**Supplementary Information for**

# Endosomal pH is an evolutionarily conserved driver of phenotypic plasticity in colorectal cancer

Hari Prasad[1*#], Harshavardhan BV[2], Ayalur Raghu Subbalakshmi[2], Susmita Mandal[2], Mohit Kumar Jolly[2], Sandhya S. Visweswariah[1]

[1]Department of Developmental Biology and Genetics, Indian Institute of Science, Bengaluru, 560012, India.

[2]Department of Bioengineering, Indian Institute of Science, Bengaluru, 560012, India.

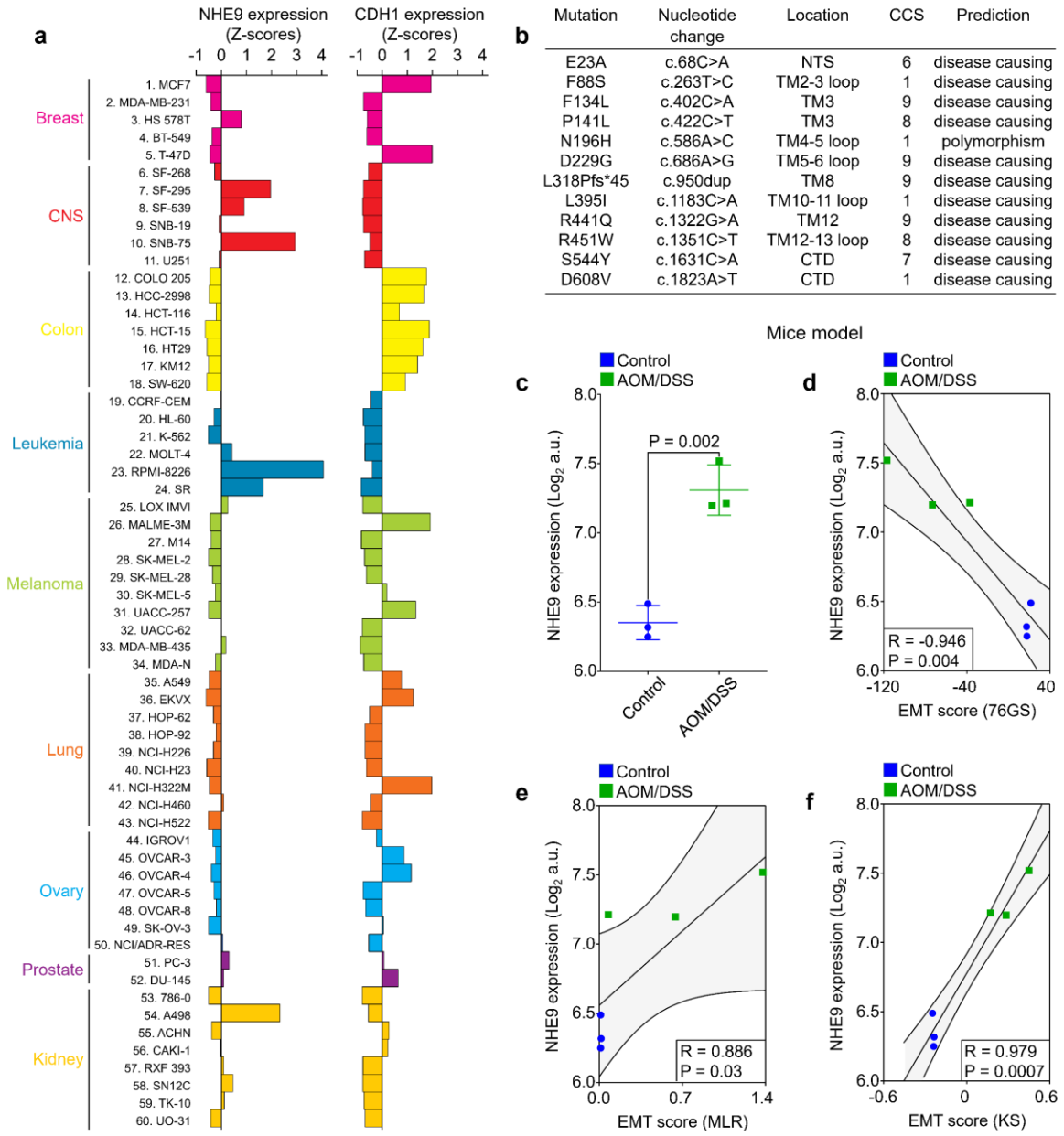
[*]Correspondence to Dr. Hari Prasad, Indian Institute of Science, Bengaluru, 560012, India.
hariprasad@iisc.ac.in

[#]Present address: Centre for Brain Research, Indian Institute of Science, Bengaluru, 560012, India.
harip@cbr-iisc.ac.in

Running title: Dysregulated endosomal pH fuels cancer progression.

## Tables of Contents

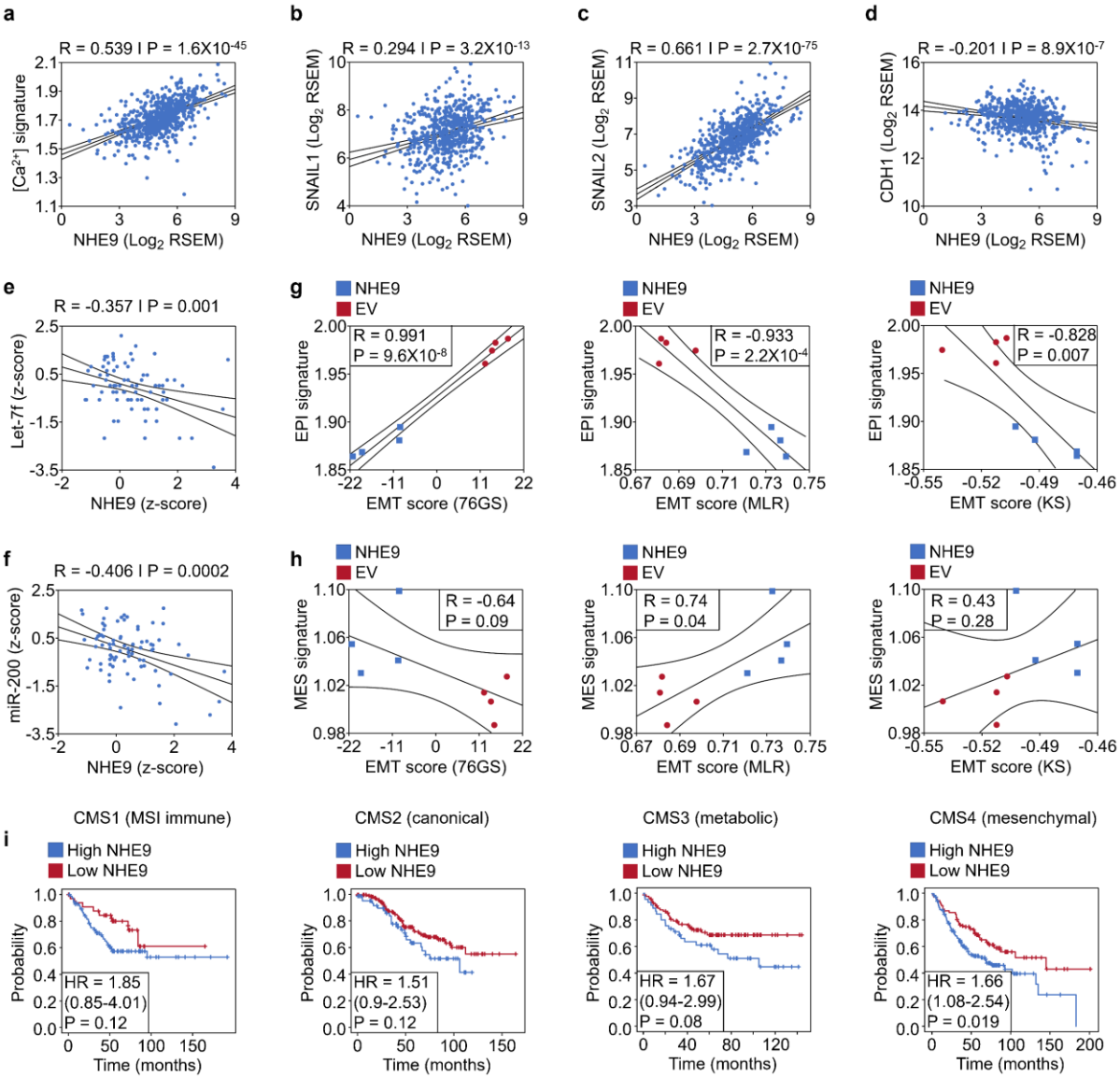
**Supplementary Figure 1: NHE9 correlates with the loss of epithelial character.**




| Mutation | Nucleotide change | Location | CCS | Prediction |
|---|---|---|---|---|
| E23A | c.68C>A | NTS | 6 | disease causing |
| F88S | c.263T>C | TM2-3 loop | 1 | disease causing |
| F134L | c.402C>A | TM3 | 9 | disease causing |
| P141L | c.422C>T | TM3 | 8 | disease causing |
| N196H | c.586A>C | TM4-5 loop | 1 | polymorphism |
| D229G | c.686A>G | TM5-6 loop | 9 | disease causing |
| L318Pfs*45 | c.950dup | TM8 | 9 | disease causing |
| L395I | c.1183C>A | TM10-11 loop | 1 | disease causing |
| R441Q | c.1322G>A | TM12 | 9 | disease causing |
| R451W | c.1351C>T | TM12-13 loop | 8 | disease causing |
| S544Y | c.1631C>A | CTD | 7 | disease causing |
| D608V | c.1823A>T | CTD | 1 | disease causing |

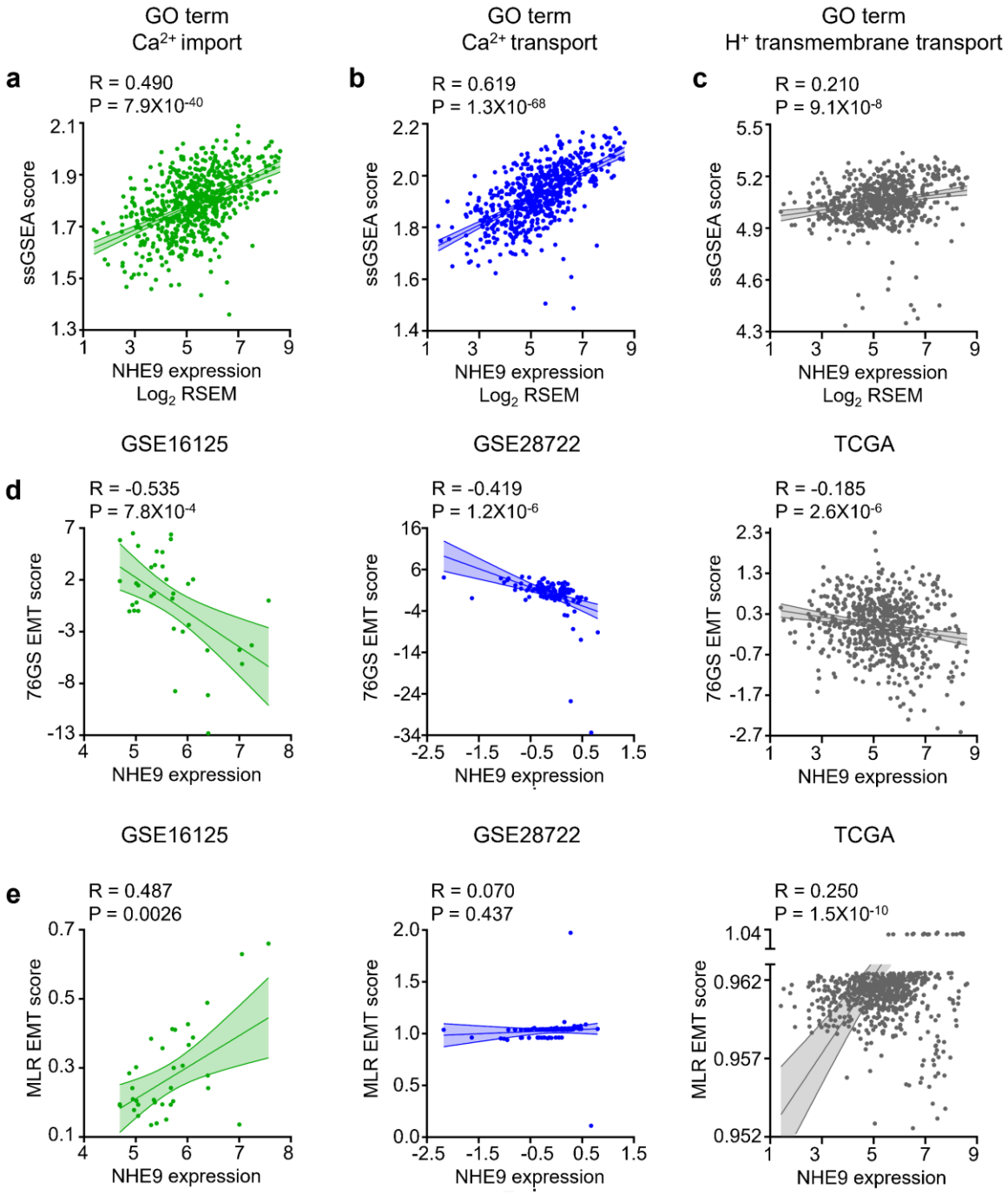
(**a**) NCI-60 gene expression profiles for NHE9 and the adherens junction gene *CDH1*, which encodes the epithelial marker e-cadherin. Note that nearly all of the cell lines that up-regulate NHE9 down-regulate CDH1. (**b**) Survey of NHE9 mutations identified in hypermutated CRC samples from the TCGA database. The ConSurf analysis was used to determine the evolutionary conservation of mutated residues on a scale ranging from 1 (highly variable) to 9 (invariant), while the MutationTaster analysis was used to predict functional impacts. (**c**) NHE9 expression in AOM/DSS mice model of CRC relative to healthy control samples at 2 weeks (GSE31106) shown as scatter plots displaying mean±SD. (**d-f**) Scatter plots depicting the relationship between NHE9 expression and 76GS (**d**), MLR (**e**), and KS (**f**) EMT scores in AOM/DSS mice model of CRC and healthy control samples (GSE31106). P values were calculated by unpaired, two-tailed *t* test (**c**) and Pearson's correlation (**d,e,f**). Abbreviations: NTS, N-terminal segment; TM, transmembrane segment; CTD, C-terminal domain; CCS, ConSurf conservation score; AOM/DSS, azoxymethane/dextran sodium sulfate. Relates to Figure 1.

**Supplementary Figure 2: NHE9 promotes the acquisition of a hybrid E/M state.**



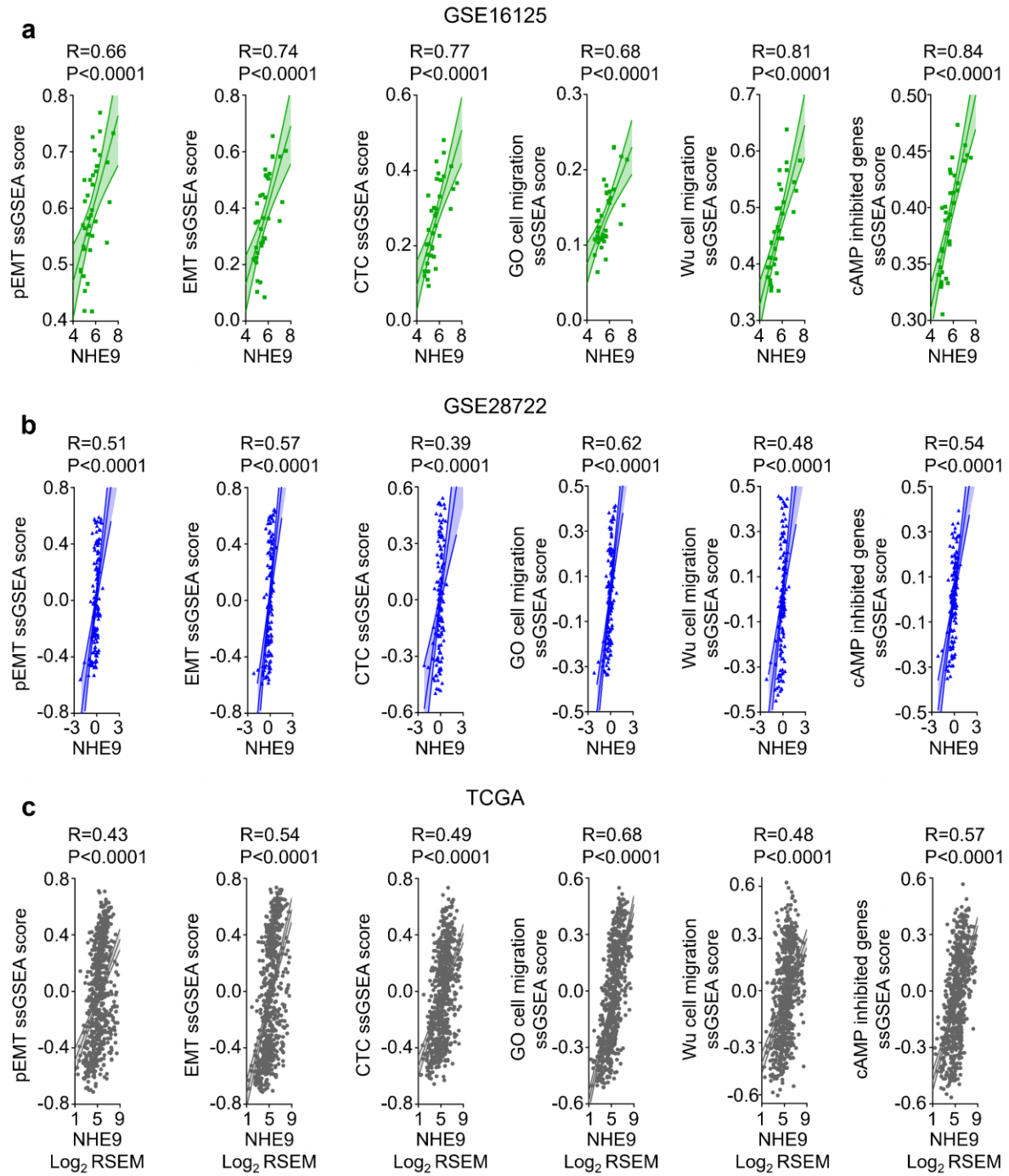

(**a-d**) Scatter plots illustrating the relationship between NHE9 expression and [$Ca^{2+}$] signature scores (as determined by ssGSEA) (**a**), SNAIL1 expression (**b**), SNAIL2 expression (**c**), and CDH1 expression (**d**) in CRC samples from the TCGA RNAseq database. (**e-f**) Scatter plots illustrating the relationship between NHE9 expression and the expression of tumour suppressor miRNAs Let-7f (**e**) and miR-200 (**f**) in CRC samples from the TCGA microarray database. (**g**) Scatter plots illustrating the relationship between epithelial signature (EPI) and 76GS (*left*), MLR (*centre*), and KS (*right*) EMT scores in colorectal cancer cells with NHE9 expression and EV control. (**h**) Scatter plots illustrating the relationship between mesenchymal signature (MES) and 76GS (*left*), MLR (*centre*), and KS (*right*) EMT scores in colorectal cancer cells with NHE9 expression and EV control. Each plot shows the linear fit and 95% confidence interval, Pearson's correlation coefficient (R), and P-value. (**i**) Overall survival in patients with CRC with low and high NHE9 expression, with tumours classified into four consensus molecular subtypes (CMS1-4) determined by using Kaplan–Meier plotter. P values were calculated by Pearson's correlation (**a-h**) and log-rank test (**i**). Relates to Figure 2.

**Supplementary Figure 3: NHE9 expression correlates with pH and $Ca^{2+}$-related pathways.**



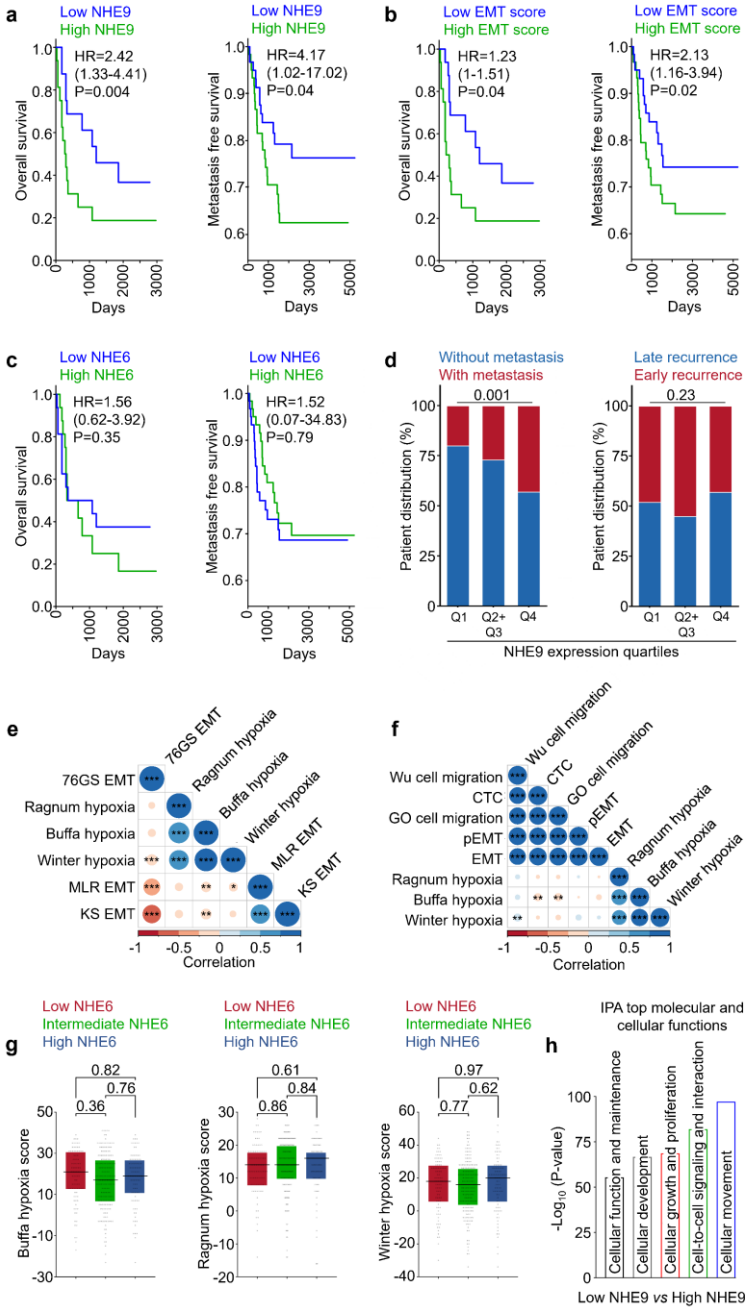

(**a-c**) Scatter plots illustrating the relationship between NHE9 expression and the activity levels (ssGSEA scores) of the $Ca^{2+}$ import (**a**), $Ca^{2+}$ transport (**b**), and $H^+$ transmembrane transport (**c**) gene sets in CRC tumours from the TCGA database. (**d**) Scatter plots depicting the relationship between NHE9 expression and 76GS EMT score in CRC samples from the GSE16125 (*n*=36) (*left*), GSE28722 (*n*=125) (*centre*), and TCGA (*n*=637) (*right*) databases. (**e**) Scatter plots depicting the relationship between NHE9 expression and MLR EMT score in CRC samples from the GSE16125 (*n*=36) (*left*), GSE28722 (*n*=125) (*centre*), and TCGA (*n*=637) (*right*) databases. Each plot shows the linear fit and 95% confidence interval, Pearson's correlation coefficient (R), and P-value. Relates to Figures 2-3.

**Supplementary Figure 4: NHE9 associates with aggressiveness and migratory potential.**




(**a-c**) Scatter plots illustrating the relationship between NHE9 expression and the activity levels (ssGSEA scores) of three independent migration gene sets (GO cell migration, Wu cell migration, and circulating tumour cell or CTC) as well as partial EMT (pEMT), EMT, cAMP inhibited, and cAMP activated gene sets in CRC samples from the GSE16125 (n=36) (**a**), GSE28722 (n=125) (**b**), and TCGA (n=637) (**c**) databases. See text for details. Each plot shows the linear fit and 95% confidence interval, Pearson's correlation coefficient (R), and P-value. Relates to Figure 4.

**Supplementary Figure 5: NHE9 affects overall survival and metastasis-free survival.**




(**a-c**) Overall (GSE16125; *left*) and metastasis-free (GSE28722; *right*) survival in patients with CRC with low and high NHE9 expression (**a**), low and high EMT state (VIM/CDH1 expression) (**b**), and low and high NHE6 expression (**c**). (**d**) Plots of patients with and without metastasis (*left*) and patients with early (<5years) and late (≥5 years) recurrence (*right*) sorted as low (Q1), intermediate (Q2, Q3), or high (Q4) NHE9 expression categories (GSE28722). (**e-f**) Correlation of Buffa, Ragnum, and Winter hypoxia scores with three EMT scores (**e**) and activity levels (ssGSEA) of three migration gene sets (GO cell migration, Wu cell migration, and circulating tumour cell or CTC), and partial EMT (pEMT) and EMT gene sets (**f**) in CRC samples from the TCGA database. Pearson's correlation value for each pairwise comparison is represented by the size of the circle, filled with the corresponding colour from the colour bar at the bottom. *P<0.05, **P<0.01, ***P<0.001. (**g**) Box plots of three hypoxia scores in CRC samples from the TCGA database, sorted as low (Q1), intermediate (Q2, Q3), or high (Q4) NHE6 expression categories. (**h**) Top enriched IPA molecular and cellular functions in CRC samples with low and NHE9 expression from the TCGA database. P values were calculated by log-rank test (**a,b,c**), chi-square test (**d**), Pearson's correlation (**e,f**), and one-way ANOVA (**g**). Relates to Figures 4-5.

**Supplementary Table 1: Parameter estimation for the minimal model.**

The model setup extends the previous experimental findings [1]. The parameters were adopted from published literature for the molecular species of the EMT regulatory circuit (miR-200, Snail, Zeb, E-cadherin), NHE9, Let-7f, and NFATc ($Ca^{2+}$) interactions. $\mu_{200}$, miR-200; $S$, Snail; $Z$, Zeb; $E$, E-cadherin; $N$, NFATc; $l7$, Let-7f; $nh$, NHE9; K, $10^3$ molecules.

| Parameter | Value | Reference |
|---|---|---|
| gμ200 (Molecules/Hour) | 2.1K | [2] |
| gmZ (Molecules/Hour) | 11 | [2] |
| $Z^0\mu_{200}$ (Molecules) | 220K | [2] |
| $Z^0m_Z$ (Molecules) | 25K | [2] |
| $n_{Z,\mu_{200}}$ | 3 | [2] |
| $n_{Z,m_Z}$ | 2 | [2] |
| $n_{\mu_{200}}$ | 6 | [2] |
| $n_{S,\mu_{200}}$ | 2 | [2] |
| $n_{S,m_Z}$ | 2 | [2] |
| $\lambda_{Z,\mu_{200}}$ | 0.1 | [2] |
| $\lambda_{Z,m_Z}$ | 7.5 | [2] |
| $\lambda_{S,\mu_{200}}$ | 0.1 | [2] |
| $\lambda_{S,m_Z}$ | 10 | [2] |
| $k_{\mu_{200}}$(Hour$^{-1}$) | 0.05 | [2] |
| $k_{m_Z}$(Hour$^{-1}$) | 0.5 | [2] |
| $k_Z$(Hour$^{-1}$) | 0.1 | [2] |
| $g_Z$(Hour$^{-1}$) | 0.1K | [2] |
| $S^0_{\mu_{200}}$(Molecules) | 180K | [2] |
| $S^0_{m_Z}$(Molecules) | 180K | [2] |
| S(Molecules) | 200K | [2] |
| $\mu^0_{200}$(Molecules) | 10K | [2] |
| $g_E$(Molecules/Hour) | 5000 | [3] |
| $k_E$(Hour$^{-1}$) | 0.1 | [3] |
| $\lambda_{E,m_Z}$ | 0.8 | [3] |
| $n_{E,m_Z}$ | 2 | [3] |
| $E^0_{m_Z}$(Molecules) | 80000 | [3] |
| $\lambda_{Z,E}$ | 0.1 | [3] |
| $n_{Z,E}$ | 2 | [3] |
| $Z^0_E$ | 100000 | [3] |
| $\lambda_{N,E}$ | 8 | [4] |
| $n_{N,E}$ | 5 | [5] |
| $N^0_E$ | 100000 | Estimated |
| $g_N$(Molecules/Hour) | 80000 | [5] |
| $k_N$(Hour$^{-1}$) | 0.1 | [5] |
| $\lambda_{N,Z}$ | 2 | [5] |
| $n_{N,Z}$ | 2 | [5] |
| $N^0_Z$ | 800000 | Estimated |
| $\lambda_{N,\mu_{200}}$ | 4 | Estimated |
| $n_{N,\mu_{200}}$ | 4 | Estimated |
| $N^0_{\mu200}$ (Molecules) | 500000 | Estimated |

| | | |
|---|---|---|
| $\lambda_{N,N}$ | 7 | [6] |
| $n_{N,N}$ | 3 | [6] |
| $N_N^0$ | 100000 | Estimated |
| $k_{l7}$ | 0.05 | [7] |
| $k_{nh}$ | 0.1 | Estimated |
| $g_{l7}$ | 200 | [7] |
| $g_{nh}$ | 80000 | Estimated |
| $\lambda_{S,l7}$ | 0.6 | [8] |
| $\lambda_{l7,nh}$ | 0.6 | [9] |
| $\lambda_{l7,l7}$ | 11 | [7] |
| $n_{S,l7}$ | 2 | [8] |
| $n_{l7,nh}$ | 2 | [9] |
| $n_{l7,l7}$ | 3 | [7] |
| $l7_{l7}^0$ | 1200 | [7] |
| $S_{l7}^0$ | 180000 | Estimated |
| $l7_{nh}^0$ | 200000 | Estimated |

**Supplementary References:**


1 Prasad, H. *et al.* An Endosomal Acid-Regulatory Feedback System Rewires Cytosolic cAMP Metabolism and Drives Tumor Progression. *Mol Cancer Res* **22**, 465-481, doi:10.1158/1541-7786.MCR-23-0606 (2024).

2 Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N. & Ben-Jacob, E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc Natl Acad Sci U S A* **110**, 18144-18149, doi:10.1073/pnas.1318192110 (2013).

3 Mooney, S. M., Jolly, M. K., Levine, H. & Kulkarni, P. Phenotypic plasticity in prostate cancer: role of intrinsically disordered proteins. *Asian J Androl* **18**, 704-710, doi:10.4103/1008-682X.183570 (2016).

4 Gould, R., Bassen, D. M., Chakrabarti, A., Varner, J. D. & Butcher, J. Population Heterogeneity in the Epithelial to Mesenchymal Transition Is Controlled by NFAT and Phosphorylated Sp1. *PLoS Comput Biol* **12**, e1005251, doi:10.1371/journal.pcbi.1005251 (2016).

5 Oikawa, T. *et al.* Acquired expression of NFATc1 downregulates E-cadherin and promotes cancer cell invasion. *Cancer Res* **73**, 5100-5109, doi:10.1158/0008-5472.CAN-13-0274 (2013).

6 Serfling, E., Chuvpilo, S., Liu, J., Hofer, T. & Palmetshofer, A. NFATc1 autoregulation: a crucial step for cell-fate determination. *Trends Immunol* **27**, 461-469, doi:10.1016/j.it.2006.08.005 (2006).

7 Jolly, M. K. *et al.* Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **7**, 27067-27084, doi:10.18632/oncotarget.8166 (2016).

8 Unternaehrer, J. J. *et al.* The epithelial-mesenchymal transition factor SNAIL paradoxically enhances reprogramming. *Stem Cell Reports* **3**, 691-698, doi:10.1016/j.stemcr.2014.09.008 (2014).

9 Chen, G. *et al.* Post-transcriptional Gene Regulation in Colitis Associated Cancer. *Front Genet* **10**, 585, doi:10.3389/fgene.2019.00585 (2019).