



OPEN

DATA DESCRIPTOR

# Bulk RNA sequencing dataset of Claudin-low breast cancer cell lines with Neuropilin-1 knockdown

Layla-Rose Lynam<sup>1</sup>, Anja Rockstroh<sup>1</sup>, Melanie Lehman<sup>1,2</sup>, Yu Hin Tang<sup>1</sup>,  
Minh Long Ngoc Nguyen<sup>1</sup>, Philip A. Gregory<sup>3,4</sup>, Colleen C. Nelson<sup>1</sup>, Marianna Volpert<sup>1</sup> &  
Brett G. Hollier<sup>1</sup>✉

Triple-negative breast cancers (TNBC) are a particularly aggressive breast cancer subtype with poor prognosis and high relapse rates. Due to a lack of identified targeted therapies, chemotherapy currently remains as the primary treatment for TNBC. Approximately 25–39% of TNBC are claudin-low breast cancers, which are mainly defined by low expression of cell-cell adhesion proteins and enrichment of mesenchymal signatures. Functional studies have demonstrated the potential role of the transmembrane-coreceptor, Neuropilin-1 (NRP1) in regulating the progression of these tumours. However, there have been no high-throughput studies to date that comprehensively investigate NRP1-modulated cell-signalling across multiple claudin-low cell lines. Therefore, we treated HS578T, MDA-MB-231 and SUM159PT claudin-low cell lines with either a non-targeting (NT) control or two NRP1-targeting small-interfering RNA (siRNA) or short-hairpin RNA (shRNA) sequences and followed this with bulk-RNA sequencing. We present this comprehensive transcriptomic dataset which provides a valuable resource for understanding both the transcriptomic landscape of claudin-low breast cancer and NRP1-regulated signalling pathways. Therefore, paving the way for future studies of its potential as a therapeutic target.

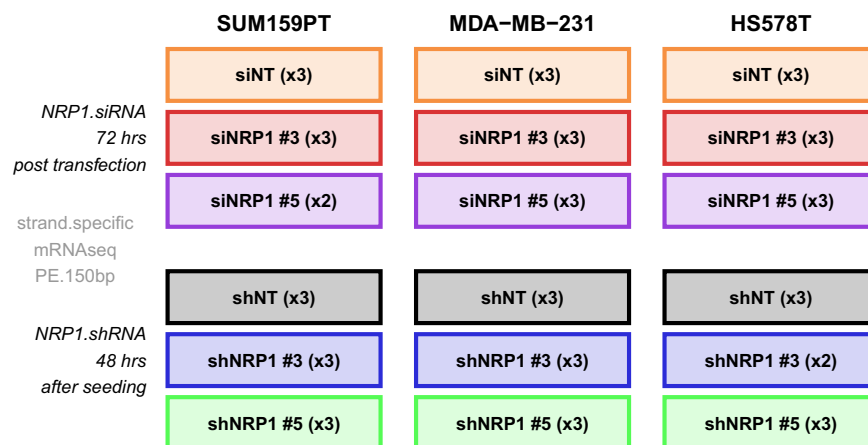
## Background & Summary

Breast cancer (BrCa) is a highly heterogeneous disease that can be divided into five intrinsic subtypes according to a 50-gene signature (PAM-50) classification which are luminal A, luminal B, human epidermal growth factor receptor 2 (HER2)-positive, basal-like and normal-like. Further molecular sub-classification is by estrogen receptor (ER), progesterone receptor (PR) and HER2 expression, which is defined as ER/PR-positive, HER2-enriched or triple-negative (TNBC), which do not express ER, PR or HER2<sup>1,2</sup>.

Triple-negative breast cancer (TNBC) comprises between 10–20% of all diagnosed breast cancers and is particularly aggressive, with metastatic disease developing in more than 30% of patients and tumour recurrence as soon as 2–3 years from diagnosis<sup>2–7</sup>. Because TNBCs lack expression of ER, PR and HER2, the available HER2 and hormone-receptor targeted therapies are ineffective for TNBC. This consequently results in chemotherapy remaining as the primary treatment regime<sup>7</sup>. Poly ADP ribose polymerase (PARP) inhibitors and Programmed cell death-ligand 1 (PD-L1) immunotherapy treatments have demonstrated significant clinical benefit, however only for subsets of TNBC that either carry BRCA mutations or are PD-L1 positive<sup>8–10</sup>. Hence, there is clinical demand for further research into potential targeted therapies for TNBC.

Approximately 25–39% of TNBC are further sub-classified as claudin-low breast cancer<sup>2,11,12</sup>. Claudin-low breast cancer was identified in 2007 and was thought to be a distinct intrinsic subtype, however research over recent years lead to a redefinition of claudin-low as an additional molecular subgroup that is acquired by the main intrinsic breast cancer subtypes<sup>12–15</sup>. Claudin-low breast cancer is defined by low expression of

<sup>1</sup>Australian Prostate Cancer Research Centre - Queensland, Centre for Genomics and Personalised Health, Faculty of Health, School of Biomedical Sciences, Translational Research Institute, Queensland University of Technology, Brisbane, QLD, Australia. <sup>2</sup>University of British Columbia, Vancouver Prostate Centre, Department of Urologic Sciences, Vancouver, BC, Canada. <sup>3</sup>Centre for Cancer Biology, University of South Australia and SA Pathology, Adelaide, SA, Australia. <sup>4</sup>Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, SA, Australia. ✉e-mail: [b.hollier@qut.edu.au](mailto:b.hollier@qut.edu.au)



**Fig. 1** Schematic summary of dataset. Schematic summarising all cell lines and NRP1-targeting siRNA and shRNA conditions within this dataset, in addition to the number of biological replicates and experiment timepoints.

cell-cell adhesion proteins such as claudin-3, -4 -7 (*CLDN3*, *CLDN4*, *CLDN7*) and E-cadherin (*CDH1*) and enrichment of epithelial to mesenchymal transition (EMT) and cancer stem-cell signatures including Zinc finger E-box-binding homeobox 1 (*ZEB1*), Snail Family Transcriptional Repressor 1 (*SNAIL1*), Vimentin (*VIM*), Integrin alpha 6 (*ITGA6*) and Aldehyde Dehydrogenase 1 (*ALDH1*)<sup>6,12,13</sup>. These tumours additionally exhibit aberrant activation of oncogenic signalling pathways including RAS-MAPK<sup>6,12-15</sup>. These characteristics result in claudin-low breast cancer being the most primitive and least differentiated subtype, with close resemblance to mammary epithelial stem cells<sup>6,12-15</sup>.

Neuropilin-1 (NRP1) is a 120 kDa transmembrane co-receptor protein that has been associated with several physiological processes including immunological and cardiovascular development, neuronal guidance, cell migration and angiogenesis<sup>16</sup>. NRP1 is known to have two main protein forms: transmembrane and soluble<sup>16,17</sup>. Transmembrane NRP1 is the most well characterised and consists of a small intracellular domain, transmembrane domain and an extracellular domain<sup>18-20</sup>. The extracellular domain function is well characterised as the binding site of most extracellular ligands and is divided into the N-terminal complement-binding CUB domain (a1/a2), coagulation factor V/VIII (b1/b2) domain, and a meprin or MAM domain (c)<sup>16,19</sup>. The transmembrane and meprin domains are essential for NRP1 dimerization to maintain functionality of co-receptor activity whereas the function of the intracellular domain remains unclear<sup>18</sup>. Soluble-NRP1 exists without the intracellular/cytoplasmic or transmembrane domains but still expresses an extracellular domain and therefore can bind extracellular ligands<sup>16,17</sup>. However, the functions of soluble-NRP1 are not well understood<sup>16,17</sup>. As a co-receptor, NRP1 dimerises to a broad spectrum of growth factor receptors including VEGFR, EGFR and PDGFR to enhance binding of the corresponding ligands<sup>21-26</sup>. In cancer, this consequently results in the activation of various oncogenic signalling cascades that promote tumorigenic processes such as angiogenesis, metastasis and invasion<sup>16,27-30</sup>.

We previously reported that NRP1 is more highly expressed in claudin-low breast cancers in comparison to Luminal A, Luminal B, HER2, Basal-like and Normal-like breast cancer<sup>2</sup>. Upon knockdown of NRP1 in triple-negative claudin-low cell lines, *in-vivo* tumour growth and *in-vitro* proliferation were significantly reduced<sup>2</sup>. Additionally, we saw decreased *ZEB1* and *ITGA6* expression as well as reduced PDGFR and EGFR activation in response to *NRP1* knockdown *in-vitro*<sup>2</sup>. This implicated NRP1 as a key driver of claudin-low breast cancer progression, namely through EMT and RAS-MAPK regulation<sup>2</sup>. However, the knowledge of the role of NRP1 in claudin-low breast cancer remains limited. Aside from a study by Al-Zeheimi *et al.*<sup>31</sup> which performed transcriptional analysis of the effect of NRP1 knockout by CRISPR-Cas9 in the claudin-low MDA-MB-231 cell line, there have been no high-throughput studies to date that comprehensively investigate NRP1-modulated cell signalling across multiple claudin-low cell lines<sup>31</sup>.

Therefore, in this study, we knocked-down *NRP1* by two small-interfering RNA (siRNA) and two short-hairpin RNA (shRNA) sequences in each of the HS578T, MDA-MB-231 and SUM159PT claudin-low cell lines<sup>31</sup> and followed this with bulk-RNA sequencing. This generated a comprehensive transcriptomic dataset (consisting of 52 samples, Fig. 1) which provides a valuable resource for in depth analysis of the cellular pathways that are altered in the absence of NRP1 in these cell lines. Understanding the NRP1-regulated signalling cascades using this dataset could pave the way for future studies of NRP1 as a potential targeted therapeutic for claudin-low breast cancers, which could be administered for a substantial portion of TNBCs.

## Methods

The following sections detail the methods involved from RNA sample collection to sequencing and processing of the sequencing data. For a schematic summary of this dataset see Fig. 1. For a list and description of all 52 samples in this dataset see Tables 2–4. Two to three independent biological replicates are provided for each treatment condition.

tool version	parameters
TrimGalore.v.0.6.5	trim_galore --fastqc --paired --retain_unpaired --length 75 --clip_r1 10 --clip_r2 10
STAR.v.2.7.2b	STAR-runMode alignReads \ --readFilesCommand zcat \ --limitBAMsortRAM 80000000000 \ --quantMode TranscriptomeSAM \ --twopassMode Basic \ --alignIntronMax 1000000 \ --alignIntronMin 20 \ --alignMatesGapMax 1000000 \ --alignSJDBoverhangMin 10 \ --alignSJstitchMismatchNmax 5 -1 5 5 \ --alignSJoverhangMin 10 \ --chimJunctionOverhangMin 12 \ --chimNonchimScoreDropMin 10 \ --chimMultimapNmax 20 \ --chimMultimapScoreRange 3 \ --chimOutJunctionFormat 1 \ --chimSegmentMin 12 \ --chimScoreJunctionNonGTAG -4 \ --peOverlapMmp 0.1 \ --peOverlapNbasesMin 12 \ --outFilterMismatchNmax 999 \ --outFilterMismatchNoverReadLmax 0.04 \ --outFilterMultimapNmax 20 \ --outFilterType BySJout \ --outReadsUnmapped Fastx \ --outSAMtype BAM SortedByCoordinate \ --outWigNorm None \ --outWigType wiggle \ --outSAMunmapped Within \ --outSAMstrandField intronMotif
RSEM.v.1.3.3	rsem-calculate-expression --bam --paired-end --alignments --no-bam-output --seed 12345 -p 12 --forward-prob 0
Kraken2.v2.0.9beta	kraken2 \ --threads \$task.cpus \ --db \$db \ --paired \ --use-names \ --classified-out \${sample_name}-classified#.fastq \ --unclassified-out \${sample_name}-unclassified#.fastq \ --output \${sample_name}-kraken2-output.txt \ --report \${sample_name}-kraken2-report.txt \ \$read1 \$read2

**Table 1.** Tools, tool versions and parameters used for raw data processing.

**Cell culture.** MDA-MB-231, HS578T and SUM159PT (SUM159) claudin-low breast cancer cell lines were obtained from the American Type Culture Collection (ATCC) and maintained in Dulbecco's Modified Eagle Medium (DMEM), High Glucose, Pyruvate supplemented with 10% Fetal Bovine Serum (FBS) both from Gibco, Thermo Fisher Scientific. These cell lines were authenticated by short tandem replicate analysis at the Genomics Research Centre (Queensland University of Technology, Australia) and tested for mycoplasma using the Lonza MycoAlert mycoplasma detection kit at the Translational Research Institute (Brisbane, Australia). HEK293T cells were sourced from the ATCC and cultured in DMEM, High Glucose, Pyruvate supplemented with 5% FBS (Gibco, Thermo Fisher Scientific). All cells were maintained at 37 °C, 5% CO<sub>2</sub>.

**NRP1 knockdown by siRNA.** Following the same reverse transfection methodology as previously reported<sup>2</sup>, 20 nM siRNA was reverse transfected into  $0.7 \times 10^6$  SUM159PT, MDA-MB-231 and HS578T cells using 1:500 Lipofectamine<sup>™</sup> RNAiMAX Transfection Reagent (Invitrogen, Thermo Fisher Scientific) and 1:10 Opti-MEM<sup>™</sup> Reduced Serum Medium (Gibco, Thermo Fisher Scientific) to DMEM supplemented with 10% FBS (Gibco, Thermo Fisher Scientific). RNA was harvested at 72 hrs post-transfection. A predesigned Silencer<sup>™</sup> Select non-targeting siRNA (siNT) (Negative Control siRNA No. 1, #4390844, Thermo Fisher Scientific) was used in addition to two custom Silencer<sup>™</sup> Select NRP1-targeting siRNAs (Thermo Fisher Scientific). The NRP1 siRNA sequences (sense) were as follows: siNRP1#3, 5' UAACCACAUUUCACAAGAA 3' and siNRP1#5, 5' CAGCCUUGAUGCACUUAU 3'. siNRP1#3 targets transmembrane-NRP1 only, whereas siNRP1#5 targets both transmembrane and soluble NRP1 (Fig. 6c).

**NRP1 knockdown by shRNA.** Following the same shRNA transduction methodology as previously reported<sup>2</sup>, lentiviral media was produced by seeding  $1.5 \times 10^6$  HEK293T cells per shRNA in 10 mL of DMEM, High Glucose, Pyruvate supplemented with 10% heat inactivated FBS (Gibco, Thermo Fisher Scientific). FBS heat inactivation was done by a 30-minute incubation in a 56 °C water bath. At 50–60% confluency, the HEK293T cells were transfected with 200 µL of lentiviral transfection mix. Transfection mix was prepared as follows: serum-free DMEM (Gibco), 12 µL X-tremeGENE<sup>™</sup> HP DNA Transfection Reagent (Sigma Aldrich), 1.8 µg pCMV Delta 8.2 R (Addgene), 0.2 µg pCMV-VSV-G (Addgene) and 2.0 µg of shRNA cloned to the pLKO.1 lentiviral vector (Addgene) was combined to a 200 µL final volume and incubated for 15 minutes at room temperature before adding to the HEK293T cells. Following an overnight incubation at 37 °C with 5% CO<sub>2</sub>, all media/transfection mix

Sample Name	Description	GEO ID
NRP1-SUM159-siNT-RNA.1	SUM159PT with non-targeting siRNA, biological replicate 1	GSM8250668
NRP1-SUM159-siNT-RNA.2	SUM159PT with non-targeting siRNA, biological replicate 2	GSM8250669
NRP1-SUM159-siNT-RNA.3	SUM159PT with non-targeting siRNA, biological replicate 3	GSM8250670
NRP1-SUM159-siNRP1.3-RNA.1	SUM159PT with NRP1 siRNA sequence 3, biological replicate 1	GSM8250663
NRP1-SUM159-siNRP1.3-RNA.2	SUM159PT with NRP1 siRNA sequence 3, biological replicate 2	GSM8250664
NRP1-SUM159-siNRP1.3-RNA.3	SUM159PT with NRP1 siRNA sequence 3, biological replicate 3	GSM8250665
NRP1-SUM159-siNRP1.5-RNA.2	SUM159PT with NRP1 siRNA sequence 5, biological replicate 2	GSM8250666
NRP1-SUM159-siNRP1.5-RNA.3	SUM159PT with NRP1 siRNA sequence 5, biological replicate 3	GSM8250667
NRP1-SUM159-shNT-RNA.1	SUM159PT with non-targeting shRNA, biological replicate 1	GSM8250660
NRP1-SUM159-shNT-RNA.2	SUM159PT with non-targeting shRNA, biological replicate 2	GSM8250661
NRP1-SUM159-shNT-RNA.3	SUM159PT with non-targeting shRNA, biological replicate 3	GSM8250662
NRP1-SUM159-shNRP1.3-RNA.1	SUM159PT with NRP1 shRNA sequence 3, biological replicate 1	GSM8250654
NRP1-SUM159-shNRP1.3-RNA.2	SUM159PT with NRP1 shRNA sequence 3, biological replicate 2	GSM8250655
NRP1-SUM159-shNRP1.3-RNA.3	SUM159PT with NRP1 shRNA sequence 3, biological replicate 3	GSM8250656
NRP1-SUM159-shNRP1.5-RNA.1	SUM159PT with NRP1 shRNA sequence 5, biological replicate 1	GSM8250657
NRP1-SUM159-shNRP1.5-RNA.2	SUM159PT with NRP1 shRNA sequence 5, biological replicate 2	GSM8250658
NRP1-SUM159-shNRP1.5-RNA.3	SUM159PT with NRP1 shRNA sequence 5, biological replicate 3	GSM8250659

**Table 2.** Overview of the 17 SUM159PT samples.

was aspirated from the HEK293T cells and replaced with 7 mL of regular culture media. Viral supernatant was then harvested at 48 hrs and 72 hrs post-transfection.

For transduction of the shRNA,  $0.7 \times 10^6$  HS578T, MDA-MB-231 and SUM159PT cells per shRNA sequence were cultured to 30–50% confluency. The cells were then incubated at 37 °C overnight in 4.6 mL of previously harvested lentiviral supernatant with 6 µg/mL of protamine sulfate (Sigma Aldrich). Successfully transduced cells were selected with 1 µg/mL puromycin (Gibco, Thermo Fisher Scientific) for three days before experimental use. Once 80–90% confluent post-selection, cells were seeded at  $0.7 \times 10^6$  and RNA was harvested after 48 hrs. Consecutive cell passages were collected as biological replicates.

The shRNAs used were a non-targeting shRNA (shNT) (Thermo Fisher Scientific), and two NRP1-targeting shRNA sequences obtained from Thermo Fisher Scientific. Sequences (sense) were as follows: shNRP1#3, 5' GCUGUGGAUGACAUAUAGUAUU 3' and shNRP1#5, 5' CAGCCUUGAAUGCACUUAU 3'. Similarly to the siRNA sequences, shNRP1#3 targets transmembrane-NRP1 only, whereas shNRP1#5 targets transmembrane and soluble NRP1. Note that the sequences for siNRP1#5 and shNRP1#5 are identical, whereas siNRP1#3 and shNRP1#3 sequences differ from each other, with their mapping sites being offset by about 20 nucleotides and hence, are targeting a slightly different subset of transcript variants (Fig. 6c).

**RNA Isolation, cDNA synthesis and Reverse-Transcription quantitative Polymerase-Chain-Reaction (RT-qPCR).** Cells were lysed with TRIzol reagent (Thermo Fisher Scientific, Invitrogen) and RNA extraction was performed as per manufacturer's instructions using the Direct-zol™ RNA Miniprep Plus kit (Zymo Research). RNA concentration and purity was determined using the NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific). A 260/280 ratio of ~2.0 was considered pure. Synthesis of 1 µg of cDNA was performed using the SensiFAST™ cDNA Synthesis Kit (Meridian Bioscience) as per manufacturer's instructions. RT-qPCR was prepared using the SYBR™ Green PCR Master Mix (Applied Biosystems, Thermo Fisher Scientific), 10 µM of reverse and forward primers and 1:10 cDNA. The QuantStudio 6 Real-Time PCR System (Thermo Fisher Scientific) was used with the default standard run settings. Gene expression was determined using the comparative Ct method, with RPL32 as the housekeeping gene. Custom primer sequences used from Sigma Aldrich were as follows: Total NRP1 (FWD 5' AGGACAGAGACTGCAAGTATGAC 3', REV 5' AACATTCAGGACCTCTCTTGA 3', see Fig. 6c for mapping site), Transmembrane NRP1 (FWD 5' CGAGGGCGAAATCGGAAAAGG 3', REV 5' CTTTCGTATCCTGGCGTGCT 3', see Fig. 6c for mapping site) and RPL32 (FWD 5' GCACCAGTCAGACCGATATG 3', REV 5' ACTGGGCAGCATGTGCTTTG 3'). Statistics were determined by one-way analysis of variance (ANOVA) and Dunnett's post-hoc multiple comparisons test within GraphPad Prism v10.0.2.

Sample Name	Description	GEO ID
NRP1-MDA.MB.231-siNT-RNA.1	MDA-MB-231 with non-targeting siRNA, biological replicate 1	GSM8250651
NRP1-MDA.MB.231-siNT-RNA.2	MDA-MB-231 with non-targeting siRNA, biological replicate 2	GSM8250652
NRP1-MDA.MB.231-siNT-RNA.3	MDA-MB-231 with non-targeting siRNA, biological replicate 3	GSM8250653
NRP1-MDA.MB.231-siNRP1.3-RNA.1	MDA-MB-231 with NRP1 siRNA sequence 3, biological replicate 1	GSM8250645
NRP1-MDA.MB.231-siNRP1.3-RNA.2	MDA-MB-231 with NRP1 siRNA sequence 3, biological replicate 2	GSM8250646
NRP1-MDA.MB.231-siNRP1.3-RNA.3	MDA-MB-231 with NRP1 siRNA sequence 3, biological replicate 3	GSM8250647
NRP1-MDA.MB.231-siNRP1.5-RNA.1	MDA-MB-231 with NRP1 siRNA sequence 5, biological replicate 1	GSM8250648
NRP1-MDA.MB.231-siNRP1.5-RNA.2	MDA-MB-231 with NRP1 siRNA sequence 5, biological replicate 2	GSM8250649
NRP1-MDA.MB.231-siNRP1.5-RNA.3	MDA-MB-231 with NRP1 siRNA sequence 5, biological replicate 3	GSM8250650
NRP1-MDA.MB.231-shNT-RNA.1	MDA-MB-231 with non-targeting shRNA, biological replicate 1	GSM8250642
NRP1-MDA.MB.231-shNT-RNA.2	MDA-MB-231 with non-targeting shRNA, biological replicate 2	GSM8250643
NRP1-MDA.MB.231-shNT-RNA.3	MDA-MB-231 with non-targeting shRNA, biological replicate 3	GSM8250644
NRP1-MDA.MB.231-shNRP1.3-RNA.1	MDA-MB-231 with NRP1 shRNA sequence 3, biological replicate 1	GSM8250636
NRP1-MDA.MB.231-shNRP1.3-RNA.2	MDA-MB-231 with NRP1 shRNA sequence 3, biological replicate 2	GSM8250637
NRP1-MDA.MB.231-shNRP1.3-RNA.3	MDA-MB-231 with NRP1 shRNA sequence 3, biological replicate 3	GSM8250638
NRP1-MDA.MB.231-shNRP1.5-RNA.1	MDA-MB-231 with NRP1 shRNA sequence 5, biological replicate 1	GSM8250639
NRP1-MDA.MB.231-shNRP1.5-RNA.2	MDA-MB-231 with NRP1 shRNA sequence 5, biological replicate 2	GSM8250640
NRP1-MDA.MB.231-shNRP1.5-RNA.3	MDA-MB-231 with NRP1 shRNA sequence 5, biological replicate 3	GSM8250641

**Table 3.** Overview of the 18 MDA-MB-231 samples.

**RNA Quality validation.** RNA quality and integrity was validated using the 2100 Bioanalyzer (Agilent Technologies) to determine the RNA integrity number (RIN). Sample preparation was done using the RNA 6000 Nano Kit (Agilent Technologies) according to manufacturer’s instructions. Bioanalyzer results showed that each RNA sample submitted had a RIN > 9.

**Library preparation and RNAseq.** RNA was submitted to the QUT Central Analytical Research Facility (CARF), Brisbane, Queensland for library preparation and sequencing. Library preparation was performed using the Illumina TruSeq Stranded mRNA Sample Prep Kit (Illumina, strand-specific, polyA enriched) with an input of 500 ng total RNA. This was followed by paired-end sequencing using the MGI DNBSEQ-G400 sequencer with a read length of 150 bp aiming for a depth of ~40 M read pairs per sample. All samples were multiplexed across all flow cells and lanes.

**Raw data processing, alignment and quality control.** For each sample, the de-multiplexed raw reads underwent quality control using the FastQC v0.11.9 tool<sup>32</sup>, after which the (good quality) FASTQ files from separate flow cells and/or lanes were combined, respectively. The combined FASTQ raw reads for each sample were then trimmed using TrimGalore v0.6.5<sup>33</sup>, followed by another quality check with FastQC. STAR aligner v2.7.2b<sup>34</sup> was used for alignment to the human genome (GRCh38/hg38) and transcriptome (Ensembl.v114/Gencode.v48, May-2025). RSEM v1.3.3<sup>35</sup> was used for read quantification, where transcript-level counts, isoform percentages and TPM values, as well as gene-level counts and TPM values were determined. Detailed tool parameters are provided in Table 1 below. Downstream data processing was performed using the R Statistical Software (version 4.4.3 2025-02-28 ucrt)<sup>36</sup> scripted in the RStudio Integrated Development Environment (version 2024.12.1.563)<sup>37</sup>. Between-sample ‘Trimmed Mean of M-values’ (TMM) normalisation followed by counts per million (CPM) and Fragments Per Kilobase of transcript per Million mapped reads (FPKM) quantification was then completed using the R package edgeR (v.4.4.2)<sup>38,39</sup>. To quality control for microbial contamination, Kraken2 v2.0.9beta<sup>40</sup> was used with default settings to align the unmapped reads from the STAR output to a comprehensive microbiome reference. MultiQC v1.9<sup>41</sup> was used to generate STAR aligner, FastQC and Kraken2 data reports. For Multidimensional scaling (MDS) analysis, the R package edgeR (v.4.4.2)<sup>38,39</sup> was used. The transcript-level and gene-level counts after TMM normalisation were used for transcript-level and gene-level MDS analysis. All plots were generated using the R package ggplot2 (v.4.0.0)<sup>42</sup>, as well as cowplot (v.1.2.0)<sup>43</sup> to arrange composite figures.



Sample Name	Description	GEO ID
NRP1-HS578T-siNT-RNA.1	HS578T with non-targeting siRNA, biological replicate 1	GSM8250633
NRP1-HS578T-siNT-RNA.2	HS578T with non-targeting siRNA, biological replicate 2	GSM8250634
NRP1-HS578T-siNT-RNA.3	HS578T with non-targeting siRNA, biological replicate 3	GSM8250635
NRP1-HS578T-siNRP1.3-RNA.1	HS578T with NRP1 siRNA sequence 3, biological replicate 1	GSM8250627
NRP1-HS578T-siNRP1.3-RNA.2	HS578T with NRP1 siRNA sequence 3, biological replicate 2	GSM8250628
NRP1-HS578T-siNRP1.3-RNA.3	HS578T with NRP1 siRNA sequence 3, biological replicate 3	GSM8250629
NRP1-HS578T-siNRP1.5-RNA.1	HS578T with NRP1 siRNA sequence 5, biological replicate 1	GSM8250630
NRP1-HS578T-siNRP1.5-RNA.2	HS578T with NRP1 siRNA sequence 5, biological replicate 2	GSM8250631
NRP1-HS578T-siNRP1.5-RNA.3	HS578T with NRP1 siRNA sequence 5, biological replicate 3	GSM8250632
NRP1-HS578T-shNT-RNA.1	HS578T with non-targeting shRNA, biological replicate 1	GSM8250624
NRP1-HS578T-shNT-RNA.2	HS578T with non-targeting shRNA, biological replicate 2	GSM8250625
NRP1-HS578T-shNT-RNA.3	HS578T with non-targeting shRNA, biological replicate 3	GSM8250626
NRP1-HS578T-shNRP1.3-RNA.1	HS578T with NRP1 shRNA sequence 3, biological replicate 1	GSM8250619
NRP1-HS578T-shNRP1.3-RNA.2	HS578T with NRP1 shRNA sequence 3, biological replicate 2	GSM8250620
NRP1-HS578T-shNRP1.5-RNA.1	HS578T with NRP1 shRNA sequence 5, biological replicate 1	GSM8250621
NRP1-HS578T-shNRP1.5-RNA.2	HS578T with NRP1 shRNA sequence 5, biological replicate 2	GSM8250622
NRP1-HS578T-shNRP1.5-RNA.3	HS578T with NRP1 shRNA sequence 5, biological replicate 3	GSM8250623

**Table 4.** Overview of the 17 HS578T samples.

## Data Records

This bulk mRNA-seq dataset is available for download from the NCBI Gene Expression Omnibus under GEO accession number **GSE266566**<sup>44</sup> and is summarised in Fig. 1 and Tables 2–4. The GEO entry includes 104 de-multiplexed raw FASTQ files after combining lanes (R1 and R2 from paired-end sequencing), a metadata file describing the samples and the experimental details, as well as RSEM-derived transcript and gene-level raw counts, TPM values and isoform percentages. Raw data is provided on GEO as a link to the Sequence Read Archive (SRA) database.

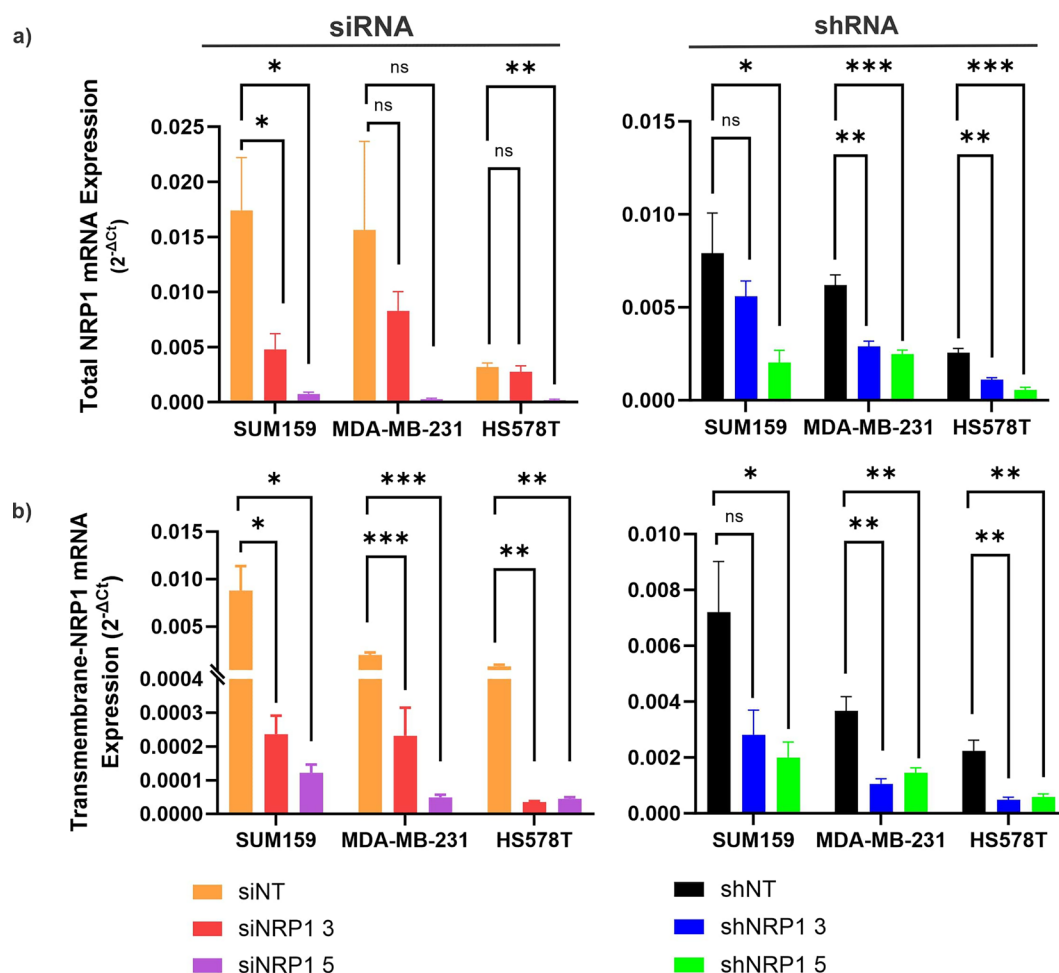
Our dataset comprises 52 samples in total, derived from the 3 claudin-low breast cancer cell lines HS578T, MDA-MB-231 and SUM159PT. For each model, NRP1 was knocked down using 2 siRNA and 2 shRNA sequences, as well as the non-targeting control RNAs, respectively. Two to three independent biological replicates (2 outlier samples were removed during data quality control) are provided for each treatment condition, resulting in 17–18 samples per cell line. See Fig. 1 for a schematic overview of the experimental setup and Tables 2–4 for a list and description of the individual samples within this claudin-low breast cancer dataset<sup>39</sup>.

## Technical Validation

**Validation of NRP1 knockdown by RT-qPCR.** Before submitting for RNA sequencing, NRP1 knockdown in the RNA samples of HS578T, MDA-MB-231 and SUM159PT cells was validated by RT-qPCR (Fig. 2). RT-qPCR revealed that total NRP1 expression in the siRNA and shRNA treated samples was significantly reduced from their respective siNT or shNT controls in most cases. However, HS578T siNRP1#3 NRP1 levels were only slightly reduced and almost equal to the siNT control (fold-change of  $-1.15$ ) (Fig. 2a). The siRNA and shRNA #3 sequences are identical and as they are designed to only target transmembrane-NRP1 (Fig. 6c), it was possible that soluble-NRP1 expression affected the total NRP1 knockdown levels in this sample. Therefore, we designed an additional primer set specifically for detecting transmembrane-NRP1 only (Fig. 6c). After repeating the RT-qPCR, results confirmed that transmembrane-NRP1 was repressed by shRNA and siRNA in all cell lines with statistical significance in all except SUM159PT shNRP1#3 (Fig. 2b).

**Data quality validation.** To analyse the quality of the sequencing data, the %GC content results generated by FastQC after trimming was first evaluated. The average %GC content per sample was plotted as a histogram and showed that all samples had a similar %GC content of ~50% (Fig. 3a).

As further validation of sequencing data quality, the number of total reads mapped to the human reference (post-trimming) as determined by the STAR aligner were analysed. Summarised as a percentage, ~85% to 95%



**Fig. 2** RT-qPCR validation of NRP1 knockdown in claudin-low breast cancer cell lines. mRNA expression ( $2^{-\Delta C_t}$ ) of (a) total NRP1 and (b) transmembrane-NRP1 in SUM159PT, HS578T and MDA-MB-231 cells following 72 hr transfection of siRNA (siNRP1#3 or #5) or lentiviral transduction of shRNA (shNRP1#3 and #5) and corresponding non-targeting (NT) controls, determined by RT-qPCR. Normalised to RPL32. N = 3. P-value determined by one-way analysis of variance (ANOVA) and Dunnett's post-hoc multiple comparisons test. NS = non-significant, \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ . Error bars = SEM.

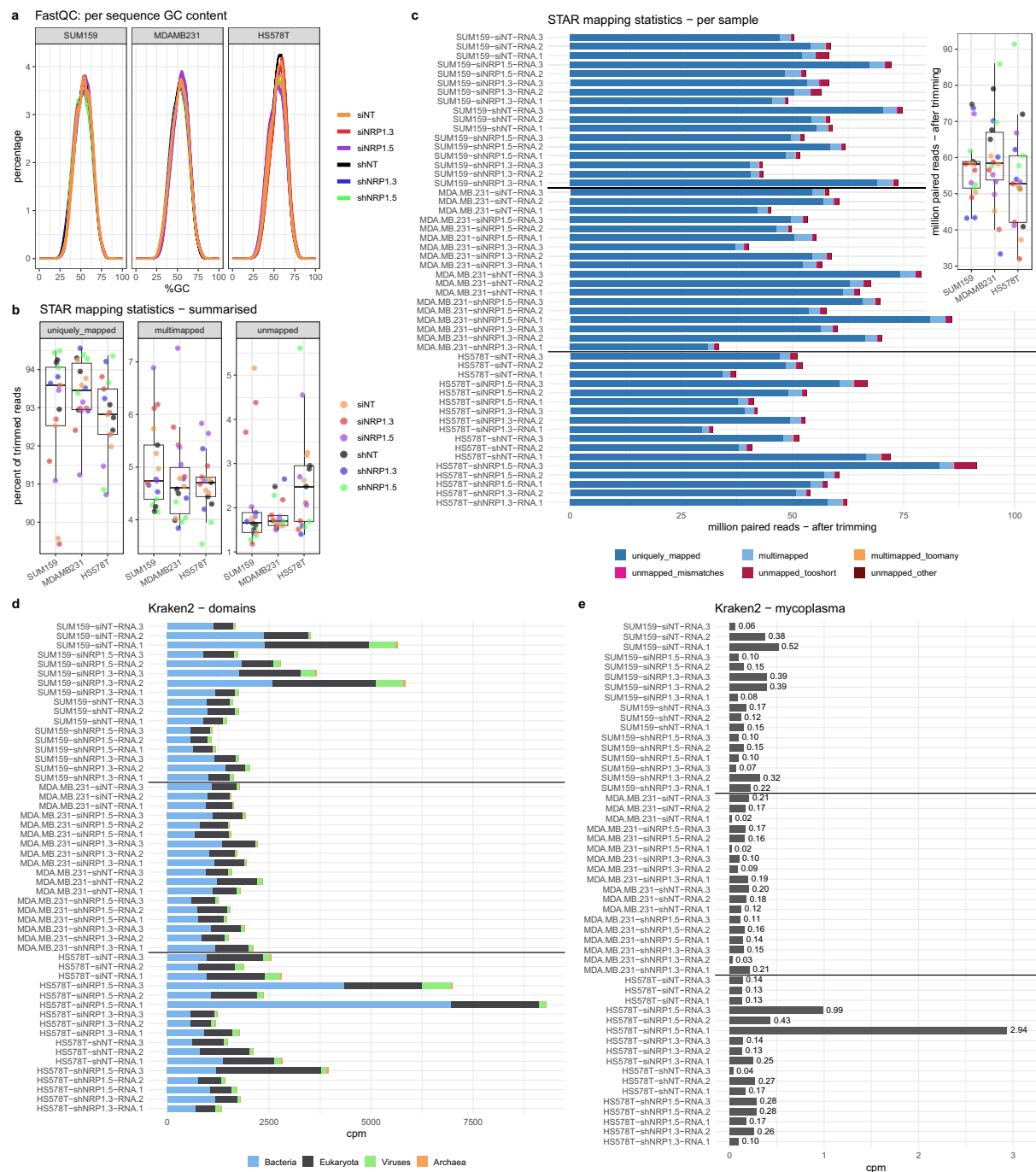
of reads were uniquely mapped, 3%–8% were multimapped and 1%–6% were unmapped (Fig. 3b). Altogether, there were ~30–92 M total paired reads for each sample (Fig. 3c). The mapping categories in Fig. 3c reflect the default categories provided in the STAR log files.

Additionally, the Kraken2<sup>40</sup> results were analysed to screen for contamination with mycoplasma or other microbes in the reads that did not map to the human reference. Microbial domain counts per million (cpm, calculated with respect to the total STAR input reads) revealed only low-level contamination (less than 2500 cpm) with Bacteria, Eukaryota, Viruses or Archaea in most samples (Fig. 3d). Mycoplasma was detected with less than 3 cpm in all samples (Fig. 3e), confirming that the cell lines were mycoplasma free, further supporting the results obtained with the Lonza MycoAlert mycoplasma detection kit on the live cultures.

Next, Multidimensional Scaling (MDS) was performed following TMM normalisation of the raw counts using the R package edgeR (v.4.4.2) to evaluate the similarity or dissimilarity of the samples within this dataset. Evaluating gene- as well as transcript-level MDS analyses of all samples (Fig. 4a,b) revealed that clustering was predominantly driven by cell line, indicating the large magnitude of transcriptomic differences between the models. This is followed by a separation of the samples based on the modality of knockdown (siRNA versus shRNA) in dimension 3, and a trend towards separating by target sequence #3 versus #5 in dimension 4 on the gene level.

Subsequent MDS analysis per individual cell line (Fig. 5) allowed a clearer visualisation of the treatment-group effect in each model. Although the sample clustering by type was not always tight and clean (on the transcript level in particular), gene- (Fig. 5a,b,c) and transcript-level (Fig. 5d,e,f) results showed that samples separate primarily by the siRNA versus shRNA group. In the second dimension a trend towards separation by target sequence #3 versus #5 is visible in the gene-level analysis.

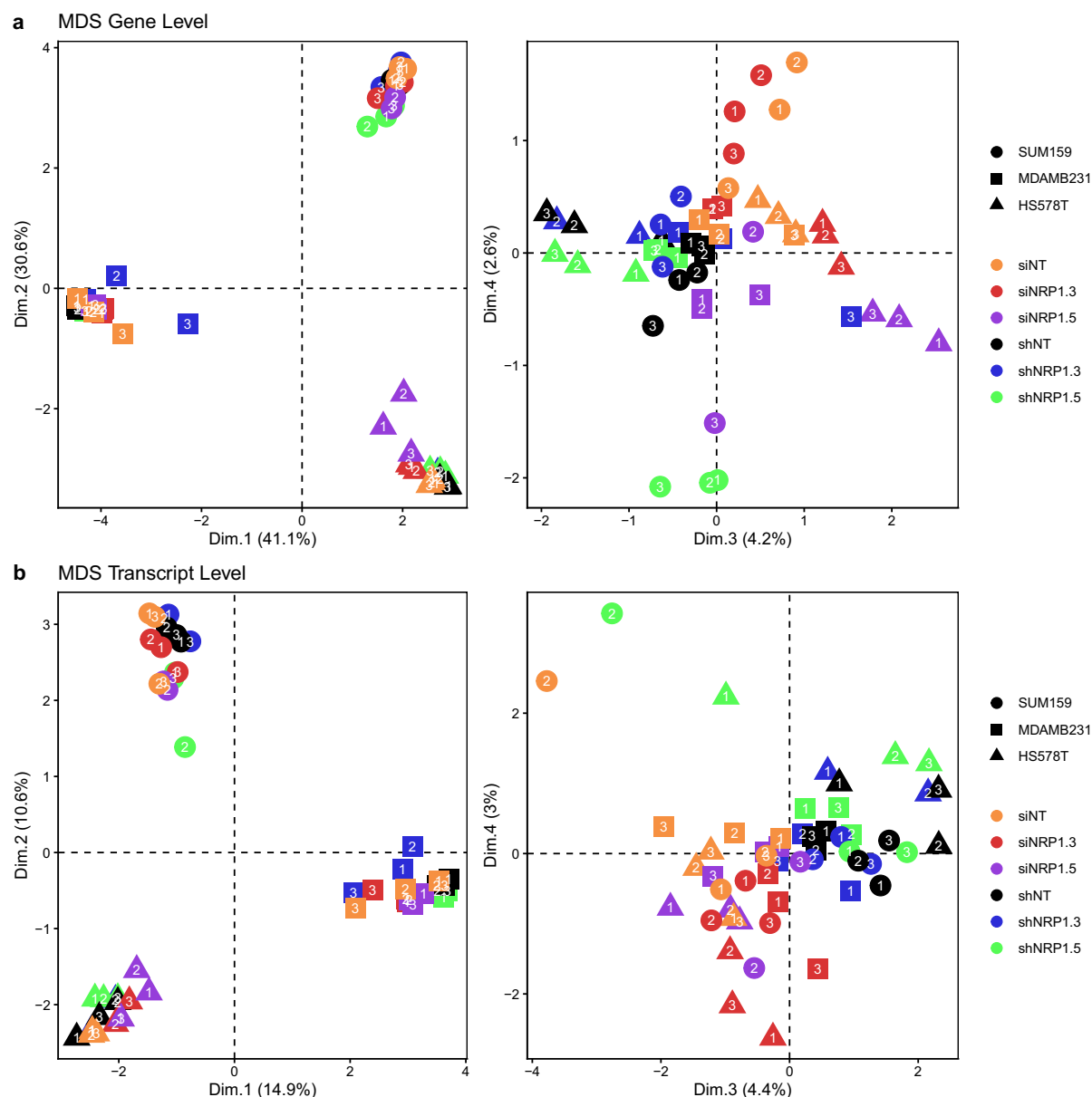
Subsequently, we evaluated NRP1 expression in our dataset at the gene and transcript level (Fig. 6). Analysis of NRP1 gene expression (Fig. 6a) revealed that in each cell line, the NRP1 expression level was lower in the



**Fig. 3** Dataset QC validation of %GC content, STAR mapping and Kraken2. Data shown is post-trimming. **(a)** Histogram of average %GC content for each sample generated by FastQC v0.11.9. **(b)** Box plot of percentage (%) of uniquely mapped, multimapped or unmapped reads to the human reference derived from the STAR aligner output. **(c)** Total STAR input reads (million paired reads) per sample with individual mapping categories as derived from the STAR aligner output. **(d)** Counts per million (cpm, calculated with respect to the total STAR input reads) for microbial domains (Bacteria, Eukaryota, Viruses, Archaea) and **(e)** mycoplasma as determined by Kraken2 v2.0.9beta. All plots were generated using the R package ggplot2 (v4.0.0).

siRNA and shRNA samples in comparison to their corresponding non-targeting (NT) RNAi controls, validating knockdown. Analysis of the isoform percentages of NRP1 transcript variants (Fig. 6b) showed that NRP1-206 (ENST00000374867) and NRP1-204 (ENST00000374822) had the highest expression percentage compared to other transcript variants. NRP1-206 and NRP1-204 are encoding the canonical transmembrane and soluble NRP1 isoform<sup>45,46</sup>, respectively (Fig. 6c). The isoform percentages further highlight that NRP1-204 (soluble variant) becomes the dominant transcript variant in the siNRP1#3 and shNRP1#3 samples, where the transmembrane-NRP1 is knocked down (Fig. 6b). This validates the specificity of the NRP1-targeting shRNA#3



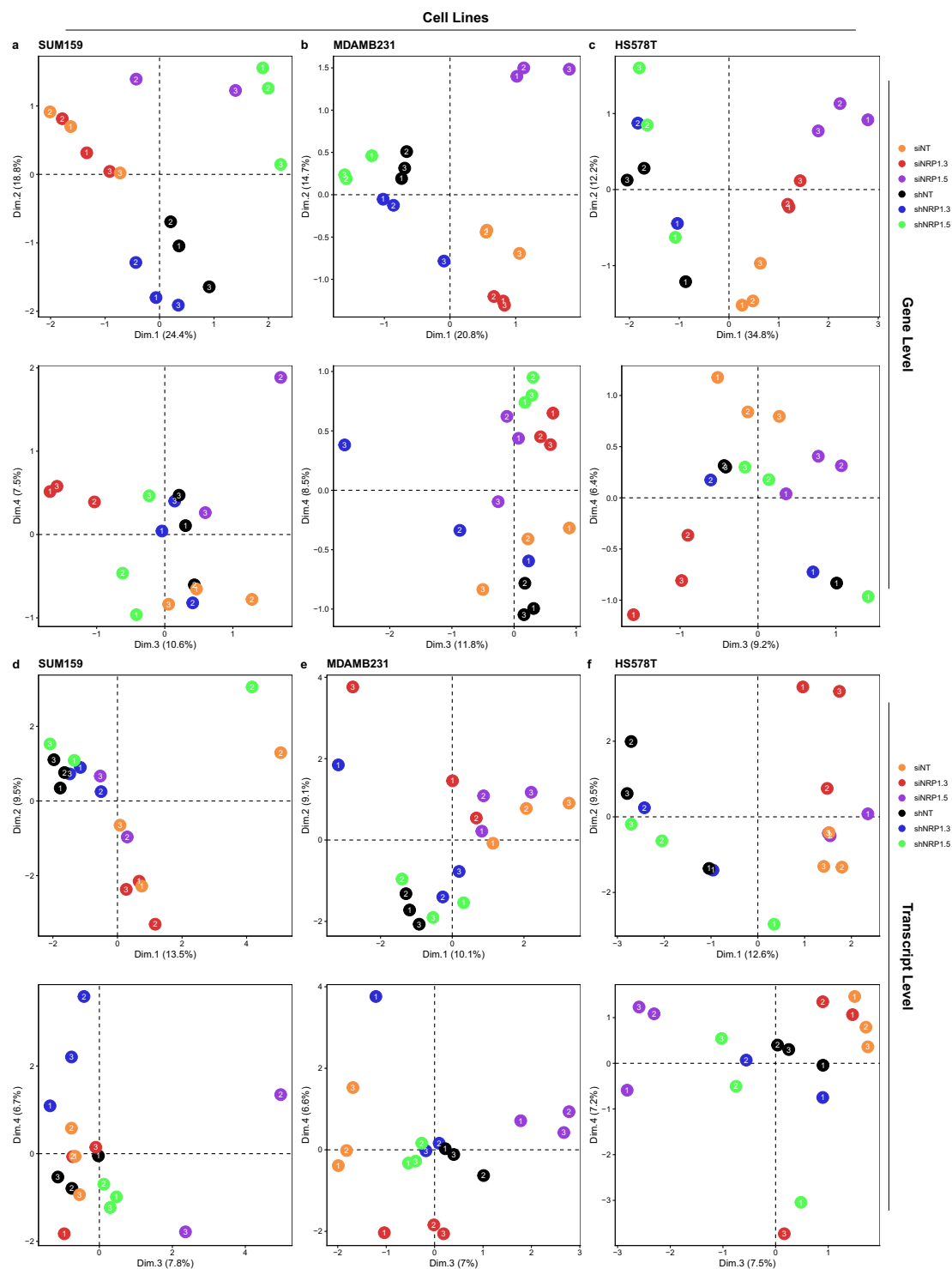


**Fig. 4** Multidimensional scaling (MDS) of all samples in the claudin-low BrCa dataset. (a) Gene-level and (b) transcript-level MDS plot based on TMM-normalised counts. Circles, squares and triangles represent SUM159PT, MDA-MB-231 and HS578T claudin-low breast cancer cell line samples, respectively. Colouring is by treatment group, with either siNT, siNRP1#3, siNRP1#5, shNT, shNRP1#3 or shNRP1#5. Data point numbers represent independent biological replicates. MDS analysis was performed using the R package edgeR (v.4.4.2) and plotted with ggplot2 (v.4.0.0).

and siRNA#3 sequences in knocking down only the transmembrane isoform of NRP1, and not soluble-NRP1 (Fig. 6b,c). Despite this specificity, note that the siNRP1#3 and shNRP1#3 sequences differ from each other, with their mapping sites being offset by about 20 nucleotides and hence, are targeting a slightly different subset of transcript variants (Fig. 6c). On the other hand, the sequences for siNRP1#5 and shNRP1#5 are identical, and thus, target exactly the same subset of transcript variants including transmembrane- and soluble-NRP1 isoforms.

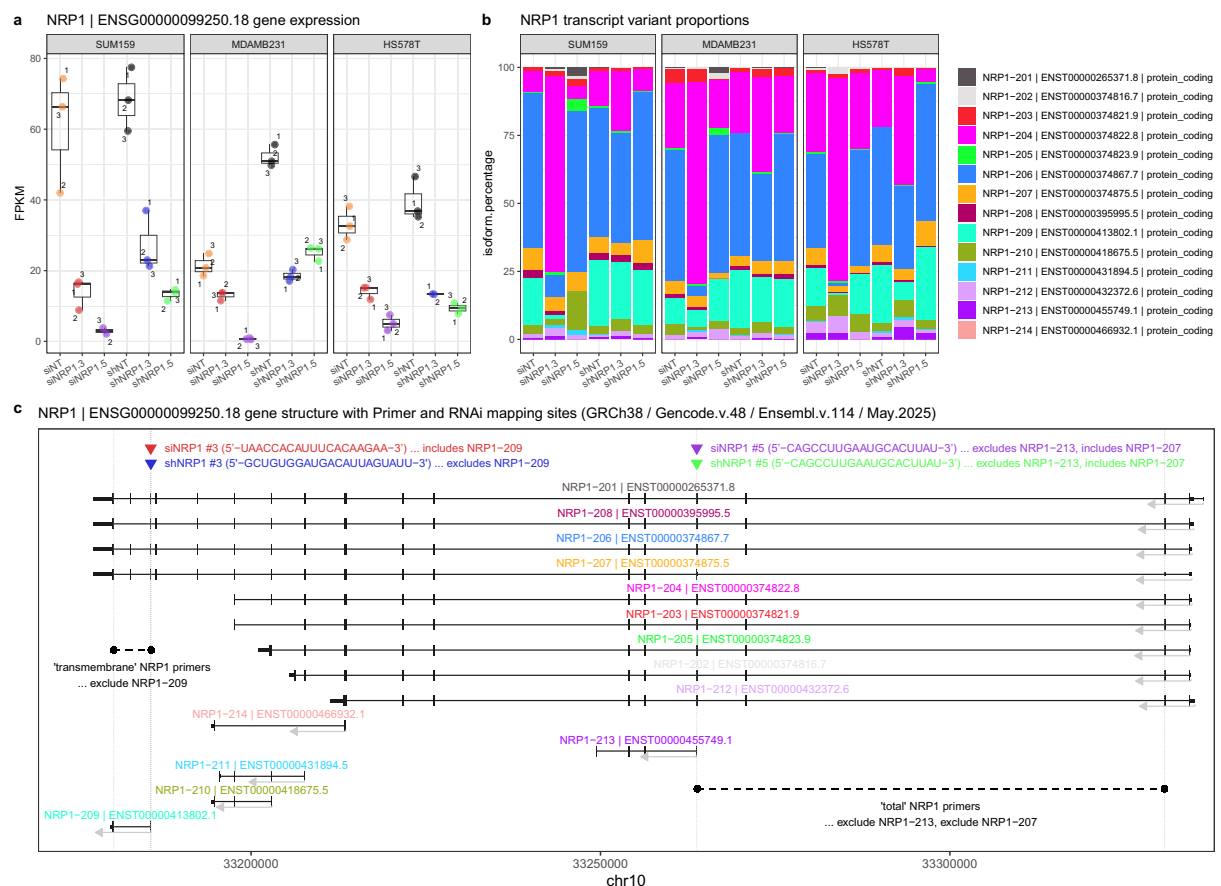
### Usage Notes

The quality validation results demonstrated that this is a robust and reliable dataset for exploring the transcriptional changes in response to NRP1 knockdown across multiple claudin-low breast cancer cell lines on the level of genes or individual transcript variants. Additionally, the non-targeting (NT) control RNAi samples enable the investigation and comparison of the transcriptional landscape of these claudin-low cell lines. However, it should



**Fig. 5** Multidimensional scaling (MDS) of the claudin-low BrCa dataset by cell line. Gene-level MDS plots for (a) SUM159PT, (b) MDA-MB-231, (c) HS578T as well as transcript-level MDS plots for (d) SUM159PT, (e) MDA-MB-231, (f) HS578T claudin-low cell lines, based on TMM-normalised counts, coloured by treatment group, with either siNT, siNRP1#3, siNRP1#5, shNT, shNRP1#3 or shNRP1#5. Data point numbers represent independent biological replicates. MDS analysis was performed using the R package edgeR (v.4.4.2) and plotted with ggplot2 (v.4.0.0).

be noted that lentiviral transduction and reverse-transfection alone can alter cellular transcriptomes. Raw data files, as well as processed/normalised data has been provided in the GEO entry so users can choose to either run their preferred raw data processing pipeline with custom parameters or utilise the processed data directly.



**Fig. 6** RNA-seq validation of NRP1 gene and transcript variant knockdown in claudin-low breast cancer cell lines. **(a)** Box plot of NRP1 gene expression (FPKM) and **(b)** bar chart of the isoform percentages (%) of NRP1 transcript variants in SUM159PT, MDA-MB-231 and HS578T cell lines treated with either siNT, siNRP1#3, siNRP1#5 or shNT, shNRP1#3 or shNRP1#5. **(c)** Schematic of NRP1 gene locus (as per Ensembl.v.114), showing the mapping sites of the total and transmembrane NRP1 primers, as well as the RNAi mapping sites. Note that the sequences for siNRP1#5 and shNRP1#5 are identical, whereas siNRP1#3 and shNRP1#3 sequences differ from each other, with their mapping sites being offset by about 20 nucleotides and hence, are targeting a slightly different subset of transcript variants. Plots were generated using the R packages ggplot2 (v.4.0.0) and transPlotR (v.0.0.2)<sup>47</sup>.

### Data availability

This bulk mRNA-seq dataset with 52 samples is available from the NCBI Gene Expression Omnibus as GEO accession number [GSE266566](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266566)<sup>44</sup>. The GEO entry includes 104 de-multiplexed raw FASTQ files from paired-end mRNA-seq, a metadata file describing the samples and the experimental details, as well as RSEM-derived transcript- and gene-level raw counts, TPM values and isoform percentages. Raw data is provided on GEO as a link to the SRA database.

### Code availability

The code used to process the data and generate the figures shown in the manuscript is provided on GitHub: [https://github.com/rocanja/GSE266566\\_NRP1\\_KD\\_BrCa\\_llynam](https://github.com/rocanja/GSE266566_NRP1_KD_BrCa_llynam).

Received: 23 October 2024; Accepted: 18 November 2025;

Published online: 01 December 2025

### References

1. Veerla, S., Hohmann, L., Nacer, D. F., Vallon-Christersson, J. & Staaf, J. Perturbation and stability of PAM50 subtyping in population-based primary invasive breast cancer. *NPJ Breast Cancer* **9**, 83, <https://doi.org/10.1038/s41523-023-00589-0> (2023).
2. Tang, Y. H. *et al.* Neuropilin-1 is over-expressed in claudin-low breast cancer and promotes tumor progression through acquisition of stem cell characteristics and RAS/MAPK pathway activation. *Breast Cancer Res* **24**, 8, <https://doi.org/10.1186/s13058-022-01501-7> (2022).
3. Jiang, Y. Z. *et al.* Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell* **35**, 428–440 e425, <https://doi.org/10.1016/j.ccell.2019.02.001> (2019).
4. Collignon, J., Lousberg, L., Schroeder, H. & Jerusalem, G. Triple-negative breast cancer: treatment challenges and solutions. *Breast Cancer (Dove Med Press)* **8**, 93–107, <https://doi.org/10.2147/BCTT.S69488> (2016).
5. Guo, L. *et al.* Local treatment for triple-negative breast cancer patients undergoing chemotherapy: breast-conserving surgery or total mastectomy? *BMC Cancer* **21**, 717, <https://doi.org/10.1186/s12885-021-08429-9> (2021).

6. Pommier, R. M. *et al.* Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nat Commun* **11**, 3431, <https://doi.org/10.1038/s41467-020-17249-7> (2020).
7. Agostinetto, E., Eiger, D., Punie, K. & de Azambuja, E. Emerging Therapeutics for Patients with Triple-Negative Breast Cancer. *Curr Oncol Rep* **23**, 57, <https://doi.org/10.1007/s11912-021-01038-6> (2021).
8. Liedtke, C. *et al.* Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* **26**, 1275–1281, <https://doi.org/10.1200/JCO.2007.14.4147> (2008).
9. Bianchini, G., De Angelis, C., Licata, L. & Gianni, L. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. *Nat Rev Clin Oncol* **19**, 91–113, <https://doi.org/10.1038/s41571-021-00565-2> (2022).
10. Badve, S. S. *et al.* Determining PD-L1 Status in Patients With Triple-Negative Breast Cancer: Lessons Learned From IMpassion130. *J Natl Cancer Inst* **114**, 664–675, <https://doi.org/10.1093/jnci/djab121> (2022).
11. Sabatier, R. *et al.* Claudin-low breast cancers: clinical, pathological, molecular and prognostic characterization. *Mol Cancer* **13**, 228, <https://doi.org/10.1186/1476-4598-13-228> (2014).
12. Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8**, R76, <https://doi.org/10.1186/gb-2007-8-5-r76> (2007).
13. Fougner, C., Bergholtz, H., Norum, J. H. & Sorlie, T. Re-definition of claudin-low as a breast cancer phenotype. *Nat Commun* **11**, 1787, <https://doi.org/10.1038/s41467-020-15574-5> (2020).
14. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352, <https://doi.org/10.1038/nature10983> (2012).
15. Pereira, B. *et al.* The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun* **7**, 11479, <https://doi.org/10.1038/ncomms11479> (2016).
16. Chaudhary, B., Khaled, Y. S., Ammori, B. J. & Elkord, E. Neuropilin 1: function and therapeutic potential in cancer. *Cancer Immunol Immunother* **63**, 81–99, <https://doi.org/10.1007/s00262-013-1500-0> (2014).
17. Schuch, G. *et al.* In vivo administration of vascular endothelial growth factor (VEGF) and its antagonist, soluble neuropilin-1, predicts a role of VEGF in the progression of acute myeloid leukemia *in vivo*. *Blood* **100**, 4622–4628, <https://doi.org/10.1182/blood.V100.13.4622> (2002).
18. Roth, L. *et al.* Transmembrane domain interactions control biological functions of neuropilin-1. *Mol Biol Cell* **19**, 646–654, <https://doi.org/10.1091/mbc.e07-06-0625> (2008).
19. Gu, C. *et al.* Characterization of neuropilin-1 structural features that confer binding to semaphorin 3A and vascular endothelial growth factor 165. *J Biol Chem* **277**, 18069–18076, <https://doi.org/10.1074/jbc.M201681200> (2002).
20. Lanahan, A. *et al.* The neuropilin 1 cytoplasmic domain is required for VEGF-A-dependent arteriogenesis. *Dev Cell* **25**, 156–168, <https://doi.org/10.1016/j.devcel.2013.03.019> (2013).
21. Prud'homme, G. J. & Glinka, Y. Neuropilins are multifunctional coreceptors involved in tumor initiation, growth, metastasis and immunity. *Oncotarget* **3**, 921–939, <https://doi.org/10.18632/oncotarget.626> (2012).
22. Glinka, Y. & Prud'homme, G. J. Neuropilin-1 is a receptor for transforming growth factor beta-1, activates its latent form, and promotes regulatory T cell activity. *J Leukoc Biol* **84**, 302–310, <https://doi.org/10.1189/jlb.0208090> (2008).
23. Matsushita, A., Gotze, T. & Korc, M. Hepatocyte growth factor-mediated cell invasion in pancreatic cancer cells is dependent on neuropilin-1. *Cancer Res* **67**, 10309–10316, <https://doi.org/10.1158/0008-5472.CAN-07-3256> (2007).
24. West, D. C. *et al.* Interactions of multiple heparin binding growth factors with neuropilin-1 and potentiation of the activity of fibroblast growth factor-2. *J Biol Chem* **280**, 13457–13464, <https://doi.org/10.1074/jbc.M410924200> (2005).
25. Ball, S. G., Bayley, C., Shuttleworth, C. A. & Kielty, C. M. Neuropilin-1 regulates platelet-derived growth factor receptor signalling in mesenchymal stem cells. *Biochem J* **427**, 29–40, <https://doi.org/10.1042/BJ20091512> (2010).
26. Rizzolio, S. *et al.* Neuropilin-1-dependent regulation of EGF-receptor signaling. *Cancer Res* **72**, 5801–5811, <https://doi.org/10.1158/0008-5472.CAN-12-0995> (2012).
27. Parikh, A. A. *et al.* Neuropilin-1 in human colon cancer: expression, regulation, and role in induction of angiogenesis. *Am J Pathol* **164**, 2139–2151, [https://doi.org/10.1016/S0002-9440\(10\)63772-8](https://doi.org/10.1016/S0002-9440(10)63772-8) (2004).
28. Miao, H. Q., Lee, P., Lin, H., Soker, S. & Klagsbrun, M. Neuropilin-1 expression by tumor cells promotes tumor angiogenesis and progression. *FASEB J* **14**, 2532–2539, <https://doi.org/10.1096/fj.00-0250com> (2000).
29. Bielenberg, D. R. *et al.* Semaphorin 3F, a chemorepulsant for endothelial cells, induces a poorly vascularized, encapsulated, nonmetastatic tumor phenotype. *J Clin Invest* **114**, 1260–1271, <https://doi.org/10.1172/JCI21378> (2004).
30. Ellis, L. M. The role of neuropilins in cancer. *Mol Cancer Ther* **5**, 1099–1107, <https://doi.org/10.1158/1535-7163.MCT-05-0538> (2006).
31. Al-Zeheimi, N., Gao, Y., Greer, P. A. & Adham, S. A. Neuropilin-1 Knockout and Rescue Confirms Its Role to Promote Metastasis in MDA-MB-231 Breast Cancer Cells. *Int J Mol Sci* **24**, <https://doi.org/10.3390/ijms24097792> (2023).
32. FastQC (Babraham Bioinformatics, 2010).
33. Trim Galore (Babraham Bioinformatics, 2012).
34. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, <https://doi.org/10.1093/bioinformatics/bts635> (2013).
35. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, <https://doi.org/10.1186/1471-2105-12-323> (2011).
36. R Core Team. *R: A Language and Environment for Statistical Computing*, <https://www.R-project.org/> (2025).
37. R Posit Team. *RStudio: Integrated Development Environment for R*. Posit Software, <http://www.posit.co/> (2025).
38. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).
39. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25, <https://doi.org/10.1186/gb-2010-11-3-r25> (2010).
40. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257, <https://doi.org/10.1186/s13059-019-1891-0> (2019).
41. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048, <https://doi.org/10.1093/bioinformatics/btw354> (2016).
42. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).
43. Wilke, C. *\_cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'\_*. R package version 1.2.0, <https://CRAN.R-project.org/package=cowplot> (2025).
44. Lynam, L. R. *et al.* RNA Sequencing dataset of Claudin-low Breast Cancer Cell Lines with Neuropilin-1 Knockdown. *NCBI GEO* <https://identifiers.org/geo/GSE266566> (2024).
45. EMBL's European Bioinformatics Institute. Gene: NRP1 (ENSG00000099250) summary. *Ensembl* <http://identifiers.org/ensembl:ENSG00000099250> (2024).
46. EMBL's European Bioinformatics Institute. O14786 NRP1\_HUMAN. *Uniprot* <http://identifiers.org/uniprot/O14786> (2024).
47. Lao, J. *\_transPlotR: Visualize Transcript Structures in Elegant Way\_*. R package version 0.0.2, <https://CRAN.R-project.org/package=transPlotR> (2022).

## Acknowledgements

We greatly acknowledge Victoria Coyne and the QUT Central Analytical Research Facility (CARF) team for performing the RNA library preparation and sequencing as well as the Translational Research Institute (TRI) for the laboratory resources necessary in the generation of this dataset. We acknowledge the funding for this project by the US Department of Defence Breast Cancer Research Program Breakthrough Level 2 Award (BC191350) to Brett Hollier.

## Author contributions

L.L. wrote the manuscript and performed all *in-vitro* laboratory work involved in the collection of the RNA samples within this dataset, including siRNA and shRNA knockdown of NRP1 in the three cell lines, RNA extraction, RT-qPCR and bioanalyzer analysis. A.R. contributed to the development and writing of this manuscript and performed all sequencing data processing. M.L. contributed to the design of the RNA sequencing data analysis pipeline and the development of this manuscript. Y.T. contributed to manuscript development, experimental/protocol design and data interpretation. M.L.N. contributed to manuscript editing and RT-qPCR data. M.V. guided the planning and procedures for the *in-vitro* work and contributed to the development of this manuscript. B.H. guided experimental planning and contributed to the development of this manuscript. C.N. and P.G. contributed to the development of this manuscript and provided the resources necessary for creating this dataset.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to B.G.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025