyondcell calculates an SP for every drug, providing a measure of drug response homogeneity and sensitivity in each cell line. For instance, cell line NCIH1568 (NSCLC, metastatic) presented higher drug response heterogeneity than RERFLCAD1 (NSCLC, primary), evidencing a more variable drug-response behaviour in the metastatic cell line. We found that 48% of cancer cell lines show > 60% of drug homogeneity (SP = 0 or SP = 1) suggesting low cell variability in drug response. A total of 88% of the cell lines have a median SP > 0.7, meaning that they contain a greater number of insensitive than sensitive cells, and 17% of cell lines have a median SP > 0.9, indicating that none of the cells would exhibit sensitivity against half of the therapeutic options (Additional file 1: Fig. S12; Additional file 9: Table S8).

To explore the functional properties of the TCs, we correlated BCS and 12 expression programs known to be recurrently heterogeneous within cancer cell lines (RHPs). Cells enriched in the epithelial senescence-associated (EpiSen) program correlated ($r > 0.6$) with high sensitivity to EGFRi, in agreement with experimental validations performed by Kinker and colleagues [18] (Fig. 3b). The EMT program (EMT-II) presented higher activity in TC2 cells and correlated with high sensitivity to PI3K and HMGCR inhibitors. In contrast, TC0 and TC3 cells correlated with low EMT-II activity and high sensitivity to HDAC inhibitors, in agreement with previous studies [35, 36] (Additional file 1: Fig. S13; Additional file 10: Table S9). Interestingly, EMT-high TC2—mostly represented by brain cancer cell lines—presents a more undifferentiated transcriptional state in contrast to the rest of the therapeutic clusters ($p = 8.3e{-}13$) [36]. TC2 differential expression analysis showed an upregulated mesenchymal profile and enrichment of the EMT pathway (Additional file 1: Fig. S14; Additional file 11: Table S10), with Beyondcell results also showing a lower sensitivity to EGFRi (Additional file 5: Table S4), confirming previous results where undifferentiated cell lines showed decreased sensitivity to EGFRi (Additional file 12: Table S11) [37]. This result shows that distinct cell functional states lead to different drug responses which are successfully detected by Beyondcell TCs. Overall, our study reveals the therapeutic landscape in multiple cancer cell lines, finding recurrent patterns of drug heterogeneity shared by specific cancer types, cell lines, as well as their relationship with functional status.

Beyondcell characterises single-cell variability in drug response in cancer patients

We also employed our method to study more heterogeneous samples such as primary tumour samples. First, we used an scRNA-seq dataset from 16 melanoma patients treated with immune checkpoint inhibitors (ICIs) and 15 untreated patients. This prospective study aimed to detect transcriptional cell states related to responsiveness to ICIs and identify a transcriptional ICI-resistance program expressed by malignant cells associated with T cell exclusion and immune evasion which predicts clinical responses to immunotherapy in melanoma patients [19]. Beyondcell revealed 7 TCs, where TC2 and TC5 mainly contained cells from the untreated patients Mel79 and Mel81 respectively, whereas TC4 and TC6 included ICI-resistant patients (Additional file 1: Fig. S15). Non-responder patients were predicted by calculating the activation level of the ICI-resistance program with Beyondcell, thus tumour cells showing high BCS would have low response to ICIs. As expected, TC4 and TC6 defined by ICI-resistant patients showed high BCS values. TC5, which included the untreated patient
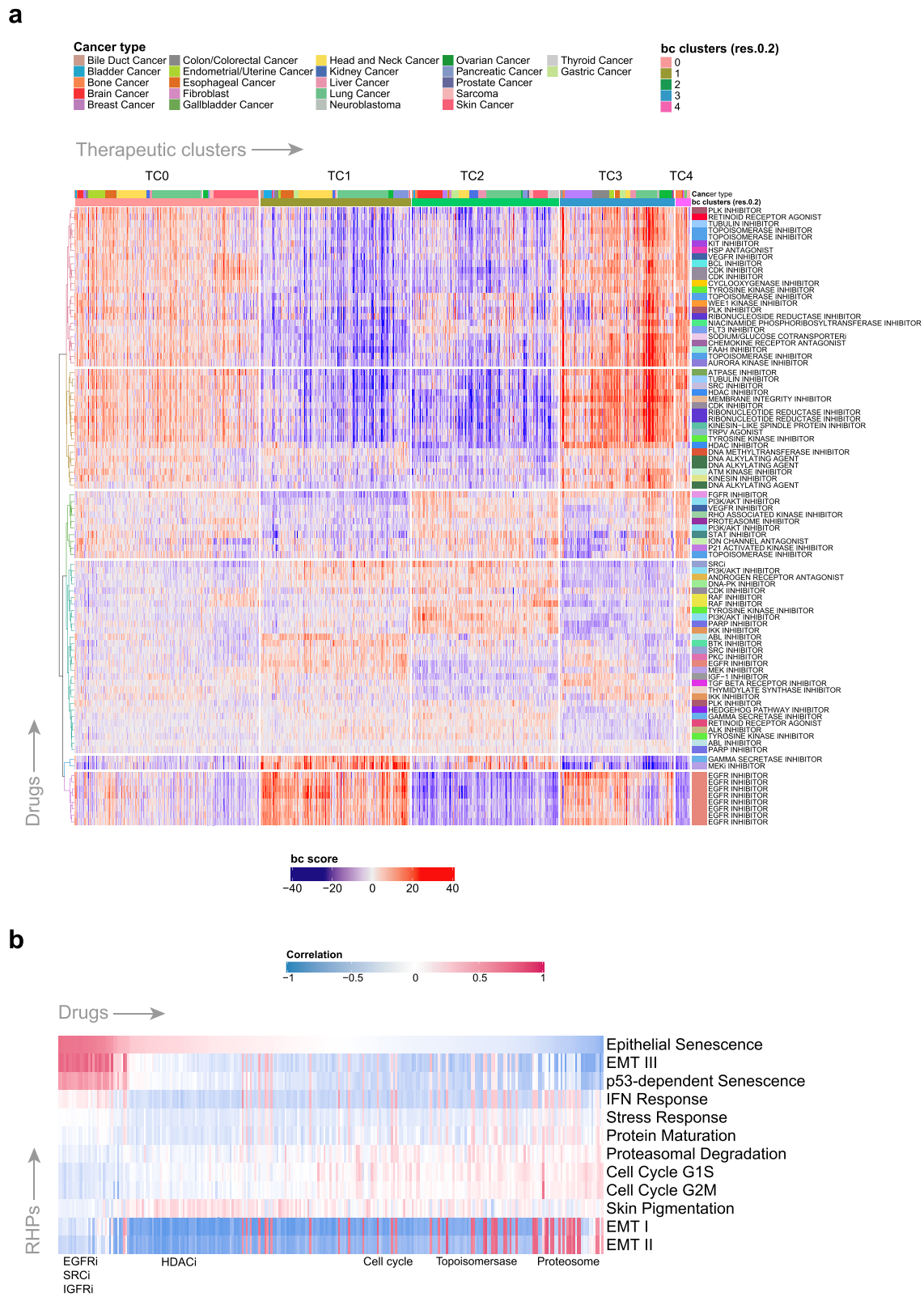
Fig. 3 (See legend on next page.)

(See figure on previous page.)

Fig. 3 Beyondcell characterises single-cell variability in drug response in pan-cancer cell line data. **a** Heatmap depicting both the cluster and cancer-specific drugs with a heterogeneous sensitivity pattern found in Kinker et al. [18]. The heatmap has been ordered according to the detected therapeutic clusters, while drug signatures have been clustered based on their BCS. Furthermore, the distinct cancer types have been colour coded. Only drugs with a mechanism-of-action (MoA) annotation are shown. **b** The heatmap depicts the correlation score between each BCS and the described RHPs. Here, we found EGFR inhibitors to be highly correlated with the Epithelial Senescence, EMT III and p53-dependent senescence programs, while tyrosine kinase inhibitors or topoisomerase inhibitors were anticorrelated. Further details are available in Additional file 1: Supplementary Methods and Figures

Mel81, also had a high BCS, so we concluded that this patient could not respond to ICIs (Fig. 4a). Beyondcell showed that Mel81 would respond to CDK inhibitor (CDKi) drugs such as alvocidib, which has been proposed for use in combination with immunotherapy to improve the response in ICI-resistant patients. Conversely, ICIs in monotherapy would be the preferred treatment for Mel79 (Fig. 4b; Additional file 1: Fig. S16). This validation demonstrated that Beyondcell correctly identified non-responders to ICIs, confirming that CDKi can be a therapeutic option to overcome ICI-resistance.

In a second study, we wanted to dissect therapeutic heterogeneity and propose novel therapeutic strategies in small-cell lung cancer patients (SCLC) patients treated with cisplatin [5]. In general, SCLC patients' initial responses occur in > 60% of patients; however, most patients relapse within 6 months. After relapse, approved therapies are effective in < 20% of patients with a median overall survival of about 10 months, indicating a dramatic shift towards resistance [38]. This remarks the need both to improve first-line treatments and to offer second-line therapies for refractory SCLC patients. With this aim, we applied Beyondcell to analyse a selection of SCLC circulating tumour cell-derived xenografts (CDX) models that included both platinum-sensitive and platinum-resistant samples. After Beyondcell analysis, we observed how the TCs were primarily driven by the patient/CDX origin regardless of response to platinum, thus underscoring the SCLC intertumoural heterogeneity (Fig. 4c). In addition, Beyondcell proposed drugs to target platinum-resistant cells including known inhibitors as well as approved drugs for repurposing (Additional file 13: Table S12). Interestingly, we found DNA repair inhibitor (bendamustine), AURKA inhibitors (GSK1070916), CHEK inhibitors (BX-912) and BCL inhibitors (navitoclax) inhibiting known therapeutic targets in SCLC [39, 40]. Beyondcell also proposes to target the MYC signalling pathway and EMT process using mTOR and PI3K inhibitors to overcome platinum-resistance (Fig. 4c).

Discussion

Addressing TH is a critical factor for the design of effective treatments in cancer and a current challenge for precision oncology [41]. Consequently, there is a need

for methodologies to detect tumour clone-specific drug sensitivities in order to properly characterise drug responses in cancer and, more importantly, to prioritise those treatments that could be clinically more effective for each patient. Here, we aim to address this challenge by introducing Beyondcell, the first method to define tumour cell subpopulations of differential drug response and propose cancer-specific treatments using single-cell RNA-seq data. A key concept in Beyondcell is the "therapeutic cluster", defined as a group of cells sharing a common drug response, which aims to address tumour therapeutic heterogeneity. For this purpose, Beyondcell uses a single-cell expression matrix and LINCS/CMap, GDSC, CCLE and CTRP drug-signature collections to calculate the Beyondcell score, an indicator for each cell of its degree of sensitivity to a drug. Then, the method also calculates the SP, a quantitative measure of the drug response homogeneity throughout the single-cell dataset reflecting the cellular variability. Finally, Beyondcell makes it possible to uncover and visualise TCs, providing a drug sensitivity ranking that prioritises the best hits to target TCs, cell subpopulations, experimental conditions or phenotypes.

As a proof of concept, we applied Beyondcell in five independent datasets comprising cancer cell lines [16–18] and patients [5, 19]. We observed that Beyondcell successfully identifies tumour subpopulations based on therapeutic response and can detect differential treatments amongst a range of experimental conditions or cell clusters. In addition, Beyondcell recapitulates the biology of the datasets analysed and easily reveals sensitive, innate and acquired drug-resistant cell subpopulations in cell lines under drug exposure. Furthermore, Beyondcell is able to propose a possible treatment strategy to overcome such resistance and identifies drug-response markers [16, 17]. We also successfully applied Beyondcell to characterise single-cell variability in drug response in a pan-cancer cell line dataset, identifying cellular functional activities and associating them with common patterns of drug response shared by specific cancer types and cell lines [18]. In addition, Beyondcell allowed us to explore tumour heterogeneity in patients relating it to clinical drug response data to successfully predict responders and non-responders to immunotherapy in melanoma [19] and suggest second-line therapies

Fustero-Torre *et al. Genome Medicine*   (2021) 13:187

Page 11 of 15
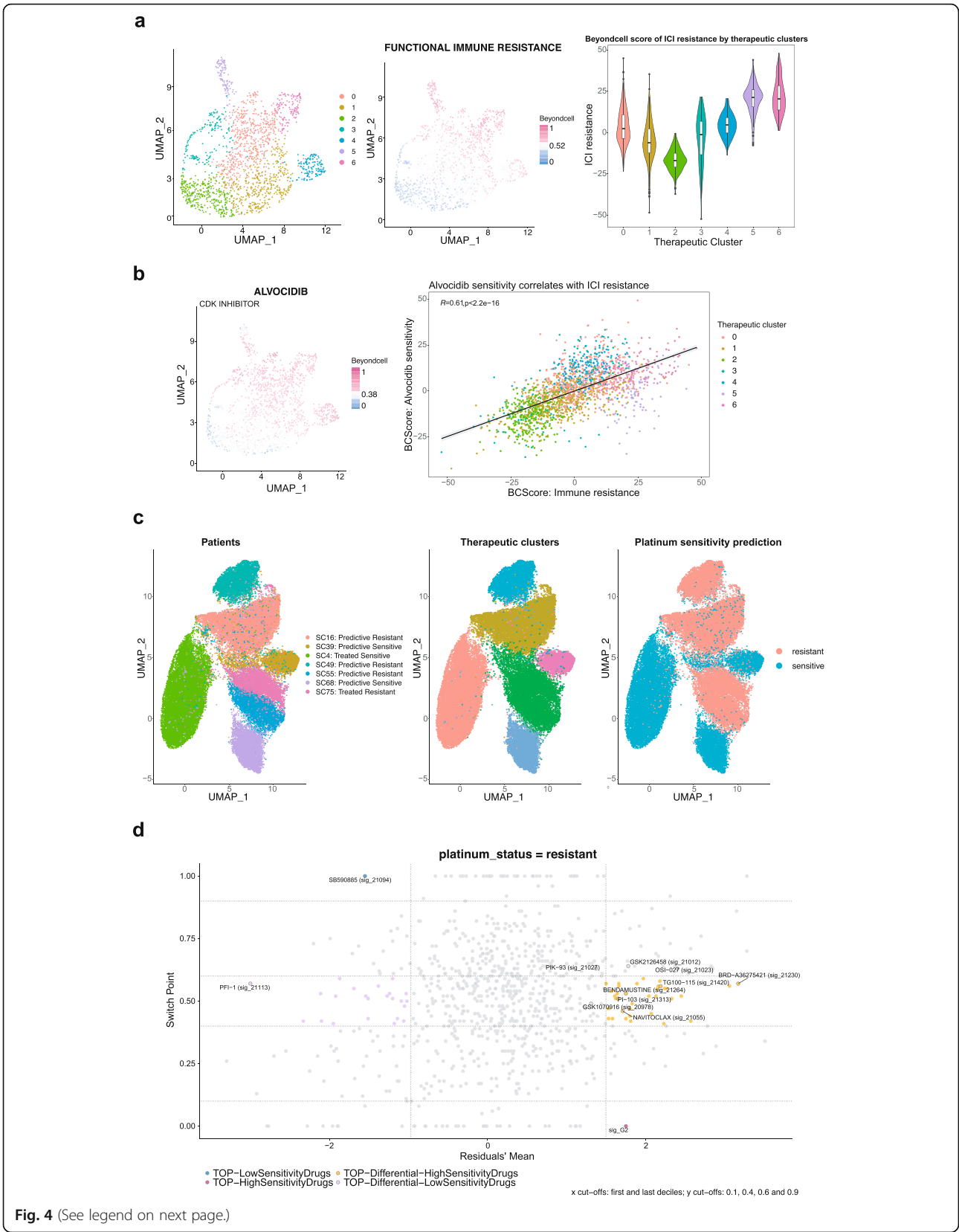
Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Beyondcell characterises single-cell variability in drug response in cancer patients. **a** Using the Jerby-Arnon dataset [19], we were able to recapitulate the melanoma cells' resistance to immune checkpoint inhibitors (ICI) at single-cell resolution. *Left:* therapeutic SS_C clusters, each cell is coloured according to the cluster it belongs to. *Centre:* UMAP representing Beyondcell's predicted affinity to the functional ICI resistance signature. Higher scores indicate a higher concordance between a cell transcriptome and the signature. *Right:* A violin plot of the normalised ICI resistance score by therapeutic cluster. **b** Beyondcell proposes alvocidib as a therapeutic alternative to ICI cells. *Left:* UMAP of alvocidib drug score (SS_C collection), where higher scores indicate cells whose transcriptome is concordant with alvocidib sensitivity. *Right:* A scatterplot of alvocidib drug score and ICI resistance score. Alvocidib sensitivity positively correlates with immune resistance ($R = 0.61$, $p < 0.01$). **c** Beyondcell recapitulates the heterogeneous response of SCLC patients [5]. *Top right:* UMAP plot of the patient's distribution after obtaining Beyondcell's therapeutic clusters. *Top centre*: Beyondcell's detected therapeutic clusters. The UMAP shows how the TCs were primarily driven by the patient/CDX origin regardless of response to platinum. *Top right*: UMAP coloured by the predicted sensitivity status of the patients. *Bottom*: summary of Beyondcell's proposed drugs to target platinum-resistant cells (further details are available in Additional file 13: Table S12)

to overcome platinum-resistance in SCLC [5]. Overall, Beyondcell allows an in-depth exploration of the TH impact in cancer therapeutics at the transcriptional level, addressing the variability of drug response in cancer cell lines and patients.

Nevertheless, the therapeutic heterogeneity revealed in our analysis pinpoints where the integration of new information is most needed. Other single-cell measurements such as cell imaging, genetic and epigenetic state or measuring gene expression at various time points could reveal differences in the drug responses of tumour cell populations [42]. Our results also reinforce the importance of tumour evolution for cancer diagnosis and treatment. Since tumour heterogeneity is related to the processes of clonal evolution, the identification of clonal populations, the clonal dynamics under selective pressure and the potential for competitive release of drug-resistant tumour subclones should be considered to prescribe more effective therapies [2]. A key aspect here is to incorporate the cells' genetic states, including information on druggable mutations and tumour vulnerabilities that could be used to target drug-tolerant cells in addition to considering the clonal evolution, in order to determine the best therapeutic opportunities to prevent or overcome tumour resistance [43].

The tumour microenvironment represents an additional layer in cancer therapeutics. Immunoediting processes—that is, the dynamic interactions between tumour cells and the immune system—drive tumour evolution and contribute to therapeutic failure [44]. Several single-cell studies have suggested targeting immune cells to overcome drug resistance in some refractory tumours [45, 46]. Beyondcell application could be extended to the context of immune cells as a therapeutic target in cancer to study its relationship with drug treatments [47]. Further short-term applications of Beyondcell would include its application to study the relationship between tumour spreading behaviour, development of metastatic phenotypes [48] and Beyondcell adaptation to the cancer single-cell spatial transcriptomics scenario [49].

Conclusions

In summary, understanding how TH leads drug response in patients will help to design precise therapeutic regimens (single-agent, combination or sequential treatments) anticipating the appearance of relapse, managing drug resistance mechanisms, delaying tumour growth or even inducing complete tumour regression. Therefore, there is an urgent need to develop methodologies to directly address TH in the design of anticancer treatment regimens. Beyondcell provides a valuable resource for better understanding of the biological and therapeutic impact of TH. Our results highlight the utility of Beyondcell to reveal, from single-cell transcriptomics, the cellular heterogeneity in drug response in cancer cell lines and patients, identifying resistant and sensitive cellular subpopulations and proposing drugs to target them. Finally, Beyondcell has been implemented as an open-source software package that can be easily combined with scRNA-seq gold-standard methodologies in custom automated analysis pipelines. Our software is complementary to current single-cell analysis approaches, opening up a discovery space to support the design of more effective lines of therapy.

Abbreviations
TCs: Therapeutic clusters; TH: Tumour heterogeneity; scRNA-seq: Single-cell RNA sequencing; CCLE: Cancer Cell Line Encyclopedia; CDX: Circulating tumour cell-derived xenografts; GDSC: Genomics of Drugs Sensitivity in Cancer; CTRP: Cancer Therapeutic Response Portal; LINCS: Library of Integrated Network-based Cellular Signatures; PS_C: Drug perturbation signature collection; SS_C: Drug sensitivity signature collection; SCLC: Small-cell lung cancer; MSigDB: Molecular Signatures Database; EMT: Epithelial-mesenchymal transition; SP: Switch Point; BCS: Beyondcell score; DSS: Drug specificity score; KNN: k-nearest neighbours; UMAP: Uniform Manifold Approximation and Projection; PCA: Principal components analysis; HSP: Heat shock proteins; RHPs: Recurrently heterogeneous programs; EpiSen: Epithelial senescence-associated; ICIs: Immune checkpoint inhibitors; CDKi: CDK inhibitor

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13073-021-01001-x.

Additional file 1: Supplementary Methods and Figures.
Supplementary Methods and Supplementary Figures S1–S16.

Fustero-Torre *et al. Genome Medicine*   (2021) 13:187

Page 13 of 15

Additional file 2: Table S1. Beyondcell results for clusters and conditions with Ho *et al.* dataset.

Additional file 3: Table S2. Genes differentially expressed in cluster 5 for Ho *et al.* dataset.

Additional file 4: Table S3. Cell distribution in clusters, cancer types and cell lines for Kinker *et al.* dataset.

Additional file 5: Table S4. Differential sensitivity drugs in cancer type and clusters for Kinker *et al.* dataset.

Additional file 6: Table S5. Beyondcell drug scores correlation modules for Kinker *et al.* dataset.

Additional file 7: Table S6. Differential drug vulnerability analysis for TCs for Kinker *et al.* dataset using PRISM dataset.

Additional file 8: Table S7. Differential EGFRi vulnerability analysis for cell lines in TC1 in comparison to TC2 for Kinker *et al.* dataset using PRISM dataset.

Additional file 9: Table S8. Cell lines drug homogeneity for Kinker *et al.* dataset.

Additional file 10: Table S9. Correlation of drug sensitivity and Recurrent Heterogeneous Programs (RHP) Beyondcell scores.

Additional file 11: Table S10. Differential gene expression analysis of TC2 cells vs rest of TCs.

Additional file 12: Table S11. Differential EGFRi vulnerability analysis for undifferentiated cell lines in comparison to differentiated cells.

Additional file 13: Table S12. Differential sensitivity drugs for platinum-resistant patients for Stewart *et al.* dataset.

Availability of data and materials
The Beyondcell algorithm is implemented in R (v. 4.0.0 or greater). All code and full documentation are available online at: https://gitlab.com/bu_cnio/beyondcell [15].
The Ben-David et al. dataset [16] was obtained from GEO (GSE114461). Ho et al. data [17] was downloaded from the SRA database (SRP127328). Both the pan-cancer samples from Kinker et al. [18] and the Jerby-Arnon et al. study [19] were obtained from the Single Cell Data Portal website (https://singlecell.broadinstitute.org/single_cell). The Stewart et al. dataset [5] was obtained from GEO (GSE138474). Sample re-analysis was done with the bollito pipeline (https://gitlab.com/bu_cnio/bollito) [20].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Bioinformatics Unit, Spanish National Cancer Research Centre (CNIO), Calle Melchor Fernandez Almagro, 3, 28029 Madrid, Spain. [2]Laboratorio de Oncología Clínico-Traslacional, Unidad de Investigación en tumores Digestivos, Instituto de Investigación I+12, Hospital 12 de Octubre, Av. de Córdoba, 28041 Madrid, Spain. [3]Lung-H120 Group, Spanish National Cancer Research Centre (CNIO), Calle Melchor Fernandez Almagro, 3, 28029 Madrid, Spain.

References
1. Turajlic S, Sottoriva A, Graham T, Swanton C. Resolving genetic heterogeneity in cancer. Nat Rev Genet. 2019;20(7):404–16 Available from: https://doi.org/10.1038/s41576-019-0114-6.
2. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. Cell. 2017;168(4):613–28 Available from: https://doi.org/10.1016/j.cell.2017.01.018.
3. Li X, Francies HE, Secrier M, Perner J, Miremadi A, Galeano-Dalmau N, et al. Organoid cultures recapitulate esophageal adenocarcinoma heterogeneity providing a model for clonality studies and precision therapeutics. Nat Commun. 2018;9(1):2983 Available from: https://doi.org/10.1038/s41467-018-05190-9.
4. Saeed K, Ojamies P, Pellinen T, Eldfors S, Turkki R, Lundin J, et al. Clonal heterogeneity influences drug responsiveness in renal cancer assessed by ex vivo drug testing of multiple patient-derived cancer cells. Int J Cancer. 2019;144(6):1356–66 Available from: https://doi.org/10.1002/ijc.31815.
5. Stewart CA, Gay CM, Xi Y, Sivajothi S, Sivakamasundari V, Fujimoto J, et al. Single-cell analyses reveal increased intratumoral heterogeneity after the onset of therapy resistance in small-cell lung cancer. Nat Cancer. 2020;1(4):423–36. Available from: https://doi.org/10.1038/s43018-019-0020-z.
6. Sharma A, Merritt E, Hu X, Cruz A, Jiang C, Sarkodie H, et al. Non-genetic intra-tumor heterogeneity is a major predictor of phenotypic heterogeneity and ongoing evolutionary dynamics in lung tumors. Cell Rep. 2019;29(8):2164–74.e5 Available from: https://doi.org/10.1016/j.celrep.2019.10.045.
7. Hoffman JA, Papas BN, Trotter KW, Archer TK. Single-cell RNA sequencing reveals a heterogeneous response to glucocorticoids in breast cancer cells. Commun Biol. 2020;3(1):126 Available from: https://doi.org/10.1038/s42003-020-0837-0.
8. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018;556(7702):463–8 Available from: https://doi.org/10.1038/s41586-018-0040-3.
9. Helmink BA, Reddy SM, Gao J, Zhang S, Basar R, Thakur R, et al. B cells and tertiary lymphoid structures promote immunotherapy response. Nature. 2020;577(7791):549–55 Available from: https://doi.org/10.1038/s41586-019-1922-8.
10. Cohen YC, Zada M, Wang S-Y, Bornstein C, David E, Moshe A, et al. Identification of resistance pathways and therapeutic targets in relapsed multiple myeloma patients through single-cell sequencing. Nat Med. 2021;27(3):491–503 Available from: https://doi.org/10.1038/s41591-021-01232-w.
11. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483(7391):603–7 Available from: https://doi.org/10.1038/nature11003.
12. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012;483(7391):570–5 Available from: https://doi.org/10.1038/nature11005.

Fustero-Torre *et al. Genome Medicine*   (2021) 13:187

Page 14 of 15

13. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell. 2013;154(5):1151–61 Available from: https://doi.org/10.1016/j.cell.2013.08.003.

14. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. Cell. 2017;171(6):1437–52.e17 Available from: https://doi.org/10.1016/j.cell.2017.10.049.

15. Fustero-Torre C, Jiménez-Santos MJ, García-Martín S, Carretero-Puche C, García-Jimeno L, Ivanchuk V, et al. Beyondcell https://gitlab.com/bu_cnio/beyondcell.git. 2021. Available from: https://zenodo.org/record/5602819

16. Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, et al. Genetic and transcriptional evolution alters cancer cell line drug response. Nature. 2018;560(7718):325–30 Available from: https://doi.org/10.1038/s41586-018-0409-3.

17. Ho Y-J, Anaparthy N, Molik D, Mathew G, Aicher T, Patel A, et al. Single-cell RNA-seq analysis identifies markers of resistance to targeted BRAF inhibitors in melanoma cell populations. Genome Res. 2018;28(9):1353–63 Available from: https://doi.org/10.1101/gr.234062.117.

18. Kinker GS, Greenwald AC, Tal R, Orlova Z, Cuoco MS, McFarland JM, et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. Nat Genet. 2020;52(11):1208–18 Available from: https://doi.org/10.1038/s41588-020-00726-6.

19. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su M-J, Melms JC, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell. 2018;175(4):984–97.e24 Available from: https://doi.org/10.1016/j.cell.2018.09.006.

20. Jimeno L, Fustero-Torre C, Jiménez-Santos MJ, Gómez-López G, Di Domenico T, Al-Shahrour F. bollito: a flexible pipeline for comprehensive single-cell RNA-seq analyses. Bioinformatics 2021. In press. Available from: https://doi.org/10.1093/bioinformatics/btab758

21. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. Cell. 2019;177(7):1888–902.e21 Available from: https://doi.org/10.1016/j.cell.2019.05.031.

22. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929–35 Available from: https://doi.org/10.1126/science.1132939.

23. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249–64 Available from: https://doi.org/10.1093/biostatistics/4.2.249.

24. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell. 2017;170(3):564–76. https://doi.org/10.1016/j.cell.2017.06.010.

25. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, Ferriero R, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci U S A. 2010;107(33):14621–6 Available from: https://doi.org/10.1073/pnas.1000138107.

26. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47 Available from: https://doi.org/10.1093/nar/gkv007.

27. Jang IS, Neto EC, Guinney J, Friend SH, Margolin AA. Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. Pac Symp Biocomput. 2014:63–74 Available from: https://www.ncbi.nlm.nih.gov/pubmed/24297534.

28. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739–40 Available from: https://doi.org/10.1093/bioinformatics/btr260.

29. Hodos R, Zhang P, Lee H-C, Duan Q, Wang Z, Clark NR, et al. Cell-specific prediction and application of drug-induced gene expression profiles. Pac Symp Biocomput. 2018;23:32–43 Available from: https://www.ncbi.nlm.nih.gov/pubmed/29218867.

30. Chen D, Frezza M, Schmitt S, Kanwar J, Dou QP. Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. Curr Cancer Drug Targets. 2011;11(3):239–53 Available from: https://doi.org/10.2174/156800911794519752.

31. Dummer R, Ascierto PA, Gogas HJ, Arance A, Mandala M, Liszkay G, et al. Encorafenib plus binimetinib versus vemurafenib or encorafenib in patients with BRAF-mutant melanoma (COLUMBUS): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2018;19(5):603–15 Available from: https://doi.org/10.1016/S1470-2045(18)30142-6.

32. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER 3rd, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569(7757):503–8 Available from: https://doi.org/10.1038/s41586-019-1186-3.

33. Corsello SM, Nagari RT, Spangler RD, Rossen J, Kocak M, Bryan JG, et al. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling. Nat Cancer. 2020;1(2):235–48 Available from: https://doi.org/10.1038/s43018-019-0018-6.

34. Siddiqui MR, Railkar R, Sanford T, Crooks DR, Eckhaus MA, Haines D, et al. Targeting epidermal growth factor receptor (EGFR) and human epidermal growth factor receptor 2 (HER2) expressing bladder cancer using combination photoimmunotherapy (PIT). Sci Rep. 2019;9(1):2084 Available from: https://doi.org/10.1038/s41598-019-38575-x.

35. Ishikawa T, Hosaka YZ, Beckwitt C, Wells A, Oltvai ZN, Warita K. Concomitant attenuation of HMG-CoA reductase expression potentiates the cancer cell growth-inhibitory effect of statins and expands their efficacy in tumor cells with epithelial characteristics. Oncotarget. 2018;9(50):29304–15 Available from: https://doi.org/10.18632/oncotarget.25448.

36. Tang HM, Kuay KT, Koh PF, Asad M, Tan TZ, Chung VY, et al. An epithelial marker promoter induction screen identifies histone deacetylase inhibitors to restore epithelial differentiation and abolishes anchorage independence growth in cancers. Cell Death Discov. 2016;2(1):16041. Available from: https://doi.org/10.1038/cddiscovery.2016.41.

37. Warren A, Chen Y, Jones A, Shibue T, Hahn WC, Boehm JS, et al. Global computational alignment of tumor and cell line transcriptional profiles. Nat Commun. 2021;12(1):22 Available from: https://doi.org/10.1038/s41467-020-20294-x.

38. Farago AF, Keane FK. Current standards for clinical management of small cell lung cancer. Transl Lung Cancer Res. 2018;7(1):69–79 Available from: https://doi.org/10.21037/tlcr.2018.01.16.

39. Byers LA, Wang J, Nilsson MB, Fujimoto J, Saintigny P, Yordy J, et al. Proteomic profiling identifies dysregulated pathways in small cell lung cancer and novel therapeutic targets including PARP1. Cancer Discov. 2012; 2(9):798–811 Available from: https://doi.org/10.1158/2159-8290.CD-12-0112.

40. Galetta D, Cortes-Dericks L. Promising therapy in lung cancer: spotlight on aurora kinases. Cancers. 2020;12(11) Available from: https://doi.org/10.3390/cancers12113371.

41. Qiu Z, Li H, Zhang Z, Zhu Z, He S, Wang X, et al. A pharmacogenomic landscape in human liver cancers. Cancer Cell. 2019;36(2):179–93.e11 Available from: https://doi.org/10.1016/j.ccell.2019.07.001.

42. Suvà ML, Tirosh I. Single-cell RNA sequencing in cancer: lessons learned and emerging challenges. Mol Cell. 2019;75(1):7–12 Available from: https://doi.org/10.1016/j.molcel.2019.05.003.

43. Piñeiro-Yáñez E, Jiménez-Santos MJ, Gómez-López G, Al-Shahrour F. In silico drug prescription for targeting cancer patient heterogeneity and prediction of clinical outcome. Cancers. 2019;11(9) Available from: https://doi.org/10.3390/cancers11091361.

44. Rosenthal R, Cadieux EL, Salgado R, Bakir MA, Moore DA, Hiley CT, et al. Neoantigen-directed immune escape in lung cancer evolution. Nature. 2019;567(7749):479–85 Available from: https://doi.org/10.1038/s41586-019-1032-7.

45. Wang Q, Guldner IH, Golomb SM, Sun L, Harris JA, Lu X, et al. Single-cell profiling guided combinatorial immunotherapy for fast-evolving CDK4/6 inhibitor-resistant HER2-positive breast cancer. Nat Commun. 2019;10(1):3817 Available from: https://doi.org/10.1038/s41467-019-11729-1.

46. Lee HW, Chung W, Lee H-O, Jeong DE, Jo A, Lim JE, et al. Single-cell RNA sequencing reveals the tumor microenvironment and facilitates strategic choices to circumvent treatment failure in a chemorefractory bladder cancer patient. Genome Med. 2020;12(1):47 Available from: https://doi.org/10.1186/s13073-020-00741-6.

47. Troulé K, López-Fernández H, García-Martín S, Reboiro-Jato M, Carretero-Puche C, Martorell-Marugán J, et al. DREIMT: a drug repositioning database and prioritization tool for immunomodulation. Bioinformatics. 2021;37(4):578–9 Available from: https://doi.org/10.1093/bioinformatics/btaa727.

48. Quinn JJ, Jones MG, Okimoto RA, Nanjo S, Chan MM, Yosef N, et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. Science. 2021;371(6532) Available from: https://doi.org/10.1126/science.abc1944.

49. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial single cell analysis. bioRxiv. 2021 [cited 2021 Nov 2]. p. 2021.02.19.431994. Available from: https://www.biorxiv.org/content/10.1101/2021.02.19.431994v2

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.