Faculty of Biology
Neurobiology

Expression of the \textit{Oprm1} gene in the mouse striatum and prefrontal cortex

Ekspresja genu \textit{Oprm1} w prążkowiu oraz korze przedczołowej myszy

Magdalena Chróścz

Master thesis prepared under the supervision of
Jan Rodriguez Parkitna, Ph.D.
at the Department of Molecular Neuropharmacology
Maj Institute of Pharmacology, Polish Academy of Sciences

Krakow 2022
Pragnę serdecznie podziękować Promotorowi, Panu dr hab. Janowi Rodriguezowi Parkitnie, za życzliwość, niewyczerpaną cierpliwość, oraz za wszelkie wskazówki udzielone podczas tworzenia tej pracy.

Dziękuję Panu mgr Łukaszowi Szumcowi oraz Panu dr Sławomirowi Goldzie za wprowadzenie w techniki biologii molekularnej i pomoc podczas przeprowadzania doświadczeń.

Pani mgr Klaudii Misiołek oraz Panu mgr Marcinowi Siwcowi dziękuję za współpracę przy doświadczeniach z hybrydzacją in situ.

Dziękuję Panu mgr Jackowi Hajto oraz Panu dr Marcinowi Piechocie z Pracowni Farmakogenomiki Instytutu Farmakologii im. Jerzego Maja PAN za przeanalizowanie danych z sekwencjonowania Nanopore.

Serdeczne podziękowania składam wszystkim pracownikom Zakładu Neurofarmakologii Molekularnej za miłą atmosferę pracy, otwartość i dzielenie się doświadczeniem.
# Table of contents

Abbreviations .................................................................................................................. 2  
Abstract .............................................................................................................................. 3  
Streszczenie ..................................................................................................................... 4  

1. Introduction .................................................................................................................. 5  
   1.1. Endogenous opioid system .................................................................................. 5  
   1.2. μ opioid receptor mechanisms of action ......................................................... 7  
   1.3. The Oprml gene ............................................................................................... 9  
   1.4. Expression of the μ opioid receptor in the central nervous system ............ 12  

2. Aim ............................................................................................................................... 16  

3. Materials and Methods .............................................................................................. 17  
   3.1. Materials ........................................................................................................... 17  
   3.2. Animals ............................................................................................................. 18  
   3.3. Quantitative PCR ............................................................................................ 19  
   3.4. RT-PCR and product sequencing .................................................................... 22  
   3.5. Oxford Nanopore Sequencing ........................................................................ 24  
   3.6. RNAscope in situ hybridization ...................................................................... 25  
   3.7. Image acquisition and processing ................................................................... 26  
   3.8. Statistical analysis ......................................................................................... 26  

4. Results .......................................................................................................................... 27  
   4.1. Oprml transcript abundance in the murine striatum, spleen, and heart ....... 27  
   4.2. Expression of specific Oprml transcripts in the murine striatum ............... 27  
   4.3. Identification of RT PCR products .................................................................. 30  
   4.4. Long-read transcriptome sequencing ............................................................. 35  
   4.5. Visualization of the Oprml 3’UTR in RNAscope ........................................... 39  

5. Graphical summary ...................................................................................................... 42  

6. Discussion .................................................................................................................... 44  

List of figures .................................................................................................................... 50  
List of tables ..................................................................................................................... 51  
References ......................................................................................................................... 51  
Appendix .......................................................................................................................... 58
**Abbreviations**

bp – base pair

cAMP – 3', 5'-cyclic adenosine monophosphate

CREB – cAMP response element-binding protein

DAMGO – [D-Ala2-MePhe4-Gly5-ol] enkephalin

GDP – guanosine-5'-diphosphase

GIRK – G protein-gated inwardly rectifying potassium channels

GRK – G-protein-coupled receptor kinases

GTP – guanosine-5'-triphosphate

IBNtxA – 3-iodobenzoyl-6β-naltrexamide

IGV – Integrative Genome Viewer

ISH – *in situ* hybridization

LTR – long tandem repeat

MAPK – mitogen-activated protein kinases pathway

MOR-1 – μ opioid receptor

ONS – Oxford Nanopore Sequencing

PKC – protein kinase C

qPCR – quantitative PCR

RT-PCR – reverse transcription PCR

SINE – short interspersed nuclear elements

TM – transmembrane

UTR – untranslated region

VGCC – voltage-gated calcium channels
Abstract

The murine μ opioid receptor is encoded by the *Oprm1* gene located in the qA1 region of chromosome 10. The gene undergoes extensive alternative splicing, with approximately 30 transcript variants and multiple protein isoforms reported so far. Here, I investigated the expression of *Oprm1* transcripts in selected brain regions, specifically the striatum and the prefrontal cortex of adult mice bred on C57BL/6N background. In the experiments, several mRNA detection assays were used, among them, qPCR, RT-PCR, Oxford Nanopore Sequencing (ONS), and RNAscope *in situ* hybridization (ISH). The research confirmed the expression of variant types that contain either exon 4 or exon 7 at their 3’ end. Furthermore, RT-PCR revealed the presence of *Oprm1* mRNA containing exon 19, similar to previously reported MOR-1U, MOR-1Eii, and MOR-1Eiv. No evidence for transcription of exon 11 was found. Interestingly, the sequencing data showed the existence of a previously unidentified 10 kbp-long contig downstream of exon 4 – a putative long 3’UTR of *Oprm1*. To confirm that the contig is a part of the μ opioid receptor’s mRNA, I have performed a preliminary ISH analysis using the RNAscope method. Initial results confirmed the presence of the putative long *Oprm1* 3’ UTR in the striatum and the prefrontal cortex, though due to technical issues transcript variant overlap could not be determined. In future research, I would like to repeat the ISH experiments with protocol modifications, as well as assess the distribution of the putative long *Oprm1*-3’UTR in various cell types.

Key words: *Oprm1*, alternative transcription, 3’UTR, μ-opioid receptor, sequencing
Streszczenie

Słowa kluczowe: Oprm1, alternatywna transkrypcja, 3’UTR, receptor opioidowy μ, sekwencjonowanie
1. Introduction

Murine μ opioid receptor is encoded by the *Oprm1* gene. The receptor is the main target of opioid analgesic drugs, including morphine, oxycodone, and fentanyl. The receptor is best known for its involvement in mediating analgesia, but it also plays a significant role in modulating affective states, reward, motivation, and stress response (Emery & Akil, 2020). Activation of the receptor is associated with numerous adverse effects of opioid analgesics, such as tolerance, physical dependence, sedation, respiratory depression, miosis, and constipation, that limit clinical applications of opioids. Despite extensive research, many gaps remain in our understanding of the μ opioid receptor activity, such as the molecular mechanisms behind adverse effects, unexpected differences in pharmacological properties between various agonists, or even the existence of multiple receptor subtypes (Pasternak & Pan, 2013). Due to the well-known clinical and physiological importance of the receptor, it remains necessary to expand our knowledge and try to fill the remaining gaps. In this thesis, I focus on the variability that arises from multiple variants of the μ opioid receptor transcripts.

1.1. Endogenous opioid system

The existence of common receptor binding sites for opioid analgesics was first postulated in 1954 by Beckett and Casy, based on the research on structural requirements for the activity of examined antinociceptive opiates (Beckett & Casy, 1954). In the 1960s the existence of multiple opioid receptors or different modes of interaction with ligands was proposed (Portoghese, 1965). Within a decade, the first specific opiate binding sites were identified in the mouse brain and guinea pig intestine using radioligand assays (Pert & Snyder, 1973). And the following studies allowed for experimental distinction of three putative types of opioid receptors, which were named after the pharmacological profiles of the compounds used in the studies or after their anatomical location: μ (after morphine), δ (after vas deferens), and κ (after ketocyclazocine) (Lord et al., 1977; Martin et al., 1976). This was followed by studies investigating potentially different subtypes of opioid receptors, which culminated in the identification of cDNAs of three distinct receptors (Chen et al., 1993; Evans et al., 1992; Meng et al., 1993). These discoveries paved the way for the generation of mice with 'knockout genes', providing conclusive evidence for the hypothesis of single genes encoding each type of opioid receptors and expanded our understanding of the functions of each receptor.
Opioid receptors are members of the metabotropic, rhodopsin-like G-protein coupled receptor (GPCR) superfamily. In the mouse, they are encoded by separate genes: \textit{Oprm1}, \textit{Oprd1}, \textit{Oprk1} for μ, δ, and κ receptors, respectively. These genes, along with \textit{Oprl1}, which encodes for the nociceptin/orphanin FQ receptor, are paralogous – they originate from a single ancestral gene that underwent quadruplication about 500 million years ago during tetraploidisation of the genome (Larhammar et al., 2015). The opioid receptors possess an extracellular N-terminus, seven transmembrane domains, 3 extracellular and 3 intracellular loops, and an intracellular C-terminus. The transmembrane regions are highly conserved, with 70-90% amino acid identity among the four receptors in the sequences encoding the intracellular loops and transmembrane domains, but only 20% identity in N-, C- terminal regions and extracellular loops (Law et al., 2000).

The opioid receptors are activated by endogenous ligands: enkephalins, dynorphins, neoendorphins, and β-endorphin. The peptides are derived from three precursors, preproenkephalin, preprodynorphin, and proopiomelanocortin, encoded in the mouse genome by single genes: \textit{Penk}, \textit{Pdyn}, and \textit{Pomc}, respectively (Akil et al., 1984). To produce active opioid peptides, the protein precursors undergo extensive enzymatic cleavage and modifications. All the ligands share a common N-terminal signature sequence that consists of 4 amino acids, YGGF, and is critical for receptor binding. The endogenous ligands are characterized by limited selectivity towards their receptors, with short pentapeptides: Met- and Leu-enkephalins binding preferentially to δ opioid receptor, other members of the enkephalin family towards μ, but also κ and δ receptors, dynorphins binding mainly to κ opioid receptor, and β-endorphin displaying a similar affinity towards μ and δ opioid receptor (Akil et al., 1984; Mansour et al., 1995). The affinities for specific receptors are graphically summarised in Figure 1.
Figure 1. Endogenous opioid peptides and their receptors. The diagram summarizes the affinities of enkephalins, dynorphins, and beta-endorphin to the μ, δ, and κ receptors. The outer rings represent the peptide precursors, and the inner rings correspond to the processed, active peptides. The numbers in the spokes are reported $K_d$ values (nM), the greater the spoke's thickness the higher the peptide's affinity for the receptor. Source: Emery and Akil, 2020, adapted from Mansour et al., 1995.

It should be noted that an additional group of peptides that specifically bind μ opioid receptor was reported, the endomorphins (Zadina et al., 1997). Endomorphin-1 and endomorphin-2 are tetrapeptides with amino acid sequences: YPDG and YPFF, respectively. They are characterized by high affinity ($K_i = 0.36$ nM and 0.69 nM) and selectivity to the μ opioid receptor (> 4000-fold preference over the δ and κ receptors). They were suggested to participate in processing of nociceptive information; however, relatively little is known about this group, as their protein precursor remains unknown (Gu et al., 2017).

### 1.2. μ opioid receptor mechanisms of action

The opioid receptors are canonically coupled to the effector heterotrimeric $G_{i/o}$ protein that consists of membrane-anchored subunits: $G\alpha$ and $G\beta\gamma$. Studies using cryomicroscopy revealed that the main interaction sites between μ opioid receptor and $G\alpha$ subunit occur between the intracellular loops 2, 3 and the transmembrane (TM) loops 3, 5, and 6 of the receptor and the α-helices of the $G\alpha$ (Koehl et al., 2018).

Activation of the receptor induces changes in its conformation, which in turn facilitates conformational changes in the $G\alpha$ protein and promotes the exchange of guanosine-5'-diphosphate (GDP) with guanosine-5'-triphosphate (GTP) (Figure 2). The exchange induces the dissociation of the $G_{o}$ from the receptor and its split into $G\alpha$ and $G\beta\gamma$ subunit. Dissociated $G\beta\gamma$ subunit activates $G$ protein-gated inwardly rectifying potassium channels (GIRK) (Al-Hasani & Bruchas, 2011). The outward potassium...
current causes hyperpolarization and acts to inhibit tonic neuronal firing. Additionally, the Gβγ suppresses the opening of voltage-gated calcium channels (VGCC). The process blocks the release of neurotransmitters and also contributes to the inhibition of neuronal firing (Siegelbaum et al., 2013). Concurrently, the Ga subunit inhibits adenylyl cyclase attenuating production of the second messenger, 3', 5'-cyclic adenosine monophosphate (cAMP). A low level of cAMP yields a decrease in the cAMP-dependent influx of Ca²⁺ and downregulation of protein kinase A, which contributes to changes in the activity of transcription factors and regulatory proteins, for example, the cAMP response element-binding protein (CREB) (Al-Hasani & Bruchas, 2011; Law et al., 2000).

Upon activation, the C-terminus of the μ opioid receptor becomes a target for phosphorylation by serine/threonine kinases: protein kinase C (PKC) and G-protein-coupled receptor kinases (GRK) (Figure 2) (Allouche et al., 2014). The phosphorylation pattern is strongly dependent on the ligand and determines β-arrestin binding. β-arrestins 1 and 2 are scaffold proteins that promote receptor uncoupling from the G-protein and enables the recruitment of alternate signal transduction cascades, including the mitogen-activated protein kinases pathway (MAPK): ERK1/2, JNK1-3 and p38. These processes further contribute to changes in the activity of numerous transcriptional factors and regulatory proteins, e.g., CREB (Al-Hasani & Bruchas, 2011; Allouche et al., 2014). β-arrestins are also crucial for receptor internalization via endocytosis. After endocytosis, μ opioid receptors are returned to the cellular membrane.
Earlier research suggested that signalling bias towards G-protein or β-arrestin-associated signalling pathways may explain discrepancies in the pharmacological profiles of various μ opioid compounds. It was postulated that ligands preferentially activating the ‘classical’ G-protein signalling pathway would be devoid of adverse effects, e.g. respiratory depression and tolerance, which were attributed to β-arrestin signalling (Bohn et al., 2000). Due to difficulties in replication of the research, the hypothesis is criticised (Gillis et al., 2020) and a more unequivocal picture of the role of downstream signalling pathways and receptor trafficking in the regulation of μ opioid receptor is supported (Allouche et al., 2014; Gillis et al., 2020; Williams et al., 2013).

Regardless, many other factors have been hypothesized to contribute to the complexity of the μ opioid activity, like allosteric modulation, receptor’s dimerization or the existence of receptor isoforms. In the following subsection, I will focus on the transcriptional diversity of the Oprml gene, that leads to generation of multiple isoforms of the μ opioid receptor.

1.3. The Oprml gene

The μ opioid receptor is encoded by the Oprml gene in rodents (OPRM1 in humans). The gene is evolutionarily conserved and has been sequenced in more than 200 species including chicken, zebrafish, and sea lamprey. In humans the gene is located on chromosome 6 (6q25.2), in rat on chromosome 1 (1q11), and in mouse on chromosome 10 (10qA1). Its locus spans over 210, 264, and 270 kbp, respectively. Up to date, there are at least 11 exons reported in humans, 13 in rats (Rattus norvegicus), and 19 in mice (Mus musculus) that comprise for the Oprml gene (Doyle et al., 2007; Pasternak & Pan, 2013). In these species, exon 1 encodes for N-terminal end and the first transmembrane domain. Exons 2 and 3 encode for the transmembrane regions 2-7 as well as intra- and extracellular loops. They are highly preserved and present in nearly all reported variants. Additionally, exon 3 also contains a part of the carboxy terminal encoding sequence. The downstream exons encode for the remaining parts of C-terminus and the 3’UTRs of the synthetised mRNA. Putative STOP codons are present in exons 4, 6, or 9 (Pasternak & Pan, 2013).

The Oprml gene undergoes extensive alternative splicing. Over 30 transcript variants of the murine gene were reported (Figure 3). These variants generate multiple receptor isoforms that differ substantially in their N- and C- termini. Most of the transcript variants contain exons 1-3 and one or a few additional exons at the 3’ end. Such variants
produce functional 7-TM receptors, . In addition to the 'classical' 7-TM isoforms, proteins with 6- or 1-TM domains were postulated (Pasternak, 2001). The truncated 6-TM receptors are derived from variants that contain exon 11 instead of exon 1, while the 1-TM isoforms arise from transcripts that lack exon 2 or 3 (Pasternak & Pan, 2013). According to evidence from mice, deletion of exon 1 results in the generation of 6-TM protein isoforms (Majumdar et al., 2011). Such proteins are functional receptors with region-specific distribution in the brain (Abbadie et al., 2004; Xu et al., 2014) and distinct affinity and selectivity towards μ opioid ligands compared to 7-TM proteins. For example, mice with exon-1 deletion did not produce morphine-induced analgesia, while the analgesia produced by another μ opioid agonist, 3-iodobenzoyl-6β-naltrexamide (IBNtxA) was preserved. On the contrary, deletion of exon 11 did not affect morphine analgesia, however, it strongly reduced IBNtxA-induced analgesia (Lu et al., 2018; Majumdar et al., 2011). The 1-TM isoforms were not shown to bind opioid ligands. They were suggested to have chaperon-like function and dimerise with 7-TM receptors, preventing them from misfolding and degradation in the endoplasmic reticulum (Xu et al., 2013).

The expression of exon 1- or exon 11-containing variants is dependent on promoter. There are two independent promoter sites in the *Oprm1* gene, the first is located upstream of exon 11 and encoding mainly 6-TM isoforms, while the second localize upstream exon 1, encoding mainly 7-TM isoforms (Pan, 2002). The exon 1-associated promoter consists of the proximal and the distal elements and lacks a TATAbox, which is important for the initiation of transcription. It possesses several transcription starts and includes a few GC-rich cis-acting motifs that bind transcriptional factors like STAT1/3, AP1, Ik, NF-κB. The structure of the promoter suggests an expression pattern more typical for housekeeping genes (Pasternak & Pan, 2013). On the other hand, exon 11-associated promoter is the only promoter among the opioid receptor gene family that contains a TATAbox sequence (Pasternak & Pan, 2013). It also includes cis-acting elements like NF-1 or cMyc/Max that modulate its activity, however, because the promoter requires the formation of a preinitiation complex and recruitment of RNA polymerase II, it is classified as eukaryote class II promoter (Pan, 2002; Pasternak & Pan, 2013).
Figure 3. Schematic representation of the mouse Oprm1 gene and reported transcript variants. Variants divided into three groups according the number of encoding transmembrane domains. Exons are depicted as rectangles. Schematic representation of putative translated proteins is shown on the right. Adapted from Pasternak & Pan, 2013.
Oprm1 expression is also modulated by multiple transcriptional factors, repetitive elements and epigenetic factors like methylation of GpCs, histone methylation or histone acetylation (Cuitavi et al., 2021; Wu et al., 2019). For instance, it was observed that exon 1-associated promoter contains multiple GpC sites. High methylation in this region was shown to suppress the expression of the Oprm1 in the dorsal root ganglia (Sun et al., 2021). Also in humans, it was shown to contribute to various pathological conditions from increased risk of alcohol or opioid dependence (Ebrahimi et al., 2018; Zhang et al., 2012) to Alzheimer Disease diagnosis (Xu et al., 2018).

1.4. Expression of the μ opioid receptor in the central nervous system

The μ opioid receptor is abundant in the telencephalon, mesencephalon, pons, medulla, and spinal cord. Relatively high protein levels were observed in the laminar parts of the frontal cortex, piriform cortex and entorhinal cortex, the hippocampal formation and subiculum, as well as in the amygdala, especially the medial and lateral nuclei (Mansour et al., 1995). The high level of the receptors is present in the basal ganglia, dorsal and ventral striatum, ventral pallidum, globus pallidus, and the bed nucleus of the stria terminalis. The highest abundance of the receptor were observed in the thalamus and medial habenula. The hypothalamus is characterized by a relatively low receptor level in the dorsomedial nucleus and the lateral hypothalamic area. Furthermore, the receptors were observed in the mesencephalic structures, the interpeduncular nucleus, the substantia nigra, the ventral tegmental area, the superior and inferior colliculi, as well as the dorsal raphe nucleus. μ opioid receptor is also abundant in the pons, especially the parabrachial nucleus, the nucleus tractus solitarius, the spinal trigeminal nucleus, and the nucleus raphe magnus. High levels of μ opioid receptors were observed in the substantia gelatinosa of the spinal cord (Mansour et al., 1988; . The presence of μ opioid receptors in the radioligand-binding assays does not necessarily correlate with high levels of Oprm1 mRNA. The apparent discrepancies in distribution occur in regions of the neocortex, olfactory bulb, superior colliculus, and spinal cord. The discrepancies may result from receptor mRNAs transport along the axons (Mansour et al., 1995).

The expression of the Oprm1 mRNA in the striatum was confirmed by ISH assays, qPCRs, and northern blotting (Banghart et al., 2015; Minami et al., 1994; Oude Ophuis et al., 2014). Within the dorsal striatum, Oprm1 transcripts co-localize with Pdyn mRNA. Expression of μ opioid receptors occurs in both D1- and D2-positive spiny projection neurons, with slight preference towards D1-positive cells (Banghart et al., 2015; Oude
The μ opioid receptor proteins create characteristic patches that converge with inputs from the limbic cortex (Banghart et al., 2015) Additionally, the proteins were also observed in 'giant' cholinergic interneurons and presynaptic terminals of neurons projecting from the cortex (Ponterio et al., 2013; H. Wang & Pickel, 1998). In the ventral striatum, *Oprm1* transcripts were observed in the nucleus accumbens core, especially the dorsomedial part referred to the ‘hedonic hot-spot’, the nucleus accumbens shell, and the olfactory tubercle. In the prefrontal cortex μ opioid reactivity was found in all layers I-V of the cortex, in layer I, the receptor was expressed together with Neuropeptide Y, whereas as in layers II-IV with enkephalin and vasoactive intestinal peptide (Férézou et al., 2007) In another study the μ opioid receptor reactivity was observed on small, bipolar, non-pyramidal neurons, located in layers II-IV (Taki et al., 2000). The distribution of *Oprm1* mRNA in the striatum and the cortex can be observed in images of spatial transcriptomics (Ståhl et al., 2016), generated in previous research at the Department of Molecular Neuropharmacology (data not published). Spatial transcriptomics utilizes short-read sequencing that captures the poli-A tails of the mRNAs. Usually, the reads cover 150-300 nucleotides, and through specific barcodes, they can be assigned to the specific spots on the prepared slide, each spot 55 μm in diameter. The visualization of the *Oprm1* transcripts obtained from spatial transcriptomics is presented in Figure 4. The Figure depicts predominant expression of μ opioid receptor mRNA in the dorsal striatum, nucleus accumbens, ventral pallidum, islands of Calleja, lateral septal nucleus and dorsal endopiriform nucleus with lower expression in the cortex, especially the cingulate and piriform cortex, which remains in line with the literature (Mansour et al., 1995)
Figure 4. A. The visualization of *Oprm1* transcripts in spatial transcriptomics. The color of the dots represents expression levels according to the scale shown on the left. B. The corresponding brain slice using ISH with probe targeting the *Oprm1* gene (*Gene :: Allen Brain Atlas: Mouse Brain, 2011*).

After discarding spatial information, the data revealed a few contigs suggesting the existence of several types of transcripts with diverse polyadenylation sides. The peaks of reads are schematically visualized in Figure 5. Whereas the contigs located in close proximity to exons 4 and 9 may be annotated to the *Oprm1* transcripts containing exon 4 and exon 9, the other peaks are difficult to annotate to the specific *Oprm1* exon.

![Figure 5 Polyadenylation sites of the Oprm1 gene in the striatum and the prefrontal cortex revealed by the spatial transcriptomics method. Scheme of exons 4 to 9 of the Oprm1 gene (GRCm38.v6). The double-slanted lines represent parts of the gene that are omitted in the scheme. Observed reads of peaks are indicated by arrows.](image)

In summary, accumulated evidence indicates that the *Oprm1* gene has multiple transcripts that may generate receptor proteins of distinct functions, different affinities for ligands, and particularly high variability at the carboxy terminus, a region involved in
signal transduction pathways and internalization of the receptor. The existence of multiple types of the receptor implies that it could be possible to develop selective medications targeting specific subpopulations of μ opioid receptors and thus to develop improved analgesics. However, existing data on variants is largely based on just one analysis method (i.e., PCR) and there is limited data on the cell-specificity of expression of the reported variants.
2. Aim
The primary aim of the study was to determine the presence of selected transcripts of the *Oprm1* gene in the striatum of adult mice. The second aim was to determine the distribution of specific transcript variants in the striatum and the prefrontal cortex.
3. Materials and Methods

3.1. Materials

3.1.1. Chemicals

<table>
<thead>
<tr>
<th>Name</th>
<th>Manufacturer</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 × PBS Buffer</td>
<td>Serva, Germany</td>
<td>RNAscope ISH</td>
</tr>
<tr>
<td>50 × Wash Buffer</td>
<td>Advanced Cell Diagnostics, Inc., USA</td>
<td>RNAscope ISH</td>
</tr>
<tr>
<td>6 × DNA Loading Dye</td>
<td>VWR Life Science, USA</td>
<td>RT-PCR</td>
</tr>
<tr>
<td>Agarose Bio Standard</td>
<td>Prona Agarose, Spain</td>
<td>RT-PCR</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>POCH, Poland</td>
<td>RNA isolation</td>
</tr>
<tr>
<td>Nuclease-free Water</td>
<td>Ambion Inc., USA</td>
<td>RNA isolation, RT-PCR</td>
</tr>
<tr>
<td>Oligo-dT Primers</td>
<td>Invitrogen, USA</td>
<td>RNA isolation</td>
</tr>
<tr>
<td>Optical Cutting Temperature</td>
<td>Cell Path, UK</td>
<td>RNAscope ISH</td>
</tr>
<tr>
<td>Paraformaldehyde Reinst, DAC</td>
<td>Carl Roth, Germany</td>
<td>RNAscope ISH</td>
</tr>
<tr>
<td>Prolong Gold Antifade Mountant</td>
<td>Invitrogen, USA</td>
<td>RNAscope ISH</td>
</tr>
<tr>
<td>RNA later Stabilization Solution</td>
<td>Qiagen Inc., USA</td>
<td>Tissue preparation</td>
</tr>
<tr>
<td>RNAscope® Probes</td>
<td>Advanced Cell Diagnostics, Inc., USA</td>
<td>RNAscope ISH</td>
</tr>
<tr>
<td>RNAscope® Protease IV</td>
<td>Advanced Cell Diagnostics, Inc., USA</td>
<td>RNAscope ISH</td>
</tr>
<tr>
<td>TaqMan Probes</td>
<td>Thermo Fisher Scientific, UK</td>
<td>Quantitative PCR</td>
</tr>
<tr>
<td>TEA buffer</td>
<td>A&amp;A Biotechnology, Poland</td>
<td>RT-PCR</td>
</tr>
<tr>
<td>Trizol Reagent</td>
<td>Invitrogen, USA</td>
<td>RNA isolation</td>
</tr>
</tbody>
</table>

3.1.2. Kits and ready-made solutions

<table>
<thead>
<tr>
<th>Name</th>
<th>Manufacturer</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>DreamTaq Green PCR Master Mix (2 ×)</td>
<td>Thermo Fisher Scientific, UK</td>
<td>PCR</td>
</tr>
<tr>
<td>Gel Purification Kit</td>
<td>Qiagen Inc., USA</td>
<td>Short-read sequencing</td>
</tr>
<tr>
<td>Omniscript RT Kit</td>
<td>Qiagen Inc., USA</td>
<td>Reverse transcription</td>
</tr>
<tr>
<td>PCR Product Purification Kit</td>
<td>Qiagen Inc., USA</td>
<td>Short-read sequencing</td>
</tr>
<tr>
<td>RNA 6000 Nano Methods LabChip</td>
<td>Agilent, USA</td>
<td>Nanopore sequencing</td>
</tr>
<tr>
<td>RNAscope Fluorescent Multiplex Detection Kit</td>
<td>Advanced Cell Diagnostics, Inc., USA</td>
<td>RNAscope ISH</td>
</tr>
</tbody>
</table>
RNAscope Intro Pack For Multiplex Fluorescent Reagent Kit
TaqMan Universal PCR Master Mix (2 ×)  
Advanced Cell Diagnostics, Inc., USA  
Thermo Fisher Scientific, UK  
RNAscope ISH  
Quantitative PCR

3.1.3. PCR primers’ sequences

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5’-3’</th>
<th>Targeted exon/exon junction</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF1</td>
<td>GACCCATGCGGTCTCTAACC</td>
<td>1</td>
</tr>
<tr>
<td>PF11</td>
<td>TCTGGGCGATGATGGAAGC</td>
<td>11</td>
</tr>
<tr>
<td>PR2</td>
<td>TCTGCCAGAGCAAGGTTGAA</td>
<td>2</td>
</tr>
<tr>
<td>PR11/2</td>
<td>TCTGAGGTAACTCTTCCCCTTTGA</td>
<td>11/2</td>
</tr>
<tr>
<td>PF3</td>
<td>TCCTCCACAATCGAAGCACA</td>
<td>3</td>
</tr>
<tr>
<td>PF2/3</td>
<td>CAGGCAGGGGTCCATAGATTGC</td>
<td>2/3</td>
</tr>
<tr>
<td>PR_6860kb</td>
<td>GATCTGTGTAGGATGGAGCTATT</td>
<td>-</td>
</tr>
<tr>
<td>PR_6990kb</td>
<td>GCCACATTTCATGCCCCAAG</td>
<td>19</td>
</tr>
</tbody>
</table>

3.1.4. RNAscope probes

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Accession/region</th>
<th>Reactivity/species</th>
<th>Catalogue Number</th>
<th>Channel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penk</td>
<td>NM_001002927.2</td>
<td>Mus Musculus</td>
<td>318761-C2</td>
<td>2</td>
</tr>
<tr>
<td>Oprm1</td>
<td>NM_001039652.1</td>
<td>Mus Musculus</td>
<td>315841-C3</td>
<td>3</td>
</tr>
<tr>
<td>Oprm1-O7</td>
<td>chr10:6858476-6861848 (GRCm38)</td>
<td>Mus Musculus</td>
<td>1178981-C1</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2. Animals

In the experiment tissue from adult, C57BL/6 mice (Mus musculus L.) was used. The qPCR and Oxford Nanopore Sequencing (ONS) were performed on adult wild type mice from the Penk<sup>tm1Rpa</sup> strain (Misiołek, 2020). The RNAscope experiments were conducted on 14-week old, wild type male mice from the Pdyn<sup>tm1Zim</sup> strain (Zimmer et al., 2001). Genetically modified mice were congenic to the C57BL/6N strain. The animals were kept in the in-house breeding facility of the Maj Institute of Pharmacology of the Polish Academy of Sciences. Animal genotypes were confirmed by PCR on DNA
isolated from distal tail tissue biopsy. Primers as follows were used, for Penk KO: primer forward – CTACAGGCGCTTCTTCTCT and primer reverse – GGATGGAGCAG GTAGTGG, for Pdyn KO: primer forward – CTTCAAGTAGTTATGGGGTTC TCCTGGG and primer reverse – CGCACCCTCTTTTAATGAGGAGGACTTG.

Mice were group-housed (2-6 animals per cage) in standard type II L cages, with ad libitum access to food and water. The breeding facility was kept in a 12/12h light/dark cycle with the light phase starting at 7:00 a.m. The temperature and humidity were controlled and maintained at 22 ±2 ºC, 55±10%, respectively. The ARRIVE guidelines were followed in planning of the experiments and reporting the results. A summary of animals used in the experiments is shown in Table 1.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Number of Animals</th>
<th>Strain, Genotype</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT-PCR and product sequencing</td>
<td>5</td>
<td>C57BL/6</td>
<td>males</td>
<td>adult</td>
</tr>
<tr>
<td>qPCR</td>
<td>3</td>
<td>Penk&lt;sup&gt;tm1Rpa&lt;/sup&gt; wt wt</td>
<td>males</td>
<td>14-18 weeks</td>
</tr>
<tr>
<td>ONS</td>
<td>2</td>
<td>Penk&lt;sup&gt;tm1Rpa&lt;/sup&gt; wt wt</td>
<td>males females</td>
<td>16-18 weeks</td>
</tr>
<tr>
<td>RNAScope ISH</td>
<td>2</td>
<td>Pdyn&lt;sup&gt;tm1Zim&lt;/sup&gt; wt wt</td>
<td>males</td>
<td>14 weeks</td>
</tr>
</tbody>
</table>

Table 1. Number, strain (genotype), sex, and age of the animals used for specific experiments

Mice that were used for the RNAScope experiment underwent mild behavioral testing prior to sacrifice.

### 3.3. Quantitative PCR

#### 3.3.1. Tissue preparation

Animals were euthanized by cervical dislocation. Brains, hearts, and spleens were removed immediately after decapitation, submerged in falcons with RNAlater Stabilization Solution (Qiagen Inc., USA) and placed at 4ºC. The next day, the striatum (caudate putamen and nucleus accumbens) was isolated under stereo microscope (StemiDV4, Carl Zeiss, Germany) with 23G sterile needles. Small sections containing heart apexes and spleens were excised. All tissue samples were placed in 2.0 ml round bottom Eppendorf® tubes and kept at -20ºC.

#### 3.3.2. RNA isolation

RNA was isolated using the single-step acid guanidinium thiocyanate-phenol-chloroform extraction (AGPC) method (Chomczynski, 1993). Samples were thawed at
room temperature, then the RNAlater solution was removed, and stainless steel balls were placed in the tubes. First, tissue samples were homogenized in 1 ml of TRIzol reagent (Invitrogen, USA) in the TissueLyser II apparatus (Qiagen Inc., USA). Homogenization was performed twice for 3 minutes at 25 Hz. To precipitate RNA, 200 µl of chloroform was added to each tube. Samples were incubated for 10 minutes on ice, mixed by vortexing, and centrifuged (12 000g, 20 minutes, 4°C). The aqueous phase (approx. 450 µl) was transferred to the new 1.5 ml tubes, and an equal volume of isopropanol (POCH, Poland) was added to each tube. Samples were mixed, placed at -70°C for 20 minutes and then centrifuged (12 000g, 30 minutes, 4°C). Supernatant was discarded and the RNA-containing pellets were resuspended in 1 ml of 70% (v/v) ethanol and further centrifuged (12 000g, 10 minutes, 4°C). After washing with ethanol, the pellets were air-dried and dissolved in 30µl of sterile, nuclease-free water (Ambion Inc., USA). Finally, Samples were incubated in thermomixer (Thermomixer Comfort, Eppendorf, Germany) set at 65°C for 5 minutes, then mixed and stored at -20 ºC. RNA concentration was measured using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., USA). Total RNA concentration was calculated based on the values obtained at A$_{260}$. The obtained results are presented in Table 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Blank</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Heart1</td>
<td>689.92</td>
<td>1.97</td>
</tr>
<tr>
<td>Heart2</td>
<td>1016</td>
<td>2</td>
</tr>
<tr>
<td>Heart3</td>
<td>1585.76</td>
<td>2</td>
</tr>
<tr>
<td>Spleen1</td>
<td>2280.11</td>
<td>1.98</td>
</tr>
<tr>
<td>Spleen2</td>
<td>1562.77</td>
<td>2.01</td>
</tr>
<tr>
<td>Spleen3</td>
<td>3607.2</td>
<td>1.87</td>
</tr>
<tr>
<td>Striatum1</td>
<td>792.75</td>
<td>1.97</td>
</tr>
<tr>
<td>Striatum2</td>
<td>180.28</td>
<td>1.87</td>
</tr>
<tr>
<td>Striatum3</td>
<td>360.62</td>
<td>1.95</td>
</tr>
</tbody>
</table>


3.3.3. Reverse transcription

Reverse transcription was performed using the Omniscript RT Kit (Qiagen Inc., USA). RNA samples were thawed on ice, mixed by vortexing and briefly spun down at 4°C. 10 × Buffer RT, dNTP Mix, RNase-free water and oligo-dT primers (Invitrogen, USA) were thawed in RT, mixed, short spin centrifuged, and kept on ice. The volumes of
the components were calculated to obtain 2 µg of template RNA in 20 µl of the reaction mixture, as in Table 3. 10 × Buffer RT, dNTP Mix and oligo-dT primers and the Omniscript Reverse Transcriptase enzyme were mixed together and added to the tubes with diluted template RNA, the prepared mixtures were gently stirred with a pipette. The samples were incubated for 60 minutes at 37°C and then diluted 20 times with sterile water.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume/Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 × Buffer Rt</td>
<td>2 µl</td>
</tr>
<tr>
<td>dNTP Mix (5 mM each dNTP)</td>
<td>2 µl</td>
</tr>
<tr>
<td>Oligo-dT (10 µm)</td>
<td>2 µl</td>
</tr>
<tr>
<td>Omniscript Reverse Transcriptase</td>
<td>1 µl</td>
</tr>
<tr>
<td>RNase-free Water</td>
<td>variable</td>
</tr>
<tr>
<td>Template RNA</td>
<td>variable</td>
</tr>
<tr>
<td>Total Volume:</td>
<td>20 µL</td>
</tr>
</tbody>
</table>

Table 3. Volumes of the reverse-transcription reaction components

3.3.4. Quantitative PCR (qPCR)

The abundance of µ opioid receptor mRNA was analysed using the TaqMan Mm01188089_m1 probe (Thermo Fisher Scientific, UK). The probe target junction between exons 2 and 3 of the MOR-1C transcript of the murine Oprm1 gene (NCBI reference sequence – NM_001039652.2). The Hprt1 probe (NM_013556.2) was used as a control. Both probes were labelled with FAM dye (Carboxyfluorescein).

After thawing, the probes were mixed with TaqMan Universal PCR Master Mix (2 ×) (Thermo Fisher Scientific, UK) in the volumes indicated in Table 4. The mixture was gently stirred, vortexed, and aliquoted into a 6-well PCR plate (BIO-RAD, USA). Subsequently, the template cDNA was dispensed to the wells. The plate was sealed with an adhesive film (BIO-RAD, USA) and briefly centrifuged. The reaction was performed on the C1000 CFX96 Touch Real-Time PCR cycler (BIO-RAD, USA). The reaction parameters are presented in Table 5. Cycle threshold values (Ct) were automatically calculated by the PCR machine software, relative gene expression was determined using ΔCt method.
Table 4. Volumes of the qPCR reaction components

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume/Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaqMan Universal PCR Master Mix (AmpliTaq Gold® DNA Polymerase, Uracil-N glycosylase (UNG), dNTPs with dUTP, ROX™ Passive Reference, optimized buffer components) (2 ×)</td>
<td>10 µl</td>
</tr>
<tr>
<td>TaqMan Probe</td>
<td>1 µl</td>
</tr>
<tr>
<td>Template cDNA</td>
<td>9 µl</td>
</tr>
<tr>
<td>Total Volume:</td>
<td>20µl</td>
</tr>
</tbody>
</table>

Table 5. Cycling parameters of the qPCR

<table>
<thead>
<tr>
<th>Stage</th>
<th>Temperature</th>
<th>Time</th>
<th>Repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNG Activation</td>
<td>50°C</td>
<td>2 min</td>
<td></td>
</tr>
<tr>
<td>Polymerase Activation, DNA Denaturation</td>
<td>95 °C</td>
<td>10 min</td>
<td></td>
</tr>
<tr>
<td>Amplification</td>
<td>95 °C, 60 °C</td>
<td>15s, 60s</td>
<td>40 ×</td>
</tr>
</tbody>
</table>

3.4. RT-PCR and product sequencing

3.4.1. Polymerase chain reaction

To assess the presence of selected *Oprm1* transcripts in the mouse striatum RT-PCR was performed. The cDNA used as templates was prepared as described in Subsections 3.3.1 - 3.3.3. Primers targeting different variants of murine *Oprm1 splicing* were designed based on transcript sequences available in the Ensemble Genome Browser (https://www.ensembl.org/Mus_musculus/Gene/TranscriptComparison?db=core;g=ENSMUSG00000000766;r=10:6708506-6988198) using the NCBI Primer BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The primers were supplied by the DNA Sequencing and Oligonucleotide Synthesis Lab at the Institute of Biochemistry and Biophysics, PAS and prepared according to manufacturer’s instructions.

Samples were thawed at room temperature. Reaction reagents: dH₂O, primers, and Master Mix (DreamTaq Green PCR Master Mix (2 ×), Thermo Scientific, USA) including optimized DreamTaq Green buffer, MgCl₂, and dNTPs were mixed together and parcelled into a 96-Well PCR Plate (BIO-RAD, USA). The samples were added to the wells. The volumes of specific components per well are shown in Table 6. The plate was then sealed with an adhesive film (BIO-RAD, USA), briefly centrifuged and placed in the Thermocycler (C100 Touch™ Thermal Cycler, BIO-RAD, USA). The temperature profiles and cycling parameters were adjusted to the expected lengths of the amplified
products and the thermodynamic properties of the primers. To increase the specificity of obtained products, the annealing temperature was gradually decreased with each cycle (touchdown PCR). Cycling parameters are summarised in Table 7.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume/Reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>DreamTaq Green PCR Master Mix (2 ×) (DreamTaq DNA polymerase, 2x DreamTaq Green buffer, dNTP, 0.4 mM each, 4 mM MgCl2)</td>
<td>10 µl</td>
</tr>
<tr>
<td>dH2O</td>
<td>8 µl</td>
</tr>
<tr>
<td>Primer Forward (10 mM)</td>
<td>1 µl</td>
</tr>
<tr>
<td>Primer Reverse (10 mM)</td>
<td>1 µl</td>
</tr>
<tr>
<td>Template cDNA (50-150 ng/µl)</td>
<td>1 µl</td>
</tr>
<tr>
<td>Total Volume:</td>
<td>21 µl</td>
</tr>
</tbody>
</table>

Table 6. Volumes of the PCR reactions components

<table>
<thead>
<tr>
<th>Stage</th>
<th>Temperature</th>
<th>Time</th>
<th>Repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Denaturation</td>
<td>95ºC</td>
<td>3 min</td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>95 ºC</td>
<td>15s</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>65-60 ºC</td>
<td>30s</td>
<td>35-40 ×</td>
</tr>
<tr>
<td>Elongation</td>
<td>72 ºC</td>
<td>60s-90s</td>
<td></td>
</tr>
<tr>
<td>Final Elongation</td>
<td>72 ºC</td>
<td>10 min</td>
<td></td>
</tr>
<tr>
<td>Cooling</td>
<td>4 ºC</td>
<td>∞</td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Cycling parameters of the PCRs

3.4.2. Agarose gel electrophoresis
In order to identify the length of the products obtained in PCR reactions, electrophoretic separation was performed. Electrophoresis was carried out in 2% (w/v) agarose gel (Nu Micropore, Prona Agarose, Spain) with ethidium bromide and TAE buffer (A&A Biotechnology, Poland). The samples were loaded into gel wells mixed with a loading buffer (6 × DNA Loading Dye, VWR Life Science, USA). The gels were run for approx. 45 minutes at 120V. After separation, the gels were placed on a UV Transilluminator and photographed (BioDoc-It™ Imaging System, UVP, UK).

3.4.3. DNA purification and preparation for sequencing
To confirm the specificity of the products obtained in the PCR reaction, the bands were excised from the gel and prepared for sequencing. First, to extract DNA fragments from agarose gels, the Gel Extraction Kit (Qiagen Inc., USA) was used according to the manufacturer’s instructions. 0.4 µl of each product was further used as a template cDNA in a second round of PCRs, the other reagents and primer pairs were added as in
Subsection 3.1.3. The cycling parameters were as in Table 7, except the annealing temperature, which was 60ºC (constant) and the number of cycle repeats (20 cycles). For each sample, five repeats were performed to obtain an adequate amount of product. After the PCR, 88 µl of the total volume for each sample was processed for further purification and 10 µl was used in electrophoresis. For purification, the PCR Purification Kit (Qiagen Inc., USA) was used following the manufacturer's instructions. After purification, the samples were stored in 1.5 ml Eppendorf tubes. Sanger sequencing was performed by the DNA Sequencing and Oligonucleotide Synthesis Lab at the Institute of Biochemistry and Biophysics, PAS.

3.5. Oxford Nanopore Sequencing

In order to reconstruct the entire length of splicing variants of the Oprm1 gene, Oxford Nanopore Sequencing (ONS) was performed. The technology is a third generation sequencing that allows for obtaining long reads up to 2 Mb, which is not possible for most of available sequencing strategies, e.g. Illumina (Amarasinghe et al., 2020). In ONS, a nanometer-sized pore (nanopore) is inserted into an electrically resistant membrane created by synthetic polymers. The membrane is bathed in an electrochemical solution so that, when a potential is applied, an ionic current can be generated through the nanopore. Molecules can be identified by measuring the disruptions in the current, which are characteristic of a specific compound that enters the nanopore.

For this experiment, the RNA was prepared as described in Subsections 3.3.1 - 3.3.3. with few modifications. Brains immersed in RNAlater were cut on a vibratome VT1200 (Leica, Germany) into 200 µm sections. The entire slices containing the striatum and the prefrontal cortex (1.34-0.74 anterior to Bregma, Paxinos & Franklin, 2001) as well as the thalami isolated from the slices at -0.94 - -1.82 posterior to Bregma were selected to reverse transcription. After RNA isolation, its integrity was determined using capillary chip-based electrophoresis with an RNA 6000 Nano Methods 59 LabChip Kit and an Agilent Bioanalyzer 2100 (Agilent, USA) according to the manufacturer's instructions. The samples underwent reverse transcription and were prepared for shipment. The ONS was performed by Novogene UK using the Nanopore PromethION platform.

<table>
<thead>
<tr>
<th>Sample Id</th>
<th>ng/μl</th>
<th>260/280</th>
<th>260/230</th>
<th>RIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH_7570</td>
<td>139.68</td>
<td>1.92</td>
<td>0.38</td>
<td>7.9</td>
</tr>
<tr>
<td>TH_7566</td>
<td>128.24</td>
<td>1.91</td>
<td>0.46</td>
<td>7.8</td>
</tr>
<tr>
<td>STR_7570</td>
<td>302.7</td>
<td>1.95</td>
<td>0.59</td>
<td>8.1</td>
</tr>
<tr>
<td>STR_7566</td>
<td>339.35</td>
<td>1.99</td>
<td>0.86</td>
<td>7.9</td>
</tr>
</tbody>
</table>

3.6. RNAscope in situ hybridization

The RNAscope assay is an advanced method of ISH that allows the analysis of RNA target expression within tissues and cell cultures (Wang et al., 2012). In the experiments, the RNAscope™ Fluorescent Multiplex Assay was performed according to the manufacturer’s instructions.

The animals were killed by cervical dislocation. Brains were removed immediately from the skulls, fresh frozen on dry ice, and stored at -80°C for up to 1 month. The day before sectioning, the brains were embedded in OCT Compound (Optimal Cutting Temperature, Cell Path, UK) and then stored at -80°C. The frozen brains were sliced into 10 μm coronal sections on cryostat (CM 3050 S, Leica, Germany) with both the object and chamber temperature set at -20°C. Slices were thaw-mounted on positively charged Superfrost® plus microscope slides and stored at -80°C for up to one week. Slides representing the dorsal striatum, nucleus accumbens, and prefrontal cortex (the cingulate cortex and infralimbic cortex) were selected for ISH.

The RNAscope assay started with pre-fixing the slides with ice cold 4% paraformaldehyde (PFA), followed by dehydration at increasing concentrations of ethanol. After the dehydration, the slides were air-dried and the hydrophobic barrier was drawn. To permeabilize the tissue, Protease IV was applied for 30 minutes at room temperature. The slides were then washed in 1 x PBS and hybridized with a specific probe for 2 h at 40°C. After this step, four hybridizations with subsequent amplifiers were performed, in turns, with washing the slides in the buffer. In the final hybridization, fluorophores were attached to the targets. After hybridization, DAPI stain was applied to the slides for 30 seconds. The coverslips were mounted with ProLong Gold Antifade Mountant (Invitrogen, USA), the slides were stored at 4°C in the opaque slide box.
3.7. **Image acquisition and processing**

The image acquisition was performed one month after the ISH. For imaging a fluorescent microscope Axio Imager.Z2 (Carl Zeiss, Germany) and Plan-Apochromat 63×/1.40 Oil M27 lens were used. Laser lengths equal to 644 (At647), 553 (At550), 493 (AF488), and 353 nm (DAPI) were set to excite the fluorophores. The laser power was adjusted to the brightest sample and remained the same throughout the experiment. For each stage, 5 μm-thick Z-stacks containing 21 slices were acquired. Image processing was carried out using ZEN lite software (3.5.093.00002 version, Carl Zeiss, Germany). For each Z-stack, the maximum orthogonal projection method was applied.

3.8. **Statistical analysis**

Statistical analysis was conducted in R version 4.0.3 using stats package. In order to compare the relative expression of *Oprm1* mRNA in qPCR experiment, Kruskal-Wallis test was performed. The statistical significance level was set at 0.05.
4. Results

4.1. *Oprm1* transcript abundance in the murine striatum, spleen, and heart

The expression level of *Oprm1* was evaluated by quantitative PCR. I have evaluated the abundance of gene transcripts in the mouse striatum and compared it with expression levels in peripheral tissues, spleen, and heart. I used the Mm01188089_m1 TaqMan probe that targeted the region within exons 2 and 3. The expression of *Oprm1* in the spleen and the heart was close to the detection limit, with one sample from the heart and two samples from the spleen not crossing the cycle threshold. The results showed that the relative expression level in the striatal samples was ~1000 times higher than in the peripheral organs (Figure 6).

The discrepancies were not likely the result of technical problems, as the values of the cycle thresholds for the housekeeping gene were similar in all samples. The statistical test used to compare the relative *Oprm1* expression suggested that there were no significant differences between the organs (Kruskal-Wallis, $\chi^2=5.6092$, $p=0.06053$), though it should be noted that the sample size was small (n=3/group).

4.2. Expression of specific *Oprm1* transcripts in the murine striatum

In order to assess the expression of selected *Oprm1* transcripts RT-PCR followed by gel electrophoresis was performed. The identity of selected products was confirmed by sequencing.

Figure 6. Relative expression of the *Oprm1* gene in the murine striatum, spleen, and heart. Relative expression was assessed by ΔCt method with *Hprt1* used as a reference gene and standardization to mean values obtained in the striatum. The box depicts the interquartile range with a black thick line representing median, and whiskers ranging between minimum and maximum values.
Determination of 5'-terminal exons of *Oprm1*

First, we wanted to determine the existence of transcripts containing exon 1 or exon 11. A reaction with two primer pairs targeting these exons was performed on the material from 5 separate striatal samples obtained from adult male mice. The expected product sizes were 221 bp for the primer pair that flanked exon 1, and 90 bp for the pair that targeted exon 11. The results obtained are presented in Figure 7. Although the primers for exon 1 gave clear bands of expected size, the primers targeting exon 11 produced a weak, diffuse signal that did not allow for product identification.

![Primer sequences and electropherogram](image)

**Figure 7.** Determination of the 5'-terminal *Oprm1* exons in the striatum. A. The sequences of the primers used in the experiment with their localization in the *Oprm1* gene. The double-slanted lines represent parts of the gene that are omitted in the scheme. B. Electropherogram of the products of RT PCR. A 100 bp DNA ladder was run in the first lane from the left. Products of interest are indicated by arrows. The bands visible on the left side, at 230 bp, suggest the existence of variants containing exon 1. On the right, a weak band visible at ~100 bp in one of the samples may represent transcripts that contain exon 11.
Determination of 3’-terminal exons of Oprm1

Spatial transcriptomics sequencing revealed the existence of several potential 3’ terminations of transcripts within the Oprm1 locus. Based on these results, we decided to design reverse primers targeting regions that were not previously annotated to Oprm1 gene, at 6 860 000 and 6 986 000 bp (GRCm38.v6). To increase the reproducibility of the experiments, the previously-used forward primers were chosen. No potential amplicons of the Oprm1 gene were produced using the PR_6860kb primer (Figure 8).

![Diagram showing primers and electropherograms](image)

**Figure 8.** Evaluation of the possible exon at 6860 kbp (GRCm38). A. Sequences of the primers used in the reaction with their localization in the Oprm1 gene. The double-slanted lines represent parts of the gene that are omitted in the scheme. B. Electropherograms of the products of the RT PCR reactions. A 100 bp DNA ladder was run in the peripheral lanes. The electropherograms suggest lack of the products of expected length.

In contrast, multiple bands were produced using PR_6990kb primer (Figure 9). Although the expected product lengths were not possible to predict, the span between forward primers suggested that the products of PF3 should be 433 bases longer than PF2/3, which might be applicable to bands appearing at 950 and 510 bp. To confirm the results, the selected PCR products were purified, amplified, and sequenced.
Figure 9. Evaluation of the possible exon at 6990 kbp (GRCm38). A. Sequences of the primers used in the reaction with their localization in the Oprm1 gene. The double-slanted lines represent parts of the gene that are omitted in the scheme. B. Electropherograms of the products of the RT PCR reactions. A 100 bp DNA ladder was run in the peripheral lanes. Possible Oprm1 transcripts are indicated by arrows at 510, 630, 690 and 950 bp.

4.3. Identification of RT PCR products

As the 630 bp and 950 bp bands provided most reproducible results, they were excised from the gel, cleaned up and then amplified in a secondary PCR using the same primers as previously. Small fractions of the amplicons were separated by electrophoresis to assess the accuracy of the procedures (Figure 10). The remaining material underwent sequencing. The cDNA concentrations in the samples were as follows: 75 ng/μl for PCR1, 88 ng/μl for PCR2, and 132 ng/μl for PCR3.

Figure 10. Evaluation of PCR clean-up and amplification procedures in samples prepared for Sanger sequencing. A 100 bp DNA ladder was run in the 1 lane from the left. Products of interest are indicated by arrows. The second lane corresponds to samples possessed from 630 bp bands of the previous experiment, and the other two lanes correspond to 950 bp bands. The discrepancy in length between the third and fourth lane stems from different forward primers used in the samples.

Sequences were aligned to the genome using the BLAT tool (University of California, Santa Cruz). In general, the results confirmed that the excised bands matched the specific
regions within the *Oprm1* gene. The sequences showed a relatively high alignment score and identity with the genomic sequence (≥ 96.0%) (Table 9). It is worth to note that the alignment did not include the terminal regions of the query sequences, due to the high noise level in these regions.

<table>
<thead>
<tr>
<th>Query sequence</th>
<th>Query length</th>
<th>Alignment score</th>
<th>Query cover</th>
<th>Sequence identity</th>
<th>Chromosome</th>
<th>Genome start</th>
<th>Genome end</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR1-PF3</td>
<td>606</td>
<td>572</td>
<td>17-592</td>
<td>100.0%</td>
<td>chr10</td>
<td>6830444</td>
<td>6985996</td>
</tr>
<tr>
<td>PCR1-PR_6990kb</td>
<td>602</td>
<td>578</td>
<td>8-591</td>
<td>99.9%</td>
<td>chr10</td>
<td>6830411</td>
<td>6985971</td>
</tr>
<tr>
<td>PCR2-PF3</td>
<td>589</td>
<td>429</td>
<td>16-487</td>
<td>96.0%</td>
<td>chr10</td>
<td>6830444</td>
<td>6986007</td>
</tr>
<tr>
<td>PCR2-PR_6990kb</td>
<td>883</td>
<td>435</td>
<td>10-479</td>
<td>96.6%</td>
<td>chr10</td>
<td>6830411</td>
<td>6985971</td>
</tr>
<tr>
<td>PCR3-PF2/3</td>
<td>920</td>
<td>862</td>
<td>25-899</td>
<td>99.5%</td>
<td>chr10</td>
<td>6830017</td>
<td>6985988</td>
</tr>
<tr>
<td>PCR3-PR_6990kb</td>
<td>918</td>
<td>814</td>
<td>17-897</td>
<td>96.4%</td>
<td>chr10</td>
<td>6829988</td>
<td>6985965</td>
</tr>
</tbody>
</table>

Table 9. BLAT alignments of sequenced RT PCR samples. 'Alignment score' represents the number of nucleotides matched to the genome with a penalty for mismatches and gaps. 'Query cover' describes the location of the beginning and end of the alignment in the query sequence. 'Sequence identity' indicated the number of nucleotides and gaps in percent. 'Genome start' and 'Genome end' represent the beginning and end of the alignment in the genome (GRCm38.v6).

BLAT analysis showed that all sequenced fragments contained exon 3, 7, and part of exon 19, the PCR1 sample also contained exon 6. Reconstruction of the 'partial' transcripts is presented in Figure 11, while alignment with the genome and the whole sequences are shown in Appendix. As reads that were generated from forward primers (assigned as PCR_-PF_) had better quality, they were used for further analysis.

![Figure 11](image.png)

Figure 11. Schematic presentation of alignments for the sequenced PCR samples. On the top, *Oprm1* scheme is presented. On the bottom, the products of the PCR sequencing are presented, and recognized exons are marked with vertical lines.
In order to further characterize the transcripts, results were analysed with the blastn tool (National Center for Biotechnology Information) using the reference RNA sequences database (refseq_RNA). Sequences were matched with multiple transcripts, with the highest query coverage for MOR-1U, MOR-1M, and MOR-1C (Table 10). Detailed analysis revealed common exon 3, exon 7 between the samples and the aforementioned variants. A part of the sequenced exon 19 was identical to MOR-1U. After comparison with the literature, PCR1 was identified as MOR-1Eii or MOR-1Eiv. PCR2 and PCR3 showed the greatest similarity to MOR-1U; however, their exon 19 was longer (Doyle et al., 2007).

<table>
<thead>
<tr>
<th>Query sequence</th>
<th>Transcript variant</th>
<th>NCBI accession</th>
<th>Query cover (%)</th>
<th>Query cover (bp)</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR1</td>
<td>MOR-1U</td>
<td>NM_001304955.1</td>
<td>40%</td>
<td>17-69, 188-381</td>
<td>97.49%</td>
</tr>
<tr>
<td></td>
<td>MOR-1M</td>
<td>NM_001302795.1</td>
<td>23%</td>
<td>17-69, 188-279</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>MOR-1C</td>
<td>NM_001039652.2</td>
<td>23%</td>
<td>17-69, 188-279</td>
<td>100%</td>
</tr>
<tr>
<td>PCR2</td>
<td>MOR-1U</td>
<td>NM_001304955.1</td>
<td>44%</td>
<td>2-265</td>
<td>94.05%</td>
</tr>
<tr>
<td></td>
<td>MOR-1M</td>
<td>NM_001302795.1</td>
<td>26%</td>
<td>2-157</td>
<td>96.27%</td>
</tr>
<tr>
<td></td>
<td>MOR-1C</td>
<td>NM_001039652.2</td>
<td>26%</td>
<td>2-157</td>
<td>96.27%</td>
</tr>
<tr>
<td>PCR3</td>
<td>MOR-1U</td>
<td>NM_001304955.1</td>
<td>73%</td>
<td>25-696</td>
<td>98.82%</td>
</tr>
<tr>
<td></td>
<td>MOR-1M</td>
<td>NM_001302795.1</td>
<td>61%</td>
<td>25-594</td>
<td>99.82%</td>
</tr>
<tr>
<td></td>
<td>MOR-1C</td>
<td>NM_001039652.2</td>
<td>61%</td>
<td>25-594</td>
<td>99.82%</td>
</tr>
</tbody>
</table>

Table 10. Comparison of RT-PCR sequences with existing transcripts.

Next, putative protein products were predicted using Expasy Translate Tool (SIB Swiss Institute of Bioinformatics) (Figure 12). The analysis revealed the presence of putative STOP codons in exons 6 and 19. PCR1 encoded 15 amino acids downstream exon 3, while PCR2 and PCR3 – 44 amino acids. A non-synonymous A125C mutation was observed in the PCR3 sample. The substitution was an artefact and did not appear in any other sample nor the reference genome.
Figure 12. Partial sequences of PCR samples (downstream of exon 3) with their translation. To simplify the view, only sequences downstream exon 3 are included. Vertical lines mark exon junctions, and numbers indicate exons. The amino acid sequences are underlined in pink. STOP codons are marked in red.

Finally, I compared the putative amino-acid sequences with protein isoforms of mouse and human orthologues of Oprm1 listed in the UniProt database (Figure 13). The alignment was prepared using ClustalW2 (EMBL’s European Bioinformatics Institute) with the “Gap Opening Penalty” set at 90.00 and the “Delay divergent sequences” at 0%. The analysis revealed that the predicted amino acid sequence of PCR1 was identical with mMOR-1E, mMOR-1Eiii and mMOR-1Eiv, while the PCR2 and PCR3 appeared to be partially convergent with mMOR-1O, mMOR-1C, mMOR-1M, mMOR-1U, and hMOR-1C. Despite high similarity, the putative PCR1 and PCR2 protein sequences differ significantly in the carboxy terminus from already reported isoforms. This shows that if
the spotted variant encodes a functional protein, it produce an unique isoform of μ opioid receptor.

PRC1

mMOR-1E

mMOR-1Eii

mMOR-1Eiv

mMOR-1P

mMOR-1B5

mMOR-1F

mMOR-1Y

mMOR-1V

hP35372-14

PCR2

PCR3

mMOR-1O

mMOR-1C

mMOR-1M

mMOR-1U

hMOR-1C

hMOR-1B3

hMOR-1A

hMOR-1W

hMOR-1B4

hMOR-1R

hMOR-1B2

hMOR-1B

hMOR-1B2

hMOR-1B5

hMOR-1

mMOR-1

mMOR-1H

mMOR-1J

mMOR-1T

hMOR-1R

hMOR-1X

mMOR-1D

mMOR-1N

PRC1
Figure 13. Comparison of human - mouse protein sequences for μ opioid receptor. Human isoforms are assigned with the prefix ‘h’, while mouse isoforms are assigned with the ‘m’. Translation of the sequenced samples was presented in red.

4.4. **Long-read transcriptome sequencing**

In order to perform an unbiased analysis of *Oprm1* transcripts, I have used Oxford Nanopore Sequencing (ONS). ONS is one of the third-generation sequencing methods. The technology enables sequencing of long DNA or unamplified RNA molecules. The ONS utilizes transmembrane proteins – nanopores, that are introduced into a synthetic membrane characterized by high electrical resistance. When an electric current is applied on both sides of the membrane, it creates a potential difference that attracts negatively charged nucleic acid polymers. Nucleotides passing through the nanopores create changes in the electric potential between both sides of the membrane. Such a variation is specific to each type of nucleotide and therefore can be recognized by the sequencing machine. In the experiment, samples from two regions of the brain were analysed, the whole slices containing the anterior striatum and prefrontal cortex (1.34-0.74 anterior to Bregma) and the isolated thalamus (at the -0.94 - -1.82 posterior to Bregma). Sequencing was performed using the PromethION platform, which is a high-throughput system. Each
sample was sequenced three times. The raw statistics that describe the quality and characteristics of the procedure are presented in Table 11. The number of sample bases varied between 5.2 and 11.5 Gbp per sample, while the number of reads fluctuated between 3.8 and $8.2 \times 10^6$. The median read length in all samples was relatively consistent and oscillated between 1 002 and 1 165 bp. The N50 read length, so the length of the shortest read in the subset of the longest sequences that together represent $\geq 50\%$ of the nucleotides in the sample varied from 1 875 to 2 172 bp with higher values obtained in the samples containing the striatum and prefrontal cortex than those from the thalamus (Table 11). The median read quality was high and consistent between samples, reaching values between 12.1-13.0. The percentage of reads with relatively low quality (<7) did not exceed 3.5% in any of the samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample Bases (bp)</th>
<th>Number of reads</th>
<th>Median read length</th>
<th>Read length N50</th>
<th>Median read quality</th>
<th>&gt;Q7 number(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH_7566</td>
<td>11 489 387 734.0</td>
<td>8 233 012.0</td>
<td>1 062.0</td>
<td>1 957.0</td>
<td>12.2</td>
<td>7987049(97.0%)</td>
</tr>
<tr>
<td>TH_7566</td>
<td>9 993 495 631.0</td>
<td>6 851 893.0</td>
<td>1 117.0</td>
<td>2 055.0</td>
<td>13.0</td>
<td>6633785(96.8%)</td>
</tr>
<tr>
<td>TH_7566</td>
<td>11 489 387 734.0</td>
<td>8 233 012.0</td>
<td>1 062.0</td>
<td>1 957.0</td>
<td>12.2</td>
<td>7987049(97.0%)</td>
</tr>
<tr>
<td>TH_7570</td>
<td>6 057 697 607.0</td>
<td>4 577 351.0</td>
<td>1 002.0</td>
<td>1 875.0</td>
<td>12.1</td>
<td>4436861(96.9%)</td>
</tr>
<tr>
<td>TH_7570</td>
<td>5 218 273 416.0</td>
<td>3 785 962.0</td>
<td>1 047.0</td>
<td>1 955.0</td>
<td>12.9</td>
<td>3658243(96.6%)</td>
</tr>
<tr>
<td>TH_7570</td>
<td>6 057 697 607.0</td>
<td>4 577 351.0</td>
<td>1 002.0</td>
<td>1 875.0</td>
<td>12.1</td>
<td>4436861(96.9%)</td>
</tr>
<tr>
<td>STR_7566</td>
<td>6 057 697 607.0</td>
<td>4 750 245.0</td>
<td>1 019.0</td>
<td>1 972.0</td>
<td>12.1</td>
<td>4598738(96.8%)</td>
</tr>
<tr>
<td>STR_7566</td>
<td>6 535 407 003.0</td>
<td>3 933 810.0</td>
<td>1 067.0</td>
<td>2 072.0</td>
<td>12.9</td>
<td>3797426(96.5%)</td>
</tr>
<tr>
<td>STR_7566</td>
<td>5 649 055 794.0</td>
<td>4 750 245.0</td>
<td>1 019.0</td>
<td>1 972.0</td>
<td>12.1</td>
<td>4598738(96.8%)</td>
</tr>
<tr>
<td>STR_7570</td>
<td>6 535 407 003.0</td>
<td>6 142 736.0</td>
<td>1 106.0</td>
<td>2 068.0</td>
<td>12.1</td>
<td>5970938(97.2%)</td>
</tr>
<tr>
<td>STR_7570</td>
<td>8 968 497 314.0</td>
<td>5 110 330.0</td>
<td>1 165.0</td>
<td>2 172.0</td>
<td>12.9</td>
<td>4951231(96.9%)</td>
</tr>
<tr>
<td>STR_7570</td>
<td>8 968 497 314.0</td>
<td>6 142 736.0</td>
<td>1 106.0</td>
<td>2 068.0</td>
<td>12.1</td>
<td>5970938(97.2%)</td>
</tr>
<tr>
<td>TH_mean</td>
<td>8 384 323 288.0</td>
<td>6 043 096.8</td>
<td>1048.7</td>
<td>1945.7</td>
<td>12.4</td>
<td>5856641(96.9%)</td>
</tr>
<tr>
<td>STR_mean</td>
<td>7 119 093 672.5</td>
<td>5 138 350.3</td>
<td>1 080.3</td>
<td>2 054.0</td>
<td>12.4</td>
<td>4981335(96.9%)</td>
</tr>
</tbody>
</table>

Table 11. MinION Quality Control statistics for all reads. Read length and the average quality for all samples, each sample was sequenced 3 times. The means and t tests for each parameter were presented in the bottom rows. Samples containing the striatum and the prefrontal cortex are annotated as STR_, and samples containing the thalamus are annotated as TH_.

36
All FASTQ files for each brain region were merged and the data were aligned to the GRCm38_102/mm10 genome with the Minimap2 program (v. 2.24) by Jacek Hajto and Dr. Marcin Piechota from the Laboratory of Pharmacogenomics, Maj Institute of Pharmacology, PAS. Alignment results were filtered for strand mismatches and invalid 3' ends. The final number of valid alignments was ~35 million.

The output files were visualized using the Integrative Genome Viewer v2.12.3 (IGV; Thorvaldsdóttir et al., 2013; Figure 14). Initially, only the reads that contained exons 2 and 3 were classified as *Oprm1* transcripts. The analysis revealed that the vast majority of the transcripts (25 reads in the STR_ samples and 72 in the TH_ samples) contained exons 1, 2, 3, and 4. Four transcripts were observed that contained exon 7 instead of exon 4 (one in the striatum and prefrontal cortex and three in the thalamus). Three of them expanded through exon 8 to exon 9 and one terminated in exon 7. Additionally, in the thalamus, a variant with exon 6 was detected. No *Oprm1* transcripts comprised of exon 11 were observed.

In addition to the reads aligned to both exons 2 and 3, and defined as the *Oprm1* transcripts, multiple ‘shorter’ reads were detected. More than 100 reads of 150-250 kb length were reported in the region of exon 9 in both of the analysed areas. Notably, between 8850-8860 kbp, in the region adjacent to exon 4, a long contig was observed. The reads formed a shape described here as a ‘cat with a tail’. The genome coverage of the contig was similar to those observed in exons 2 and 3, except the region between 6 851 700 and 6 852 700 bp, where the number of reads was five times higher. Due to the localization in genome and high coverage with reads, the ‘cat with a tail’ region was hypothesised to be a part of the *Oprm1* gene. To verify the hypothesis, RNAscope ISH with probe targeting the region was performed.
Figure 14. Representation of BAM files for the *Oprm1* gene in the striatum and the prefrontal cortex (upper panel) and in the thalamus (lower panel), as depicted in IGV. The overall coverage and reads are displayed as independent tracks. The top track represents the genome *locus*, the lowest track represents the reference *Oprm1* sequence. Small nucleotide deletions are marked in purple, clipped parts of the sequences are marked with red lines and angle brackets. In the picture, 4 exons (from 1 to 4) are captured. A 10 kb-long fragment of ‘a cat with a tail’ was marked in red.
4.5. Visualization of the Oprml 3’UTR in RNAscope

The method is characterized by high sensitivity and specificity, which is provided by characteristic probe design and signal amplification strategy. Typically, for each target RNA, about 20 pairs of Z-shaped probes are designed that specifically hybridize with the target molecule. Both probes from single ZZ pair have to hybridize to the target sequence in order for signal amplification to occur. Hybridized ZZ probes pairs can bind a preamplifier which possesses numerous binding sites for amplifiers; such design enables for visualization of single RNA molecules and provides a high signal-to-noise ratio.

The aim of the experiment was to observe whether region spotted in the ONS experiment is a part of Oprml transcripts and to compare the distribution of mRNA targeted by a standard probe (Mm-Oprm1-C3) with a custom-made probe (Mm-Oprm1-O7) hybridizing to the region between 6 858 476 and 6 861 848 bp (GRCm38.v6). Additionally, expression of the proenkephalin gene was assessed as a marker of the striato-pallidal medium spiny neurons. Three areas of the brain were analysed, the prefrontal cortex (the infralimbic cortex together with the cingulate cortex), the dorsal striatum and the nucleus accumbens core (Figure 15B, 16A) experiment revealed robust proenkephalin expression in many of the dorsostriatal and accumbal cells but not in the prefrontal cortex. There were also characteristic signal dots from the custom-made probe observed on slices. Unfortunately, the standard probe fluorescence exhibited low signal-to-noise ratio, what precluded localization of Oprml transcripts (Figure 15C, D). Even though localization of Oprml with a standard probe was not possible, the characteristics of the signal produced by the custom-made probe may suggest targeting the receptor’s transcripts. The signal from the probe enveloped the nuclei, with lower expression in the background. The Oprml signal was observed in both Penk-positive and Penk-negative cells within the dorsal and ventral striatum. The expression in the cortex was rather low in the outer layers, and slightly higher within inner layers (Figure 16B).
Figure 15. *Oprm1* transcript detection in the Dorsal Striatum and the Nucleus Accumbens as assessed by RNAscope ISH. A. Region targeted by the *Oprm1*-O7 customized probe. B. Reference section 1.34 mm from bregma, based on *The Mouse Brain in Stereotaxic Coordinates* (Paxinos & Franklin, 2001), area matching the regions presented on the photos delineated in red. C, D. Maximum intensity projections for *Penk* (red), *Oprm1*-O7 (orange), *Oprm1* (green) in the Dorsal Striatum (C) and the Nucleus Accumbens (D), with corresponding photos for each transcript shown separately. *Oprm1* mRNA-expressing cells (white arrowheads), *Penk/Oprm1* mRNA expressing cells (yellow arrowheads). Scale bars 10 μ.
Figure 16. 

A. 

Oprm1 transcript detection in the Cortex as assessed by RNAscope ISH. A. Reference section 1.34 mm from bregma, based on The Mouse Brain in Stereotaxic Coordinates (Paxinos & Franklin, 2001), area matching the region presented on the photos delineated in red. 

B. 

Maximum intensity projections for Penk (red), Oprm1-O7 (orange), Oprm1 (green) merged photos from the intersection of prefrontal cortex. Scale bar 20 μm.
5. Graphical summary

Figure 17. Graphical summary of the experiments. In the upper row - schematic visualization of the *Oprm1* gene (based on Ensembl.org). Summary of the results obtained from Spatial Transcriptomics in orange, peaks of reads are indicated by arrows. The ONS data from the striatum and the prefrontal cortex samples are indicated in red, the distal ends of the reads are indicated by arrows, long 3’UTR is labelled by as ‘cat with a tail’. PCR product sequencing sources: ensembl.org, GRCm38.v6, nov2020 Doyle et al., 2007 Pasternak and Pan, 2013
visualization is presented in grey. Exons are indicated with vertical lines or rectangles. Sources: (Ensembl.org archive (GRCm38.v6, nov2020; Doyle et al., 2007; Pasternak & Pan, 2013)
6. Discussion

In this thesis I aimed to determine the presence of various transcripts of the Oprm1 gene in the striatum and the prefrontal cortex of adult mice. The RT-PCR and ONS experiments enabled me for a detailed analysis of the Oprm1 gene structure and provided evidence for the expression of specific transcript variants of the μ opioid receptor. Next, I tried to confirm the hypothesis of the long Oprm1 transcript in the forebrain using RNAscope ISH. The results obtained are preliminary and further experiments should permit evaluating the specificity of distribution of the Oprm1 transcripts.

Expression of Oprm1 exon 4 and the long 3’UTR

The ONS results strongly suggest the existence of a long 3’UTR downstream exon 4. The observed contig expands beyond 10 kbp, and has not been mapped so far. The postulated length of exon 4 in the Oprm1 transcripts varies between 53 and 808 bp, as reported in the Ensembl database. The length does not affect the produced protein size since the open reading frame is conserved, and a STOP codon occurs 39 bp downstream the beginning of the exon. There are several transcript variants reported that contain exon 4, among them those producing the canonical MOR-1 isoform. In the ONS experiment, the 5’ end of exon 4 is clearly visible and converges with the Ensembl data. The 3’ terminus, however, seems more diverse. There are multiple reads that contain exons 1 to 4 and then gradually diminish in counts between 6 850 050 kb and 6 851 400 kb. Some of them may be recognized as the reported variants Oprm1-215, Oprm1-208, and Oprm1-203. However, an early termination of exon 4 is not consistent with the data provided by Spatial Transcriptomics, where peaks of reads indicating 3’ends of mRNAs appear at 6 852 900 bp (40 reads) and at 6 860 000 bp (314 reads). Interestingly, despite the fact that full-length Oprm1 transcripts terminate between 6 850 - 6 851 kbp, the high coverage of reads generated by the ONS expands up to 6 860 050 bp, creating specific contig described here as a ‘cat with a tail’. The distal end of the contig matches the Spatial Transcriptomics results and is also consistent with a single-cell transcriptomics data acquired from neurons in the Tabula Muris Project (Schaum et al., 2018, Figure 18). Therefore, it is hypothesised that the visualized contig comprises for a 3’UTR of Oprm1 transcripts that possess exons 1-4.
Several factors may contribute to the lack of ‘full’ Oprm1 reads within the contig. First, the region analysed contains several repetitive elements, including one long-tandem repeat (LTR), three short interspersed nuclear elements (SINE; one mammalian-wide interspersed repeat and two Alu sequences), and 8 simple repeats of various sequences, as indicated by the RepeatMasker Open 4.0 tool (Smit et al., 2013). Both, LTR and SINE are transposable elements that can contribute to improper genome mapping and ‘breakage’ of a continuous sequence assembly (Tørresen et al., 2019). Additionally, even though, the long-read sequencing techniques should be able to span the repetitive sequences, they generate a relatively high error rate, caused mainly by frequent indels (insertion and deletions), which can negatively affect the correctness of genome alignment (Amarasinghe et al., 2020; Tørresen et al., 2019).

It should be noted that the discontinuity of reads in the ONS experiment was also observed, to some extent, in other genes, e.g. Oprk1 and Oprd1. Although the genes were confirmed to contain 3 347 bp and 2 801 bp long 3 'UTRs, in both cases the last exons were not equally covered with reads. This may indicate that numerous transcripts were not mapped or sequenced as one read in the current experiment. While the ONS is known to generate 10-30 kb genomic libraries (with the longest reported length of 2.3 Mb), in our experiment only 4 826 out of ~68 million raw reads exceeded the length of 10 kbp. The possible explanation for the low long-read abundance includes fragmentation of the transcripts during the library preparation or nanopore blocking. The ONS is based on a protocol in which transcripts are sequenced towards the 5’ end of the nucleic acid. This results in generation of truncated reads if blockade of the nanopore appears. Truncated
reads may also result from a coverage bias towards the 3 'end' observed in this type of sequencing (Amarasinghe et al., 2020). During library preparation, transcription is carried out by a reverse transcriptase, which displays 3'-5' exonuclease activity. The enzyme types vary in their specifications and efficiency. Moreover, transcriptases display limited capability to deal with secondary structure mRNA, and stop their activity while recognizing such structure. Omniscript reverse transcriptase used to prepare the library in my experiment was found to have the highest transcription efficiency for mRNA fragments up to 1.2 kb (Martín-Alonso et al., 2021). Therefore, to improve the quality of cDNA, it would be beneficial to use enzyme that is dedicated to longer fragments and with higher detection sensitivity. An additional factor that may contribute to a vague picture of the ‘cat with a tail’ contig is a phenomenon of internal poli(A)s. It was shown that more than one poli(A) sequence can be included into a single mRNA molecule. During the library preparation, oligo(dT) primers used for transcription of mRNAs into cDNA can hybridize with the internal poli(A) and contribute to generation of truncated sequences (Nam et al., 2002). The phenomenon of internal poli(A) may be associated with the existence of alternative 3’UTRs, which were observed in neurons and will be described in the last section of the discussion.

Although the ONS results provide a massive amount of data, the aforementioned factors may account for a blurred overview of the real structure of the Oprm1 gene. To finally confirm the ONS observations another suitable mRNA assay should be applied.

**Confirmation of the long Oprm1 3’UTR in RNAscope ISH**

To confirm the existence of the long 3’UTR, RNAscope ISH was performed. The main aim of the experiment was to colocalize the standard Oprm1 probe with the custom-made probe targeting putative 3’UTR. The results did not provide satisfactory colocalization evidence, as the standard Oprm1 probe produced a very low signal-to-noise ratio, impossible to distinguish by manual or automatic approaches. This may be partially explained by the properties of the fluorophore paired with the probe. The 3 'UTR targeting probe was conjugated with Alexa488, which is known to produce relatively high autofluorescence (Abuduwali et al., 2013). However, it cannot explain the complete lack of puncta. The other factor that can affect the results is the much longer time elapsed between the RNAscope assay and slides imaging than suggested (RNAscope Fluorescent Multiplex Reagent Kit Part 2, 2019).
The signal from the new, 3’UTR-targeting probe has nuclear localization, suggesting that the probe properly targets mRNAs. Furthermore, some characteristics of Oprml expression were detected. Firstly, the puncta were observed to be more abundant in the dorsal striatum and the nucleus accumbens than in the cortical regions; secondly, the signal was present in both Penk-positive and Penk-negative cells, which is consistent with the literature on Oprml gene expression (Banghart et al., 2015; Mansour et al., 1995; Xu et al., 2014). To unequivocally confirm the existence of a long Oprml 3’UTR and analyse its distribution, it is necessary to repeat the RNAscope experiments with another set of fluorophores, following the suggested image acquisition time.

Expression of Oprml exons 7-9

It should be noted that the discontinuity of reads does not neccessarily affect the ‘cat with a tail’ region only. In the ONS experiment few 'full' transcripts that contain exons 7-9 were observed. This may be surprising, as the variants with exon 7 were reported to account for a considerable fraction of transcribed Oprml mRNA in the analysed brain regions (Xu et al., 2014). Although few ‘full’ transcripts were observed, there were many ‘short’ reads aligned to exon 9 in the ONS. Also, the Spatial Transcriptomics data show a robust genome coverage in this region (771 reads at 7 038 800 bp). This suggests a high abundance of Oprml transcripts that possess exon 9.

In addition, PCR reactions provide evidence for transcription of several variants that expand beyond exon 4 and contain either exon 7 or the preceding exon 6. Noteworthily, the PCR revealed an unique transcriptional variant related to MOR-1U, but with a longer exon 19. The variant was not reported so far in the NCBI or Ensembl databases. If the revealed variant is transcribed into functional protein, then the carboxy terminal domain of the receptor would differ considerably from those present in the previously reported isoform, which, in turn, may affect its properties, for example, phosphohorylation sites (Narayan et al., 2021). A caveat of this interpretation is that only ‘partial’ cDNA was sequenced. To assess the entire transcript sequence, the use of the RACE PCR (Rapid amplification of cDNA ends PCR) method would be beneficial.

Expression of Oprml 5’ terminal exons

According to the literature, exon 11 is predominantly expressed in transcripts encoding 6-TM isoforms – MOR-1G, MOR-1K, MOR-1L, MOR-1M and MOR-1N. Previous research showed robust exon 11-like immunoreactivity in the murine striatum (Abbadie et al., 2004), whereas mRNA levels for these isoforms in the striatum and
prefrontal cortex were scant (Xu et al., 2014). Here, neither the RT-PCR nor the ONS method provided evidence for exon 11 transcription in any of the regions examined. It may be due to the insufficient sensitivity of the methods used. However, the results do not exclude the expression of such transcripts in other regions of the central nervous system, e.g., the perigluecular gray or the dorsal horn of a spinal cord.

**Putative role of long 3’UTR in neuronal transcripts**

In recent years, thousands of genes in both rodents and humans have been updated to have novel distant 3’UTRs. Substantially longer 3’UTRs were observed in neurons. Many of such transcripts were shown to possess >10kbp 3’UTR, and some of them exceeded >18kbp. The function of such a long untranslated region has not been fully established yet. It is suggested that the region contains sequences responsible for regulation of gene expression, stability and cellular localization of the mRNAs (Bae & Miura, 2020). For instance, research on slides from rodent hippocampus showed that the 3’UTRs found in the neuropil influenced longer mRNAs half-lives than the 3’UTRs in the somatic compartment (Tushev et al., 2018). Conversely, a knock-out of this region in the mTOR gene significantly limited the distribution of mRNAs to specific neuronal compartments (Terenzio et al., 2018).

Neuronal transcripts commonly contain several alternate 3’UTRs. Diversity is caused by alternative polyadenylation sites used during the generation of mRNAs. Up to 70% of protein-coding genes are estimated to have multiple poly(A) sites. Multiple poly(A)s may be associated with alternative splicing and variability in the last exon included in the transcript. However, in most events, alternative polyadenylation occurs in variants with the same open reading frame, but different 3’ UTRs, the so-called ‘tandem’ 3’UTRs (Bae & Miura, 2020). ‘Tandem’ 3’UTRs were shown to generate mRNAs of different lengths, with a particular preference for longer transcripts observed in the central nervous system (Miura et al., 2013).

Such long 3’UTRs usually possess several microRNAs (miRNAs) binding sites, which contribute to a more precise regulation of gene expression, especially important in neurons. If the ‘cat with a tail’ is included in *Oprm1* transcripts, then the expression of μ opioid receptor can be modulated by miRNA targeting the region. According to the miRNA database, mirbase.org (Griffiths-Jones et al., 2006), the region contains two possible miRNA binding sites, targeted by mmu-mir-3470b and mmu-mir-3470a. The
miRNAs were identified in the ovary of newborn mice; their existence in the central nervous system was not evaluated (Ahn et al., 2010).

**Conclusion**

To conclude, my results provide evidence for a long 3'UTR downstream exon 4 in the murine *Oprm1* gene, which according to the data obtained in the ONS experiment represents for a considerable fraction of the transcribed μ opioid receptor. Furthermore, the data enabled me to determine the existence of splicing variants containing exon 19, including 'partial' cDNA of a form that was not previously reported. Finally, the results show a predominant expression of mRNAs containing exon 1, and no evidence of exon 11 transcription in the striatum and the prefrontal cortex. The data obtained so far confirm a highly complex architecture of the *Oprm1* gene. In future research, I would like to focus on assessing the cellular distribution of the long *Oprm1* 3’UTR in the brain, by continuing the experiments utilizing RNAscope assay. As long as the significance of 3’UTR sequences in the μ opioid receptor has not been fully elucidated yet, I would also like to investigate the putative role of the sequence in the context of the receptor function and potential role in its pathologies.
List of figures

Figure 1. Endogenous opioid peptides and their receptors.................................................. 7
Figure 2. μ opioid receptor signalling pathways after activation by agonist.......................... 8
Figure 3. Schematic representation of the mouse Oprml gene and reported transcription variants................................................................. 11
Figure 4. A. The visualization of Oprml transcripts in spatial transcriptomics......................... 14
Figure 5 Polyadenylation sites of the Oprml gene in the striatum and the prefrontal cortex revealed by the spatial transcriptomics method................................................................. 14
Figure 6. Relative expression of the Oprml gene in the murine striatum, spleen, and heart...... 27
Figure 7. Determination of the 5'-terminal Oprml exons in the striatum.............................. 28
Figure 8. Evaluation of the possible exon at 6860 kbp (GRCm38). ........................................ 29
Figure 9. Evaluation of the possible exon at 6990 kbp (GRCm38). ........................................ 30
Figure 10. Evaluation of PCR clean-up and amplification procedures in samples prepared for Sanger sequencing........................................................................................................... 30
Figure 11. Schematic presentation of alignments for the sequenced PCR samples............... 31
Figure 12. Partial sequences of PCR samples (downstream of exon 3) with their translation.... 33
Figure 13. Comparison of human - mouse protein sequences for μ opioid receptor.............. 35
Figure 14. Representation of BAM files for the Oprml gene in the striatum and the prefrontal cortex and in the thalamus, as depicted in IGV................................................................. 38
Figure 15. Oprml transcript detection in the Dorsal Striatum and the Nucleus Accumbens as assessed by RNAscope ISH ................................................................. 40
Figure 16. Oprml transcript detection in the Cortex as assessed by RNAscope ISH.............. 41
Figure 17. Graphical summary....................................................................................... 42
Figure 18. Visualization of Tabula Muris single cell sequencing on neuronal cells at 10:6 850 000 – 10:6 860 000 bp of the mouse genome (GRCm38). ................................................................. 45
List of tables

Table 1. Number, strain (genotype), sex, and age of the animals used for specific experiments 19

Table 2. mRNA concentrations (ng/μl) in the samples from the murine striata, hearts and spleens after reverse transcription ................................................................. 20

Table 3. Volumes of the reverse-transcription reaction components ............................................ 21

Table 4. Volumes of the qPCR reaction components .................................................................. 22

Table 5. Cycling parameters of the qPCR .................................................................................. 22

Table 6. Volumes of the PCR reactions components ................................................................. 23

Table 7. Cycling parameters of the PCRs .................................................................................. 23

Table 8 Results of capillary chip-based electrophoresis for the mRNA samples for the ONS experiment ............................................................................................................. 25

Table 9. BLAT alignments of sequenced RT PCR samples. ....................................................... 31

Table 10. Comparison of RT-PCR sequences with existing transcripts. ................................... 32

Table 11. MinION Quality Control statistics for all reads .......................................................... 36

References


Appendix

Raw data obtained in PCR product sequencing

>PCR1-PR_6990kb
GGACTCCGAGAAGGAGGCCCTCCTACCTCATCTCAGGATGGATGGGGCGCGATCTCAGGACATTCCCATCACCGGCTATTCCGAT
AAATGTTAATTTTGTGGATGCTGTGGAAAAGGGGGGCAA

>PCR2-PR_6990kb
CCTGCCATCTCGATCGTCAACACTAGGGACACCCCTCCACGGCTAATACAGTGGATCGAACTAACCCAGAAGAAAAAGCTGGAC
TCAGAAGGAGGAGGAGAACAAATATCCTACTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGC
AAGCATGTTGAAAAACCCTGCAAGAGTTGCATGGACAGACTGCTACAACAGAAAACAACAGACTGGGTCACTCAGAAAAATAA
AAAAGAAGAAAGAGGAGGAAGAACAAATATCCTAGAAGCTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTC
CTGACTTCTGTTGGGTTGTTTGCTGTTCGATGGGGGAAGAAA

>PCR3-PR_6990kb
AAACAGGCAAATGCGTGATGCTGTTCGAGATGTGTTGAGAGGTCATAGGTCATGAGGGTGGCGTCCTCATCGATGGCAGTTGTGC
CTTTATCTGGTGAACAGGAATTTCCTATTTGCCATGTGAGGGTTATATTAAGATCATAGATGTCTGCAAGTCAGGAAGAGGGTC
ATCCTTCATTAACTAATCCTCAATGCACCAAAAATATAAAAGAAAAGAAAGAGGAGGAAGAACAAATATCCTATAAGCTGGAATAT
CTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGCCTTCTTCTTCTTTTTATTTTTTCTGAGTGACCCAGTCT
GGTTTTTCTGTTGTAGCAGTCTGTCCATGCAACTCTTGCAGGGTTTATTTCAACATGAGTCGGAAAAGGATATCCTGTAAAGATC
TGGGCCACCCTGACGCTGACTGCCAGGGTTGGGTGGTTAGTTGGAAATTTTTCCCCCCCCCCCCCTGGTTTGTCATTATTACAAAT
TAGGAGAAGTGATCAATTTCCTTTCTCGCACATGATTAATATTTATTATAGGTCAGCATCATCCGGGTATCGATGGAAACACATCC
ACATTATATCCATTCGAGTCGATCA

}>PCR1-PCR_6990kb
GGACTCCGAGAAGGAGGCCCTCCTACCTCATCTCAGGATGGATGGGGCGCGATCTCAGGACATTCCCATCACCGGCTATTCCGAT
AAATGTTAATTTTGTGGATGCTGTGGAAAAGGGGGGCAA

}>PCR2-PCR_6990kb
CCTGCCATCTCGATCGTCAACACTAGGGACACCCCTCCACGGCTAATACAGTGGATCGAACTAACCCAGAAGAAAAAGCTGGAC
TCAGAAGGAGGAGGAGAACAAATATCCTACTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGC
AAGCATGTTGAAAAACCCTGCAAGAGTTGCATGGACAGACTGCTACAACAGAAAACAACAGACTGGGTCACTCAGAAAAATAA
AAAAGAAGAAAGAGGAGGAAGAACAAATATCCTAGAAGCTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTC
CTGACTTCTGTTGGGTTGTTTGCTGTTCGATGGGGGAAGAAA

}>PCR3-PCR_6990kb
AAACAGGCAAATGCGTGATGCTGTTCGAGATGTGTTGAGAGGTCATAGGTCATGAGGGTGGCGTCCTCATCGATGGCAGTTGTGC
CTTTATCTGGTGAACAGGAATTTCCTATTTGCCATGTGAGGGTTATATTAAGATCATAGATGTCTGCAAGTCAGGAAGAGGGTC
ATCCTTCATTAACTAATCCTCAATGCACCAAAAATATAAAAGAAAAGAAAGAGGAGGAAGAACAAATATCCTATAAGCTGGAATAT
CTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGCCTTCTTCTTCTTTTTATTTTTTCTGAGTGACCCAGTCT
GGTTTTTCTGTTGTAGCAGTCTGTCCATGCAACTCTTGCAGGGTTTATTTCAACATGAGTCGGAAAAGGATATCCTGTAAAGATC
TGGGCCACCCTGACGCTGACTGCCAGGGTTGGGTGGTTAGTTGGAAATTTTTCCCCCCCCCCCCCTGGTTTGTCATTATTACAAAT
TAGGAGAAGTGATCAATTTCCTTTCTCGCACATGATTAATATTTATTATAGGTCAGCATCATCCGGGTATCGATGGAAACACATCC
ACATTATATCCATTCGAGTCGATCA

}>PCR1-PCR_6990kb
GGACTCCGAGAAGGAGGCCCTCCTACCTCATCTCAGGATGGATGGGGCGCGATCTCAGGACATTCCCATCACCGGCTATTCCGAT
AAATGTTAATTTTGTGGATGCTGTGGAAAAGGGGGGCAA

}>PCR2-PCR_6990kb
CCTGCCATCTCGATCGTCAACACTAGGGACACCCCTCCACGGCTAATACAGTGGATCGAACTAACCCAGAAGAAAAAGCTGGAC
TCAGAAGGAGGAGGAGAACAAATATCCTACTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGC
AAGCATGTTGAAAAACCCTGCAAGAGTTGCATGGACAGACTGCTACAACAGAAAACAACAGACTGGGTCACTCAGAAAAATAA
AAAAGAAGAAAGAGGAGGAAGAACAAATATCCTAGAAGCTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTC
CTGACTTCTGTTGGGTTGTTTGCTGTTCGATGGGGGAAGAAA

}>PCR3-PCR_6990kb
AAACAGGCAAATGCGTGATGCTGTTCGAGATGTGTTGAGAGGTCATAGGTCATGAGGGTGGCGTCCTCATCGATGGCAGTTGTGC
CTTTATCTGGTGAACAGGAATTTCCTATTTGCCATGTGAGGGTTATATTAAGATCATAGATGTCTGCAAGTCAGGAAGAGGGTC
ATCCTTCATTAACTAATCCTCAATGCACCAAAAATATAAAAGAAAAGAAAGAGGAGGAAGAACAAATATCCTATAAGCTGGAATAT
CTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGCCTTCTTCTTCTTTTTATTTTTTCTGAGTGACCCAGTCT
GGTTTTTCTGTTGTAGCAGTCTGTCCATGCAACTCTTGCAGGGTTTATTTCAACATGAGTCGGAAAAGGATATCCTGTAAAGATC
TGGGCCACCCTGACGCTGACTGCCAGGGTTGGGTGGTTAGTTGGAAATTTTTCCCCCCCCCCCCCTGGTTTGTCATTATTACAAAT
TAGGAGAAGTGATCAATTTCCTTTCTCGCACATGATTAATATTTATTATAGGTCAGCATCATCCGGGTATCGATGGAAACACATCC
ACATTATATCCATTCGAGTCGATCA

}>PCR1-PCR_6990kb
GGACTCCGAGAAGGAGGCCCTCCTACCTCATCTCAGGATGGATGGGGCGCGATCTCAGGACATTCCCATCACCGGCTATTCCGAT
AAATGTTAATTTTGTGGATGCTGTGGAAAAGGGGGGCAA

}>PCR2-PCR_6990kb
CCTGCCATCTCGATCGTCAACACTAGGGACACCCCTCCACGGCTAATACAGTGGATCGAACTAACCCAGAAGAAAAAGCTGGAC
TCAGAAGGAGGAGGAGAACAAATATCCTACTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGC
AAGCATGTTGAAAAACCCTGCAAGAGTTGCATGGACAGACTGCTACAACAGAAAACAACAGACTGGGTCACTCAGAAAAATAA
AAAAGAAGAAAGAGGAGGAAGAACAAATATCCTAGAAGCTGGAATATCTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTC
CTGACTTCTGTTGGGTTGTTTGCTGTTCGATGGGGGAAGAAA

}>PCR3-PCR_6990kb
AAACAGGCAAATGCGTGATGCTGTTCGAGATGTGTTGAGAGGTCATAGGTCATGAGGGTGGCGTCCTCATCGATGGCAGTTGTGC
CTTTATCTGGTGAACAGGAATTTCCTATTTGCCATGTGAGGGTTATATTAAGATCATAGATGTCTGCAAGTCAGGAAGAGGGTC
ATCCTTCATTAACTAATCCTCAATGCACCAAAAATATAAAAGAAAAGAAAGAGGAGGAAGAACAAATATCCTATAAGCTGGAATAT
CTAGGCAAGGTCAAGTTGCTAACAGAATGACCTCTTCCTGACTTGCCTTCTTCTTCTTTTTATTTTTTCTGAGTGACCCAGTCT
GGTTTTTCTGTTGTAGCAGTCTGTCCATGCAACTCTTGCAGGGTTTATTTCAACATGAGTCGGAAAAGGATATCCTGTAAAGATC
TGGGCCACCCTGACGCTGACTGCCAGGGTTGGGTGGTTAGTTGGAAATTTTTCCCCCCCCCCCCCTGGTTTGTCATTATTACAAAT
TAGGAGAAGTGATCAATTTCCTTTCTCGCACATGATTAATATTTATTATAGGTCAGCATCATCCGGGTATCGATGGAAACACATCC
ACATTATATCCATTCGAGTCGATCA
The PCR product sequencing alignment to the genome (GRCh38). Analysis generated using the BLAT tool, data presented for both forward and reverse primers for each reaction:

**PCR1-PF3 (+ strand)**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAGAACATGAGGGTGCATCATGCCCCCCCCCCCCCCAA</td>
<td>0000066</td>
</tr>
<tr>
<td>CAGGTTCCTGTCCTTTTCTTTGGAGCGCGACAGCATGCGGACACTCTTGAGTCGTAAGATCATCTCTCCATACCACACAG</td>
<td>6830493</td>
</tr>
<tr>
<td>TGATGATGAGGACCGGCATGATGAGAGCAAAGATGAAAACACAGATTTTGAGCAGGTTCTCCCATCACCATGTA</td>
<td>6830494</td>
</tr>
</tbody>
</table>

**PCR1-PF3 (- strand)**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>6972299 tccttgtatcctctctgaag</td>
<td>0000190</td>
</tr>
<tr>
<td>6972299 tccttgtatcctctctgaag</td>
<td>6972318</td>
</tr>
<tr>
<td>6972199 aagaaaaagctggactcccagagagggtgtgtacagcatccagtgtgacc</td>
<td>00000119</td>
</tr>
<tr>
<td>6979616 ccaaccctggcagtcagcgtggcccagatctttacaggatatccttctcc</td>
<td>0000240</td>
</tr>
<tr>
<td>6985683 actgctacaacagaaaacacagactggctgactcgacagaaaaatagga</td>
<td>0000328</td>
</tr>
<tr>
<td>6985373 taggcaaggtcaggtgctcaataaagttgatatagcttgtgacttgtgactgta</td>
<td>0000428</td>
</tr>
<tr>
<td>6985833 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>0000478</td>
</tr>
<tr>
<td>698583 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985832</td>
</tr>
<tr>
<td>6985683 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985833</td>
</tr>
<tr>
<td>6985833 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985882</td>
</tr>
<tr>
<td>6985683 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985832</td>
</tr>
<tr>
<td>6985683 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985833</td>
</tr>
<tr>
<td>6985683 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985882</td>
</tr>
<tr>
<td>6985683 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985832</td>
</tr>
<tr>
<td>6985683 gacactctctctgtgacatcataatatactctacatccagtatggaatattctctg</td>
<td>6985833</td>
</tr>
</tbody>
</table>
PCR2-PF3 (+ strand)

0000016 aacactagg.aacaccctccacggctaatcacggtgatcgaactaacc 0000064
0000065 ccag 0000068
0000494 ccag 6830497

0000069 ccaccccttgccagtctcggtgcccagatctttacaggtatctcttc 0000118
0000119 gactcatgtgtgaaaaccctgcagagttgcatggcacag 0000157
0000158 tgtgc 0000162
0000363 tcaccaaataaaggcacaactgccatcgatgacgacccccccctcatgac 6985882
0000463 aggacttgg.catgtggaatgtggca 0000487

PCR2-PR_6990kb (- strand)

0000016 aatgggatgctgttcagagatgtgttgagaggtcataggtcatgagggt 0000059
0000059 aatgggatgctgttcagagatgtgttgagaggtcataggtcatgagggt 6985971
0000363 tcaccaataaaaggcacaactgccatcgatgacgacccccccctcatgac 6985883
0000463 aggacttgg.catgtggaatgtggca 6985973

0000060 ggcgtcctcatcgatggcagttgtgcctttatctggtgaacaggaatttc 0000109
0000109 ggcgtcctcatcgatggcagttgtgcctttatctggtgaacaggaatttc 6985921
0000110 ctatgacctctccatcactctctctgacaccatccatcctagttgaaatttt 6985933

0000060 ggcgtcctcatcgatggcagttgtgcctttatctggtgaacaggaatttc 6985921
0000463 aggacttgg.catgtggaatgtggca 6985983

0000119 gactcatgtgtgaaaaccctgcagagttgcatggcacag 0000157
6830494 gacatctatgatcttaatataacctcacatggcaaataggaaattcctgt 6985973
6830494 gacatctatgatcttaatataacctcacatggcaaataggaaattcctgt 6985973
60000213 aaaaagaagaagaagaagaacaaatatctataaaactggaatgctc 6985833
60000213 aaaaagaagaagaagaagaacaaatatctataaaactggaatgctc 6985833
6000313 cacttctatgtcttaaataacccatcagggcaataggaatctcccc 6985833
6000313 cacttctatgtcttaaataacccatcagggcaataggaatctcccc 6985833
6000069 ccaccccttgccagtctcggtgcccagatctttacaggtatctcttc 6979616
6000069 ccaccccttgccagtctcggtgcccagatctttacaggtatctcttc 6979666
6000119 gactcatgtgtgaaaaccctgcagagttgcatggcacag 6979704
CTATTTGCCATGTGAGGTATATTAGATATACATAGATGTCTGCAAGTCCAGG

AAAGGTCATCTCATTGTAACATGCGTCAGCATATTCCACG

TTCCTAGGATATTGCTGTCTCTCCTTCTCTCTCTCTCTCTTATT

TTCTAGGATATTGCTGTCTCTCCTTCTCTCTCTCTCTCTTATT

ACCTGAGTCACCCCATGCTGTCTGTGTTCTGTTGTCAGTCTAC

CCCTGCAATCGCTATCTTGCTGGTTTTCAACTTGAATCAGAAAAGAT

ATCCTGTTAGTACCGCCACCTGACTGCCAGGT

GGGTAATTCAATCATTGTATTTAGCCGGGGAGGGGTGTTCCCTATTT

TTTTGACGATTGCACAGAGTTTTGCCTCTGATT

GTAACGAGACTCTGCTCAAAATCTGTGTTTTCATCTCTGCCTTCA

TGCCGCTTATCATCATTGTTATGACTAGCTGTATTACGACTC

AAGAGTGTCCGCAATGCTGTCCGGCTCCAAAGAAAGACAGGAACCTGC

CAGGATCCGCCAGATGTTGCTGGTTTCTGATTATATATGGCTCT

GGACCCCATCCACATCTATGTCATCATCAAAAGCACGTGACGATCC

GGACCACCATCCTTCACTTCGACTTGTAGGCTTGGTTA

CACAACAGCTGCTCAACCCAGTTTCTTTATGCGTCTCTGGATGAACACT
0000839 acacagatttgagcaggttctcccatcaccatgtgatgagagaacat 0000888
6830046 acacagatttgagcaggttctcccatcaccatgtgatgagagaacgt 6829997
0000889 gagggtgca 0000897
6829996 gagggtgca 6829988