

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1JP97>

Title: Estudo de análise de associação genômica ampla (GWAS) de fenótipos de sono e ritmos circadianos na Coorte de Baependi

Creator: Daniel Vartanian - **ORCID:** [0000-0001-7782-759X](https://orcid.org/0000-0001-7782-759X)

Affiliation: Universidade de São Paulo (www5.usp.br)

Principal Investigator: Mario Pedrazzoli, Ana Amélia Benedito-Silva

Data Manager: Daniel Vartanian, Ana Amélia Benedito-Silva

Project Administrator: Mario Pedrazzoli

Funder: São Paulo Research Foundation (fapesp.br)

Funding opportunity number: 56445

Template: Template USP - Mínimo

Project abstract:

O sono é essencial à vida. As consequências da privação de sono crônica são devastadores para a saúde, mas silenciosas; obesidade, problemas cardiovasculares, câncer, doenças psiquiátricas, principalmente distúrbios de humor, estão todos associados à privação de sono. A organização social contemporânea, especialmente nos grandes centros urbanos, estimula um estilo de vida no qual se dorme tarde e se acorda cedo, de forma que em geral as pessoas dormem menos do que gostariam dormir. Este conjunto de fatos torna a questão do sono um problema de saúde pública, o qual merece investimentos em pesquisa. O objeto de pesquisa do presente projeto é buscar entender, do ponto de vista genético e cultural, as interações entre o ambiente contemporâneo nas

idades e o sono (como processo evolutivo adaptativo com uma história na natureza). Realizaremos a pesquisa na cidade de Baependi, localizada no estado de Minas Gerais, Brasil. O Projeto se trata de um estudo genético epidemiológico sobre estilo de vida, fenótipos de sono e ritmos circadianos na população. O cenário experimental a ser explorado é uma coorte de aproximadamente 2 mil pessoas. Essa população será explorada usando tecnologias genéticas (plataforma de Genome-wide SNP 31 Human Array 6.0 - Genechip 6.0), tecnologia de aquisição do ciclo sono/vigília (actimetria com sensores para temperatura e para luz espectral).

Start date: 02-01-2020

End date: 02-01-2022

Last modified: 09-14-2023

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Estudo de análise de associação genômica ampla (GWAS) de fenótipos de sono e ritmos circadianos na Coorte de Baependi - Descrição dos Dados e Metadados produzidos pelo projeto

Descrição dos dados e metadados produzidos

Que dados serão coletados ou criados?

O projeto prevê a coleta de dados pessoais, actimétricos/actigráficos e genotípicos da amostra correspondente à coorte Corações de Baependi. Trata-se de uma amostra da população de Baependi, uma cidade em área rural (752 km², 18.307 habitantes no Censo de 2010), localizada no estado de Minas Gerais, Brasil (21,95 °S; 44,88 °W).

Os dados pessoais contemplam dados de registro na coorte (*e.g.* id do participante), identificação (*e.g.* nome completo), contato (*e.g.* telefone, e-mail) e sobre hábitos do participante (*e.g.* dias de estudo e/ou trabalho). O volume desses dados é pequeno (cerca de 20 dados/indivíduo) e tem como principal finalidade o controle e descrição da amostra.

Durante o registro actimétrico serão coletadas séries temporais (data e hora) em intervalos (épocas) de 60 segundos com duração de aproximadamente 15 dias consecutivos por indivíduo. Essas séries serão associadas aos seguintes dados:

- (1) atividade motora, coletada por meio de um acelerômetro triaxial cuja saída de dados é pré-processada pelo aparelho e entregue nos formatos PIM (*Proportional Integration Mode*), PIM normalizado, TAT (*Time Above Threshold*), TAT normalizado, ZCM (*Zero Crossing Mode*) e ZCM normalizado;
- (2) exposição à luz ambiente (em lux e microwatts/metro quadrado) e luz nos espectros vermelho, verde, azul e infravermelho (em microwatts/centímetro quadrado), coletado por meio de um sensor RGB-IR;
- (3) temperatura do pulso e do ambiente (em Celsius), coletado por meio de sensores térmicos;
- (4) dados de evento, variável binária que indica a ocorrência de eventos durante a coleta, conforme orientação dos pesquisadores, coletado por meio de um botão na superfície externa do actímetro.

Cada registro actimétrico, nas configurações mencionadas, contém cerca de 1.000.000 de dados por indivíduo.

Será realizada a genotipagem de novos indivíduos incluídos no acompanhamento, uma vez que a coorte iniciou uma nova onda de coleta de dados. Adicionalmente, será também realizada a

genotipagem do polimorfismo VNTR do gene PER3.

A partir dos dados actimétricos serão gerados outros conjuntos de dados relacionados a suas análises, como pontuação de sono (classificação de estado sono-vigília), estatísticas relacionadas ao sono e estatísticas relacionadas ao ritmo circadiano. O mesmo irá ocorrer para os dados genotípicos, com a criação de novos conjuntos de dados a partir do cruzamento de dados com bancos de dados genéticos.

A coorte conta com dados previamente coletados dos participantes, como dados antropométricos, bioquímicos e genotípicos. Esses dados podem ser utilizados durante a pesquisa, mas não serão aqui detalhados por não pertencerem ao escopo deste documento.

Como os dados serão coletados ou criados

Os dados actimétricos serão coletados por meio de um aparelho de monitoramento compacto, semelhante a um relógio, denominado actímetro ou actígrafo, posicionado no punho não dominante do voluntário. O modelo de actímetro utilizado será o ActTrust 1 da empresa Condor Instruments.

Os dados genotípicos serão coletados a partir de amostras de DNA provenientes de leucócitos, oriundos de amostras de sangue. A plataforma de genotipagem adotada foi a Genome-wide SNP 31 Human Array 6.0 (Genechip 6.0), comercialmente disponível pela empresa Affymetrix, capaz de descrever mais de 900.000 marcadores genéticos do tipo SNP espalhados pelo genoma humano.

A criação e processamento dos dados será realizada utilizando a linguagem de programação R junto ao IDE (*Integrated Development Environment*) RStudio, podendo contar também com o auxílio dos softwares *web-based* Google Sheets e Google Forms para alguns procedimentos. Estes últimos serão utilizados a partir do serviço Google Suite contratado pela Universidade de São Paulo (USP).

A organização e estrutura das pastas e arquivos seguirá as convenções utilizadas para a criação de pacotes na linguagem de programação R (*cf.* <https://r-pkgs.org/>) e pelo guia de estilo tidyverse (*cf.* <https://style.tidyverse.org/>). A estrutura dos dados processados seguirá de acordo com um *Entity-relationship model* previamente elaborado.

Cada indivíduo que compõe a coorte tem um identificador único (ID) associado. Para fins de controle e organização, os arquivos de cada indivíduo serão nomeados com o prefixo "id_" mais o número de ID. Os dados serão separados em pastas de acordo com seu tipo.

O versionamento dos dados será realizado pela tecnologia Git, a partir de um repositório privado na nuvem hospedado na página do Grupo Interdisciplinar de Pesquisa em Sono (GIPSO) na plataforma GitHub.

Como medida para garantir a qualidade e reprodutibilidade dos dados, todos os conjuntos de dados

serão processados de acordo com as fases da cadeia de valor estatístico proposta por Mark van der Loo e Edwin de Jonge (*Raw data > Input data > Valid data > Statistics > Output*). Desta forma, os dados *raw* são mantidos seguros e sem alteração e é possível reproduzir todos os dados criados a partir dos algoritmos desenvolvidos.

Os dados provenientes de análises laboratoriais serão criados de acordo com protocolos explícitos. O banco de dados final do projeto conterà também os dados de controle de cada procedimento.
