

Legends to the Supplement

Supplementary Figure S1. Deduced primary structures of *Defcr* and *Defcr-rs* gene products in the NIH and Celera mouse genome assemblies. The products of α -defensin genes identified in the NIH C57BL/6J mouse (panel A) and the mixed-strain Celera mouse genome assemblies (panel B) were deduced by translation of their coding sequences and are noted as *DefcrX^N* and *DefcrX^C*, respectively, where X = 1-27. The products of CRS1C and CRS4C genes from both assemblies, denoted by superscripted Ns and Cs as in panels A & B, are shown in panel C. The sequences are shown aligned with stop codons denoted with bold asterisks, frameshifted sequences underlined, and gap positions hyphenated. Provisional gene symbols are based on the identity of the deduced mature peptides encoded by the 3' exons with previously described mouse Paneth cell α -defensins, and because certain genes exist as multiple copies coding for the same α -defensin, e.g., Crp5, their proposed gene symbols incorporate peptide identity with lower case letters to distinguish them at the locus. Deviations from these sequences are reflected in the gene name with an initial v for variant form. Gene sequences with disrupted coding sequences are labeled -ps to identify them as pseudogenes.

Supplementary Figure S2. Organization of the rat α -defensin gene locus. The organization of the α -defensin gene locus within the rat genome on chromosome 16q12.5 is shown schematically (See Fig 4). Since the NIH reference assembly and the Celera alternate assembly for the rat genome involves the same strain, Sprague-Dawley, only one assembly for the rat genome is depicted. Consistent with the genome of the mouse, the rat genome is displayed with the telomere, denoted Tel, toward the left of the chromosomal region and the centromere, denoted Cen, toward the right. Again, white line segments represent regions of the chromosome that are currently mapped and annotated, while those regions that have yet to be described are represented by white line segments. Colored arrows indicate individual genes and their transcriptional orientations as follows: β -defensin genes (green), enteric α -defensin genes (red), and myeloid α -defensin genes (black).

Supplementary Figure S1
NIH C57BL/6

A

| Gene Symbol | Deduced Primary Structure |
|-----------------------------|---|
| <i>Defcr1N</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPG EEDQAVSVSFGDPEG TSLQEESLRDLVCYCRSRGCKGRERMNGTCRKGHLLYTLCCR |
| <i>vDefcr2N, 18N</i> | MKTLVLLSALALLAFQVQADPIQNRDEESKIDEQPGKEDQAVSVSFGDPEGSSLQEESLRDLRICYCRTS-CKKRERLNGTCRKGHLMYKLCR |
| <i>vDefcr2-psN*</i> | MKTLVLLSALALLASQVQADPIQNRDEESKIDEQPGKEDQAVSVSFGDPEGSSLQEESLRDLICYCRTRGCKRRERLNGTCRKGHLMYTLCCR |
| <i>vDefcr2N, 16N-ps</i> | SLRDLICYCRTRGCKGRERMNGTCRKGHLMYTLCCR |
| <i>vDefcr2N, 7N, 18N-ps</i> | MKTLVLLSALALLAFQVQADPIQNRDEESKIDEQPGKEDQAVSVSFGDPEGSSLQEESLRDLRICYCRTS-CKKKERLIGTCRKGHLMYKLCR |
| <i>Defcr3N</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPG EDDQAVSVSFGDPEGSSLQEESLRDLVCYCRKRKGRCKRRERMNGTCRKGHLMYTLCCR |
| <i>Defcr5aN</i> | MKTFVLLSALVLLAFQVQADPIHKTD EETNTEEQPG EEDQAVSISFGGQEGSALHDEL SKKLI CYCRIRGCKRRERVFGT CRNLFLT FVFCCS |
| <i>Defcr5bN</i> | MKTFVLLSALVLLAFQAQADPIHKTD EETNTEEQPG EEDQAVSISFGGQEGSALHEEL SKKLI CYCRIRGCKRRERVFGT CRNLFLT FVFCCS |
| <i>Defcr5cN</i> | MKTFVLLSALVLLAYQVQADPIHKTD EETNTEEQPG EEDQAVSISFGGQEGSALHEEL SKKLI CYCRIRGCKRRERVFGT CRNLFLT FVFCCS |
| <i>Defcr5dN</i> | MKTI VLLSALVLLAFQVQADPIQKT D EETNTEEQPG EEDQAVSISFGGQEGSALHEEL SKKLI CYCRIRGCKRRERVFGT CRNLFLT FVFCCS |
| <i>vDefcr5N</i> | MKTFVLLSALVLLAFQAQADPIHKTD EETNTEEQPG EEDQAVSISGGQEGSALHEEL SKKLI CYCRIRGCKRRERVFGT CRNLFLT FVFCCS |
| <i>Defcr17N</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPG EDDQAVSVSFGDPEGSSLQEESLRDLVCYCRKRKGRCKRRERMNGTCRKGHLMYTLCCR |
| <i>vDefcr18a-psN</i> | MKTLVLLSALALLAFQVQADPIQNTDEETNTEEQPGKEDQAVSVSFGDPEGSSLQEESLRDLI*YCRTRGCKRRERLNGTYRKGHLMYMLCC |
| <i>vDefcr18b-psN</i> | MKTLVLLSALALLAFQVQADPIQNTDEETNTEEQPGKEDQAVSVSFGDPEGSSLQEESLRDLI*YCRTRGCKRRERLNGTYRKGHLMYMLCC |
| <i>Defcr20aN</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETNTEEQPG EEDQAVSVSFGDPEGSSALHEKSRDLICYCRKGGCNRGEQVYGTCSGRLLFCRRRRRH |
| <i>Defcr20bN</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETNTEEQPG EEDQAVSVSFGDPEGSSALHEKSRDLICYCRKGGCNRGEQVYGTCSGRLLFCRRRRRH |
| <i>Defcr21N</i> | MKTLVLLSALILLAYQVQADPIQNTDEETNTEEQPG EDDQAVSVSFGGQEGSALHEKLSRDLICLNRNRCNRGELFYGT CAGPFLRCCRRRR |
| <i>Defcr22N</i> | MKTLVLLSALILLAYQVQADPIQNTDEETNTEEQPG EEDQAVSVSFGGQEGSALHEKLSRDLICLNRNRCNRGELFYGT CAGPFLRCCRRRR |
| <i>Defcr23aN</i> | MKTLVLLSALILLAFQVQADPIQNTDEETKTEEQPGKEDQAVSVSFGDPEGSSLQEESLRDLVCYCRTRGCKRRERMNGTCRKGHLLYTLCCR |
| <i>Defcr23bN</i> | MKTLVLLSALILLAFQVQADPIQNTDEETKTEEQPGKEDQAVSVSFGDPEGSSLQEESLRDLVCYCRTRGCKRRERMNGTCRKGHLLYTLCCR |
| <i>Defcr24N</i> | MKTLILL S ALVLLAFQVQADPIQNTDEETKTEEQPG EEDQAVSVSFGDPEGASLQEESLRDLVCYCRARGCKGRERMNGTCSKGHLLYMLCCR |
| <i>vDefcr24N</i> | MKTLILL S ALVLLAFQVQADPIQNTDEETKTEEQPG EEDQAVSVSFGDPEGSSLQEESLRDLVCYCRARGCKGRERMNGTCSKGHLMYMLCCR |
| <i>Defcr25N</i> | MKTLVLLSALALLAFQVQADPIQNRDEESKIDEQPGKEDQAVSVSFGDPEGSSLQEESLRDLICYCRTRGCKRRERLNGTCRKGHLMYMLWCC |
| <i>Defcr26aN**</i> | MKTLVLLSALFLLAFQVQADPIQKT D EETNTEVQPE EEEQAVSVSFGNPEGSDLQEESLRDLGCYCRKRKGRCTRRERLNGTCRKGHLMYTLCC |
| <i>Defcr26b**</i> | MKTLVLLSALFLLAFQVQADPIQNTDEETNTEVQPE EEDQAVSVSFGNPEGSDLQEESLRDLGCYCRKRKGRCTRRERLNGTCRKGHLMYTLCC |

Supplementary Figure S1 CELERA

B

| Gene Symbol | Deduced Primary Structure |
|-------------------|--|
| <i>Defcr1C</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPG EEDQAVSVSFGDPEGTSLQ EES LRDLCYCRSRGCKGRERMNGTCRKGHLLYTLCCR |
| <i>Defcr2C</i> | MKPLVLLSALVLLSFQVQADPIQNTDEETKTEEQSG EEDQAVSVSFGDREGASLQ EES LRDLCYCRTRGCKRRERMNGTCRKGHLMYTLCCR |
| <i>vDefcr2C*</i> | LRDLICYCRTRGCKGRERMNRTC RKGHLMYTLCCR |
| <i>Defcr4C</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETNTEEQPG EEDQAVSISFGGQEGSALHEKS LRGLLCYCRKGHCGRGERVGTGIRFLYCCPRR |
| <i>Defcr5aC</i> | MKTFVLLSALVLLAFQVQADPIHKTDEETNTEEQPG EEDQAVSISFGGQEGSALHEE LSKKLICYCRIRGCKRRERVFGT CRNLFLTFVFCCS |
| <i>Defcr5bC*</i> | LSDKLLICYCRIRGCKRRERVFGT CRNLFLTFVFCCS |
| <i>Defcr7C</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETNTEEQPG EEDQAVSVSFGDPEGSSLQ EES LRDLCYCRTRGCKRRERMNGTCRKGHLMYTLCCR |
| <i>vDefcr18C</i> | MKILVLLCALALLAFQVQADPIQNRDEESKIDEQPG KEDQAVSVSFGDPEGSSLQ EES LRDLCYCRTRGCKRRERLNGTCRKGHLMYMLCCR |
| <i>Defcr22C</i> | MKTLVLLSALILLAYQVQTDPIQNTDEETNTEEQPG EEDQAVSVSFGGQEGSALHEK LSRDLICLCRKRRCNRGELFYGT CAGPFLRCCRRRR |
| <i>Defcr23C</i> | MKTLVLLSALILLAFQVQADPIQNTDEETKTEEQPG KEDQAVSVSFGDPEGSSLQ EES LRDLCYCRTRGCKRRERMNGTCRKGHLYTLCCR |
| <i>Defcr24C</i> | MKTLILLVLLAFQVQADPIQNTDEETKTEEQPG EEDQAVSVSFGDPEGTSLQ EES LRDLCYCRARGCKGRERMNGTCSKGHLLYMLCCR |
| <i>Defcr26C**</i> | MKTLVLLSALFLLAFQVQADPIQKTDEETNTEVQPR VEDQAVSVSFGNAEGSDLQ EES LRDLCYCRKRGCTRRERINGT CRKHLMYTLCCCL |

C

NIH C57BL/6

Supplementary Figure S1

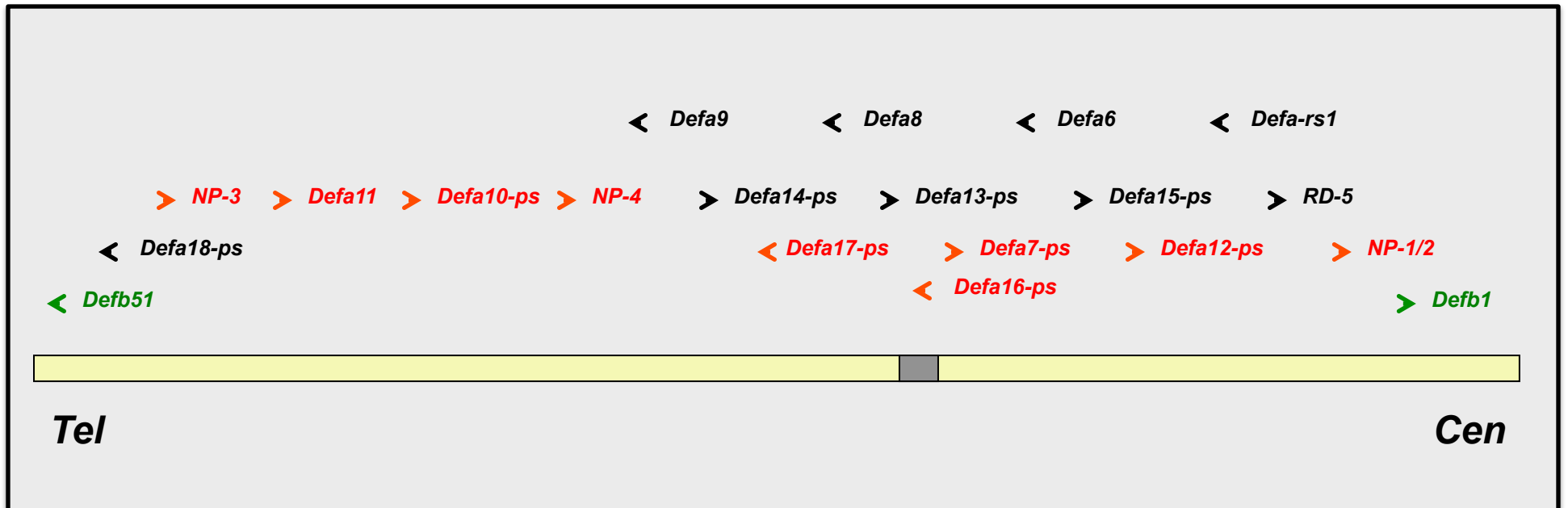
| Gene Symbol | Deduced Primary Structure |
|---------------------|--|
| <i>CRS1CN</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPEEEDQAVSVSFGGTEGSALQDVAQRRFPWCRKCRVCQKCEVCQKCPVCPTCPQCPKQPLCEERQNKTAITTTQAPNTQHKGC |
| <i>CRS1C-1aN</i> | MKTLVLLSALALLALQVQADPIQNTDEETKTQEOPGEEEDQAVSVSFGGTEGSALQDVAQRRFPWCRKCRVCQKCEVCQKCPVCPTCPQCPKQPLCKERQNKTAITTTQAPNTHHKGC |
| <i>CRS1C-1bN</i> | MKTLVLLSALALLALQVQADPIQNTDEETKTQEOPGEEEDQAVSVSFGGTEGSALQDVAQRRFPWCRKCRVCQKCEVCQKCPVCPTCPQCPKQPLCKERQNKTAITTTQAPNTHHKGC |
| <i>CRS1C-2N</i> | MKTLVLLSALALLAFQVQADPIQNTDEETKTEEQPEEEDQAVSVSFGGTEGSALQDVAQRRFLWCRKCPVCQKCEVCQKCPVCPTCPQCPKQPLCEERQNKTAITTTQAPNTQHKGC |
| <i>CRS1C-3N</i> | MKTLVLLSALALLAFQVQADPIKNTDEETKTGEEQPEEEDQAVSVSFGGTEGSALQYVAQRRFPWCRKCPVCQKCEVCQKCPVCPTCPQCPKQPLCKERQNKSAITTTQAPNTQHKGC |
| <i>proCRS1C-psN</i> | MKTLVLLSALALLAFQVQADPIKNTDEETNTEEQPEEEDQAVSVSFGGTEGSALQDV |
| <i>CRS4C-1N</i> | MKKLVLLFALVLLAFQVQADSIQNTDEEIKTEEQPGREENQAVSVSFGDPQGSALQDAALGWGRRCPQCPRCPSCPSPRCPRCPR - - -CKCNP |
| <i>CRS4C-6N</i> | MKTLVLLSALVLLAFYVQADSTQETDEETKTDDQPGEEEDQGVSVSFGDPERYVQLQVSGLGKPPQCPKCPVCSKCPQCPQCPQCPGCPRCNCMTK |

CELERA

| Gene Symbol | Deduced Primary Structure |
|---------------------|--|
| <i>CRS1CC</i> | MKTLVLLSALVLLAFQVQADPIQNTDEETKTEEQPEEEDQAVSVSFGGTEGSALQDVAQRRFPWCRKCRVCQKCEVCQKCPVCPTCPQCPKQPLCEERQNKTAITTTQAPNTQHKGC |
| <i>CRS1C-1C</i> | MKTLVLLSALALLALQVQADPIQNTDEETKTQEOPGEEEDQAVSVSFGGTEGSALQDVAQRRFPWCRKCRVCQKCEVCQKCPVCPTCPQCPKQPLCKERQNKTAITTTQAPNTHHKGC |
| <i>CRS1C-2C</i> | MKTLVLLSALALLAFQVQADPIQNTDEETKTEEQPEEEDQALSVSFGGTEGSALQDVAQRRFLWCRKCPVCQKCEVCQKCPVCPTCPQCPKQPLCEERQNKTAITTTQAPNTQHKGC |
| <i>CRS4C-1C</i> | MKKLVLLFALVLLAFQVQADSIQNTDEEIKTEEQPGREENQAVSVSFGDPQGSALQDAALGWGRRCPQCPRCPSCPSPRCPRCPRCKCNP |
| <i>CRS4C-4C</i> | MKKLVLLSAFVLLAFQVQADSIQNTDEETKTEEQPGREENQAMSVSFGDPEGSALQDAAVGMARPCPPCSPSPCPWCPMCPRCPSCKCNP |
| <i>CRS4C-5C</i> | MKKLVLLFALVLLAFQVQADSIQNTDEETKTEEQPGEEEDQAVSISFGDPEGYALQDAAIRARRCPSPCLSCPWCPRCLRCPICKCNP |
| <i>CRS4C-6C</i> | MKTLVLLSALVLLAFYVQADSTQETDEETKTDDQPGEEEDQGVSVSFGDPERYVQLQVSGLGKPPQCPKCPVCSKCPQCPQCPGCPRCNCMTK |
| <i>vCRS4C-psC</i> | MKTFVLLSALVLLAFQVQADSIQNTDEETKTEEQPGEEEDQIVSVSFGGPEGSALQDAAQRRSLLCEPCPCPC*RCCLT - - -CNCNP |
| <i>proCRS4C-psC</i> | MKKLVLLFALVLLAFQVQADSIQNTDEETKTEEQPGEEEDQAVSVSFGDPQGSALQDA |

Supplementary Figure S2

α -Defensin Gene Locus in Rattus norvegicus Chromosome 16 Sprague-Dawley Genome Assembly



Supplementary Table S1

| Gene Symbol | Peptide | Mouse Strain | Accession Number (mRNA) |
|----------------|---------|---------------------------------|-------------------------|
| <i>Defcr1</i> | Crp1 | Swiss, C3H/HeN, 129/SvJ, BALB/c | NM_010031 |
| <i>Defcr2</i> | Crp2 | Swiss, C3H/HeN, 129/SvJ, BALB/c | NM_007849 |
| <i>Defcr3</i> | Crp3 | Swiss, C3H/HeN, 129/SvJ, BALB/c | NM_007850 |
| <i>Defcr4</i> | Crp4 | Swiss, C3H/HeN, 129/SvJ, BALB/c | NM_010039 |
| <i>Defcr5</i> | Crp5 | Swiss, C3H/HeN, 129/SvJ, BALB/c | NM_007851 |
| <i>Defcr6</i> | Crp6 | Swiss, C3H/HeN, 129/SvJ, BALB/c | NM_007852 |
| <i>Defcr7</i> | Crp7 | Swiss, C3H/HeN, 129/SvJ | U03035 |
| <i>Defcr8</i> | Crp8 | Swiss, C3H/HeN, 129/SvJ | U03036 |
| <i>Defcr9</i> | Crp9 | Swiss, C3H/HeN | U03037 |
| <i>Defcr10</i> | Crp10 | Swiss, C3H/HeN | U03061 |
| <i>Defcr11</i> | Crp11 | Swiss, C3H/HeN | U03062 |
| <i>Defcr12</i> | Crp12 | Swiss, C3H/HeN | U03063 |
| <i>Defcr13</i> | Crp13 | Swiss, C3H/HeN | U03064 |
| <i>Defcr14</i> | Crp14 | Swiss, C3H/HeN | U03067 |
| <i>Defcr15</i> | Crp15 | Swiss, C3H/HeN | U03065 |
| <i>Defcr16</i> | Crp16 | Swiss, C3H/HeN | U03066 |
| <i>Defcr17</i> | Crp17 | Swiss | None |
| <i>Defcr18</i> | Crp18 | Swiss | None |
| <i>Defcr19</i> | Crp19 | Swiss | None |
| <i>Defcr20</i> | Crp20 | C57BL/6L | NM_183268 |
| <i>Defcr21</i> | Crp21 | C57BL/6 | NM_183253 |
| <i>Defcr22</i> | Crp22 | C57BL/6 | NM_207658 |
| <i>Defcr23</i> | Crp23 | C57BL/6 | NM_001012307 |
| <i>Defcr24</i> | Crp24 | C57BL/6 | NM_001024225 |
| <i>Defcr25</i> | Crp25 | C57BL/6 | NM_007849 |
| <i>Defcr26</i> | Crp26 | C57BL/6 | NM_001079933 |
| ND | Crp27 | C57BL/6 | ND |

Supplementary Table S2

a

| Gene Symbol | Peptide | Mouse Strain | Accession Number (mRNA) |
|------------------|-----------|--------------|-------------------------|
| <i>Defcr-rs1</i> | CRS1C,-4 | Swiss, FvB | NM_007844, EU760893 |
| ND | CRS1C-1 | C57BL/6* | XM_001006437 |
| ND | CRS1C-2 | C57BL/6 | AY761183 |
| ND | CRS1C-3,6 | C57BL/6, FVB | AY761184, EU760895 |
| ND | CRS1C-5 | FvB | EU760894 |

b

| Gene Symbol | Peptide | Mouse Strain | Accession Number (mRNA) |
|----------------------|-------------------|------------------------------|-------------------------------|
| <i>Defcr-rs2,rs3</i> | CRS4C-1/d/e/f/h/j | Swiss, C3H/HeN, 129/SvJ, FvB | NM_007847 AJ564863-5,-7,-9 |
| <i>Defcr-rs2,rs3</i> | CRS4C-1g/i | C3H/HeN, 129/SvJ, FvB | AJ564866,-8 |
| <i>Defcr-rs4</i> | CRS4C-1a | C3H/HeN, 129/SvJ, FvB | NM_001005418 |
| <i>Defcr-rs5</i> | CRS4C-1b | 129/SvJ, FvB | MGI:22620 |
| <i>Defcr-rs6</i> | CRS4C-1c | 129/SvJ, FvB | MGI:22621 |
| <i>Defcr-rs7</i> | CRS4C-2/b/c/d | | |
| | CRS4C-3a/b/c/d/e | C3H/HeN, Swiss, FvB | AJ564870-9 |
| <i>Defcr-rs8</i> | CRS4C-2a | 129/SvJ | MGI:22623 |
| <i>Defcr-rs9</i> | CRS4C-3 | 129/SvJ | MGI:22624 |
| ND | CRS4C-3f | FvB | EU760896 |
| <i>Defcr-rs10</i> | CRS4C-4 | 129/SvJ | NM_007845 |
| <i>Defcr-rs11</i> | CRS4C-4a | 129/SvJ | MGI:22626 |
| <i>Defcr-rs12</i> | CRS4C-5 | 129/SvJ | NM_007846 |
| ND | CRS4C-6 | C57BL/6* | NM_001012640 |