



ESCOLA BAHIANA DE MEDICINA E SAÚDE PÚBLICA
CURSO BIOMEDICINA

JOYCE KAROLINE DA SILVA

SARA: DESENVOLVIMENTO E VALIDAÇÃO DE FLUXO DE ANÁLISES
SEMIAUTOMÁTICO PARA ESTUDOS DE EXPRESSÃO GÊNICA

SALVADOR – BA

2022

JOYCE KAROLINE DA SILVA

SARA: DESENVOLVIMENTO E VALIDAÇÃO DE UM FLUXO DE ANÁLISES
SEMIAUTOMÁTICO PARA ESTUDOS DE EXPRESSÃO GÊNICA

Trabalho de Conclusão de Curso apresentado à Escola Bahiana de Medicina e Saúde Pública, como parte dos requisitos para obtenção do título de Bacharel em Biomedicina.

Orientadora: M.Sc. Laise Eduarda Paixão de Moraes

SALVADOR – BA
2022

AGRADECIMENTOS

Gostaria que a vivência de faculdade tivesse sido mais plena, que os últimos anos tivessem sido outros, infelizmente não podemos escolher muito da vida, agradeço infinitamente não ter perdido alguém que amo nesse período complexo de pandemia. Não poderia deixar isto passar.

Entendo a sorte de, entre tantas possibilidades, estar numa família que me aceita, me ama e me dá todo suporte possível dentre as limitações que são colocadas. Sei que só pude chegar a este momento por ter pessoas que acreditaram em mim e se dedicaram ao meu sonho, mesmo sem entender direito esse caminho, em especial minha mãe que se dedicou a esse sonho tanto quanto eu e suportou meus conflitos quando nem mesmo eu conseguia.

Agradeço a todos os meus amigos por entenderem meu estresse, cansaço, exaustão e ainda me amarem mesmo com todas as faltas.

Por fim, este trabalho só pode existir graças ao grupo Khouriosos, obrigada.

SUMÁRIO

1. ARTIGO CIENTÍFICO.....	5
2. PROPOSTA DE SUBMISSÃO.....	24
3. ANEXOS.....	38

1. ARTIGO CIENTÍFICO

TÍTULO DO MANUSCRITO

SARA: Desenvolvimento e validação de um fluxo de análises semiautomático para estudos de expressão gênica para não bioinformatas

TÍTULO RESUMIDO

SARA: Fluxo semiautomático de análises de RNA-Seq

Joyce Silva^{1,2}, Laise de Moraes^{2,3}

¹Escola Bahiana de Medicina e Saúde Pública, Salvador, BA, Brasil

²Fundação Oswaldo Cruz. Instituto Gonçalo Moniz. Laboratório de Enfermidades Infecciosas Transmitidas por Vetores. Salvador, BA, Brasil

³Universidade Federal da Bahia. Faculdade de Medicina. Programa de Pós-graduação em Ciências da Saúde. Salvador, BA, Brasil

Autor para correspondência: Joyce Karoline da Silva

Endereço: R. Silveira Martins, Número 100

Bairro: Cabula

Salvador – BA

Cep: 41150-100

E-mail: joyce.karol@hotmail.com

RESUMO

Com os ensaios de sequenciamento de nova geração (*Next Generation Sequencing – NGS*) e o aumento da capacidade de processamento e armazenamento dos computadores, vem crescendo a quantidade de dados gerados por RNA-Seq para avaliação do transcriptoma. Esta abordagem avalia a expressão gênica possibilitando a identificação detalhada de isoformas de RNA, variações de nucleotídeos, fusão gênica e mutações. As análises destes dados gerados pelo RNA-Seq exigem uma grande infraestrutura computacional e conhecimentos de programação e das diferentes ferramentas de análise. Para este estudo, a proposta é gerar um fluxo de análise (*pipeline*) capaz de unificar as ferramentas de análise de transcriptoma para que seja aplicável a uma quantidade relevante de abordagens de dados e ser usado por não bioinformatas. Este *pipeline* semiautomatizado foi escrito em programação Shell e R, organizado em subcomandos, os quais permitem que o usuário defina os programas e informe os grupos de interesse através de um arquivo separado por vírgula. O processamento semiautomatizado inclui: (i) controle de qualidade das sequências (FastQC); (ii) sumarização dos dados de qualidade (MultiQC); (iii) filtragem dos dados de baixa qualidade (para escolher Trimmomatic ou Fastp); (iv) mapeamento com o genoma referência (Bowtie2, STAR ou Hisat2); (v) contagem dos transcritos (featureCounts); (vi) determinação do perfil de genes diferencialmente expressos (DESeq2). Este fluxo de análises está armazenado em um repositório de códigos aberto, o qual permite fácil implementação em computadores, *notebooks* ou servidores. O SARA (*Semi Automatic RNA-Seq Analysis*) é um fluxo de trabalho simples que reúne os *softwares* de análise para RNA-Seq e capaz de ser utilizado em diversas aplicações e utilizado por pessoas com pouco conhecimento em programação.

INTRODUÇÃO

Os avanços das metodologias de sequenciamento de nova geração (*Next Generation Sequencing – NGS*), adjacente ao aumento da capacidade de processamento e armazenamento dos computadores, possibilitou a produção em larga escala de dados relacionados à abundância de proteínas e metabólitos, interação proteína-proteína¹ e sequenciamento de RNA (RNA-Seq)².

O RNA-Seq é uma metodologia versátil, podendo ser aplicada para estudos de identificação detalhada de isoformas de RNA, variações de nucleotídeos³, fusão gênica e mutações⁴. O processamento dos dados gerados a partir do RNA-Seq envolve etapas como: avaliação da qualidade das sequências obtidas, corte de partes e/ou filtragem das sequências de baixa qualidade, alinhamento com o genoma referência, contagem dos transcritos, determinação do perfil de genes diferencialmente expressos (*Differential Expression Genes – DEG*) para cada um dos grupos (controle e experimental) e enriquecimento de vias⁵.

É possível encontrar disponível muitos tutoriais e fluxos de trabalhos para análise de RNA-Seq^{6,7}, no entanto, desafios como a pouca familiaridade com o uso de programas baseado em linhas de comando, a falta ou o pouco conhecimento em diferentes linguagens de programação e recursos humanos não disponíveis nestes quesitos, podem contribuir para uma maior dificuldade em aplicar o uso do NGS na rotina em grupos de pesquisa com pouco recurso disponível⁸.

Nesse cenário, este estudo teve como objetivo o desenvolvimento de um fluxo de análises de processamento de RNA-Seq para pesquisadores não bioinformatas, o qual fosse de fácil execução e reprodutibilidade. E neste fluxo, denominado SARA (**S**emi-**A**utomatic **R**NA-Seq **A**nalysis), nós implementamos ferramentas até a etapa da determinação dos perfis de genes diferencialmente expressos.

METODOLOGIA

Implementação e instalação

O fluxo de análises SARA foi desenvolvido utilizando a linguagem Shell para o núcleo do código e implementação de todas as etapas, com exceção da diferenciação de expressão gênica, a qual foi escrita em R. Para a instalação e gerenciamento das dependências, foram utilizados o Conda (<https://conda.io>) e Mamba (<https://mamba.readthedocs.io>); e para um dos programas, foi utilizado o repositório de códigos fonte SourceForge (<https://sourceforge.net>) para a instalação.

Todos os processos presentes no SARA foram implementados para funcionar em segundo plano, desta forma, evitando que a análise seja interrompida. Além disto, o SARA também integra para cada fase, comandos condicionais e variáveis para verificação se a expressão lógica é verdadeira, e assim, o conjunto de comandos relacionado à etapa é executado. Com estes comandos, o SARA permite evitar repetições de processos previamente completos.

O SARA está disponível na plataforma GitHub, no endereço <https://github.com/khourious/sara>, e nele, constam as instruções necessárias para a instalação. É necessário um computador com o sistema operacional baseado em Linux ou o subsistema do Linux (*Windows Subsystem for Linux – WSL*) implementado ao Windows versão 10 ou 11. De forma breve, a instalação consiste em: clonar o repositório, entrar na pasta criado do repositório no computador, dar as permissões necessárias ao arquivo de instalação e iniciar a instalação. Todos estes passos podem ser copiados e executados de uma vez de acordo com as instruções no site do SARA (Figura 1A).

Utilização do SARA

O SARA requer um arquivo de entrada em formato de texto separado por vírgulas (*Comma-separated values – CSV*). Neste arquivo, deve constar na primeira linha as ferramentas e o genoma de referência que o usuário deseja utilizar seguindo a ordem: trimagem, referência, mapeamento, contagem, análise da expressão diferencial. O controle de qualidade é realizado pelo FastQC⁹ e

sumarizados pelo MultiQC¹⁰, não sendo necessário informar a ferramenta para estas etapas.

A partir da segunda linha, devem estar listados os nomes das sequências e a identificação do grupo de análise: EXP para as amostras do grupo experimental e CTRL para as amostras pertencentes ao grupo controle, a ordem dos grupos deve ser mantida (Figura 1B). É importante que não tenha espaços, nem linhas em branco nos nomes dentro do arquivo CSV para que não ocorra quebra do tratamento dos dados. Este arquivo é de extrema importância, uma vez que o SARA interpreta as decisões do usuário de acordo com este arquivo.

Além disto, o SARA requer que os dados brutos para análise estejam dispostos dentro da sua página do repositório, na pasta “raw”. O SARA permite a análise de dados de sequenciamento gerados pela plataforma Illumina *paired-end* de fita reversa, sejam dados gerados pelo usuário ou obtidos no *Sequence Read Archive (SRA)* (<https://www.ncbi.nlm.nih.gov/sra>).

Dentro do SARA, foram implementadas as ferramentas FastQC⁹ para avaliação da qualidade; MultiQC¹⁰ para sumarizar os dados de qualidade; Fastp¹¹ e o Trimmomatic¹² para a etapa de corte e filtragem das sequências de baixa qualidade; Bowtie2¹³, STAR¹⁴ e Hisat2¹⁵ para o alinhamento com genoma referência; featureCounts¹⁶ para estimar a quantidade de transcritos; e DeSeq2¹⁷ para determinar o perfil de genes diferencialmente expressos para cada um dos grupos (Figura 1C).

Como genomas referência, três versões do genoma humano foram implementados: GRCh37 e GRCh38.p13 (<https://www.gencodegenes.org/human>); e CHM13hTERT, disponibilizado pela *Maryland Bioinformatics Lab* no GitHub (<https://github.com/marbl/CHM13>). Também foi implementado 1 versão de genoma de camundongo, o GRCm39 (Release M29), disponível no GENCODE (<https://www.gencodegenes.org/mouse>) (Figura 1C).

Após os processos de instalação e organização dos dados brutos, a utilização do SARA requer apenas, na linha de comando escrever o nome do fluxo de análises acrescido do argumento “-c” e colocar o nome do arquivo CSV, caso o usuário deseje é possível definir a quantidade de processos de execução

(*Threads*) dedicados para realizar a análise adicionando “-t” seguido do número (Figura 1D).

Validação

Para a validação do SARA, utilizamos o conjunto de dados disponível no SRA de Nakaya *et. al.* (2019) (BioProject: PRJNA507472) com a finalidade de avaliar a aplicabilidade de todas as combinações de *softwares* implementados. Os testes foram conduzidos utilizando como genoma humano referência o GRCh38.p13.

O conjunto de dados consistiu em uma biblioteca de 59 amostras, geradas utilizando a plataforma Illumina HiSeq 1500. Destas sequências, 39 são de indivíduos infectados por Chikungunya (grupo EXP) e 20 de indivíduos não infectados (grupo CTRL).

As análises foram realizadas em um servidor dedicado para bioinformática com as seguintes especificações: Servidor Torre Laniaq Riigel; com processador Intel Core i9-10900H; memória RAM de 128 gigabytes (GB) DDR4; armazenamento de SSD de 500 GB e HDD de 2 terabytes (TB); e placa de vídeo NVIDIA GeForce GTX 1660 Super. Foram utilizados 12 threads para a análise dos dados.

Os resultados da determinação dos genes diferencialmente expressos do pipeline passaram por um enriquecimento de vias utilizando o Gene Set Enrichment Analysis (GSEA). Foram utilizados genes que tinham \log_2 fold-change < 2 e valor de $p < 0.05$ para identificar classes de genes e proteínas para fins de comparação com os resultados de Nakaya *et. al.* (2019).

RESULTADOS

O SARA é um pipeline que integra *softwares* de controle de qualidade, alinhamento das sequências com o genoma de referência, quantificação transcritos, normalização e DEG por meio de linha de comando tendo parâmetros restritos.

Para avaliar aplicabilidade, tempo e o espaço de armazenamento de todas as possibilidades combinatórias de softwares, utilizamos o conjunto de dados de Nakaya et. al. (2019) (Tabela I). O genoma de referência e a anotação ocuparam espaço de armazenamento de 3168 *megabytes* (MB) e 1422 MB, respectivamente.

Na trimagem, o Trimmomatic finalizou seu processamento em 3 horas, 40 minutos e 48 segundos, enquanto Fastp em 49 minutos e 35 segundos, sendo cerca de 4,5 vezes mais rápido, em questão de espaço de armazenamento, Trimmomatic utilizou 77290 MB e Fastp 102921 MB (Tabela II).

O mapeamento é a etapa mais longa e computacionalmente intensa, e ocorre em duas partes: criação de arquivos de índice (index) e o mapeamento das sequências com o genoma de referência. As ferramentas Bowtie2, STAR e Hisat2 realizaram o index em 28 minutos e 8 segundos, 38 minutos e 52 segundos, 14 minutos e 10 segundos utilizando 4205, 29136 e 4531 MB, e finalizam o mapeamento em 12 horas e 15 minutos, 4 horas e 57 minutos, e 4 horas e 1 minuto horas utilizando de espaço de armazenamento 579048, 136393 e 853356 MB, respectivamente.

A contagem dos transcritos feita com o FeatureCounts utilizou 48780 MB de espaço e foi executada em 1 hora e 20 minutos, enquanto o DEG utilizou 21744 MB e operacionaliza em 6 minutos (Tabela II). O resultado da determinação dos genes diferencialmente expressos é entregue em uma tabela em formato CSV. O DESeq2 inclui a visualização dos resultados com diversos gráficos auxiliando o usuário a entender os dados obtidos e garantir que a análise foi realizada apropriadamente (Figura 2).

Realizando o enriquecimento de vias com GSEA com a lista de genes identificados no SARA foram encontrado regulação positiva de genes

associados a quimosinas e moléculas inflamatórias em células mieloides, imunidade e sensibilidade viral, assinatura antiviral IFN, enriquecimento na ativação de células dendríticas e monócitos, ativação na superfície da célula dendrítica, transporte transmembrana, divisão e proliferação celular, resposta imune inata em mucosas, transdução de sinal, processamento e apresentação de antígeno de macrófagos. E regulação negativa de genes associados a ativação célula T, enriquecimento em células T, assinatura de células B Naive, assinatura superfícies célula T, diferenciação célula T.

DISCUSSÃO

Com o fluxo de análises SARA é possível realizar de forma simples, o controle de qualidade, trimagem, *download* da referência e mapeamento de todas as amostras, e a partir do arquivo CSV definir subgrupos para passar pela contagem dos transcritos e DEG.

O conjunto de dados de Nakaya *et. al.* (2019) foi utilizado para testar aplicabilidade, e em confronto com os achados do SARA, comparando os resultados de DEG, em ambos há regulação positiva de genes que participam da resposta viral, o enriquecimento de vias pelo GSEA encontrou associação positiva de enriquecimento genes associados a quimosinas e moléculas inflamatórias em células mieloides, imunidade e sensibilidade viral, assinatura antiviral IFN, enriquecimento na ativação de células dendríticas e monócitos, ativação na superfície da célula dendrítica, transporte transmembrana e regulação negativa de genes associados a ativação célula T, enriquecimento em células T, assinatura de células B Naive, assinatura superfícies célula T, diferenciação célula T.

Contudo, os achados de SARA apontaram regulação positiva também para divisão e proliferação celular, resposta imune inata em mucosas, transdução de sinal, processamento e apresentação de antígeno de macrófagos e negativa para organização da matriz celular e transporte de ácido monocarboxílico, não identificando imunoglobulinas associadas ao plasma das células B, enriquecimento em células B, assinatura de superfície células B como Nakaya *et. al.* (2019) apontou em seus resultados. Para suas análises Nakaya *et. al.* (2019) utilizou Trimmomatic para remoção das sequências de baixa qualidade, o mapeamento foi realizado com o TopHat2 utilizando o genoma GRCh38.p10 e a determinação dos genes diferencialmente expressos com EdgeR.

De acordo com Costa-Silva, Domingues e Lopes (2017), a identificação de DEG por cada ferramenta apresenta resultados distintos, em seus resultados demonstraram que a ferramenta DESeq2 obteve 95% de especificidade e 81% de valor preditivo verdadeiro para performando próximo ao padrão-ouro de análise perfis de expressão gênica, transcriptase reversa seguida de reação em cadeia da polimerase quantitativa (*Reverse Transcriptase Quantitative Polymerase*

Chain Reaction – RT-qPCR), enquanto EdgeR foi determinado 94% de especificidade e 71% de valor preditivo verdadeiro, tendendo a uma menor confiabilidade. Schurd *et. al.* (2016) aponta que EdgeR tem alta sensibilidade para análise de pequenos grupos experimentais, enquanto DESeq2 demonstra ser mais apropriado para grupos experimentais maiores sendo mais pertinente para o conjunto de dados utilizado na validação, composto de 20 amostras do grupo controle e 39 no grupo experimental o que explica os resultados a mais encontrados com SARA.

Como limitações do nosso pipeline, o SARA até o momento não consegue identificar o sentido que as sequências foram geradas, sendo especificado para processar apenas fitas reversas. Além disto, há apenas o featureCounts para contagem dos transcritos e DESeq2 para realizar a etapa final da análise, e sabendo disto, planejamos implementar outras possibilidades.

SARA é voltado para pesquisadores que não tem conhecimento aprofundado com programas baseados em linhas de comando ou com pouco conhecimento em diferentes linguagens de programação para aplicar análises de NGS, assim este fluxo de análises implementa diversas ferramentas utilizando códigos-fonte em linguagem Shell e R, com este pipeline torna-se mais fácil aplicar a técnica de RNA-Seq para determinação de genes diferencialmente expressos para diversos grupos experimentais dentro do mesmo conjunto de dados, sendo um processo que requer apenas alterações no arquivo CSV. O fluxo de análises está em desenvolvimento contínuo e aberto às contribuições da comunidade, o pipeline está disponível em: <https://github.com/khourious/sara>.

REFERÊNCIAS

1. ALTAF-UL-AMIN, M. et al. Systems Biology in the Context of Big Data and Networks. *BioMed Research International*, v. 2014, n. 3, p. 1–11, 2014. DOI: 10.1155/2014/428570
2. FINOTELLO, F.; DI CAMILLO, B. Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Briefings in Functional Genomics*, 2015. DOI: 10.1093/bfpg/elu035
3. RAPAPORT, F. et al. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, v. 14, n. 9, 2013. DOI: 10.1186/gb-2013-14-9-r95
4. PARK, S. T.; KIM, J. Trends in next-generation sequencing and a new era for whole genome sequencing. *International Neurology Journal*, v. 20, p. 76–83, 2016. DOI: 10.5213/inj.1632742.371
5. PARRISH, N.; HORMOZDIARI, F.; ESKIN, E. Assembly of non-unique insertion content using next-generation sequencing. *Bioinformatics: The Impact of Accurate Quantification on Proteomic and Genetic Analysis and Research*, p. 21–40, 2014. DOI: 10.1186/1471-2105-12-S6-S3
6. LATARETU, M.; HÖLZER, M. Rnaflow: An effective and simple rna-seq differential gene expression pipeline using nextflow. *Genes*, v. 11, n. 12, p. 1–17, 2020. DOI: 10.3390/genes11121487
7. CORCHETE, L. A. et al. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Scientific Reports*, v. 10, n. 1, p. 1–15, 2020. DOI: 10.1038/s41598-020-76881-x
8. KATSANTONI, M. et al. ZARP: An automated workflow for processing of RNA-seq data. *bioRxiv*, p. 2021.11.18.469017, 2021. DOI: <https://doi.org/10.1101/2021.11.18.469017>
9. ANDREWS, S. FasnQC: a quality control tool for high throughput sequence data, 2010. Disponível em: <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>>
10. EWELS, P. et al. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, v. 32, n. 19, p. 3047–3048, 2016. DOI: 10.1093/bioinformatics/btw354
11. CHEN, S. et al. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, v. 34, n. 17, p. i884–i890, 2018. DOI: 10.1093/bioinformatics/bty560
12. BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114–2120, 2014. DOI: 10.1093/bioinformatics/btu170
13. LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, v. 9, n. 4, p. 357–359, 2012. DOI: 10.1038/nmeth.1923

14. DOBIN, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, v. 29, n. 1, p. 15–21, 2013. DOI: 10.1093/bioinformatics/bts635
15. KIM, D.; LANGMEAD, B.; SALZBERG, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, v. 12, n. 4, p. 357–360, 2015. DOI: 10.1038/nmeth.3317
16. LIAO, Y.; SMYTH, G. K.; SHI, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, v. 30, n. 7, p. 923–930, 2014. DOI: 10.1093/bioinformatics/btt656
17. LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, v. 15, n. 12, p. 550, dez. 2014. DOI: 10.1186/s13059-014-0550-8

TABELAS

Corte e filtragem das sequências	Mapeamento	Quantificação dos transcritos	Determinação genes diferencialmente expressos	Tempo	Espaço de armazenamento (MB)
Trimmomatic	Bowtie2	Featurecounts	DESeq2	21 horas e 21 minutos	686880
Fastp	Bowtie2	Featurecounts	DESeq2	18 horas e 18 minutos	661249
Trimmomatic	Hisat2	Featurecounts	DESeq2	12 horas e 52 minutos	940088
Fastp	Hisat2	Featurecounts	DESeq2	9 horas e 49 minutos	965719
Trimmomatic	STAR	Featurecounts	DESeq2	13 horas e 33 minutos	247730
Fastp	STAR	Featurecounts	DESeq2	10 horas e 30 minutos	273361

Tabela I. Combinações possíveis de softwares de corte e/ou filtragem das sequências, mapeamento, quantificação dos transcritos e DEG comparando o espaço de armazenamento ocupado e o tempo de processamento.

Trimagem				
	Espaço de armazenamento (MB)		Tempo	
Fastp		102921		49 minutos e 35 segundos
Trimmomatic		77290		3 horas, 40 minutos e 48 segundos
Mapeamento				
	Espaço de armazenamento - Index (MB)	Tempo - Index	Espaço de armazenamento (MB)	Tempo - Mapeamento
Bowtie2	4205	28 minutos e 8 segundos	579048	12 horas e 15 minutos
Hisat2	4531	14 minutos e 10 segundos	853356	4 horas e 1 minuto
STAR	29136	38 minutos e 52 segundos	136393	4 horas e 17 minutos
Contagem dos Transcritos				
	Espaço de armazenamento (MB)		Tempo	
FeatureCounts		49		1 hora e 20 minutos
DEG				
	Espaço de armazenamento (MB)		Tempo	
DESeq2		22		6 minutos

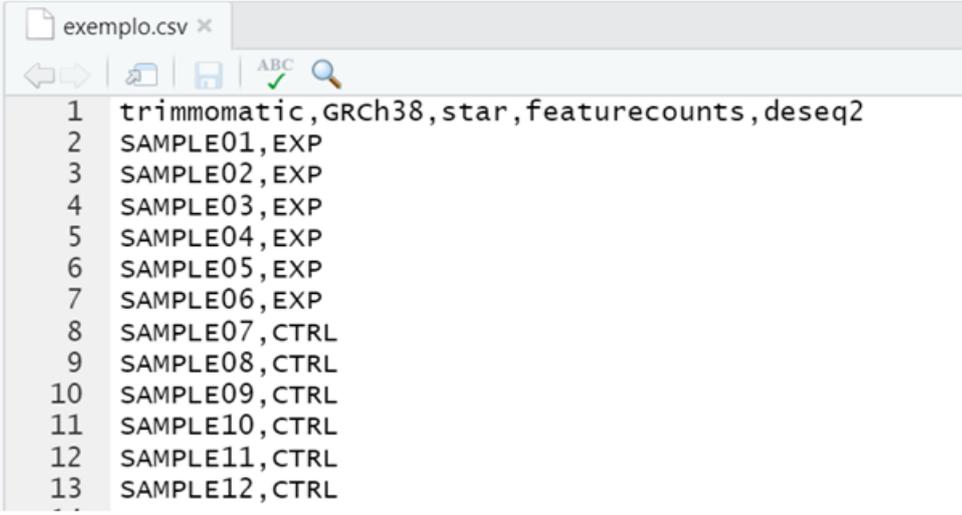
Tabela II. Comparação de espaço de armazenamento e tempo de processamento entre os softwares de mapeamento nas etapas de construção dos arquivos de índice e mapeamento das sequências e das ferramentas de contagem dos transcritos e análise de expressão diferencial, utilizando o software Trimmomatic para corte das sequências.

FIGURAS

A

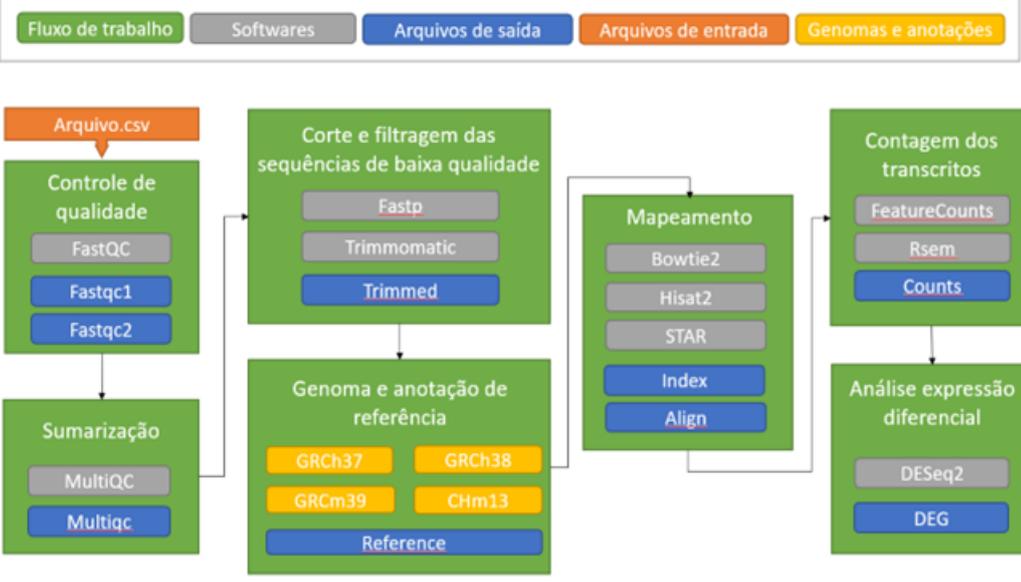
```
git clone --recursive https://github.com/khourious/sara.git; cd sara
chmod 700 -R INSTALL
bash INSTALL
```

B



```
1 trimmomatic, GRCh38, star, featurecounts, deseq2
2 SAMPLE01, EXP
3 SAMPLE02, EXP
4 SAMPLE03, EXP
5 SAMPLE04, EXP
6 SAMPLE05, EXP
7 SAMPLE06, EXP
8 SAMPLE07, CTRL
9 SAMPLE08, CTRL
10 SAMPLE09, CTRL
11 SAMPLE10, CTRL
12 SAMPLE11, CTRL
13 SAMPLE12, CTRL
```

C



Fluxo de trabalho Softwares Arquivos de saída Arquivos de entrada Genomas e anotações

Arquivo.csv

Controle de qualidade

- FastQC
- Fastqc1
- Fastqc2

Sumarização

- MultiQC
- Multiqc

Corte e filtragem das sequências de baixa qualidade

- Fastp
- Trimmomatic
- Trimmed

Genoma e anotação de referência

- GRCh37
- GRCh38
- GRCm39
- CHm13
- Reference

Mapeamento

- Bowtie2
- Hisat2
- STAR
- Index
- Align

Contagem dos transcritos

- FeatureCounts
- Rsem
- Counts

Análise expressão diferencial

- DESeq2
- DEG

D

```
c:\> sara -c exemplo -t 12 █
```

Figura 1. Instalação, arquivo de entrada e fluxo de trabalho do SARA. (A) Instruções de instalação disponibilizado no site do SARA. (B) Exemplo do arquivo de entrada que SARA necessita, em formato CSV, onde na primeira linha

consta as ferramentas e genoma de referência escolhido seguindo a ordem: trimagem, referência, mapeamento, contagem, DEG. Da segunda linha em diante estão listados os nomes das amostras e o grupo experimental ao qual pertencem, EXP para grupo experimental, sendo este sempre o primeiro grupo a ser apresentado, e CTRL para grupo controle. (C) Fluxo de trabalho e diretórios dos arquivos de saída do SARA. Em verde, as etapas que são realizadas no pipeline; em laranja o arquivo de entrada que vai indicar as amostras a serem processadas, as ferramentas e genoma de referência que o usuário selecionou; em azul, os diretórios para os arquivos de saída de cada etapa; em cinza, as ferramentas disponibilizadas; em amarelo os genomas e anotações de referências disponíveis no repositório. (D) Comandos para utilização do SARA contendo o argumento para identificar o arquivo CSV para análise dos dados brutos e, optativo, o usuário pode definir a quantidade de processos de execução dedicados para o fluxo de análises.

A

A	B	C	D	E	F	G	H	I	J	K
1	ENSG00000162772.17	133.62142121072	5.6142902799563	0.37938863743055	2.52720080769950	1.76171268304041e-47	483.3488222956	5.998521476	5.998521477	5.998521480
2	ENSG00000148131.17	167.607367629448	5.1388352598429	0.3454242437006	1.403731085279e-51	1.76171268304041e-47	370.407900150286	364.25303027095	364.25303027095	364.25303027095
3	ENSG00000118003.16	1933.30573050561	4.80800389405771	0.33323979787293	6.409007878011e-49	3.00184867431459e-45	2318.14008298137	2814.60870428296	4518.18035433425	2323.08950963936
4	ENSG00000186030.13	439.855651777933	8.006972854398	0.59028948552122	5.61765217300070e-48	1.74892066420521e-44	1034.65361705739	213.22793214305	161.26261020232	33.7657629390487
5	ENSG00000136689.18	1395.18517092118	4.35471306719564	0.32442911912217	2.69140520543489e-42	7.0305954883306e-39	4015.3995327486	1721.02386438903	3522.95736348174	1542.61420912167
6	ENSG00000149138.1	178.688264149	4.2006111864953	0.303334872919715	4.036761291381e-42	7.0305954883306e-39	229.710331261573	236.408300179403	538.368886652123	173.62640406216
7	ENSG00000138572.4	62.1447701394491	5.3655388248735	0.40193452636454	3.2180314838954e-42	7.0305954883306e-39	193.309528615024	40.1121852545578	300.34074046638	60.7784261228256
8	ENSG00000109860.9	1348.61761303496	4.21125087921418	0.314444838860772	3.90251321860712e-42	7.0305954883306e-39	3699.29348060158	1670.35864070558	2104.6238707595	1328.7220889684
9	ENSG00000136141.14	480.22644424999	4.5882637059867	0.346637662270917	3.6642527671234e-41	4.74867127169438e-38	1024.1252287451	1081.34006778708	1626.43444389704	81.0379014871028
10	ENSG00000128512.24	830.31470218942	3.3482170313482	0.254862837919744	5.3488593993096e-41	7.45499137451753e-38	1361.0317127482	968.603715515324	1717.93860779004	713.90527125955
11	ENSG00000181938.9	115.178631844479	4.2428304854957	0.323889819624998	1.906789825727e-40	1.8114768379117e-37	343.65837911233	397.18370403827	388.15804296399	54.990048571184
12	ENSG00000135114.13	1803.6154473563	3.7940494946629	0.29148343139658	3.4713015643551e-40	4.2657908419013e-37	3605.49507442844	3672.14500187252	413.45300412396	232.834236016184
13	ENSG00000160812.11	213.596942471385	3.8927591008549	0.307252489040107	6.0072214842477e-40	6.442513995054e-37	505.36773875881	437.85616968174	497.37048774867	67.531584089173
14	ENSG00000146212.17	448.10451278674	3.36712318600538	0.260882702994322	1.0547135206e-39	1.0491060704432e-36	832.69958025202	657.436838174749	739.42255402003	515.68973885142
15	ENSG00000130541.15	105.827095948752	3.7482038782639	0.29194889813329	4.22825270471542e-38	3.68016968973624e-35	287.03820091579	324.27830286426	418.15370366784	287.481802900191
16	ENSG00000184194.3	477.321906102929	3.1342408994451	0.245688961326252	4.2417638444958e-39	3.68016968973624e-35	987.754242424764	795.78821892488	1265.55757712342	521.82876184233
17	ENSG00000188817.13	147.468992938598	4.7704748015108	0.372120922944809	6.4027488176685e-38	6.4027488176685e-38	272.78103373118	201.40538333078	4.8236446128287	4.8236446128287
18	ENSG00000125462.6	196.70060536484	5.182272996258	0.40031307328455	7.2448005190441e-39	5.81304966020842e-36	408.405000748095	233.07200805431	386.11707056877	232.501583428074
19	ENSG00000109088.11	1044.6021448869	4.1195200294882	0.322851348975702	1.0358612328057e-38	7.5897932488324e-36	22610.1964808006	16595.4237847504	22601.58662523104	856.259649681
20	ENSG00000109088.11	740.35204797029	3.5420718624297	0.27875455050396	1.7862908908107e-38	1.21034612109941e-35	1818.45782539409	882.486075960273	1149.7208944101	1214.60378553393
21	ENSG00000137965.11	3216.28220647499	4.02145043823399	0.325913881094	2.1873340789393e-38	1.49882025959383e-35	7121.97739548932	3607.26129595913	11063.23634791485	798.11691177048
22	ENSG00000188209.6	1136.22929482099	3.4737543378681	0.29194889813329	9.866251415254e-38	6.32886781046979e-35	1363.90508188658	2222.09681818165	2389.77392031444	1087.18524717825
23	ENSG00000113311.14	1530.29763899733	4.18423893173888	0.33393875667036	2.83194810346967e-37	1.7720934237279e-34	2711.5390326862	3240.22010151029	5338.3543632832	323.18886800819
24	ENSG00000187608.10	7723.8102100002	5.351511178675	0.403282063888807	4.44200807970787e-37	2.5854787668763e-34	16756.4115393153	2784.1454581179	1524.3397142696	785.2095496831
25	ENSG00000110047.16	447.32664231522	3.3231499547124	0.29832976271822	6.79064127468887e-37	3.78700482710611e-34	1166.72705732174	849.53880338836	803.397388284	448.60260901898
26	ENSG00000110688.10	258.23864831839	3.2187319444259	0.29595386705032	1.6773382632967e-35	5.6169139388796e-33	281.395156795427	454.74551077788	923.90020596878	663.03374118985
27	ENSG00000128374.12	2484.67134520981	3.3604898217421	0.276227483948731	1.907768414547e-36	5.8189139837878e-35	6182.6782907708	9843.39051413849	7963.95327821718	1343.8783316028
28	ENSG00000143874.11	387.05622945677	3.7164741359027	0.305201599666117	1.43871134012376e-35	8.1598132052079e-33	793.45726232683	408.596523125488	413.25013074744	570.1595121247458
29	ENSG000001109088.11	247.704112165697	0.88280783062774	0.07208183065718	1.70487339212858e-34	8.18953428817772e-32	348.384800313386	296.430170883728	371.09175260254	271.09175260254
30	ENSG000001109088.11	4834.3613365863	2.8906646993883	0.22854087815978	1.94583005096868e-34	9.0429580306439e-32	10547.5327104272	5428.1230811019	10199.8867153617	6134.7620001505
31	ENSG00000122861.16	95.930181620163	5.98812263881477	0.5071258233057	3.0061377088973e-34	1.35188619138219e-31	87.6268807861695	119.814322246678	66.8118888200411	66.8118888200411
32	ENSG00000181938.9	38.1230963246693	3.899070421479	0.327356472910844	3.21713692894696e-34	1.44534847674065e-31	89.9698797448121	167.1351313206	141.68578988512	21.2242129986597
33	ENSG00000124888.14	209.6410496467	1.7724758616781	0.15180290211317	6.80776030197448e-34	2.9218930004489e-31	348.28132783452	268.807722980202	382.2557956488	183.80936616786
34	ENSG00000105480.13	2112.39621149484	3.20701114749688	0.212355512227416	7.90708838028496e-34	3.2423815671341e-31	2522.98513836188	4434.140291881323	8374.74828200979	5198.96727580805
35	ENSG00000178081.8	157.79944072799	5.20078903191521	0.44443296331808	9.8109642561959e-34	3.9478018188395e-31	432.62123879362	406.188854893253	582.96782495764	44.378964388885
36	ENSG00000120861.16	62.321182248338	4.79131020985865	0.404352938473399	1.0457674914586e-33	1.11.02686510976	111.02686510976	108.514016951804	119.547359144701	142.78164542511
37	ENSG00000136784.16	7795.88846931847	4.20662318307421	0.346807887677777	1.1787628764075e-33	4.7514883816879e-31	8055.18311010642	6258.36778884279	7898.380200779	878.028916760071

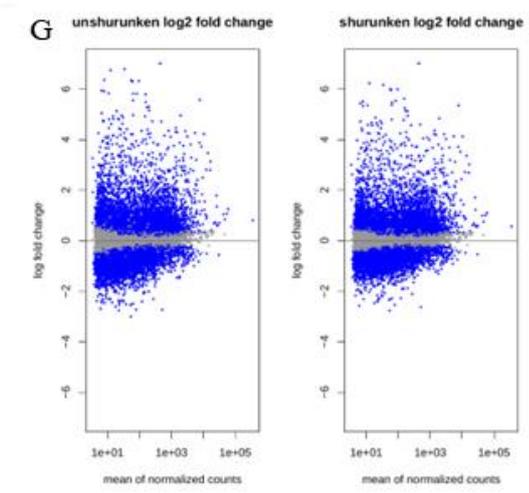
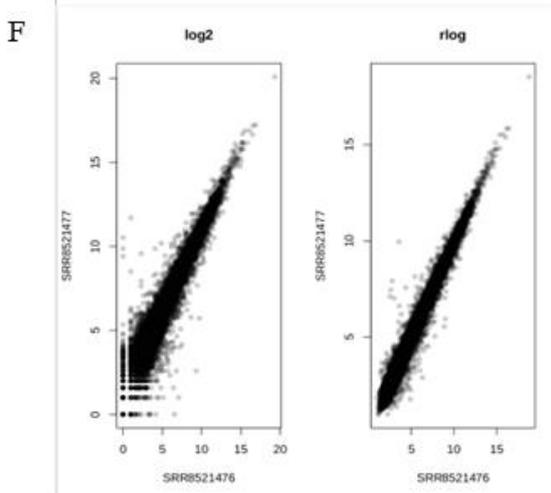
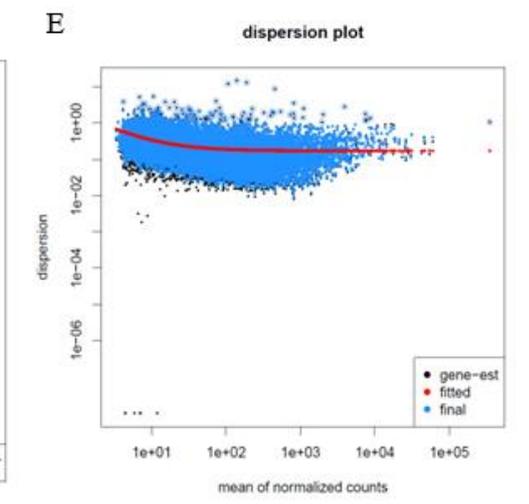
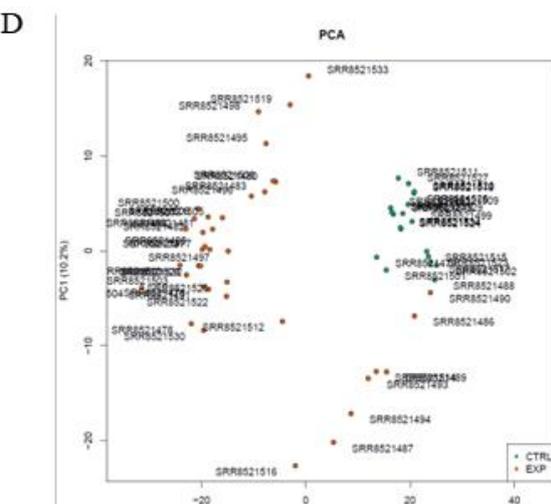
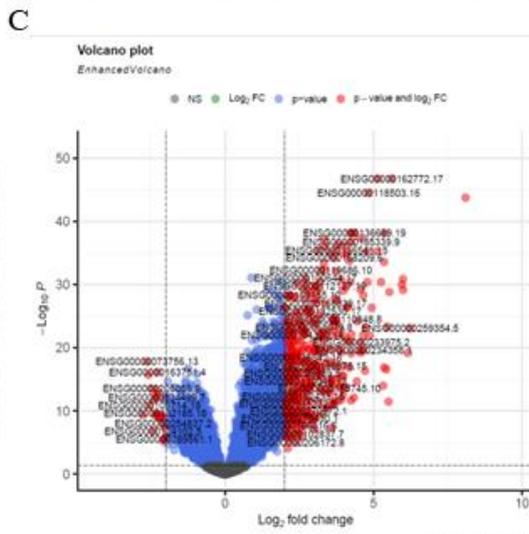
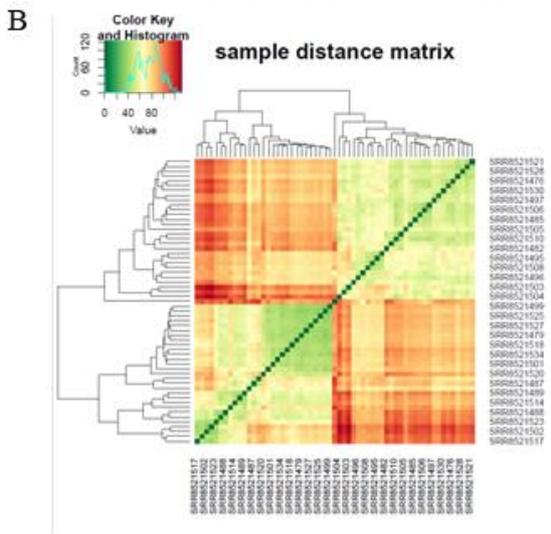


Figura 2. Resultados gerados pelo DESeq2. (A) Tabela resultados gerado em formato CSV com a determinação do perfil de genes diferencialmente expressos, identificando os genes e os valores de probabilidade de significância. (B) Mapa de calor que demonstra, baseado na coloração, agrupamento entre amostras com base na semelhança de seu padrão de expressão. (C) Gráfico de dispersão em formato de vulcão, em vermelho são os genes identificados como diferencialmente expressos; em azul e cinza foram genes que foram encontrados, mas que não tinham valores estatísticos significantes. (D) Gráfico da análise componente principal; em laranja, as amostras do grupo experimental; em verde, as amostras do grupo controle, organizadas no plano cartesiano e agrupadas segundo observações de variáveis possivelmente correlacionadas. (E) Plot de dispersão da estimativa da contagem dos transcritos; em vermelho é a tendência geral; em azul, genes que seguem a tendência central; em preto, genes distantes da tendência central. (F) Gráfico de dispersão da variância ajustado em função de \log_2 e $r\log$, respectivamente, dos genes encontrados da amostra, cada ponto preto é um gene encontrado. (G) Gráfico de dispersão das mudanças de dobra em \log_2 para a média das contagens de todas as amostras, e ao lado o mesmo gráfico com a remoção de interferências técnicas.

2. PROPOSTA DE SUBMISSÃO

2.1. Revista: Revista de Saúde Pública

2.2. Regras para Submissão:

2.2.1. Informações gerais

São aceitos manuscritos nos idiomas: português, espanhol e inglês. Artigos submetidos em português ou espanhol são traduzidos para o inglês e publicados nesses dois idiomas. Para artigos submetidos em inglês, não há tradução para o português ou espanhol.

O texto de manuscrito de pesquisa original deve seguir a estrutura conhecida como IMRD: Introdução, Métodos, Resultados e Discussão. Manuscritos baseados em pesquisa qualitativa podem ter outros formatos, admitindo-se Resultados e Discussão em uma mesma seção e Considerações Finais ou Conclusões. Outras categorias de manuscritos (revisões, comentários etc.) seguem os formatos de texto a elas apropriados.

Os estudos devem ser apresentados de forma que qualquer pesquisador interessado possa reproduzir os resultados. Para isso estimulamos o uso das seguintes recomendações, de acordo com a categoria do manuscrito submetido:

- CONSORT – checklist e fluxograma para ensaios controlados e randomizados;
- STARD – checklist e fluxograma para estudos de acurácia diagnóstica;
- MOOSE – checklist e fluxograma para metanálises e revisões sistemáticas de estudos observacionais;
- PRISMA – checklist e fluxograma para revisões sistemáticas e metanálises;
- STROBE – checklist para estudos observacionais em epidemiologia;
- RATS – checklist para estudos qualitativos.

Por menores sobre os itens exigidos para apresentação do manuscrito estão descritos de acordo com a categoria de artigos.

Como forma de avaliação da ocorrência de plágio, todos os manuscritos recebidos são submetidos à programa de detecção de similaridade entre textos.

O ORCID do primeiro autor e de todos os coautores deverá ser informado no momento da submissão dos manuscritos, na carta de apresentação.

2.2.1.1. Resolução de conflitos de interesse e violações éticas

Os editores tomarão as medidas necessárias para identificar e prevenir a publicação de artigos onde ocorra má conduta de pesquisa ou violações éticas, incluindo plágio, manipulação de citações e falsificação / fabricação de dados, ausência de autorizações pertinentes, discriminação, entre outros. As situações e alegações que chegarem ao conhecimento de editores e avaliadores serão levadas ao Comitê Editorial, que tomará as providências cabíveis, incluindo o encaminhamento a instâncias superiores da Universidade, se necessário.

2.2.2. Categorias de artigos

a) Artigos Originais

Incluem estudos observacionais, estudos experimentais ou quase-experimentais, avaliação de programas, análises de custo-efetividade, análises de decisão e estudos sobre avaliação de desempenho de testes diagnósticos para triagem populacional. Cada artigo deve conter objetivos e hipóteses claras, desenho e métodos utilizados, resultados, discussão e conclusões.

Incluem também ensaios teóricos (críticas e formulação de conhecimentos teóricos relevantes) e artigos dedicados à apresentação e discussão de aspectos metodológicos e técnicas utilizadas na pesquisa em saúde pública. Neste caso, o texto deve ser organizado em tópicos para guiar o leitor quanto aos elementos essenciais do argumento desenvolvido.

Instrumentos de aferição em pesquisas populacionais

Manuscritos abordando instrumentos de aferição podem incluir aspectos relativos ao desenvolvimento, a avaliação e à adaptação transcultural para uso em estudos populacionais, excluindo-se aqueles de aplicação clínica, que não se incluem no escopo da RSP.

Aos manuscritos de instrumentos de aferição, recomenda-se que seja apresentada uma apreciação detalhada do construto a ser avaliado, incluindo seu possível gradiente de intensidade e suas eventuais subdimensões. O desenvolvimento de novo instrumento deve estar amparado em revisão de literatura, que identifique explicitamente a insuficiência de propostas prévias e justifique a necessidade de novo instrumental.

Devem ser detalhados: a proposição, a seleção e a confecção dos itens e o emprego de estratégias para adequá-los às definições do construto, incluindo o uso de técnicas qualitativas de pesquisa (entrevistas em profundidade, grupos focais etc.), reuniões com painéis de especialistas, entre outras; o trajeto percorrido na definição

da forma de mensuração dos itens e a realização de pré-testes com seus conjuntos preliminares; e a avaliação das validades de face, conteúdo, critério, construto ou dimensional.

Análises de confiabilidade do instrumento também devem ser apresentadas e discutidas, incluindo-se medidas de consistência interna, confiabilidade teste-reteste ou concordância inter-observador. Os autores devem expor o processo de seleção do instrumento final e situá-lo em perspectiva crítica e comparativa com outros instrumentos destinados a avaliar o mesmo construto ou construtos semelhantes.

Para os manuscritos sobre adaptação transcultural de instrumentos de aferição, além de atender, de forma geral, às recomendações supracitadas, é necessário explicitar o modelo teórico norteador do processo. Os autores devem também justificar a escolha de determinado instrumento para adaptação a um contexto sociocultural específico, com base em minuciosa revisão de literatura. Finalmente, devem indicar explicitamente como e quais foram as etapas seguidas do modelo teórico de adaptação no trabalho submetido para publicação.

O instrumento de aferição deve ser incluído como anexo dos artigos submetidos.

2.2.2.1. Organização do manuscrito

Além das recomendações mencionadas, verificar as seguintes instruções de formatação:

a) Artigo original:

- Devem conter até 3.500 palavras (excluindo resumos, tabelas, figuras e referências).
- Número máximo de tabelas e figuras: 5.
- Número máximo recomendado de referências: 30.
- Resumos no formato estruturado com até 300 palavras.

b) Comunicação breve

São relatos curtos de achados que apresentam interesse para a saúde pública, mas que não comportam uma análise mais abrangente e uma discussão de maior fôlego.

Formatação: Sua apresentação deve acompanhar as mesmas normas exigidas para artigos originais.

- Devem conter até 1.500 palavras (excluindo resumos, tabelas, figuras e referências).
- Número máximo de tabelas e figuras: 1.
- Número máximo recomendado de referências: 5.

- Resumos no formato narrativo com até 100 palavras.

c) Artigos de revisão

Revisão sistemática e meta-análise – Por meio da síntese de resultados de estudos originais, quantitativos ou qualitativos, objetiva responder a uma pergunta específica e de relevância para a saúde pública. Descreve com pormenores o processo de busca dos estudos originais, os critérios utilizados para seleção daqueles que foram incluídos na revisão e os procedimentos empregados na síntese dos resultados obtidos pelos estudos revisados. Consultar:

MOOSE checklist e fluxograma para meta-análises e revisões sistemáticas de estudos observacionais;

PRISMA checklist e fluxograma para revisões sistemáticas e meta-análises.

Revisão narrativa ou crítica – Apresenta caráter descritivo-discursivo e dedica-se à apresentação compreensiva e à discussão de temas de interesse científico no campo da saúde pública. Deve apresentar formulação clara de um objeto científico de interesse, argumentação lógica, crítica teórico-metodológica dos trabalhos consultados e síntese conclusiva. Deve ser elaborada por pesquisadores com experiência no campo em questão ou por especialistas de reconhecido saber.

- Devem conter até 4.000 palavras (excluindo resumos, tabelas, figuras e referências).
- Número máximo de tabelas e figuras: 5.
- Número máximo de referências: ilimitado.
- Resumos no formato estruturado com até 300 palavras, ou narrativo com até 150 palavras.

d) Comentários

Visam a estimular a discussão, introduzir o debate e oxigenar controvérsias sobre aspectos relevantes da saúde pública. O texto deve ser organizado em tópicos ou subitens. A introdução deve destacar o assunto e sua importância. As referências citadas devem dar sustentação aos principais aspectos abordados no artigo.

- Devem conter até 2.000 palavras (excluindo resumos, tabelas, figuras e referências).
- Número máximo de tabelas e figuras: 5.
- Número máximo recomendado de referências: 30.
- Resumos no formato narrativo com até 150 palavras.

Publicamos também Cartas Ao Editor com até 600 palavras e até 5 referências.

2.2.3. Dados de identificação do manuscrito

2.2.3.1. Autoria

O conceito de autoria está baseado na contribuição substancial de cada uma das pessoas listadas como autores, no que se refere sobretudo à concepção do projeto de pesquisa, análise e interpretação dos dados, redação e revisão crítica do manuscrito. A contribuição de cada autor deve ser explicitada em declaração para esta finalidade. Não se justifica a inclusão de nome de autores cuja contribuição não se enquadre nos critérios mencionados.

2.2.3.2. Dados de identificação dos autores (cadastro)

Nome e sobrenome: O autor deve seguir o formato pelo qual já é indexado nas bases de dados e constante no ORCID.

Correspondência: Deve constar o nome e endereço do autor responsável para troca de correspondência.

Instituição: Podem ser incluídas até três hierarquias institucionais de afiliação (por exemplo: universidade, faculdade, departamento).

Coautores: Identificar os coautores do manuscrito pelo nome, sobrenome e instituição, conforme a ordem de autoria.

Financiamento da pesquisa: Se a pesquisa foi subvencionada, indicar o tipo de auxílio, o nome da agência financiadora e o respectivo número do processo.

Apresentação prévia: Tendo sido apresentado em reunião científica, indicar o nome do evento, local e ano da realização.

2.2.4. Conflito de interesses

A confiabilidade pública no processo de revisão por pares e a credibilidade de artigos publicados dependem, em parte, de como os conflitos de interesses são administrados durante a redação, revisão por pares e tomada de decisões pelos editores.

Conflitos de interesses podem surgir quando autores, revisores ou editores possuem interesses que, aparentes ou não, possam influenciar a elaboração ou avaliação de manuscritos. O conflito de interesses pode ser de natureza pessoal, comercial, política, acadêmica ou financeira.

Quando os autores submetem um manuscrito, eles são responsáveis por reconhecer e revelar conflitos financeiros ou de outra natureza que possam ter influenciado seu trabalho. Os autores devem reconhecer no manuscrito todo o apoio financeiro para o trabalho e outras conexões financeiras ou pessoais com relação à pesquisa. O relator deve revelar aos editores quaisquer conflitos de interesses que possam influir em sua opinião sobre o manuscrito e, quando couber, deve se declarar não qualificado para revisá-lo.

Se os autores não tiverem certos do que pode constituir um potencial conflito de interesses, devem contatar a secretaria editorial da RSP.

2.2.5. Declarações e documentos

Em conformidade com as diretrizes do *International Committee of Medical Journal Editors*, são solicitados alguns documentos e declarações do(s) autor(es) para a avaliação de seu manuscrito. Observe a relação dos documentos abaixo e, nos casos em que se aplique, anexe o documento ao processo. O momento em que tais documentos serão solicitados é variável:

a) Carta de apresentação

A carta deve ser assinada por todos os autores ou, ao menos, pelo primeiro autor. O ORCID de todos os autores deverá ser informado nessa carta. A carta de apresentação deve conter:

- Informações sobre os achados e as conclusões mais importantes do manuscrito e esclarecimento de seu significado para a saúde pública;
- Informação sobre a novidade do estudo e porque ele deve ser publicado nesta revista;
- Menção de até três artigos, se houver, publicados pelos autores na linha de pesquisa do manuscrito; Declaração de potenciais conflitos de interesses dos autores;
- Atestado de exclusividade da submissão do manuscrito à RSP;
- Contribuição ao manuscrito por parte de cada autor.

Segundo o critério de autoria do *International Committee of Medical Journal Editors*, autores devem contemplar todas as seguintes condições: (1) contribuir substancialmente para a concepção e planejamento, ou análise e interpretação dos dados; (2) contribuir significativamente na elaboração do

rascunho ou na revisão crítica do conteúdo; e (3) participar da aprovação da versão final do manuscrito.

Nos casos de grupos multicêntricos ou grande número de autores terem desenvolvido o trabalho, o grupo deve identificar os indivíduos que aceitam a responsabilidade direta pelo manuscrito. Esses indivíduos devem contemplar totalmente os critérios para autoria definidos anteriormente. Nesse caso os editores solicitarão a eles as declarações exigidas na submissão de manuscritos. O autor correspondente deve indicar claramente a forma de citação preferida para o nome do grupo e identificar seus membros. Estes serão listados no final do texto do artigo.

Não justificam autoria: aqueles que realizaram apenas a aquisição de financiamento, a coleta de dados, ou supervisão geral do grupo de pesquisa.

b) Agradecimentos

Devem ser mencionados os nomes de pessoas que, embora não preencham os requisitos de autoria, prestaram colaboração ao trabalho. Será preciso explicitar o motivo do agradecimento (por exemplo, consultoria científica, revisão crítica do manuscrito, coleta de dados etc). Deve haver permissão expressa dos nomeados e o autor responsável deve anexar a Declaração de Responsabilidade pelos Agradecimentos. Também pode constar agradecimentos a instituições que prestaram apoio logístico.

c) Transferência de direitos autorais

Todos os autores devem ler, assinar e enviar documento transferindo os direitos autorais. O artigo só será liberado para publicação quando esse documento estiver de posse da RSP.

O documento de transferência de direitos autorais será solicitado após a aprovação do artigo.

MODELOS DOS DOCUMENTOS

a) CARTA DE APRESENTAÇÃO

Cidade, _[dia]__ de Mês de Ano.

Prezado Sr. Editor, Revista de Saúde Pública

Assim, submetemos à sua apreciação o trabalho “_____ [título] _____”, o qual se encaixa nas áreas de interesse da RSP. A revista foi escolhida [colocar justificativa da escolha da revista para a publicação do manuscrito].

Contribuição dos autores (exemplo): concepção, planejamento, análise, interpretação e redação do trabalho: autor 1; interpretação e redação do trabalho: autor 2. Ambos os autores aprovaram a versão final encaminhada.

Certifico que este manuscrito representa um trabalho original e que nem ele, em parte ou na íntegra, nem outro trabalho com conteúdo substancialmente similar, de minha autoria, foi publicado ou está sendo considerado para publicação em outra revista, quer seja no formato impresso ou no eletrônico.

Os autores não possuem conflitos de interesse ao presente trabalho. (Se houver conflito, especificar).

_____ nome completo do autor 1
+ assinatura + ORCID

_____ nome completo do autor 2
+ assinatura + ORCID

b) DECLARAÇÃO DE RESPONSABILIDADE PELOS AGRADECIMENTOS

Eu, (nome por extenso do autor responsável pela submissão), autor do manuscrito intitulado (título completo do artigo):

- Certifico que todas as pessoas que tenham contribuído substancialmente à realização deste manuscrito, mas que não preencheram os critérios de autoria, estão nomeadas com suas contribuições específicas em Agradecimentos no manuscrito.
- Certifico que todas as pessoas mencionadas nos Agradecimentos forneceram a respectiva permissão por escrito.

_____/_____/_____

DATA NOME COMPLETO E ASSINATURA

c) DECLARAÇÃO DE TRANSFERÊNCIA DE DIREITOS AUTORAIS

Concordo que os direitos autorais referentes ao manuscrito [TÍTULO], aprovado para publicação na Revista de Saúde Pública, serão propriedade exclusiva da Faculdade de Saúde Pública, sendo possível sua reprodução, total ou parcial, em qualquer outro meio de divulgação, impresso ou eletrônico, desde que citada a fonte, conferindo os devidos créditos à Revista de Saúde Pública.

Autores: _____

Local, data NOME COMPLETO + Assinatura

2.2.6 Preparo do manuscrito

Título no idioma original do manuscrito

O título deve ser conciso e completo, contendo informações relevantes que possibilitem a recuperação do artigo nas bases de dados. O limite é de 90 caracteres, incluindo espaços.

Título resumido

É o título que constará no cabeçalho do artigo. Deve conter a essência do assunto em até 45 caracteres.

Descritores

Para manuscritos escritos em português ou espanhol, devem ser indicados entre 3 a 10 descritores extraídos do vocabulário “Descritores em Ciências da Saúde” (DeCS), da BVS/Bireme, no idioma original. Para manuscritos em inglês, utilizar o Medical Subject Headings (MeSH) da National Library of Medicine (EUA). Se não forem encontrados descritores adequados para a temática do manuscrito, poderão ser indicados termos livres.

Resumo

O resumo deve ser escrito em seu idioma original. As especificações quanto ao tipo de resumo estão descritas em cada uma das categorias de artigos. Como regra geral, o resumo deve incluir: objetivo do estudo, principais procedimentos metodológicos (população em estudo, local e ano de realização, métodos observacionais e analíticos), principais resultados e conclusões.

Estrutura do texto

Introdução – Deve relatar o contexto e a justificativa do estudo, apoiados em referências pertinentes. O objetivo do manuscrito deve estar explícito no final da introdução.

Métodos – É imprescindível a descrição clara dos procedimentos adotados, das variáveis analisadas (com a respectiva definição, se necessário) e da hipótese a ser testada. Descrever também a população, a amostra e os instrumentos de medida, com a apresentação, se possível, de medidas de validade. É necessário que haja informações sobre a coleta e o processamento de dados. Devem ser incluídas as devidas referências para as técnicas e métodos empregados, inclusive os métodos estatísticos; é fundamental que os métodos novos ou substancialmente modificados sejam descritos, justificando-se as razões para seu uso e mencionando-se suas limitações. Os critérios éticos de pesquisa devem ser respeitados. Os autores devem explicitar que a pesquisa foi conduzida dentro dos padrões éticos e aprovada por comitê de ética.

Resultados – É preciso que sejam apresentados em uma sequência lógica, iniciando-se com a descrição dos dados mais importantes. Tabelas e figuras devem ser restritas àquelas necessárias para argumentação e a descrição dos dados no texto deve ser restrita aos mais importantes. Os gráficos devem ser utilizados para destacar os resultados mais relevantes e resumir relações complexas. Dados em gráficos e tabelas não devem ser duplicados, nem repetidos no texto. Os resultados numéricos devem especificar os métodos estatísticos utilizados na análise.

Discussão – A partir dos dados obtidos e resultados alcançados, os aspectos novos e importantes observados devem ser interpretados à luz da literatura científica e das teorias existentes no campo. Argumentos e provas baseadas em comunicação de caráter pessoal ou divulgadas em documentos restritos não podem servir de apoio às argumentações do autor. Tanto as limitações do trabalho quanto suas implicações para futuras pesquisas precisam ser esclarecidas. É necessário incluir somente hipóteses e generalizações baseadas nos dados do trabalho. As Conclusões devem finalizar esta parte, retomando o objetivo do trabalho.

Referências

Listagem: As referências devem ser normatizadas de acordo com o estilo Vancouver – Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication, listadas por ordem de citação. Os títulos de periódicos devem ser referidos de forma abreviada, de acordo com o PubMed. No caso de publicações com até seis autores, todos devem ser citados; acima de seis, devem ser citados apenas os seis primeiros, seguidos da expressão latina “et al.”. Sempre que possível, incluir o DOI do documento citado.

Exemplos:

Artigo de periódicos

Brüggemann OM, Osis MJD, Parpinelli MA. Apoio no nascimento: percepções de profissionais e acompanhantes escolhidos pela mulher. Rev Saude Publica. 2007;41(1):44-52. <https://doi.org/10.1590/S0034-89102006005000015>

Livro

Wunsch Filho V, Koifman S. Tumores malignos relacionados com o trabalho. In: Mendes R, coordenador. Patologia do trabalho. 2. ed. São Paulo: Atheneu; 2003. v.2, p. 990-1040.

Foley KM, Gelband H, editors. Improving palliative care for cancer Washington: National Academy Press; 2001[citado 2003 jul 13]. Disponível em: http://www.nap.edu/catalog.php?record_id=10149

Para outros exemplos recomendamos consultar as normas (Citing Medicine) da National Library of Medicine, disponível em <http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=citmed>.

Citação no texto:

É necessário que a referência seja indicada pelo seu número na listagem, na forma de **expoente (sobrescrito)** antes da pontuação no texto, sem uso de parênteses, colchetes ou similares. Nos casos em que a citação do nome do autor e ano for relevante, o número da referência deve ser colocado seguido do nome do autor. Trabalhos com dois autores devem fazer referência aos dois autores ligados por “e”. Nos outros casos de autoria múltipla, apresentar apenas o primeiro autor, seguido de “et al.”

Exemplos:

A promoção da saúde da população tem como referência o artigo de Evans e Stoddart⁹, que considera a distribuição de renda, desenvolvimento social e reação individual na determinação dos processos de saúde-doença.

Segundo Lima et al.⁹ (2006), a prevalência de transtornos mentais em estudantes de medicina é maior do que na população em geral.

Tabelas

Devem ser apresentadas no final do texto, após as referências bibliográficas, numeradas consecutivamente com algarismos arábicos, na ordem em que foram citadas no texto. A cada uma deve-se atribuir um título breve. Não utilizar traços internos horizontais ou verticais. As notas explicativas devem ser colocadas no rodapé das tabelas e não no cabeçalho ou no título. Se houver tabela extraída de

outro trabalho publicado previamente, os autores devem solicitar formalmente autorização da revista que a publicou para sua reprodução. Para composição de uma tabela legível, o número máximo é de 10 colunas, dependendo da quantidade do conteúdo de cada casela. Notas em tabelas devem ser indicadas por letras e em sobrescrito.

Quadros

Diferem das tabelas por conterem texto em vez de dados numéricos. Devem ser apresentados no final do texto, após as referências bibliográficas, numerados consecutivamente com algarismos arábicos, na ordem em que foram citados no texto. A cada um deve-se atribuir um título breve. As notas explicativas devem ser colocadas no rodapé dos quadros e não no cabeçalho ou no título. Se houver quadro extraído de trabalho publicado previamente, os autores devem solicitar formalmente autorização da revista que o publicou para sua reprodução.

Figuras

As ilustrações (fotografias, desenhos, gráficos etc.) devem ser citadas como Figuras e numeradas consecutivamente com algarismos arábicos, na ordem em que foram citadas no texto e apresentadas após as tabelas. Elas também devem conter título e legenda apresentados em sua parte inferior. Só serão admitidas para publicação figuras suficientemente claras e com qualidade digital, preferencialmente no formato vetorial. No formato JPEG, a resolução mínima deve ser de 300 dpi. Não se aceitam gráficos apresentados com as linhas de grade, e os elementos (barras, círculos) não podem apresentar volume (3D). Se houver figura extraída de trabalho publicado previamente, os autores devem solicitar formalmente autorização da revista que a publicou para sua reprodução.

2.2.7 Checklist para submissão

1. Nome e instituição de afiliação de cada autor, incluindo e-mail e telefone.
2. Título do manuscrito em inglês e em seu idioma original (português ou espanhol), se aplicável.
3. Título resumido com 45 caracteres.
4. Texto apresentado em letras arial, corpo 12, em formato Word ou similar (doc, docx e rtf).
5. Resumos estruturados para trabalhos originais de pesquisa no idioma original do manuscrito.
6. Resumos narrativos para manuscritos que não são de pesquisa no idioma original do manuscrito.

7. Carta de Apresentação, constando a responsabilidade de autoria. Deve ser assinada por todos os autores ou, pelo menos, pelo primeiro autor e conter o ORCID de todos os autores.
8. Nome da agência financiadora e número(s) do(s) processo(s).
9. Referências normatizadas segundo estilo Vancouver, apresentadas por ordem de citação. É necessário verificar se todas estão citadas no texto.
10. Tabelas numeradas sequencialmente, com título e notas, com no máximo 10 colunas.
11. Figura no formato vetorial ou em pdf, tif, jpeg ou bmp, com resolução mínima 300 dpi. Gráficos devem estar sem linhas de grade e sem volume.
12. Tabelas e figuras não devem exceder a cinco, no conjunto.
13. Não publicamos materiais suplementares, anexos e apêndices.

2.2.8 Processo editorial

a) Avaliação

Pré-análise: : o Editor Científico avalia os manuscritos com base na qualidade e interesse para a área de saúde pública e decide se seleciona o manuscrito para avaliação por pares externos ou não. Análise por pares: se selecionado na pré-análise, o manuscrito é encaminhado a um dos Editores Associados cadastrados no sistema segundo a respectiva área de especialização. O Editor Associado seleciona os revisores (dois) de acordo com a área de especialização e envia o manuscrito para avaliação. Caso o Editor Associado considere que os pareceres recebidos são insuficientes para uma conclusão, deverá indicar outro(s) relator(es). Com base nos pareceres, o Editor Associado decide por: recusa, no caso de o manuscrito ter deficiências importantes; aceite; ou possibilidade de nova submissão, devendo neste caso indicar nos seus comentários as modificações importantes para eventual reformulação, que será reavaliada por relatores.

b) Revisão da redação científica

Para ser publicado, o manuscrito aprovado é editado por uma equipe que fará à revisão da redação científica (clareza, brevidade, objetividade e solidez), gramatical e de estilo. A RSP se reserva o direito de fazer alterações visando a uma perfeita comunicação aos leitores. O autor

responsável terá acesso a todas as modificações sugeridas até a última prova enviada.

c) Provas

O autor responsável pela correspondência receberá uma prova, em arquivo de texto (doc, docx ou rtf), com as observações e alterações feitas pela equipe de leitura técnica. O prazo para a revisão da prova é de dois dias.

Caso ainda haja dúvidas nessa prova, a equipe editorial entrará em contato para que seja feita a revisão, até que seja alcançada uma versão final do texto.

Artigos submetidos em português ou espanhol serão vertidos para o inglês. Aproximadamente 20 dias após o autor ter finalizado a prova do artigo, a RSP enviará a versão em inglês do artigo para apreciação do autor. Nessa revisão, o autor deverá atentar-se para possíveis erros de interpretação, vocabulário da área e, principalmente, equivalência de conteúdo com a versão original aprovada. O prazo de revisão da versão em inglês é de dois dias.

A RSP adota o sistema de publicação contínua. Dessa forma, a publicação do artigo se torna mais rápida: não depende de um conjunto de artigos para fechamento de um fascículo, mas do processo individual de cada artigo. Por isso, solicitamos o cumprimento dos prazos estipulados.

2.2.9. Taxa de publicação

A RSP iniciou em 2012 a cobrança de taxa de publicação, fato este imperioso para garantir sua continuidade, com qualidade e recursos tecnológicos.

Em consideração aos autores cuja situação financeira foi prejudicada pela pandemia, decidimos manter inalterado o valor da taxa nos últimos 2 anos e meio, absorvendo os aumentos nos custos de publicação. Entretanto, dado os cortes expressivos nos apoios recebidos de instituições de fomento e as pressões inflacionárias do setor, a partir de **1º de janeiro de 2022** o valor da taxa de publicação será alterado para: R\$ 3.800,00 para artigo original, revisão e comentário e R\$ 3.200,00 para comunicação breve. Os pagamentos provenientes do exterior serão de US\$ 1.200,00 e US\$ 1.000,00, respectivamente. Os artigos submetidos anteriormente a essa data não sofrerão reajuste na taxa de publicação.

A RSP fornece aos autores os documentos necessários para comprovar o pagamento da taxa perante instituições

empregadoras, programas de pós-graduação ou órgãos de fomento à pesquisa.

Após aprovação do artigo, os autores deverão aguardar o envio da fatura proforma com as informações sobre como proceder quanto ao pagamento da taxa.

3. ANEXOS

CARTA DE APRESENTAÇÃO

Cidade, 04 de junho de 2022.

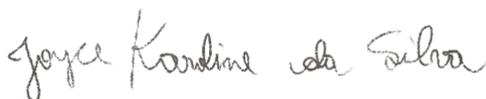
Prezado Sr. Editor, Revista de Saúde Pública

Assim, submetemos à sua apreciação o trabalho “SARA: DESENVOLVIMENTO E VALIDAÇÃO DE UM FLUXO DE ANÁLISES SEMIAUTOMÁTICO PARA ESTUDOS DE EXPRESSÃO GÊNICA”, o qual se encaixa nas áreas de interesse da RSP. A revista foi escolhida por ser de qualis A2 que realiza contribuições relevantes para a saúde, alcançando estudantes da área que podem se beneficiar com aplicações de técnicas de análise de dados de sequenciamento de nova geração, importante para as ciências ômicas.

Contribuição dos autores: análise, interpretação e redação do trabalho: autor 1; concepção, planejamento, e redação do trabalho: autor 2. Ambos os autores aprovaram a versão final encaminhada.

Certifico que este manuscrito representa um trabalho original e que nem ele, em parte ou na íntegra, nem outro trabalho com conteúdo substancialmente similar, de minha autoria, foi publicado ou está sendo considerado para publicação em outra revista, quer seja no formato impresso ou no eletrônico.

Os autores não possuem conflitos de interesse ao presente trabalho. (Se houver conflito, especificar).



Joyce Karoline da Silva

ORCID: orcid.org/0000-0002-8039-9706



Laise Eduarda Paixão de Moraes

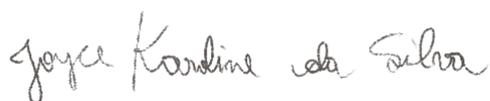
ORCID: orcid.org/0000-0001-9095-5860

DECLARAÇÃO DE RESPONSABILIDADE PELOS AGRADECIMENTOS

Eu, Joyce Karoline da Silva, autor do manuscrito intitulado (título completo do artigo): SARA: DESENVOLVIMENTO E VALIDAÇÃO DE UM FLUXO DE ANÁLISES SEMIAUTOMÁTICO PARA ESTUDOS DE EXPRESSÃO GÊNICA

- Certifico que todas as pessoas que tenham contribuído substancialmente à realização deste manuscrito, mas que não preencheram os critérios de autoria, estão nomeadas com suas contribuições específicas em Agradecimentos no manuscrito.
- Certifico que todas as pessoas mencionadas nos Agradecimentos forneceram a respectiva permissão por escrito.

23/05/2022 Joyce Karoline da Silva



DECLARAÇÃO DE TRANSFERÊNCIA DE DIREITOS AUTORAIS

Concordo que os direitos autorais referentes ao manuscrito SARA: DESENVOLVIMENTO E VALIDAÇÃO DE UM FLUXO DE ANÁLISES SEMIAUTOMÁTICO PARA ESTUDOS DE EXPRESSÃO GÊNICA aprovado para publicação na Revista de Saúde Pública, serão propriedade exclusiva da Faculdade de Saúde Pública, sendo possível sua reprodução, total ou parcial, em qualquer outro meio de divulgação, impresso ou eletrônico, desde que citada a fonte, conferindo os devidos créditos à Revista de Saúde Pública.

Autores:

Joyce Karoline da Silva

Laise Eduarda Paixão de Moraes

Salvador, 23/05/2022

Joyce Karoline da Silva

Laise Eduarda Paixão de Moraes