Assessing Cell Activities rather than Identities to Interpret Intra-Tumor Phenotypic Diversity and Its Dynamics

Laloé Monteiro, Lydie da Silva, Boris Lipinski, Frédérique Fauvet, Arnaud Vigneron, Alain Puisieux, Pierre Martinez

Version of Record: https://www.sciencedirect.com/science/article/pii/S2589004220302467
Manuscript_08983af14f2884ed9ea01bc006816df0

Assessing cell activities rather than identities to interpret intra-tumour phenotypic diversity and its dynamics

Laloé Monteiro[1+], Lydie Da Silva[1+], Boris Lipinski[1+], Frédérique Fauvet[1], Arnaud Vigneron[1], Alain Puisieux[1], Pierre Martinez[1,2]*.

[1]Univ Lyon, Université Claude Bernard Lyon 1, INSERM 1052, CNRS 5286, Centre Léon Bérard, Cancer Research Center of Lyon, Lyon, 69008, France, INSERM U1052, Cancer Research Center of Lyon, Lyon, F-69008, France.

[2]Lead Contact.

[+]These authors contributed equally.

*Correspondence: pierre.martinez@lyon.unicancer.fr

Abstract

Despite advances in single-cell and molecular techniques, it is still unclear how to best quantify phenotypic heterogeneity in cancer cells that evolved beyond normal, known classifications. We present an approach to phenotypically characterise cells based on their activities rather than static classifications. We validated the detectability of specific activities (Epithelial-mesenchymal transition, glycolysis) in single cells, using targeted RT-qPCR analyses and *in vitro* inductions. We analysed 50 established activity signatures as a basis for phenotypic description in public data, and computed cell-cell distances in 28,513 cells from 85 patients and 8 public datasets. Despite not relying on any classification, our measure correlated with standard diversity indices in populations of known structure. We identified bottlenecks as phenotypic diversity reduced upon colorectal cancer initiation. This suggests that focusing on what cancer cells do rather than what they are can quantify phenotypic diversity in universal fashion, to better understand and predict intra-tumour heterogeneity dynamics.

Introduction

Somatic evolution naturally occurs in all multicellular organisms, as cells accumulate genetic alterations upon replication and exposure to mutagenic environments (Gatenby and Brown, 2017). This can eventually select for highly adapted cells breaking free of the constraints imposed by homeostatic regulation on proliferation and motility, leading to cancer (Greaves and Maley, 2012; Trigos *et al.*, 2018). This evolutionary nature implies that cancer cells originating from a common ancestor can display extensive diversity at both the genetic and phenotypic level (Gerlinger *et al.*, 2012). This diversity, known as intra-tumour heterogeneity (ITH) (McGranahan and Swanton, 2015), can foster resistance and facilitate adaptation upon the environmental changes induced by therapeutic regimens (Nowell, 1976). To limit the risk of resistant populations emerging upon treatment and predict cancer evolution, it is thus necessary to better understand the dynamics of ITH (Maley *et al.*, 2006; Lässig, Mustonen and Walczak, 2017).

Being able to follow the evolution of ITH first implies to be able to reliably quantify it. Although there exist multiple methods for genetic ITH thanks to alteration frequencies in the population (Nik-Zainal *et al.*, 2012; Andor *et al.*, 2014; Fischer *et al.*, 2014; Martinez *et al.*, 2017; Williams *et al.*, 2018), phenotypic ITH is more challenging. Many studies have relied on the identification of static classifications (Frazer *et al.*, 2007; Patel *et al.*, 2014; Zhang *et al.*, 2019), often based on lineage markers (Almendro *et al.*, 2014; Nguyen *et al.*, 2018), allowing the calculation of standard diversity metrics such as the Shannon (Bertucci *et al.*, 2019), Simpson (Martinez *et al.*, 2016) or GINI (Ferrall-Fairbanks *et al.*, 2019) indices. While these classifications make perfect sense in the context of normal tissue homeostasis, they may not be relevant in cancer cells bypassing the host's regulatory mechanisms through abnormal transcriptional programmes. Cancer cells drift away from well-characterised normal phenotypes according to evolutionary trajectories specific to each tumour. They however display strong convergence at the phenotypic level, with key pathways and cellular activities recurrently dysregulated across both patients and tumour types (Hanahan and Weinberg, 2000, 2011). Aside from static subtype classifications, other methods have focused on expression variation among specific gene sets (Davis-Marcisak *et al.*, 2019) and uneven repartition of expressed transcripts per gene (Hinohara *et al.*, 2018). Yet, there is no golden standard approach to quantify phenotypic diversity in cancer.

Here we investigated the feasibility of predicting the activities that a single cell partakes in, and the relevance of considering them as traits to describe the cell's overall phenotypic profile. We performed targeted single-cell experiments on 3 cellular activities induced *in vitro* (Epithelial-mesenchymal transition, DNA repair, glycolysis), which suggested that targeted panels can reliably identify the presence of a given activity from single cell RNA expression data. To expand on this limited data, we then analysed 50 hallmark activity signatures from the Molecular Signature database (MSigDB) in 8 publicly available single-cell tumour datasets. We used leave-one-out procedures to avoid overfitting, along with Principal Component and clustering analyses to account for the redundancy among the 50 activities. By using activity-based phenotypic profiles to quantify cell-cell divergence and sample-wise phenotypic diversity, we report that such an approach is relevant in pan-cancer fashion. It could furthermore recapitulate diversity indices based on known population structures, independently of tissue and cell types. Finally, such a method allowed a glimpse into the evolutionary dynamics of phenotypic diversity, hinting at the existence of evolutionary bottlenecks reducing phenotypic diversity upon colorectal cancer initiation. Although more work is necessary to provide specific and accurate quantitative tools and software, our results suggest that focusing on cell activities to measure phenotypic ITH can provide a more relevant angle than standard classification and marker-based methods.

Results

Detecting hallmark signatures in single cells

We assessed the relevance of 3 MSigDB hallmark gene signatures in single cells via *in vitro* inductions: Epithelial-mesenchymal transition (EMT), DNA repair and glycolysis. We aimed to take advantage of the higher accuracy of single-cell RT-qPCR compared to whole transcriptome scRNA-seq (Mojtahedi *et al.*, 2016), and designed reduced panels of 9 to 13 marker genes to detect each activity in single cells (see Methods). To do so, we first analysed gene expression in 1,036 cell lines samples from the CCLE (Barretina *et al.*, 2012) for marker gene discovery and 10,885 pan-cancer samples from the TCGA (Chang *et al.*, 2013) for cross-validation. The activity-specific markers respectively achieved Areas Under the Curve (AUC) of 0.96, 086 and 0.79 in teasing out the top and bottom

scoring TCGA samples for EMT, DNA repair and glycolysis, respectively (Supplementary Table 3).This suggested that these reduced gene panels satisfactorily recapitulated the signal from whole-geneset enrichment analyses, implying that analysing the expression of few marker genes could help quantify the presence of activity-based phenotypic traits in single cells.

We analysed the expression of 48 selected markers genes in 48 single epithelial mammary cells (MCF10A), in which each activity had been induced or not (12 EMT-induced, 12 DNA-repair-induced, 12 Glycolysis-induced, 12 control cells with no-induction, Figure 1a). Significantly differentially expressed genes could be identified in all experiments (Figure 1b). We inferred Beta-Poisson expression distributions for each gene in active/inactive conditions, which we used to calculate the likelihood that expression values from marker genes corresponded to cells in which the related activity was induced (Figure 1c). Differentially expressed genes, generalised linear models and leave-one-out procedures were used to predict cells undergoing each activity induction (see Supplementary Methods). We could achieve AUCs of 0.99, 0.72 and 0.86 for respectively the EMT, DNA repair and glycolysis activities (Figure 1d, Supplementary Figures 2-3, Supplementary Table 4). The absence of expression patterns clearly separating DNA repair cells from the other 3 types, for most DNA repair genes, impaired prediction for this activity (Supplementary Figure 3). This targeted experiment however suggests that the expression of adequate marker genes can be used to identify whether an activity is present in a given cell with satisfying accuracy.

Whole transcriptome cell activity scores

Following these targeted *in vitro* results supporting the feasibility of predicting the activities of single cells, we investigated the relevance of an activity-centred approach to quantify phenotypic diversity in high-throughput patient datasets. In absence of single-cell inference methods tailored to each of the 50 hallmark cell activities, we used standard tools to investigate the behaviour of the related signatures in patient data. We used the AUCell (Aibar *et al.*, 2017) software to score the enrichment of all MSigDB hallmark gene sets in all cells from the 8 datasets. We normalised these data per set and merged them into a meta-dataset of 50 activity scores per cell in 28,513 cells from different cancer types (See Methods). No major batch effect could be observed as samples didn't specifically cluster according to their sets of origin , while similar cell types appear to cluster together (Figure 2). However, the most common cell types (T cells, macrophages and malignant cells) segregated into more than one cluster each. This suggests that cells with similar identity tend to behave similarly across batches and tissues, but that different subset of activity profiles could also be observed among cells of identical classification.

Our analysis however revealed extensive redundancy among the 50 activities scored (Figure 3a), suggesting that the signal from the hallmark signatures likely corresponded to fewer than 50 distinct activity-based phenotypic traits. We furthermore assigned cell cycle phases (G1/S/G2M) to cells using the cyclone software (Scialdone *et al.*, 2015). The cell-cycle phase in which a cell is influences its transcriptome, which can in turn bias cell-type assignment. However, because our approach is cancer-oriented and based on cellular activities rather than identities, we considered this information as part of the phenotypic state of a cell and purposely did not correct for it. Cell-cycle phase assignment was found to correlate with the *G2M Checkpoint*, *E2F Targets* and *Mitotic Spindle* signatures, highlighting that such cycle phase information was indeed taken into account in our phenotypic profiling of cells (Supplementary Figure 4).

Redundancy reduction to obtain phenotypic profiles

We designed two methods to tackle redundancy, based on Principal Component (PC) and clustering analyses (see Methods). The first 3 PCs of the entire meta-dataset respectively explained 25.9%,

12.8% and 7.7% of the variance in the data, while 11 PCs explained more than 2% of the variance (Figure 3b). For the clustering analyses, we investigated the relevance of splitting the data into 2 to 15 clusters. Using the consensus indices from bootstrapping experiments, we defined an optimal range between 6 and 10 clusters, after which increasing the number of clusters would not improve consensus (Figure3c-d).

We defined phenotypic profiles for each cell based on either the PC scores or the average activity scores per cluster. We analysed the 6 sets that provided metadata describing the predicted (sub)type of each cell (Filbin, Li, Neftel, Tirosh melanoma & oligodendroglioma, Venteicher), using leave-one-out procedures to prevent overfitting. In line with our observations that cells clustered according to their type rather than set of origin, defining PC weights and optimal cluster compositions on all sets but the one analysed still allowed to identify patterns differentiating cell types (Supplementary Figures 5-10).

Cell-cell divergence across tissue and cancer types

Pairwise Euclidean distances between phenotypic profiles then served to measure the phenotypic divergence between cells. We used different thresholds to calculate PCA- and cluster-based divergence, respectively based on the minimum percentage of variance for a PC to be included in phenotypic profiles (0, 1, 2, 3 and 5%), and on the numbers of clusters to summarise all 50 activities (6 to 10 clusters).Phenotypic heterogeneity measures were highly correlated regardless of the thresholds in both methods (all Spearman's rho >= 0.72, all p< 0.001, Supplementary Tables 5-5), suggesting they are nearly equivalent. However, we observed less redundancy between PC scores than between cluster scores, independently of the number of clusters (Supplementary Figure 10). We therefore use PCA-based phenotypic heterogeneity measures hereafter, with a 2% minimum threshold on explained variance for PC inclusion.

We investigated the pan-cancer relevance of our activity-based phenotypic divergence measure, using the 6 datasets for which cell type metadata were available. We report differences in cell-cell divergence distributions, according to whether two cells are of the same type or not and what that cell type is (Figure 4, Supplementary Figure 11). In agreement with our pan-cancer observations that cells clustered by type more than dataset, the divergence between cells of different cell types was always the highest distribution (compared to same-type distributions) in all 6 datasets. This suggests that our metric will assign smaller divergence scores to cells from the same cell type. Using bootstrapped clustering analyses, we also investigated if different recurrent activity profiles could be observed among cancer cells only, in each set (Supplementary Figures 12-16, see Methods). Clusters related to proliferation and immune response could be observed in most analyses, while the most discriminant activities, and PC scores derived from them, varied between datasets. In the Venteicher astrocytoma dataset, a discernible sub-population tied to immune activities can be distinguished on the left, with marked differences in interferon alpha and gamma signatures (Figure 5). A separate sub-population with strong proliferation signalling can be observed in the centre, whereas cells on the right side do not display particularly strong proliferation nor immune-related signal. This suggests that activity-based distances can separate distinct subpopulations of malignant cells presenting different phenotypic characteristics.

Phenotypic diversity quantification

We further analysed the relevance of activity-based approaches on two subsets with extended characterization in a large number of patients: 7 non-malignant cell types (Tcell, Bcell, Macrophage, Endothelial, Fibroblast, NK, Undefined) in 19 patients from the Tirosh melanoma dataset; 6 malignant subtypes (AC-like, OPC-like, MES1-like, MES2-like, NPC1-like, NPC2-like) in 28 patients from the

Neftel glioma dataset. The average divergence in a group of cells was used as a surrogate for the group's phenotypic heterogeneity. We observed differences across the average profiles calculated for the distinct cell types, suggesting they are each characterized by specific activity patterns. The differences between the most divergent cells in each category however exemplify that individual cells can strongly deviate from these overall profiles (Figure 6a-b, Supplementary Figure 17). Such variability, possibly due to the stochastic nature of gene expression, would be absent from standard classifying methods.

We proceeded to reclassify all cells according to the smallest Euclidean distance between their PCA-based profiles and the average profiles of each classification in both datasets. We observed a stronger concordance (p=0.022, Wilcoxon test) when reclassifying cells from established normal cell types in melanoma samples according to their activities (Figure 6c, 82% ± 14 correctly reclassified samples), compared to subtypes of malignant glioma cells (Figure 6d, 54% ± 23). This confirmed that cells of similar type tend to partake in similar activities. However, in the glioma samples we analysed, the differences between marker-based malignant subtypes were not as closely reflected by activity profiles as was observed in normal cell types.

We then computed the standard Simpson diversity index on a per-patient basis, according to the repartition of all cells from a patient into the relevant categories in both subsets. We found that it correlated very significantly with our divergence-based phenotypic heterogeneity score in both non-malignant cells from melanoma samples and malignant glioma cells (Figure 6e-f, Spearman's rho=0.73 and rho=0.49; p=0.001 and p=0.009, respectively). This suggests that this approach, although not relying on cell classification, can accurately capture the diversity of populations whose structure is known, both for malignant and normal cells from different tissues. Similar observations were reported using cluster-based distances (Supplementary Figure 18).

Using the average activity-based divergence between malignant cells, we quantified intra-tumour phenotypic heterogeneity in all samples from the 6 datasets with metadata and compared them (Figure 7a). The mean phenotypic divergence of colorectal cancers (Li *et al.*) was significantly higher than others datasets, while melanoma heterogeneity was significantly lower (Wilcoxon test, Benjamini-Hochberg correction, p<0.001 and p=0.004, respectively). We furthermore report that between-samples variation in phenotypic diversity was the highest in melanoma (*i.e.* most heterogeneous in heterogeneity levels) and the lowest in oligodendroglioma (Figure 7b-c).

Phenotypic diversity evolution

We finally took advantage of cancer samples paired with normal tissue in the colorectal dataset, to investigate the evolution of phenotypic diversity. In the 5 colorectal cancer patients from Li *et al.* for which we could find paired tumour-normal data, diversity stayed at similar levels in 3 cases (CRC04, CRC06, CRC10), while it decreased very significantly in the tumour material in 2 cases (Figure 7d, CRC05, CRC08, p<0.001, Wilcoxon test). Such decrease in diversity was not observed in other cell types in these patients (Supplementary Figure 19). This fits a scenario in which cells go through a phenotypic bottleneck at tumour initiation, followed by the expansion of few selected clones.

Discussion

Better understanding the dynamics of intra-tumour heterogeneity will help tailor better therapeutics to control and funnel cancer evolution. During malignant somatic evolution, cells drift away from their well-characterised normal ancestors by following trajectories unique to each patient (Tokutomi

et al., 2019), while there is convergence across patients to (de)activate the necessary cellular activities (Hanahan and Weinberg, 2000, 2011). Consequently, we investigated the relevance of focusing on what cancer cells do, rather than what they are, to measure phenotypic diversity in the cancer context. We considered cellular activities as traits describing the phenotypic state of cells and used pairwise distances to quantify cell-cell divergence and overall diversity. Unlike many existing methods (Almendro *et al.*, 2014; Ferrall-Fairbanks *et al.*, 2019; Zhang *et al.*, 2019), such an approach does not rely on classifying cells into putative, static identities that cancer cells drift away from in patient-specific fashion. It furthermore encompasses the temporal variability inherent to populations of cells replicating asynchronously and exhibiting stochastic differences in gene expression, which can itself foster resistance (Shaffer *et al.*, 2017). In addition, such a method is not tissue-type specific and was relevant in all investigated datasets.

We first performed *in vitro* analyses, which revealed that it was possible to reliably predict in which cells a given activity had been induced, using targeted panels based on the MSigDB hallmark gene sets and the literature. This was done using single-cell RT-qPCR technology, which is more precise than RNA-seq on specific genes of interest (Mojtahedi *et al.*, 2016). Our analysis however revealed that some of the best markers for activity detection were absent from the hallmark gene sets. Although this is likely to be attenuated when using entire gene sets rather than targeted panels, it exemplifies the need for more reliable gene signatures, particularly ones taking into account single-cell level specificities (Hwang, Lee and Bang, 2018; Larsson *et al.*, 2019).

We then scored 50 hallmark activity signatures in 28,513 cells from 8 publicly available datasets using the AUCell software. AUCell is based on a ranking procedure, which efficiently deals with normalisation and is not affected by the dissimilarity in using either FPKM or TPM units across the datasets (Aibar *et al.*, 2017). This was illustrated by cells not clustering according to their dataset of origin in the meta-dataset. “Dropouts” occurring when transcripts are not captured before sequencing, can however affect ranking in low-expressed genes (Davis-Marcisak *et al.*, 2019). Gene set enrichment analyses, in which multiple genes can contribute to the overall enrichment signal for an activity in each cell, are however less affected by dropouts than gene-specific differential expression analyses.

We reported high redundancy among the 50 activities scored, which we addressed by using PC and clustering analyses. We found that both methods were by and large equivalent. Importantly, hallmark activities do not focus on lineage-specific markers. Using their output, which summarises multiple genes, is thus less likely to separate cells according to the expression of few highly discriminating lineage markers, such as can occur when focusing on the entire transcriptome. This is particularly relevant for cancer cells that broke free of homeostatic control and differentiation hierarchies, in which lineage markers inherited from ancestors may no longer correlate with phenotype and behaviour.

We applied such an activity-based approach to investigate the divergence between and among cell types in 6 datasets with available metadata. We found that cells of the same type were less divergent than cells of different types. This can be explained by the fact that most reported cell types are non-malignant, with cells from the same type thus likely to partake in similar activities. We also observed that activity profiles recapitulated normal cell types better than malignant subtypes, although with very limited data (n=1 in both cases). Furthermore, we could identify distinct clusters of malignant cells showing marked differences in their activity profiles in all datasets. Therefore, although same cell identity often implied similar activities, it was not always the case, especially in cancer cells for which our activity-based approach was aimed. This also indicated that this approach could reflect the

divergence between cells similarly considered malignant using a blanket classification, but who appeared to engage in different activities.

Interestingly, the divergence between malignant cells were not recurrently higher nor lower than those between normal subtypes, and patterns varied according to tumour type. In the two datasets with high numbers of both patients and cell type sub-classifications, the mean cell-cell divergence correlated significantly with standard diversity indices based on the repartition of individuals into subpopulations. These results suggest that avoiding the use of known lineage markers did not hamper the relevance of this approach across the investigated tissue types. Although we used a leave-one-out design to avoid overfitting, it is however worth noticing that brain cell and tumour data are likely overrepresented in this study. Finally, using this approach on 5 patients with paired tumour-normal data, suggested the existence of evolutionary bottlenecks on phenotypic diversity at tumour initiation. This would be in agreement with the genetic diversity decrease observed at this stage in orthogonal studies (Cross *et al.*, 2020).

In this work, we focused on the quantification of phenotypic diversity according to cancer's atavistic evolutionary nature, as cells deviate from normal healthy cell types and regress towards ancestral unicellular growth (Davies and Lineweaver, 2011). We used single-cell expression analyses to quantify activity-based traits for each cell to create individual phenotypic profiles differing from static subtype classifications. This provides an alternative to marker-based methods, which can rely on markers not relevant anymore in the cancer context, and that often cannot allow to quantify the differences between cells classified similarly. Not relying on markers furthermore bypasses tissue-specificity and provides a universal approach applicable to all tumour types.

Limitations of the study

In this study we used pre-defined activity signatures based on bulk data that were not specifically designed for relevance in cancer studies. More work is therefore needed to provide standardised tools to reproducibly measure phenotypic ITH from single-cell RNA data. The development of accurate single-cell-specific expression signatures for the most recurrently dysregulated pathways in cancer would provide enhanced precision to build per-cell phenotypic profiles. This will require to determine the most relevant activities that contribute to the convergence towards the "cancer hallmarks" (Hanahan and Weinberg, 2011) dysregulation common to most cancer types. It will also be necessary to reliably assess their predictability in single cells, taking into account the specificity of single cell expression data, and design methods accounting for the redundancy among them. Finally, it will also be critical to understand how intra-tumour heterogeneity at single-cell level can be extrapolated from bulk samples, how this reflects inter-patient heterogeneity and how it ties to genetic and clinical features.

Successful implementations will improve future similar activity-based approaches to quantify phenotypic diversity in the evolutionary context of cancer. This will in turn allow to better monitor the evolution of phenotypic diversity over time and space, and facilitate the identification of therapeutic opportunities to control intra-tumour heterogeneity. This would ultimately help thwart the emergence of resistant populations and thereby enhance clinical outcomes.

Data and code availability

The R scripts and data for this project are available on github: https://github.com/pierremartinez/PhDiv.

Acknowledgements

The authors wish to thank Anne-Pierre Morel and Christelle Chassot for valuable discussions and assistance on the experimental setup. **Author contributions**

LM and FF performed *in vitro* experiments. FF, AV, AP and PM designed *in vitro* experiments. LDS, BL and PM performed bioinformatics analyses. AV, AP and PM supervised the work. PM wrote the manuscript. All authors revised the manuscript.

Competing interests

The authors declare no competing interests.

References

Aibar, S. *et al.* (2017) 'SCENIC: Single-cell regulatory network inference and clustering', *Nature Methods*. Nature Publishing Group, 14(11), pp. 1083–1086. doi: 10.1038/nmeth.4463.

Almendro, V. *et al.* (2014) 'Genetic and Phenotypic Diversity in Breast Tumor Metastases', *Cancer Research*, 74(5), pp. 1338–1348. doi: 10.1158/0008-5472.CAN-13-2357-T.

Andor, N. *et al.* (2014) 'EXPANDS: expanding ploidy and allele frequency on nested subpopulations.', *Bioinformatics (Oxford, England)*, 30(1), pp. 50–60. doi: 10.1093/bioinformatics/btt622.

Barretina, J. *et al.* (2012) 'The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity', *Nature*. Nature Research, 483(7391), pp. 603–307. doi: 10.1038/nature11003.

Bertucci, F. *et al.* (2019) 'Genomic characterization of metastatic breast cancers', *Nature*. Nature Publishing Group, 569(7757), pp. 560–564. doi: 10.1038/s41586-019-1056-z.

Chang, K. *et al.* (2013) 'The Cancer Genome Atlas Pan-Cancer analysis project', *Nature Genetics*. Nature Research, 45(10), pp. 1113–1120. doi: 10.1038/ng.2764.

Cross, W. *et al.* (2020) 'Stabilising selection causes grossly altered but stable karyotypes in metastatic colorectal cancer', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.03.26.007138. doi: 10.1101/2020.03.26.007138.

Davies, P. C. W. and Lineweaver, C. H. (2011) 'Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors.', *Physical biology*, 8(1), p. 015001. doi: 10.1088/1478-3975/8/1/015001.

Davis-Marcisak, E. F. *et al.* (2019) 'Differential variation analysis enables detection of tumor heterogeneity using single-cell RNA-sequencing data', *Cancer Research*. American Association for Cancer Research, p. canres.3882.2018. doi: 10.1158/0008-5472.can-18-3882.

Fan, J. *et al.* (2018) 'Linking transcriptional and genetic tumor heterogeneity through allele analysis of single-cell RNA-seq data', *Genome Research*. Cold Spring Harbor Laboratory Press, 28(8), pp. 1217–1227. doi: 10.1101/gr.228080.117.

Ferrall-Fairbanks, M. C. *et al.* (2019) 'Leveraging Single-Cell RNA Sequencing Experiments to Model Intratumor Heterogeneity', *JCO Clinical Cancer Informatics*. American Society of Clinical Oncology,

3(3), pp. 1–10. doi: 10.1200/CCI.18.00074.

Filbin, M. G. *et al.* (2018) 'Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq', *Science*. American Association for the Advancement of Science, 360(6386), pp. 331–335. doi: 10.1126/science.aao4750.

Fischer, A. *et al.* (2014) 'High-definition reconstruction of clonal composition in cancer.', *Cell reports*. Elsevier, 7(5), pp. 1740–52. doi: 10.1016/j.celrep.2014.04.055.

Frazer, K. A. *et al.* (2007) 'A second generation human haplotype map of over 3.1 million SNPs', *Nature*. Nature Publishing Group, 449(7164), pp. 851–861. doi: 10.1038/nature06258.

Gatenby, R. A. and Brown, J. (2017) 'Mutations, evolution and the central role of a self-defined fitness function in the initiation and progression of cancer', *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. doi: 10.1016/j.bbcan.2017.03.005.

Gerlinger, M. *et al.* (2012) 'Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.', *The New England journal of medicine*, 366(10), pp. 883–92. doi: 10.1056/NEJMoa1113205.

Greaves, M. and Maley, C. C. (2012) 'Clonal evolution in cancer.', *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 481(7381), pp. 306–13. doi: 10.1038/nature10762.

Hanahan, D. and Weinberg, R. A. (2000) 'The hallmarks of cancer.', *Cell*, 100(1), pp. 57–70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10647931 (Accessed: 17 May 2011).

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646–674. doi: 10.1016/j.cell.2011.02.013.

Hinohara, K. *et al.* (2018) 'KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance.', *Cancer cell*. Elsevier, 34(6), pp. 939-953.e9. doi: 10.1016/j.ccell.2018.10.014.

Hwang, B., Lee, J. H. and Bang, D. (2018) 'Single-cell RNA sequencing technologies and bioinformatics pipelines', *Experimental & Molecular Medicine*. Nature Publishing Group, 50(8), p. 96. doi: 10.1038/s12276-018-0071-8.

Larsson, A. J. M. *et al.* (2019) 'Genomic encoding of transcriptional burst kinetics', *Nature*. Nature Publishing Group, 565(7738), pp. 251–254. doi: 10.1038/s41586-018-0836-1.

Lässig, M., Mustonen, V. and Walczak, A. M. (2017) 'Predicting evolution', *Nature Ecology and Evolution*, 1(3). doi: 10.1038/s41559-017-0077.

Maley, C. C. *et al.* (2006) 'Genetic clonal diversity predicts progression to esophageal adenocarcinoma', *Nature Genetics*, 38(4), pp. 468–473. doi: 10.1038/ng1768.

Martinez, P. *et al.* (2016) 'Dynamic clonal equilibrium and predetermined cancer risk in Barrett's oesophagus', *Nature Communications*. Nature Publishing Group, 7, p. 12158. doi: 10.1038/ncomms12158.

Martinez, P. *et al.* (2017) 'Quantification of within-sample genetic heterogeneity from SNP-array data', *Scientific Reports*, 7(1), p. 3248. doi: 10.1038/s41598-017-03496-0.

McGranahan, N. and Swanton, C. (2015) 'Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution', *Cancer Cell*. Elsevier, 27(1), pp. 15–26. doi: 10.1016/j.ccell.2014.12.001.

Mojtahedi, M. *et al.* (2016) 'Cell Fate Decision as High-Dimensional Critical State Transition', *PLoS*

Biology. Allen Unwin Ltd, 14(12), p. e2000640. doi: 10.1371/journal.pbio.2000640.

Nguyen, Q. H. *et al.* (2018) 'Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity', *Nature Communications*, 9(1), p. 2028. doi: 10.1038/s41467-018-04334-1.

Nik-Zainal, S. *et al.* (2012) 'The life history of 21 breast cancers.', *Cell*, 149(5), pp. 994–1007. doi: 10.1016/j.cell.2012.04.023.

Nowell, P. C. (1976) 'The clonal evolution of tumor cell populations.', *Science (New York, N.Y.)*, 194(4260), pp. 23–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/959840 (Accessed: 18 October 2013).

Patel, A. P. *et al.* (2014) 'Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma', *Science*. American Association for the Advancement of Science, 344(6190), pp. 1396–1401. doi: 10.1126/science.1254257.

Scialdone, A. *et al.* (2015) 'Computational assignment of cell-cycle stage from single-cell transcriptome data', *Methods*. Academic Press, 85, pp. 54–61. doi: 10.1016/j.ymeth.2015.06.021.

Shaffer, S. M. *et al.* (2017) 'Reprogramming As a Mode of Cancer Drug Resistance', *Nature Publishing Group*. Macmillan Publishers Limited, part of Springer Nature. All rights reserved., 546(7658), pp. 431–435. doi: 10.1038/nature22794.

Tokutomi, N. *et al.* (2019) 'Quantifying local malignant adaptation in tissue-specific evolutionary trajectories by harnessing cancer's repeatability at the genetic level', *Evolutionary Applications*. John Wiley & Sons, Ltd (10.1111), 12(5), pp. 1062–1075. doi: 10.1111/eva.12781.

Trigos, A. S. *et al.* (2018) 'How the evolution of multicellularity set the stage for cancer.', *British journal of cancer*. Nature Publishing Group, 118(2), pp. 145–152. doi: 10.1038/bjc.2017.398.

Williams, M. J. *et al.* (2018) 'Quantification of subclonal selection in cancer from bulk sequencing data', *Nature Genetics*. Nature Publishing Group, 50(6), pp. 895–903. doi: 10.1038/s41588-018-0128-6.

Zhang, A. W. *et al.* (2019) 'Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling', *Nature Methods*. Nature Publishing Group, pp. 1–9. doi: 10.1038/s41592-019-0529-1.

Figure legends

Figure 1 – Detection of selected activities induced *in vitro* using single-cell expression of targeted genes. a) Overall scheme. EMT (blue), DNA repair (green) and glycolysis (red) activities are induced *in vitro* in MCF10A cells, prior to single-cell analysis and RNA quantification. Targeted marker genes expression is used to assess the likelihood that an activity, considered as a phenotypic trait, is present in a cell. All quantified traits are used to create cell-specific phenotypic profiles and serve as a basis to calculate pairwise cell-cell divergence and overall phenotypic diversity. b) Row-normalised single-cell expression for the marker genes of EMT (left), DNA repair (centre) and glycolysis (right). Blue: lower expression; red: higher expression. Cells in which the activity was induced are on the left and indicated by coloured bars below. Control cells having undergone no induction are on the right and indicated by a grey bar. Significantly differentially expressed genes in bold ($p<0.05$, BPglm function). c) PFKM marker gene expression in glycolysis and control conditions. Blue curve: number of transcripts in cells in which glycolysis cells was induced; grey curve: control conditions. Confidence intervals around the observed values are used to calculate the probability that a value comes from glycolytic (p_{glyco}, blue) and control (p_{ctrl}, grey) conditions. The p_{glyco} / (p_{glyco}+p_{ctrl}) ratio gives the likelihood that the observed value comes from a cell in which glycolysis was induced. d) Glycolysis prediction in single-cells from all 4 populations: glycolysis (red), EMT (blue) and DNA repair (green) inductions and control (grey). Black and white bar underneath indicates the reported probability of each cell to be glycolytic (log10 scale). Black: missing values.

Figure 2 – Normalised activity scores in the meta-dataset. Heatmap of activity scores in the metadataset, normalised per activity per set. Dendrograms highlight relationships between activities (left) and cells (top). The dataset of origin of each cell is reported by the bottom colour bar. The top row below the score heatmap indicates the dataset of origin of each cell, while the bottom one indicates its reported type.

Figure 3 – Principal Component and clustering analyses to circumvent hallmark activity redundancy. a) Correlation heatmap between all 50 MSigDB hallmark activities on a meta-dataset comprising 28,513 cells from 8 different datasets. b) Importance of the 15 Principal Components (PC) for each activity (squared cosine, indicated by increasing circle size and blueness). Below, the proportion of total variance in the dataset explained by each PC. c) Relative increase in measure of clustering consensus as the number of clusters is increased. CDF: cumulative distribution function. d) Cluster assignment of all 50 activities, for a number of 2 to 15 clusters.

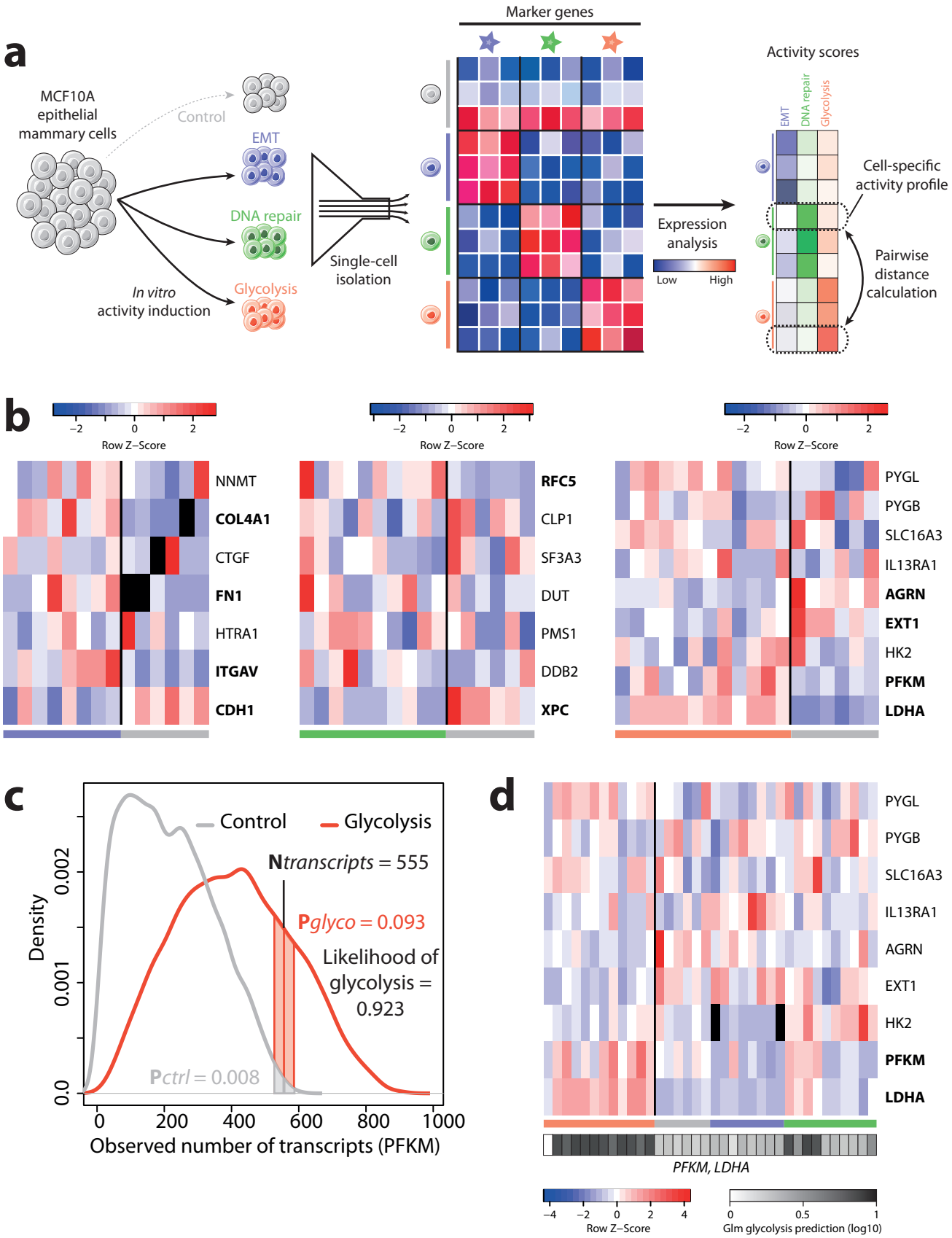
Figure 4 – Pan-cancer phenotypic cell-cell divergence. Pairwise cell-cell divergence distributions per cell type in each of the 6 datasets with curated metadata. Inter: inter-type divergence (between cells of different subtypes). All other distributions are between cells of the reported type. Dashed horizontal line: total average; broad horizontal lines: individual distribution averages.

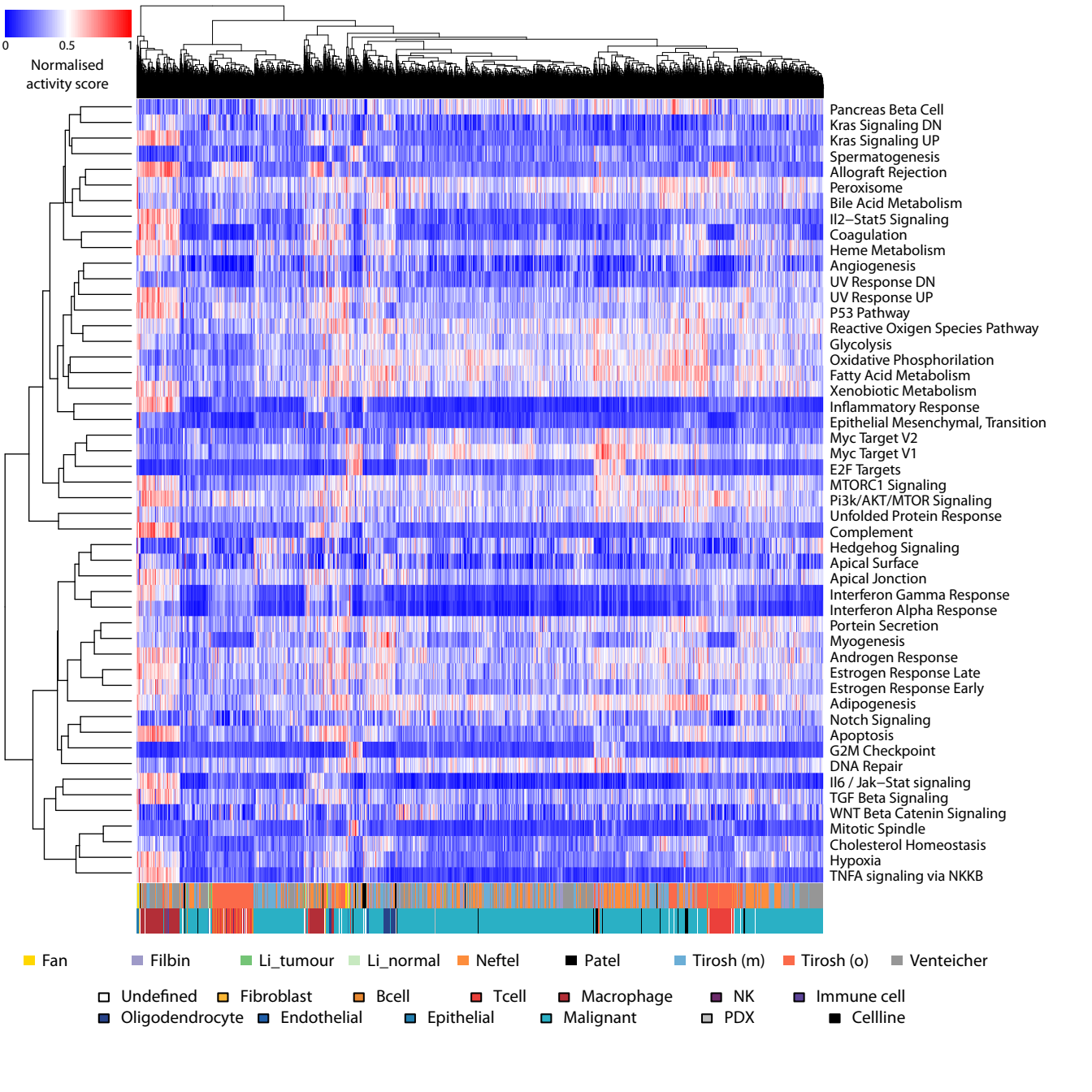
Figure 5 – Isolated activity profiles of significant clusters of malignant cells in the Venteicher *et al.* astrocytoma dataset. Top: distinct significant clusters are identified by alternating black and grey colour bars. Cells are ordered left to right according to the overall cluster data including all cells, although only significant clusters of 5 cells or more are displayed. Middle: Heatmap of PCA-based activity scores. All principal components were used for clustering analyses, but only those explaining >3% of total variance are displayed. PCA scores are ordered top to bottom according to complete hierarchical clustering based on Euclidean distances. Bottom: Heatmap of normalised activity scores, ordered top to bottom according to complete hierarchical clustering based on Euclidean distances.

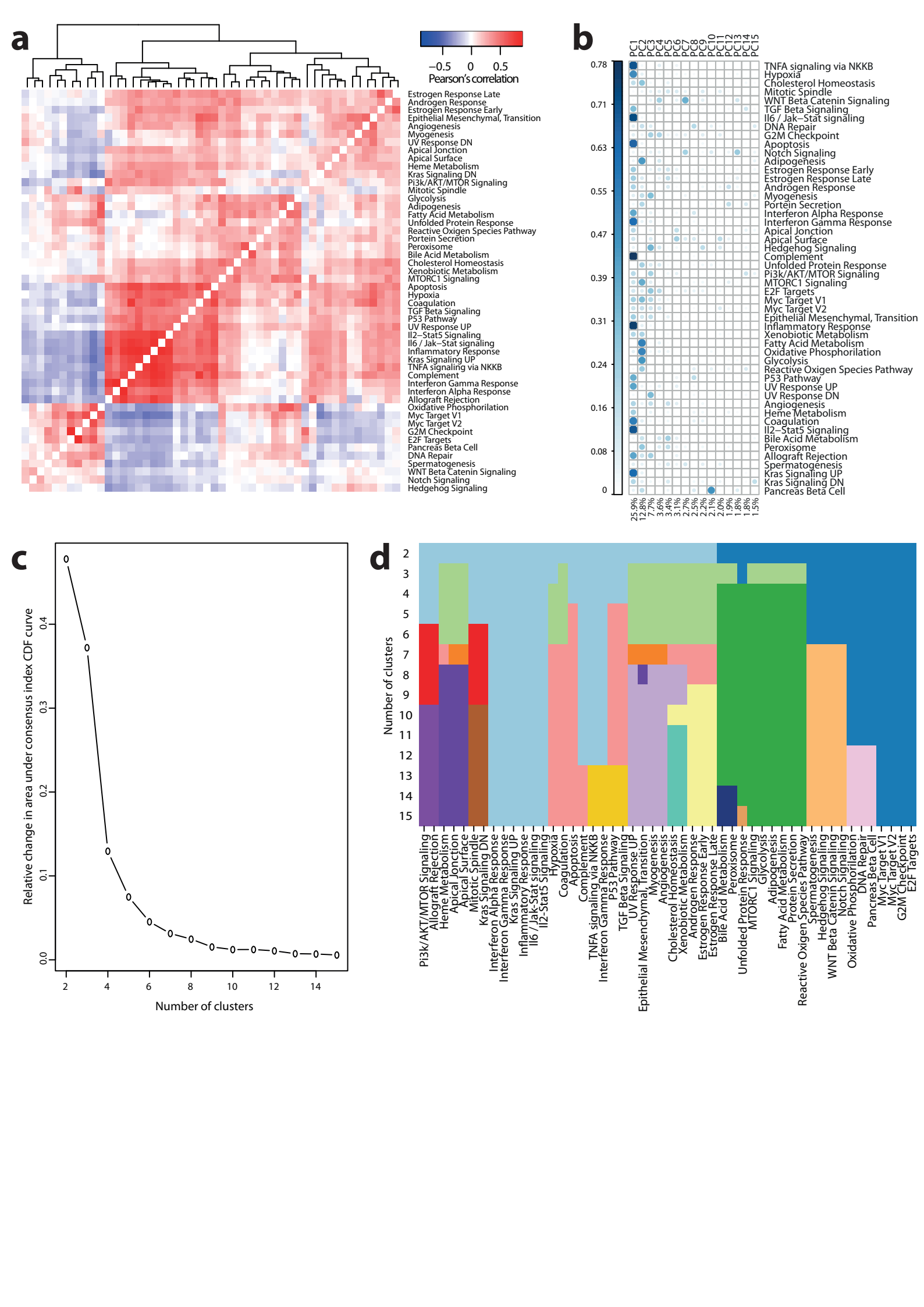
Figure 6 – Phenotypic diversity in populations of known structure. PCA-based phenotypic profiles of a) 7 non-malignant cell types from the Tirosh *et al.* melanoma dataset and b) 6 glioma subtypes from

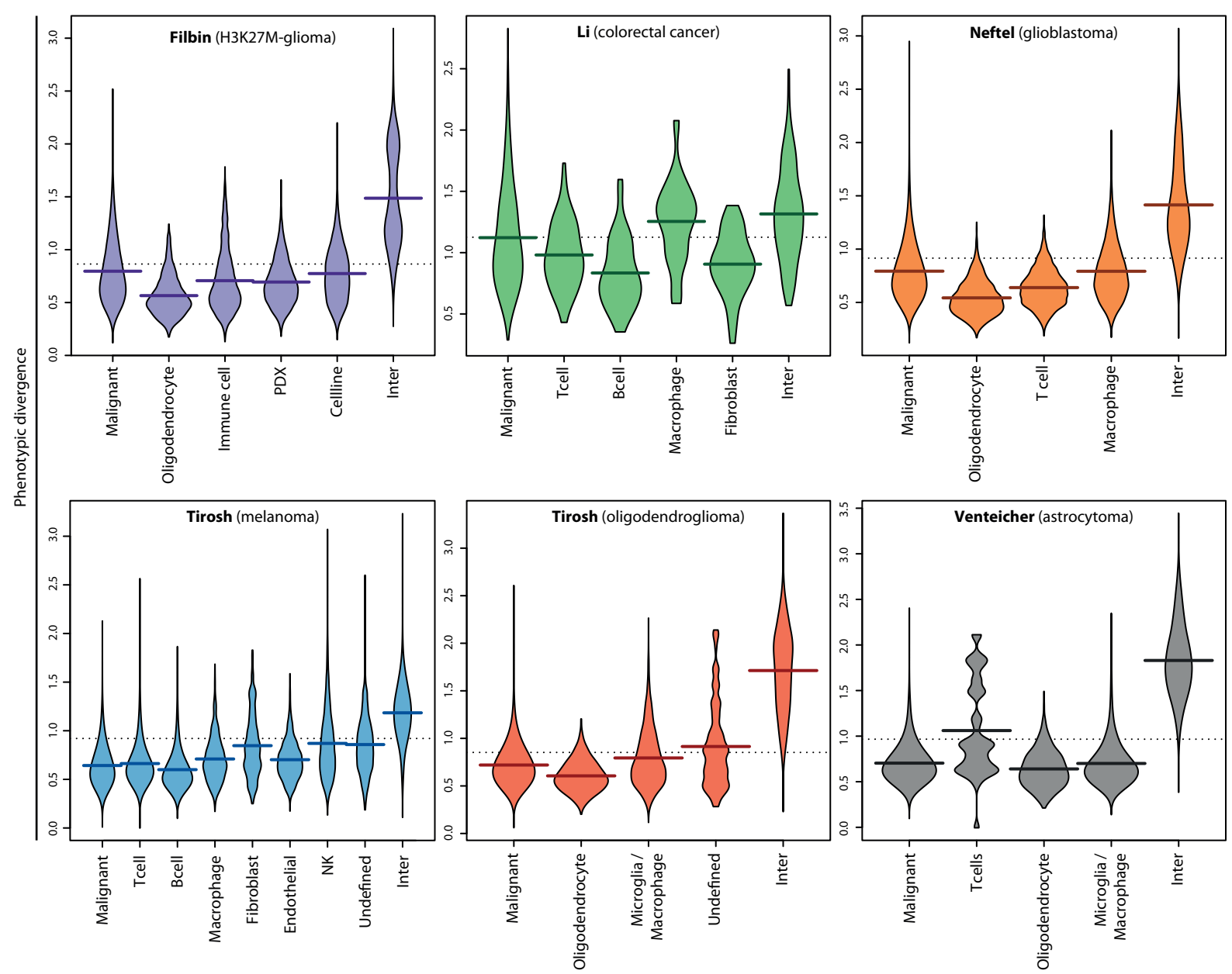
the Neftel *et al.* H3K27M-glioma dataset. Average profiles on top were obtained by averaging all cells from a given subtype across all patients. The outlier profiles at the bottom were obtained from the same-type cell pairs displaying the highest activity-based divergence for each cell type. Only the first five principal components are shown. c) Barplots showing the breakdown of how non-malignant cells from melanoma samples would be re-categorised, based on the average activity profiles of each category in the Tirosh melanoma dataset. d) Barplots showing the breakdown of how malignant glioblastoma cells would be re-categorised, based on the average activity profiles of each category in the Neftel dataset. e) Relationship between mean phenotypic divergence between non-malignant cells in the melanoma dataset and the Simpson diversity index calculated on the repartition of cells into the 7 non-malignant classes. f) Relationship between mean phenotypic divergence between malignant cells in the glioma dataset and the Simpson diversity index calculated on the classification of cells into the 6 glioma subtypes. Black lines: linear models.

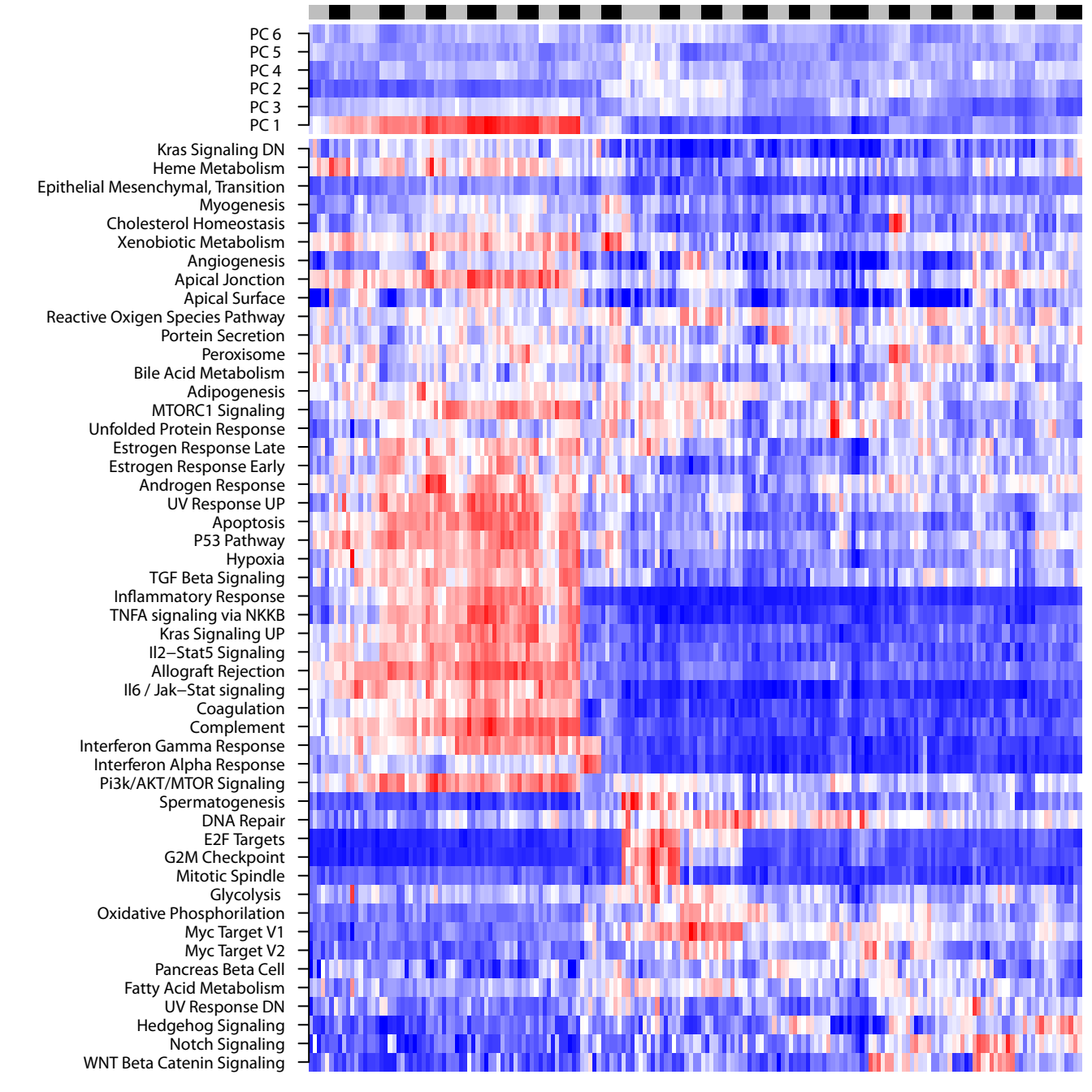
Figure 7 – Differences and dynamics of phenotypic diversity. a) Distribution of phenotypic divergence between malignant cells in each sample across 6 datasets. Samples ordered by sample-wise phenotypic diversity (average divergence). ***: $p<0.001$; *: $p<0.05$ (Wilcoxon test, BH correction). Boxes represent the middle quartiles; black horizontal bars represent the median of each distribution; whiskers extend up to 1.5 times the interquartile range (box height) away from the box. Outliers (beyond the whiskers) are not displayed. b) Per-sample phenotypic diversity in all 6 sets. c) Coefficient of variation in phenotypic diversity across samples in each set. d) Phenotypic divergence distributions in normal and cancerous epithelia in 5 patients from the Li *et al.* dataset. Dashed horizontal line: total average; broad horizontal lines: individual distribution averages.

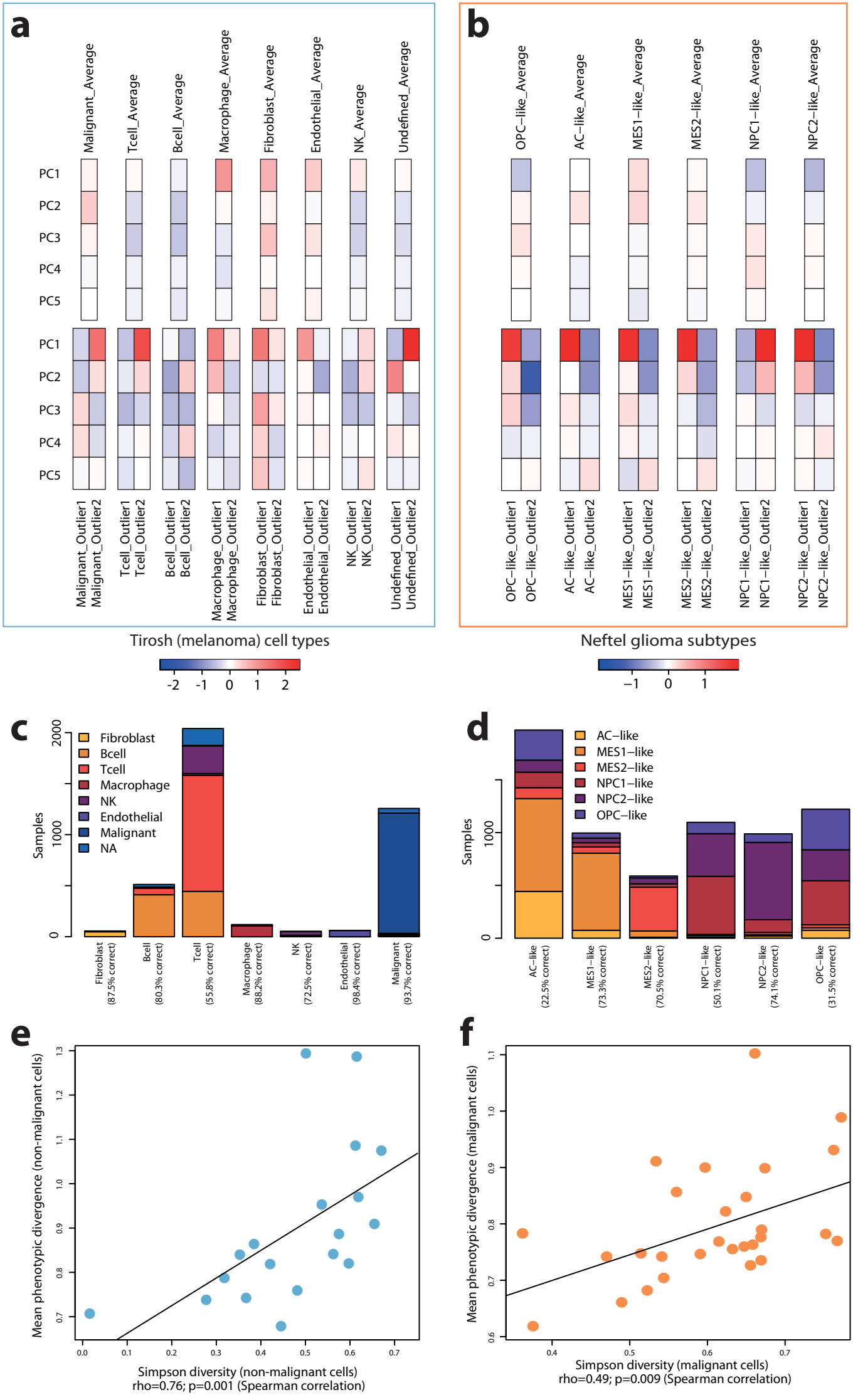

a
MCF10A epithelial mammary cells
Control
EMT
DNA repair
Glycolysis
In vitro activity induction
Single-cell isolation
Marker genes
Expression analysis
Low
High
Activity scores
EMT
DNA repair
Glycolysis
Cell-specific activity profile
Pairwise distance calculation
b
Row Z–Score
NNMT
COL4A1
CTGF
FN1
HTRA1
ITGAV
CDH1
RFC5
CLP1
SF3A3
DUT
PMS1
DDB2
XPC
PYGL
PYGB
SLC16A3
IL13RA1
AGRN
EXT1
HK2
PFKM
LDHA
c
Control
Glycolysis
Ntranscripts = 555
Pglyco = 0.093
Likelihood of glycolysis = 0.923
Pctrl = 0.008
Density
Observed number of transcripts (PFKM)
d
PFKM, LDHA
Row Z–Score
Glm glycolysis prediction (log10)

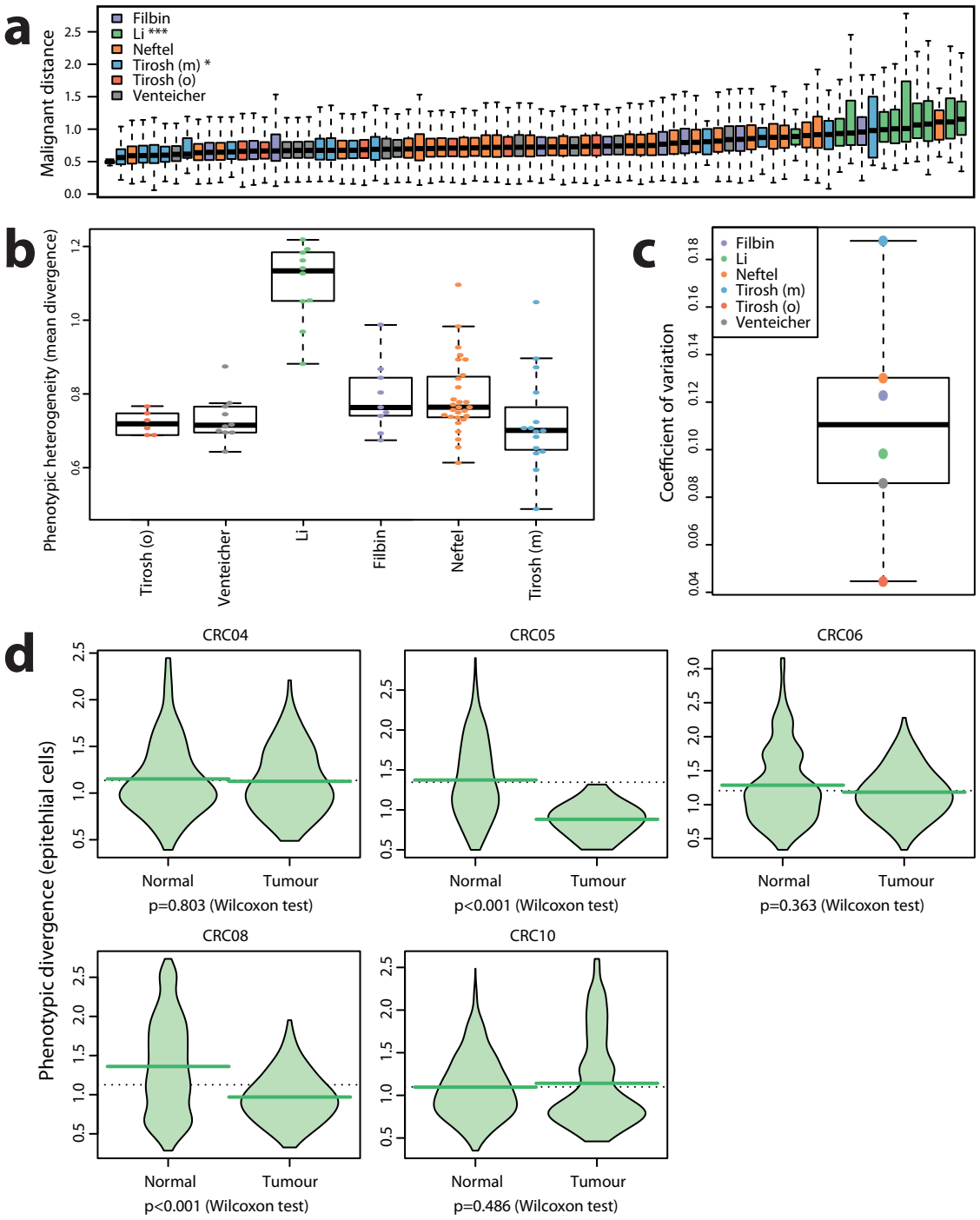

0
0.5
1
Normalised activity score
Pancreas Beta Cell
Kras Signaling DN
Kras Signaling UP
Spermatogenesis
Allograft Rejection
Peroxisome
Bile Acid Metabolism
Il2–Stat5 Signaling
Coagulation
Heme Metabolism
Angiogenesis
UV Response DN
UV Response UP
P53 Pathway
Reactive Oxigen Species Pathway
Glycolysis
Oxidative Phosphorilation
Fatty Acid Metabolism
Xenobiotic Metabolism
Inflammatory Response
Epithelial Mesenchymal, Transition
Myc Target V2
Myc Target V1
E2F Targets
MTORC1 Signaling
Pi3k/AKT/MTOR Signaling
Unfolded Protein Response
Complement
Hedgehog Signaling
Apical Surface
Apical Jonction
Interferon Gamma Response
Interferon Alpha Response
Portein Secretion
Myogenesis
Androgen Response
Estrogen Response Late
Estrogen Response Early
Adipogenesis
Notch Signaling
Apoptosis
G2M Checkpoint
DNA Repair
Il6 / Jak–Stat signaling
TGF Beta Signaling
WNT Beta Catenin Signaling
Mitotic Spindle
Cholesterol Homeostasis
Hypoxia
TNFA signaling via NKKB
Fan
Filbin
Li_tumour
Li_normal
Neftel
Patel
Tirosh (m)
Tirosh (o)
Venteicher
Undefined
Fibroblast
Bcell
Tcell
Macrophage
NK
Immune cell
Oligodendrocyte
Endothelial
Epithelial
Malignant
PDX
Cellline


a
−0.5 0 0.5
Pearson's correlation
Estrogen Response Late
Androgen Response
Estrogen Response Early
Epithelial Mesenchymal, Transition
Angiogenesis
Myogenesis
UV Response DN
Apical Jonction
Apical Surface
Heme Metabolism
Kras Signaling DN
Pi3k/AKT/MTOR Signaling
Mitotic Spindle
Glycolysis
Adipogenesis
Fatty Acid Metabolism
Unfolded Protein Response
Reactive Oxigen Species Pathway
Portein Secretion
Peroxisome
Bile Acid Metabolism
Cholesterol Homeostasis
Xenobiotic Metabolism
MTORC1 Signaling
Apoptosis
Hypoxia
Coagulation
TGF Beta Signaling
P53 Pathway
UV Response UP
Il2−Stat5 Signaling
Il6 / Jak−Stat signaling
Inflammatory Response
Kras Signaling UP
TNFA signaling via NKKB
Complement
Interferon Gamma Response
Interferon Alpha Response
Allograft Rejection
Oxidative Phosphorilation
Myc Target V1
Myc Target V2
G2M Checkpoint
E2F Targets
Pancreas Beta Cell
DNA Repair
Spermatogenesis
WNT Beta Catenin Signaling
Notch Signaling
Hedgehog Signaling
b
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 PC10 PC11 PC12 PC13 PC14 PC15
0.78 0.71 0.63 0.55 0.47 0.39 0.31 0.24 0.16 0.08 0
TNFA signaling via NKKB
Hypoxia
Cholesterol Homeostasis
Mitotic Spindle
WNT Beta Catenin Signaling
TGF Beta Signaling
Il6 / Jak−Stat signaling
DNA Repair
G2M Checkpoint
Apoptosis
Notch Signaling
Adipogenesis
Estrogen Response Early
Estrogen Response Late
Androgen Response
Myogenesis
Portein Secretion
Interferon Alpha Response
Interferon Gamma Response
Apical Jonction
Apical Surface
Hedgehog Signaling
Complement
Unfolded Protein Response
Pi3k/AKT/MTOR Signaling
MTORC1 Signaling
E2F Targets
Myc Target V1
Myc Target V2
Epithelial Mesenchymal, Transition
Inflammatory Response
Xenobiotic Metabolism
Fatty Acid Metabolism
Oxidative Phosphorilation
Glycolysis
Reactive Oxigen Species Pathway
P53 Pathway
UV Response UP
UV Response DN
Angiogenesis
Heme Metabolism
Coagulation
Il2−Stat5 Signaling
Bile Acid Metabolism
Peroxisome
Allograft Rejection
Spermatogenesis
Kras Signaling UP
Kras Signaling DN
Pancreas Beta Cell
25.9% 12.8% 2.7% 2.7% 3.6% 3.4% 3.1% 2.7% 2.5% 2.2% 2.1% 2.0% 1.9% 1.8% 1.8% 1.5%
c
Relative change in area under consensus index CDF curve
0.0 0.1 0.2 0.3 0.4
2 4 6 8 10 12 14
Number of clusters
d
Number of clusters
2 3 4 5 6 7 8 9 10 11 12 13 14 15
Pi3k/AKT/MTOR Signaling
Allograft Rejection
Heme Metabolism
Apical Jonction
Apical Surface
Mitotic Spindle
Kras Signaling DN
Interferon Alpha Response
Interferon Gamma Response
Kras Signaling UP
Inflammatory Response
Il6 / Jak−Stat signaling
Il2−Stat5 Signaling
Hypoxia
Coagulation
Apoptosis
Complement
TNFA signaling via NKKB
Interferon Gamma Response
P53 Pathway
TGF Beta Signaling
UV Response UP
Epithelial Mesenchymal, Transition
Myogenesis
Angiogenesis
Cholesterol Homeostasis
Xenobiotic Metabolism
Androgen Response
Estrogen Response Early
Estrogen Response Late
Bile Acid Metabolism
Peroxisome
Unfolded Protein Response
MTORC1 Signaling
Glycolysis
Adipogenesis
Fatty Acid Metabolism
Protein Secretion
Reactive Oxigen Species Pathway
Spermatogenesis
Hedgehog Signaling
WNT Beta Catenin Signaling
Notch Signaling
Oxidative Phosphorilation
DNA Repair
Pancreas Beta Cell
Myc Target V1
Myc Target V2
G2M Checkpoint
E2F Targets


Phenotypic divergence
Filbin (H3K27M-glioma)
Malignant
Oligodendrocyte
Immune cell
PDX
Cellline
Inter
Li (colorectal cancer)
Malignant
Tcell
Bcell
Macrophage
Fibroblast
Inter
Neftel (glioblastoma)
Malignant
Oligodendrocyte
T cell
Macrophage
Inter
Tirosh (melanoma)
Malignant
Tcell
Bcell
Macrophage
Fibroblast
Endothelial
NK
Undefined
Inter
Tirosh (oligodendroglioma)
Malignant
Oligodendrocyte
Microglia / Macrophage
Undefined
Inter
Venteicher (astrocytoma)
Malignant
Tcells
Oligodendrocyte
Microglia / Macrophage
Inter


PC 6
PC 5
PC 4
PC 2
PC 3
PC 1
Kras Signaling DN
Heme Metabolism
Epithelial Mesenchymal, Transition
Myogenesis
Cholesterol Homeostasis
Xenobiotic Metabolism
Angiogenesis
Apical Jonction
Apical Surface
Reactive Oxigen Species Pathway
Portein Secretion
Peroxisome
Bile Acid Metabolism
Adipogenesis
MTORC1 Signaling
Unfolded Protein Response
Estrogen Response Late
Estrogen Response Early
Androgen Response
UV Response UP
Apoptosis
P53 Pathway
Hypoxia
TGF Beta Signaling
Inflammatory Response
TNFA signaling via NKKB
Kras Signaling UP
Il2−Stat5 Signaling
Allograft Rejection
Il6 / Jak−Stat signaling
Coagulation
Complement
Interferon Gamma Response
Interferon Alpha Response
Pi3k/AKT/MTOR Signaling
Spermatogenesis
DNA Repair
E2F Targets
G2M Checkpoint
Mitotic Spindle
Glycolysis
Oxidative Phosphorilation
Myc Target V1
Myc Target V2
Pancreas Beta Cell
Fatty Acid Metabolism
UV Response DN
Hedgehog Signaling
Notch Signaling
WNT Beta Catenin Signaling


a
Malignant_Average
Tcell_Average
Bcell_Average
Macrophage_Average
Fibroblast_Average
Endothelial_Average
NK_Average
Undefined_Average
PC1
PC2
PC3
PC4
PC5
Malignant_Outlier1
Malignant_Outlier2
Tcell_Outlier1
Tcell_Outlier2
Bcell_Outlier1
Bcell_Outlier2
Macrophage_Outlier1
Macrophage_Outlier2
Fibroblast_Outlier1
Fibroblast_Outlier2
Endothelial_Outlier1
Endothelial_Outlier2
NK_Outlier1
NK_Outlier2
Undefined_Outlier1
Undefined_Outlier2
Tirosh (melanoma) cell types
-2 -1 0 1 2
b
OPC−like_Average
AC−like_Average
MES1−like_Average
MES2−like_Average
NPC1−like_Average
NPC2−like_Average
OPC−like_Outlier1
OPC−like_Outlier2
AC−like_Outlier1
AC−like_Outlier2
MES1−like_Outlier1
MES1−like_Outlier2
MES2−like_Outlier1
MES2−like_Outlier2
NPC1−like_Outlier1
NPC1−like_Outlier2
NPC2−like_Outlier1
NPC2−like_Outlier2
Neftel glioma subtypes
−1 0 1
c
Samples
0
1000
2000
Fibroblast
Bcell
Tcell
Macrophage
NK
Endothelial
Malignant
NA
Fibroblast (87.5% correct)
Bcell (80.3% correct)
Tcell (55.8% correct)
Macrophage (88.2% correct)
NK (72.5% correct)
Endothelial (98.4% correct)
Malignant (93.7% correct)
d
Samples
0
1000
AC−like
MES1−like
MES2−like
NPC1−like
NPC2−like
OPC−like
AC−like (22.5% correct)
MES1−like (73.3% correct)
MES2−like (70.5% correct)
NPC1−like (50.1% correct)
NPC2−like (74.1% correct)
OPC−like (31.5% correct)
e
Mean phenotypic divergence (non-malignant cells)
0.7 0.8 0.9 1.0 1.1 1.2 1.3
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Simpson diversity (non-malignant cells)
rho=0.76; p=0.001 (Spearman correlation)
f
Mean phenotypic divergence (malignant cells)
0.6 0.7 0.8 0.9 1.0 1.1
0.4 0.5 0.6 0.7
Simpson diversity (malignant cells)
rho=0.49; p=0.009 (Spearman correlation)


a
Malignant distance
Filbin
Li ***
Neftel
Tirosh (m) *
Tirosh (o)
Venteicher
b
Phenotypic heterogeneity (mean divergence)
Tirosh (o)
Venteicher
Li
Filbin
Neftel
Tirosh (m)
c
Coefficient of variation
Filbin
Li
Neftel
Tirosh (m)
Tirosh (o)
Venteicher
d
Phenotypic divergence (epitehlial cells)
CRC04
CRC05
CRC06
CRC08
CRC10
Normal
Tumour
p=0.803 (Wilcoxon test)
p<0.001 (Wilcoxon test)
p=0.363 (Wilcoxon test)
p<0.001 (Wilcoxon test)
p=0.486 (Wilcoxon test)

MCF10A epithelial mammary cells

Control

EMT

DNA repair

Glycolysis

In vitro activity induction

Single-cell isolation

Marker genes

Activity scores

Expression analysis

Low

High

EMT

DNA repair

Glycolysis

Cell-specific activity profile

Pairwise distance calculation