

INFORME PEC1

Análisis de datos ómicos (M0-157)

Alejandro Parra

Tabla de contenido

Resumen.....	2
Objetivos.....	2
Materiales y métodos.....	3
Origen y naturaleza de los datos	3
Metodología empleada.....	3
Importación y depuración de los datos:	3
Análisis exploratorio inicial:	3
Preprocesado de los datos:.....	3
Análisis exploratorio:	3
Herramientas utilizadas	4
Resultados	4
Exploración inicial.....	4
Preprocesado de los datos	4
PCA: agrupación de muestras en función del perfil metabólico.....	5
Heatmap: patrones de abundancia de metabolitos	6
Análisis de <i>clustering</i> (PomaClust).....	6
Estudio de metabolitos clave.....	7
Discusión	8
Conclusiones	9
Referencias	9

Resumen

En este estudio, se realiza un análisis de metabolómica dirigido a entender la implicación de la proteína lisina metiltransferasa KMT9 en la regulación de la lipogénesis *de novo* en células de cáncer de próstata (CaP). Partiendo de los datos procedentes del estudio MTBLS2452, se desarrolla un *pipeline* bioinformático que incluye la preparación de los datos y un posterior análisis exploratorio de los patrones de abundancia de metabolitos entre las muestras de CaP tratadas con siKMT9α y muestras de CaP control. Posteriormente, se seleccionaron metabolitos clave relacionados con la vía del complejo piruvato deshidrogenasa (PDC) y la lipogénesis (piruvato, lactato, NAD⁺, acetil-CoA) y se evaluó su distribución entre los grupos tratados y control.

Los resultados muestran diferencias en la abundancia relativa de los metabolitos entre ambos grupos de muestras, apoyando la hipótesis de que KMT9 afecta la actividad de la PDC y la proliferación celular en CaP.

Objetivos

El objetivo principal de trabajo es la realización de un análisis exploratorio del conjunto de datos metabólicos procedentes del estudio MTBLS2452 de Metabolights. Concretamente, se pretende elucidar si la depleción de KMT9 en células de CaP afecta al metabolismo de los lípidos. Para llevar a cabo dicho análisis, el objetivo principal se descompone en los siguientes objetivos específicos:

1. Explorar la capacidad discriminante del perfil metabólico de las muestras en función del grupo al que pertenecen las muestras (tratadas con siKMT9α o tratadas con un siARN control) mediante un análisis de componentes principales (PCA).
2. Evaluar la variación global de los metabolitos (patrones) entre las muestras incluidas en este estudio mediante la elaboración de un HeatMap.
3. Analizar la distribución específica de diversos metabolitos relevantes para el contexto del cáncer de páncreas y la lipogénesis: piruvato, lactato, NAD⁺ y acetil-CoA.

Materiales y métodos

Origen y naturaleza de los datos

Los datos metabólicos utilizados provienen del estudio MTBLS2452, disponible en MetaboLights, y fueron obtenidos mediante espectrometría de masas (MS). Concretamente, los datos proceden de 12 muestras celulares de CaP. Seis de estas fueron tratadas con un ARN silenciador (siKMT9α) del gen que codifica para la proteína KMT9. Las seis muestras restantes fueron utilizadas como control negativo.

Para la realización de este estudio, se han empleado dos tipos de datos: (1) el fichero *m_MTBLS2452_metabolite_profiling_mass_spectrometry_v2_maf.tsv*, el cual contiene los datos metabólicos para cada una de las muestras y (2) el fichero *s_MTBLS2452.txt*, el cual contiene los metadatos.

Metodología empleada

El análisis exploratorio de este estudio se llevó a cabo en R (v4.4.2). El flujo de trabajo consistió en:

Importación y depuración de los datos:

Los datos metabólicos y metadatos han sido importados a R usando la función *read.table*. Posteriormente se han modificado ambos conjuntos de datos para la correcta creación del objeto *SummarizedExperiment* (SE).

A continuación, se empleó Bioconductor para la creación del objeto SE mediante la función *SummarizedExperiment*, el cual recibió el nombre de Objeto_SE.

Análisis exploratorio inicial:

Se analizó el Objeto_SE con el fin de identificar sus dimensiones (número de muestras estudiadas y cantidad de metabolitos incluidos en el estudio), el tipo de muestras y el número de valores NA.

Preprocesado de los datos:

Mediante el uso del paquete POMA [1], diseñado para el análisis de datos ómicos, se llevó a cabo la imputación de los valores NA y la normalización de los datos. La imputación de los valores NA se realizó mediante el método KNN, se eliminaron los metabolitos con un alto porcentaje de datos faltantes (>20%) (*PomaImpute*). Se aplicó una normalización mediante log-pareto (*PomaNorm*). A continuación, se estudió la presencia de *outliers* mediante el uso del comando *PomaOutliers*.

Análisis exploratorio:

Se realizó un análisis de componentes principales (PCA) para observar la separación entre muestras tratadas y controles en base al perfil global de metabolitos.

Después, se generó un HeatMap que muestra la distribución diferencial de la abundancia de metabolitos entre los dos grupos experimentales. Para la realización de este análisis exploratorio se utilizaron las funciones de POMA: *PomaPCA* y *PomaHeatmap*.

Finalmente se estudiaron, diversos metabolitos relevantes para el contexto del cáncer de páncreas y la lipogénesis (piruvato, lactato, NAD⁺ y acetil-CoA) mediante la creación de gráficos de cajas.

Herramientas utilizadas

Los paquetes principales de R empleados en este estudio han sido:

- **POMA_1.16.0**: para preprocesamiento y análisis exploratorio de los datos metabolómicos.
- **SummarizedExperiment_1.36.0**: para la estructuración de los datos.
- **ggplot2_3.5.1** y **ggtext_0.1.2**: para la creación de gráficos a partir de los datos
- **dplyr_1.1.4**: para la manipulación de datos.

Resultados

Exploración inicial

La exploración inicial de los datos metabólicos y metadatos nos permitió identificar un total de 662 metabolitos y 12 muestras. Seis de las muestras fueron tratadas con siARN control (FREI-01736, FREI-01737, FREI-01738, FREI-01739, FREI-01740 y FREI-01741) y las otras seis restantes (FREI-01742, FREI-01743, FREI-01744, FREI-01745, FREI-01746 y FREI-01747) con si siKMT9a. A su vez, se identificaron un total de 939 valores NA en el Objeto_SE.

Preprocesado de los datos

Tras la imputación y limpieza de valores NA, se redujo el número de metabolitos a 598, obteniendo el Objeto_SE_NA. Por otro lado, tras aplicar la normalización, la dispersión se redujo significativamente, logrando una distribución más homogénea y comparable entre grupos (Figura 1). Obteniendo así el Objeto_SE_norm.

El análisis de valores atípicos detectó una posible muestra *outlier* (FREI-01741), ya que se queda fuera del polígono de convexidad, aunque en una posición marginal (Figura 2). Sin embargo, se optó por mantenerla en el análisis, considerando que su desviación es mínima. Tras la realización de estos procesos, el conjunto de datos está ahora preparado para avanzar con el análisis, asegurando que los datos sean adecuados y comparables entre grupos.

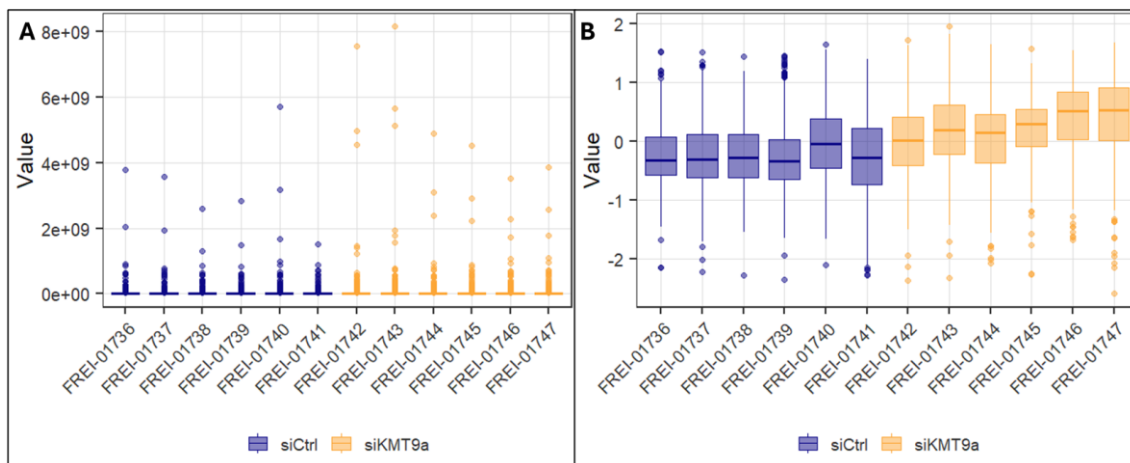


FIGURA 1: (A) DIAGRAMA DE CAJAS DE LA INTENSIDAD DE SEÑAL DE METABOLITOS POR MUESTRA REALIZADO A PARTIR DEL OBJETO_SE_NA. (B) DIAGRAMA DE CAJAS DE LA INTENSIDAD DE SEÑAL DE METABOLITOS POR MUESTRA REALIZADO A PARTIR DEL OBJETO_SE_NORM.

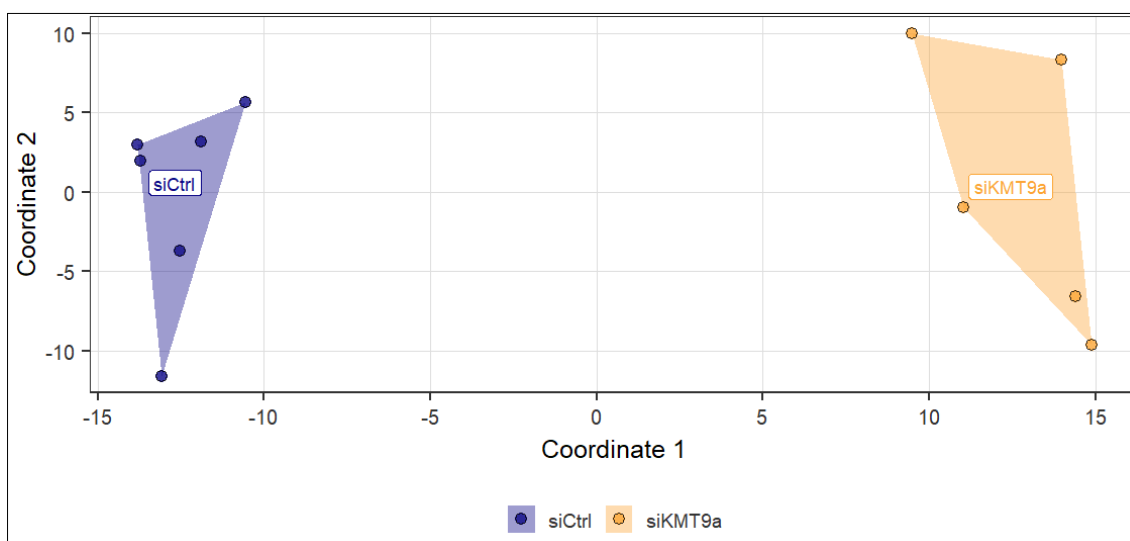


FIGURA 2: ESTUDIO DE VALORES ATÍPICOS MEDIANTE DETECTADOS MEDIANTE POLÍGONO DE CONVEXIDAD.

PCA: agrupación de muestras en función del perfil metabólico

El análisis de componentes principales (Figura 3A) mostró una clara separación entre las muestras tratadas con siKMT9a y las tratadas con siARN control en el espacio bidimensional de las primeras dos componentes (79.58% de la varianza explicada). Esto sugiere que el perfil metabólico global de las células tratadas con siKMT9a difiere significativamente del de las células tratadas con siARN control.

A su vez, mediante el “método del codo” (Figura 3B) se estipuló que eran tres los componentes principales que explicaban más del 80% de la variabilidad del conjunto de datos.

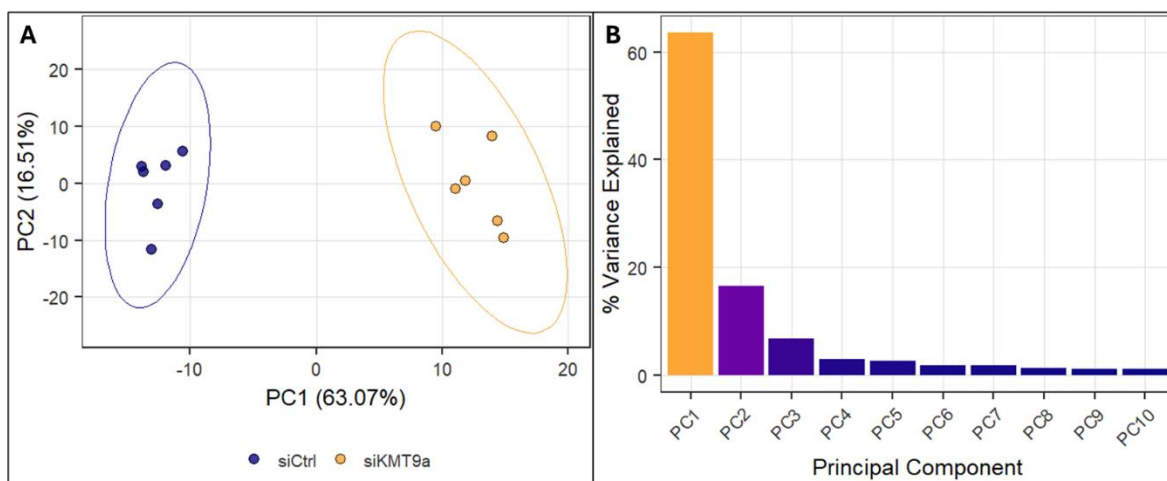


FIGURA 3: (A) ANÁLISIS DE COMPONENTES PRINCIPALES (PCA): SEPARACIÓN ENTRE MUESTRAS TRATADAS CON SIKMT9A Y SIARN CONTROL EN EL ESPACIO DE LAS DOS PRIMERAS COMPONENTES PRINCIPALES. (B) SCREE PLOT EN EL QUE SE PUEDE IDENTIFICAR QUE CON TRES COMPONENTES PRINCIPALES SE EXPLICA MÁS DEL 80% DE LA VARIANZA.

Heatmap: patrones de abundancia de metabolitos

El heatmap creado a partir del Objeto_SE_norm (Figura 4) mostró dos patrones opuestos de abundancia de metabolitos entre los grupos. Los metabolitos con alta abundancia en las muestras tratadas con siKMT9a tienden a tener baja abundancia en las muestras control, y viceversa, indicando una clara disociación metabólica inducida por la depleción de KMT9.

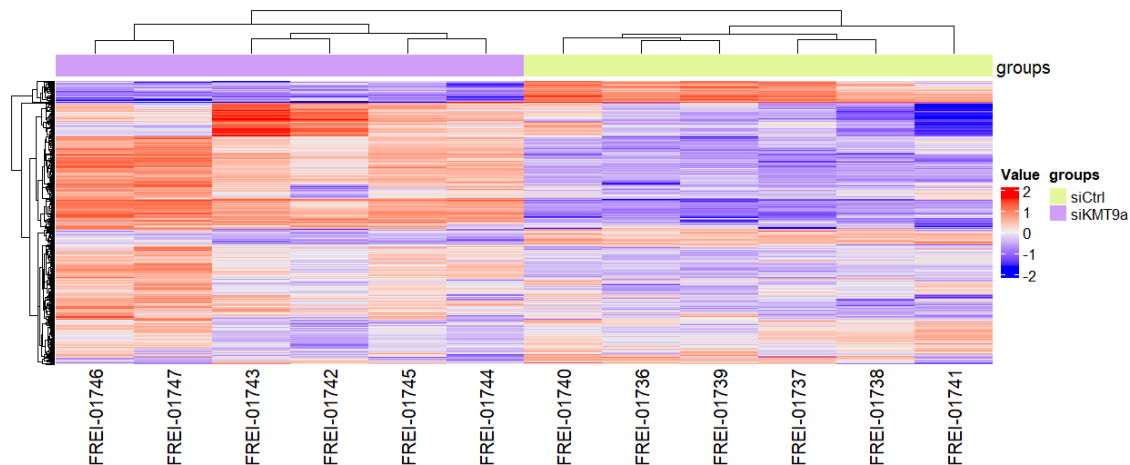


FIGURA 4: HEATMAP DE METABOLITOS NORMALIZADOS POR MUESTRA. SE PUEDE IDENTIFICAR QUE EL DENDOGRAMA AGRUPA PERFECTAMENTE A LOS DOS TIPOS DE MUESTRAS.

Análisis de clustering (PomaClust)

El Análisis de clustering (Figura 5), realizado mediante *PomaClust*, identificó dos grupos de muestras en base a sus perfiles metabólicos. Estos grupos se corresponden con los tratados con siKMT9a y los controles. Lo cual refuerza la hipótesis de que ambos tipos de muestras presentan una expresión diferencial de metabolitos.

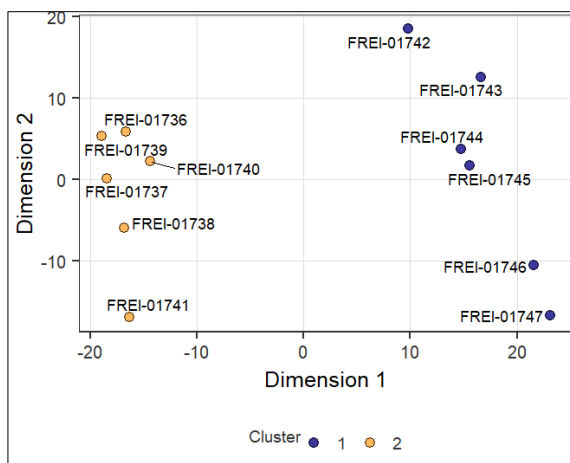


FIGURA 5: REPRESENTACIÓN GRÁFICA DEL ANÁLISIS DE CLUSTERING.

Estudio de metabolitos clave

Para concluir este estudio se decidió analizar la abundancia relativa de ciertos metabolitos clave para la lipogénesis *de novo* en el contexto de CaP. Estos metabolitos son el piruvato, el lactato, el NAD⁺ y el Acetil CoA. A continuación, se muestra los niveles de abundancia relativa de estos metabolitos:

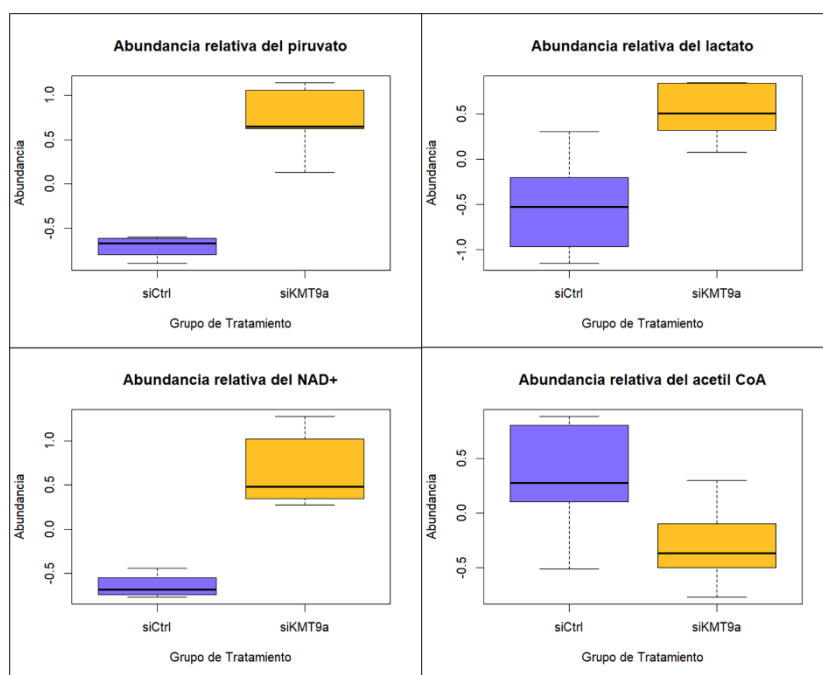


FIGURA 6: DIAGRAMAS DE CAJAS DE LA ABUNDANCIA RELATIVA DEL PIRUVATO, LACTATO, NADH⁺ Y ACETIL COA EN AMBOS GRUPOS DE MUESTRAS. SE PUEDE OBSERVAR QUE EL PIRUVATO, EL LACTATO Y EL NADH⁺ PRESENTAN MAYOR ABUNDANCIA RELATIVA EN LAS MUESTRAS TRATADAS CON SIKMT9a, MIENTRA QUE EL ACETIL COA PRESENTAN UNA ABUNDANCIA DISMINUIDA EN COMPARACIÓN CON LAS MUESTRAS CONTROL

Discusión

En ese estudio se ha realizado un análisis metabólico aplicado a 12 muestras de CaP. En seis de ellas se llevó a cabo la depleción de KMT9 mediante un siARN con el objetivo de estudiar si esta proteína interfiere en el proceso de lipogénesis de novo en células de CaP. Puesto que la síntesis de ácidos grasos desempeña un papel clave en el proceso de crecimiento de las células de CaP [2].

El análisis realizado confirma que la depleción de KMT9 en células de CaP induce alteraciones metabólicas significativas, particularmente en metabolitos asociados a la actividad de la PDC y la lipogénesis de novo. El PCA y el PomaClust confirma una clara separación entre los grupos control y tratados, sugiriendo que el perfil metabólico de ambos grupos es significativamente diferente. Además, el HeatMap de los niveles de abundancia normalizados de los metabolitos estudiados confirma la existencia de un patrón metabólico diferencial entre estos grupos de muestras. Entre las limitaciones del estudio destaca el tamaño reducido de la muestra ($n=12$), lo que puede limitar la generalización de los resultados y la robustez estadística.

El estudio de los metabolitos piruvato, lactato, NADH⁺ y acetil CoA reveló mayor abundancia de de piruvato, lactato, NADH⁺ en las muestras con depleción del KMT9, mientras que las muestras control presentaban mayor abundancia de acetil CoA. Estos datos, indican que KMT9, regula a la PDC; y su depleción reduce la generación de Acetil-CoA, limitando la síntesis de ácidos grasos necesarios para el crecimiento de las células de CaP. Por lo tanto, la dependencia de las células de PaC de la proteína KMT9 para la síntesis de ácidos grasos podría utilizarse para el desarrollo de estrategias terapéuticas contra el CaP.

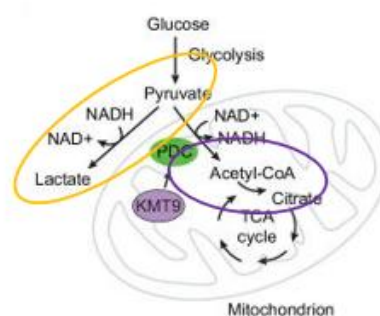


FIGURA 7: ESQUEMA QUE MUESTRA LA DISPOSICIÓN DE LOS METABOLITOS EN LA CÉLULA. EL CÍRCULO AMARILLO ENGLOBA LOS METABOLITOS CON ALTA ABUNDANCIA EN CÉLULAS CON KMT9 DEPLECIONADO, MIENTRAS QUE EL MORADO ENGLOBA AL ACETIL COA, CUYA ABUNDANCIA RELATIVA ES MAYOR EN CONTROLES. IMAGEN ADAPTADA [3]

Conclusiones

Las conclusiones de este análisis metabólico son:

1. La depleción de KMT9 en células de CaP produce alteraciones significativas en el perfil metabólico de dichas células.
2. El perfil metabólico diferencial entre las muestras con depleción de KMT9 y las muestras control permite diferenciar a ambos grupos de células mediante técnicas de clustering.
3. El perfil metabólico de ambos tipos de muestras es opuesto.
4. La proteína KMT9 es una buena candidata como diana terapéutica para el CaP.

Referencias

1. Castellano-Escuder P, González-Domínguez R, Carmona-Pontaque F, Andrés-Lacueva C, Sánchez-Pla A (2021). "POMAShiny: A user-friendly web-based workflow for metabolomics and proteomics data analysis." *PLOS Computational Biology*, **17**(7), 1-15. <https://doi.org/10.1371/journal.pcbi.1009148>.
2. Wu, X., Daniels, G., Lee, P., & Monaco, M. E. (2014). Lipid metabolism in prostate cancer. *American journal of clinical and experimental urology*, **2**(2), 111.
3. Jia, Y., Wang, S., Urban, S., Müller, J. M., Sum, M., Wang, Q., ... & Schüle, R. (2025). Mitochondrial KMT9 methylates DLAT to control pyruvate dehydrogenase activity and prostate cancer growth. *Nature Communications*, **16**(1), 1191.

Enlace al repositorio de GitHub que contiene el código utilizado para abordar el análisis:

- https://github.com/AlejandroParra72/Parra-Parra-Alejandro_PEC1.git