Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated single-cell transcriptomics reveal E2F1-driven multidrug resistance in medulloblastoma and identify Palbociclib as a therapeutic candidate

Yueliang Yao
Fuzhou Medical College of Nanchang University
Junying Zhang
Chongqing Medical and Pharmaceutical College
Ping Liang
Children's Hospital of Chongqing Medical University
Hang Ji
jihang0723@163.com

West China Hospital Sichuan University

Research Article

Keywords: Medulloblastoma, chemoresistance, E2F1, palbociclib, intratumoral heterogeneity

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11043118/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

Medulloblastoma (MB) is a malignant pediatric brain tumor where tumor recurrence driven by acquired multidrug resistance remains a paramount clinical challenge. Resolving the intratumoral heterogeneity underlying this resistance is urgently required to identify novel therapeutic targets.

Methods

Single-cell RNA sequencing data and large-scale bulk transcriptomes were integrated and comprehensively analyzed. Multidrug-resistant subpopulations were identified by mapping ridge regression-based drug sensitivity predictions to single-cell profiles. Developmental trajectories and core regulatory networks were reconstructed, and essential transcription factors were functionally evaluated by computational vector field perturbations. The prognostic value of specific cellular states was assessed, and potential therapeutic antagonists were screened using computational databases and spatial molecular docking.

Results

Two distinct multidrug-resistant reservoirs were identified, an *E2F1*-driven DNA repair state exhibiting cross-resistance to genotoxic agents, and a *ZIC5*-regulated neuronal differentiation state associated with Hedgehog signaling. Trajectory inference and in silico perturbations demonstrated that the continuous accumulation of *E2F1* is critical for maintaining the chemoresistant DNA repair phenotype. Clinical deconvolution revealed that elevated *E2F1* expression acts as a robust indicator of poor overall survival. Subsequently, computational pharmacogenomic screening and molecular docking identified palbociclib as a highly potent inhibitor capable of precisely binding and targeting the CDK4/6-E2F1 axis.

Conclusion

An *E2F1*-driven DNA repair program and a *ZIC5*-associated differentiation state directly govern distinct multidrug resistance mechanisms in MB. Targeting the CDK4/6-E2F1 axis with palbociclib represents a promising precision therapeutic strategy to overcome chemoresistance in high-risk patients.

Introduction

Medulloblastoma (MB) represents the most prevalent malignant pediatric brain tumor and remains a profound clinical challenge in pediatric oncology. Current standard therapeutic regimens involve maximal safe surgical resection followed by craniospinal irradiation and intensive cytotoxic chemotherapy[1]. Despite these aggressive multi-modal interventions, tumor recurrence driven by

acquired multidrug resistance continues to be the primary cause of treatment failure and disease-associated mortality. Clinical outcomes are particularly devastating for patients harboring Group 3 and Group 4 tumors, which frequently exhibit high rates of dissemination and intrinsic evasion of conventional DNA-damaging agents[2]. Consequently, elucidating the molecular mechanisms underlying multidrug resistance is urgently required to discover novel vulnerabilities and develop precision-targeted therapies for refractory MB.

Tumor intratumoral heterogeneity serves as the fundamental basis for therapeutic evasion, allowing rare resistant clones to survive and repopulate under pharmacological pressure. Recent advancements in single-cell RNA sequencing have revolutionized our understanding of the medulloblastoma microenvironment and mapped intricate developmental hierarchies within malignant compartments[2,3]. Previous single-cell profiling studies have successfully delineated diverse cellular states, spanning uncommitted progenitor-like proliferation to distinct neurogenic differentiation trajectories. However, the precise temporal regulatory networks coordinating the emergence of specific chemoresistant phenotypes are incompletely understood. Furthermore, how disparate multidrug-resistant cellular states evolve under continuous therapeutic stress and dynamically dictate poor clinical prognosis remains a prominent knowledge gap[4]. Addressing these unresolved transcriptional dynamics is critical for intercepting evolutionary bottlenecks during MB progression.

In the present study, we integrated single-cell and large-scale bulk transcriptomic datasets to systematically map the multidrug resistance landscape of MB. We identified two mechanistically distinct chemoresistant populations including an *E2F1* driven DNA repair subpopulation mediating resistance through enhanced cell-cycle programs, and a *ZIC5* regulated neuronal differentiation state associated with Hedgehog signaling. By utilizing pseudotime trajectory inference and computational vector field perturbations, we computationally defined the essential transcription factor cascades required for the maintenance of these chemoresistant phenotypes. Subsequent deconvolution of clinical cohorts confirmed that the hyperactivation of *E2F1* acts as a robust indicator of poor overall survival. Driven by these molecular insights, we applied computational drug screening and molecular docking to identify palbociclib as a highly potent candidate capable of disrupting the essential CDK4/6 and E2F1 axis. Together, this study delineates a high-resolution regulatory map of MB chemoresistance and propose a clinically translatable therapeutic strategy to overcome multidrug-resistant microenvironments in high-risk patients.

Materials and Methods

Data collection and preprocessing

This study included nine MB samples. Seven publicly available scRNA-seq datasets were obtained from the Gene Expression Omnibus: six samples (GSM6537659-GSM6537664) from a study by Riemondy et al. (2022) comprising two Group 3 and four Group 4 tumors[5], and one SHH tumor (GSM6607007) from a

study by Maxwell et al. (2021)[6]. In addition, two surgically resected MB specimens (MB-1, MB-2) were collected at with informed consent and institutional review board approval. Fresh tumor tissue was dissociated into single-cell suspensions, and libraries were generated using the 10x Genomics Chromium platform (3' v3 chemistry). Sequencing was performed on an Illumina NovaSeq 6000 system. For each sample, raw gene expression matrices were processed with Cell Ranger (v7.0.0) and aligned to the GRCh38 reference genome.

For each of the nine samples, an independent quality control procedure was applied. Cells with fewer than 300 or more than 5,500 detected genes, and cells with a mitochondrial gene percentage exceeding 15%, were removed. Potential doublets were identified using the DoubletFinder algorithm (v2.0.3)[7]. Briefly, artificial doublets were simulated by randomly combining existing cell profiles, and a cell-specific doublet score was computed based on the local density of artificial doublets in PCA space. The optimal pK parameter was determined by maximizing the recovery of simulated doublets. Cells with doublet scores above the threshold set by the expected doublet rate (approximately 1% per 1,000 cells sequenced) were excluded. The parameters pN = 0.25 and nPCs = 30 were used. After doublet removal, expression data for each sample were normalized (LogNormalize), and variable features were identified. Samples were then integrated using the Harmony algorithm (v1.0) to correct for batch effects while preserving biological variation[8]. The integrated dataset was scaled, and principal component analysis (PCA) was performed on the top 2,000 variable features. UMAP was computed on the harmony embeddings for visualization.

Cell clustering and annotation

Unsupervised clustering was performed using the FindClusters function in Seurat (v5) with a resolution parameter tuned to identify biologically distinct populations. Cell types were annotated based on the expression of canonical markers described in prior MB single-cell studies[9,10]. Clusters were assigned to malignant subgroups (MB group3, MB group4, MB SHH, proliferative, and differentiation states) or tumor microenvironment components (immune and stromal cells) following these established signatures.

Drug sensitivity prediction and multidrug resistance assessment

Drug sensitivity prediction was performed using the GDSC2 dataset (Genomics of Drug Sensitivity in Cancer, https://www.cancerrxgene.org/)[11], which contains gene expression profiles and half-maximal inhibitory concentration (IC_{50}) values for 969 cell lines across 198 compounds. Seven agents clinically relevant to MB treatment were selected: cisplatin, vincristine, cyclophosphamide, carmustine, temozolomide, irinotecan, and teniposide. For each drug, a ridge regression model ($\alpha = 0$) was trained on GDSC2 expression data using the R package glmnet (v4.1) with 10-fold cross-validation to select the optimal regularization parameter (λ)[12]. Common genes between GDSC2 and the Cavalli cohort expression matrix (GSE85217) were retained[13], and both datasets were Z-score standardized per gene

prior to modeling. The trained model was then applied to predict IC_{50} values for each sample in the Cavalli cohort (n = 763).

For each drug, Cavalli samples were stratified into resistant (top 25% of predicted IC_{50}) and sensitive (remaining 75%) groups. Differentially expressed genes between these groups were identified using the limma package (v3.58) with the model ~ 0 + group[14]. Genes with $\log_2$ fold change > 0.5 and Benjamini-Hochberg-adjusted P < 0.05 were defined as the drug-specific resistance upregulated gene set. AUC-based resistance scoring in single-cell data. For each drug, a single-cell resistance score was calculated as the mean log-normalized expression of the corresponding resistance upregulated gene set, yielding an AUC-like metric. For composite multidrug resistance scoring, core resistance genes were first identified as those present in the resistance gene sets of at least four of the seven drugs (n = 450). The composite multidrug resistance score for each cell was then defined as the mean expression of these core genes. Scores were summarized at the cell-type level for comparison.

Pathway activity analysis

Gene set enrichment analysis was performed on the malignant cell subpopulations using single-sample Gene Set Enrichment Analysis (ssGSEA) as implemented in the GSVA package (v1.50)[15]. Hallmark gene sets (h.all.v2026.1.Hs.symbols) were obtained from the Molecular Signatures Database (MSigDB)[11]. Enrichment scores were calculated per cell and averaged within each cell type, followed by row-wise Z-score normalization for visualization.

Single cell regulatory network inference

To map the gene regulatory networks governing the chemoresistant phenotypes, we applied the SCENIC pipeline to our scRNA-seq dataset[16]. Coexpression modules between transcription factors (TFs) and candidate target genes were inferred first. We then performed cis regulatory motif enrichment analysis to filter out indirect targets and define high confidence regulons. The activity of each regulon within individual cells was quantified using the AUCell algorithm. To isolate the regulatory drivers specific to the DNA repair phenotype, we performed differential regulon activity analysis between the MB DNA repair subpopulation and all other malignant clusters using a Wilcoxon rank sum test. Regulons demonstrating significant upregulation (typically defined by an adjusted P-value ≤ 0.05 and the highest positive log2 fold change) were selected.

Pseudotime trajectory analysis

Cellular developmental trajectories were reconstructed utilizing the Monocle 3 package[17]. A cell data set object was constructed from the normalized expression matrix of the relevant cellular subsets. Following principal component analysis and data integration, the global developmental graph was learned across the UMAP embedding. Cells were then ordered in pseudotime by manually designating

the biological root node for each defined trajectory. To evaluate the expression dynamics of the identified regulatory networks, the normalized expression matrices of the candidate transcription factors were extracted and mapped along the pseudotime continuum. Cells were grouped into sequential pseudotime bins, and the smoothed average expression of each transcription factor was calculated and scaled to generate temporal expression heatmaps.

In silico gene perturbation and vector field prediction

To functionally evaluate target TFs, we performed in silico genetic perturbations using CellOracle[18]. The single cell expression matrix, metadata, and UMAP coordinates were imported into a Python computational environment. A base human gene regulatory network was customized to our dataset through ridge regression models to build cluster specific regulatory links. To establish an accurate directional flow of cellular fate, Monocle 3 derived pseudotime values were incorporated as a prior parameter to explicitly guide the cell state transition probabilities. For simulation, the expression level of candidate TFs, such as *E2F1*, was artificially set to zero to mimic complete gene knockout. The resulting cascade of transcriptomic changes was formulated into a vector field and projected as grid arrows and streamlines onto the UMAP space, allowing quantitative visualization of fate perturbation away from the original MB DNA repair trajectory.

Estimation of cell fractions by transcriptomic deconvolution

To infer the relative proportions of specific cell subpopulations within the bulk tumor microenvironment, we employed the Multi-subject Single Cell deconvolution (MuSiC) algorithm[19]. The paired scRNA-seq dataset was utilized as a high-resolution reference to compute the cell-type specific gene expression profiles. Subsequently, these reference profiles were integrated with the large-scale bulk RNA sequencing cohort. The algorithm mathematically transformed the bulk transcriptomic signals into precise cell composition estimations, allowing us to quantify the relative abundance of distinct malignant populations, including the MB DNA repair and MB group 4 neuron differentiation states, for each individual bulk sample.

Survival analysis

Clinical outcome correlations underlying distinct cell fractions and driver genes were evaluated utilizing standard survival analysis techniques. Patients were systematically stratified into high and low groups taking the median cell fraction values or the median gene expression levels as the cut-off thresholds. Overall survival curves were subsequently generated utilizing the Kaplan-Meier (K-M) method and tested using the long-rank test. The statistical significance of survival differences between the stratified patient groups was determined applying the log-rank test. Results with a p value less than 0.05 were considered statistically significant unless otherwise specified in the specific comparisons.

Drug susceptibility prediction and molecular docking

To translate the prognostic gene signatures into clinical therapeutic strategies, the top 100 DEGs characteristic of the MB DNA repair subgroup were submitted to the Connectivity Map (CMap) database to screen for potential small-molecule inhibitors[20]. The compound with strict mechanistic associations to the target was selected for downstream validation.

Molecular docking was performed to simulate the ligand-receptor binding affinity. The 3D conformer of the selected compound (Palbociclib) was retrieved from the PubChem database in SDF format. The crystalline structure of the target protein (PDB ID: 1H24) was obtained from the RCSB Protein Data Bank. Cavity detection and blind docking were executed using the CB-Dock2 web server[21], which integrates the AutoDock Vina algorithm optimized for maximized binding precision. The conformation displaying the lowest binding energy (Vina score in kcal/mol) was considered the most stable and visualized to identify critical interacting residues.

Results

Single-cell transcriptomic landscape of MB

We integrated single-cell RNA-seq data from MB samples (seven public, two in-house) using Harmony after independent quality control (Fig. 1A, Table 1). The integrated dataset corrected batch effects while preserving biological variation, as shown by the UMAP embedding (Fig. 1B). Unsupervised clustering identified 16 transcriptionally distinct populations (Fig. 1C). Malignant clusters included MB group3, MB group4, MB SHH, their proliferative counterparts (MB proliferative, MB SHH proliferative), differentiation states (MB group3 differentiation, MB group4 neuron differentiation), and a DNA-repair-associated cluster (MB DNA repair). The tumor microenvironment (TME) contained glia progenitor like, endothelial, microglia, macrophage, neutrophil, cytotoxic T cell, and B cell. No WNT-subgroup cells were detected, consistent with cohort composition.

Table 1
Sample information of scRNA-seq data included.

GEO accession	GSM accession	Tumor type	Platform	Quantified cell
GSE212559	GSM6537659	G3	Illumina NextSeq 500	13,713
GSE212559	GSM6537660	G3	Illumina NextSeq 500	638
GSE212559	GSM6537661	G4	Illumina NextSeq 500	4,111
GSE212559	GSM6537662	G4	Illumina NextSeq 500	2,271
GSE212559	GSM6537663	G4	Illumina NextSeq 500	3,981
GSE212559	GSM6537664	G4	Illumina NextSeq 500	7,219
GSE214469	GSM6607007	SHH	Illumina NovaSeq 6000	4,671
MB-1	In-house data	SHH	Illumina NovaSeq 6000	7,215
MB-2	In-house data	SHH	Illumina NovaSeq 6000	2,231

Cell identities were confirmed by canonical marker expression (Fig. 1D). Malignant subgroups expressed expected genes such as *GABRA5* and *IMPG2* (group 3), *STMN2* and *NEUROD1* (group 4), and *SFRP1* and *SOX9* (SHH). Proliferative clusters expressed *MKI67* and *TOP2A*. Immune populations showed lineage-appropriate markers: myeloid cells expressed *CD14* and *CSF1R*, T cells expressed *CD3E* and *CD8A*, B cells expressed *MS4A1*, and neutrophils expressed *CXCR2*. Stromal components were marked by *VWF* (endothelial) and *OLIG1* (glial progenitors), consistent with prior MB atlases[9,10].

Sample composition varied markedly (Fig. 1E). GSM6537659 and GSM6537660 were dominated by MB group3 cells. GSM6537661 through GSM6537664 were predominantly MB group4 with a large MB proliferative fraction. GSM6607007 displayed a mixed profile: MB group4 and MB proliferative cells together comprised approximately 45% of the population, accompanied by MB SHH cells and a notable immune infiltrate (Macrophage, Microglia). In contrast, the in-house samples MB-1 and MB-2 were heavily enriched for immune cells, especially neutrophils and macrophages, with malignant cells representing a minor proportion.

Hierarchical clustering of average expression profiles using Spearman correlation revealed a clear separation between malignant and TME compartments (Fig. 1F). Within the malignant branch, group3/4 clusters and their derivatives grouped together, distinct from the SHH branch, recapitulating the known molecular divergence between SHH-driven and group3/4 medulloblastoma. Immune and stromal cells formed a separate major clade with further subclustering by lineage. Together, these results provide a global cellular framework for downstream analyses.

Multi-drug resistance profiling identifies MB DNA repair and neuronally differentiated group 4 cells as the principal chemoresistant populations

To systematically characterize chemoresistance across MB subgroups, we predicted the half-maximal inhibitory concentration (IC_{50}) values for seven clinically relevant agents, including cisplatin, vincristine, cyclophosphamide, carmustine, temozolomide, irinotecan, and teniposide, based on the Cavalli cohort using ridge regression models trained on GDSC2 data. Resistance gene sets were defined for each drug by identifying genes significantly upregulated in the top 25% of predicted IC_{50} samples relative to the remaining 75%. An AUC-based resistance score was then calculated for each tumor cell subpopulation in our single-cell dataset. The resulting heatmap revealed heterogeneous drug sensitivity patterns across malignant subtypes (Fig. 2A). MB group 4 neuron-differentiated cells, MB group 3, MB group 3 differentiated cells, MB group4, and MB DNA repair cells exhibited elevated resistance scores for cisplatin, irinotecan, and teniposide, indicating cross-resistance to DNA-damaging agents and topoisomerase inhibitors.

We next integrated the resistance gene sets from all seven drugs, identifying 450 core resistance genes that were consistently upregulated in at least four of the seven drug-resistant groups (Fig. 2B). A composite multidrug resistance score, derived from the mean expression of these core genes, was highest in the MB DNA repair subpopulation, followed by MB group4 neuron-differentiated cells (Fig. 2C). This result suggests that these two subpopulations are the major reservoirs of multidrug resistance within MB.

Examination of the top 50 most frequently occurring core resistance genes across MB malignant subsets revealed distinct expression patterns (Fig. 2D). Approximately two-thirds of these genes, including *AGT*[22], *ABCC8*[23] and *GABBR2*[24], were predominantly upregulated in MB group 4 neuron-differentiated cells. Many of these encode neurotransmitter receptors and ion channels, suggesting that the neuronal differentiation program may confer resistance through altered cell-surface signaling. In addition, roughly half of the core genes were highly expressed in MB DNA repair cells, among them the muscarinic acetylcholine receptor genes *CHRM3* and *CHRM4*, the adhesion molecule *CDH23*, the G protein-coupled receptors *GRM8* and *GPR12*, and *GFAP*. The enrichment of GPCRs and cholinergic receptors points to a potential role for microenvironmental cues in sustaining the resistant state. Focusing on cisplatin as a paradigm DNA-damaging agent, we projected the cisplatin resistance AUC score onto the UMAP embedding. High resistance scores were predominantly localized to MB DNA repair, MB group 4 neuron-differentiated cells, and a fraction of MB SHH proliferative cells (Fig. 2E). Quantitative comparison confirmed that the top three subpopulations with the highest cisplatin resistance scores were MB DNA repair, MB group 4, and MB group 4 neuron-differentiated cells (Fig. 2F), consistent with the multidrug resistance pattern.

To elucidate the molecular programs driving chemoresistance, we performed ssGSEA analysis of Hallmark pathways across malignant cell subsets (Fig. 2G). HALLMARK E2F TARGETS and HALLMARK G2M CHECKPOINT were highly and selectively enriched in MB DNA repair and MB proliferative cells, showing remarkably similar enrichment profiles. This co-activation of cell-cycle and DNA replication programs likely underlies the enhanced DNA repair capacity and cisplatin resistance of MB DNA repair cells. In contrast, HALLMARK HEDGEHOG SIGNALING was enriched in MB group4 neuron-differentiated

cells and, as expected, in MB SHH cells, suggesting that Hedgehog pathway activity may contribute to the chemoresistant phenotype of neuronally differentiated group 4 cells.

Collectively, these results delineate two mechanistically distinct modes of chemoresistance in MB: an E2F/G2M-driven program associated with heightened DNA repair activity in MB DNA repair cells, and a Hedgehog-associated, neuronally differentiated program in group 4 cells. The coexistence of these programs within individual tumors likely provides a composite defense against cytotoxic therapy.

Transcriptional trajectory and in silico perturbation identify critical drivers of the MB DNA repair subgroup

To uncover the regulatory mechanisms orchestrating the chemoresistant DNA repair phenotype, we performed SCENIC specifically on the MB DNA repair subpopulation. This analysis identified a core regulon network comprising 73 TFs with significantly upregulated activity in this subset. The resulting network highlighted extensive regulatory interactions between these TFs and their high confidence downstream target genes, suggesting a highly coordinated transcriptional program underlying this specific state (Fig. 3A).

Given the transcriptomic and pathway level similarities between the MB group 4, MB proliferative, and MB DNA repair cells, we hypothesized that these populations form an evolutionary continuum. To test this, we applied Monocle 3 to reconstruct their developmental trajectory. The pseudotime ordering revealed a continuous biological transition where uncommitted MB group 4 cells progress through a highly proliferative intermediate stage before terminating in the MB DNA repair state (Fig. 3B), supporting a model wherein acquisition of the chemoresistant DNA repair phenotype is intimately linked to the active cell cycle machinery.

We next mapped the expression dynamics of the 73 SCENIC identified TFs along this defined pseudotime trajectory. As a result, a distinct module of TFs whose expression steadily increased and peaked at the late stages of pseudotime within the MB DNA repair cluster was identified (Fig. 3C). This late-stage regulatory module included notable cell cycle regulators, chromatin remodelers, and DNA damage response factors, including *NFYB*, *CBX3*, *HMGB3*, *RAD21*, *E2F1*, *EZH2*, *E2F3*, *MYBL1*, *E2F7*, *MBD4*, *PAXIP1*, *CEBPG*, *ZNF682*, *GABPB1*, *SMARCA4*, and *CEBPZ*. The persistent enrichment of E2F family members (*E2F1*, *E2F3*, *E2F7*) along the trajectory is consistent with prior mechanistic studies establishing *E2F1* as a dual regulator that not only drives cell cycle progression but also directly promotes homologous recombination and DNA repair survival upon genotoxic stress[25]. Furthermore, key epigenetic and structural regulators such as *EZH2* and *RAD21* have been well documented as essential components for maintaining genomic stability and therapeutic resistance in high grade pediatric malignancies[26,27].

To functionally validate the requirement of these terminal TFs in sustaining the MB DNA repair trajectory, we utilized CellOracle to perform transcriptomic vector field predictions and in silico gene perturbations. We selected *E2F1* as a representative candidate given its prominent position in both the regulatory

network and the time aligned expression cascade. Virtual knockout of *E2F1* resulted in a distinct disruption of the predicted cell fate flow, as reflected by the directional vectors that originally guided cells from the MB proliferative state toward the MB DNA repair state were significantly attenuated and disorganized upon *E2F1* suppression (Fig. 3D).

To further characterize the biological functions associated with this regulatory network, we performed functional enrichment analysis on the core downstream target genes characterizing the MB DNA repair state (Fig. 3E). Consistently, the analysis revealed a highly significant enrichment in biological processes and pathways related to cell cycle machinery and DNA damage survival. Notably, terms such as S Phase, Synthesis of DNA and Double-strand Break Repair via Break-Induced Replication were profoundly enriched. Additionally, the prominent enrichment of epigenetic signaling, including HATs Acetylate Histones, further corroborated the structural chromatin rewiring required for this state. These functional annotations substantiate that the dynamic *E2F1*-driven transcriptional module strictly coordinates aggressive DNA replication and repair machineries to confer multidrug resistance in this specific subpopulation.

Trajectory and computational manipulation reveal the Hedgehog associated regulatory axis in differentiating group 4 MB

To elucidate the transcriptional drivers coordinating the unique neuronal differentiation program within group 4 MB, we focused on the MB group 4 neuron differentiation subpopulation. By intersecting our SCENIC regulatory predictions with differential expression analysis, we refined a specific module of 26 TFs intrinsically enriched in this cellular subset (Fig. 4A).

Given the structural transcriptomic similarities between MB group 4 cells and the MB group 4 neuron differentiation population, we employed Monocle 3 to continuously align these distinct states. The resulting trajectory defined an ordered progression where cells transitioning towards the differentiated state exhibited sequential waves of transcriptional rewiring (Fig. 4B). Expression mapping of the 26 candidate transcription factors along this defined pseudotime continuum revealed a group of regulators whose induction closely paralleled the terminal differentiation stage. This dynamic core included potent neural developmental regulators and transcription factors such as *NEUROD2*, *EN2*, *ZIC5*, *TEAD1*, *POU2F1*, and *JUN*. Notably, the composition of these late pseudotime TFs matches the biological hallmark of this group. The upregulation of *ZIC5* and *EN2* acts as a molecular bridge to the pathway analysis previous observation of Hedgehog signaling enrichment in this subpopulation. Prior biochemical literature establishes that ZIC family proteins function as essential physical interactors and TFs for GLI transcription factors, thereby acting as obligatory components of the canonical Hedgehog signaling cascade[28]. Furthermore, *EN2* operates as a classical developmental node functionally responsive to Hedgehog pathway activation in MB pathogenesis[29].

To quantitatively test whether these Hedgehog associated transcription factors are required for establishing or maintaining the differentiated state, we also performed analytical vector field simulations. We selected *ZIC5* as the primary candidate for in silico perturbation due to its dual structural role in cerebellar neurogenesis and Hedgehog signal transduction. Simulation of *ZIC5* transcript removal induced a prominent collapse of the directional pseudotime vectors (Fig. 4D). Specifically, the fluid transcriptomic trajectories that originally directed the development from the uncommitted MB group 4 state to the specialized neuron differentiation terminal were largely disorganized and attenuated upon *ZIC5* interference. Moreover, functional enrichment analysis of high-confidence targets suggests an enrichment of terms including Response to ethanol and Regulation of monoatomic ion transmembrane transport (Fig. 4E). Therefore, these computational validations suggest that the coordination of Hedgehog pathway coupled regulators is required for the targeted progression of the neuronal differentiation phenotype.

Clinical and prognostic significance of distinct cell subpopulations and key driver genes

To further explore the potential implications of specific cell subpopulations in the real-world tumor microenvironment, we evaluated their distribution patterns across various clinical features in a large-scale transcriptomic cohort via a deconvolution approach. Clinical correlation analysis revealed that the distribution of the MB DNA repair cell subpopulation was highly specific among tumor subgroups, with a significant enrichment observed in Group 3 and Group 4 patients (Fig. 5A). Regarding age demographics, these specific clones were similarly enriched in the 4 to 10 and 10 to 17 age ranges. Notably, the proportion of MB DNA repair cells was substantially elevated in patients presenting with metastasis (M+). Evaluation of the MB group4 neuron differentiation cell population demonstrated an upward trend in the Group 3 subtype ($p = 0.053$). Although the distribution of this cellular cluster was not restricted by age, its abundance was significantly increased in the M+ group (Fig. 5B). These distribution patterns preliminarily suggest that these two distinct malignant states might be closely involved in the clinical progression and the acquisition of distal metastatic potential in MB.

Subsequently, we conducted Kaplan-Meier survival models to assess the direct impact of specific cell abundances on patient prognosis. For the MB DNA repair population, the results indicated that its enrichment level did not constitute a significant prognostic factor for overall survival in either the overall cohort or the Group 4 sub-cohort ($p = 0.14$ and $p = 0.66$, respectively, Fig. 5C). In turn, the proportional variation of MB group4 neuron differentiation cells exhibited a certain predictive value for survival. In the overall population, patients with a lower infiltration ratio of this cell subpopulation showed a tendency toward survival benefit ($p = 0.087$), and this trend was further pronounced in the Group 4 clinical subtype cohort ($p = 0.055$, Fig. 5D).

Given that our previous transcriptional regulatory analysis identified *E2F1* and *ZIC5* as direct drivers in maintaining the associated drug-resistant phenotypes, we further extended the prognostic evaluation to the molecular level. Survival validation demonstrated that the expression of *E2F1* serves as a robust

adverse prognostic indicator, as the overall survival time of patients in the *E2F1* low-expression group was significantly prolonged (p = 0.004), and this survival benefit conferred by low expression remained highly consistent within the specific Group 4 subpopulation (p = 0.0037, Fig. 5E). Conversely, the upregulation of *ZIC5*, a target gene characterizing another differentiation state, only revealed a marginal trend toward survival improvement in the overall cohort (p = 0.086) and failed to achieve any statistically significant prognostic impact within Group 4 (Fig. 5F). Taken together, the abundance signatures derived from specific invasive cellular states and the core driving networks represented by *E2F1* provide robust evidence for elucidating the progression trajectory and clinical outcomes of MB.

Identification of Palbociclib as a therapeutic candidate targeting the CDK4/6 and E2F1 axis

Based on the highly expressed *E2F1* and the top 100 DEGs related to DNA repair in MB, we utilized the CMap database to identify potential antagonist drugs capable of reversing this transcriptomic signature. The analysis yielded the top five candidate compounds: etacrynic acid, palbociclib, chlorpromazine, mitoxantrone, and GSK1070916 (Fig. 6A). Given the classical regulatory relationship between the CDK4/6 signaling pathway and the cell cycle transcription factor *E2F1*[30], palbociclib, a highly selective CDK4/6 inhibitor, was selected for further investigation.

To validate the clinical applicability of palbociclib, we performed a ridge regression model to predict the drug sensitivity among MB patients. The results demonstrated that patients in the *E2F1* high-expression group exhibited significantly lower predicted IC50 values for palbociclib compared to the low-expression group (p value = 0.027), indicating enhanced drug sensitivity in patients with elevated *E2F1* levels (Fig. 6B).

Furthermore, to elucidate the spatial binding capacity, we structuralized a molecular docking analysis between palbociclib and the target protein. The analysis revealed a robust binding affinity with a minimum binding energy of -9.2 kcal/mol. The visualization of the top-ranked conformation demonstrated that palbociclib was stably embedded within the hydrophobic pocket of the receptor via specific interactions with amino acid residues including GLU12, LYS33, LEU55 (Fig. 6C). Taken together, the integrated mechanistic network highlights that palbociclib potentially targets the CDK4/6-E2F1 axis, highlighting its value in reversing the MB DNA repair phenotype associated with MB chemoresistance (Fig. 6D).

Discussion

The evolution of multidrug resistance in MB is a dynamic and multifaceted process that fundamentally limits the efficacy of conventional chemotherapy. While previous transcriptomic profiling has mapped the static heterogeneity of pediatric brain tumors, the precise transitional frameworks driving the acquisition of treatment refractoriness have remained elusive. In this study, we addressed this challenge by constructing an integrated computational pipeline that bridges high-resolution single-cell transition analysis with bulk clinical prognosis and computational pharmacology. Rather than simply cataloging

static cell clusters, we employed pseudotime inference and vector field simulations to capture the continuous transcriptomic rewiring utilized by malignant cells under selective pressure. These efforts allowed us to pinpoint specific evolutionary trajectories toward therapeutic resistance and successfully map these computational findings to actionable clinical strategies.

A central finding of our investigation is the identification of an *E2F1*-driven, enhanced DNA repair trajectory that serves as a major multidrug-resistant reservoir in Group 3 and Group 4 MB. The prominent co-activation of cell-cycle progression and DNA damage repair pathways within this cluster perfectly mirrors the clinical behavior of highly aggressive tumors capable of rapidly mitigating genotoxic stress induced by agents such as cisplatin. Current literature strongly supports the dual functionality of the E2F transcription factor family. Beyond its classical role in the G1/S transition, *E2F1* directly coordinates the transcription of crucial homologous recombination and DNA repair factors upon genomic insult[2]. The steady accumulation of *E2F1* along our projected developmental continuum further corroborates its role as an indispensable master regulator of this resistant state. Importantly, our computational drug screening and spatial docking analysis identified the CDK4/6 inhibitor palbociclib as a highly potent agent capable of reversing this *E2F1*-dependent DNA repair phenotype. Given that CDK4/6 precisely governs the activation of *E2F1*, our results align well with emerging preclinical evidence suggesting that dual cell-cycle and DNA repair targeting presents a robust vulnerability in pediatric high-grade malignancies[31].

In parallel with the DNA repair program, our analysis uncovered a structurally distinct mode of resistance driven by a neuronal differentiation state intrinsically linked to *ZIC5* and Hedgehog pathway components. Traditionally, Group 4 MB is considered clinically independent of classical Sonic Hedgehog (SHH) genomic alterations[1]. However, our single-cell network inference and computational perturbation revealed that uncommitted Group 4 cells might selectively hijack aspects of the Hedgehog signaling cascade, coordinated by the neural developmental transcription factor *ZIC5*, to sustain a transient differentiated phenotype. The ZIC family proteins are well-documented physical interactors and obligate components acting parallel to the canonical GLI transcription factors involved in neural tube development and medulloblastoma pathogenesis[32]. The acquisition of resistance through this differentiated, potentially quiescent-like state highlights the remarkable biologic plasticity of MB cells. This observation underscores the necessity of considering developmental pathway reprogramming as an alternative mechanism of therapeutic evasion, complementing the hyper-proliferative DNA repair mode.

Conclusion

Despite the comprehensive nature of this multi-omics approach, several limitations should be acknowledged. Firstly, although state-of-the-art computational algorithms such as single-cell vector field perturbations provide robust probabilistic validations of transcription factor dependencies, our findings are primarily derived from in silico modeling. The inherent biases associated with transcriptomic deconvolution algorithms and public microarray datasets also dictate the need for cautious

interpretation. Future investigations should prioritize rigorous functional validation using patient-derived xenograft (PDX) models and three-dimensional tumor organoids to physically dissect the roles of *E2F1* and *ZIC5*. Secondly, prospective clinical trials or sophisticated in vivo pharmacological assays are required to definitively validate the therapeutic efficacy of palbociclib in reversing multidrug resistance within the specialized medulloblastoma microenvironment. Continuing these experimental trajectories will be paramount for translating our single-cell specific vulnerabilities into personalized clinical therapies globally.

Abbreviations

CDK4/6
cyclin-dependent kinase 4/6
CMap
Connectivity Map
DEG
differentially expressed gene
E2F1
E2F transcription factor 1
GDSC2
Genomics of Drug Sensitivity in Cancer
GPCR
G protein-coupled receptor
GSVA
Gene Set Variation Analysis
HAT
histone acetyltransferase
IC_{50}
half-maximal inhibitory concentration
K-M
Kaplan-Meier
MB
Medulloblastoma
MSigDB
Molecular Signatures Database
MuSiC
Multi-subject Single Cell deconvolution
PCA
principal component analysis
PDB
Protein Data Bank

PDX
patient-derived xenograft
SCENIC
Single-Cell Regulatory Network Inference and Clustering
scRNA-seq
single-cell RNA sequencing
SHH
Sonic Hedgehog
ssGSEA
single-sample Gene Set Enrichment Analysis
TF
transcription factor
TME
tumor microenvironment
UMAP
Uniform Manifold Approximation and Projection
ZIC5
Zic family member 5.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Declaration of Helsinki and its later amendments. Approval for using the two MB samples for single-cell transcriptome sequencing in OMIX005152 was granted by the Ethics Committee of Children's Hospital of Chongqing Medical University, PR China (Approval No. KY201916). Written informed consent was obtained from all participants (or their legal guardians) prior to the collection and use of the samples for research purposes.

Consent for publication

Informed consent for publication of clinical data and sequencing results was obtained from all participants.

Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

Natural Science Foundation of Jiangxi Provine(20224BAB206067), Science and Technology Research Project of the Jiangxi Provincial Department of Education (GJJ218110).

Author Contribution

Manuscript conception, design, data collection and analysis, drafting and reviewing: Yueliang Yao, Junying Zhang. Data collection: Hang Ji, Ping Liang. Study Supervise: Hang Ji, Ping Liang. Yueliang Yao and Junying Zhang contribute equally to this work.

Acknowledgement

Not applicable.

Availability of data and materials

The raw sequencing data of the two in-house samples (MB-1 and MB-2) are not publicly available at this time because they are part of ongoing research that will be reported in a separate publication. The publicly available single-cell RNA-seq datasets used in this study are accessible from the Gene Expression Omnibus under accession numbers GSM6537659–GSM6537664 and GSM6607007. All processed data, intermediate files, and custom analysis scripts are available from the corresponding authors upon reasonable request.

References

1. Northcott PA, Robinson GW, Kratz CP et al. Medulloblastoma. Nat Reviews Disease Primers 2019;5(1).
2. Hovestadt V, Ayrault O, Swartling FJ, Robinson GW, Pfister SM, Northcott PA. Medulloblastomics revisited: biological and clinical insights from thousands of patients. Nat Rev Cancer. 2019;20(1):42–56.
3. Ruchiy Y, Tsea I, Preka E et al. Genomic tumor evolution dictates human medulloblastoma progression. Neuro-Oncology Adv 2024;6(1).
4. Smith KS, Bihannic L, Gudenas BL, et al. Unified rhombic lip origins of group 3 and group 4 medulloblastoma. Nature. 2022;609(7929):1012–20.
5. Funke VLE, Walter C, Melcher V, et al. Group-specific cellular metabolism in Medulloblastoma. J Transl Med. 2023;21(1):363.

6. Gold MP, Ong W, Masteller AM, et al. Developmental basis of SHH medulloblastoma heterogeneity. Nat Commun. 2024;15(1):270.
7. McGinnis CS, Murrow LM, Gartner ZJ, DoubletFinder. Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019;8(4):329–e337324.
8. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289–96.
9. Hovestadt V, Smith KS, Bihannic L, et al. Resolving medulloblastoma cellular architecture by single-cell genomics. Nature. 2019;572(7767):74–9.
10. Vladoiu MC, El-Hamamy I, Donovan LK, et al. Childhood cerebellar tumours mirror conserved fetal transcriptional programs. Nature. 2019;572(7767):67–73.
11. Yang W, Soares J, Greninger P, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2012;41(D1):D955–61.
12. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.
13. Cavalli FMG, Remke M, Rampasek L, et al. Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell. 2017;31(6):737–.
14. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
15. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics 2013;14(1).
16. Aibar S, González-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083–6.
17. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;566(7745):496–502.
18. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature. 2023;614(7949):742–51.
19. Wang X, Park J, Susztak K, Zhang NR, Li M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun 2019;10(1).
20. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171(6):1437–e14521417.
21. Liu Y, Yang X, Gan J, Chen S, Xiao Z-X, Cao Y. CB-Dock2: improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. Nucleic Acids Res. 2022;50(W1):W159–64.
22. Vaidyanathan G, White B, Affleck DJ, McDougald D, Zalutsky MR. Radioiodinated O(6)-Benzylguanine derivatives containing an azido function. Nucl Med Biol. 2011;38(1):77–92.
23. Begicevic RR, Falasca M. ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance. Int J Mol Sci 2017;18(11).

24. Young SZ, Bordey A. GABA's control of stem and cancer cell proliferation in adult neural and peripheral niches. Physiol (Bethesda). 2009;24:171–85.
25. Biswas AK, Johnson DG. Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage. Cancer Res. 2012;72(1):13–7.
26. Losada A. Cohesin in cancer: chromosome segregation and beyond. Nat Rev Cancer. 2014;14(6):389–93.
27. Alimova I, Birks DK, Harris PS, et al. Inhibition of EZH2 suppresses self-renewal and induces radiation sensitivity in atypical rhabdoid teratoid tumor cells. Neuro Oncol. 2013;15(2):149–60.
28. Koyabu Y, Nakata K, Mizugishi K, Aruga J, Mikoshiba K. Physical and functional interactions between Zic and Gli proteins. J Biol Chem. 2001;276(10):6889–92.
29. Steinbach JP, Kozmik Z, Pfeffer P, Aguzzi A. Overexpression of Pax5 is not sufficient for neoplastic transformation of mouse neuroectoderm. Int J Cancer. 2001;93(4):459–67.
30. Asghar U, Witkiewicz AK, Turner NC, Knudsen ES. The history and future of targeting cyclin-dependent kinases in cancer therapy. Nat Rev Drug Discovery. 2015;14(2):130–46.
31. Ramaswamy V, Taylor MD. Medulloblastoma: From Myth to Molecular. J Clin Oncol. 2017;35(21):2355–63.
32. Olson James M. Therapeutic Opportunities for Medulloblastoma Come of Age. Cancer Cell. 2014;25(3):267–9.

Figures

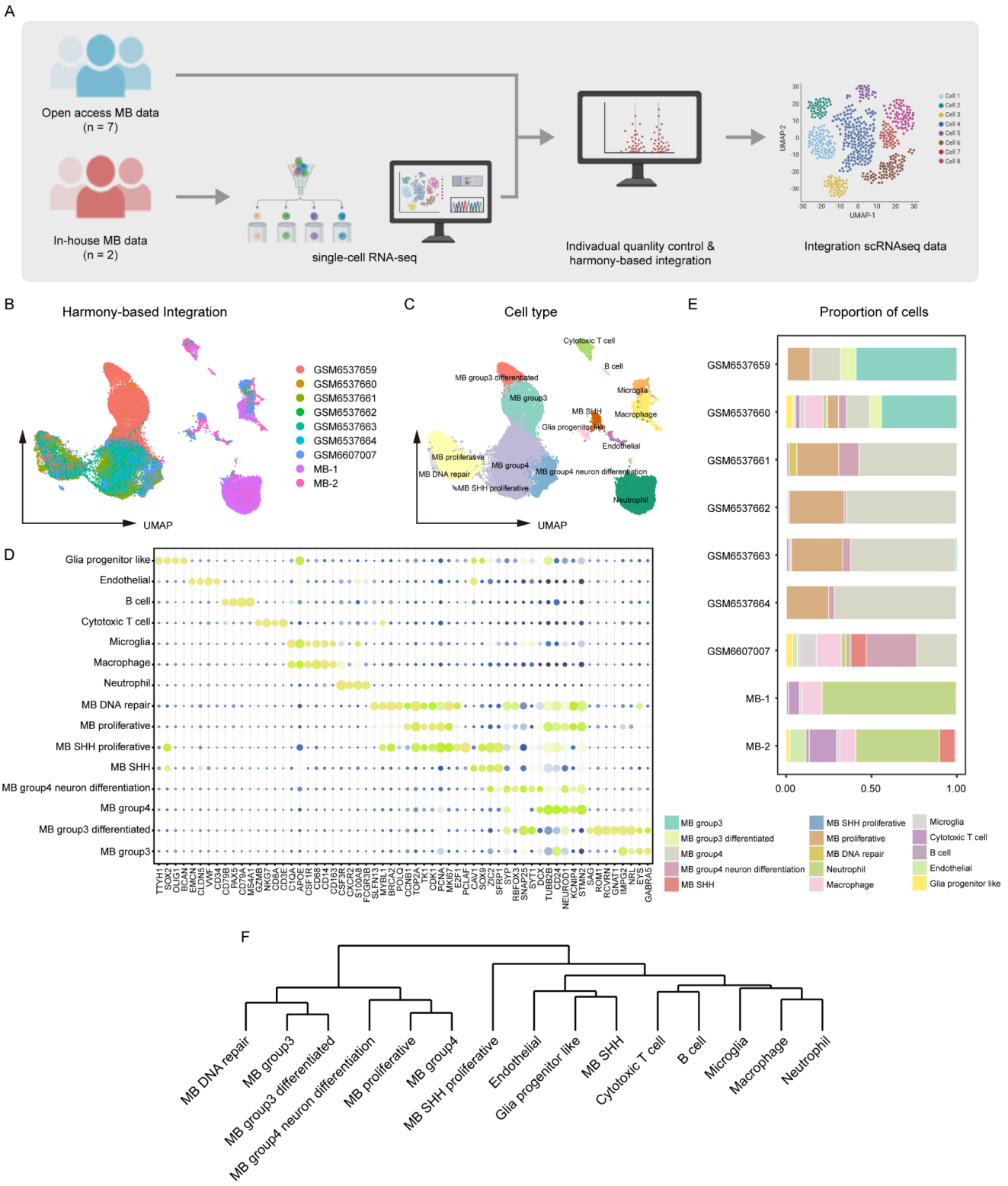


Figure 1

Single-cell transcriptomic landscape of medulloblastoma. (**A**) Schematic workflow of the study. (**B**) UMAP visualization of the integrated samples, colored by sample of origin. (**C**) UMAP visualization of the 15 identified cell populations, colored by cell type. (**D**) Bubble plot of canonical marker genes confirming the identity of each cell population. Dot size represents the proportion of expressing cells and color intensity indicates average expression level. (**E**) Cell type composition across individual samples shown

as stacked bar plots. (**F**) Hierarchical clustering dendrogram of cell types based on average transcriptomic profiles. The distance metric is defined as 1 minus Spearman correlation coefficient, and the complete linkage method was used. Malignant and tumor microenvironment compartments form two distinct major clades.

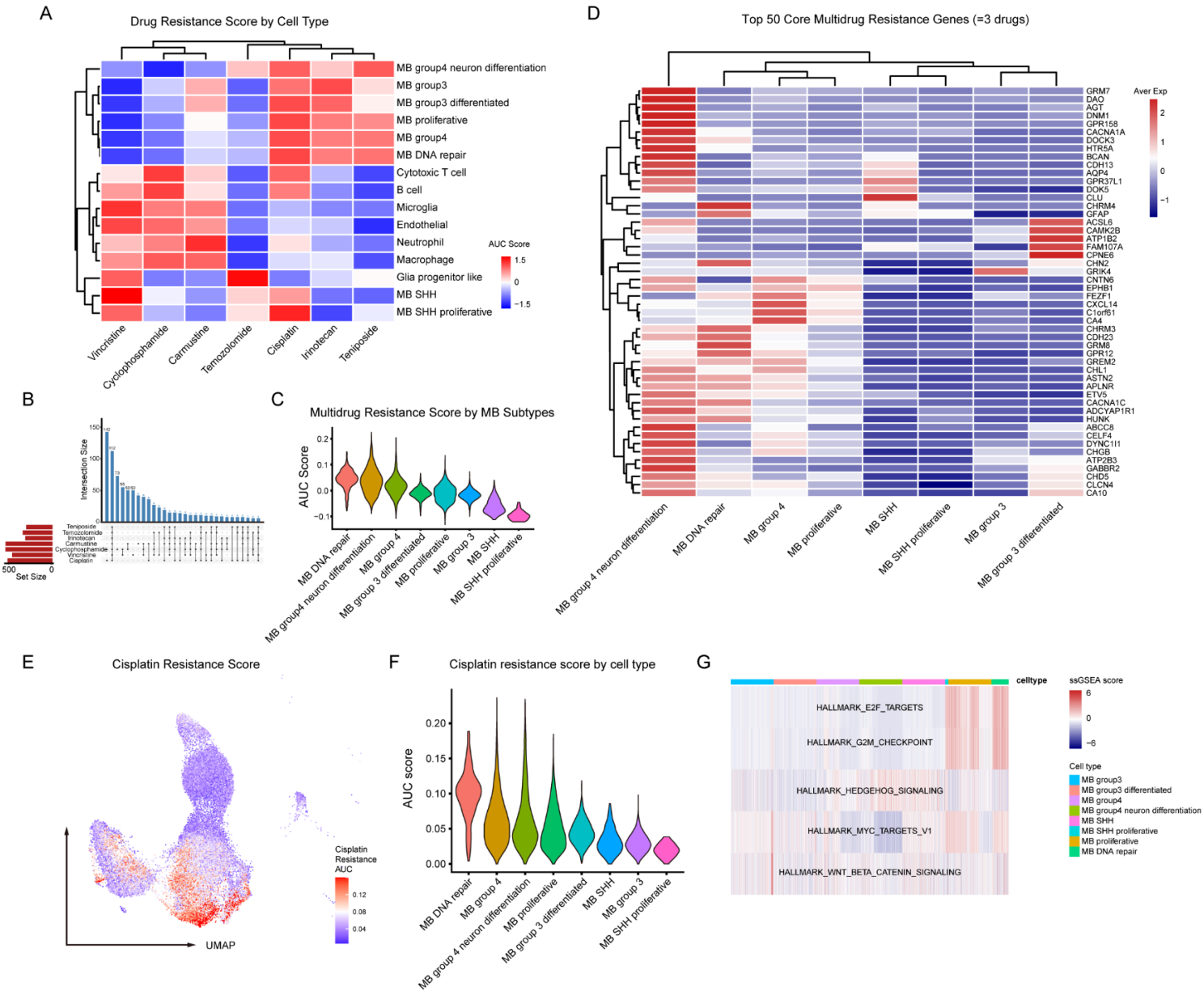


Figure 2

Multi-drug resistance profiling identifies MB DNA repair and neuronally differentiated Group 4 cells as principal chemoresistant populations. (**A**) Heatmap showing AUC-based resistance scores for seven clinically relevant agents across MB malignant subpopulations. Resistance scores were calculated as the mean expression of drug-specific resistance gene sets (genes upregulated in the top 25% vs. bottom 75% of predicted IC_{50} samples; $log_2FC > 0.5$, adjusted $P < 0.05$). Scores are row-scaled (Z-score). (**B**) Violin plot of the composite multidrug resistance score across MB subtypes. The composite score was derived from the mean expression of 450 core resistance genes that were consistently upregulated in at

least four of the seven drug resistance signatures. Center lines denote median values; box limits indicate the first and third quartiles. (**C**) UpSet plot illustrating the overlap of resistance-associated upregulated genes among the seven drugs. The horizontal bar chart (left) shows the total number of upregulated genes per drug. The vertical bar chart (top) displays the number of genes shared across specific drug combinations. The 450 core genes, defined as those shared by ≥4 drugs, are highlighted. (**D**) Heatmap of the top 50 most frequently occurring core resistance genes across MB malignant subpopulations. Expression values represent log-normalized average expression per cell type. (**E**) UMAP visualization of the cisplatin resistance AUC score across all cells. (**F**) Violin plot comparing cisplatin resistance scores across cell types. (**G**) Heatmap of Hallmark pathway enrichment scores (ssGSEA) across malignant cell subpopulations.

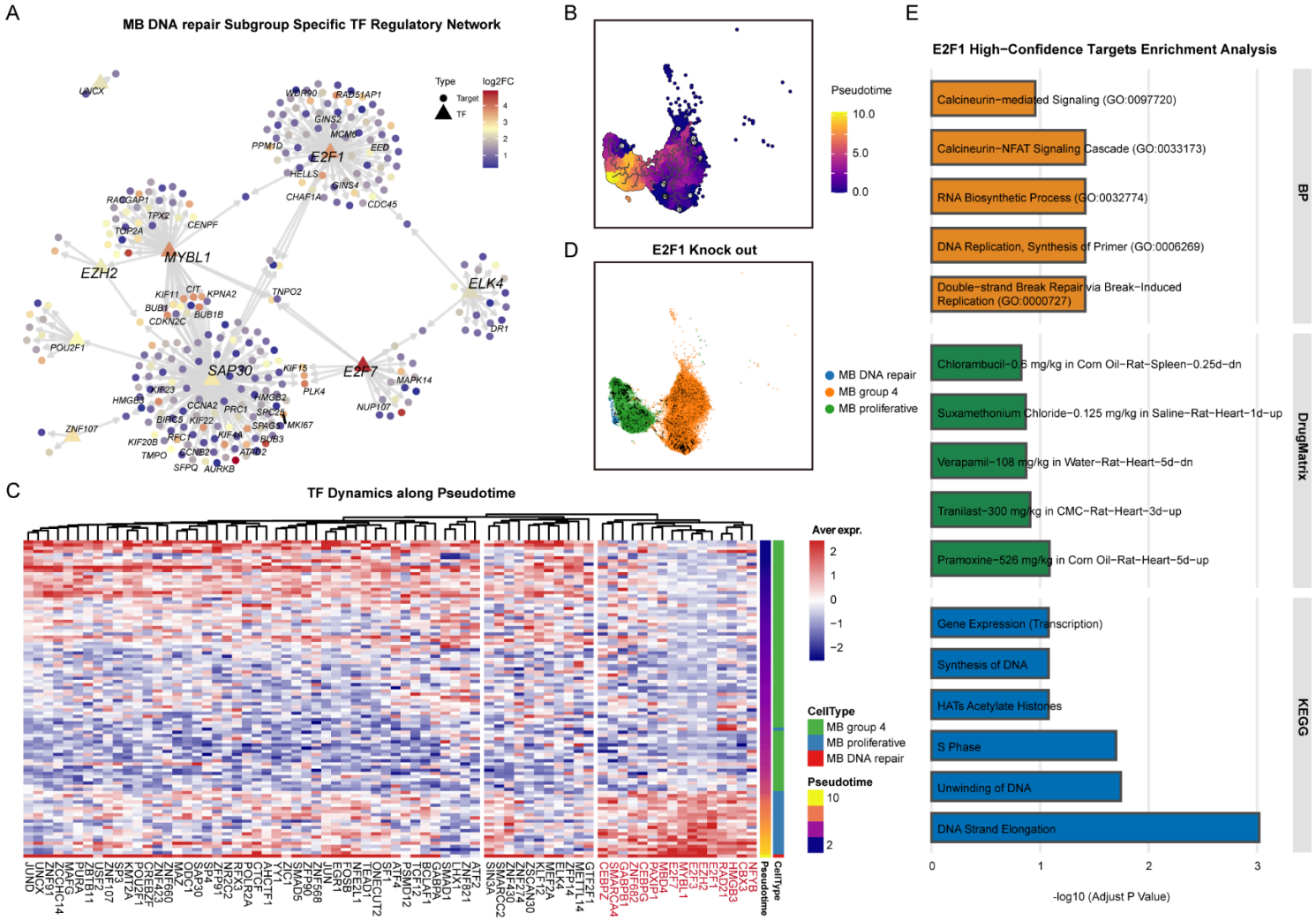


Figure 3

Transcriptional network and trajectory analysis reveals regulatory drivers of the DNA repair phenotype. (**A**) Gene regulatory network inferred by SCENIC for the MB DNA repair subpopulation. The network visualizes 73 significantly upregulated TFs and their respective interactions with high confidence downstream target genes. Genes involved in the hallmark T2F and G2M pathways were annotated. (**B**) Cellular developmental trajectory reconstructed utilizing Monocle 3. The UMAP projection illustrates a

continuous biological progression starting from uncommitted MB group 4 cells, transitioning through the MB proliferative intermediate, and terminating at the MB DNA repair state. (**C**) Heatmap depicting the smoothed expression dynamics of the 73 identified transcription factors along the computed pseudotime. The temporal ordering highlights a specific module of transcription factors that peak at the late evolutionary stage corresponding to the DNA repair phenotype. (**D**) Vector field visualization of in silico gene perturbation via CellOracle. The streamlines projected onto the UMAP space demonstrate the predicted attenuation and disruption of the cellular transition from the proliferative state toward the DNA repair state following the virtual knockout of *E2F1*. (**E**) Functional enrichment analysis using Enrichr.

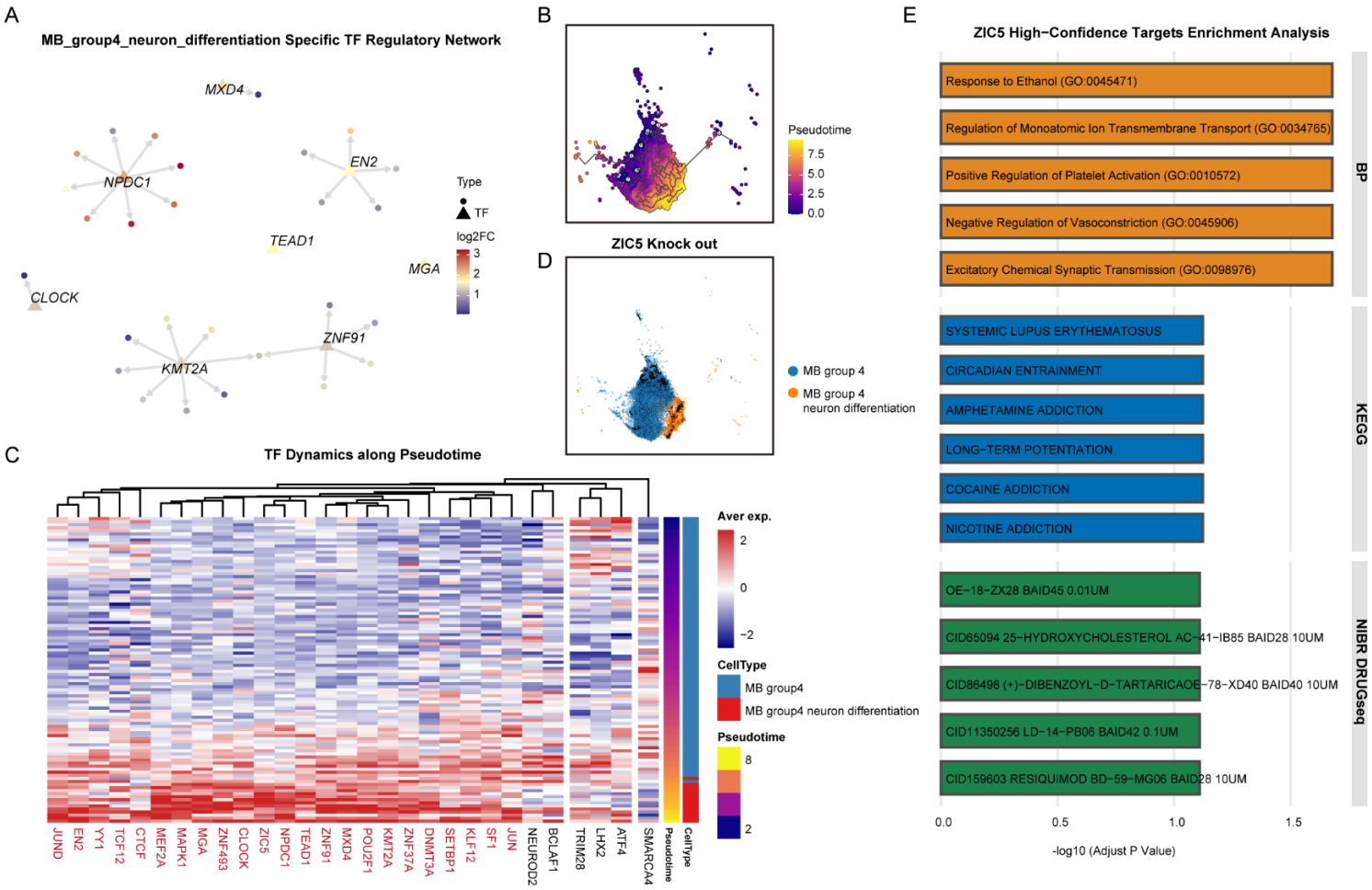


Figure 4

Trajectory and computational manipulation reveal the Hedgehog associated regulatory axis in differentiating group 4 medulloblastoma. (**A**) Transcriptional regulatory module identified by combining SCENIC predictions and differential expression analysis. The data highlights TFs specific to the MB group 4 neuron differentiation subpopulation. (**B**) Cellular trajectory analysis. The projection illustrates an ordered biological transitioning process from MB group 4 cells toward the specialized neuron differentiation state. (**C**) Pseudotime expression mapping of the 26 candidate transcription factors along the developmental continuum. The dynamic ordering reveals sequential waves of transcriptional rewiring with Hedgehog associated regulators prominently defining the terminal differentiation stage. (**D**) Vector

field visualization of in silico *ZIC5* perturbation. The projected streamlines demonstrate the prominent disorganization and attenuation of the original fluid transcriptomic trajectories required for the neuronal differentiation phenotype following virtual *ZIC5* removal. (**E**) Functional enrichment analysis using Erichr.

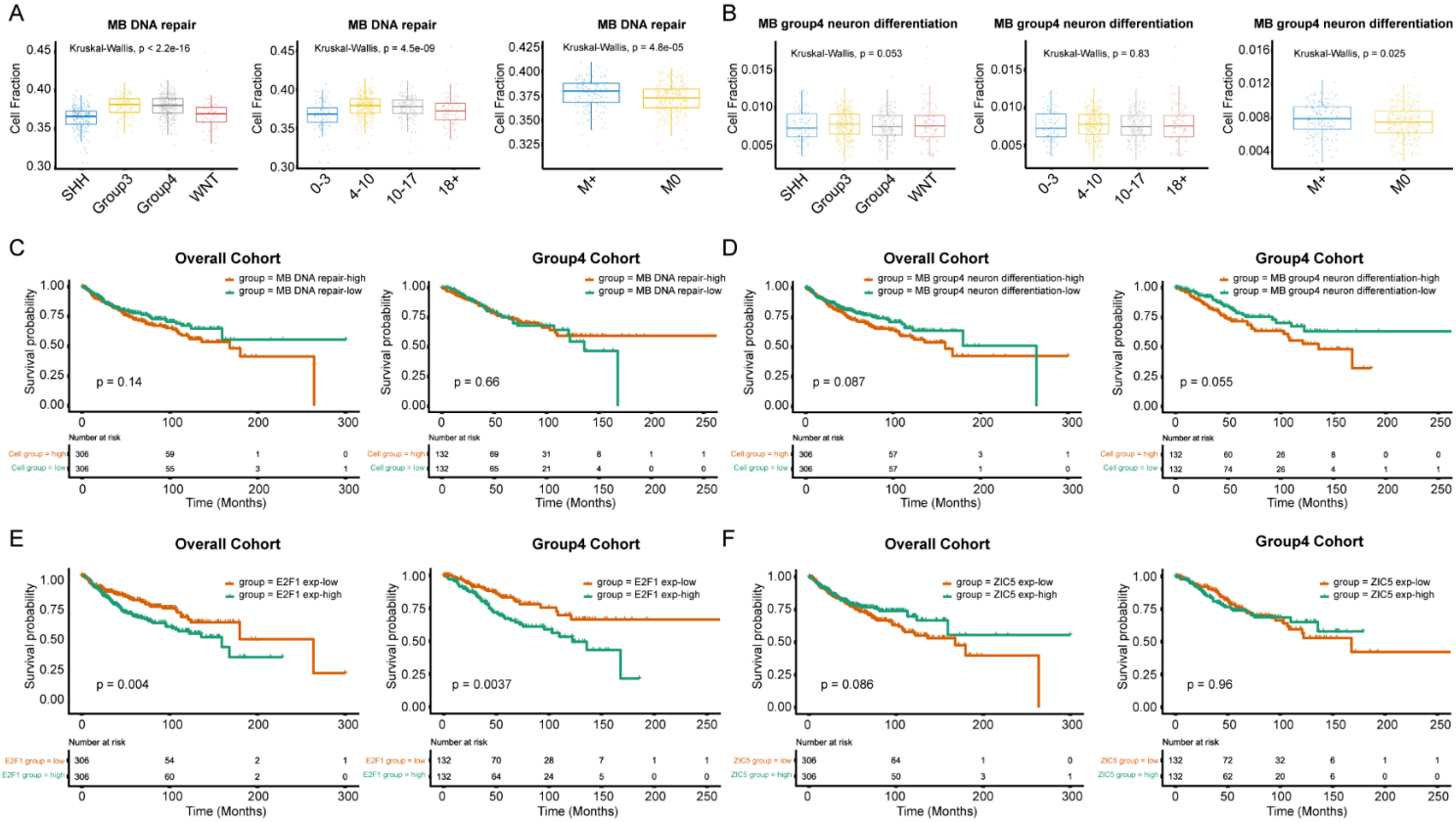


Figure 5

Clinical characteristics and prognostic evaluation of distinct cellular subpopulations and core driver genes. Boxplots illustrating the relative abundance of the MB DNA repair cell subpopulation (**A**) and MB group4 neuron differentiation cell subset (**B**) across different clinical strata. The deconvolutionally estimated fractions are evaluated stratified by tumor subgroups, patient age groups, and metastasis status. K-M overall survival curves based on the estimated fraction of MB DNA repair cells (**C**) and MB group4 neuron differentiation cell fraction (**D**). The prognostic evaluations are presented for both the overall clinical cohort and the predefined Group 4 sub-cohort. K-M survival analysis revealing the association between *E2F1* (**E**) and *ZIC5* (**F**) gene expression levels and total survival periods in both the entire cohort and Group 4 patients.

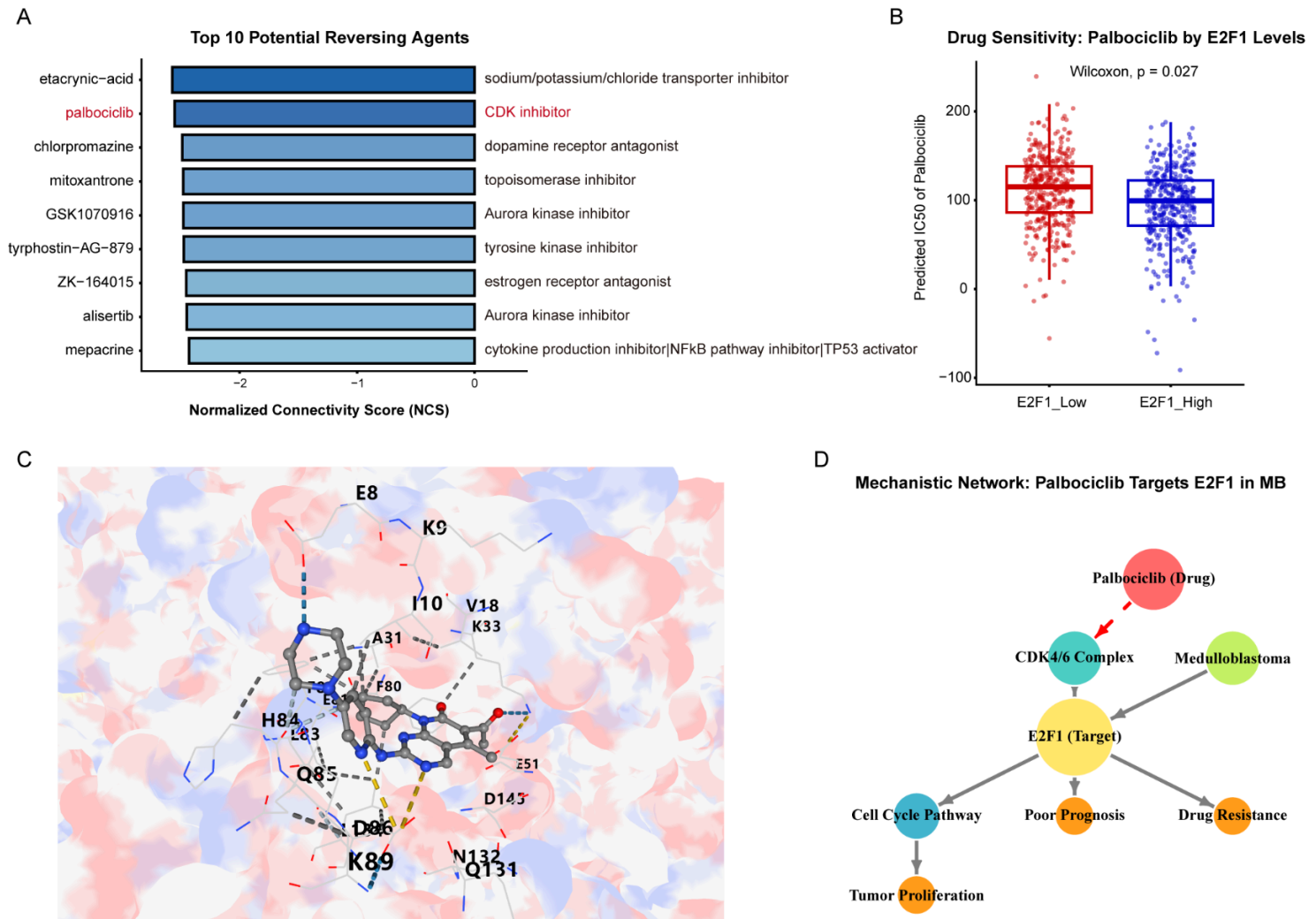


Figure 6

Identification and prediction of Palbociclib as a therapeutic agent for E2F1-driven MB. (**A**) Bar plot illustrating the top ten candidate pharmaceutical compounds predicted by the CMap database to reverse the disease transcriptomic signature. (**B**) Boxplot comparing the predicted IC50 of Palbociclib between the *E2F1* high and low expression groups based on the Cavalli dataset.

(**C**) Computational molecular docking model of Palbociclib interacting with the target protein (PDB ID: 1H24). (**D**) An integrative network diagram summarizing the mechanistic interrelationships among the target *E2F1*, clinical prognosis, biological pathways, and the specific intervention node of Palbociclib.