UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

## Individual project in Bioinformatics

Lucía Sánchez Schneider

# Analysing opioid prescription patterns from the UK Biobank to uncover genetic determinants of the *OPRM1* gene in drug response

Internal Supervisor: Anders Albrechtsen

External Supervisors: Alexander Sebastian Hauser, Jakob Sture Madsen;
Department of Drug Design and Pharmacology, Pharmaceutical Informatics
(University of Copenhagen)

Date: 20 January 2025

# Table of Contents


*Abstract* ..... 3

*Introduction* ..... 4

*Methods* ..... 6

**Data source** ..... 6

**Exploratory data analysis using the OMOP data model** ..... 7

**Phenotype definition** ..... 8

**Association analysis using REGENIE** ..... 8

Step 1: Ridge regression model fitting ..... 8

Step 2: Association testing with variants of interest ..... 9

Quality control of variants in the *OPMR1* gene ..... 10

Mask selection ..... 12

Covariates selection ..... 12

*Results* ..... 12

**Exploratory data analysis of opioid prescriptions** ..... 12

**WGS association analysis with morphine/fentanyl usage phenotypes** ..... 14

Single-variant association testing ..... 14

Gene-based testing ..... 15

*Discussion* ..... 16

**Conclusion** ..... 18

*References* ..... 18

*Supplemental Material* ..... 21

# Abstract

Opioid medications are critical for pain management but exhibit wide inter-individual variability in efficacy and adverse effects, often mediated by genetic factors [1]. This study aimed to investigate the genetic determinants of opioid response, with a focus on the *OPRM1* gene, which encodes the μ-opioid receptor, the primary target for opioids. Leveraging the extensive genomic and phenotypic data available in the UK Biobank, I conducted an integrative analysis to explore the association between genetic variants and opioid usage phenotypes, specifically for morphine and fentanyl. With an initial exploratory data analysis three phenotypes—morphine-only, fentanyl-only, and combined usage—were defined. Using the REGENIE framework, a two-step genome-wide association analysis was performed, including both single-variant and gene-based tests. While no statistically significant associations given the low sample size were detected, gene-based burden tests revealed trends linking genetic variants with opioid efficacy, suggesting potential influences on switching behaviours between morphine and fentanyl. This study underscores the complexity of pharmacogenomic research and highlights the need for refined phenotype definitions and expanded gene analyses to unravel the genetic mechanisms underlying opioid response. Moreover, it could serve as a solid foundation for exploring and advancing pharmacogenomic applications in other therapeutic domains.

# Introduction

Opioids are a class of drugs primarily used for their analgesic properties to manage pain. Beyond their therapeutic applications, opioids also carry a significant risk of adverse effects, including tolerance and addiction [2].

Pharmacogenomics, the study of genetic influences on drug response, offers a promising pathway to optimize opioid therapy, addressing challenges such as variability in efficacy and the risk of adverse effects. Genetic variations in drug targets like *OPRM1* (μ-opioid receptor) and metabolizing enzymes like *CYP2D6* significantly affect opioid pharmacodynamics and pharmacokinetics, influencing drug efficacy, side effects, and dependence potential [3], [4]. These insights have driven the development of precision medicine approaches, such as genotype-guided dosing, which aim to tailor opioid prescriptions to individual genetic profiles [1]. Despite its potential to reduce opioid-related adverse events and enhance safety, translating pharmacogenomic findings into clinical practice remains a challenge, requiring further validation and integration into standardized prescribing practices.

One key genetic player in opioid response is the *OPRM1* gene, which encodes the μ-opioid receptor. This receptor is the primary molecular target for most opioid medications (Figure 1A), mediating their effects such as pain relief, reward, and addiction-related behaviours. Given its role in the variability of opioid responses and the development of addiction, studying *OPRM1* is essential for understanding inter-individual differences in opioid therapy outcomes and advancing pharmacogenetic strategies for safer and more effective treatments [5], [6], [7].

OPRM1 is part of a large protein family known as GPCRs (G protein-coupled receptors), which are extensively mutated in the human population [8]. Therefore, studying the *OPRM1* gene in large-scale datasets like the UK Biobank offers valuable insights into genetic factors influencing opioid efficacy, enabling more personalized treatments. The UK Biobank, containing genetic, phenotypic, and health data from over 500,000 participants, is an unparalleled resource for investigating opioid-related genetic factors. It enables the analysis of opioid use patterns and the identification of associated genetic variations [9]. Specifically, in this study, the OMOP *drug eras* model was utilized, this model is applied to the prescription data from the UK Biobank to analyze standardized data and identify patterns of opioid use (Figure 1B).

To deepen our understanding of the functional implications of *OPRM1* mutations, I have utilized unpublished cell-based *in vitro* data assessing the effects of the ten most common missense mutations in *OPRM1* on opioid efficacy, using signalling assays (Figure 2). The colour scale represents normalized changes in potency (ΔLogEC50) compared to the wild-type receptor (WT),

providing a comparative view of how these mutations alter the pharmacological efficacy of various opioids, including morphine, fentanyl, and others. For instance, the black cells represent reduced potency compared to WT, yellow cells indicate comparable potency to the wild type, while crossed cells denote no measurable activity suggesting no effect when treated with the respected opioid. Our study aims to bridge these laboratory findings with population genotype-to-phenotype data, investigating whether these *OPRM1* variants, shown to affect opioid efficacy *in vitro*, are also associated with opioid-related phenotypes within the UK Biobank cohort. Through this integrative approach—combining *in vitro* functional data with large-scale genomic and phenotypic datasets—I seek to uncover the genetic mechanisms of opioid response (Figure 1C).

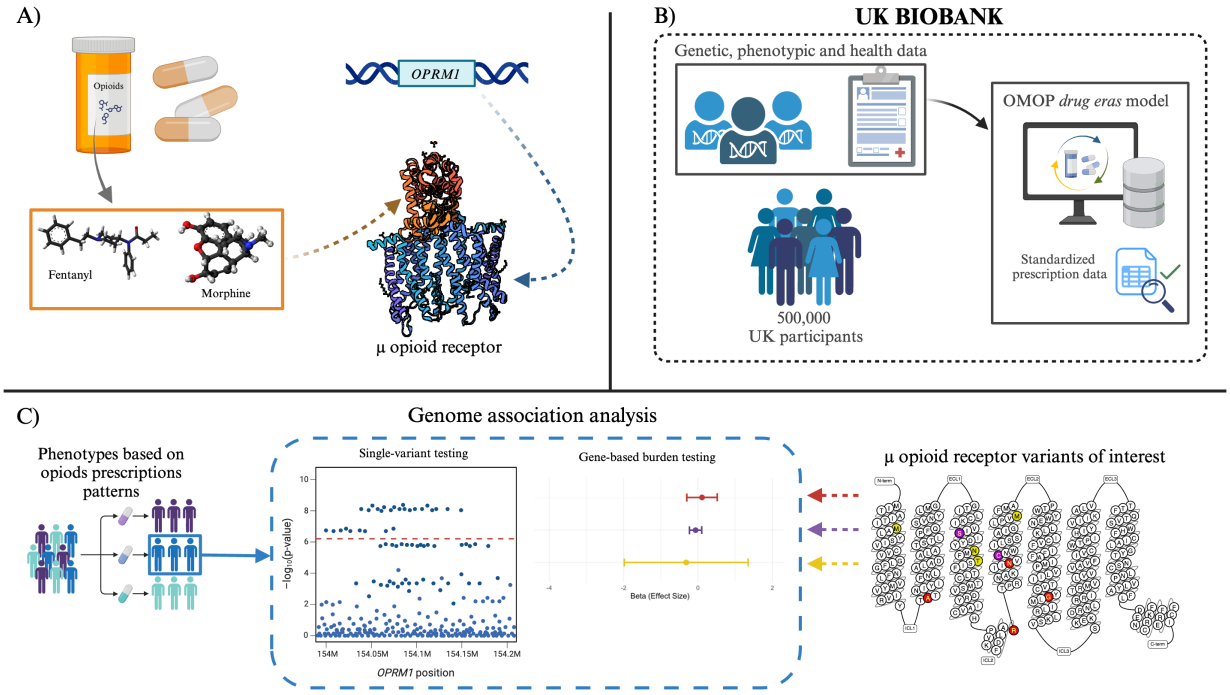



**Figure 1: Overview of Study Design and Methodology.**
**A)** The OPRM1 gene encodes the μ-opioid receptor, which serves as the primary molecular target for opioid medications, including morphine and fentanyl. The figure highlights the molecular structures of these drugs and the crystal structure of the μ-opioid receptor bound to a morphinan antagonist (PDB ID: 4DKL). This receptor's genetic variations are key to understanding differential responses to opioids. **B)** The study leverages the UK Biobank, a resource comprising genetic, phenotypic, and health data from over 500,000 participants. The analysis specifically utilizes the OMOP *drug eras* data model, a standardized prescription data framework, to explore opioid-related prescription patterns. **C)** The analysis focuses on *OPRM1* variants with known in vitro effects on opioid efficacy, aiming to determine their association with opioid-related prescription phenotypes. Statistical genome association analyses were performed, including single-variant testing (illustrated with a Manhattan plot) and gene-based burden testing (illustrated with a forest plot). The structural representation on the right highlights μ-opioid receptor variants of interest. These variants are used to create masks for gene-based burden testing. The colour pattern of the variants corresponds to their effect on morphine efficacy: red indicates no efficacy, purple indicates low efficacy compared to the wild type, and yellow indicates similar efficacy to the wild type.
Created in https://BioRender.com

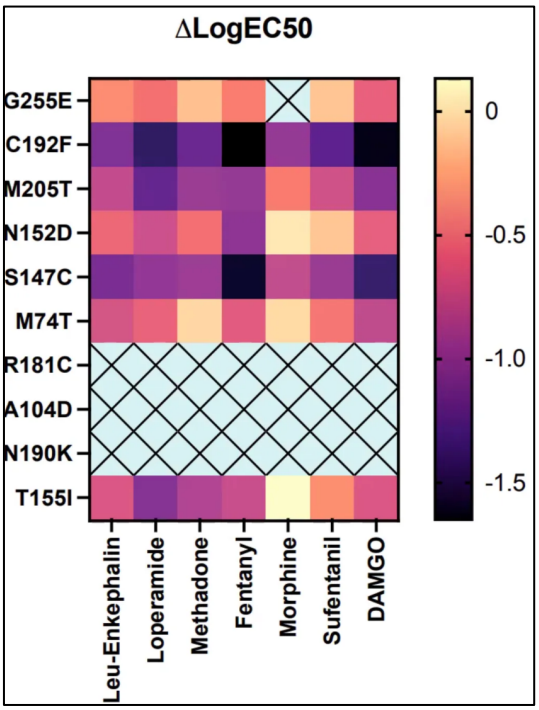



**Figure 2: Heatmap of the functional implications of *OPRM1* missense mutations on opioid efficacy.**
This heatmap displays the effects of the 10 most frequent missense mutations in the *OPRM1* gene on the pharmacological efficacy of various opioids, measured by changes in potency (ΔLogEC50) in an *in vitro* signalling assay. The colour scale represents normalized potency changes relative to the wild-type (WT) receptor. The mutations are shown on the y-axis, while the tested opioids—including Leu-Enkephalin, Loperamide, Methadone, Fentanyl, Morphine, Sufentanil and DAMGO—are on the x-axis. Figure provided by collaborators.

# Methods

## Data source

For this study the UK Biobank was used, which is a large-scale biomedical biomedical resource that includes extensive genetic, phenotypic, and health-related data from approximately 500,000 participants across the United Kingdom. This cohort offers a wide range of information, such as detailed medical records, lifestyle factors, and genetic data, including imputed genotype data, whole-exome sequencing, and whole-genome sequencing [9].

Given the large scale of the UK Biobank dataset, Apache Spark, a distributed computing framework, is employed to handle and process the data efficiently. Spark enables parallel processing and in-memory computation, making it suitable for managing extensive and complex datasets [10].

## Exploratory data analysis using the OMOP data model

For this study, I utilized the UK Biobank Research Analysis Platform (RAP) to analyse the more than 50 million prescription data records, with the goal of identifying patterns of drug exposure or switching, in order to define opioid-related phenotypes and examine medication usage behaviours. For the analysis, the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), version 5.3 was used. The OMOP-CDM provides a standardized framework for organizing healthcare data including a comprehensive range of information, such as prescription records, diagnoses, surgical procedures, and demographic details, all structured into a unified format [11]. In this study, I specifically focus on *drug eras*, which aggregates prescription records into periods of drug exposure. A drug era is defined as a continuous period during which an individual is assumed to be exposed to a specific active ingredient. A drug era begins with the start date of the first recorded drug exposure and requires a minimum gap of 31 days from the last exposure to the same ingredient to qualify as a new era. The end date of the drug era is determined by the final recorded drug exposure within that period. This aggregation is a key tool for examining opioid usage patterns and the duration of exposure, providing a standardized and reliable method to study medication usage behaviours across the cohort [12]. This standardized approach facilitates the examination of opioid usage patterns and exposure durations.

My exploratory data analysis began by identifying the opioids of interest from Figure 2 — loperamide, methadone, fentanyl, morphine, and sufentanil—within a mapping dataset, which linked drugs to their respective codes in the OMOP *drug eras* table. These codes were then used to extract the relevant drug data from the *drug eras* table. Separate tables were created for each drug, containing records of participants who had been prescribed that drug. These tables included details such as participant identifiers, the start and end dates of drug exposure periods, the number of exposures, and the gap days between exposures.

The extracted data were then summarized by grouping records for each participant. Key metrics were computed, including the total number of drug exposure periods, the distribution of start and end dates for each exposure, the number of exposures during each period, and the gaps between consecutive exposures. To further explore the data, visualizations such as an upset plot of drug combinations and histograms were created. This analysis guided the selection of phenotypes that were considered most significant for further investigations.

## Phenotype definition

Following the exploratory data analysis on opioid prescription data from the UK Biobank, key opioid-related phenotypes were defined based on the observed patterns of drug usage. Specifically, the binary phenotypes were defined based on exposure to morphine and fentanyl as follows:

- Control group: Individuals with any drug prescriptions present in the OMOP *drug eras* table who have not been exposed to either morphine or fentanyl.
- Phenotype cases:
  - Individuals taking only morphine.
  - Individuals taking only fentanyl.
  - Individuals taking both morphine and fentanyl.
- Other participants: Classified as NA, i.e., individuals who do not fit into any of the above categories.

A script was created to assign these phenotypes based on participant and drug exposure data. Initially, participant identifiers were retrieved from the OMOP *drug eras* table. Then lists of participants exposed to only morphine, fentanyl, and both drugs were created. Using these lists binary values were assigned to the phenotype columns, with a value of 1 indicating exposure to a specific phenotype and 0 for individuals in the control group.

After assigning the phenotypes, the data were prepared for export in a format compatible with REGENIE's input requirements, so two additional columns, 'FID' **(**Family ID) and 'IID' (Individual ID), were added to the dataset. Both columns were filled with the participant identifiers. The final output includes participant identifiers and the corresponding phenotype values.

## Association analysis using REGENIE

Once the opioid-related phenotypes were defined, we performed association analysis using the REGENIE framework. REGENIE is a computationally fast and memory-efficient tool for genome-wide association studies (GWAS), designed to handle large datasets and complex phenotypes, including binary and quantitative traits. Its key advantage lies in its two-step methodology, which minimizes computational demands while maintaining statistical power [13].

#### Step 1: Ridge regression model fitting

The first step of the REGENIE framework involves fitting a ridge regression model to estimate the genetic component of the phenotype. For this, I utilized genotype data from the UK Biobank, specifically the dataset titled "Genotype Calls" (ukb22418). This dataset provides high-quality genome-wide genetic data generated from custom genotyping arrays. This step was run

following UK Biobank recommendations (https://rgcgithub.github.io/regenie/recommendations/).

To prepare the data, genotype files for chromosomes 1–22 were merged using PLINK, ensuring all files were processed to generate a single unified dataset. Quality control (QC) was then performed on this dataset using PLINK2 to filter variants based on minor allele frequency (MAF $\geq 0.01$), minor allele count (MAC $\geq 20$), genotype missingness (geno $\leq 0.1$), and Hardy-Weinberg equilibrium (p-value $\geq$ 1e-15). Samples with excessive missingness (mind $\leq 0.1$) and those lacking sex information were also excluded.

Following QC, the resulting dataset of filtered genotype variants was utilized in Step 1 of the REGENIE framework, where a whole genome regression model is fit to this subset of filtered genetic markers. The process begins by applying ridge regression, where the data is divided into blocks of consecutive markers. For each block, a set of predictors is generated using several shrinkage parameters to estimate the genetic contribution to the phenotype while regularizing the model to prevent overfitting. This approach is applied to reduce the dimensionality of the data, focusing on key genetic variants within each block.

Next, these predictors are used to build a genome-wide regression model, in our case for binary traits, using logistic ridge regression. This model combines the predictors to estimate the probability of being a case. Cross-validation is applied to select the optimal shrinkage parameter. Covariates such as age, sex, and 20 first principal components (to account for population structure) are included in the model (Covariates selection).

A Leave-One-Chromosome-Out (LOCO) approach is applied when building the full model. By excluding the chromosome being tested from the model during each prediction, the LOCO approach avoids proximal contamination ensuring that genetic predictions are independent of the chromosome being analysed. These LOCO predictions are crucial in Step 2 of REGENIE when testing each marker for association. [13]

Finally, the output of this step, which includes null phenotype predictions, was stored for use in Step 2.

#### Step 2: Association testing with variants of interest

In the second step, I conducted association testing using Whole Genome Sequencing (WGS) data from the UK Biobank's "DRAGEN Population-Level WGS Variants" dataset (pVCF format, 500k release). This dataset contains high-quality genomic data for approximately 500,000 participants, processed using the DRAGEN 3.7.8 platform. DRAGEN is a genomic analysis tool that efficiently processes WGS data to generate accurate genetic variant calls. The dataset includes both individual

and joint-called data across all participants. The data is provided in phased Variant Call Format (pVCF), which not only captures genetic variants but also includes phase information, indicating the specific inheritance patterns of alleles across the genome. The filtered and annotated variants of interest from the *OPRM1* gene, obtained after extensive QC (see Quality control of variants in the *OPMR1* gene), were tested for association with the phenotype.

The association analysis began with single-variant testing. To optimize memory usage, markers were read in blocks and tested sequentially, avoiding the need to load all markers into memory at once. A logistic regression score test was used to evaluate the association between each genetic marker and the phenotype. The logistic regression model included LOCO predictions from Step 1 as an offset, while covariates were again incorporated to control for potential confounders. To address the challenges of imbalanced case-control ratios, which can lead to type I error inflation at rare markers, REGENIE employed the Firth correction. This correction reduced bias in maximum likelihood estimates and resulted in better-calibrated test statistics. P-values for single-variant tests were calculated using a likelihood ratio test (LRT) [13].

Following the single-variant tests, a burden test was performed as the default statistical test in REGENIE's gene-based association analysis. The burden test aggregates the effects of multiple genetic variants within a group and evaluates whether their combined effect is associated with the phenotype of interest. By collapsing variants into a single score, the burden test increases statistical power, particularly when the variants within a group have similar biological effects. This is especially helpful when analysing rare variants, which are often underpowered in single-variant tests. By default, the burden tests in REGENIE include cutoffs based on 1% alternate allele frequency (AAF) and singletons ensuring a focus on rare and low-frequency variants. The grouping of variants for the burden test is defined using a mask file, which contains the criteria for selecting and grouping variants [14], [15]. Details on the construction and content of the mask file are provided in the Mask selection section.

#### Quality control of variants in the *OPMR1* gene

In Step 2 of our analysis, the pipeline from Kizilkaya et al. 2024 [16], originally applied to Whole-Exome Sequencing (WES), was expanded upon by conducting rigorous QC on the Whole Genome Sequencing (WGS) data from the UK Biobank's "DRAGEN Population-Level WGS Variants" dataset. The QC process was implemented using Hail, a computational framework specifically designed for scalable genomic data analysis (Hail Team. Hail 0.2.116 https://github.com/hail-is/hail/).

The QC process began by obtaining gene annotations from the GENCODE database [17], which were used to define genomic intervals for the *OPMR1* gene. Using Hail, these gene intervals were intersected with genomic blocks from the DRAGEN pVCF dataset to ensure that only the relevant genomic regions were included in the analysis. The genomic blocks were filtered and parsed, focusing on variants within the defined intervals.

Next, VCF files corresponding to the selected blocks were imported into Hail, where singletons were filtered out to reduce the number of non-significant variants, and multi-allelic variants with more than six alleles were removed to simplify the analysis. The remaining variants were split into single alleles, ensuring consistency and compatibility with downstream workflows. Variants were further filtered based on quality metrics, retaining only those marked as "PASS" in the FILTER field and with at least one non-missing genotype.

To ensure data quality, variants with fewer than 95% unfiltered genotypes and samples with fewer than 95% unfiltered variants were removed.

Additionally, a predefined list of samples was used to exclude problematic samples, including those with outliers for heterozygosity or missing genotype rates (UK Biobank field 22027), sex chromosome aneuploidy (field 22019), excessive genetic kinship (field 22021), and those from specific ethnic backgrounds, such as *White and Black Caribbean*, *White and Black African*, *White and Asian*, or *Any other mixed background* (field 21000). An ancestry filter further removed samples not belonging to the *United Kingdom* group to reduce population stratification. Samples that had withdrawn from the study or were related to others with a kinship coefficient above 0.176 (second-degree relatives or closer) were excluded using a maximal independent set approach to retain only one representative per closely related group. This filtering ensured that only high-quality and independent samples remained for the analysis.

The Variant Effect Predictor (VEP) [18] was integrated with Hail to annotate the filtered variants, providing detailed information about their potential functional consequences. Variants were filtered to retain only those with meaningful annotations, excluding intronic and downstream/upstream gene variants. Other filters were applied, retaining only variants with a high genotype quality (mean GQ $\geq$ 20), call rates $\geq$ 95%, and at least one non-reference allele. A final annotation was generated with gene symbols and the most relevant consequences for each variant.

Finally, the cleaned and annotated dataset was exported in various formats, including a variant QC file, an annotation file, a setlist file specific to the *OPRM1* gene, and a BGEN file to store genotype probabilities. These outputs were prepared for input into the REGENIE association testing pipeline, ensuring that only high-confidence, biologically relevant variants were analysed.

#### Mask selection

For gene-based association testing in REGENIE Step 2, the missense variants of the *OPRM1* gene shown in Figure 1 that passed QC were grouped into three functional masks that specifically reflect how they alter morphine efficacy in an *in vitro* signalling assay. These masks were defined as follows:

- *No efficacy*: Variants G255E, R181C, A104D, and N190K, which completely abolish morphine efficacy.
- *Low efficacy*: Variants C192F and S147C, reduce morphine efficacy compared with WT.
- *WT efficacy*: Variants N152D and T155I, retain morphine efficacy like WT.

#### Covariates selection

Covariates are essential components in genetic association studies, as they account for factors that might influence the relationship between genetic variants and the phenotype of interest. These factors help to control for confounding effects, ensuring that the genetic signals detected are not biased by population structure or other external variables like biological age-sex differences. In the REGENIE pipeline, covariates are used in both Step 1, where they help in generating accurate phenotype predictions from genetic data, and Step 2, where they adjust for confounding variables during association testing. In this study, covariates were generated following the methodology outlined in [13]. These covariates include genetic sex, age, and the first 20 principal components (PCs), as well as interaction terms like $age^2$, age × sex, and $age^2$ × sex, which account for more complex relationships between variables and non-linear effects. The covariates were obtained from the UK Biobank dataset, specifically from fields 21022 (age at recruitment), 22001 (genetic sex), and 22009 (genetic principal components). After cleaning the data to remove missing values, the data was processed to generate relevant covariates that were exported in the required format, with the 'FID' and 'IID' columns, for the REGENIE pipeline.

# Results

## Exploratory data analysis of opioid prescriptions

To identify opioid-related phenotypes for further association analysis, I conducted an exploratory data analysis of opioid prescription data using the OMOP *drug eras* table from the UK Biobank dataset. The focus of this analysis was on five opioids: loperamide, methadone, fentanyl, morphine, and sufentanil, as these were the ones tested in the *in vitro* signalling assay presented in Figure 2.

However, sufentanil did not appear in the OMOP database, so subsequent analyses were conducted using the remaining four drugs.

Initially, I generated histograms of drug usage for each drug (Supplementary Figure 1: Histogram of drug usage.), where I observed that loperamide was the drug taken by the largest number of unique individuals. Furthermore, when analysing total drug exposure across the population (Supplementary Figure 2: Histogram of total exposure per drug.), morphine emerged as the drug with the highest cumulative exposure.

Next, patterns of drug combinations were explored using an upset plot (Figure 3), this visualization revealed that more than half of the people taking fentanyl have also taken morphine. Based on this observation, I decided to focus the study on morphine and fentanyl, as the patterns of switching between these two analgesics are particularly interesting.

Further analysis included plotting the distribution of total drug exposure for morphine and fentanyl (Supplementary Figure 3). These plots highlighted that for both drugs, a large proportion of individuals had only a single exposure to the drug, indicating short-term or sporadic usage in many cases.

To better understand the interplay between morphine and fentanyl, analysed drug-switching patterns. The results showed that 6,240 individuals were exclusively exposed to morphine, while 707 individuals were exclusively exposed to fentanyl. Notably, 977 individuals were exposed to both drugs, with 563 of them switching from morphine to fentanyl and 414 switching from fentanyl to morphine. Among those exposed to both drugs, 294 individuals switched between morphine and fentanyl more than once.

Based on the insights derived from this exploratory analysis, three simple phenotypes were defined for the subsequent association analysis. These included individuals taking only morphine, individuals taking only fentanyl, and individuals taking both morphine and fentanyl.

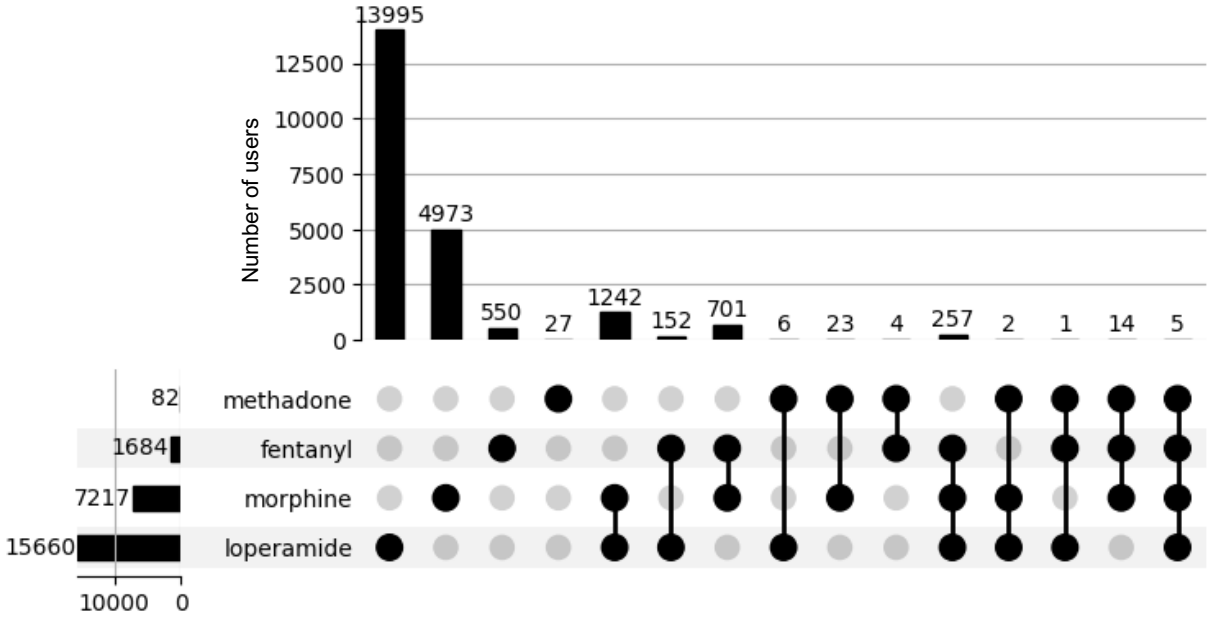



**Figure 3: UpSet plot of drug combinations.**
This UpSet plot illustrates the number of individuals who were prescribed different combinations of opioids from the UK Biobank dataset. The horizontal bars on the left represent the total number of users for each drug: loperamide (15,660 users), morphine (7,217 users), fentanyl (1,684 users), and methadone (82 users). The dots and connecting lines below the bar graph indicate specific drug combinations, while the vertical bars above show the number of individuals corresponding to each combination.

## WGS association analysis with morphine/fentanyl usage phenotypes

### Single-variant association testing

Single-variant association testing was conducted on 4,756 variants from the *OPRM1* gene that passed the previously described quality-control filters. REGENIE Step 2 was used to assess the relationship between these variants and each of the three binary phenotypes. To visualize the results from REGENIE Step 2 of the association between these variants and each of the three binary phenotypes, Manhattan plots were generated for each phenotype (Supplementary Figure 4). A Bonferroni correction was applied to account for multiple comparisons, but no variants surpassed the significance p-value threshold ($0.05/4{,}756 = 1.051 \times 10^{-5}$). Furthermore, attempts were made to annotate the variants of interest from Figure 2; however, all these variants were excluded from single-variant tests because, by default, REGENIE omits variants with a minor allele count (MAC) below 5. Given that single-variant association methods are underpowered to detect associations in these rare variants, further gene-based testing was performed to capture the combined effects of multiple variants.

**Gene-based testing**

In the final stage of the association analysis, a gene-based burden test was conducted in REGENIE Step 2. Variants of interest in the *OPRM1* gene were grouped according to three functional masks based on their impact on morphine efficacy —*No efficacy*, *Low efficacy*, and *WT efficacy*— (see Mask selection). This grouping collapses rare variants into biologically relevant categories, thereby increasing statistical power to detect associations that single-variant tests may overlook. The results were visualized with a forest plot (Figure 4) that displays the estimated effect size (Beta) for each combination of mask and phenotype, along with 95% confidence intervals that reflect the uncertainty around those estimates. Beta indicates the direction and magnitude of association, while the standard error (SE) quantifies the uncertainty in the estimate; confidence intervals are derived by subtracting and adding 1.96 times the SE from Beta. Positive Beta values indicate an elevated likelihood of association with a given phenotype relative to control. The phenotypes analyzed in this study are binary and include individuals taking only morphine, individuals taking only fentanyl, and individuals taking both morphine and fentanyl.

The expectation, based on the variants affecting morphine efficacy from Figure 2, is that variants with a greater impact on morphine efficacy should be less associated with morphine consumption and more associated with fentanyl consumption. The figure revealed a potential trend suggesting a gradual decrease in Beta values from No efficacy to WT efficacy masks across all phenotypes., with the effect being more pronounced in the phenotypes of both morphine and fentanyl users as well as only fentanyl users.

- *No efficacy* Mask: This mask includes also variants with varying effects on fentanyl. While R181C, A104D, and N190K abolish both morphine and fentanyl efficacy, G255E retains near-WT fentanyl efficacy. This heterogeneity might suggest that the positive Beta value observed for the *No efficacy* mask in the only fentanyl users phenotype is potentially driven by the presence of G255E.
- *Low efficacy* Mask: This mask contains variants that reduce both morphine and fentanyl efficacy. This could contribute to the generally negative close to zero Beta values observed across phenotypes. Particularly a more noticeably negative value was observed in both morphine and fentanyl users phenotype, suggesting that individuals taking both drugs might be more likely to carry these variants that reduce both drug efficacy.
- *WT efficacy* Mask: This mask includes variants that primarily affect fentanyl efficacy while preserving morphine's efficacy. This could explain the lower Beta values observed for this mask in only fentanyl and both morphine and fentanyl users phenotypes, potentially due to the underrepresentation of these variants in individuals primarily using fentanyl.

Although these results may hint that individuals requiring fentanyl or morphine could be associated with variants that cause reduced efficacy of one or both of these drugs, the wide confidence intervals indicate that none of the observed effects can be statistically distinguished from zero. Additionally, all p-values exceed 0.05, confirming the lack of significant associations.

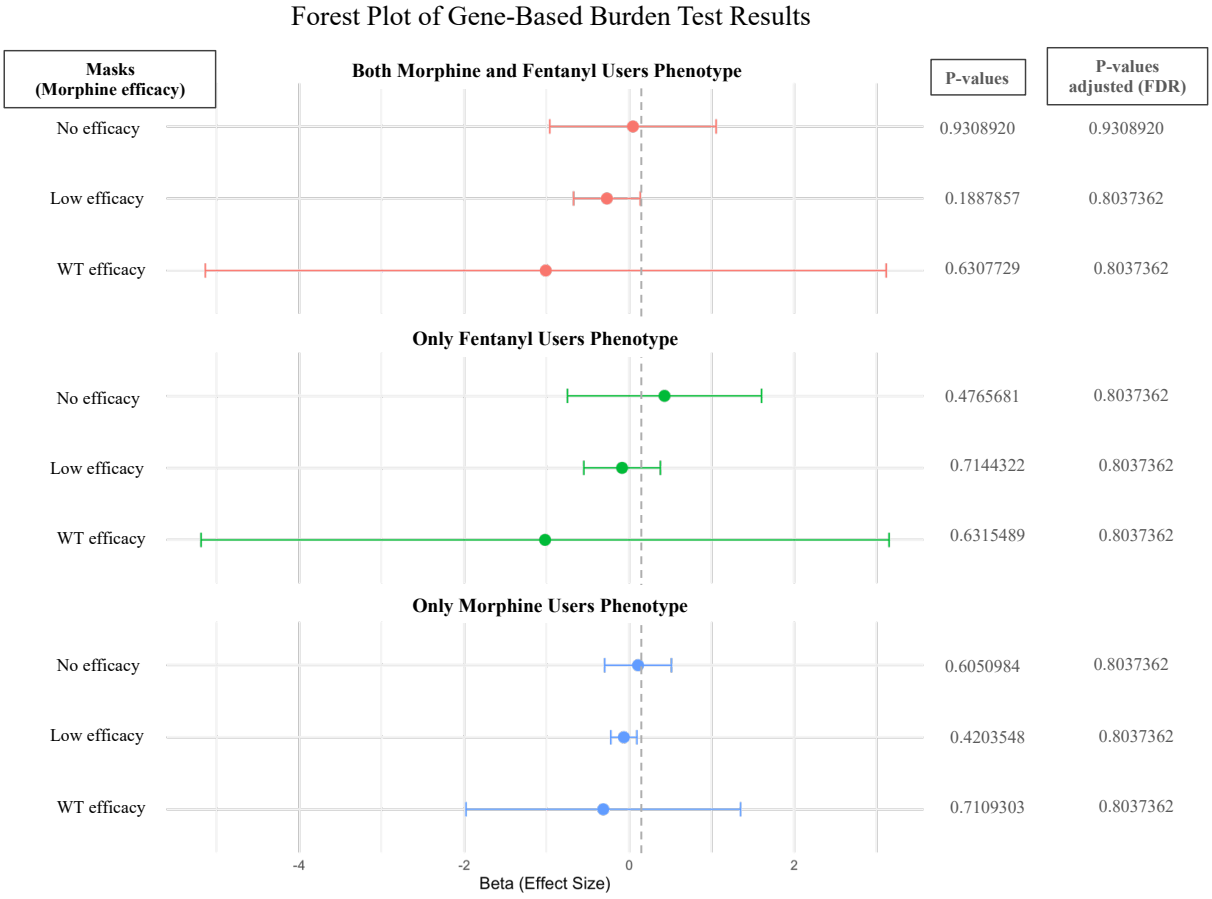



**Figure 4: Forest plot of gene-based testing results for *OPRM1* variants grouped by functional masks.**
This forest plot visualizes the results of gene-based association testing for OPRM1 variants grouped into three functional masks based on their impact on morphine efficacy: *No efficacy*, *Low efficacy*, and *Wild-Type (WT)* efficacy. The results are shown for three opioid-related phenotypes: Both morphine and fentanyl users, only fentanyl users, and only morphine users.
Each point represents the estimated effect size (Beta) for the association, with horizontal lines showing 95% confidence intervals derived by adding and subtracting 1.96 times the standard error (SE) from Beta. The p-values corresponding to each test are displayed alongside the plot for each mask and phenotype. Additionally, the p-values have been adjusted using the False Discovery Rate (FDR) method to control for multiple comparisons, ensuring that the reported p-values reflect a reduced likelihood of false-positive findings.

# Discussion

In this study, the extensive resources of the UK Biobank dataset were leveraged to perform an exploratory data analysis (EDA) and whole-genome sequencing (WGS) association analysis focused on opioid use, specifically examining morphine and fentanyl prescription patterns. Despite observing potential trends in gene-base burden testing, no statistically significant associations between genetic variants that affect morphine and fentanyl efficacy in the *OPRM1* gene, and the

defined phenotypes were identified. However, these findings contribute to the development of a robust pipeline for pharmacogenetics studies, highlighting key methodological considerations and opportunities for future research.

The UK Biobank provided a robust platform for this investigation, offering a large, diverse cohort and comprehensive genomic and clinical data. This enabled us to examine opioid use patterns across a wide population, identify phenotypes, and conduct detailed genetic analyses. However, the construction of phenotypes in this study may have impacted our ability to detect associations. For instance, individuals with only a single exposure were included to both morphine and fentanyl in the phenotype groups, potentially masking signals in the association analysis. Future research could refine phenotype definitions by excluding these single-exposure individuals or using them as control. Such adjustments, however, must balance the trade-off between improved phenotype specificity and reduced statistical power due to smaller sample sizes. A more refined approach to phenotype selection, such as the method used by Song et al., 2024, which categorized individuals based on the count of opioid prescriptions, could also be applied in our study [19].

Additionally, incorporating drug-switching behaviours into phenotype definitions could enhance the analysis. For instance, distinguishing individuals who switched from morphine to fentanyl from those who switched in the opposite direction could yield more precise associations with genetic variants.

Exploring quantitative phenotypes presents an additional opportunity for improvement. Metrics such as treatment duration, exposure counts, and gaps between exposures could be used to create a morphine adherence score. Such a score would better capture an individual's opioid usage patterns and facilitate the identification of genetic variants influencing morphine efficacy and adherence. Furthermore, using data from other biobanks, such as All of Us [20] and FinnGen [21], could increase the sample size and enhance statistical power.

Our focus on the OPRM1 gene, while driven by its central role in encoding the μ-opioid receptor, its direct involvement in opioid binding and efficacy, and the availability of additional experimentally informed mask possibilities, may have limited the scope of our findings. Expanding the analysis to include other genes implicated in opioid pharmacodynamics and pharmacokinetics could uncover additional insights. For instance, genes like *CYP2D6* (opioid metabolism), *CYP3A4* and *CYP3A5* (involved in opioid clearance and variability in drug response), *UGT2B7* (morphine clearance), *ABCB1* (drug transport), *COMT* (pain modulation), and *DRD2* (implicated in pain and reward pathways) could provide a more comprehensive understanding of genetic factors influencing opioid efficacy, metabolism, and adverse effects [1], [22].

Moreover, using alternative functional masks, such as those emphasizing loss-of-function variants or other relevant annotations, could provide a better understanding of the genetic basis of opioid use.

### Conclusion

This study leveraged the UK Biobank dataset to explore the genetic determinants of opioid use, focusing on the *OPRM1* gene. Despite observing no statistically significant associations, this study lays the foundation for future pharmacogenomic investigations into other drugs and prescription phenotypes, providing a framework for exploring the genetic underpinnings of medication response and usage patterns. The findings highlight potential trends linking genetic variants to opioid efficacy and switching behaviours between morphine and fentanyl. Future studies could consider redefining phenotypes to account for single-exposure individuals, incorporating drug-switching patterns, and utilizing quantitative measures such as morphine adherence scores. Expanding the scope of genetic analysis beyond the *OPRM1* gene and employing alternative functional masks could further enhance the understanding of opioid efficacy and adherence. This work provides a foundation for advancing precision medicine approaches to optimize opioid therapy and minimize adverse effects.

# References


[1] L. Magarbeh, I. Gorbovskaya, B. Le Foll, R. Jhirad, and D. J. Müller, "Reviewing pharmacogenetics to advance precision medicine for opioids," *Biomedicine & Pharmacotherapy*, vol. 142, p. 112060, Oct. 2021, doi: 10.1016/J.BIOPHA.2021.112060.

[2] H. C. S. Chan, D. McCarthy, J. Li, K. Palczewski, and S. Yuan, "Designing Safer Analgesics via μ-Opioid Receptor Pathways," *Trends Pharmacol Sci*, vol. 38, no. 11, pp. 1016–1037, Nov. 2017, doi: 10.1016/J.TIPS.2017.08.004.

[3] M. Benjeddou and A. M. Peiró, "Pharmacogenomics and Prescription Opioid Use," *Pharmacogenomics*, vol. 22, no. 4, pp. 235–245, Mar. 2021, doi: 10.2217/PGS-2020-0032.

[4] A. K. Paul *et al.*, "Opioid Analgesia and Opioid-Induced Adverse Effects: A Review," *Pharmaceuticals 2021, Vol. 14, Page 1091*, vol. 14, no. 11, p. 1091, Oct. 2021, doi: 10.3390/PH14111091.

[5] M. M. Taqi, M. Faisal, and H. Zaman, “ OPRM1 A118G Polymorphisms and Its Role in Opioid Addiction: Implication on Severity and Treatment Approaches ,” *Pharmgenomics Pers Med*, vol. 12, pp. 361–368, Nov. 2019, doi: 10.2147/PGPM.S198654.

[6] R. C. Crist and W. H. Berrettini, “Pharmacogenetics of OPRM1,” *Pharmacol Biochem Behav*, vol. 123, pp. 25–33, Aug. 2014, doi: 10.1016/J.PBB.2013.10.018.

[7] C. Bond *et al.*, “Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction,” *Proc Natl Acad Sci U S A*, vol. 95, no. 16, pp. 9608–9613, Aug. 1998, doi: 10.1073/PNAS.95.16.9608.

[8] M. D. Thompson *et al.*, “G protein-coupled receptor (GPCR) pharmacogenomics,” *Crit Rev Clin Lab Sci*, Nov. 2024, doi: 10.1080/10408363.2024.2358304.

[9] C. Sudlow *et al.*, “UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age,” *PLoS Med*, vol. 12, no. 3, Mar. 2015, doi: 10.1371/JOURNAL.PMED.1001779.

[10] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, “Spark: Cluster Computing with Working Sets,” 2010.

[11] V. Papez *et al.*, “Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond,” *Journal of the American Medical Informatics Association*, vol. 30, no. 1, pp. 103–111, Dec. 2022, doi: 10.1093/JAMIA/OCAC203.

[12] “OMOP CDM v5.3.” Accessed: Jan. 15, 2025. [Online]. Available: https://ohdsi.github.io/CommonDataModel/cdm53.html

[13] J. Mbatchou *et al.*, “Computationally efficient whole-genome regression for quantitative and binary traits,” *Nature Genetics 2021 53:7*, vol. 53, no. 7, pp. 1097–1103, May 2021, doi: 10.1038/s41588-021-00870-7.

[14] S. Lee, G. R. Abecasis, M. Boehnke, and X. Lin, “Rare-Variant Association Analysis: Study Designs and Statistical Tests,” *The American Journal of Human Genetics*, vol. 95, no. 1, pp. 5–23, Jul. 2014, doi: 10.1016/J.AJHG.2014.06.009.

[15] J. D. Backman *et al.*, “Exome sequencing and analysis of 454,787 UK Biobank participants,” *Nature 2021 599:7886*, vol. 599, no. 7886, pp. 628–634, Oct. 2021, doi: 10.1038/s41586-021-04103-z.

[16] H. S. Kizilkaya *et al.*, “Characterization of genetic variants of GIPR reveals a contribution of β-arrestin to metabolic phenotypes,” *Nature Metabolism 2024 6:7*, vol. 6, no. 7, pp. 1268–1281, Jun. 2024, doi: 10.1038/s42255-024-01061-4.

[17] “GENCODE - Human Release 47.” Accessed: Jan. 15, 2025. [Online]. Available: https://www.gencodegenes.org/human/

[18] W. McLaren *et al.*, “The Ensembl Variant Effect Predictor,” *Genome Biol*, vol. 17, no. 1, pp. 1–14, Jun. 2016, doi: 10.1186/S13059-016-0974-4/TABLES/8.

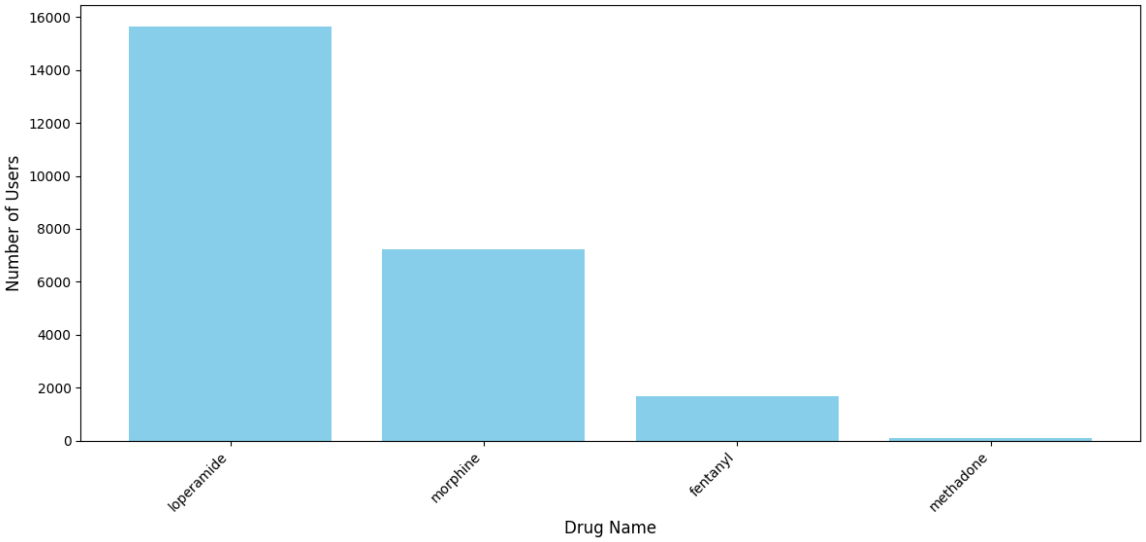
[19] W. Song *et al.*, “A genome-wide Association study of the Count of Codeine prescriptions,” *Sci Rep*, vol. 14, no. 1, p. 22780, Dec. 2024, doi: 10.1038/S41598-024-73925-4.

[20] A. G. Bick *et al.*, “Genomic data in the All of Us Research Program,” *Nature 2024 627:8003*, vol. 627, no. 8003, pp. 340–346, Feb. 2024, doi: 10.1038/s41586-023-06957-x.

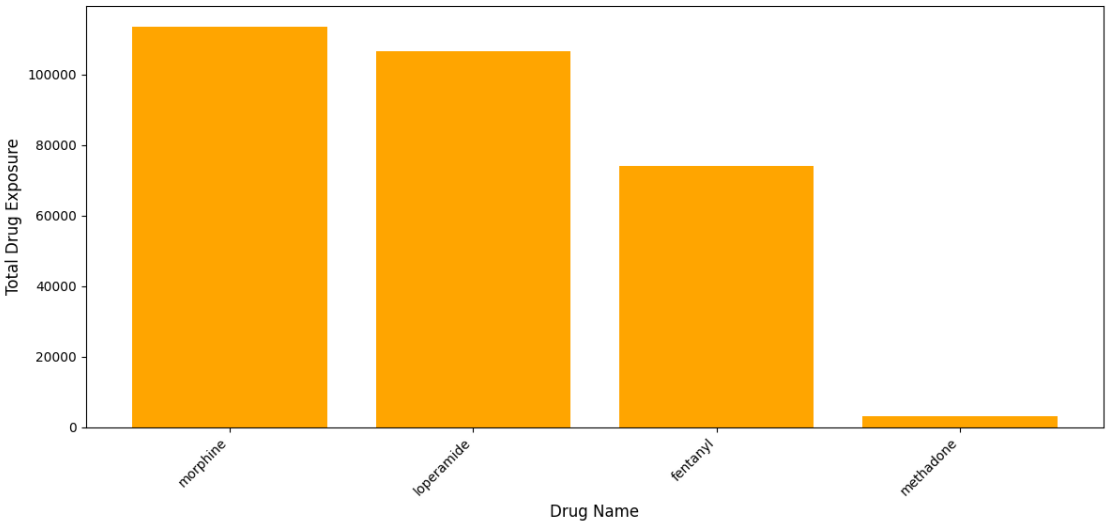
[21] M. I. Kurki *et al.*, “FinnGen provides genetic insights from a well-phenotyped isolated population,” *Nature 2023 613:7944*, vol. 613, no. 7944, pp. 508–518, Jan. 2023, doi: 10.1038/s41586-022-05473-8.

[22] A. K. Wong, A. A. Somogyi, J. Rubio, and J. Philip, “The Role of Pharmacogenomics in Opioid Prescribing,” *Curr Treat Options Oncol*, vol. 23, no. 10, pp. 1353–1369, Oct. 2022, doi: 10.1007/S11864-022-01010-X.

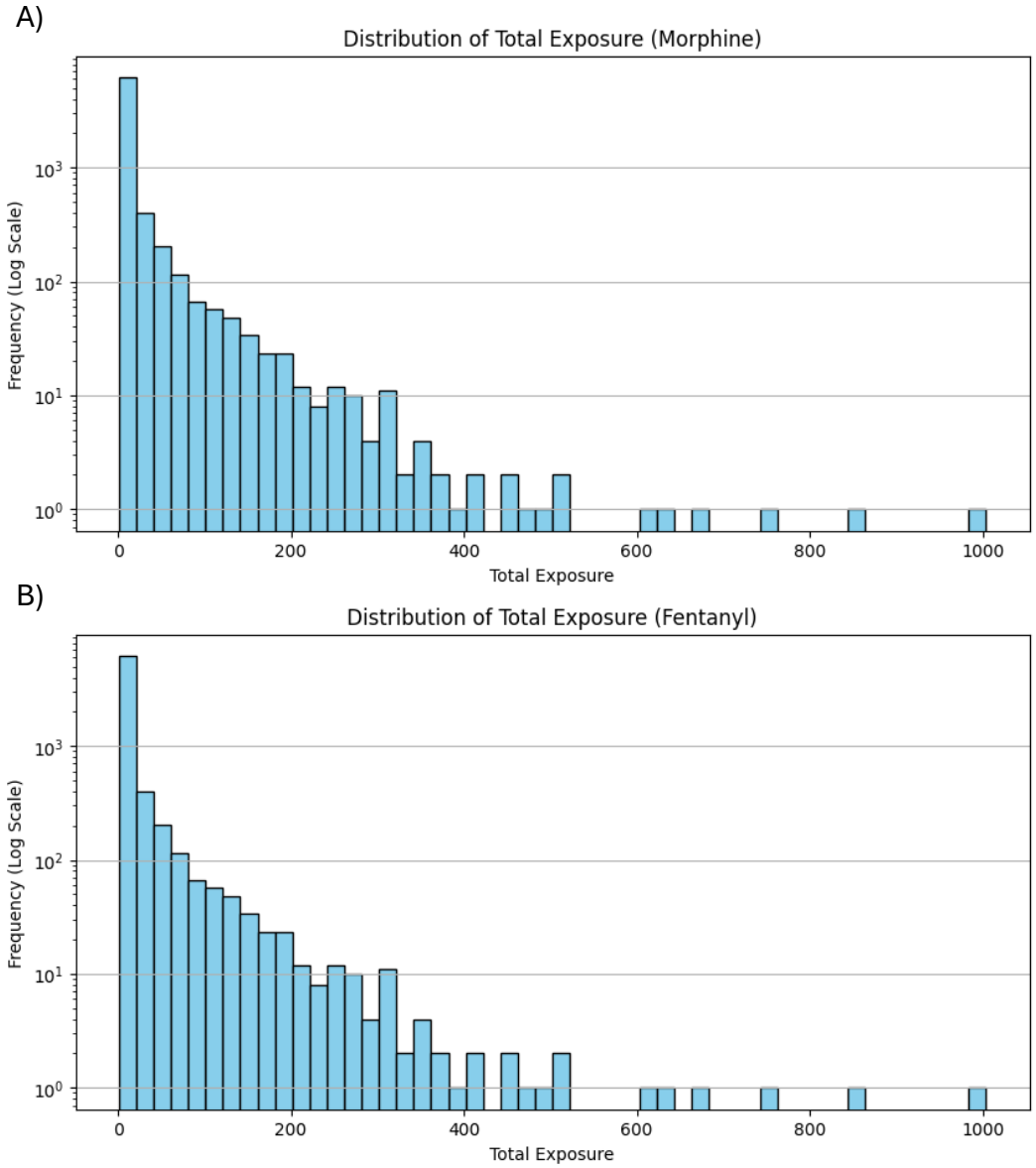
# Supplemental Material



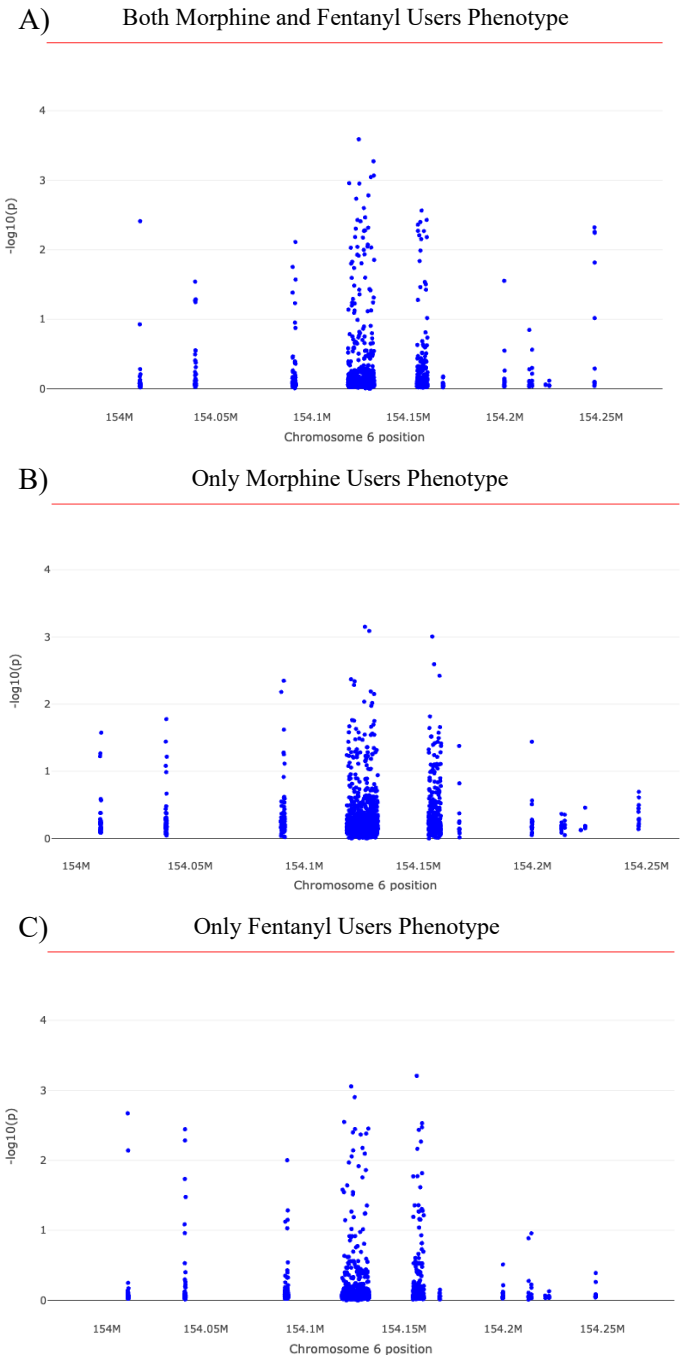

**Supplementary Figure 1: Histogram of drug usage.**
This histogram illustrates the number of individuals prescribed each opioid in the UK Biobank dataset. The x-axis represents the drug names—loperamide, morphine, fentanyl, and methadone—while the y-axis indicates the number of users.




**Supplementary Figure 2: Histogram of total exposure per drug.**
This histogram illustrates the total number of exposures for each opioid across the cohort in the UK Biobank dataset. The x-axis lists the drug names—morphine, loperamide, fentanyl, and methadone—while the y-axis represents the total drug exposures.




**Supplementary Figure 3: Distribution of Total Exposure for Morphine (A) and Fentanyl (B)**
This figure presents histograms illustrating the distribution of total exposures to morphine (A) and fentanyl (B) across individuals in the UK Biobank dataset. The x-axis represents the total exposure counts, while the y-axis (logarithmic scale) indicates the frequency of individuals within each exposure range.




**Supplementary Figure 4: Manhattan plots of single-variant association testing results for *OPRM1* variants across three phenotypes.**

This figure presents Manhattan plots summarizing the results of single-variant association testing specifically for *OPRM1* variants across three opioid-related phenotypes: A) Both morphine and fentanyl users, B) Only morphine users, and C) Only fentanyl users. Each plot displays the chromosomal position of variants on the x-axis and the negative log10-transformed p-values (-log10(p)) on the y-axis, representing the statistical significance of the associations. The red horizontal line indicates the Bonferroni-corrected significance threshold, accounting for the 4,756 variants tested. Variants exceeding this threshold would be considered statistically significant.