

# NRG1 Fusions in Breast Cancer

**Karen D Howarth**

University of Cambridge

**Tashfina Mirza**

University of Cambridge

**Susanna L Cooke**

University of Glasgow

**Suet-Feung Chin**

University of Cambridge

**Jessica C Pole**

University of Cambridge

**Ernest Turro**

University of Cambridge

**Matthew D Eldridge**

University of Cambridge

**Raquel Manzano Garcia**

University of Cambridge

**Oscar M Rueda**

University of Cambridge

**Chris Boursnell**

University of Cambridge

**Jean E Abraham**

University of Cambridge

**Carlos Caldas**

University of Cambridge

**Paul A W Edwards** (✉ [pawe1@cam.ac.uk](mailto:pawe1@cam.ac.uk))

University of Cambridge <https://orcid.org/0000-0002-4789-3374>

---

## Research article

**Keywords:** NRG1 fusions, breast cancer, NRG1, clinically actionable, MDA-MB-175, whole-genome sequencing, transcriptome sequencing

**Posted Date:** September 18th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-75553/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Breast Cancer Research on January 7th, 2021. See the published version at <https://doi.org/10.1186/s13058-020-01377-5>.

# Abstract

**Background** *NRG1* gene fusions may be clinically actionable, since cancers carrying the fusion transcripts can be sensitive to tyrosine kinase inhibitors. The *NRG1* gene encodes ligands for the HER2(ERBB2)-ERBB3 heterodimeric receptor tyrosine kinase, and the gene fusions are thought to lead to autocrine stimulation of the receptor. The *NRG1* fusion expressed in the breast cancer cell line MDA-MB-175 serves as a model example of such fusions, showing the proposed autocrine loop and exceptional drug sensitivity. However, its structure has not been properly characterised, its oncogenic activity has not been fully explained, and there is limited data on such fusions in breast cancer.

**Methods** We analysed genomic rearrangements and transcripts of *NRG1* in MDA-MB-175 and a panel of 572 breast cancers.

**Results** We found that the MDA-MB-175 fusion—originally reported as a *DOC4(TENM4)-NRG1* fusion, lacking the cytoplasmic tail of *NRG1*—is in reality a double fusion, *PPP6R3-TENM4-NRG1*, producing multiple transcripts, some of which include the cytoplasmic tail. We hypothesize that many *NRG1* fusions may be oncogenic not for lacking the cytoplasmic domain but because they do not encode *NRG1*'s nuclear-localised form. The fusion in MDA-MB-175 is the result of a very complex genomic rearrangement, which we partially characterized, that creates additional expressed gene fusions, *RSF1-TENM4*, *TPCN2-RSF1*, and *MRPL48-GAB2*.

We searched for *NRG1* rearrangements in 572 breast cancers subjected to genome sequencing and transcriptome sequencing and found four cases (0.7%) with fusions, *WRN-NRG1*, *FAM91A1-NRG1*, *ARHGEF39-NRG1* and *ZNF704-NRG1*, all splicing into *NRG1* at the same exon as in MDA-MB-175. However, the *WRN-NRG1* and *ARHGEF39-NRG1* fusions were out of frame. We identified rearrangements of *NRG1* in many more (8% of) cases that seemed more likely to inactivate than to create activating fusions, or whose outcome could not be predicted because they were complex, or both. This is not surprising because *NRG1* can be pro-apoptotic and is *inactivated* in some breast cancers.

**Conclusions** Our results highlight the complexity of rearrangements of *NRG1* in breast cancers, and confirm that some do not activate but inactivate. Careful interpretation of *NRG1* rearrangements will therefore be necessary for appropriate patient management.

## Introduction

The *NRG1* gene encodes ligands for the ERBB2(HER2)-ERBB3 heterodimeric receptor tyrosine kinase [1–3]. Gene fusions of *NRG1* such as *CD74-NRG1* and *SLC33A2-NRG1* have been found at low frequency in a wide range of carcinomas including lung, breast, colorectal, ovarian and pancreatic cancers and with a wide range of fusion partners [4–9, reviewed in 10]. Although the frequency is generally low, 0.2% overall in a wide range of carcinomas according to Jonna et al [9], it rises to roughly 25% in the rare invasive mucinous adenocarcinoma of lung [6].

The fusions that retain the EGF-like receptor-binding domain are thought to be activating and oncogenic by creating an autocrine loop in which the fused NRG1 protein stimulates the heterodimeric ERBB2-ERBB3 receptor [3, 7,11,12]. Such cancers might be very sensitive to inhibition of ERBB2-ERBB3 [13], and there are several recent reports of good responses in patients to anti-ERBB2 or anti-ERBB3 therapy, including anti-ERBB3 antibody and HER-family kinase inhibitors such as the pan-ERBB inhibitor afatinib. This has led to the proposal that *NRG1* fusions are clinically actionable [7, 8,14–18] and, as a result, NRG1 fusions have been included in the TAPUR study matching patients that have driver mutations to appropriate therapy (Clinical trials ref NCT02693535).

The *NRG1* fusion of the breast cancer cell line MDA-MB-175 [11,19, 20] was the first *NRG1* fusion reported and serves as a model of such fusions and the proposed autocrine loop. MDA-MB-175 cells secrete a fused NRG1 protein that was originally thought to be an isoform of NRG1 [11] but subsequently was reported to be a *DOC4* (now renamed *TENM4*)-*NRG1* fusion [19,20]. The cells secrete a fusion protein that stimulates ERBB3 phosphorylation when added to other cells [13], and they are very sensitive to tyrosine kinase inhibitors, being the cancer cell line most sensitive to a dual ERBB2-ERBB3 inhibitor in the survey of Wilson et al [13].

However, the structure of this *NRG1* fusion has not been completely described. It was reported to be a *DOC4/TENM4-NRG1* fusion, but the original cDNA sequence [11] is in fact a double fusion *PPP6R3-TENM4-NRG1* implying a complex genomic rearrangement. Furthermore, normal *NRG1* has many splice variants: the original fusion cDNA—and indeed many of the fusion transcripts described in clinical samples—represents only one of potentially many isoforms. In particular, the original cDNA lacked the cytoplasmic tail, and this has been hypothesized to enhance its oncogenic activity, since the cytoplasmic tail has been linked to pro-apoptotic activity of *NRG1* [21] (see also Discussion).

We report here the full transcript structure with alternatively-spliced variants, partial characterisation of the underlying complex genomic rearrangement, and other gene fusions from the same genomic regions that presumably result from the same complex rearrangement. We also report a search for *NRG1* fusions and rearrangements in nearly 600 breast cancers.

## Results

### The *NRG1* fusion in MDA-MB-175 is a double fusion *PPP6R3-TENM4-NRG1*

The *NRG1* fusion transcript of breast cancer cell line MDA-MB-175 was originally described as a fusion of *DOC4* (now *TENM4*, encoding teneurin transmembrane protein 4, also called *ODZ4*) to *NRG1* [11, 19, 20]. However, the original cDNA sequence Genbank AF009227 [11] is in fact a double fusion, 5' *PPP6R3-TENM4-NRG1* 3' (Supplementary Table 1; Fig. 1). The fusion cDNA comprises the first, non-coding exon of *PPP6R3* (protein phosphatase 6, regulatory subunit 3; formerly *SAPS3*), correctly spliced to exons 3 to 12 of *TENM4*, and then to exon 3 of *NRG1* (exons numbered according to their order in the genome, exon 3 is the second exon of many normal *NRG1* transcripts; Fig. 1, Supplementary Table 2). The *PPP6R3*-

*TENM4* (formerly *SAPS3-ODZ4*) cDNA junction was previously detected by Robinson et al [Supplementary data of ref. 22] in transcriptome sequencing, but recorded as a separate fusion.

We confirmed the double-fusion structure, first by RT-PCR and Sanger sequencing (Fig. 1 panels C, D) and paired-end sequencing of cDNA fragments ('RNA-seq') (Supplementary Figure 1), and finally by amplifying, cloning and Sanger sequencing complete cDNAs extending from exon 1 of *PPP6R3* through the *TENM4* component to various exons of *NRG1* (Fig. 1; Supplementary Textfile 1).

### **The fusion extends to the 3' end of *NRG1*, including the cytoplasmic tail.**

*NRG1* has many alternative isoforms [1]. Although the original cDNA, AF009227 [11], lacks the cytoplasmic exons 12 to 18, terminating in exon 11ext (Fig. 1; Supplementary Tables 1, 2), which defines normal  $\beta 3$  isoforms such as Heregulin- $\beta 3$  (RefSeq NM\_013958) and *NRG1* Type III- $\beta 3$ , we detected expression in MDA-MB-175 of all the later exons as well, in fusion transcripts (Fig. 1). We amplified complete cDNAs extending from *PPP6R3* exon 1 to the last *NRG1* exon, exon 18, which on cloning included at least three isoforms: we detected both the alpha and beta forms of *NRG1* (respectively including *NRG1* exon 10 or 11), and two transcripts included an additional, unannotated exon we designated 'exon int15'. This exon is in reverse orientation within intron 15 of *TENM4*, at hg19/chr11:78,506,385-78,506,462 (hg38 chr11:78795340-78795417) and therefore must be transcribed from an inverted fragment of chromosome 11 inserted into the main *TENM4-NRG1* junction (Fig. 1; Supplementary Table 2; Supplementary Textfile 1). RNAseq confirmed presence of this exon (Supplementary Figure 2 panel C), but showed it was a minor variant, as only 24% (38/160) of split reads across the junction with *NRG1* exon 3 were from this exon (the others were all from *TENM4* exon 12). This exon would insert 26 amino acids and preserve the reading frame downstream.

We also amplified the originally-reported transcript that terminates in an extension of exon 11, and a variant that included the inverted exon int15 (Fig. 1; Supplementary Textfile 1).

### **Other gene fusion transcripts: *RSF1-TENM4*, *TPCN2-RSF1*, *MRPL48-GAB2***

RNAseq revealed other cDNA junctions in MDA-MB-175 from further gene fusions caused by related rearrangements within chromosome 11, and we confirmed them by RT-PCR: *RSF1* exon 3 fused to exon 15 of *TENM4*; *TPCN2* exon 5 fused to *RSF1* exon 4; and *MRPL48* exon 3 fused to *GAB2* exon 2 (Supplementary Figure 3; Supplementary Figure 4). A normal transcript of *RSF1*, exons 1 to 5 was also detected by PCR, and RNAseq showed expression of all *RSF1* exons with normal splicing, the apparently-normal expression exceeding that of the fusions.

### **Expression of *NRG1*, *TENM4*, and *PPP6R3* other than in the fusion**

*NRG1*, *TENM4*, and *PPP6R3* are all expressed in normal breast, *PPP6R3* strongly, *NRG1* and *TENM4* weakly (GTEx RNAseq database accessed via UCSC Genome browser and [23]). We asked whether there was expression of unrearranged *NRG1*, *TENM4*, and *PPP6R3* in MDA-MB-175, or of a hypothetical precursor fusion, *PPP6R3-TENM4* or *TENM4-NRG1*, a copy of which might still be present. For *NRG1*, no

expression was detected from the major transcription start sites of *NRG1* exon1, exon 2, and exon 7, either by RT-PCR or transcriptome sequencing (RNA-seq; (Supplementary Figure 2)), and all splicing into *NRG1* exon 3 was from *TENM4* exons 12 or int15 (total 160 split reads). *NRG1* exon 2 is the main transcription start site in normal breast epithelial cells and carcinoma cell lines [23]. We previously showed that in MDA-MB-175 the CpG island at exon 2 is methylated [23]. However, a recently-described alternative minor transcription start site, exon 8, found in isoform ndf43c/VI-1 (Refseq NM\_001159996 and NM\_001322197) was weakly expressed (Supplementary Figures 1 and 2), and we amplified a cDNA extending from exon 8 to exon 18, comprising exons 8 to 10 and 13 to 18, and a second cDNA with exon 17 omitted. We presume these are normal transcripts, since splicing into exon 8 has not been described, and PCR failed to amplify between *TENM4* and exon 8 of *NRG1*.

Similarly, almost no expression was detected from *TENM4* exons 1 and 2 (Supplementary Figures 1 and 2; manual search found 5 split reads joining exons 1 and 2), although data from the Cancer Genome Atlas (TCGA) shows that these are expressed in other breast cancer cell lines such as MCF7 and MDA-MB-134, and we failed to amplify cDNA joining exon 2 of *TENM4* to *NRG1*. By RNAseq, only 1 of 145 split reads that included the splice acceptor of *TENM4* exon 3 matched exon 2 splice donor; all others were consistent with splicing in from *PPP6R3* (two included an alternative exon 2 of *PPP6R3* and two reads included intronic sequence from upstream of exon 3 in *TENM4*). This seems to rule out significant expression of both normal *TENM4* and a *TENM4-NRG1* precursor. However, there was good expression of *TENM4* exons beyond the last exon involved in the fusion, exon 12, and splicing was normal, including exon12-exon13. Although some of this expression would be the *RSF1-TENM4* fusion (Supplementary Figures 1 and 2), this only starts at *TENM4* exon 15 and there is splicing exon14 to exon 15, and no other splicing anomalies were detected: a *PPP6R3-TENM4* fusion may therefore remain. We failed to amplify between *PPP6R3* exon 1 and *TENM4* exon 16, but this might have been technical failure.

Expression of *PPP6R3* exons not in the fusion was also evident, presumably from a normal copy (Supplementary Figure 1).

### **Genome rearrangements associated with the fusions**

We were able to demonstrate some of the complexity of the underlying genomic rearrangements that formed the *PPP6R3-TENM4-NRG1* fusion, by constructing a plausible, though incomplete, model of them from the gene fusions, data from FISH (fluorescence-in situ hybridisation), paired-end sequencing of part of the *NRG1* region, and copy number.

The fusion cDNA suggests that it was formed by a large inversion of chromosome 11 that joined *PPP6R3* to *TENM4* and a translocation joining chromosome 11 to chromosome 8 upstream of exon 3 of *NRG1*, with an additional local inversion that allows exon int15 to be included in some transcripts (Fig. 2).

Cytogenetically, a typical metaphase of MDA-MB-175 has two copies of an unbalanced 8;11 translocation, plus apparently-normal chromosomes 8 (three copies) and 11 (two copies) [19,24]. FISH with BAC clones showed that, as expected, the 8;11 translocation chromosome harbours all the three

major genomic segments that form the fusion, apparently in more than one copy, with the *PPP6R3*, *TENM4* and *NRG1* regions colocalised, while the normal chromosomes retain single copies (Supplementary Figure 5).

We identified a number of rearrangement junctions in genomic DNA that are consistent with the fusions, by paired-end sequencing of DNA captured by hybridisation from around exons 2 to 6 of *NRG1* (hg19/chr8: 31.7 to 31.874 and 32.14 to 32.50 Mb). This identified junctions within chromosome 8 and between this region and chromosome 11 (Fig. 2). These were curated manually using the Integrative Genomics Viewer (IGV), revealing an additional small inversion on chromosome 11 apparently encompassing the 8-11 junction that forms the fusion. Genomic copy number analysis by counting reads from the capture-sequencing (Supplementary Figure 6) confirmed the presence of unbalanced junctions corresponding to the rearrangement junctions in the region captured: up at junction marked F in Fig. 2; down at junction D; and up at the cluster of breakpoints marked A1, A2, B, C and E.

The rearrangements shown are only part of a more complex picture. Array-CGH (array-comparative genomic hybridization) from [25] suggests multiple copy number steps in the rearranged regions (Supplementary Figure 6), and the additional fusions found suggest additional rearrangements, shown as dotted lines in Fig. 2. Our capture sequencing would not have found junctions that did not include the captured *NRG1* region—notably the *TENM4-RSF1* and *PPP6R3-TENM4* junctions.

**Overexpression of the fusion.** We transfected, into HEK293 cells, FLAG-tagged coding sequences of the principal isoforms of the fusion, with and without the cytoplasmic tail. FLAG-tagged protein of the expected size was detected on harvesting cells at 48 hrs (Supplementary Figure 7), showing that such isoforms can be expressed at least transiently. We also transfected isoforms that included the extra inverted exon int15, but these were not detectably expressed.

### ***NRG1* fusions in primary breast cancers**

To put the MDA-MB-175 fusion in the context of breast cancer, we surveyed 572 consecutive consented cases of breast cancer subjected to whole-genome DNA and RNA sequencing. We identified four *NRG1* fusions of the form (*geneA*)-*NRG1*, that were both predicted from DNA rearrangement junctions and found in RNA sequence reads: *WRN-NRG1*, *FAM91A1-NRG1*, *ARHGEF39-NRG1* and *ZNF704-NRG1* (Supplementary Table 3). However, while the *FAM91A1-NRG1* fusion was in frame, the *WRN-NRG1* and *ARHGEF39-NRG1* fusions were not. The *ZNF704-NRG1* fusion was to an undocumented exon in *ZNF704*, so its reading frame is unknown, but the fused sequence was in frame with *NRG1*. All four fusions spliced into *NRG1* at exon 3, as in MDA-MB175, and many of the fusions described by others; no fusions that spliced into exons 4 to 9 (the receptor-binding exon) were detected. All were created by internal rearrangement of chromosome 8 (Supplementary Table 3). A *WRN-NRG1* fusion has been described before—*WRN* is the gene immediately 5' to *NRG1* so the fusion is typically formed by genomic deletion—but the previous example included no *WRN* coding sequence, so would presumably have resulted in expression of *NRG1* protein [5, 9]. A fusion of *ZNF704* has also been reported, but to *MYC*, in lung [5].

## Discussion

We have shown that the *NRG1* fusion of MDA-MB-175 is more complex than previously described, being a double fusion *PPP6R3-TENM4-NRG1* with multiple alternative transcripts, some including the cytoplasmic tail; and it is the result of complex genomic rearrangements. We also confirmed that similar fusions—coding sequence of another gene splicing into genomic exon 3 of *NRG1*—are found in breast cancers, supporting the use of this fusion as a model example. The structure of these fusions has implications for clinical identification of *NRG1* fusions, for understanding the subcellular location and secretion of *NRG1* fusion proteins, and explanations of their oncogenicity.

### Identifying *NRG1* fusions is challenging

Our search for *NRG1* fusions in breast cases, and the complexity of the fusion in MDA-MB-175, illustrate that identifying *NRG1* fusions in clinical cases is not straightforward. Although we found 4 examples in 572 cases (0.7%), in rough agreement with the 2/120 found by Kim et al [26] but a substantially higher prevalence than others [8,9], there might well have been more. None of the four fusions in cancers had been called from RNA sequencing alone by the fusion detection software STAR-fusion, presumably because of insufficient read coverage—indeed, in two of four cases, our RNA sequencing yielded only one or two split fusion reads (Supplementary Table 3). Without corresponding evidence from DNA sequencing these fusions would not have been discovered. Similarly, the DNA rearrangements alone would not have specifically identified these cases. The prediction of the *ZNF704-NRG1* fusion was tentative because the genomic rearrangement is complex. A plausible reconstruction (Supplementary Figure 8) is that in addition to the *ZNF704-NRG1* junction, there is a tandem duplication of about 57kb of *NRG1*, encompassing the unused exon 7, with insertion of 24kb of inverted sequence into the duplication junction. Furthermore, splicing into exon 3 of *NRG1* is from an undocumented exon in *ZNF704*, so the fusion would not have been found in the RNA sequencing using software that assumed splicing would be to known splice sites. A further 20 of the first 250 breast cancer cases had breakpoints within *NRG1* by DNA sequencing, but without detectable fusion, 13 of which had multiple breakpoints, which would make fusion prediction difficult (Supplementary Table 4). No anomalous splicing into *NRG1* was detected in these cases. MDA-MB-175 itself is a case in point: the complex rearrangements of *NRG1*, with 7 breakpoints called within *NRG1* (Fig. 2), would not have led to a confident prediction of a fusion.

The importance of correct interpretation is underlined by the probability that some *NRG1* rearrangements—including presumably the out-of-frame fusions—are *inactivating* events as discussed below. Probably our [28] and others' [29] estimates of around 5% of breast cancers having breaks within *NRG1* by FISH includes many cases where there is no fusion. In conclusion, (as noted before [6,8-10]) RNA analysis is probably necessary, and combining with DNA sequencing improves sensitivity and specificity, but, even with both, sensitive identification of fusions is challenging.

### Structure of the MDA-MB-175 fusion

The fusion partners *TENM4* and *PPP6R3* have not been seen in *NRG1* fusions in tissue samples, but this is not surprising, because there are already upwards of 30 known fusion partners (e.g. [9]). *TENM4*, teneurin4, has been identified as a probable driver target of structural variation, notably in breast [30] and Fig. 3b of ref. [31] and its relative *TENM1/ODZ1* was identified as an oncogene target of the mouse mammary tumour virus (MMTV) [32]. It is a transmembrane protein with a cytoplasmic N-terminus and a large extracellular domain, most of which is lost in the fusion (Fig. 3).

An important feature of the *PPP6R3-TENM4-NRG1* fusion is that, paralleling wild-type *NRG1*, we found multiple isoforms, including isoforms with the cytoplasmic tail (Fig. 1). The original cDNA cloned by Schaefer et al [11] lacked the cytoplasmic domain of *NRG1*, and it has often been assumed that this was a feature of *NRG1* fusions in general. The isoforms we found (Fig. 1) all had the Ig-like domain, and they included both alpha and beta forms (alternative exons 10 and 11). Some had the full transmembrane and cytoplasmic C terminus designated 1a and 2a forms [1] while others, including the original cDNA of Schaefer et al [11], terminated in an extended exon 11, designated -β3.

### **Alternative splicing of other *NRG1* fusions**

Many fusions have been presented as lacking the C-terminal, cytoplasmic exons, and terminating in the β3, non-transmembrane terminus (genomic exon 11ext). But the multiple splice forms in MDA-MB-175 suggests that these other fusions will also come in multiple isoforms, including forms with the cytoplasmic tail. Their absence from the literature is probably an oversight: partly a legacy of the original reports [4,11]; and partly technical, because short-read sequencing only shows the fusion junction, not downstream splicing patterns, and PCR or single-primer amplification of cDNA has often used primers in the β3 terminus (extended exon 11) or the EGF-like domain (exon 9) (e.g. [5, 9]). Our RNA sequencing shows expression of the cytoplasmic tail exons—clearly in two cases, *FAM91A1-NRG1* and *ARHGEF39-NRG1* but not conclusively in the other two cases where there were too few reads to be confident—but we cannot tell whether these reads are from fusion transcripts, because there is also expression from the normal transcription start site, genomic exon 2 (possibly from contaminating normal tissue).

Further confusion arises because some *NRG1* fusions have been described as derivatives of *NRG1* TypeIII-β3, but this is misleading: no fusions involve the transcription start site, genomic exon 7, that defines TypeIII neuregulins/hereregulins; and many of the fusions include the Ig domains which are not in TypeIII-β3 [1].

It has also been assumed that the form of *NRG1* secreted into the medium by MDA-MB-175 is encoded by the original cDNA of Schaefer et al [11] but this may not be correct—it might be a cleaved fragment of a transmembrane isoform (Fig. 3).

### **Oncogenic Function of *NRG1* fusions is paradoxical.**

The oncogenic function of *NRG1* fusions is paradoxical and remains to be fully explained. The fusions apparently form an autocrine loop, stimulating the co-expressed ERBB-ERBB2/HER2 heterodimer [12,13].

But normal epithelia produce both NRG1 and its receptors [23,33], so why would *NRG1* fusions be oncogenic? And *NRG1* expression is pro-apoptotic when cDNAs are transfected into cells, including the breast cancer cell line MCF7 [34].

A possible resolution of this puzzle would be that NRG1 and its ERBB-family receptors are, in normal epithelium, produced by different cells, and/or on different faces of the cell [33], with co-expression in the same cell prevented by strong controls—perhaps leading to the apoptotic activity of transfected *NRG1* [34].

So why are *NRG1* fusions oncogenic? One previous hypothesis was that the cytoplasmic domain of *NRG1* is pro-apoptotic and is absent from the *PPP6R3-TENM4-NRG1* fusion [21]; our analysis rules this out.

We suggest two alternative explanations: alteration of expression, or alteration of subcellular localisation. Simplest would be altered regulation of *NRG1* expression, by placing it downstream of an unrelated promoter, allowing one cell to express ligand and receptor. This would be consistent with the wide range of fusion partners.

### **Loss of nuclear signalling?**

A more intriguing hypothesis is that the fusion proteins have a different subcellular distribution, and, specifically, that one route of nuclear signalling is lost.

*NRG1* encodes many isoforms and proteolytically-cleaved forms, secreted, membrane-bound, cytoplasmic and nuclear [1,35]. Among these, two entirely unrelated forms can signal to the nucleus: the cleaved cytoplasmic tail and the Type1- $\beta$ 3 form which includes the Ig-like and EGF-like domains (Fig. 3). The latter is intracellular because it lacks a transmembrane domain or signal sequence [2, 35], and it has been shown to translocate to the nucleus and alter gene expression [36,37]. Translocation is mediated by sequences around the Ig-like domain [36,38] (Breuleux et al [38] used a truncated 'heregulin-alpha' cDNA that lacked a transmembrane domain).

The *PPP6R3-TENM4-NRG1* fusion proteins consist of the intracellular part of *TENM4* and its transmembrane domain, joined to a range of essentially intact *NRG1* isoforms: the only exons of *NRG1* lost are the first two transcription-start exons (Fig. 3). Thus, *TENM4* brings a transmembrane domain to the fusion, and the Type1- $\beta$ 3 forms that would normally be intracellular presumably become extracellular (Fig. 3). Similarly, as noted by others (e.g. [5]), several fusion partners bring a transmembrane domain, including two of the commonest, *CD74* and *SLC3A2*. Other fusion partners have a signal sequence, e.g. *SDC4* [5], *CLU* [8], *ADAM9* [9], or fuse with loss of the Ig-like domain, e.g. some *CD74* fusions, *RBPMS*, *TSHZ2* [9], again denying  $\beta$ 3 forms access to the nucleus. However, this is not a universal feature of the fusions, e.g. the *FOXA1* and *ROCK1* fusions [8], so it may not be essential for oncogenicity.

### ***NRG1* can be a tumour suppressor or oncogene**

Although *NRG1* appears to be oncogenic in some tumours, it is *inactivated* in carcinomas at least as often as it is activated. *NRG1* is silenced by methylation in some breast and other carcinomas [23,39,40]; and seems to be at least one target of distal 8p loss, which is one of the most frequent large-scale events in carcinomas [41]. Many of the rearrangements in *NRG1* appear not to fuse the gene, or create a fusion that lacks the EGF-like, receptor-binding domain, or are simply out of frame. Examples include a deletion in a breast cancer that removes the ligand-binding domain [42] and three further inactivating deletions [43]; fusions that retain only the 5' end of *NRG1*, e.g. two described by Drilon et al [8]; and 3' fusions that splice in at the transmembrane domain [43]. Of 16 *NRG1* fusions found in TCGA RNA-seq data by Hu et al [44], only 6 appeared to be in-frame fusions of 3' *NRG1* that included the EGF-like domain: four retained only the 5' end, and most of the others appeared out-of-frame. Many of the rearrangements of *NRG1* that we found in breast cancers did not, or were unlikely to create an activating fusion (Supplementary Table 4), including the two fusions that were out of frame.

This dual role could be because high ERBB3 activity can be achieved in two ways: either *NRG1* is inactivated to permit high ERBB3 activity in all cells or at both faces, or to prevent *NRG1*'s pro-apoptotic activity [34] (which may be a manifestation of the same control); or *NRG1* can form an oncogenic autocrine loop, if control preventing co-expression can be broken. The lack of nuclear signalling by the Ig-like domain might be part of the control mechanism.

Whether or not this is the explanation, because many *NRG1* rearrangements seem to be inactivating, the correct identification of activating fusions may require care.

## Methods

MDA-MB-175 (ATCC catalogue HTB-25) was obtained from the collection of Dr. M. J. O'Hare, Ludwig Cancer Research (who also provided the immortalised normal breast cell line HB4A [45]) and is the same stock as we used for karyotyping [24]. Its authenticity is confirmed by the presence of the fusion. It was maintained in DMEM Glutamax +15% FBS.

Genome positions are given relative to hg19/GRCh37 unless otherwise marked.

Paired-end transcriptome sequencing (RNA-seq) for MDA-MB-175 used Illumina's stranded RNA kit with polyA selection. 41 million mapped reads were obtained after removing duplicates and were analysed with TopHat-Fusion [27]. Additional cell line transcriptome (RNA-seq) data was downloaded from the Cancer Genome Atlas (TCGA) project, using the Cancer Genomics Hub (now superseded by [www.cancer.gov](http://www.cancer.gov)) as mapped sequence reads.

Genomic DNA sequencing of MDA-MB-175: DNA was captured by hybridisation using Nextera Custom Target Enrichment kit (Illumina, Great Chesterford, UK). Nextera uses a modified Tn5 transposase to simultaneously fragment DNA and attach a transposon sequence to both ends of the fragments. Fragments were PCR amplified and barcoded in 11 cycles of PCR; quantified using Qubit HS dsDNA assay (Life Technologies, CA.) and 500ng pooled into a pool of twelve samples. 80-mer enrichment

probes were designed by Illumina to *NRG1* genomic regions, from hg19/chr8:31696790-31873798 and 32140458-32310458, both within intron 1, at intermediate probe density; and 32320000-32500000, at dense probe density, 85kb upstream of exon 2 to just beyond exon 6. Capture was performed twice to increase specificity. Enriched libraries were amplified using universal primers in 11 cycles of PCR, their quality assessed using Bioanalyser (Agilent Technologies, Ca.) and quantified using KAPA Library Quantification Kits (Kapa Biosystems, Ma.). Four capture reactions (48 samples) were pooled for 125-bp paired-end sequencing in a lane of Illumina HiSeq 2000. Structural variants were called as described [46] and calls manually inspected using the IGV (Integrative Genomics Viewer). Copy number was estimated from read counts using geneCN [47,48].

## **Breast cancers**

Data was from consecutive consented patients with a successful DNA extraction, in the Cambridge Personalised Breast Cancer Program, led by JEA and CC, to be described elsewhere. 89% tumours were primary, 11% metastatic. DNA and RNA from tumours and DNA from matched blood were paired-end sequenced by Illumina, Great Chesterford, UK, using respectively the TruSeq® DNA PCR-Free Library Preparation kit or the TruSeq® Stranded Total RNA Library Preparation Gold kit. For DNA, reads were 150bp to minimum 75X, typically 100X, coverage of tumour after alignment and removal of duplicates; for matched normal minimum 30X, typically near 40X. RNA sequencing was approximately 100 million 75-bp read pairs per tumour. Structural variants and copy number aberrations were called by Illumina using Manta [49] following alignment with Isaac [50] to GRCh38 with decoy sequences. Structural variants were further filtered to remove calls with any supporting reads in the matched normal; calls also found in the pooled matched blood normal samples; and calls involving unassembled or mitochondrial chromosomes [46].

Fusion transcripts were identified in individual RNA reads by text searching the original sequence (fastq) files for *NRG1* splice acceptor and splice donor sequences, and extracting adjacent sequences. DNA structural variant calls were available for 250 patients of which 235 had RNA sequencing; RNA sequences from a further 337 cases were searched for additional fusions and, in the two cases identified, the matching DNA sequences were analysed individually for rearrangements.

## **Abbreviations**

RNA-seq: short-read whole transcriptome sequencing

BAC: bacterial artificial chromosome

FISH: fluorescence-in situ hybridisation

## **Declarations**

**Ethics approval and consent to participate**

Ethical approval for the Cambridge Personalised Breast Cancer Program was granted by the East of England (Cambridge Central) Health Research Authority (refs 16/EE/0100 and 18/EE/0251).

### **Consent for publication**

Not applicable

### **Availability of data and materials**

Sequence data for MDA-MB-175 are available from the corresponding authors on reasonable request. The breast cancer data are not publicly available because of patient confidentiality, but application for access may be made jointly to CC Carlos.Caldas@cam.ac.uk and JEA ja344@medschl.cam.ac.uk.

### **Competing interests**

Subsequent to this work, KDH and JCP were employed respectively by Inviata and Illumina, which have interests in cancer diagnostics.

### **Funding**

Supported by grants from Cancer Research UK and Breast Cancer Now (as Breast Cancer Campaign) to PAWE and from Cancer Research UK, Addenbrookes Charitable Trust and the Mark Foundation to CC and JEA.

### **Authors' contributions**

KDH, TM, SLC, S-FC and JCP performed experiments. MDE and ET analysed cell line sequencing. JEA and CC lead the Personalised breast cancer program and RMG, OR, and CB analysed that data. CC and PAWE managed the project. KDH and PAWE conceived the study, analysed data and wrote the manuscript. All authors read and approved the final manuscript.

### **Acknowledgements**

We thank the patients and staff of the Cambridge Breast Cancer Research Unit for providing and collecting tumours for sequencing, and CRUK Cambridge Institute Core Genomics for cell-line sequencing.

## **References**

1. Falls DL. Neuregulins: functions, forms, and signaling strategies. *Exp Cell Res.* 2003;284:14–30.
2. Hayes NVL, Gullick WJ. The neuregulin family of genes and their multiple splice variants in breast cancer. *J Mammary Gland Biol Neoplasia.* 2008;13:205–14.
3. Lemmon MA, Schlessinger J, Ferguson KM. The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harbor Perspectives in Biology.* Cold Spring Harbor Lab; 2014;6:a020768–8.

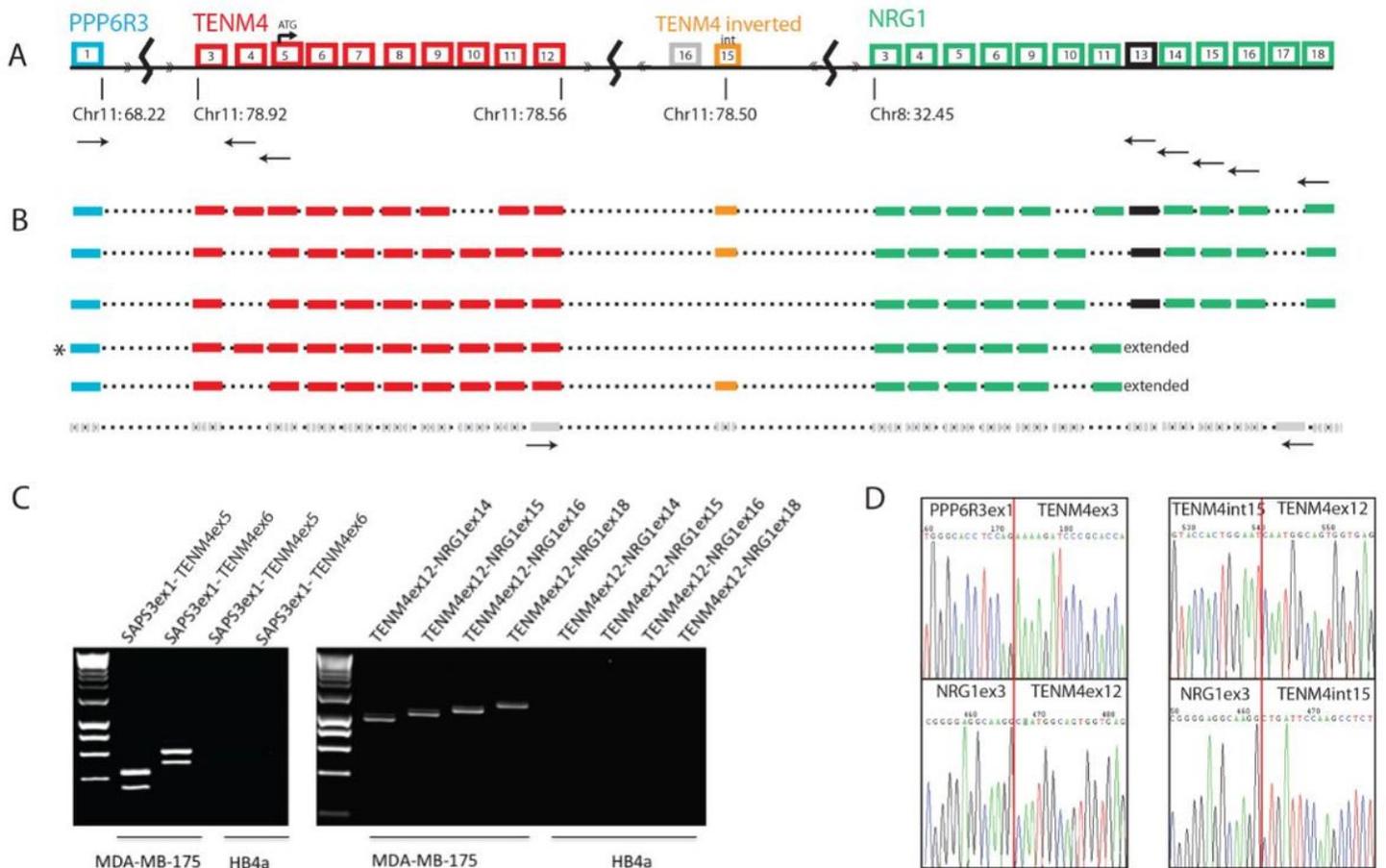
4. Fernandez-Cuesta L, Plenker D, Osada H, Sun R, Menon R, Leenders F, et al. CD74-NRG1 fusions in lung adenocarcinoma. *Cancer Discovery*. 2014;4:415–22.
5. Dhanasekaran SM, Alejandro Balbin O, Chen G, Nadal E, Kalyana-Sundaram S, Pan J, et al. Transcriptome meta-analysis of lung cancer reveals recurrent aberrations in NRG1 and Hippo pathway genes. *Nat Commun*. 2014;5:5893.
6. Shin DH, Lee D, Hong DW, Hong SH, Hwang J-A, Lee BI, et al. Oncogenic function and clinical implications of SLC3A2-NRG1 fusion in invasive mucinous adenocarcinoma of the lung. *Oncotarget*. 2016;7:69450–65.
7. Heining C, Horak P, Uhrig S, Codo PL, Klink B, Hutter B, et al. NRG1 Fusions in KRAS Wild-Type Pancreatic Cancer. *Cancer Discovery*. 2018;8:1087–95.
8. Drilon A, Somwar R, Mangatt BP, Edgren H, Desmeules P, Ruusulehto A, et al. Response to ERBB3-Directed Targeted Therapy in NRG1-Rearranged Cancers. *Cancer Discovery*. 2018;8:686–95.
9. Jonna S, Feldman RA, Swensen J, Gatalica Z, Korn WM, Borghaei H, et al. Detection of NRG1 Gene Fusions in Solid Tumors. *Clin Cancer Res*. American Association for Cancer Research; 2019;25:4966–72.
10. Nagasaka M, Ou S-HI. Neuregulin 1 Fusion-Positive NSCLC. *J Thorac Oncol*. 2019;14:1354–9.
11. Schaefer G, Fitzpatrick VD, Sliwkowski MX. Gamma-heregulin: a novel heregulin isoform that is an autocrine growth factor for the human breast cancer cell line, MDA-MB-175. *Oncogene*. 1997;15:1385–94.
12. Sheng Q, Liu X, Fleming E, Yuan K, Piao H, Chen J, et al. An activated ErbB3/NRG1 autocrine loop supports in vivo proliferation in ovarian cancer cells. *Cancer Cell*. 2010;17:298–310.
13. Wilson TR, Lee DY, Berry L, Shames DS, Settleman J. Neuregulin-1-Mediated Autocrine Signaling Underlies Sensitivity to HER2 Kinase Inhibitors in a Subset of Human Cancers. *Cancer Cell*. 2011;20:158–72.
14. Cheema PK, Doherty M, Tsao M-S. A Case of Invasive Mucinous Pulmonary Adenocarcinoma with a CD74-NRG1 Fusion Protein Targeted with Afatinib. *J Thorac Oncol*. 2017;12:e200–2.
15. Gay ND, Wang Y, Beadling C, Warrick A, Neff T, Corless CL, et al. Durable Response to Afatinib in Lung Adenocarcinoma Harboring NRG1 Gene Fusions. *J Thorac Oncol*. 2017;12:e107–10.
16. Jones MR, Lim H, Shen Y, Pleasance E, Ch'ng C, Reisle C, et al. Successful targeting of the NRG1 pathway indicates novel treatment strategy for metastatic cancer. *Ann Oncol*. 2017;28:3092–7.
17. Wilson FH, Politi K. ERBB Signaling Interrupted: Targeting Ligand-Induced Pathway Activation. *Cancer Discovery*. 2018;8:676–8.
18. Jones MR, Williamson LM, Topham JT, Lee MKC, Goytain A, Ho J, et al. NRG1 Gene Fusions Are Recurrent, Clinically Actionable Gene Rearrangements in KRAS Wild-Type Pancreatic Ductal Adenocarcinoma. *Clin Cancer Res*. 2019;25:4674–81.
19. Wang XZ, Jolicoeur EM, Conte N, Chaffanet M, Zhang Y, Mozziconacci M-J, et al. gamma-heregulin is the product of a chromosomal translocation fusing the DOC4 and HGL/NRG1 genes in the MDA-MB-

- 175 breast cancer cell line. *Oncogene*. 1999;18:5718–21.
20. Liu X, Baker E, Eyre HJ, Sutherland GR, Zhou M. Gamma-herregulin: a fusion gene of DOC-4 and neuregulin-1 derived from a chromosome translocation. *Oncogene*. 1999;18:7110–4.
21. Weinstein EJ, Leder P. The extracellular region of heregulin is sufficient to promote mammary gland proliferation and tumorigenesis but not apoptosis. *Cancer Res*. 2000;60:3856–61.
22. Robinson DR, Kalyana-Sundaram S, Wu Y-M, Shankar S, Cao X, Ateeq B, et al. Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat Med*. 2011;17:1646–51.
23. Chua YL, Ito Y, Pole JCM, Newman S, Chin SF, Stein RC, et al. The NRG1 gene is frequently silenced by methylation in breast cancers and is a strong candidate for the 8p tumour suppressor gene. *Oncogene*. 2009;28:4041–52.
24. Davidson JM, Gorringer KL, Chin SF, Orsetti B, Besret C, Courtay-Cahen C, et al. Molecular cytogenetic analysis of breast cancer cell lines. *Br J Cancer*. 2000;83:1309–17.
25. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, Andrews JM, et al. Signatures of mutation and selection in the cancer genome. *Nature*. 2010;463:893–8.
26. Kim J, Kim S, Ko S, In Y-H, Moon H-G, Ahn SK, et al. Recurrent fusion transcripts detected by whole-transcriptome sequencing of 120 primary breast cancer samples. *Genes Chromosomes Cancer*. 2015;54:681–91.
27. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol*. 2011;12:R72.
28. Huang H-E, Chin S-F, Ginestier C, Bardou V-J, Adélaïde J, Iyer NG, et al. A recurrent chromosome breakpoint in breast cancer at the NRG1/neuregulin 1/herregulin gene. *Cancer Res*. 2004;64:6840–4.
29. Prentice LM, Shadeo A, Lestou VS, Miller MA, deLeeuw RJ, Makretsov N, et al. NRG1 gene rearrangements in clinical breast cancer: identification of an adjacent novel amplicon associated with poor prognosis. *Oncogene*. 2005;24:7281–9.
30. Rebolledo-Jaramillo B, Ziegler A. Teneurins: An Integrative Molecular, Functional, and Biomedical Overview of Their Role in Cancer. *Front Neurosci*. 2018;12:683–20.
31. Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*. 2020;578:102–11.
32. Theodorou V, Kimm MA, Boer M, Wessels L, Theelen W, Jonkers J, et al. MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. *Nat Genet*. 2007;39:759–69.
33. Vermeer PD, Einwalter LA, Moninger TO, Rokhlina T, Kern JA, Zabner J, et al. Segregation of receptor and ligand regulates activation of epithelial growth factor receptor. *Nature*. 2003;422:322–6.
34. Weinstein EJ, Grimm S, Leder P. The oncogene heregulin induces apoptosis in breast epithelial cells and tumors. *Oncogene*. 1998;17:2107–13.

35. Talmage DA. Mechanisms of neuregulin action. In: Chadwick DJ, Goode J, editors. Growth Factors and Psychiatric Disorders: Novartis Foundation Symposium 289. John Wiley & Sons, Ltd; 2008. p. 74-86
36. Golding M, Ruhrberg C, Sandle J, Gullick WJ. Mapping nucleolar and spliceosome localization sequences of neuregulin1-beta3. *Exp Cell Res*. 2004;299:110–8.
37. Wang M, Trim CM, Gullick WJ. Localisation of Neuregulin 1-β3 to different sub-nuclear structures alters gene expression. *Exp Cell Res*. 2011;317:423–32.
38. Breuleux M, Schoumacher F, Rehn D, Küng W, Mueller H, Eppenberger U. Heregulins implicated in cellular functions other than receptor activation. *Mol Cancer Res*. 2006;4:27–37.
39. Oster B, Thorsen K, Lamy P, Wojdacz TK, Hansen LL, Birkenkamp-Demtröder K, et al. Identification and validation of highly frequent CpG island hypermethylation in colorectal adenomas and carcinomas. *Int J Cancer*. 2011;129:2855–66.
40. Fernandez SV, Snider KE, Wu Y-Z, Russo IH, Plass C, Russo J. DNA methylation changes in a human cell model of breast cancer progression. *Mutat Res*. 2010;688:28–35.
41. Pole JCM, Courtay-Cahen C, Garcia MJ, Blood KA, Cooke SL, Alsop AE, et al. High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation. *Oncogene*. 2006;25:5693–706.
42. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*. 2010;464:999–1005.
43. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*. 2012;486:353–60.
44. Hu X, Wang Q, Tang M, Barthel F, Amin S, Yoshihara K, et al. TumorFusions: an integrative resource for cancer-associated transcript fusions. *Nucleic Acids Res*. 2018;46:D1144–9.
45. Stamps AC, Davies SC, Burman J, O'Hare MJ. Analysis of proviral integration in human mammary epithelial cell lines immortalized by retroviral infection with a temperature-sensitive SV40 T-antigen construct. *Int J Cancer*. 1994;57:865–74.
46. Paterson AL, Weaver JM, Eldridge MD, Tavaré S, Fitzgerald RC, Edwards PAW, et al. Mobile element insertions are frequent in oesophageal adenocarcinomas and can mislead paired-end sequencing analysis. *BMC Genomics*. 2015;16:473.
47. Cooke SL, Ennis D, Evers L, Dowson S, Chan MY, Paul J, et al. The Driver Mutational Landscape of Ovarian Squamous Cell Carcinomas Arising in Mature Cystic Teratoma. *Clin Cancer Res*. 2017;23:7633–40..
48. Cooke SLC, Marshall J. GeneCN. 2020. <https://github.com/wwcrc/geneCN> accessed July 2020
49. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32:1220–2.

50. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics*. 2013;29:2041–3.
51. SMART database. <http://smart.embl-heidelberg.de>.

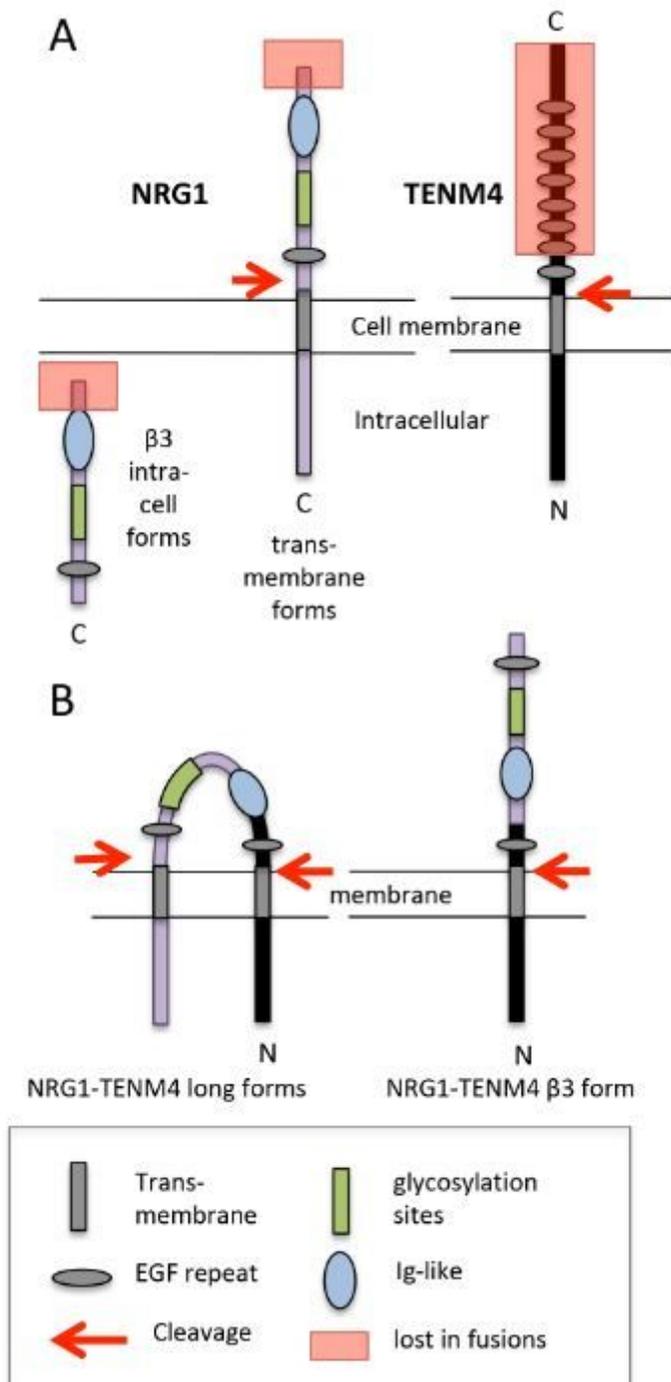
## Figures



**Figure 1**

Structure of PPP6R3-TENM4-NRG1 fusion transcripts expressed by MDA-MB-175. A, Schematic of genes involved, showing hypothetical genomic structure deduced from the transcripts. Exons from PPP6R3 are shown in blue, TENM4 exons, red, NRG1 exons, green with the transmembrane domain exon in black. Cryptic exon from TENM4 intron 15 in reverse orientation, orange. In grey, possible position of exon 16 of TENM4 for orientation, not present in fusion transcripts. NRG1 exons are numbered according to genome position. Exon 9 is the receptor-binding domain; exons 3 and 4 are the Ig-like domain. Exons 1,2,7, and 8 are omitted because they are normal transcription start sites and do not participate in the fusion. Where transcription terminates in exon 11, the exon is extended to give the  $\beta 3$  isoforms. Genomic positions are hg19. B, Transcripts detected. In colour, transcripts amplified by PCR and cloned; grey, additional isoform that includes exon 17, inferred from successful amplification between primers shown. Asterisk marks transcript matching original cDNA of Schaefer et al [11]. C, Examples of confirmation of junctions by PCR,

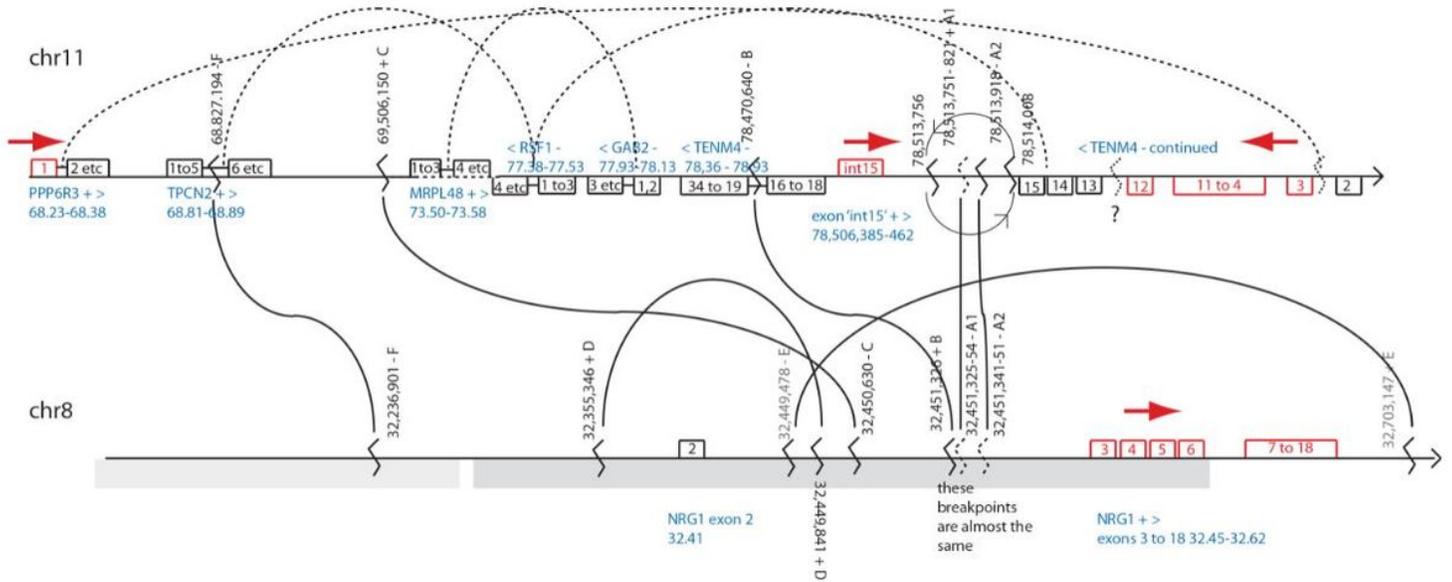
showing inclusion of cytoplasmic exons of NRG1 in fusion. Primer pairs are shown by arrows in A. HB4a is control normal breast cell line. D, sequences through junctions.



**Figure 2**

Partial genomic structure of rearrangements in MDA-MB-175 from capture sequencing and RNA fusions. Regions capture-sequenced shown as grey shading: lighter grey, 32.14 - 32.310 Mb, intermediate probe density; darker grey, 32.320 - 32.5 Mb, denser probes. All positions hg19. Solid arcs, breakpoint joins from capture sequencing, with positions of breakpoints given in bp, with direction of join and a one-letter

junction identifier, e.g. on chr8, junction marked '32,236,901 - F', is junction F at 32,236,901 bp, '-' indicating that sequence to the right of junction is retained. Dotted arcs, possible junctions in unsequenced regions, deduced from observed gene fusion transcripts. Blue, genes with their locations in Mb and +ve or -ve strand; numbered exons shown above line for + strand genes, below for - strand genes. Red, exons in fusion transcripts; red arrows, direction of transcription of fusion. Junctions on chromosome 8 correspond to the major copy number steps shown by the capture sequencing of this region of chromosome 8 (Supplementary Figure 6).



**Figure 3**

Predicted domain structure of the proteins encoded by the PPP6R3-TENM4-NRG1 fusion. Schematic diagrams showing the major domains detected by SMART [51]. A, normal structures: a typical transmembrane NRG1 and a  $\beta 3$  isoform, and TENM4 (PPP6R3 contributes only untranslated sequence). At least some  $\beta 3$  isoforms go to the nucleus [1,37]. B, Predicted structure of fusion proteins. One EGF-repeat and the transmembrane domain of TENM4 are retained, juxtaposed to the Ig-like domain of NRG1. The chimeric protein may either include the transmembrane and cytoplasmic region of NRG1, or the short  $\beta 3$  terminus.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1.docx](#)
- [AdditionalFile3sequencestextfilev2copy.pdf](#)
- [AdditionalFile2SupplementaryTables3.xlsx](#)
- [Additionalfile1combinedv2.pdf](#)