**Title: Single-Cell Integrated Analysis Reveals Comprehensive Pathway and Network in Enzalutamide (ENZ)-Resistant Prostate Cancer and Possible Combinatorial Drug Therapeutic Strategies**

**LEUNG, Chak On**

**Abstract:**

**Background:**

Androgen-independent tumor progression complicates the chemotherapy drug therapy due to tumor resistance against Enzalutamide (ENZ). A thorough understanding in molecular network is essential for new drug discovery for better clinical outcomes.

**Results:**

Herein, I employed different pathway enrichment and network analysis methods integrated with single-cell transcriptomic integrated analysis to reveal the ENZ-resistance mechanism. The results showed the development of resistance not related to epithelial regeneration and cell cycle, and enriched network from JUN and its related genes was significantly associated with pathways favouring cancer survival, such as autophagy and detoxification against specific drugs. I also applied new machine model to predict different multi-target drugs for the new direction of chemotherapy.

**Conclusions:**

Ultimately, this project showed the applications of single-cell technologies integrated with molecular system analysis for the potential improvement of clinical therapeutic decision making.

**Keywords:**

ENZ-resistance prostate cancer, Jun, pathway enrichment, network analysis, scTherapy, AR-independent pathway and network

**Introduction:**

Androgen deprivation therapy (ADT) has been one of the primary drug treatments of metastatic prostate cancer (PC) expressing androgen receptors (AR). However, the progression on ADT, such as the second-generation AR antagonist enzalutamide (ENZ), leads to ENZ-resistance prostate cancer. Several proposed mechanisms have been annotated to the transformation of its lineage on drug resistance, including noncanonical AR addiction via CXXC5 and its regulated genes which led to an emergence of AR-independent tumor progression [1-2].

Single-cell technologies have been utilized to identify driver genes from the heterogeneous population for drug resistance. However, most studies to date have focused on gene expression or gene mutations form single-cell transcriptomics without making allowance for pathway enrichment and gene regulatory network for the comprehensive overview of drug resistance [3]. Because the transcriptome relates to vast interaction among different genes, including transcription regulators, the gene expression profile may not reveal the origin of drug resistance. Therefore, the mechanism of the lineage transformation related to ENZ resistance is not well-understood.

Drug prediction has witnessed a rapid growth and a number of innovations via the rapid discovery of oncologic drivers and transcription factors obtained from single cell transcriptome. A majority of machine learning algorithms predicted monotherapy with a high accuracy and precision [4], while there has been lack of development on multiple drug prediction with high cancer-selectivity. This is due to the limited computation power for all drug permutations and hard-to-predict drug synergy [5] accounting for different drug sensitivities on targeted and normal cells, leading to low efficacy and potential safety issues. Recently, it has been possible to use monotherapy-specific algorithm for multiple drug predictions via identifying cancer-selective and low-toxic multi-targeting options based on the single cell transcriptome data [6,7].

In this project, I analyse the emergence of resistance of PC with the exposure of ENZ at the single-cell level via the single-cell transcriptome. Through pathway and network analysis, the enriched transcription networks and potential driver genes related to ENZ-resistance would be identified. The drug prediction based on the target genes would be implemented via scTherapy [7] which originate from the gene expression differences between the target cells and controlled cells, and the result would be a drug list for further validation.

**Results:**

**ENZ-resistance was developed after one-week exposure of ENZ**

To study the mechanisms for the ENZ-resistance, I utilized different LNCaP cell lines with the raw sequencing data downloaded from Gene Expression Omnibus (GEO) archive under accessions GSE168669 and GSE168733 respectively [8]. The cell lines included LNCaP parental cell lines (LNCaP-DMSO), and 4 LNCaP cell lines with exposure of ENZ for 48 hours (LNCaP-ENZ48), for 168 hours (LNCaP-ENZ168), for 9.5 months (LNCaP-RESA) and for 13.5 months (LNCaP-RESB) respectively (Fig 1a,b). The hypothesis from this study was that the ENZ-resistance would be linked with AR-independent mechanism [9].

To explore the rise of ENZ-resistance, I performed the single-cell transcriptome integration analysis for all LNCaP-cell lines. A total of 16 clusters (0-15) were divided into 4 major cell groups based on the population data with 3 “Initial” clusters, 1 “ENZ168” cluster, 2 “Resistance” clusters and 10 “Persistence” clusters (Fig 1c). To investigate the lineage transformation process, the pseudo-time analysis was implemented using Monocle3 [10] with the validation from RNA velocity analysis [11]. Note that “Initial” cell group gene expression activity was reduced through time and the gene expression “ENZ168” cell group shifted to resistance clusters. Both results highlighted the cluster 0 as the transient stage for ENZ-resistance (Fig 1d,e). The 5 cell lines was queried with 4 LNCaP cell lines (LNCaP-DMSO, LNCaP-ENZ48, LNCaP-RESA, LNCaP-RESB) to investigate the closedness between LNCaP-ENZ168 and other cell lines by transferring scRNA labels. The ENZ168 cluster was similar to 3 clusters of the 4 LNCaP cell lines containing higher proportion of RESA and RESB cells (Fig 1f,g), suggesting that the one-week ENZ treatment was sufficiently induced the ENZ-resistance clusters.

a. A colorful dots in a circle

AI-generated content may be incorrect. b. A group of colorful dots

AI-generated content may be incorrect.

c. A graph of different colored bars

AI-generated content may be incorrect.d. A colorful dots on a white background

AI-generated content may be incorrect.

e.

A graph showing different colored spots

AI-generated content may be incorrect. A diagram of a horse

AI-generated content may be incorrect.

A colorful dots and lines

AI-generated content may be incorrect. A colorful drawing of a brain

AI-generated content may be incorrect. A diagram of a colorful flow

AI-generated content may be incorrect.

f. A screenshot of a graph

AI-generated content may be incorrect.

g.A graph of data on a white background

AI-generated content may be incorrect.h. A colorful cluster of dots

AI-generated content may be incorrect.

**Fig 1. Characteristics of 5 LNCaP cell lines after single-cell transcriptomic analysis:** a,b) Uniform manifold approximation and projection (UMAP) clustering visualization of a) integrated single-cell transcriptomics of LNCaP parental, LNCaP–ENZ48, LNCaP-ENZ168, RES-A, and RES-B, and b) single-cell transcriptomics for each cell lines; c) Proportions of cells in clusters. Clusters were colored according to cluster type: Initial (most prevalent in LNCaP parental and LNCaP–ENZ48), ENZ168 (most prevalent in ENZ168) ,Resistance(most prevalent in RES-A or RES-B), or Persistence (present in similar proportions in all samples); d) Pseudo-time analysis for 5 cell lines. Referring to S. Taavitsainen et al. that the Initial group as the starting point for cancer evolution [12], the cancer cell line evolved in the direction of: Initial, ENZ168, Resistance, and Persistent; e) RNA velocities curves on 5 cell lines depicted as streamlines; f) The matching of clusters of 5 cell lines against the 4 cell lines (LNCaP-DMSO, LNCaP-ENZ48, LNCaP-RESA and LNCaP-RESB) shown in the heatmap, where the color of the cells represented the proportion of the clusters of 5 cell lines matched with the clusters of the 4 cell lines; g,h) Basic information of 4 cell lines: g) The cell distribution and h) UMAP of 4 cell lines.

To investigate the relation between ENZ-resistance and genes with regenerative and proliferative potential, the cell cycle scores were compared to different cell groups. I found that the ENZ168 cell group had statistically significantly lower scores compared to the Persistence group in terms of S score (-0.058 vs -0.039, p-val: 2.34e-104) and G2M score (-0.126 vs -0.0738, p-val: 0.00) respectively (Fig 2a,b). This was further validated by the UMAP plot with S and G2M scores, where cluster 0 has significantly lower S/G2M genes compared with clusters of Persistence group (Fig 2c). Because genes for epithelial regeneration were significantly expressed in castration-resistant PC, I compare the gene expression among clusters with the reference to the gene list from Karthaus et al. [12] with log scores based on the percentage of expression and average expression to learn whether these genes related to ENZ resistance (See Methods). Cluster 0 overall had one of the lowest scores compared with the other clusters, suggesting that the ENZ-resistance mechanism was not related to epithelial regeneration (Fig 2d,e). The cell potency assessment from CytoTRACE2 which showed only 3 Persistence cell group (clusters 8,12,13) being the only pluripotent group while others being unipotent (Fig 2f).

a. A graph of cell groups

AI-generated content may be incorrect.b. A diagram of a cell group

AI-generated content may be incorrect.

c. A diagram of a plane

AI-generated content may be incorrect.

d. A chart of different colored dots

AI-generated content may be incorrect.

e. A graph of a bar chart

AI-generated content may be incorrect.

f. A graph of different colored bars

AI-generated content may be incorrect. A graph with different colored bars

AI-generated content may be incorrect. A graph with colorful bars

AI-generated content may be incorrect.

A graph of different colored bars

AI-generated content may be incorrect. A graph of different colored bars

AI-generated content may be incorrect.

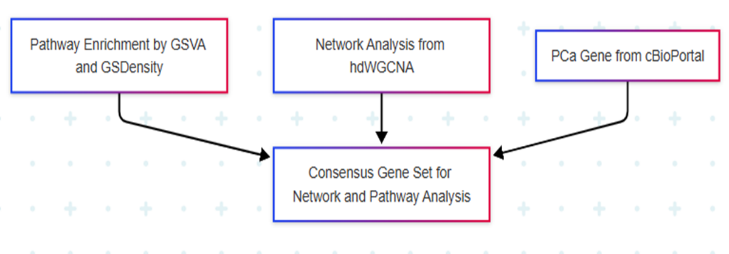
**Fig 2.** **Regenerative potential and cell potency of 5 cell lines:** a, b) a) S-score and b) G2M-score comparison among cell groups with the pairwise Wilcoxon two-sample rank-sum test; c) UMAP project with S-score and G2M-score distribution; d) Dotplot representation for the gene expression for all clusters with reference to the gene list for epithelial regeneration from Karthaus et al. [12]; e) Comparison of Log-scores in terms of average expression and expression percentage among cell clusters; f) Cell potency test across all clusters for all 5 cell lines: (From top to bottom, left to right): LNCaP-DMSO, LNCaP-ENZ48, LNCaP-ENZ168, LNCaP-RESA and LNCaP-RESB

**Pathway enrichment and network analysis showed JUN-regulated network for AR-independent mechanisms of ENZ resistance**

To explore the enriched network on ENZ-resistance, a total of 14 transcription factors (TFs) were collected from the transcription factor (TF) consensus from pathway enrichment, network analysis and prostate adenocarcinoma genes from the cBioPortal (Fig 3a).

From the pathway enrichment analysis relating to 14 transcription factors, the cluster 0 had similar pathway enrichment pattern compared to cluster 5 and 6 with the reference to Hallmark and Reactome pathways which were statistically significant (Fig 3b, See Methods). S. Taavitsainen et al. suggested that AR-independent pathways, including MYC targeting and chromatin remodelling [12], were significantly upregulated during the ENZ-resistance development. Interestingly, all mentioned pathways were downregulated in the ENZ168 clusters, highlighting different TF-related pathways may be positively enriched which caused ENZ-resistance.

Out of 7 co-expression modules, only the turquoise module significantly covers cluster 0 (p-val: < 2.2e-16) containing 7 TFs enriched in cluster 0 with JUN being the most enriched (Fig 3b,c, Table 1a). Interestingly, cluster 15, a cluster in persistence group, had a higher JUN expression than cluster 0 (Fig 3d). Yet, the network analysis only covered the JUN associated genes enriched in ENZ168 group. Furthermore, JUN regulated the expression of other TFs, including 4 (out of 6) TFs in ENZ168 clusters, and TFs from other clusters such as REST, STAT3 and MXI1 (highlighted in green circles) (Fig 3e, Table 1a), suggesting JUN a crucial regulator for different cell activities. Including JUN, a total of 892 genes targeted by the JUN transcription factor network were utilized in the Protein-Protein Interaction (PPI) Network with reference to the STRINGDB for the protein-protein interaction confidence measurement with the minimum of 0.4 (Fig 2f, Supplementary File 1). The result highlighted the potential transcription regulator role of JUN with the degree of 78 and neighbourhood connectivity of 29 (Table 1b). Out of the top 5 genes with the highest co-expression score with JUN, 2 of which, FOS and MAP1B, were the first neighbours of JUN with different connectivity score (0.999 and 0.454 respectively) (Fig 3g,h, Supplementary File 1).

a. 

b. A screenshot of a computer screen

AI-generated content may be incorrect.

c. A group of colored lines

AI-generated content may be incorrect.

d. A group of images of numbers and letters

AI-generated content may be incorrect.A graph of different colored squares

AI-generated content may be incorrect.

e. A close-up of a network

AI-generated content may be incorrect.

f. A colorful dots and lines

AI-generated content may be incorrect.

g. A graph with colored dots

AI-generated content may be incorrect.h. A network of dots and lines

AI-generated content may be incorrect.

**Fig 3. Pathway enrichment and network analysis on ENZ168 cell group:** a) Transcription factors set collected from the cis-regulatory genes consensus in GSVA, GSDensity pathway enrichment analysis, hdWGCNA network analysis and PC genes from cBioPortal; b) 42 enriched pathways related to 14 TFs from the consensus of pathway enrichment and network analysis; c) 7 co-expression modules on the UMAP, with turquoise module specifically covering the most of cluster 0; d) Expression distribution across all 14 TFs with the JUN differential expression comparison across all clusters; e) The complete overview of TF networks with TFs interaction across different modules; f) Network of JUN and its 892 genes enriched in ENZ168 cell group with STRINGDB confident score of 0.4; g) Top 5 positively and negatively correlated target genes co-expressed with JUN; h) First neighbourhood of JUN with 2 of 5 most positively correlated target genes (highlighted in red)

a. A screenshot of a computer

AI-generated content may be incorrect.

b.

A screenshot of a computer

AI-generated content may be incorrect.

**Table 1. Basic information of JUN-related network:** a) JUN-associated genes’ differential expression and its relation with JUN; b) Network parameters for JUN network

I proceeded to find enriched pathway based on the network construction via the package enrichr, where 42 positively enriched and 39 negatively enriched GO biological pathways (GOBP) were found (adjusted p.val <0.05, Supplementary File 2). The top 5 positively regulated pathways included positive protein deubiquitylation regulation, endosomal transport and inflammatory signalling pathway, while the top 5 downregulated pathways involved in mitochondrial electron transport, ribosomal and protein synthesis (Fig 4a). The pathways were divided into 7 pathway groups in the heatmap (Fig 4b, Supplementary File 2), showing 2 downregulated groups (Mitochondrial and protein targeting group), 2 upregulated groups (Protein modification and regulation) and 3 balanced groups (Assembly, Biosynthetic pathway and Others). To further investigating the pathway interactions for ENZ-resistance, a total of 571 significantly expressed genes related to the pathways were for the PPI network construction with STRING confident score of 0.8 or above with the account of their expression. The negatively regulated pathway filled with downregulated genes, while the positively regulated pathway contained both upregulated and downregulated genes (Fig 4b, c). 3 groups of subnetworks showed different hub interactions, with the common insight that ENZ-resistance associated with dysregulation of protein ubiquitination, upregulation of inflammatory response, inhibition of mitochondrial activity and ribosomal proteins assembly (Fig 4d).

a. A graph of a bar graph

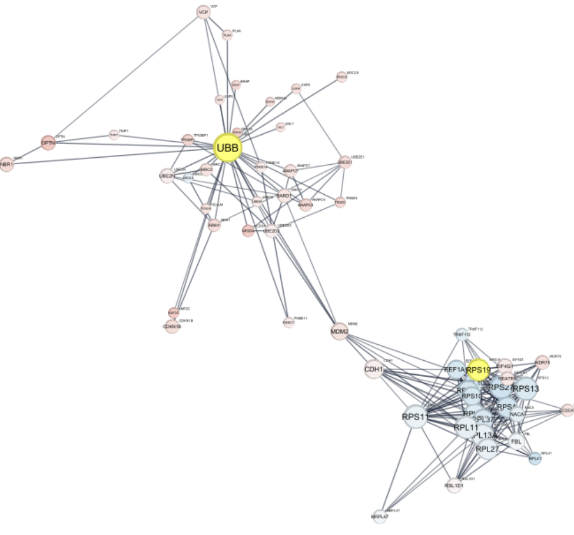
AI-generated content may be incorrect.

b. A screen shot of a computer screen

AI-generated content may be incorrect.

c. A network of connected lines and dots

AI-generated content may be incorrect.

d. A network of dots and lines

AI-generated content may be incorrect.

A network of connected lines and dots

AI-generated content may be incorrect.

**Fig 4. Pathway enrichment based on network genes**: a) Top 5 positively and negatively enriched pathways (adjusted p value<0.05); b) Relation between differential expressed cluster 0 genes and the enriched pathway; c) Major network consisting of 7 pathway groups; d) 3 groups of network associated with cancers: 1) Upregulation of Protein modification (mainly ubiquitnation) – Downregulation of protein targeting and ribosomal pathway; 2) Upregulation of protein for autophagy – Downregulation of mitochondrial activity; 3) Upregulation of autophagy and pro-inflammatory interleukin – Upregulation of Pre-mRNA splicosome assembly.

**scTherapy predicts multiple drugs for therapeutic intervention**

To investigate the target drugs for JUN related ENZ-resistance, around 250 genes in JUN network with expression larger than 1 or smaller than -1 were fed to the model (See Methods). The model outputted 6 drugs (out of 15 drugs) specifically targeted for moderate coverage of genes, including MAP1B and PTEN (Table 2). The drug list would be subjected to evaluation and validation in the future.

|  |  |  |  |
| --- | --- | --- | --- |
| **Drug name** | **Major Target genes** | **Target enriched pathway of JUN?** | **Drug Dosage (nM)** |
| GDC-0941 | PI3K | Yes, Upregulation of RPS genes | 258.7844 |
| Alpelisib | PI3K | Yes, Upregulation of RPS genes | 1000 |
| Tivantinib | PTPN | Yes, Inhibition of PTPN 11-SHP2 signalling pathway, involving PIK3R1, PTPN6 | 320.251 |
| Parbendazole | MAP1B | No. Specific for MAP1B tubulin polymerization | 313.1875 |
| Nocodazole | MAP1B | No. Specific for MAP1B tubulin polymerization | 946 |
| RITA | MDM2 | Yes. Target MDM2 p53 transactivation domain-binding cleft for inhibition | 961.742 |

**Table 2. Drug list for ENZ168 cluster from scTherapy**

**Discussion:**

In this project, I filled the research gap regarding to the AR-independent ENZ-resistance with the integration of pathway enrichment and network analysis. Several researchers, including S. Taavitsainen [8], Karthaus [12] and their colleagues, found genes with strong regeneration or chromatin remodelling capabilities increasing the proliferative activity and invasiveness of PC counteracting the drug-induced apoptosis. However, the results from the project suggested otherwise, where the development of ENZ-resistance cancer cells had poor cell potency from the Potency and Cell Cycle Scores, as well as the downregulated chromatin remodelling pathway. Previous studies explained the origin of drug resistance could be pathways not connected with proliferation potential. Luo et al. and Valerine et al. suggested the JAK2/STAT1-Dependent Pathway associating with inhibition of apoptosis via the pro-inflammatory cytokine upregulation [13, 14]. Mou et al. recently discovered hypoxia-induced drug resistance during the angiogenesis of breast cancer [15]. Furthermore, specific timeframes after the drug ingestion could observe different pathways for drug resistance. Although S. Taavitsainen et al. utilized the longitudinal studies on tumour drug resistance, there was a significant time gap between Initial cell group and resistance cell group, potentially missing different pathways for tumor progression. Here I included LNCaP-ENZ168 as a transition point between these cell groups in the single-cell analysis and I found different enriched networks for the drug resistance.

The usage of 3 distinct pathway and network analysis tools was to produce a consensus enriched transcription factors gene set for high accuracy and precision. I successfully found JUN to be the major transcription factor and different enriched JUN regulatory networks regulating a majority of cellular events related to tumor progression with the context of gene expression levels, with some networks being overrepresented in different cancer types. For example, in the MDM2-RPS network, the downregulated RPS induced the loss of binding to MDM2, an oncogenic gene suppressing p53-induced apoptosis [16, 17]. The inverse gene expression between cellular component assembly and mitochondrial activity highlighted the increase in ATG5/GABARAPL2-induced autophagy related to ER-stress via the downregulation of CYCS [18], resulting in the removal of cytotoxic effects of the drug. The activation of PI3K pathway via the upregulation of PIKC3 and PIK3R1, both of which were hotspot for breast cancer [19]. Upregulation of PTPN11 involved in PTPN11-SH2 signalling pathway via PI3KR1/PTPN6 and related to reduced Androgen receptor expression via the downregulation of ribosomal proteins assembly genes [20].

I showed that multiple drugs could be utilized to target genes responsible to AR-independent tumor progression. Without providing the longitudinal cancer gene expression data induced from different inhibitors as the input, the drug prediction was not possible. Until recently, Aleksandr et al. claimed the scTherapy success in drug prediction with high specificity and sensitivity simply with comparing the differential gene expression of the targeted and controlled regions, which predicted drug combinatorics for leukaemia with 88% specificity [7]. Using scTherapy, I successfully generated potential drugs targeting multiple genes against the ENZ-resistance development. Notably, some of the drugs predicted by the model are not restricted to a single cancer type by targeting common pathway from different cancers. For example, tivantinib, a common drug for Juvenile myelomonocytic leukaemia (JMML), targeted PTEN11-SHP2 pathway, which was commonly found in the AR-independent drug resistance pathway [21, 22].

**Limitations and future directions:**

Although this study presented plausible networks and possible drugs for treatment, there were several limitations, including the unexplained casualty among interactive genes for which co-expression network could not account and limited data for further validation on network and potential treatment.

To address the above issues, different approaches would be utilized to complete the studies. The single cell ATAC data for all LNCaP cell lines would be for the direct network analysis via DIRECT-NET, a sophisticated machine learning model for predicting cis-regulatory network from multiomics data. Furthermore, I planned to collect different patient PC cell lines with ENZ-resistance to validate the drug performance with given dosage.

**Methods**

**Algorithms for the drug prediction**

The drug prediction was done via the gradient boosting model called LightGBM model. To train the model, the data containing the change in expression in the single-cell transcriptomic dataset given a small molecule perturbation (LINC2020 dataset), and the drug structures as ECFP fingerprints together with drug-dose response data were integrated with the transcriptomic dataset. After training, the gene expression of the experimental single-cell data would be inputted to the model to query against the database, which would produce the drug list suggesting the prior drugs with dosage for the cure [7].

**Cell line collection and single-cell transcriptomic data [8]**

The LNCaP cell lines were obtained from American Type Culture Collection (ATCC; LGC Standards) and tested for mycoplasma contamination each month. 5 LNCaP cell lines from different culture medium were included in the study. For the first three cell lines, ~1 × 106 cells were cultured in DMSO (LNCaP-DMSO) or 10 μM ENZ medium for 48 hrs (LNCaP-ENZ48) and 168 hrs (LNCaP-ENZ168) respectively. RESA and RESB cell lines were obtained by prolong exposure to ENZ with increasing concentrations for 9.5 months and 13.5 months respectively.

Cells from all cell lines except LNCaP-ENZ168 were harvested with 0.05% Trypsin-EDTA (Sigma T3924), and subjected to centrifugation (300 × g for 5min) and resuspension PBS/0.5% BSA, before being sorted by FACS. Around ~3 × 104 cells were further processed via the Chromium Single Cell 3′ Library, Gel Bead & Multiplex Kit, and Chip Kit (v3, 10x Genomics). LNCaP-ENZ168 single-cell population was obtained with Drop-seq using the Dolomite cell encapsulation system (Dolomite Bio), where the cell lines were trypsinized with TrypLETM Express Enzyme (ThermoFisher Scientific, #12604021), spun down (5 min at 300 × g), and washed with 0.1% BSA–PBS. 3 × 105 cells/mL single-cell suspension was obtained via single-cell encapsulation with beads in lysis buffer and oil. Droplets were separated by centrifugation and tagmentation of cDNA was performed with the Nextera XT DNA Library Preparation Kit (Illumina, #FC-131-1024).

The raw FASTQ files from non ENZ168 cell lines were processed into single-cell feature counts via Cell Ranger v3.0.2 with reference to GRCh38 reference from Cell Ranger webpage, while the ENZ168 FASTA file were preprocessed, aligned, and processed to cell count matrices with the Drop-seq tools v2.3.0 (https://github.com/broadinstitute/Drop-seq/tree/master) using default parameters and NUM\_CORE\_BARCODES to be 1000. Note that the preprocessing procedure involved the unaligned BAM file by removing gaps of the ENZ168 FASTA file, which was aligned to GRCh38 via the STAR aligner v2.7.3a, and the annotation file was referred to Gencode annotations version 33.

**Single-cell RNA quality control, clustering and integration**

The Cell Ranger and Dropseq output were subjected to Seurat v5.2.0 for further analysis of the scRNA-seq samples. The quality cells were selected based on the number of detected genes, the total number of UMI, and the percentage of genes from mitochondria. All filtering threshold and the pre- and post-quality control result were mentioned in Supplementary File 3. Multiple single-cell RNA-seq datasets were merged into one object using the “merge” function and subjected to integration via “IntegrateData” for anchor number computation and “FindIntegrationAnchors” function for batch effect correction. The clustering was done using default parameters of uniform manifold approximation projection (UMAP) nonlinear dimensionality reduction and “FindNeighbours”, and “FindClusters” with the resolution of 1.3. The marker

genes of each cluster and between samples were calculated using the “wilcox” (Wilcoxon Rank Sum test). A gene would be considered differentially expressed if the with Bonferroni-corrected p-value < 0.01 and at least 1% cells expressing the gene. The integrated object was subjected to CellCycleScoring where the S and G2/M scores were applied to each cell of the object based on the proportion of S and G2/M genes.

**Regenerative power and cell potency test for different clusters**

The regenerative strength of cells from different clusters was examined based on the regenerative list from Karthaus et al [12]. The genes were ranked based on average expression and the percentage of expressing cells. Since the study from Karthaus et al. did not arrange the genes in terms of relatedness to epithelial regeneration, an assumption that all genes equally weighed to the epithelial regeneration was made. The rank of the clusters for each gene was directly related to gene expression, and the scores for each cluster would be calculated as follows:

For each cluster, the average scores with respect to the average expression and expression percentage would be subjected to natural log transformations, and the clusters were ranked via the addition of the scores.

The cell differential potential for each cluster were predicted using cytoTRACE 2 and respective RNA velocities was examined with scVelo v0.3.3.

**Individual pathway and network algorithms**

A total of 3 pathway enrichment and 1 network algorithm to the gene expression data. The packages for pathway enrichment included the following methods: GSVA, GSDensity and GO enrichment from hdWGCNA package. Table 3 shows pathway databases and interaction networks used by each method with different parameters and statistical significance methods [23-28]. All parameters for the analysis were default. Note that the pathway enrichment from hdWGCNA package could bring statistically significant positively and negatively enriched network-based pathways in case of GSVA and GSDensity failed to present either type of pathways.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Packages | Methodology | Reference Database | Used for | Note |
| GSVA | Non-parametric, unsupervised method to assess pathway variation | HALLMARK and REACTOME from MSigDB | Pathway enrichment | NA |
| GSDensity | Multiple correspondence analysis (MCA) to cluster cells with similar gene activity, and assigned clusters with different enriched pathways with statistical significance, where p -value was calculated with based on the distribution of kld (log-transformed KL-divergence) | HALLMARK and REACTOME from MSigDB | Pathway enrichment | Only the pathway significantly enriched (adjust p-value <0.05) in GSDensity would match with GSVA for further pathway enrichment analysis. |
| GO enrichment from hdWGCNA | Pathway enrichment analysis based on consisting of TF network measured in enrichment log score | Gene Ontology (GO), including GOBP, GOMF, GOCC | Pathway enrichment | Used when either one of the above packages only showed positive or negative pathway results |
| Weighted gene co-expression network analysis (WGCNA) | Score gene-gene interaction based on pairwise gene expression correlation | STRINGDB (For network construction) | Network analysis | NA |

**Table 3. Specific details for Pathway enrichment and Network Analysis**

**Drug prediction using scTherapy**

The clusters containing transformative genes for ENZ-resistance would be subjected for drug prediction. Their gene expression data would be split into two groups, both of which were differentially expressed: The target cell group containing genes with average log fold change >1, and the control cell group containing with the average log fold change <-1. The data served as an input to the pre-trained LightGBM model, and a drug list would be predicted from scTherapy which would be subjected to further validation.

**Statistical test:**

All statistical tests were done in R ver 4.4.1. Unless otherwise specified, all pairwise tests were done by pairwise Wilcoxon rank-sum test. The pathway enrichment from hdWGCNA used Benjamini-Hochberg multiple testing correction for the Fisher’s exact test p-values.

**Data availability:**

The single-cell RNA-seq transcriptomic data for all LNCaP-DMSO, LNCaP-ENZ48, LNCaP-ENZ168, LNCaP-RESA and LNCaP-RESB were retrieved from in the Gene Expression Omnibus (GEO) archive under accessions GSE168669 and GSE168733 respectively.

**Abbreviation:**

LNCaP: Lymph Node Carcinoma of the Prostate; ENZ: Enzalutamide; hdWGCNA: High-dimensional weighted gene co-expression network analysis; JUN: Jun Proto-Oncogene, AP-1 Transcription Factor Subunit; MAP1B: Microtubule associated protein 1B; FOS: Fos Proto-Oncogene, AP-1 Transcription Factor Subunit; PI3K: Phosphoinositide 3-kinases; MDM2: Mouse double minute 2 homolog; RP: Ribosomal Protein; GSVA: Gene Set Variation Analysis; PTPN11: Tyrosine-protein phosphatase non-receptor type 11; SH2: Src Homology 2.

**Authors' contributions:**

Leung Chak On

**Acknowledgement:**

I want to thank Prof Stephen Tsui, my supervisor for the project, for suggested improvements on specific parts of the project. I also want to give thanks to different online resources helping me on downstream analysis.

**References:**

1. Lupien, Mathieu et al., FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription, Cell, Volume 132, Issue 6, 958 – 970

2. He, Y., Wei, T., Ye, Z. et al. A noncanonical AR addiction drives enzalutamide resistance in prostate cancer. Nat Commun 12, 1521 (2021). <https://doi.org/10.1038/s41467-021-21860-7>

3. Iacono, G., Massoni-Badosa, R. & Heyn, H. Single-cell transcriptomics unveils gene regulatory network plasticity. Genome Biol 20, 110 (2019). <https://doi.org/10.1186/s13059-019-1713-4>

4. Farzaneh Firoozbakht, Behnam Yousefi, Benno Schwikowski, An overview of machine learning methods for monotherapy drug response prediction, *Briefings in Bioinformatics*, Volume 23, Issue 1, January 2022, bbab408, <https://doi.org/10.1093/bib/bbab408>

5. Pan Y, Ren H, Lan L, Li Y, Huang T. Review of Predicting Synergistic Drug Combinations. Life (Basel). 2023 Sep 7;13(9):1878. doi: 10.3390/life13091878. PMID: 37763281; PMCID: PMC10533134.

6. Cheng, F., Kovács, I.A. & Barabási, AL. Network-based prediction of drug combinations. *Nat Commun* **10**, 1197 (2019). <https://doi.org/10.1038/s41467-019-09186-x>

7. Ianevski, A., Nader, K., Driva, K. *et al.* Single-cell transcriptomes identify patient-tailored therapies for selective co-inhibition of cancer clones. *Nat Commun* **15**, 8579 (2024). <https://doi.org/10.1038/s41467-024-52980-5>

8. Taavitsainen, S., Engedal, N., Cao, S. et al. Single-cell ATAC and RNA sequencing reveal pre-existing and persistent cells associated with prostate cancer relapse. Nat Commun 12, 5307 (2021). https://doi.org/10.1038/s41467-021-25624-1

9. Watson, P., Arora, V. & Sawyers, C. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat Rev Cancer* **15**, 701–711 (2015). <https://doi.org/10.1038/nrc4016>

10. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014). “The dynamics and regulators of cell fate decisions are revealed by pseudo-temporal ordering of single cells.” Nature Biotechnology.

11. La Manno, G., Soldatov, R., Zeisel, A. et al. RNA velocity of single cells. Nature 560, 494–498 (2018). <https://doi.org/10.1038/s41586-018-0414-6>

12. Wouter R. Karthaus et al. ,Regenerative potential of prostate luminal cells revealed by single-cell analysis.Science368,497-505(2020).DOI:10.1126/science.aay0267

13. Luo Y, Yang X, Basourakos SP, Zuo X, Wei D, Zhao J, Li M, Li Q, Feng T, Guo P and Jiang Y (2021) Enzalutamide-Resistant Progression of Castration-Resistant Prostate Cancer Is Driven via the JAK2/STAT1-Dependent Pathway. *Front. Mol. Biosci.* 8:652443. doi: 10.3389/fmolb.2021.652443

14. Valerie Sloane Jones, Ren-Yu Huang, Li-Pai Chen, Zhe-Sheng Chen, Liwu Fu, Ruo-Pan Huang, Cytokines in cancer drug resistance: Cues to new therapeutic strategies, Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, <https://doi.org/10.1016/j.bbcan.2016.03.005>.

15. Jiancheng Mou, Chenhong Li, Qinghui Zheng et al., Research progress in tumor angiogenesis and drug resistance in breast cancer, Cancer Biology & Medicine Jun 2024, 20230515; DOI: 10.20892/j.issn.2095-3941.2023.0515

16. Zhang, X., Wang, W., Wang, H. et al. Identification of ribosomal protein S25 (RPS25)–MDM2–p53 regulatory feedback loop. Oncogene 32, 2782–2791 (2013). <https://doi.org/10.1038/onc.2012.289>

17. Liu, Yong et al., RP–MDM2–p53 Pathway: Linking Ribosomal Biogenesis and Tumor Surveillance, Trends in Cancer, Volume 2, Issue 4, 191 – 204

18. Deegan, S. et al. (2014) ‘Deficiency in the mitochondrial apoptotic pathway reveals the toxic potential of autophagy under ER stress conditions’, Autophagy, 10(11), pp. 1921–1936. doi: 10.4161/15548627.2014.981790.

19. Tharin, Z., Richard, C., Derangère, V. et al. PIK3CA and PIK3R1 tumor mutational landscape in a pan-cancer patient cohort and its association with pathway activation and treatment efficacy. Sci Rep 13, 4467 (2023). <https://doi.org/10.1038/s41598-023-31593-w>

20. AMA Style

Richards CE, Elamin YY, Carr A, Gately K, Rafee S, Cremona M, Hanrahan E, Smyth R, Ryan D, Morgan RK, et al. Protein Tyrosine Phosphatase Non-Receptor 11 (PTPN11/Shp2) as a Driver Oncogene and a Novel Therapeutic Target in Non-Small Cell Lung Cancer (NSCLC). International Journal of Molecular Sciences. 2023; 24(13):10545. <https://doi.org/10.3390/ijms241310545>

21. Michieli, P., Di Nicolantonio, F. Tivantinib—a cytotoxic drug in MET inhibitor's clothes? Nat Rev Clin Oncol 10, 372–374 (2013). https://doi.org/10.1038/nrclinonc.2013.86

22. Zhao, Y., He, C., Zhang, D. *et al.* Leukemogenic SHP2 mutations lead to erythropoietin independency of HCD-57 cells: a novel model for preclinical research of SHP2-mutant JMML. *Exp Hematol Oncol* **12**, 20 (2023). <https://doi.org/10.1186/s40164-023-00379-1>

23. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics 14, 7 (2013). <https://doi.org/10.1186/1471-2105-14-7>

24. Liang, Q., Huang, Y., He, S. *et al.* Pathway centric analysis for single-cell RNA-seq and spatial transcriptomics data with GSDensity. *Nat Commun* **14**, 8416 (2023). <https://doi.org/10.1038/s41467-023-44206-x>

25. Morabito, Samuel et al., hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data, Cell Reports Methods, Volume 3, Issue 6, 100498

26. Liberzon, Arthur et al., The Molecular Signatures Database Hallmark Gene Set Collection, Cell Systems, Volume 1, Issue 6, 417 – 425

27. Marc et al., The reactome pathway knowledgebase 2022, Nucleic Acids Research, Volume 50, Issue D1, 7 January 2022, Pages D687–D692, https://doi.org/10.1093/nar/gkab1028

28. Childs, Jessica E. et al. Relapse to cocaine seeking is regulated by medial habenula NR4A2/NURR1 in mice, Cell Reports, Volume 43, Issue 3, 113956