
Guia do usuário: Co-localização de parasitas de Chagas em células musculares cardíacas

Patrick. H. F. Alvares, João V. S. Guerra, José G. C. Pereira

November 13, 2024

Table of Contents

1	Introdução	3
2	Requisitos	3
3	Acessando o HPCC Marvin	3
4	Preparando o Repositório ParasiteCoLocalization	3
4.1	Instalando PyEnv no HPCC Marvin	3
4.2	Instalando as Dependências do Python	4
4.3	Enviando Imagens para o HPCC Marvin	4
4.4	Executando o Pipeline no HPCC Marvin	4
4.4.1	Monitoramento do Job SLURM	4
5	Visualizando os Resultados	5
6	Ajustando o pipeline parasitecolocalization.cppipe	6
7	Glossário	6

1 Introdução

Este guia do usuário fornece instruções detalhadas sobre como executar a análise de co-localização de parasitas de Chagas em células musculares cardíacas. O objetivo deste guia é auxiliar os usuários a configurar e executar o pipeline de análise, bem como a interpretar os resultados obtidos.

2 Requisitos

Para executar a análise de co-localização de parasitas de Chagas em células musculares cardíacas, você precisará dos seguintes requisitos:

- **Acesso ao GitHub do CNPEM:** Você deve ter acesso ao repositório do projeto (<https://github.com/cnpem/ParasiteCoLocalization>) no GitHub do CNPEM (<https://github.com/cnpem/>) para baixar os arquivos necessários. Se não tiver acesso, entre em contato com a Equipe de Dados Biológicos (edb@lnbio.cnpem.br).
- **Acesso ao HPCC Marvin:** Você deve ter acesso ao HPCC Marvin para executar o pipeline de análise. Se você não tiver acesso, acesse <https://marvindocs.cnpem.br/primeiros-passos/index.html>.

3 Acessando o HPCC Marvin

Para acessar o HPCC Marvin, comece abrindo o terminal. Se estiver usando Windows, abra o PowerShell; se estiver usando Linux ou MacOS, abra o Terminal. Para acessar o HPCC Marvin, use o comando:

```
ssh <seu.login.cnpem>@marvin.cnpem.br
```

Após inserir o comando, será solicitado que você insira sua senha, digite sua senha institucional.

4 Preparando o Repositório ParasiteCoLocalization

Para clonar o repositório [ParasiteCoLocalization](https://github.com/cnpem/ParasiteCoLocalization), execute o seguinte comando no terminal:

```
git clone https://github.com/cnpem/ParasiteCoLocalization.git
```

O comando acima cria uma pasta chamada `ParasiteCoLocalization` no diretório atual. Esta pasta contém todos os arquivos necessários para a execução da análise de co-localização de parasitas de Chagas em células musculares cardíacas, incluindo o script `run.sh`, que automatiza a execução da análise via SLURM.

Após clonar o repositório `ParasiteCoLocalization`, acesse o diretório do projeto com o comando:

```
cd ParasiteCoLocalization
```

4.1 Instalando PyEnv no HPCC Marvin

O PyEnv é uma ferramenta que permite instalar e gerenciar várias versões do Python em um ambiente virtual. Para instalar o PyEnv no HPCC Marvin, execute os seguintes comandos no terminal:

```
curl https://pyenv.run | bash
echo 'export PATH="$HOME/.pyenv/bin:$PATH"' >> ~/.bashrc
echo 'eval "$(pyenv init -)"' >> ~/.bashrc
echo 'eval "$(pyenv virtualenv-init -)"' >> ~/.bashrc
source ~/.bashrc
```

Para instalar uma versão específica do Python, execute o seguinte comando:

```
pyenv install 3.8.19
```

Para definir a versão do Python como padrão neste diretório, execute o seguinte comando:

```
pyenv local 3.8.19
```

Caso deseje utilizar outra versão do Python, você pode selecionar a versão desejada para a sessão atual com o comando:

```
pyenv shell <versão_desejada>
```

4.2 Instalando as Dependências do Python

Para instalar as dependências do Python, execute o seguinte comando no terminal:

```
pip install -r requirements.txt
```

4.3 Enviando Imagens para o HPCC Marvin

Para enviar as imagens para o HPCC Marvin, você pode usar o comando `scp` em qualquer terminal.

```
scp /caminho/para/suas/imagens/*.tiff  
↪ <seu.login.cnpem>@marvin.cnpem.br:/caminho/para/ParasiteCoLocalization/data
```

Caso você esteja usando o Windows, você pode usar o <https://winscp.net/eng/download.php> para transferir arquivos de e para o HPCC Marvin. Para isso, siga as instruções abaixo:

1. Abra o WinSCP e insira o endereço do servidor (`marvin.cnpem.br`), seu nome de usuário e senha.
2. Navegue até o diretório `ParasiteCoLocalization/data` no HPCC Marvin.
3. Arraste e solte os arquivos de imagem do seu computador para o diretório `ParasiteCoLocalization/data` no HPCC Marvin.

4.4 Executando o Pipeline no HPCC Marvin

Após enviar as imagens para o HPCC Marvin, você pode executar o pipeline de análise de co-localização de parasitas de Chagas em células musculares cardíacas. O script `run.sh` executa o pipeline de análise via SLURM, que é o sistema de gerenciamento de tarefas usado no HPCC Marvin. Para isso, execute o script `run.sh` no diretório `ParasiteCoLocalization`.

```
cd ParasiteCoLocalization  
sbatch run.sh -m marvin
```

O comando acima retorna o ID do job SLURM, que você pode usar para monitorar o progresso da análise.

4.4.1 Monitoramento do Job SLURM

Para monitorar o progresso do job SLURM, você pode usar o comando `squeue` no terminal. O comando `squeue` exibe informações sobre os jobs SLURM em execução no HPCC Marvin.

```
squeue -u <seu.login.cnpem>
```

Se o job estiver em execução, o estado será exibido como R (em execução). Se o job estiver aguardando na fila, o estado será PD (pendente).

Caso deseje cancelar o job SLURM, você pode usar o comando `squeue` no terminal. O comando `squeue` cancela um job SLURM em execução no HPCC Marvin.

```
squeue <job_id>
```

Por fim, você pode verificar o arquivo de log gerado pelo SLURM (`slurm_<job_id>.out`) para obter informações sobre o seu job.

5 Visualizando os Resultados

Após a conclusão da análise, os resultados serão armazenados no diretório `results/`, que incluem os seguintes arquivos:

- `summary.csv`: Este arquivo contém um resumo dos resultados da análise por poço, incluindo contagem total de células, contagem total de parasitas, contagem de células infectadas (> 3 parasitas), taxa de infecção, e número mediano de parasitas por célula.

	TotalCells	TotalSpots	InfectedCells	InfectionRate	MedianSpotsPerInfectedCell
A1	726.0	5993.0	155	21.349862258953166	23.0
A2	593.0	5176.0	165	27.82462057335582	16.0
A3	759.0	7076.0	232	30.566534914361	18.0
A4	679.0	6243.0	170	25.03681885125184	23.0
A5	601.0	5898.0	164	27.287853577371045	19.5
A6	715.0	6112.0	187	26.153846153846157	18.0
A7	551.0	4285.0	172	31.21597096188748	13.0
A8	625.0	5036.0	161	25.759999999999998	17.0

- `plate_map/`: Este diretório contém os mapas interativos da placa de 384 poços (Figura 1), contendo:
 - `number_of_cells.html`: contagem total de células por poço.
 - `number_of_spots.html`: contagem total de parasitas por poço.
 - `median_spots_per_infected_cell.html`: número mediano de parasitas por célula por poço.
 - `infection_rate.html`: razão entre o número de células infectadas e o número total de células por poço.

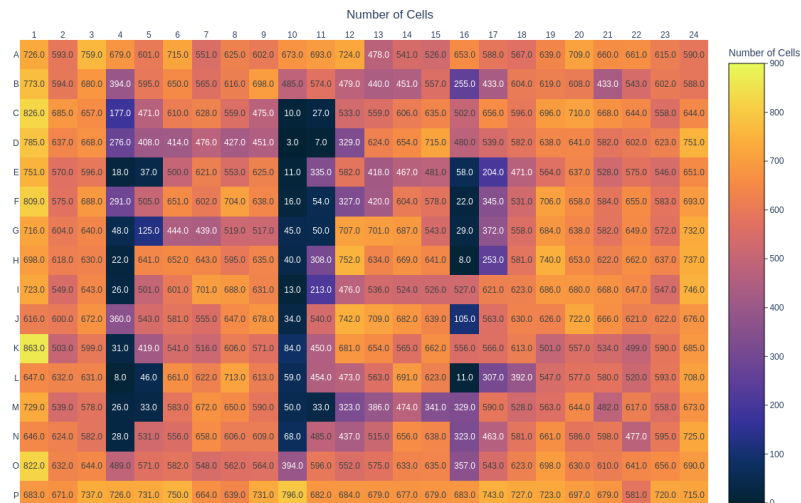


Figure 1: Exemplo de mapa interativo da placa de 384 poços.

- `scatter/`: Este diretório contém os gráficos de dispersão interativos para a placa de 384 poços (Figura 2), contendo:
 - `number_of_cells_vs_number_of_infected_cells.html`: relação entre o número de células e o número de células infectadas.
 - `number_of_cells_vs_infection_rate.html`: relação entre o número de células e a taxa de infecção.
 - `median_spots_per_infected_cell_vs_number_of_cells.html`: relação entre o número mediano de parasitas por célula infectada e o número de células.
 - `median_spots_per_infected_cell_vs_infection_rate.html`: relação entre o número mediano de parasitas por célula infectada e a taxa de infecção.

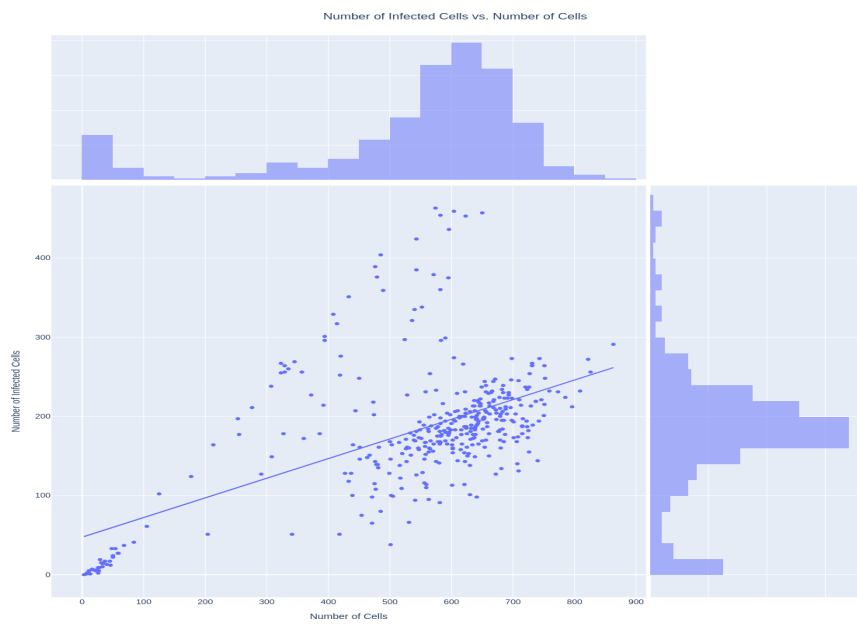


Figure 2: Exemplo de gráfico de dispersão interativo da placa de 384 poços.

6 **Ajustando o pipeline `parasitecolocalization.cppipe`**

7 **Glossário**

Here is an Example (Ex) of acronym usage.

Ex Example	6
-------------------	---