Supplementary Data File - Updates to approved nomenclature for oxytocin receptor and arginine vasopressin receptor genes

Gene nomenclature committees and abbreviations:

CGNC: Chicken Gene Nomenclature Committee HGNC: HUGO Gene Nomenclature Committee MGNC: Mouse Genome Nomenclature Committee RGD: Rat Genome Database Xenbase: The Xenopus model organism database VGNC: Vertebrate Gene Nomenclature Committee ZNC: Zebrafish Nomenclature Committee

1. Oxytocin gene nomenclature

The existing approved nomenclature (Table S1) accurately represents both the function of the gene(s), and the orthologous relationships between the genes across species. The current *OXT* symbol is widely used in the literature (1,105 results in PubMed) and highly specific – that is, its use as a search term accurately identifies publications mentioning the oxytocin gene.

The proposed symbol by Theofanopoulou *et al.*¹ (OT), in contrast, is not a specific search term. While it is in use by many papers in the literature to refer to the oxytocin gene, a PubMed search for "OT" returns over 24,000 hits, less than 15% of which refer to oxytocin. Clashes include acronyms for terms such as "operative time", "occupational therapy", and the much studied OT-I/II transgenic mice.

We will retain the existing approved nomenclature for oxytocin as shown in Table S1. OT should be included as an "alias" symbol in databases since it is used in the literature, but we encourage researchers to use the approved gene nomenclature to ensure easy identification of their papers and minimise confusion.

2. Arginine vasopressin gene nomenclature

The existing approved nomenclature (Table S2) accurately represents the function of the gene(s) and the orthologous relationships between the genes across species. The current symbol *AVP* is widely used in the literature (10,495 hits in PubMed) and is highly specific, with around 88% of PubMed hits specifically referring to the gene/gene product.

The proposed symbol by Theofanopoulou *et al.* (VT), in contrast, is not a specific search term and has not been widely used in the literature to refer to this gene. A PubMed search for "VT" returns over 35,000 results, only 244 of which mention "vasotocin" or "vasopressin".

One drawback to the current approved nomenclature that was raised by Theofanopoulou *et al.* is the inclusion of "arginine" in the gene name, which refers to a highly conserved amino acid in the AVP peptide product but which is not present in all vertebrates. We have identified only two lineages in which this arginine residue is absent: suidae and marsupials.

To avoid confusion in the literature, and in biological databases where the existing approved gene nomenclature has already propagated, we will retain the existing symbols but modify the gene names in the species where the arginine residue is not present to "vasopressin" to avoid any confusion. Of the species with approved nomenclature, this currently only affects the pig *Sus scrofa*.

We will retain all other aspects of the existing approved nomenclature for arginine vasopressin as shown in Table S2. VT should be included as an "alias" symbol in databases since it is used in the literature, but we encourage researchers to use the approved gene nomenclature to ensure easy identification of their papers and minimise confusion.

3. Oxytocin receptor gene nomenclature

The existing approved nomenclature (Table S3) accurately represents both the function of the gene(s), and the orthologous relationships between the genes across species. The current *OXTR* symbol is widely used in the literature (928 results in PubMed) and highly specific – that is, its use as a search term accurately identifies publications mentioning the oxytocin receptor gene.

The proposed symbol by Theofanopoulou *et al.* (OTR), in contrast, is not a specific search term. While it is in use by many papers in the literature to refer to the oxytocin receptor gene, it returns over 3,000 hits in PubMed, most of which refer to the abbreviation of a journal name (*Ortop. Traumatol. Rehabil.*). Other results returned for the OTR acronym include "organ transplant recipients", "oxygen transfer rate", "ocular tilt reaction", and "OXPHOS transcriptional response".

We will retain the existing nomenclature system for OXTR genes but make minor updates to the zebrafish genes (Table S3). In line with the findings of Theofanopoulou *et al.* we propose to change the zebrafish gene nomenclature to *oxtra* and *oxtrb* (with correspondingly updated gene names) and retain the existing approved nomenclature for other species. OTR should be included as an "alias" symbol in databases since it is used in the literature, but we encourage researchers to use the approved gene nomenclature to ensure easy identification of their papers and minimise confusion.

4. Vasopressin/vasotocin receptors

a. AVPR1A (referred to as VTR1A in Theofanopoulou et al.)

The existing approved nomenclature (Table S4) accurately represents the function of this gene, and accurately reflects the orthologous relationships between the genes across species. It is used in the literature and is a specific search term for this gene.

As Theofanopoulou *et al.* state, the drawback of the current approved nomenclature is that in some species it has become clear that the "arginine" residue referenced in the gene name is not present, and thus it may be misleading in those species. To avoid confusion in the literature, and in biological databases where the existing approved gene nomenclature has already propagated, we will retain the existing symbols but modify the gene names in the species where the arginine residue is not present to "vasopressin receptor" to avoid any

confusion. Of the species with approved nomenclature, this currently only affects pig. It also affects some marsupials, but these species do not have approved nomenclature at present

b. AVPR1B (referred to as VTR1B in Theofanopoulou et al.)

The existing approved nomenclature system (Table S5) accurately represents the function of this gene, and it accurately reflects the orthologous relationships between the genes across species. The symbol is used in the literature and is a specific search term for this gene.

As for other AVP* gene names, we have removed the word "arginine" from the gene name in pig, while retaining the symbol. We have also updated the gene name in rat to bring it in line with the gene name used by the other gene nomenclature committees. These updates are shown in Table S5. We will retain the existing nomenclature for all other genes.

c. AVPR2 (referred to as VTR2C in Theofanopoulou et al.)

The existing approved nomenclature (Table S6) accurately represents the function of this gene, and it accurately reflects the orthologous relationships between the genes across species. It is used in the literature and is a specific search term for this gene.

We recommend the retention of this symbol as it is currently approved, while aliasing the mammalian and Xenopus symbols as "AVPR2A". This allows the current symbols to be retained, minimizing disruption, while also encoding the orthology relationships to the zebrafish genes, which already contain the "a" in their symbols (Table S6).

As in other AVP* gene names, we have removed the word "arginine" from the gene name in pig, while retaining the symbol. We have also updated the gene name in Xenopus to bring it in line with the gene name used by the other gene nomenclature committees. These updates are shown in Table S6. We will retain the existing nomenclature for all other genes.

d. AVPR2B (referred to as VTR2B in Theofanopoulou et al.)

This gene is absent from all but one of the species with approved nomenclature. Zebrafish has only one copy of this gene (ZFIN:ZDB-GENE-131127-163) but other teleost fish have two, so one of the copies was likely lost in the lineage giving rise to zebrafish. For this reason, Theofanopoulou *et al.* propose assigning the symbol "*vtr2ba*" to this gene, i.e., "vasotocin receptor 2b, duplicate a" (although they also refer to it as "VTR2Bb" in Figure 4a, the supplementary information indicates that *vtr2ba* is the proposed symbol). We have renamed it as *avpr2b.1* (arginine vasopressin receptor 2b, tandem duplicate, 1) to retain the root symbol that is already in use across vertebrates (Table S7).

e. AVPR2C / avpr2I (referred to as VTR2A in Theofanopoulou et al.)

Although Theofanopoulou *et al.* propose the use of the "A" letter in their nomenclature system for this set of genes, we believe the least disruptive option is to reserve "A" for the orthologs of the human gene *AVPR2* (HGNC:897). This remains consistent with the evolutionary history of the AVPR2* family, and ensures that the most well studied genes retain their approved nomenclature while necessary changes are applied to paralogs that have received less attention in the literature thus far.

Since CGNC:7225 is the only paralog of *AVPR2* that chicken has retained, it has been historically named in line with the human *AVPR2*, though it is not a direct ortholog ^{2,3}. We have updated the approved nomenclature of this gene to *AVPR2C*, and also updated the symbol of the Xenopus ortholog to *avpr2c* (Table S8).

Theofanopoulou *et al.* propose that the teleost fish genes currently referred to as *avpr2l* are orthologs of the *AVPR2C* genes, due to partial shared synteny with the *AVPR2C* genes in birds, reptiles, amphibians and non-teleost fish. We are unable to find phylogenetic support for this hypothesis (Fig. S1). The Avpr2l amino acid sequences do not cluster with any of the other three AVPR2* clades with strong support in our phylogenetic analysis. Further, zebrafish has only one copy of this gene and it is not fully syntenic with its orthologs in the other teleost fish species examined, leading Theofanopoulou *et al.* to propose that it represents a duplication of the ancestral gene in teleosts (ie. that it is an out-paralog with respect to the *avpr2l* genes in other teleost species) and that it be named "VTR2Ab" with the other fish orthologs being "VTR2Aa". We find no evidence that these genes are not in single copy in teleost fish. For these reasons we are leaving the nomenclature of zebrafish *avpr2l* unchanged.

The approved nomenclature of the AVPR2C and avpr2l genes is shown in Table S8.

Gene Nomenclature Committee	Species	Database unique ID	Current Approved Symbol	Current Approved Name
HGNC	human	HGNC:8528	ΟΧΤ	oxytocin/neurophysin I prepropeptide
MGNC	mouse	MGI: 97453	Oxt	oxytocin
RGD	rat	RGD:3238	Oxt	oxytocin/neurophysin I prepropeptide
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig	VGNC:6224, VGNC:75730, VGNC:108054, VGNC:54339, VGNC:108055, VGNC:32515, VGNC:96466	ΟΧΤ	oxytocin/neurophysin I prepropeptide
CGNC	chicken	CGNC:13728	ΟΧΤ	oxytocin/neurophysin I prepropeptide
Xenbase	Xenopus	Xenbase:XB-GENE-478274	oxt	oxytocin/neurophysin I prepropeptide
ZNC	zebrafish	ZFIN:ZDB-GENE-030407-1	oxt	oxytocin

 Table S1: Approved nomenclature for the oxytocin genes in vertebrates.

 Table S2: Approved nomenclature for the arginine vasopressin genes in vertebrates.

Gene Nomenclature Committee	Species	Database unique ID	Current Approved Symbol	Current Approved Name
HGNC	human	HGNC:894	AVP	arginine vasopressin
MGNC	mouse	MGI:88121	Ανρ	arginine vasopressin
RGD	rat	RGD:2184	Ανρ	arginine vasopressin
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig	VGNC:6048, VGNC:107771, VGNC:98511, VGNC:38317, VGNC:58944, VGNC:26356, VGNC:96482	AVP	arginine vasopressin (vasopressin in pig only)
CGNC	chicken	CGNC:10532	AVP	arginine vasopressin
Xenbase	Xenopus	Xenbase:XB-GENE-478869	avp	arginine vasopressin
ZNC	zebrafish	ZFIN:ZDB-GENE-030407-2	avp	arginine vasopressin

Gene Nomenclature Committee	Species	Database unique ID	Current Approved Symbol	Previous Approved Symbol	Current Approved Name	Previous Approved Name
HGNC	human	HGNC:8529	OXTR		oxytocin receptor	
MGNC	mouse	MGI:109147	Oxtr		oxytocin receptor	
RGD	rat	RGD:3239	Oxtr		oxytocin receptor	
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig	VGNC:12118, VGNC:75731, VGNC:68667, VGNC:44206, VGNC:21105, VGNC:32516, VGNC:108052	OXTR		oxytocin receptor	
CGNC	chicken	CGNC:2274	OXTR		oxytocin receptor	
Xenbase	Xenopus	Xenbase:XB-GENE-484840	oxtr		oxytocin receptor	
ZNC	zebrafish	ZFIN:ZDB-GENE-110805-2	oxtra	oxtr	oxytocin receptor a	oxytocin receptor
		ZFIN:ZDB-GENE-110805-1	oxtrb	oxtrl	oxytocin receptor b	oxytocin receptor like

 Table S4:
 Approved nomenclature for the arginine vasopressin receptor 1A genes in vertebrates.

Gene Nomenclature Committee	Species	Database unique ID	Current Approved Symbol	Current Approved Name
HGNC	human	HGNC:895	AVPR1A	arginine vasopressin receptor 1A
MGNC	mouse	MGI:1859216	Avpr1a	arginine vasopressin receptor 1A
RGD	rat	RGD:2185	Avpr1a	arginine vasopressin receptor 1A
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig	VGNC:4943, VGNC:70197, VGNC:68838, VGNC:38318, VGNC:15710, VGNC:26358, VGNC:85703	AVPR1A	arginine vasopressin receptor 1A (vasopressin receptor 1A in pig only)
CGNC	chicken	CGNC:7451	AVPR1A	arginine vasopressin receptor 1A
Xenbase	Xenopus	Xenbase:XB-GENE-482551	avpr1a	arginine vasopressin receptor 1A
ZNC	zebrafish	ZFIN:ZDB-GENE-101028-2	avpr1aa	arginine vasopressin receptor 1Aa
		ZFIN:ZDB-GENE-041210-105	avpr1ab	arginine vasopressin receptor 1Ab

Gene Nomenclature Committee	Species	Database unique ID	Current Approved Symbol	Current Approved Name	Previous Approved Name
HGNC	human	HGNC:896	AVPR1B	arginine vasopressin receptor 1B	
MGNC	mouse	MGI:1347010	Avpr1b	arginine vasopressin receptor 1B	
RGD	rat	RGD:6502812	Avpr1b	arginine vasopressin receptor 1B	vasopressin V1b receptor-like
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig	VGNC:452, VGNC:70198, VGNC:68844, VGNC:38319, VGNC:15711, VGNC:26359, VGNC:85704	AVPR1B	arginine vasopressin receptor 1B (vasopressin receptor 1B in pig only)	
CGNC	chicken	CGNC:505	AVPR1B	arginine vasopressin receptor 1B	
Xenbase	Xenopus	Xenbase:XB-GENE-481094	avpr1b	arginine vasopressin receptor 1B	
ZNC	zebrafish		not present		

Table S5: Approved nomenclature for the arginine vasopressin receptor 1B genes in vertebrates.

Table S6: Appro	ved nomencla	ature for the	arginine va	sopressin	receptor	2 genes in	vertebrates.

Gene Nomenclature Committee	Species	Database unique ID	Current Approved Symbol	Current Approved Name	Previous Approved Name
HGNC	human	HGNC:897	AVPR2	arginine vasopressin receptor 2	
MGNC	mouse	MGI:88123	Avpr2	arginine vasopressin receptor 2	
RGD	rat	RGD:2186	Avpr2	arginine vasopressin receptor 2	
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig	VGNC:1407, VGNC:108053, VGNC:68849, VGNC:38320, VGNC:15712, VGNC:26360, VGNC:97898	AVPR2	arginine vasopressin receptor 2 (vasopressin receptor 2 in pig only)	
CGNC	chicken		not currently annotated in birds		
Xenbase	Xenopus	Xenbase:XB-GENE-482287	avpr2	arginine vasopressin receptor 2	arginine vasopressin receptor 2 (nephrogenic diabetes insipidus)
ZNC	zebrafish	ZFIN:ZDB-GENE-090313-344	avpr2aa	arginine vasopressin receptor 2a, duplicate a	
		ZFIN:ZDB-GENE-110411-48	avpr2ab	arginine vasopressin receptor 2a, duplicate b	

Gene Nomenclat ure Committee	Species	Database unique ID	Current Approved Symbol	Previous Approved Symbol	Current Approved Name	Previous Approved Name
HGNC	human		not present			
MGNC	mouse		not present			
RGD	rat		not present			
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig		not present			
CGNC	chicken		not present			
Xenbase	Xenopus		not present			
ZNC	zebrafish	ZFIN:ZDB-GENE- 121023-1	avpr2b <u>.</u> 1	si:dkey-178o16.4	arginine vasopressin receptor 2b, tandem duplicate, 1	si:dkey-178o16.4

 Table S7: Approved nomenclature for the arginine vasopressin receptor 2B genes in vertebrates.

Table S8: Approved nomenclature for the arginine vasopressin receptor 2C and 2I genes in vertebrates.	

Gene Nomenclatur e Committee	Species	Database unique ID	Current Approved Symbol	Previous Approved Symbol	Current Approved Name	Previous Approved Name
HGNC	human		not present			
MGNC	mouse		not present			
RGD	rat		not present			
VGNC	chimpanzee, macaque, cat, dog, horse, cow, pig		not present			
CGNC	chicken	CGNC:7225	AVPR2C	AVPR2	arginine vasopressin receptor 2C	arginine vasopressin receptor 2
Xenbase	Xenopus	Xenbase:XB-GENE-1219042	avpr2c	avpr2.2	arginine vasopressin receptor 2C	arginine vasopressin receptor (nephrogenic diabetes insipidus), gene 2
ZNC	zebrafish	ZFIN:ZDB-GENE-070705-429	avpr2l		arginine vasopressin receptor 2, like	



Figure S1. Maximum likelihood phylogeny of vertebrate AVPR2* amino acid sequences. The teleost Avpr2l clade does not group with the AVPR2C clade, but also doesn't group with either the AVPR2(A) or AVPR2B clades with high confidence. Node labels indicate UltraFast Bootstrap ⁴ with 1000 replicates, values <95 are not well supported. Human AVPR1A was used as the outgroup.

Methods: Amino acid sequences for AVPR2* were aligned using MUSCLE ⁵ and trimmed to remove columns with more than 20% gaps using trimAl ⁶. Maximum likelihood phylogenetic analysis was performed using the IQTree WebServer ⁷ using default parameters. Sequence accession numbers: Human AVPR1A NP_000697.1, Human AVPR2 NP_000045.1, Rat Avpr2 NP_062009.1, Mouse Avpr2 O88721, Chicken AVPR2C NP_001026650.1, Xenopus Avpr2a XP_004916778.1, Xenopus Avpr2c XP_002932869.2, Zebrafish Avpr2aa A0A2R8QNI4, Zebrafish Avpr2ab XP_001922042.4, Zebrafish Avpr2l A5WWC0, Zebrafish Avpr2b.1 E7F1C0, Southern platyfish Avpr2aa XP_023191692.1, Southern platyfish Avpr2ab XP_014325300.2, Southern platyfish Avpr2b.1 XP_005808542.2, Southern platyfish Avpr2b.2 XP_005799961.3, Southern platyfish Avpr2l XP_005799151.1. Spotted gar exonic nucleotide sequences were obtained from ¹ and translated using ExPASy translate.

Supplementary References

- Theofanopoulou, C., Gedman, G., Cahill, J. A., Boeckx, C. & Jarvis, E. D. Universal nomenclature for oxytocin-vasotocin ligand and receptor families. *Nature* 592, 747–755 (2021).
- Ocampo Daza, D., Lewicka, M. & Larhammar, D. The oxytocin/vasopressin receptor family has at least five members in the gnathostome lineage, inclucing two distinct V2 subtypes. *Gen. Comp. Endocrinol.* **175**, 135–143 (2012).
- Yamaguchi, Y. *et al.* The fifth neurohypophysial hormone receptor is structurally related to the V2-type receptor but functionally similar to V1-type receptors. *Gen. Comp. Endocrinol.* **178**, 519–528 (2012).
- Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
- Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641 (2019).
- Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973 (2009).
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44, W232– W235 (2016).