Discovery of NRG1-VII: a novel myeloid-derived class of NRG1 isoforms

Miguel Ángel Berrocal-Rubio[1], Yair David Joseph Pawer[1], Marija Dinevska[2], Ricardo De Paoli-Iseppi[1], Samuel S. Widodo[2], Nadia Rajab[1], Will De Nardo[1], Jeannette C. Hallab[1], Anran Li[1], Theo Mantamadiotis[2,3], Michael B. Clark[1], Christine A. Wells[1#].

1. Department of Anatomy and Physiology, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Australia.
2. Department of Surgery, Royal Melbourne Hospital, The University of Melbourne, Australia.
3. Department of Microbiology and Immunology, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Australia.

Correspondence to CAW wells.c@unimelb.edu.au @Mincle (Twitter)

Key words: NRG1, myeloid, growth factor, isoforms, macrophages

Abstract

Macrophages are a primary source of the growth factor Neuregulin-1 (NRG1), which has pleiotropic roles in proliferation and differentiation of the stem cell niche in different tissues, and has been implicated in gut, brain and muscle development and repair. Six isoform classes of NRG1 and over 28 protein isoforms have been previously described. Here we report a new class of NRG1, designated NRG1-VII to denote the start of this isoform class from a myeloid-specific transcriptional start site (TSS). Long-read sequencing identified up to nine different NRG1-VII transcripts that show major structural differences from one another due to use of cassette exons and alternative stop codons. Expression of NRG1-VII was confirmed in human monocytes and tissue resident macrophages. Isoform switching via cassette exon usage and alternate polyadenylation was apparent during monocyte maturation and macrophage differentiation. NRG1-VII is the major class expressed by the myeloid lineage, including tissue-resident macrophages. The size and structure of type VII isoforms suggests that they may be more diffusible than other NRG1 classes, however, the specific roles of type VII variants in tissue homeostasis and repair have not yet been determined.

Introduction

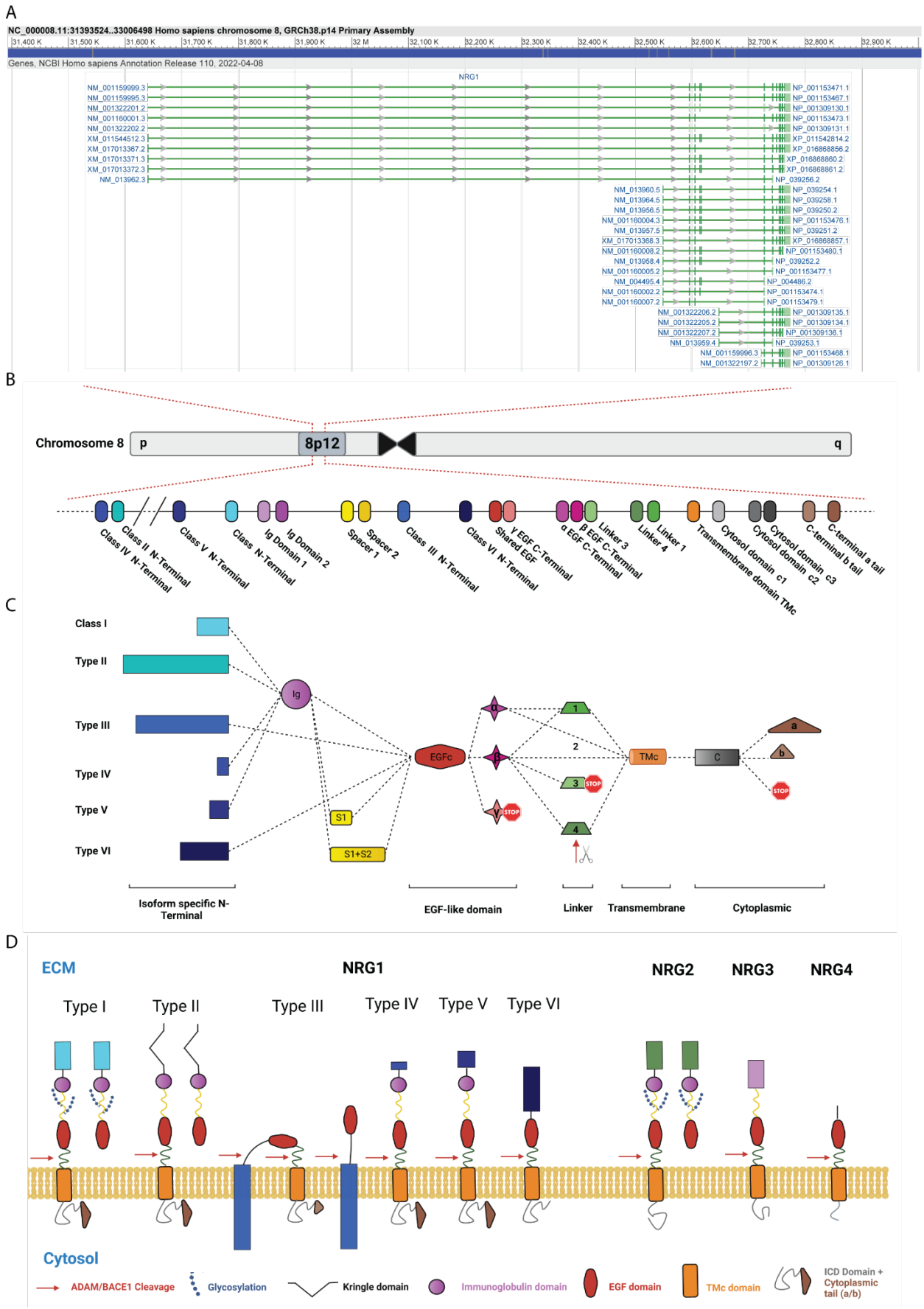
Neuregulins (NRGs) are a family of highly pleiotropic growth factors derived from four paralogous genes (*NRG1-4*) (Figure 1A). NRGs are typically synthesized as transmembrane pro-peptides cleaved by metalloproteases in the extracellular space to form a bioactive peptide with an exposed epidermal growth factor-like (EGF) domain that can bind ERBB receptors. The human *Neuregulin-1* (*NRG1*) locus, on Chromosome 8p12, generates a large number of isoforms (Figure 1A, B) which are thought to be tissue-specific, and functionally diverse (Falls, 2003). NRG1 has been implicated in the development of multiple tissues by promoting cell division within the stem cell niche and in differentiation trajectories (Yu et al., 2021, Wagner et al., 2007) of progenitor cells in the gut (Jardé et al., 2020), skeletal muscle (Gumà et al., 2010, Cheret et al., 2013) and cardiac cells (Wagner et al., 2007, Kramer et al., 1996), as well as nervous system development (Birchmeier 2009, Newbern et al., 2010). These studies support an important role for NRG1 in organogenesis and the importance of understanding how NRG1 isoforms exert tissue specific effects to maintain the adult stem cell niche.

Neuregulins interact with monomeric ERBB proteins (ERBB1-4), via an EGF-like domain, promoting dimerization and trans-phosphorylation of these receptors. Each ERBB monomer can bind different EGF members with differential ligand binding affinities, but only ERBB 3 and 4 can bind Neuregulins. Even though NRG1 and NRG2 can bind to both receptors as monomers (Carraway III et al., 1997), NRG3 and NRG4 only bind ERBB4 monomers (Harari et al., 1999, Zhang et al., 1997). Dimerized receptors can discriminate between different isoforms of NRG1, which is evidenced by differential phosphorylation on receptor tyrosine residues (Pinkas-Kramarski et al., 1998, Sweeney et al., 2000). Dimerization leads to the recruitment of different adaptor proteins, notably GRB or PI3K, which then trigger specific secondary cascades and regulate cell activity (Buonanno et al., 2001).

NRG1 is an important clinical target. Germline mutations in NRG1 are associated with developmental brain disorders such as schizophrenia (Craddock et al., 2005) as well as degenerative disorders such as Amyotrophic lateral sclerosis (Sun et al., 2020) and Alzheimer's disease (Go et al., 2005). Constitutively activated isoforms isoforms of NRG1 are implicated in cancer and blocking the NRG1 isoform Heregulin or its receptors (ERBB) are effective clinical strategies against solid tumours (Sheng et al., 2010, Zhang et al., 2022). Deficiencies in NRG1 are associated with Hirshprung's disease, leading to poor innervation of the gut (Tang et al., 2012, Garcia-Barcelo et al., 2009); as well as abnormal brain development and mental disorders like bipolar disorder or schizophrenia (Georgieva et al., 2008, Marballi et al., 2012). High levels of circulating NRG1 are associated with cardiac disease and morbidity following heart failure (Haller et al., 2022). NRG1 is also involved in modulating the immune response (Alizadeh et al., 2018, Ryzhov et al., 2017), controlling insulin related liver activity (Zhang et al., 2018), cell migration (Jumper et al.y, 2017) and cell-cell recognition and viability in the central nervous system (Garratt et al., 2000). Understanding which

cells express the different NRG1 isoforms in each tissue and how they play a role in different biological processes in said tissues is therefore of potential clinical benefit.

The diversity of roles in healthy development and disease across different organs is mirrored by the structural diversity and tissue specific transcriptional regulation of NRG1 products. The *NRG1* gene transcribes at least 28 isoforms (Figure 1B) according to the ENTREZ Gene reference transcript list (Brown et al., 2015). Others have reported higher isoform heterogeneity from this locus (Mei 2008). This extraordinary transcript diversity is due to the use of alternate transcription initiation sites, cassette exon usage, and alternate polyadenylation sequences. The locus is remarkably modular, which allows for different structural combinations of NRG1 domains depending on their start site (type I-VI), the use of alternative linker domains, which are the protease targets for ectodomain shedding (1-4), the type of EGF-like domain (α, β or γ), and whether the pro-peptide is membrane tethered or not (Figure 1A-D). One well characterized example is NRG1 type III, a nervous system specific isoform type that harbors a unique transmembrane 5' domain that is not present in any other known NRG1 isoform (Nave et al., 2006). This means that upon NRG1 processing, the growth factor domain will not be released from the source cell but will stay as a juxtracrine signal that allows neurons to establish key cell-cell interactions (with other neurons, oligodendrocytes, or muscle fibers) and controls survival of both interacting cells (Garratt et al., 2000, Wolpowitz et al., 2000).


A
NC_000008.11:31393524..33006498 Homo sapiens chromosome 8, GRCh38.p14 Primary Assembly
Genes, NCBI Homo sapiens Annotation Release 110, 2022-04-08
NRG1
B
Chromosome 8
p
8p12
q
Class IV N-Terminal
Class II N-Terminal
Class V N-Terminal
Class N-Terminal
Ig Domain 1
Ig Domain 2
Spacer 1
Spacer 2
Class III N-Terminal
Class VI N-Terminal
Shared EGF
γ EGF C-Terminal
α EGF C-Terminal
β EGF C-Terminal
Linker 3
Linker 4
Linker 1
Transmembrane domain TMc
Cytosol domain c1
Cytosol domain c2
Cytosol domain c3
C-terminal b tail
C-terminal a tail
C
Class I
Type II
Type III
Type IV
Type V
Type VI
Ig
S1
S1+S2
EGFc
α
β
γ
STOP
1
2
3
4
TMc
C
a
b
Isoform specific N-Terminal
EGF-like domain
Linker
Transmembrane
Cytoplasmic
D
ECM
NRG1
NRG2
NRG3
NRG4
Type I
Type II
Type III
Type IV
Type V
Type VI
Cytosol
ADAM/BACE1 Cleavage
Glycosylation
Kringle domain
Immunoglobulin domain
EGF domain
TMc domain
ICD Domain + Cytoplasmic tail (a/b)

95

Figure 1. Showcase of the human *NRG1* locus diversity. A) Screenshot of the NCBI RefSeq curated products from the *NRG1* locus showing 28 protein coding transcriptional variants (NCBI Gene website accessed 26/1/2023). Left and right labels correspond to transcript and protein accessions respectively. B) Schematic annotating exons in the human *NRG1* genomic locus (Chr 8p12) with modular protein coding domains. C) Schematic showing combinatorial protein domains for previously annotated NRG1 isoforms. Rows show the six major NRG1 classes, defined by alternate N-terminals. Columns show alternate domains: domains within a column are mutually exclusive. Dotted lines show known connections between domains in different NRG1 isoforms. Red stop symbols represent translation stop. Red arrows represent a protease cleavage point. D) Representation of Neuregulin protein isoforms and their domain distribution with respect to the cell membrane. Legend below the cell membrane describes symbols for ADAM/BACE1 cleavage site (red arrow), glycosylation site (blue chain), kringle domain (black V), IgG domain (purple circle), EGF domain (red oval), Transmembrane domain (orange rectangle), intracellular domain (grey squiggle, with or without brown triangle). N-terminal domains of NRG1 reference modularity shown in Fig. 1B and 1C, including the membrane-tethered N-terminal of NRG1 type III. Homologous domains in NRG2, NRG3 and NRG4 coloured by green, purple and orange respectively.

Transcripts arising from the human *NRG1* locus are differentially regulated by tissue type or developmental stage. Control over the six known isoform classes is achieved through differential proximal promoter usage at six unique TSSs, each of which is controlled by different Transcription Factors (TFs) (Frensing et al., 2008). Additional transcript processing can lead to alternative exon usage define the functional modules in the protein, like the linker or the cytoplasmic tails (Figure 1 C, D), mechanisms that are tissue-specific (Wen et al., 1994). The diversity of NRG1 isoform classes and their specificity of expression in both a temporal and anatomical manner suggest distinct roles of different isoforms in tissue patterning, especially in different brain areas (Liu et al., 2011).

NRG1 is conserved at the protein sequence level, but the function of specific isoforms has been shown to differ between species. For example, heart trabeculation failure, which leads to a lethal phenotype in mice at day 10.5, is a phenotype common to NRG1-/-, ERBB2-/- and ERBB4-/- mice (Kramer et al., 1996, Britsch et al., 1998). Although 11 different isoforms have been identified in the heart at this developmental stage, the phenotype can be attributed specifically to the absence of β-RGF types I and II, which are the isoforms containing the immunoglobulin (Ig) domain (Pentassuglia et al., 2009). However, the opposite is observed in zebrafish where Ig-like neuregulin-1 isoforms are dispensable for this process (Samsa et al., 2016). Thus, while essential processes in development can be attributed to specific subtypes of NRG1, differences among species showcase the need to characterize the different human isoforms using appropriate cell type and developmental models.

More recently, NRG1 has been described as an important factor regulating the stem cell niche in the gut (Jardé et al., 2020). In this context, NRG1 likely modulates stemness, proliferation and identity of progenitor cells in the niche, and it is required to recapitulate certain secretory and absorptive functions in human gut organoids (Kilik et al., 2021). This signal is important not only for correct development and tissue repair, but also during inflammation. NRG1 secreted by macrophages has roles in inflammation and tissue repair

(Jardé et al., 2020, Garrido-Trigo et al., 2022). However, the specific NRG1 isoforms expressed by macrophages in these contexts has not yet been described.

Understanding the cellular origins of different NRG1 isoforms is an important part of understanding how NRG1 directs these different developmental, reparative and inflammatory outcomes. Here, data mining to assess the expression profiles of *NRG1* led to the identification of a previously uncharacterized TSS that appears to be used exclusively in cells of the myeloid lineage. We propose that transcripts generated from this alternative TSS belong to a new NRG1 Class, NRG1-VII. Using Oxford Nanopore sequencing, we identified nine class VII isoforms with distinct transcript structures and predicted the protein characteristics of these isoforms. qRT-PCR targeting the unique first exon of *NRG1-VII* transcripts in human cells confirmed that type VII isoforms are expressed by monocytes and tissue resident macrophages. Immunohistochemistry using antibodies directed towards the EGF-like or intracellular domains (ICD) demonstrated that tissue-resident macrophages are the major source of NRG1 in these tissues, an observation supported by transcriptional evidence derived from single cell data collated within the Human Protein Atlas. This study therefore contributes to untangling the complexity of this already intricate locus by characterizing the structure and distribution of myeloid specific NRG1 isoforms.

Results

<u>NRG1-VII is defined by a novel TSS discovered in myeloid cells.</u>

Since the different isoforms of NRG1 have different structures and binding affinities for their receptors, we sought to investigate which of these were expressed by myeloid cells. First, we used the FANTOM5 (Functional Annotation of the Mammalian Genome) database (The FANTOM Consortium; 2014), a catalogue that maps the TSSs of genes expressed in 975 human samples (including cells, tissues, and cell lines). Here we found a previously uncharacterized TSS of NRG1 that was exclusively active in myeloid cells including monocytes, macrophages, and basophils (Figure 2A). A schematic of the locus suggested that this represents a potential new class of *NRG1* transcripts which we prospectively named *NRG1-VII* (Figure 2). The FANTOM TSS predicted that the *NRG1-VII* starting exon contained 147 bp of mRNA sequence unique to transcripts originating from this site, and an in-frame methionine was identified 162 bp downstream from the start of transcription (Figure 2B, C). Therefore, we aimed to characterize whether this newly described TSS would lead to the expression of a new class of protein coding transcripts.

The TSS for NRG1-VII initiates in the intron positioned 5’ of ENSEMBL exon ENSE00003743434 which encodes for part of the spacer sequence between the Ig and EGF-like domains in the canonical NRG1 protein (Figure 2B). The TSS adds a 5’ untranslated region (UTR) to the transcript that extends the exon from 51 bp to 190 bp (ENSE00002124728) and includes an initiating Methionine at nucleotide position 154-156 bp. Hence, a unique 20 nucleotide sequence could be found within this unique 5’ UTR that allowed

174 specific detection and amplification of the NRG1-VII isoforms (Figure 2C). The exon encodes 8 amino acids
175 in frame with the canonical *NRG1* coding regions and that are also present in other isoforms (I, II, IV and
176 V), thus NRG1-VII is the only reported NRG1 protein class that would not present a class-specific N-terminal
177 amino acid tail.
178
179 We mapped the transcriptional activity arising from this TSS and identified two Expressed Sequence Tags
180 (ESTs) that had been sequenced from the 5' end (NCBI accessions BI908144 and BI907799), and
181 originated from the *NRG1-VII* TSS. These originated from the same library (SAMN00164230) made from a
182 pool of non-activated human leukocytes from anonymous donors. Both ESTs harbored evidence of an open
183 reading frame (ORF) in frame with other NRG1 isoforms, but both were 3' truncated. To further validate the
184 potential activity of this TSS we looked for sequence evidence of conservation in other mammal species.
185 Data available from 15 non-Human primate species (Pipes et al., 2013), and cross species alignment shows
186 that the *NRG1-VII* TSS is conserved and highly specific to bone marrow and whole blood, while isoforms
187 isolated from other tissues use alternate TSSs (Figure S1A). We also found transcriptional evidence of an
188 equivalent TSS in *Mus musculus* and *Sus scrofa* derived from myeloid cells (Figure S1B-C) confirming that
189 the class VII TSS and its expression in myeloid cells are conserved through evolution.
190

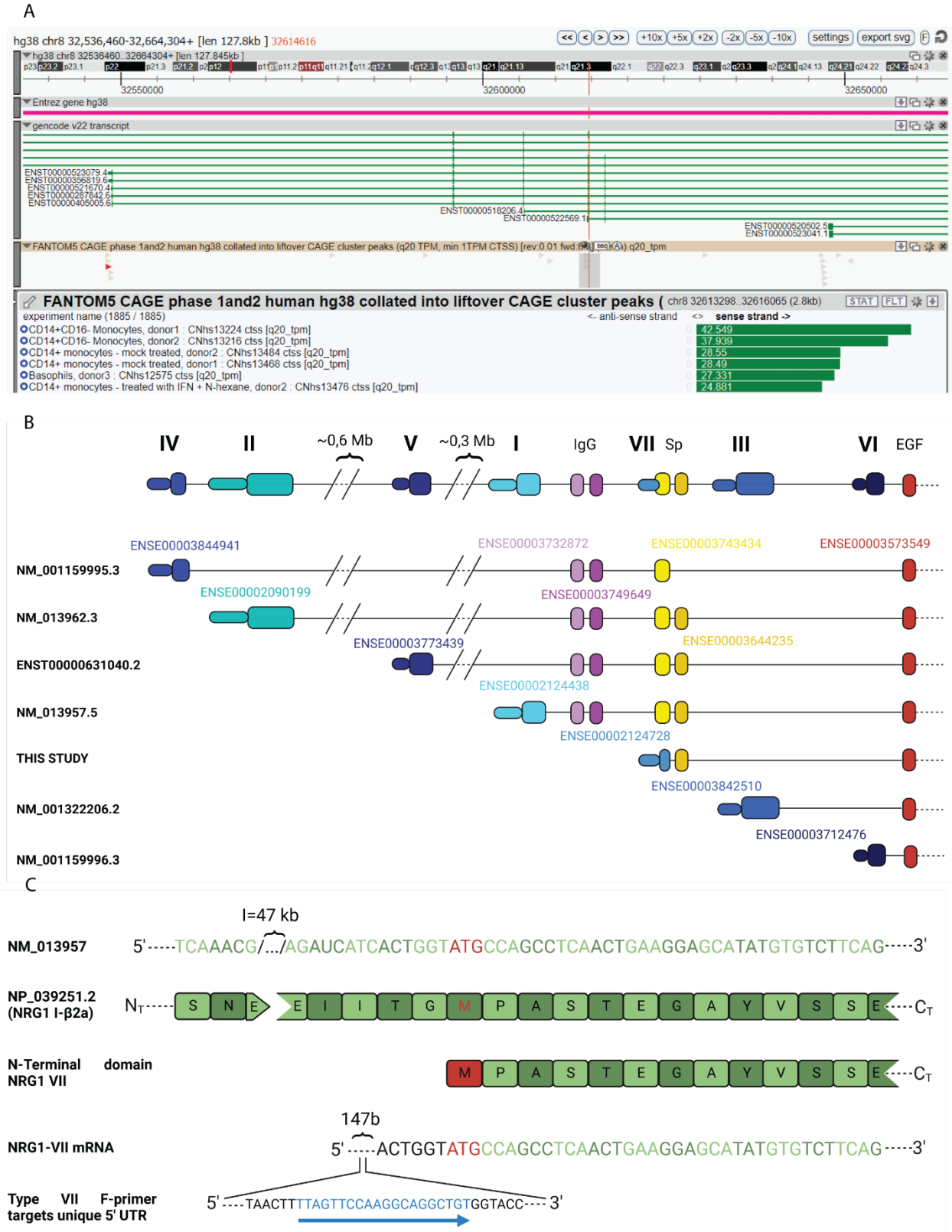
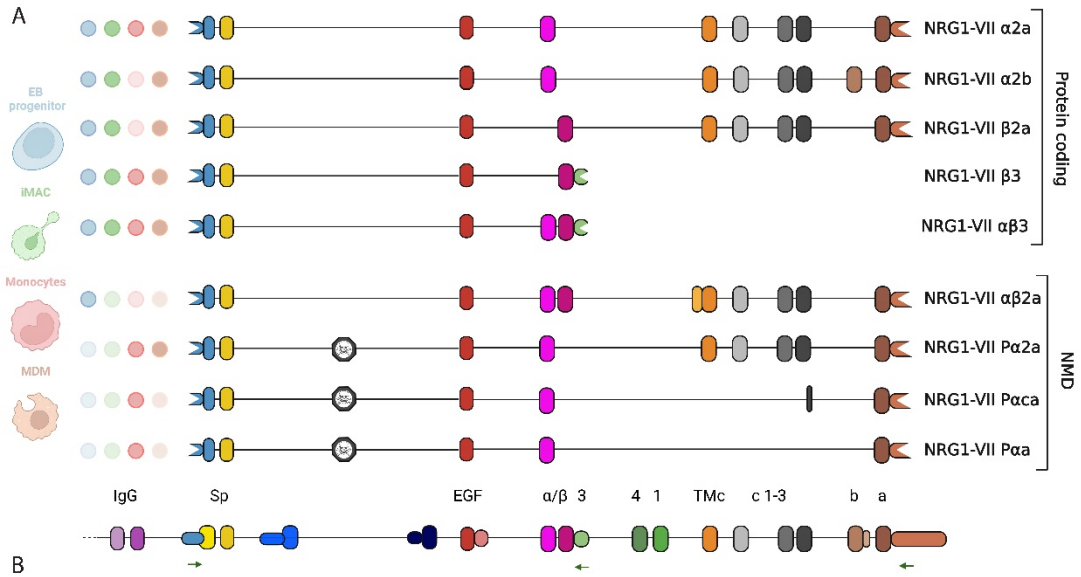

A
hg38 chr8 32,536,460-32,664,304+ [len 127.8kb] 32614616
settings
export svg
Entrez gene hg38
gencode v22 transcript
FANTOM5 CAGE phase 1and2 human hg38 collated into liftover CAGE cluster peaks (q20 TPM, min 1TPM CTSS) [rev:0.01 fwd:0.01] q20_tpm
FANTOM5 CAGE phase 1and2 human hg38 collated into liftover CAGE cluster peaks (chr8 32613298..32616065 (2.8kb)
experiment name (1885 / 1885)
<- anti-sense strand
sense strand ->
CD14+CD16- Monocytes, donor1 : CNhs13224 ctss [q20_tpm]
CD14+CD16- Monocytes, donor2 : CNhs13216 ctss [q20_tpm]
CD14+ monocytes - mock treated, donor2 : CNhs13484 ctss [q20_tpm]
CD14+ monocytes - mock treated, donor1 : CNhs13468 ctss [q20_tpm]
Basophils, donor3 : CNhs12575 ctss [q20_tpm]
CD14+ monocytes - treated with IFN + N-hexane, donor2 : CNhs13476 ctss [q20_tpm]
42.549
37.939
28.55
28.49
27.331
24.881
B
IV
II
~0,6 Mb
V
~0,3 Mb
I
IgG
VII
Sp
III
VI
EGF
NM_001159995.3
NM_013962.3
ENST00000631040.2
NM_013957.5
THIS STUDY
NM_001322206.2
NM_001159996.3
C
I=47 kb
NM_013957
5'-----TCAAACG/.../AGAUCATCACTGGTATGCCAGCCTCAACTGAAGGAGCATATGTGTCTTCAG-----3'
NP_039251.2 (NRG1 I-β2a)
N-Terminal domain NRG1 VII
147b
NRG1-VII mRNA
5'-----ACTGGTATGCCAGCCTCAACTGAAGGAGCATATGTGTCTTCAG-----3'
Type VII F-primer targets unique 5' UTR
5'----TAACTTTTAGTTCCAAGGCAGGCTGTGGTACC----3'

Figure 2. Identification of a myeloid specific TSS in the human NRG1 locus. A) Screenshot of the ZENBU genome browser showing the Gencode v22 transcript track, (accessed 26/1/23) corresponding to the FANTOM5 phase 1 and 2 TSS peaks overlapping the starting exon of NRG1 type VII. Red vertical line indicates the TSS of interest. Tag selection (grey bar, FANTOM5 CAGE track) indicates region for tag quantification in the FANTOM5 experiment table. The table is truncated to show the top 6 samples are primary myeloid cells, representative of the top 57/115 samples with >1TPM) Experimental samples are ordered by highest number of CAGE tags counted in this area on sense strand (green bars, showing counts as TPM). B) Schematic of the human *NRG1* locus highlighting alternate transcriptional starting exons that correspond to seven isoform classes. The schematic shows 5' exon composition until the first shared exon (containing the EGF domain). Exons are numbered by ENSEMBL accessions. C) Alignment of mRNA and protein sequences of NRG1 type I and type VII to show an in frame translation initiation of NRG1 type VII. Initiating methionine (M) or start codon (ATG) in red text. The blue arrow shows the target sequence of the forward primer for *NRG1-VII* mRNA's unique 5' UTR. See also Figure S1.

NRG1-VII TSS transcribes at least nine distinct transcripts in myeloid cells.

To characterize the diversity of *NRG1* transcripts that use novel *NRG1-VII* TSS, we performed Oxford Nanopore long-read amplicon sequencing on *in vitro* and *in vivo* derived myeloid cells (Figure 3A). We designed a forward primer that targets the unique 5' UTR of *NRG1-VII* transcripts. Two reverse primers were designed to target two of the known alternative transcriptional stops in which we had previously observed transcriptional activity in myeloid cells. We ran this test separately in *in vivo* and *in vitro* progenitors and differentiated cells. In total, we identified nine novel high confidence transcripts. Additionally, we identified a new exon present specifically in monocyte transcripts and positioned in the genome between TSS-3 and 6 of the *NRG1* locus. This exon introduces an early stop codon in frame; hence we classified it as a poison exon. We also report the presence of isoforms that contain both α and β EGF-like exons. This alternative splicing possibility had been previous captured in transcript NM_004495, but as this was a truncated transcript its ORF and TSS had not been defined.

We predict that five of these transcripts would be protein coding (Figure 3B). From these five proteins, three are predicted to include a transmembrane domain (NRG1-VII α2a, α2b and β2a) and undergo canonical processing through metalloproteases in order for the EGF-containing peptide to be released, while the other two lack this domain (Figure 3C). We observed that these five protein coding transcripts can be found in all our studied myeloid samples, except for monocytes in which we saw a different transcriptional profile and isoforms NRG1-VII α2a and β2a were not detected. The predicted NRG1-VII protein isoforms all share the absence of a specific N-terminal domain, but they all present domains that had been previously observed in other NRG1 isoforms. Exceptionally, the presence of both and α and β domains in NRG1-VII αβ3, causes a frame shift in the β exon, hence, changing the peptide sequence of this isoform to a unique C-terminal end. This feature might affect the binding dynamics of this isoforms to ERBB receptors, making it of especial interest.



B

NRG1 I-β2a [NP_039251.2] (637 Aa)

```
  0 MSERKEGRGK GKGKKKERGS GKKPESAAGS QSPALPPRLK EMKSQESAAG SKLVLRCETS SEYSSLRFKW FKNGNELNRK NKPQNIKIQK KPGKSELRIN
100 KASLADSGEY MCKVISKLGN DSASANITIV ESNEIITGMP ASTEGAYVSS ESPIRISVST EGANTSSSTS TSTTGTSHLV KCAEKEKTFC VNGGECFMVK
200 DLSNPSRYLC KCPNEFTGDR CQNYVMASF/Y KAEELYQKRV LTITGICIAL/ LVVGIMCVVA YCKTKKQRKK LHDRLRQSLR SERNNMMNIA NGPHHPNPPP
300 ENVQLVNQYV SKNVISSEHI VEREAETSFS TSHYTSTAHH STTVTQTPSH SWSNGHTESI LSESHSVIVM SSVENSRHSS PTGGPRGRLN GTGGPRECNS
400 FLRHARETPD SYRDSPHSER YVSAMTTPAR MSPVDFHTPS SPKSPPSEMS PPVSSMTVSM PSMAVSPFME EERPLLLVTP PRLREKKFDH HPQQFSSFHH
500 NPAHDSNSLP ASPLRIVEDE EYETTQEYEP AQEPVKKLAN SRRAKRTKPN GHIANRLEVD SNTSSQSSNS ESETEDERVG EDTPFLGIQN PLAASLEATP
600 AFRLADSRTN PAGRFSTQEE IQARLSSVIA NQDPIAV
```

NRG1 VII-α2a (504 Aa)

```
  0 MPASTEGAYV SSESPIRISV STEGANTSSS TSTSTTGTSH LVKCAEKEK KTFCVNGGEC FMVKDLSNPS RYLCKCQPGF TGARCTENVP MKVQN/QEKAE
100 ELYQKRVLTI TGICIAL/LVV GIMCVVAYCK TKKQRKKLHD RLRQSLRSER NNMMNIANGP HHPNPPPENV QLVNQYVSKN VISSEHIVER EAETSFSTSH
200 YTSTAHHSTT VTQTPSHSWS NGHTESILSE SHSVIVMSSV ENSRHSSPTG GPRGRLNGTG GPRECNSFLR HARETPDSYR DSPHSERYVS AMTTPARMSP
300 VDFHTPSSPK SPPSEMSPPV SSMTVSMPSM AVSPFMEEER PLLLVTPPRL REKKFDHHPQ QFSSFHHNPA HDSNSLPASP LRIVEDEEYE TTQEYEPAQE
400 PVKKLANSRR AKRTKPNGHI ANRLEVDSNT SSQSSNSESE TEDERVGEDT PFLGIQNPLA ASLEATPAFR LADSRTNPAG RFSTQEEIQA RLSSVIANQD
500 PIAV
```

NRG1 VII-α2b (326 Aa)

```
  0 MPASTEGAYV SSESPIRISV STEGANTSSS TSTSTTGTSH LVKCAEKEK KTFCVNGGEC FMVKDLSNPS RYLCKCQPGF TGARCTENVP MKVQN/QEKAE
100 ELYQKRVLTI TGICIAL/LVV GIMCVVAYCK TKKQRKKLHD RLRQSLRSER NNMMNIANGP HHPNPPPENV QLVNQYVSKN VISSEHIVER EAETSFSTSH
200 YTSTAHHSTT VTQTPSHSWS NGHTESILSE SHSVIVMSSV ENSRHSSPTG GPRGRLNGTG GPRECNSFLR HARETPDSYR DSPHSERHNL IAELRRNKAH
300 RSKCMQIQLS ATHLRSSSIP HLGFIL
```

NRG1 VII-β2a (501 Aa)

```
  0 MPASTEGAYV SSESPIRISV STEGANTSSS TSTSTTGTSH LVKCAEKEK KTFCVNGGEC FMVKDLSNPS RYLCKCPNEF TGDRCQNYVM ASF/YKAEELY
100 QKRVLTITGI CIAL/LVVGIM CVVAYCKTKK QRKKLHDRLR QSLRSERNNM MNIANGPHHP NPPPENVQLV NQYVSKNVIS SEHIVEREAE TSFSTSHYTS
200 TAHHSTTVTQ TPSHSWSNGH TESILSESHS VIVMSSVENS RHSSPTGGPR GRLNGTGGPR ECNSFLRHAR ETPDSYRDSP HSERYVSAMT TPARMSPVDF
300 HTPSSPKSPP SEMSPPVSSM TVSMPSMAVS PFMEEERPLL LVTPPRLREK KFDHHPQQFS SFHHNPAHDS NSLPASPLRI VEDEETETTQ EYEPAQEPVK
400 KLANSRRAKR TKPNGHIANR LEVDSNTSSQ SSNSESETED ERVGEDTPFL GIQNPLAASL EATPAFRLAD SRTNPAGRFS TQEEIQARLS SVIANQDPIA
500 V
```

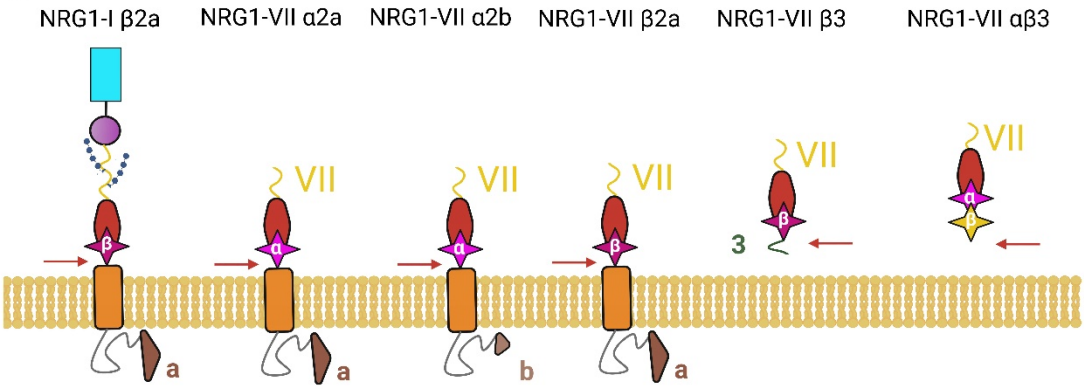
NRG1 VII-β3 (105 Aa)

```
  0 MPASTEGAYV SSESPIRISV STEGANTSSS TSTSTTGTSH LVKCAEKEK KTFCVNGGEC FMVKDLSNPS RYLCKCPNEF TGDRCQNYVM ASF/YSTSTPF
100 LSLPE
```

NRG1 VII-αβ3 (111 Aa)

```
  0 MPASTEGAYV SSESPIRISV STEGANTSSS TSTSTTGTSH LVKCAEKEK KTFCVNGGEC FMVKDLSNPS RYLCKCQPGF TGARCTENVP MKVQN/QESAQ
100 MSLLVIAAKT T
```

C



234

235 **Figure 3. Sequence and structure of NRG1-VII isoforms** A) Representation of the exon structure for nine
236 type VII isoforms sequenced using Oxford Nanopore long-read sequencing. First column (colored dots)
237 denotes presence or absence of each transcript in the different samples used in this study (opaque=
238 present, transparent= absent). Arrows underneath the reference template show target regions for the single
239 forward primer and both reverse primers used in the amplicon sequencing (See also Table 1). B) Predicted
240 translation of protein sequences of the newly characterized NRG1 type VII isoforms (β3, α2a), compared
241 to canonical NRG1 I-β2a (top sequence). Text colour indicates known protein motifs: purple Ig; red EGF;
242 orange transmembrane domains. "/" represents predicted propeptide cleavage points (green text
243 represents ADAM/BACE1 proteolysis, blue represents γ-secretase proteolysis). Dashed lines in all panels
244 represent sequences that are not shown. C) Schematic of predicted translated proteins for NRG1-VII
245 isoforms compared to canonical NRG1-I (Left), borrowing from domain schema shown in Figure 1. Alternate
246 (α or β) EGF domains annotated in pink. Intracellular triangles represent alternate a or b cytoplasmic tails.
247 The yellow β domain represents the peptide sequence that arises from the related exon but is translated to
248 a different peptide sequence due to a frame shift.
249

250 Macrophages are a major source of NRG1 in human tissue.

251 We next assessed the distribution patterns of *NRG1* in single-cell RNAseq experiments in the Human
252 Protein Atlas (Karlsson et al., 2021), which revealed that the *NRG1* locus was actively expressed by many
253 different cell types, such as neurons in brain and eye; stromal cells in colon; endothelial cells in heart and
254 liver; and epithelial cells in kidney and lung. In most of the tissues that we examined, *NRG1* expression
255 was observed in macrophages (Figure 4A). However, no isoform-specific data is available to compare
256 differential isoform expression between these tissues.
257

258 NRG1-VII expression in myeloid cells is affected by differentiation and maturation.

259 We sought to confirm if isoform VII is the only or even primary TSS used by myeloid cells. Therefore, using
260 primers that could discriminate between each unique start exon (Table 1), we investigated the patterns of
261 the seven classes of NRG1 isoforms in different *in vivo* and *in vitro* myeloid cells (Figure 4B). Expression
262 of *NRG1-VII* was detected in all myeloid cells, and it was the isoform class showing highest expression for
263 cells that belong to this lineage (except for macrophages derived from a 2D epithelium). In contrast, control
264 cell types (hiPSCs and iPSC-derived neurons) showed high expression of other *NRG1* classes, but not
265 *NRG1-VII*.
266

267 Having demonstrated that our iPSC derived myeloid cells can be used to model use of this TSS, we next
268 asked at what stage during the differentiation process could we first detect the expression of NRG1-VII
269 transcripts. Thus, we investigated the NRG1-VII expression patterns during the differentiation process
270 towards the myeloid lineage, and its concurrence with specific myeloid markers (Figure 4C). Our results
271 show that NRG1-VII transcription on *in vitro* derived myeloid cells coincides with the expression of mature
272 monocyte markers like CD16, unlike what is observed in blood monocytes from *in vivo* samples. This means
273 that the sequence of events happening *in vivo* are not recapitulated exactly, and that a more mature myeloid
274 identity is needed in order to activate transcription arising from the *NRG1-VII* TSS.
275

276 NRG1-VII expression in myeloid cells is affected by differentiation and maturation.

To confirm that *NRG1* mRNAs are translated into proteins on myeloid cells, we then performed immunohistochemical stainings of NRG1 on human glioblastoma (GBM) tissue which is enriched in bone marrow-derived macrophages (Klemm et al., 2020). Myeloid cells were identified using a CD68 antibody; antibodies detecting the extracellular (EGF-like domain) and intracellular (ICD) domains of NRG1 were used. NRG1 was detected in both myeloid and non-myeloid cells (Figure 4D-E). Not all CD68+ cells expressed NRG1, and those cells that did, exhibited different patterns of expression (Figure S2). Altogether, these results show that myeloid cells synthesize NRG1 peptides in human tissues but are not the sole contributors to the NRG1 pool. They also reveal diverse NRG1 expression patterns in myeloid cells, suggesting that regulation of NRG1 expression is different among myeloid cells.

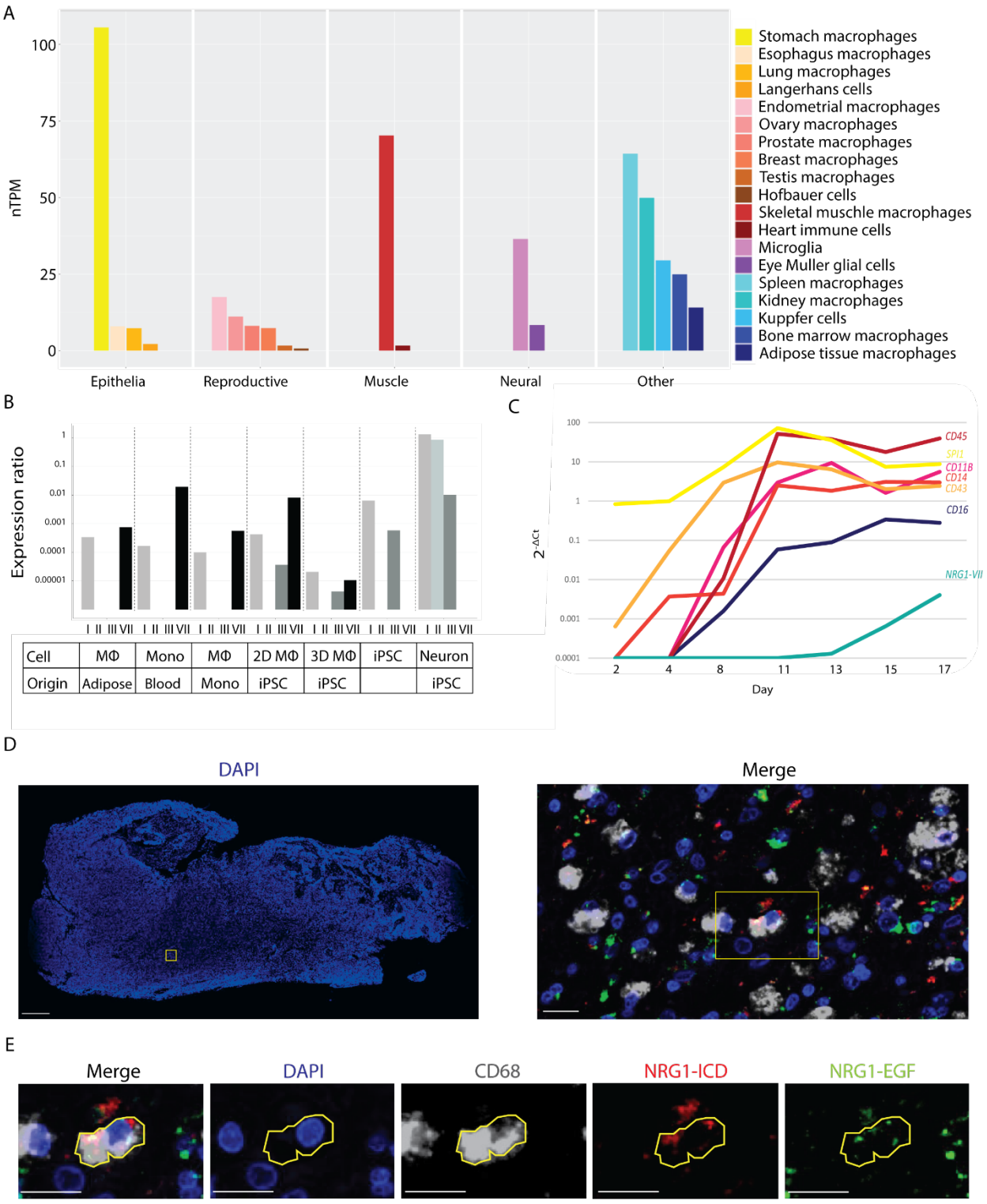


Figure 4. NRG1 expression profile in hematopoietic cells. A) Quantification of NRG1 mRNA expression in different tissue macrophages according to The Human Protein Atlas (Karlsson et al., 2021). nTPM: Number of transcripts per million. B) qRT-PCR data measuring presence of the different NRG1 classes in

292 different cell types. Classes IV-VI were not detected in any samples. Expression ratios were calculated as
293 described in Pfaffl, 2021. Mono: Monocyte, Mφ: macrophage, iPSC: Induced pluripotent stem cell. C) Time
294 series showing qRT-PCR expression of typical myeloid markers in progenitor maturation and relationship
295 with myeloid specific NRG1-VII transcripts. D) Immunostaining of GBM samples showing presence of
296 $CD68^+$ cells that synthesize NRG1 peptides *in vivo*. The left panel shows sample section stained with DAPI.
297 The boxes indicate the regions shown in the right panel in D and the cell in E, respectively. Scale bar on
298 left panel is 1 mm, scale bar on right panel represents 20 µm. E) Example of cell in which all markers (CD68,
299 NRG1-EGF and NRG1-ICD) are expressed. Scale bars = 20 µm. See also Figure S2.
300

301 **Discussion**

302 Macrophages have been reported as a major source of NRG1 in different human tissues (Karlsson et al.,
303 2021), including a recent report highlighting the importance of NRG1 in the developing gut (Jardé et al.,
304 2020). Our own review of *NRG1* expression in the public single cell atlases, including the Human Protein
305 Atlas mRNA dataset, suggested that macrophages are a source of NRG1 in most tissues. However, the
306 specific isoforms involved in the macrophage mediated NRG1 secretion were uncharacterized. Here we
307 report that myeloid cells preferentially use a novel TSS that is myeloid-specific and conserved to generate
308 a previously uncharacterized class of NRG1 isoforms. The tissue variability and functional versatility of
309 NRG1 are a consequence of the alternative promoter usage and exon retention that give rise to a high
310 diversity of isoforms. Description of a new TSS class, NRG1-VII, including at least five new protein-coding
311 isoforms, has expanded the known NRG1 protein coding isoforms from 28 to 33.
312

313 It had been previously reported that the highest NRG1 expression levels in human tissues was in circulating
314 blood (Pipes et al., 2013). This is not surprising as maturation of myeloid cells *in vivo* seems to cause the
315 downregulation of this gene across the monocyte-macrophage differentiation axis (Figure 2A), with a subset
316 of macrophages expressing NRG1 in tissues (Figure 4). Our review of the FANTOM and protein atlas
317 suggests that monocytes are the main source of NRG1 in blood. The function of monocyte-derived NRG1
318 has not been described but circulating levels of NRG1 correlate with liver metabolic activity (Zhang et al.,
319 2018) as well as post-infarct recovery and cardiovascular health improvement (Mendes-Ferreira et al.,
320 2013). We further found that monocytes express at least six NRG1-VII isoforms, including transcripts that
321 contain a novel 'poison' exon (NRG1-VII Pαca and Pαa) predicted to prematurely terminate translation.
322 This exon introduces an early stop codon and hence is likely to drive the transcript to nonsense mediated
323 decay; hence generating a monocyte specific transcriptional mechanism to regulate the levels of NRG1
324 synthesis. Additionally, transcripts containing this exon present unusual splicing on the 3' end, suggesting
325 that there are additional differences and undiscovered mechanisms controlling transcription of these
326 mRNAs. Due to the high levels of expression of this gene in monocytes compared to other cell types, we
327 hypothesize that these represent regulatory mechanisms that regulate NRG1 expression in circulation.
328

NRG1 is an important clinical target in different cancer types, for example, NRG1 fusion proteins cause pathological activation of ERBB receptors (Laskin et al., 2020). NRG1 expression correlates with a shorter survival in patients with glioma (Yin et al., 2015), which could be related to its role in cell migration and proliferation. Glioblastomas particularly exhibit high macrophage infiltration (Wei et al., 2019). Our results here show that tumor-associated myeloid cells ($CD68^+$) may contribute to the NRG1 pool seen in these tumors. Due to the lack of unique domains in type VII isoforms, the antibodies used in this study targeted common NRG1 regions (the EGF-like and intracellular domains). Even though the available antibodies lack specificity to prove that type VII isoforms are translated in this disease, the only isoforms detected in primary monocytes or macrophages in this study belong to classes I and VII.

While NRG1 double positive myeloid cells may show newly synthesized pro-peptides, we also observed $CD68^+$/NRG1-ICD^+ cells indicating that the EGF-like domain had been cleaved. On the other hand, the presence of $CD68^+$/NRG1-EGF^+ cells suggests that some cells restrict their expression only to isoforms that end transcription in linker 3 and lack a transmembrane domain. Therefore, we confirmed that macrophages *in vivo* express NRG1 and that there are distinct populations of macrophages in tissue based on their NRG1 expression patterns.

In order to understand whether *in vitro* derived myeloid cells can be used to model the activity of the NRG1-VII TSS, we used primers designed to target specific TSS usage in different cell types. We confirmed that NRG1-VII was expressed in all myeloid samples and in no others, suggesting the TSS is active. However, iPSC-derived samples showed expression of NRG1-III. This could be a result of incomplete differentiation of the culture, or retention of stem cell features in *in vitro* derived myeloid cells. Moreover, the time series data obtained during the differentiation process shows that NRG1 expression follows mature markers like CD16; *in vivo*, NRG1 expression precedes CD16, and as CD16 is upregulated, NRG1 expression decreases. Hence, while we show activity of the TSS, the *in vitro* model may not recapitulate the transcriptional and maturation sequence seen *in vivo* due to the differences in the differentiation process.

The human NRG1 locus has over 30 curated isoforms, to which this study adds nine new transcripts that are specific to myeloid cells. It is likely that the full transcriptional profile of this locus has not yet been described; for example, while long-read amplicon sequencing is a sensitive isoform recovery method (Clark et al., 2020), it is limited by the primer set(s) used. Thus, only NRG1-VII isoforms utilizing linker 3 or the a tail regions were amplifiable in this study. In addition, transcript NM_004495 presents domains α and β, but is 5' truncated. Hence it is unique but unassigned to a specific NRG1 class as its TSS has not been identified. Our results show an isoform with both exons also present, supporting the existence of splice variants, but with a 5' different to NM_004495, suggesting the presence of isoforms yet to be characterized in this locus.

366 NRG1 is involved in many different functions, this is attributed to the many isoforms which are tissue-
367 restricted during different developmental stages. Only through a thorough investigation of this locus, can
368 we better understand the specifics in each process and how to better develop clinical strategies to prevent
369 or treat the different pathologies associated with this locus. Here we showed that myeloid cells exhibit a
370 unique regulation pattern of the NRG1 locus to generate cell specific isoforms that could play important
371 roles in many diseases. Further detailed investigation on the molecular genetic features and functions of
372 these novel isoforms might uncover how NRG1-VII isoforms elicit differential receptor activity and
373 downstream effects as previously described for other NRG1 isoforms. This ultimately allows for targeted
374 therapies and better understanding of the complexity of an *in vivo* system and might help recreate the
375 processes in which appropriate signals are essential to model the desired biological mechanisms.
376
377 **Materials and methods**
378 Cell lines and ethics approvals.
379 Stem cell work was performed in accordance with The University of Melbourne ethics committee HREC
380 (approval 1851831). The line of human iPSCs used was: PB001.1 (Vlahos et al., 2019), obtained from the
381 Stem Cell Core Facility at the Murdoch Children's Research Institute. Kolf2.1 (hPSCReg accession
382 WTSIi018-B; Wellcome Trust Sanger Institute) cells were used to differentiate cortical neurons (Ethics ID:
383 12374) and 2D macrophages. Monocytes were isolated from buffy coat, which was obtained from the
384 Australian Red Cross Blood Service in accordance with The University of Melbourne ethics committee
385 HREC (approval 1646608). Ethics for adipose tissue derived samples was obtained from the University of
386 Melbourne Human Ethics Committee (ethics ID 1851533) and approved by The Avenue Hospital Human
387 Research Ethics Committee (Ramsay Health; ethics ID WD00006, HREC reference number
388 2019/ETH/0050). For human glioblastoma samples, human ethics approval was covered by project
389 application 1853511, approved by the Medicine and Dentistry Human Ethics Sub-Committee, The
390 University of Melbourne.
391
392 Stem cell culture.
393 Cells were cultured in Gibco™ Essential 8™ media with Essential 8™ supplement (Thermo Fisher
394 Scientific; A1517001) on growth-factor reduced Matrigel® Matrix (Corning®; 356234) coated dishes. Cells
395 were cultured with daily changed fresh media. Cell culture was performed in an APT.line™ C150 (E2) CO_2
396 manual incubator (BINDER; 7001-0172) in constant and stable conditions of humidity, temperature (37°C)
397 and CO_2 concentration in air (5%).
398
399 Cell passaging was performed routinely when cell confluency reached (70-80%). Cells were first washed
400 with Gibco™ PBS (Thermo Fisher Scientific; 10010023). Then, a dilution of sterile 0.5 M EDTA (Thermo
401 Fisher Scientific; 15575020) in said PBS at a concentration of 0.5 mM was used to detach the cells from
402 the plate. After 3-4 minutes in humid incubator conditions, cells were collected and replated with fresh
403 media.

404

405 3D iPSC derived macrophages.

406 As described in Rajab et al., 2021. Human iPSCs were differentiated into macrophages following (Joshi et
407 al., 2019), but with the following alterations to the protocol: harvested cells were cultured in MAGIC media,
408 which was changed as indicated in (Ng et al., 2008). During this process, cells were plated in 10cm Ø non-
409 treated Petri dishes (IWAKI; 1020-100) and placed on an orbital shaker (N-Biotek orbital shaker NB-
410 T101SRC) in a humidified incubator with 5% CO_2 at 37°C.

411

412 After 11 days of culture, cells can be observed detaching from the embryoid bodies, remaining in
413 suspension in the media as non-adherent cells (which are characterized as myeloid progenitors). When all
414 cells are collected and allowed to settle in a 15 mL Falcon tube (Corning®; CLS431470-500EA), embryoid
415 bodies pelleted in the bottom, but progenitors stayed in suspension in the supernatant, and could then be
416 collected. The supernatant was the centrifuged (Heraeus Multifuge 1S-R) at 400 rpm for 5 minutes to
417 separate the progenitors from the media. Progenitors were then resuspended in a 10% dilution in volume
418 of FBS and 100ng/mL CSF-1 (R&D Systems; 216-MC-500) in Gibco™ RPMI-1640 media (Thermo Fisher
419 Scientific; 11875093). Cells were plated in Costar® 6-well tissue-culture treated plates (Corning®; 3516)
420 for 4-7 days in stable incubator conditions (humid, 5% CO_2, 37°C), when cells showed morphological and
421 molecular features displayed by macrophages.

422

423 iPSC derived neurons

424 Kolf2.1 hiPSCs (Kao et al., 2016) were cultured under xenogenic conditions as defined in (Niclis et al.,
425 2017). The cells were then differentiated into cortical neurons following the protocol described by
426 (Gantner et al., 2021).

427

428 Human samples and cell sorting.

429 *Blood monocyte isolation.*

430

431 Buffy Coat was obtained from the Australian Red Cross Blood Service. The blood was diluted with PBS at
432 a 1:3 dilution and underlayed with Histopaque®-1077 (Sigma-Aldrich; Cat. No. 10771-100ml). The
433 underlayed blood samples were centrifuged (TECHCOMP CT1SRT) at 350g for 30 minutes at 24°C with
434 no brake. Peripheral blood mononuclear cells (PBMCs) were isolated from the interphase and washed twice
435 by using MACs buffer (Gibco™ Dulbecco's phosphate-buffered saline (DPBS) (Ca2+Mg2+ free) (Thermo
436 Fisher Scientific; Cat. No. 14190144) with 0.5% heat inactivated Fetal Bovine Serum (FBS) (Thermo Fisher
437 Scientific; Cat. No. 10082147 or 10099141) and 2mM EDTA (Invitrogen™ UltraPure™ 0.5M EDTA (Thermo
438 Fisher Scientific; Cat. No. 15575020)) and centrifuging at 400g for 5 minutes at 4°C. A cell count and
439 viability were determined using 0.4% Gibco™ Trypan Blue (Thermo Fisher Scientific; Cat. No. 15250061)
440 using a hemocytometer. Cells were centrifuged at 400g for 5 minutes at 4°C and resuspended in 40µl MACs

441 buffer per 10^7 cells. Monocytes were positively selected by a magnetic field using Human CD14 MicroBeads
442 (Miltenyi Biotec; Cat. No. 130-050-201) and LS Columns (Miltenyi Biotec; Cat. No. 130-042-401).
443
444 After isolation, monocytes could be differentiated into macrophages by culture the same conditions as iPSC
445 derived progenitors: resuspension in a 10% dilution in volume of FBS and 100ng/mL CSF-1 in Gibco™
446 RPMI-1640 media. Cells were plated in Costar® 6-well tissue-culture treated plates for 4-7 days in stable
447 incubator conditions (humid, 5% CO_2, 37°C), when cells showed morphological and molecular features
448 displayed by macrophages.
449
450 *Adipose tissue samples.*
451 Adipose tissue was obtained and processed as described in (Raajendiran et al., 2019). Myeloid cells were
452 selected by FACS using a BD FACSAria™ III system (BD Biosciences). Cells that were positive for markers
453 CD45 and CD11b were isolated for this study.
454
455 <u>RNA extractions and cDNA synthesis.</u>
456 Total RNA was extracted using the RNeasy® Plus Mini Kit (Qiagen; 74134) according to manufacturer's
457 instructions. After final total RNA elution in RNase-free water, overall RNA quality and concentration were
458 measured using an RNA ScreenTape (Agilent Technologies; 5067-5576) in an Agilent 2200 Tapestation
459 System (Agilent Technologies; G2964-90003).
460
461 cDNA synthesis was then performed using the isolated total RNA and considering the concentration values
462 assigned to each of the samples for the coming steps. cDNA was synthesized using SensiFASTTM cDNA
463 Synthesis Kit (BioLine; BIO-65053) and following all protocol specifications from the vendor. cDNA
464 concentration and quality were checked using a D5000 ScreenTape (Agilent Technologies; 5067-5588) in
465 an Agilent 2200 Tapestation System. Final cDNA samples were then stored at -20°C.
466
467 <u>q-RT PCR.</u>
468 For mRNA quantification, the Applied Biosystems Viia7 ™ real time system was used (Thermo Fisher
469 Scientific; 4453536) using a Fast 96 well plate hardware set up. The reactions in each well of the Micro
470 Amp Fast 96 well reaction plate 0.1 mL (Thermo Fisher Scientific; 4346907) were prepared as indicated for
471 the Fast SYBR™ Green Master Mix (Thermo Fisher Scientific; 4385612). Primers used in the reaction were
472 designed for each transcript class of interest (Table 1). Plates were then sealed using Optical adhesive
473 covers (Thermo Fisher Scientific; 436 0954). For quantification, n=3 technical replicates were used.
474
475 Relative expression ratio was calculated as E(HKG)Average HKG Ct/ E(GOI)Average GOI Ct as described
476 in Pfaffle, 2021. Efficiencies for each primer pair were calculated from serial dilutions of template and

477 ranged between 95 and 98 % for all NRG1 isoforms classes reported in results. Isoforms IV, V, VI were not
478 detected in any samples. B2M was used as House Keeping gene (HKG) for all samples.

Gene	F-Primer	Tm °C	R-primer	Tm °C
B2M	TAGCTGTGCTCGCGCTACT	66.5	TTCAATGTCGGATGGATGAA	60.2
NRG1 I	CAAAGAAGGCAGAGGCAAAG	62.9	AACTGGTTTCACACCGAAGG	63.9
NRG1 II	AACCTCAAGAAGGAGGTCAGC	63.9	AACTGGTTTCACACCGAAGG	63.9
NRG1 III	CCGACACCGAAGAATCGTAT	63.8	ACTCCCCTCCATTCACACAG	64.0
NRG1 IV	GCGACAGAGAGGGAGGA	63.6	AACTGGTTTCACACCGAAGG	63.9
NRG1 V	AATTCTTCTACGGAGTTTTAAGGTACAC	62.6	GCCGATTCCTGGCTTTTCAT	67.2
NRG1 VI	TCTTCAGGAACCACCTAAGCA	63.6	TCTCCTTCTCCGCACATTTT	63.6
NRG1 VII	TTAGTTCCAAGGCAGGCTGT	63.7	TTGCTCCTTCTGTGGATACTGA	63.7
CD11b	AGAACAACATGCCCAGAACC	63.9	GCGGTCCCATATGACAGTCT	63.9
CD14	GCCGCTGTGTAGGAAAGAAG	63.8	ATCGTCCAGCTCACAAGGTT	63.7
CD16	TGAGGTGTCACAGCTGGAAG	64.3	GGTTGACACTGCCAAACCTT	63.9
CD43	AGTGCTGCGTCCTTATCAGC	64.4	CAAACAGGCAGGAGCAAGAG	65.0
CD45	AGAATACTGGCCGTCAATGG	63.8	GCTGAAGGCATTCACTCTCC	63.9
SPL1	CCAGCTCAGATGAGGAGGAG	64.2	CAGGTCCAACAGGAACTGGT	64.0
NRG1-VII Nanopore	TTTCTGTTGGTGCTGATATTGC TTAGTTCCAAGGCAGGCTGT	84	ACTTGCCTGTCGCTCTATCTTCG ATGCAGCAACAAGAAAGCA And ACTTGCCTGTCGCTCTATCTTCTT TCCTGTTTTCTATTTGCAGAAC	79.4 83.9

479
480
481 **Table 1.** Primers designed for qRT-PCR and Long Read Amplicon Sequencing.
482
483 Nanopore Amplicon sequencing.
484
485 RNA was extracted from a pool of monocytes from 3 different blood donors, a pool of myeloid progenitors
486 derived *in vitro*, macrophages derived from monocytes, and macrophages differentiated from *in vitro*
487 derived progenitors. The samples were then prepared using specific primers (Table 1) and purified using
488 AMPure beads. 2 ng of purified cDNA from each sample was then barcoded following the EXP-PBC096
489 protocol from Oxford Nanopore Technologies. Samples were pooled (equimolar) and a sequencing library
490 was prepared as described in the SQK-LSK110 Oxford Nanopore protocol. Samples were then loaded on
491 a Flongle flow cell (FLG-001) and sequencing was performed using a GridION device.
492
493 Sequencing data analysis.
494 Basecalling was performed using Guppy (v6.3.8) using the super-accurate basecalling config file with a
495 Qscore threshold of 10. Passed reads were mapped to the GRCh38 (GCA_000001405) version of the
496 human genome assembly using the Minimap2 (Li, 2018) (v 2.22) using the command minimap2 -ax splice
497 --splice-flank=yes. Files were then converted and sorted into bam files using Samtools v 1.13 (Li et al.,

2009). The sequences were then collapsed into isoforms with StringTie (v2.2.1) using default parameters (Pertea et al., 2015). Raw reads and collapsed isoforms were visualized in IGV (Robinson et al., 2011).

Automated multiplex immunohistochemistry.

Formalin-fixed paraffin-embedded (FFPE) GBM tissue sections were stained using the Bond RX automated stainer (Leica Biosystems). Slides were deparaffinized in xylene followed by exposure to a graded series of ethanol solutions for rehydration. Heat-induced epitope retrieval was performed with either a Citrate pH 6 buffer or Tris Ethylenediaminietetraacetic acid (EDTA) pH 9 buffer. Slides were blocked with 3% hydrogen peroxide (H2O2) to block endogenous peroxidase activity. For multiplexed IHC staining the Opal 6-plex Detection Kit (Akoya Biosciences) was used. Serial multiplexing was performed by repeating the sequence of antigen retrieval, primary antibody, and Opal polymer incubation, followed by Opal fluorophore visualisation for all antibodies as follows. GBM tissue was stained with CD68 (Abcam ab955, 1:100), NRG1 ICD (Abcam ab191139, 1:200) and NRG1 EGF-like domain (ThermoFisher MA4-12896, 1:100). Slides were incubated for 1 hour at room temperature with primary antibodies diluted in 1x Opal blocking/antibody diluent (Akoya biosciences). Slides were subsequently incubated with the Opal Polymer HRP Ms + Rb secondary polymer for 30 minutes prior to incubation with Opal fluorophores (Opal 520, 540, 570, 620, 650 and 690) diluted at 1:150 in 1x Plus Automation Amplification Diluent (Akoya biosciences) for 10 minutes. Slides were counterstained with 10x Spectral DAPI and coverslipped with ProLong Glass Antifade Mountant (Invitrogen). Multispectral images were acquired at 20x and 40x magnification using PhenoImagerTM HT (Akoya Biosciences). inForm 2.4.8 software (Akoya Biosciences) was used for spectral deconvolution. Deconvoluted multispectral images were subsequently fused in HALO (Indica Labs).

Supplementary data

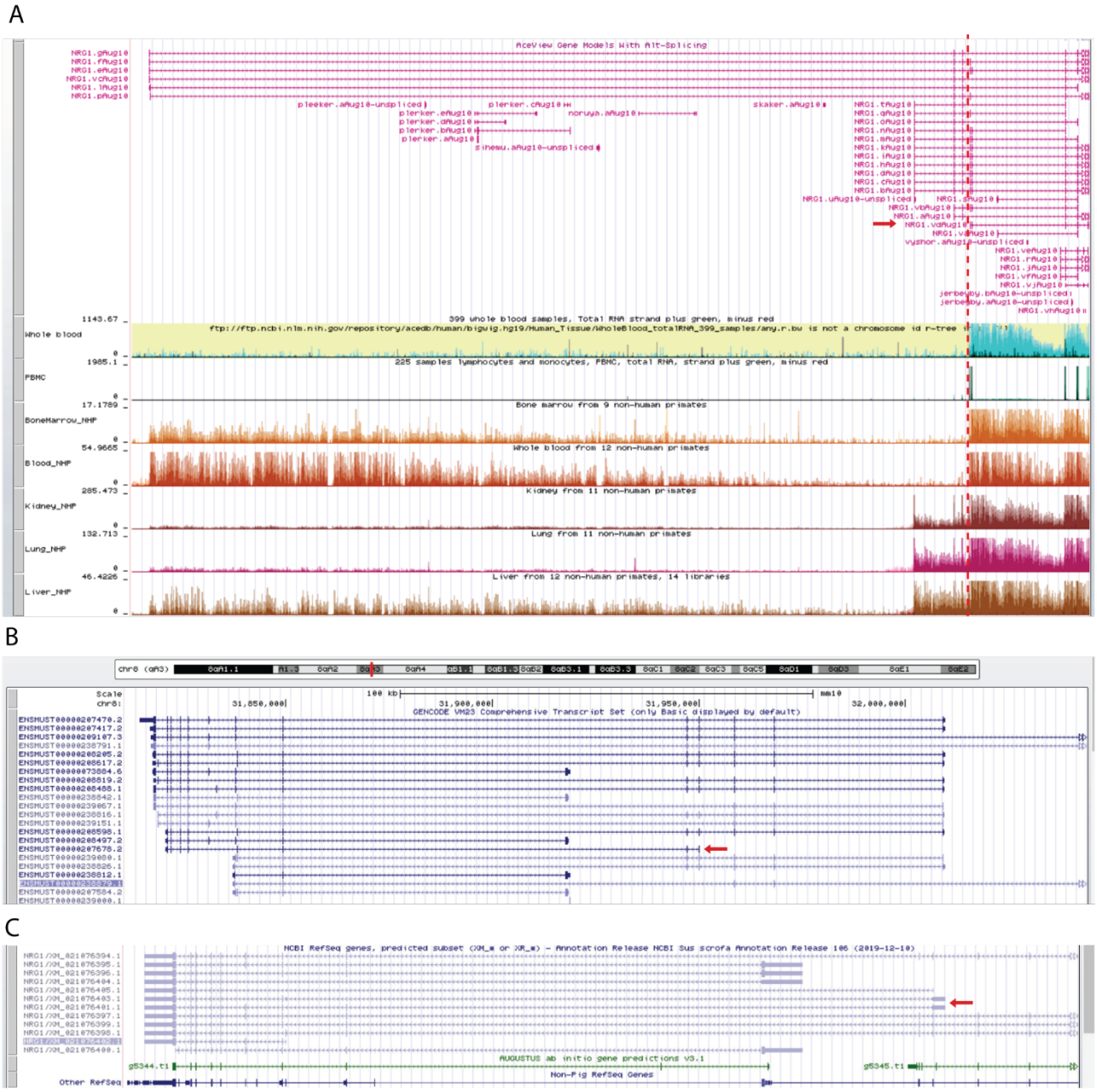


Figure S1. NRG1-VII TSS Shows conservation in primate and non-primate mammals. A) NRG1 expression profiles of human and non-human primates showing that TSS VII is conserved in bone marrow and whole blood B) Curated transcript in *Mus musculus* databases showing conservation of TSS VII in blood cells. C) Predicted transcript in *Sus scrofa* based ETSs derived from dendritic cells and other myeloid progenitor samples. Screenshot from https://www.ncbi.nlm.nih.gov1. *Accessed on 15/12/2022.*

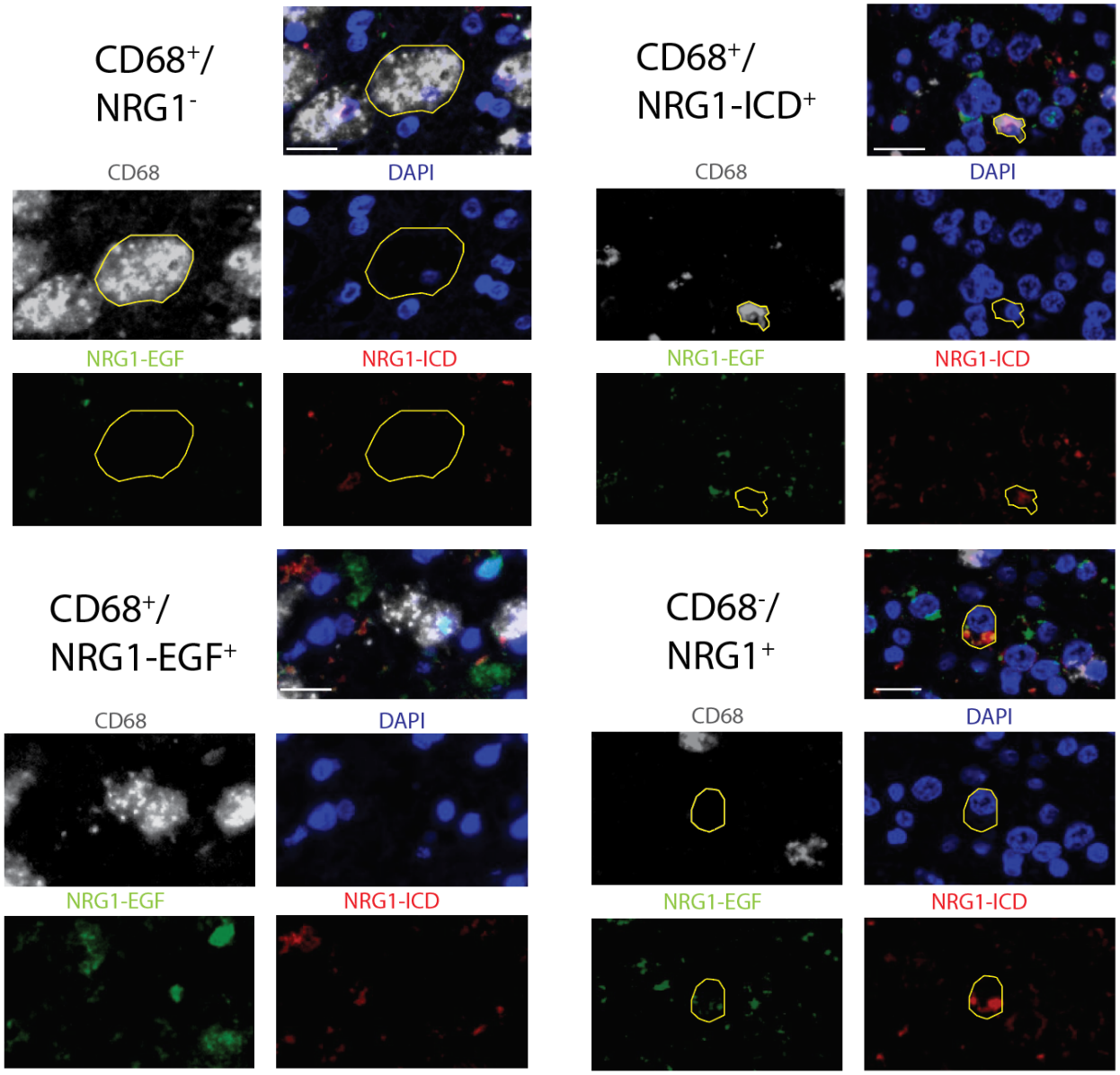


Figure S2. Panel of marker combinations present in GBM samples. CD68$^+$ cells are diverse in their NRG1 expression patterns. Panels show CD68$^+$ cells that can be NRG1$^-$, or positive for either of the NRG1 antibodies independently. CD68$^-$ cells were also observed to be positive for presence NRG1. Scale bars = 20 µm.

REFERENCES:

Alizadeh, A., Santhosh, K.T., Kataria, H., Gounni, A.S. & Karimi-Abdolrezaee, S. 2018, "Neuregulin-1 elicits a regulatory immune response following traumatic spinal cord injury", *Journal of neuroinflammation,* vol. 15, no. 1, pp. 1-21.

Birchmeier, C. 2009, "ErbB receptors and the development of the nervous system", *Experimental cell research,* vol. 315, no. 4, pp. 611-618.

Britsch, S., Li, L., Kirchhoff, S., Theuring, F., Brinkmann, V., Birchmeier, C. & Riethmacher, D. 1998, "The ErbB2 and ErbB3 receptors and their ligand, neuregulin-1, are essential for development of the sympathetic nervous system", *Genes & development,* vol. 12, no. 12, pp. 1825-1836.

Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D. & Maglott, D.R. 2015, "Gene: a gene-centered information resource at NCBI", Nucleic acids research, vol. 43, no. D1, pp. D36-D42.

Buonanno, A. & Fischbach, G.D. 2001, "Neuregulin and ErbB receptor signaling pathways in the nervous system", Current opinion in neurobiology, vol. 11, no. 3, pp. 287-296.

Carraway III, K.L., Weber, J.L., Unger, M.J., Ledesma, J., Yu, N., Gassmann, M. & Lai, C. 1997, "Neuregulin-2, a new ligand of ErbB3/ErbB4-receptor tyrosine kinases", Nature, vol. 387, no. 6632, pp. 512-516.

Cheret, C., Willem, M., Fricker, F.R., Wende, H., Wulf-Goldenberg, A., Tahirovic, S., Nave, K., Saftig, P., Haass, C. & Garratt, A.N. 2013, "Bace1 and Neuregulin-1 cooperate to control formation and maintenance of muscle spindles", *The EMBO journal,* vol. 32, no. 14, pp. 2015-2028.

Clark, M.B., Wrzesinski, T., Garcia, A.B., Hall, N.A., Kleinman, J.E., Hyde, T., Weinberger, D.R., Harrison, P.J., Haerty, W. & Tunbridge, E.M. 2020, "Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene CACNA1C in human brain", *Molecular psychiatry,* vol. 25, no. 1, pp. 37-47.

Craddock, N., O'Donovan, M.C. & Owen, M.J. 2005, "The genetics of schizophrenia and bipolar disorder: dissecting psychosis", *Journal of medical genetics,* vol. 42, no. 3, pp. 193-204.

Falls, D.L. 2003, "Neuregulins: functions, forms, and signaling strategies", *The EGF Receptor Family*, pp. 15-31.

FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. Nature 507, 462–470 (2014).

Frensing, T., Kaltschmidt, C. & Schmitt-John, T. 2008, "Characterization of a neuregulin-1 gene promoter: Positive regulation of type I isoforms by NF-κB", *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* vol. 1779, no. 2, pp. 139-144.

Gantner, C.W., Hunt, C.P., Niclis, J.C., Penna, V., McDougall, S.J., Thompson, L.H. & Parish, C.L. 2021, "FGF-MAPK signaling regulates human deep-layer corticogenesis", *Stem cell reports,* vol. 16, no. 5, pp. 1262-1275.

Garcia-Barcelo, M., Tang, C.S., Ngan, E.S., Lui, V.C., Chen, Y., So, M., Leon, T.Y., Miao, X., Shum, C.K. & Liu, F. 2009, "Genome-wide association study identifies NRG1 as a susceptibility locus for Hirschsprung's disease", *Proceedings of the National Academy of Sciences,* vol. 106, no. 8, pp. 2694-2699.

Garratt, A.N., Britsch, S. & Birchmeier, C. 2000, "Neuregulin, a factor with many functions in the life of a schwann cell", *Bioessays,* vol. 22, no. 11, pp. 987-996.

Garrido-Trigo, A., Corraliza, A., Veny, M., Dotti, I., Melon-Ardanaz, E., Rill, A., Crowell, H.L., Corbi, A., Gudino, V. & Esteller, M. 2022, "Macrophage and neutrophil heterogeneity at single-cell spatial resolution in inflammatory bowel disease", bioRxiv.

Georgieva, L., Dimitrova, A., Ivanov, D., Nikolov, I., Williams, N.M., Grozeva, D., Zaharieva, I., Toncheva, D., Owen, M.J. & Kirov, G. 2008, "Support for neuregulin 1 as a susceptibility gene for bipolar disorder and schizophrenia", *Biological psychiatry,* vol. 64, no. 5, pp. 419-427.

Go, R.C., Perry, R.T., Wiener, H., Bassett, S.S., Blacker, D., Devlin, B. & Sweet, R.A. 2005, "Neuregulin-1 polymorphism in late onset Alzheimer's disease families with psychoses", *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,* vol. 139, no. 1, pp. 28-32.

Gumà, A., Martínez-Redondo, V., López-Soldado, I., Cantó, C. & Zorzano, A. 2010, "Emerging role of neuregulin as a modulator of muscle metabolism", *American Journal of Physiology-Endocrinology and Metabolism,* vol. 298, no. 4, pp. E742-E750.

Haller, P.M., Gonçalves, I.F., Acar, E., Jäger, B., Pilz, P.M., Wojta, J., Huber, K., Kiss, A. & Podesser, B.K. 2022, "Relationship between plasma Neuregulin-1 and cardiac function in patients with ST-elevation myocardial infarction", *Reviews in* Cardiovascular Medicine, vol. 23, no. 2, pp. 63.

Harari, D., Tzahar, E., Romano, J., Shelly, M., Pierce, J., Andrews, G. & Yarden, Y. 1999, "Neuregulin-4: a novel growth factor that acts through the ErbB-4 receptor tyrosine kinase", Oncogene, vol. 18, no. 17, pp. 2681-2689.

Jardé, T., Chan, W.H., Rossello, F.J., Kahlon, T.K., Theocharous, M., Arackal, T.K., Flores, T., Giraud, M., Richards, E. & Chan, E. 2020b, "Mesenchymal niche-derived neuregulin-1 drives intestinal stem cell proliferation and regeneration of damaged epithelium", *Cell Stem Cell,* vol. 27, no. 4, pp. 646-662. e7.

Joshi, K., Elso, C., Motazedian, A., Labonne, T., Schiesser, J.V., Cameron, F., Mannering, S.I., Elefanty, A.G. & Stanley, E.G. 2019, "Induced pluripotent stem cell macrophages present antigen to proinsulin-specific T cell receptors from donor-matched islet-infiltrating T cells in type 1 diabetes", *Diabetologia,* vol. 62, no. 12, pp. 2245-2251.

Jumper, N., Hodgkinson, T., Paus, R. & Bayat, A. 2017, "A Role for Neuregulin-1 in Promoting Keloid Fibroblast Migration via ErbB2-mediated Signaling.", *Acta Dermato-Venereologica,* vol. 97, no. 6.

Kao, T., Labonne, T., Niclis, J.C., Chaurasia, R., Lokmic, Z., Qian, E., Bruveris, F.F., Howden, S.E., Motazedian, A. & Schiesser, J.V. 2016, "GAPTrap: a simple expression system for pluripotent stem cells and their derivatives", *Stem Cell Reports,* vol. 7, no. 3, pp. 518-526.

Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J. & Dusart, P. 2021a, "A single–cell type transcriptomics map of human tissues", *Science Advances,* vol. 7, no. 31, pp. eabh2169.

Kilik, U., Yu, Q., Holtackers, R., Seimiya, M., dos Santos, Aline Xavier da Silveira, Treutlein, B., Spence, J.R. & Camp, J.G. 2021, "Maturation of human intestinal epithelium from pluripotency in vitro", *bioRxiv*.

Klemm, F., Maas, R.R., Bowman, R.L., Kornete, M., Soukup, K., Nassiri, S., Brouland, J., Iacobuzio-Donahue, C.A., Brennan, C. & Tabar, V. 2020, "Interrogation of the microenvironmental landscape in brain tumors reveals disease-specific alterations of immune cells", *Cell,* vol. 181, no. 7, pp. 1643-1660. e17.

Kramer, R., Bucay, N., Kane, D.J., Martin, L.E., Tarpley, J.E. & Theill, L.E. 1996, "Neuregulins with an Ig-like domain are essential for mouse myocardial and neuronal development", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 93, no. 10, pp. 4833-4838.

Laskin, J., Liu, S., Tolba, K., Heining, C., Schlenk, R., Cheema, P., Cadranel, J., Jones, M., Drilon, A. & Cseh, A. 2020, "NRG1 fusion-driven tumors: biology, detection, and the therapeutic role of afatinib and other ErbB-targeting agents", *Annals of Oncology,* vol. 31, no. 12, pp. 1693-1703.

Li, H. 2018, "Minimap2: pairwise alignment for nucleotide sequences", *Bioinformatics,* vol. 34, no. 18, pp. 3094-3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup 2009, "The Sequence Alignment/Map format and SAMtools", *Bioinformatics (Oxford, England),* vol. 25, no. 16, pp. 2078-2079.

Liu, X., Bates, R., Yin, D.M., Shen, C., Wang, F., Su, N., Kirov, S.A., Luo, Y., Wang, J.Z., Xiong, W.C. & Mei, L. 2011, "Specific regulation of NRG1 isoform expression by neuronal activity", *The Journal of neuroscience: the official journal of the Society for Neuroscience,* vol. 31, no. 23, pp. 8491-8501.

Marballi, K., Cruz, D., Thompson, P. & Walss-Bass, C. 2012, "Differential neuregulin 1 cleavage in the prefrontal cortex and hippocampus in schizophrenia and bipolar disorder: preliminary findings", *PloS one,* vol. 7, no. 5, pp. e36431.

Mei, L. & Xiong, W. 2008, "Neuregulin 1 in neural development, synaptic plasticity and schizophrenia", *Nature Reviews Neuroscience,* vol. 9, no. 6, pp. 437-452.

Nave, K. & Salzer, J.L. 2006, "Axonal regulation of myelination by neuregulin 1", *Current opinion in neurobiology,* vol. 16, no. 5, pp. 492-500.

NCBI Gene [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [2022, Aug, 10]. Available from: https://www.ncbi.nlm.nih.gov/gene/3084

Newbern, J. & Birchmeier, C. 2010, "Nrg1/ErbB signaling networks in Schwann cell development and myelination", *Seminars in cell & developmental biology*Elsevier, , pp. 922.

675 Ng, E.S., Davis, R., Stanley, E.G. & Elefanty, A.G. 2008, "A protocol describing the use of a recombinant
676 protein-based, animal product-free medium (APEL) for human embryonic stem cell differentiation as
677 spin embryoid bodies", *Nature protocols,* vol. 3, no. 5, pp. 768-776.
678 Niclis, J.C., Gantner, C.W., Alsanie, W.F., McDougall, S.J., Bye, C.R., Elefanty, A.G., Stanley, E.G.,
679 Haynes, J.M., Pouton, C.W. & Thompson, L.H. 2017, "Efficiently specified ventral midbrain
680 dopamine neurons from human pluripotent stem cells under xeno-free conditions restore motor
681 deficits in parkinsonian rodents", *Stem cells translational medicine,* vol. 6, no. 3, pp. 937-948.
682 Pentassuglia, L. & Sawyer, D.B. 2009, "The role of Neuregulin-1β/ErbB signaling in the heart",
683 *Experimental cell research,* vol. 315, no. 4, pp. 627-637.
684 Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic
685 acids research, 29(9), pp.e45-e45.
686 Pinkas-Kramarski, R., Shelly, M., Guarino, B.C., Wang, L.M., Lyass, L., Alroy, I., Alamandi, M., Kuo, A.,
687 Moyer, J.D. & Lavi, S. 1998, "ErbB tyrosine kinases and the two neuregulin families constitute a
688 ligand-receptor network", Molecular and cellular biology, vol. 18, no. 10, pp. 6090-6101.
689 Pipes, L., Li, S., Bozinoski, M., Palermo, R., Peng, X., Blood, P., Kelly, S., Weiss, J.M., Thierry-Mieg, J. &
690 Thierry-Mieg, D. 2013, "The non-human primate reference transcriptome resource (NHPRTR) for
691 comparative functional genomics", *Nucleic acids research,* vol. 41, no. D1, pp. D906-D914.
692 Raajendiran, A., Ooi, G., Bayliss, J., O'Brien, P.E., Schittenhelm, R.B., Clark, A.K., Taylor, R.A.,
693 Rodeheffer, M.S., Burton, P.R. & Watt, M.J. 2019, "Identification of metabolically distinct adipocyte
694 progenitor cells in human adipose tissues", *Cell reports,* vol. 27, no. 5, pp. 1528-1540. e7.
695 Rajab, N., Angel, P.W., Deng, Y., Gu, J., Jameson, V., Kurowska-Stolarska, M., Milling, S., Pacheco,
696 C.M., Rutar, M. & Laslett, A.L. 2021, "An integrated analysis of human myeloid cells identifies gaps
697 in in vitro models of in vivo biology", Stem cell reports, vol. 16, no. 6, pp. 1629-1643.
698 Rajab, N., Gearing, L.J., Prawer, Y.D., Angel, P.W., Grimmond, S.M., Laslett, A.L., Choi, J. & Wells, C.A.
699 2022, "Proliferation is a driver of quorum sensing in an in vitro model of tissue resident
700 macrophages", *bioRxiv*.
701 Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. & Mesirov, J.P.
702 2011, "Integrative genomics viewer", *Nature biotechnology,* vol. 29, no. 1, pp. 24-26.
703 Ryzhov, S., Matafonov, A., Galindo, C.L., Zhang, Q., Tran, T., Lenihan, D.J., Lenneman, C.G., Feoktistov,
704 I. & Sawyer, D.B. 2017, "ERBB signaling attenuates proinflammatory activation of nonclassical
705 monocytes", *American Journal of Physiology-Heart and Circulatory Physiology,* vol. 312, no. 5, pp.
706 H907-H918.
707 Samsa, L.A., Ito, C.E., Brown, D.R., Qian, L. & Liu, J. 2016, "IgG-containing isoforms of Neuregulin-1 are
708 dispensable for cardiac trabeculation in zebrafish", *PLoS One,* vol. 11, no. 11, pp. e0166734.
709 Sheng, Q., Liu, X., Fleming, E., Yuan, K., Piao, H., Chen, J., Moustafa, Z., Thomas, R.K., Greulich, H. &
710 Schinzel, A. 2010, "An activated ErbB3/NRG1 autocrine loop supports in vivo proliferation in ovarian
711 cancer cells", *Cancer cell,* vol. 17, no. 3, pp. 298-310.
712 Sun, L., Cheng, B., Zhou, Y., Fan, Y., Li, W., Qiu, Q., Fang, Y., Xiao, S., Zheng, H. & Li, X. 2020, "ErbB4
713 mutation that decreased NRG1-ErbB4 signaling involved in the pathogenesis of amyotrophic lateral
714 sclerosis/frontotemporal dementia", *Journal of* Alzheimer's Disease, vol. 74, no. 2, pp. 535-544.
715 Sweeney, C., Lai, C., Riese, D.J.,2nd, Diamonti, A.J., Cantley, L.C. & Carraway, K.L.,3rd 2000, "Ligand
716 discrimination in signaling through an ErbB4 receptor homodimer", The Journal of biological
717 chemistry, vol. 275, no. 26, pp. 19803-19807.
718 Tang, C.S., Ngan, E.S., Tang, W., So, M., Cheng, G., Miao, X., Leon, T.Y., Leung, B.M., Hui, K.W. & Lui,
719 V.H. 2012, "Mutations in the NRG1 gene are associated with Hirschsprung disease", *Human*
720 *genetics,* vol. 131, no. 1, pp. 67-76.
721 Vlahos, K., Sourris, K., Mayberry, R., McDonald, P., Bruveris, F.F., Schiesser, J.V., Bozaoglu, K.,
722 Lockhart, P.J., Stanley, E.G. & Elefanty, A.G. 2019, "Generation of iPSC lines from peripheral blood
723 mononuclear cells from 5 healthy adults", *Stem Cell Research,* vol. 34, pp. 101380.
724 Wagner, M. & Siddiqui, M. 2007, "Signal transduction in early heart development (II): ventricular chamber
725 specification, trabeculation, and heart valve formation", *Experimental biology and medicine,* vol. 232,
726 no. 7, pp. 866-880.
727 Wen, D., Suggs, S.V., Karunagaran, D., Liu, N., Cupples, R.L., Luo, Y., Janssen, A.M., Ben-Baruch, N.,
728 Trollinger, D.B. & Jacobsen, V.L. 1994, "Structural and functional aspects of the multiplicity of Neu
729 differentiation factors", *Molecular and cellular biology,* vol. 14, no. 3, pp. 1909-1919.

Wolpowitz, D., Mason, T.B., Dietrich, P., Mendelsohn, M., Talmage, D.A. & Role, L.W. 2000, "Cysteine-rich domain isoforms of the neuregulin-1 gene are required for maintenance of peripheral synapses", *Neuron,* vol. 25, no. 1, pp. 79-91.

Yu, Q., Kilik, U., Holloway, E.M., Tsai, Y., Harmel, C., Wu, A., Wu, J.H., Czerwinski, M., Childs, C.J. & He, Z. 2021, "Charting human development using a multi-endodermal organ atlas and organoid models", *Cell,* vol. 184, no. 12, pp. 3281-3298. e22.

Zhang, C., Mei, W. & Zeng, C. 2022, "Oncogenic Neuregulin 1 gene (NRG1) fusions in cancer: A potential new therapeutic opportunities", Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, pp. 188707.

Zhang, D., Sliwkowski, M.X., Mark, M., Frantz, G., Akita, R., Sun, Y., Hillan, K., Crowley, C., Brush, J. & Godowski, P.J. 1997, "Neuregulin-3 (NRG3): a novel neural tissue-enriched protein that binds and activates ErbB4", Proceedings of the National Academy of Sciences, vol. 94, no. 18, pp. 9562-9567.

Zhang, P., Kuang, H., He, Y., Idiga, S.O., Li, S., Chen, Z., Yang, Z., Cai, X., Zhang, K., Potthoff, M.J., Xu, Y. & Lin, J.D. 2018, "NRG1-Fc improves metabolic health via dual hepatic and central action", *JCI insight,* vol. 3, no. 5, pp. 10.1172/jci.insight.98522.

Authorship contributions

Conceptualization MABR, CAW, Data curation MABR, RDP, YDJP, Formal analysis MABR, RDP, YDJP, Funding acquisition CAW, MBC, Investigation MABR, CAW, Methodology MABR, RDP, YDJP, MD, JCH, SSW, WD, NR, AL, JH, Project administration CAW, Resources CAW, MBC, TM, Software MABR, RDP, YDJP, Supervision CAW, MBC, Validation MABR, RDP, YDJP, MD, SSW, Visualization MABR, MD, SSW, writing original draft MABR, Writing – review and editing MABR, RDP, YDJP, MD, SSW, WD, NR, AL, TM, MBC, CAW.

Funding sources

MABR is funded by a University of Melbourne Research Scholarship. MD is supported by the Australian Government Research Training Program Scholarship. MBC is supported by an NHMRC Investigator Fellowship (APP11968410). The project was supported by NHMRC Synergy grant APP 1186371 to CAW.

Conflicts of Interest

YDJP, RDP and MBC have received support from Oxford Nanopore Technologies (ONT) to present their findings at scientific conferences. However, ONT played no role in study design, execution, analysis or publication. No conflicts are declared for MABR, MD, SSW, NR, WDN, TM or CAW.

Acknowledgements.

The authors thank Professor Matt Watt (Department of Anatomy and Physiology, University of Melbourne), for gifting adipose tissue samples. The authors thanks Dr Le Christy Ying and Prof Paul Hertzog (Hudson Institute, Clayton, Melbourne) for gifting 2D macrophage RNA for this study. The authors thank Mrs Zahra Elahi (Department of Anatomy and Physiology, University of Melbourne) for her logistic and emotional support, crucial to this study. The authors thank Mr Yidi Deng (Department of Maths and Statistics, University of Melbourne) for discussions about statistical approaches.

Sequence Accessions: All mRNA sequences have been deposited at NCBI GenBank nucleotide database, Accessions assigned 18/1/23: OQ272754, OQ272755, OQ272756, OQ272757, OQ272758, OQ272759, OQ272760, OQ272761, OQ272762, OQ272763, OQ272764, OQ272765, OQ272766, OQ272767, OQ272768, OQ272769, OQ272770, OQ272771, OQ272772, OQ272773, OQ272774, OQ272775, OQ272776.