

Treball de fi de grau

Estudi de mètodes de deconvolució

Nidia Barco Armengol

Grau en Biotecnologia

Tutora: Dra. Júlia Perera Bel

Co-tutora: Dra. Susanna Bodoy Salvans

Ubicació: Vic, Gener de 2023

Agraïments

La realització d'aquest treball ha suposat per mi molts mesos de treball i aprenentatge, mesos que he compartit amb persones increïbles que m'han ajudat i acompanyat al llarg dels meus anys d'universitat, i per això els vull donar les gràcies a tots.

En primer lloc, a la meva tutora, Júlia, gràcies per la confiança que m'has donat, pels consells i per tot el que m'has ensenyat, que t'asseguro no ha estat poc. M'hauria resultat impossible pensar en una tutora millor.

També donar les gràcies a la meva tutora Susanna per la guia que m'has donat, i a tots els membres del GRIB, que em van acollir tan bé, mes en concret a l'Ariadna i al Jorge, gràcies per ajudar-me i aconsellar-me aquests últims mesos.

Després també m'agradaria donar les gràcies al professorat de la UVic per ajudar-me a créixer com a biotecnòloga i com a persona.

Moltes gràcies també a les meves amigues, Danna, Iman, Giulia i Gala. Sense vosaltres els últims anys no haurien sigut iguals, i m'alegra moltíssim haver compartit tants bons i estressants moments juntes.

I per últim, però no menys important, m'agradaria donar les gràcies a la meva família, en concret als meus pares, per ajudar-me sempre de manera incondicional i per estar al meu costat, no estaria on sóc ara sense vosaltres. També a la meva tieta Neus, gràcies per donar-me suport i escoltar sempre.

Resum

Títol: Estudi de mètodes de deconvolució

Autora: Nidia Barco Armengol

Tutores: Dra. Júlia Perera Bel (IMIM) i Dra. Susanna Boday Salvans (UVic)

Paraules clau: Deconvolució, expressió gènica, tipus cel·lulars, comparativa

La Deconvolució és una eina computacional que permet identificar i quantificar els tipus cel·lulars presents en una matriu d'expressió d'una mostra bulk o mixta.

Actualment, aquesta tècnica es pot dur a terme amb diversos mètodes, però no hi ha moltes eines que et permetin obtenir els resultats de manera fàcil i visual, i que permetin fer una comparació entre els diferents mètodes.

En aquest projecte, hem creat un paquet d'R que permet utilitzar diferents mètodes i matrius de referència per fer la deconvolució, i que retorna els resultats en forma de gràfics fàcilment interpretables.

Utilitzant aquest paquet hem realitzat un estudi comparatiu fent servir dades simulades, per avaluar la precisió dels mètodes amb diferents matrius de referència. Finalment, també hem estudiat en detall les possibles confusions més comunes entre tipus cel·lulars concrets.

Com a resultats d'aquest projecte, obtenim un paquet d'R que permet fer la deconvolució amb els mètodes CIBERSORT, GEDIT, EPIC, FARDEEP i QuantIseq. A banda de la quantificació dels tipus cel·lulars, el paquet posa el focus en generar gràfics optimitzats per a poder comparar fraccions cel·lulars entre condicions. Per altra banda, els resultats de l'estudi comparatiu mostren que FARDEEP, GEDIT i CIBERSORT son els mètodes amb millors resultats a la deconvolució i que els tipus cel·lulars que s'acostumen a confondre són els macròfags i monòcits i les T CD4 i T CD8.

Com a conclusió podem observar que encara que els mètodes funcionen força bé amb diferents matrius de referència, els resultats estan lluny de ser precisos al 100%, per la qual cosa, és necessari millorar els mètodes i els perfils d'expressió dels tipus cel·lulars de les matrius de referència.

Summary

Title: Study of deconvolution methods

Author: Nidia Barco Armengol

Supervisor: Dra. Júlia Perera Bel (IMIM) and Dra. Susanna Bodoy Salvans (UVic)

Keywords: Deconvolution, gene expression, cell types, comparative

Deconvolution is a computational tool that allows you to identify and quantify the cell types present in an expression matrix of a bulk or mixed sample.

Currently, this technique can be carried out with several methods, but there are not many tools that allow you to obtain the results in an easy and visual way, and to make a comparison between the different methods.

In this project, we have created an R package that allows you to use different methods and reference matrices to perform the deconvolution, and that returns the results in the form of easily interpretable graphs.

Using this package we have performed a comparative study using simulated data, to evaluate the accuracy of the methods with different reference matrices. Finally, we have also studied in detail the most common possible confusions between specific cell types.

As a result of this project, we obtained an R package that allows deconvolution with the CIBERSORT, GEDIT, EPIC, FARDEEP and QuanTIseq methods. Apart from the quantification of cell types, the package focuses on generating optimised plots to compare cell fractions between conditions. Moreover, the results of the comparative study show that FARDEEP, GEDIT and CIBERSORT are the methods with the best deconvolution results and that the cell types that tend to be confused are macrophages and monocytes and CD4 and CD8 T cells.

In conclusion, we can see that although the methods work quite well with different reference matrices, the results are far from 100% accurate, so it is necessary to improve the methods and expression profiles of the cell types of the reference matrices.

Índex

| | |
|--|-----|
| 1. Introducció | 7 |
| 1.1. Que és la deconvolució? | 8 |
| 1.2. Matrius de referència | 9 |
| 1.3. Matriu resultant..... | 11 |
| 1.4. Aplicacions de la deconvolució..... | 11 |
| 1.5. Problemes actuals de la deconvolució..... | 11 |
| 2. Objectius..... | 13 |
| 3. Mètodes..... | 14 |
| 3.1. Selecció dels mètodes de deconvolució | 14 |
| 3.1.1. GEDIT..... | 16 |
| 3.1.2. QuanTlseq | 16 |
| 3.1.3. EPIC | 16 |
| 3.1.4. FARDEEP | 16 |
| 3.1.5. CIBERSORT | 17 |
| 3.2. Matrius de referència..... | 17 |
| 3.3. Creació de matrius mixtes amb dades simulades | 17 |
| 4. Resultats..... | 19 |
| 4.1. Paquet Deconv..... | 19 |
| 4.1.1. Estructura del paquet..... | 19 |
| 4.1.2. Funcions de deconvolució..... | 19 |
| 4.1.3. Funcions extra | 20 |
| 4.1.4. Funcions internes..... | 21 |
| 4.1.5. Paràmetres de les funcions..... | 22 |
| 4.1.6. Instal·lació i ús del paquet..... | 24 |
| 4.1.7. Resultats de les funcions de deconvolució..... | 25 |
| 4.2. Avaluació dels diferents mètodes i matrius de referència amb dades simulades | 26 |
| 4.2.1. Resultats de dades humanes..... | 26 |
| 4.2.2. Resultats de dades de ratolí | 29 |
| 4.2.3. Resultats de les dades de ratolí i matriu de referència humana | 30 |
| 4.2.4. Resultats de les mostres amb representació homogènia de tots els tipus cel·lulars | 31 |
| 4.2.5. Anàlisi detallat de tipus cel·lulars..... | 33 |
| 5. Discussió | 35 |
| 6. Conclusió..... | 37 |
| Bibliografia..... | 39 |
| Annex I: Deconvolució de dades humanes | I |
| Annex II: Deconvolució de dades de ratolí | V |
| Annex III: Deconvolució de dades de ratolí traduïdes | VII |

Lista de figures

| | |
|--|----|
| Fig 1. Esquema general de la Deconvolució | 9 |
| Fig 2. Fraccions reals utilitzades per crear les matrius simulades | 18 |
| Fig 3. Estructura de les funcions per fer deconvolució | 20 |
| Fig 4. Esquema dels gràfics resultants de la funció Box_deconv | 20 |
| Fig 5. Esquema del funcionament de la funció Mouse_Human_converter | 21 |
| Fig 6. Estructura de les matrius mixta i de referència | 22 |
| Fig 7. Figures resultants de les funcions Deconv | 25 |
| Fig 8. Gràfic resultant de la funció Box_Deconv | 26 |
| Fig 9. Comparativa de les fraccions mesurades (esquerra) i les fraccions reals (dreta) pels diferents mètodes de deconvolució | 32 |
| Fig 10. Exemples del reconeixement de macròfags | 33 |
| Fig 11. Exemples del reconeixement de monòcits | 33 |
| Fig 12. Exemple del reconeixement de cèl·lules CD4 | 34 |
| Fig 13. Exemple del reconeixement de cèl·lules CD8 | 34 |

Lista de taules

| | |
|--|------|
| Taula 1. Taula de matrius de referència..... | 10 |
| Taula 2. Mètodes comparats bibliogràficament..... | 14 |
| Taula 3. Taula introductòria als mètodes de deconvolució seleccionats per al paquet d'R | 15 |
| Taula 4. Resum dels resultats de deconvolució amb dades humanes | 27 |
| Taula 5. Resum dels resultats de deconvolució amb dades de ratolí..... | 29 |
| Taula 6. Resum dels resultats de deconvolució amb dades de ratolí traduïdes | 30 |
| Taula 7. Resultats de la deconvolució de dades humanes amb CIBERSORT | I |
| Taula 8. Resultats de la deconvolució de dades humanes amb FARDEEP | I |
| Taula 9. Resultats de la deconvolució de dades humanes amb EPIC..... | II |
| Taula 10. Resultats de la deconvolució de dades humanes amb QuanTIseq | III |
| Taula 11. Resultats de la deconvolució de dades humanes amb GEDIT | IV |
| Taula 12. Resultats de la deconvolució de dades de ratolí amb CIBERSORT | V |
| Taula 13. Resultats de la deconvolució de dades de ratolí amb FARDEEP..... | V |
| Taula 14. Resultats de la deconvolució de dades de ratolí amb EPIC | VI |
| Taula 15. Resultats de la deconvolució de dades de ratolí amb GEDIT | VI |
| Taula 16. Resultats de la deconvolució de dades de ratolí traduïdes amb CIBERSORT | VII |
| Taula 17. Resultats de la deconvolució de dades de ratolí traduïdes amb FARDEEP..... | VIII |
| Taula 18. Resultats de la deconvolució de dades de ratolí traduïdes amb EPIC..... | IX |
| Taula 19. Resultats de la deconvolució de dades de ratolí traduïdes amb QuanTIseq.. | X |
| Taula 20. Resultats de la deconvolució de dades de ratolí traduïdes amb GEDIT | XI |

1. Introducció

En l'actualitat hi ha diverses tècniques per determinar la presència o absència de diversos tipus cel·lulars dins d'una mostra composta per una barreja de cèl·lules de diferents tipus (ex. mostra de teixit limfàtic). Alguns d'aquests mètodes són experimentals, com és el cas de la citometria de flux o de la immunohistoquímica.

Aquests mètodes experimentals presenten diverses limitacions, com la incapacitat d'identificar un gran nombre de tipus cel·lulars simultàniament, i l'elevat cost que suposa el dur a terme aquestes tècniques.

El fet de poder identificar els tipus cel·lulars presents en una mostra i poder saber en quina fracció està cada tipus cel·lular, té una gran utilitat tant en l'àmbit de la recerca com en l'àmbit mèdic. Per tal de solucionar aquesta dificultat, en els últims anys s'han trobat diferents maneres de fer aquesta identificació de manera més eficient i econòmica (Le et al., 2021).

Aquestes noves tècniques s'engloben dins del terme deconvolució, i utilitzen dades òmiques per extreure les proporcions dels tipus cel·lulars presents en una mostra d'interès. Les òmiques són un grup de tecnologies que permeten perfilar i quantificar molècules biològiques presents en una mostra a gran escala. En general, aquestes mostres són complexes i estan formades per un col·lectiu de cèl·lules (Ex. cèl·lules del teixit limfàtic), per tant, la quantificació de les molècules (Ex. expressió gènica) no és sinó la suma de les expressions en diferents cèl·lules de diversos tipus cel·lulars. Aquest tipus de mostres es solen anomenar dades "bulk" (dades que no pertanyen a un tipus cel·lular concret, sinó a un conjunt de tipus cel·lulars).

Entre les aplicacions òmiques més conegudes hi trobem la proteòmica, l'epigenòmica, la genòmica o la transcriptòmica. La transcriptòmica és la més emprada actualment per quantificar la presència dels diferents tipus cel·lulars en una mostra complexa.

La transcriptòmica mesura l'RNA que transcriuen les cèl·lules. Aquesta tècnica és idònia per identificar els tipus cel·lulars, ja que cada tipus cel·lular té un perfil d'expressió gènica diferent, que es veu reflectit en el contingut d'RNA que es transcriu. Aquesta òmica comprèn dos grans tecnologies, els microarrays (que utilitza sondes que s'hibriden a les molècules d'RNA) i la seqüenciació.

La seqüenciació és actualment una tècnica molt utilitzada per obtenir aquests perfils d'expressió, això es deu a que detecta els perfils d'expressió de manera més amplia al no necessitar sondes, sinó que detecta l'expressió de tots els gens transcrits. De totes maneres, la deconvolució es pot fer tant amb dades de microarrays com de

seqüenciació, ja que parteix d'una matriu d'expressió amb perfils d'expressió de gens coneguts.

Actualment, la seqüenciació ha millorat la seva resolució i es poden arribar a seqüenciar cèl·lules úniques (single cell RNA-seq). Aquesta nova aplicació permet mesurar directament el nombre de cèl·lules de cada tipus cel·lular, així com també descobrir i caracteritzar nous subtipus cel·lulars. De totes maneres, és una tecnologia cara i complexa en alguns teixits. Però permet obtenir perfils d'expressió molt precisos dels diferents tipus cel·lulars.

1.1. Que és la deconvolució?

La deconvolució són un conjunt d'eines computacionals que resolen el problema de quantificar les fraccions de tipus cel·lulars presents en una mostra mixta (o bulk). Aquests mètodes parteixen de la formulació segons la qual l'expressió gènica d'una mostra mixta (matriu M) és la suma ponderada (matriu F) dels perfils d'expressió dels tipus cel·lulars barrejats (matriu de referència S) (Sturm et al., 2019).

Tenint en compte que el problema es formula seguint l'equació $M = F * S$ (Newman et al., 2015), la deconvolució intenta resoldre el problema invers. És a dir, intenta trobar la matriu de fraccions dels tipus cel·lulars (F), a partir d'una matriu de referència (S) i l'expressió gènica global (matriu bulk, mixta o M) (Figura 1).

Aquesta tècnica és capaç de treballar amb diferents conjunts de dades, ja siguin transcriptòmiques, epigenòmiques o proteòmiques entre d'altres. En aquest treball ens hem centrat en les dades transcriptòmiques, ja que són les més comunament utilitzades i, per tant, també les que tenen els mètodes més testats i definits.

En el context de la transcriptòmica, les matrius de referència (S), contenen els perfils d'expressió mitjana de cada tipus cel·lular. Aquests perfils es poden obtenir de diverses maneres. A partir de microarrays o a partir de seqüenciació d'una mostra purificada experimentalment, amb tècniques com pot ser el FACS, o més recentment utilitzant la tècnica de single cell RNA-seq.

Per altra banda, es necessita la matriu amb l'expressió global de la mostra bulk que es vol deconvolucionar (M), i que procedeix de microarrays o seqüenciació.

Un cop es tenen les dues matrius, es pot fer la deconvolució amb diferents mètodes, per obtenir com a resultat una matriu F que conté les fraccions de cada tipus cel·lular per mostra. Depenent del mètode, la quantificació dels tipus cel·lulars pot ser fraccions relatives (sobre 1) o absolutes (que mostren la quantitat de cada tipus cel·lular a la mostra).

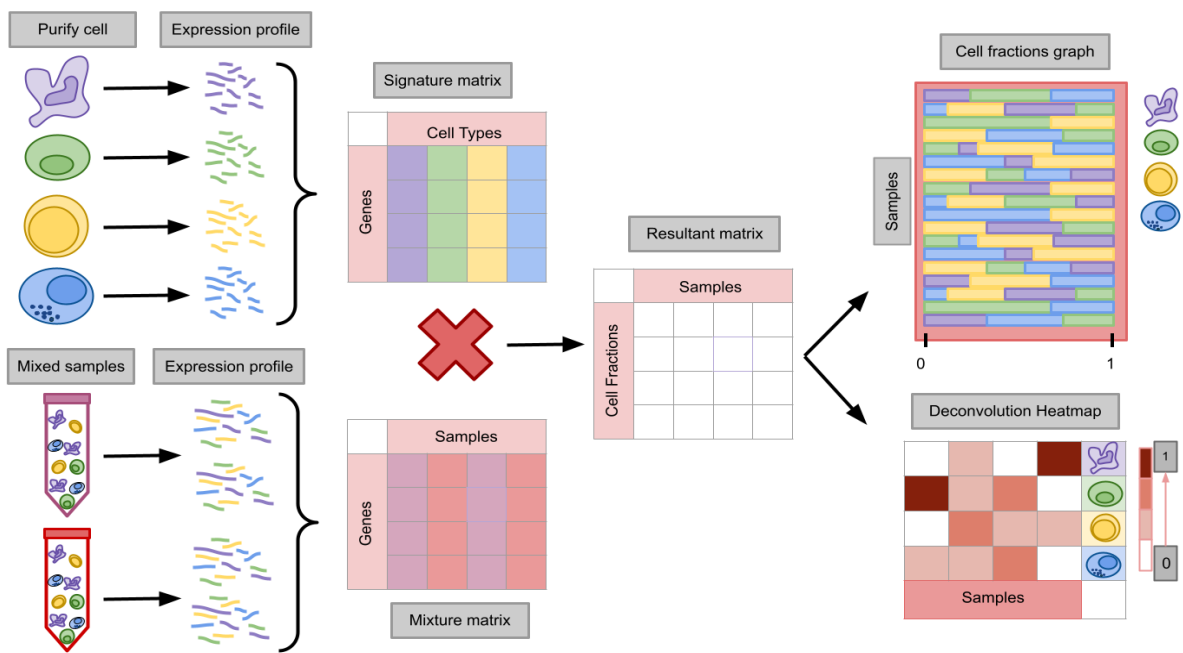


Fig 1. Esquema general de la Deconvolució

1.2. Matrius de referència

Les matrius de referència (S), són les que contenen la informació dels perfils d'expressió gènica de cada tipus cel·lular per poder fer la deconvolució.

Com ja s'ha mencionat anteriorment aquestes matrius de referència poden provenir tant de microarrays o seqüenciació de mostres purificades, o, més recentment, de single cell RNA-seq. Les diferències entre aquestes tecnologies introdueixen biaixos en les mesures d'expressió, per tant, es recomana utilitzar una matriu de referència que sigui el més similar possible a les dades que es volen deconvolucionar (Nadel, Lopez, et al., 2021).

Aquestes matrius poden ser molt variades, però alguns dels punts a tenir en compte al triar una matriu de referència poden ser:

- De quin organisme és la matriu (Ex. "*Homo sapiens*" o "*Mus musculus*")
- D'on provenen les dades (Microarrays, seqüenciació o scRNA-seq)
- Quants tipus cel·lulars contenen
- Quin és el nombre de gens de la matriu i quants d'aquests gens són diferencials per un tipus cel·lular concret.

Tenint en compte aquests punts, el més recomanat per diferents autors, és seleccionar la matriu de referència que s'adapti millor a les teves necessitats i que s'assembli més a les teves dades en organisme i procedència de les dades, ja que diversos estudis

confirmen la importància de la selecció de la matriu de referència per obtenir resultats confiablès, degut a la dependència entre els mètodes i aquestes matrius.

En alguns casos no hi ha disponibilitat de matrius específiques per un organisme concret, pel que també existeix la possibilitat de traduir una matriu utilitzant els gens homòlegs (traduint els HUGO symbols de l'organisme d'origen als de l'organisme estudiat), però en aquests casos s'ha de tenir en compte que pot afectar els resultats.

Les matrius de referència que s'utilitzen en estudis comparatius són aquelles que s'ha vist que donen bons resultats amb diversos mètodes de deconvolució. En la següent taula es mostra un recull de les matrius de referència més utilitzades en la literatura. Algunes d'aquestes matrius han estat generades pels autors de diversos mètodes, com pot ser TIL10, que és una matriu generada pels creadors del mètode QuanTISeq.

Taula 1. Taula de matrius de referència

| Nom de la matriu | Organisme | Nºtipus cel·lulars | Teixit | Referència |
|--|--------------|--------------------|-------------------|-------------------------------|
| 10X immune | Homo sapiens | 9 | Sang | (Zheng et al., 2017) |
| BLUEPRINT | Homo sapiens | 8 | Sang | (Martens & Stunnenberg, 2013) |
| BlueCode | Homo sapiens | 35 | Múltiples teixits | (Nadel, Lopez, et al., 2021) |
| Human Primary Cell Atlas | Homo sapiens | 13 | Múltiples teixits | (Mabbott et al., 2013) |
| ImmunoStates | Homo sapiens | 20 | Sang | (Vallania et al., 2018) |
| LM22 | Homo sapiens | 22 | Sang | (Newman et al., 2015) |
| Skin Signatures | Homo sapiens | 21 | Múltiples teixits | (Swindell et al., 2013) |
| Blood circulating immune cells (BCIC) | Homo sapiens | 6 | Sang | (Racle et al., 2017) |
| Tumor infiltrating cells (TIC) | Homo sapiens | 6 | Melanoma de pell | (Racle et al., 2017) |
| TIL10 | Homo sapiens | 10 | Sang | (Plattner et al., 2020) |
| Wang2020 | Homo sapiens | 6 | Cervell | (Wang et al., 2020) |
| Tabula Muris | Mus musculus | 12 | Múltiples teixits | (Schaum et al., 2018) |
| Mouse Body Atlas | Mus musculus | 20 | Múltiples teixits | (Lattin et al., 2008) |
| ImmGen | Mus musculus | 137 | Múltiples teixits | (Heng et al., 2008) |
| sig_matr_seqImmCC | Mus musculus | 10 | Múltiples teixits | (Z. Chen et al., 2018) |
| Donovan_Brain.mouse | Mus musculus | 7 | Cervell | (Donovan et al., 2020) |

1.3. Matriu resultant

Després d'utilitzar un mètode de deconvolució, el que s'obté és la matriu resultant (F) que conté les fraccions de cada tipus cel·lular dins d'una mostra mixta. Aquesta té una estructura definida, en què trobem a les columnes les diferents mostres i a les files les fraccions de cada tipus cel·lular.

Depenent dels supòsits o de les restriccions que s'apliquin als mètodes, aquestes fraccions poden ser o bé relatives (percentatge sobre 1 en què trobem cada tipus cel·lular per mostra), o bé poden ser absolutes (quantitat total de cada tipus cel·lular a la mostra).

1.4. Aplicacions de la deconvolució

El fet de poder determinar les fraccions de diversos tipus cel·lulars presents en una mostra té una gran rellevància en l'àmbit de l'oncologia (Le et al., 2021). Conèixer quina és la infiltració de cèl·lules immunes dins de l'ambient tumoral és essencial, per poder identificar biomarcadors o per poder oferir un tractament personalitzat per un tumor concret.

La immunoteràpia és una tècnica molt prometedora en el camp de l'oncologia, però la seva eficiència depèn d'una resposta immunitària de base i de l'activació de la immunitat preexistent, per la qual cosa, per utilitzar la immunoteràpia per al tractament contra el càncer és necessari identificar inicialment la infiltració de cèl·lules immunes dins del tumor (Galon & Bruni, 2019).

Per tant, les tècniques de deconvolució, que permeten obtenir aquesta informació amb un menor cost que les tècniques experimentals esmentades inicialment, podrien donar aquesta informació per l'aplicació de la immunoteràpia com a tractament contra el càncer (Kuksin et al., 2021).

1.5. Problemes actuals de la deconvolució

Encara que actualment la deconvolució ja se sol utilitzar en diversos treballs d'investigació, encara no és una tècnica 100% precisa.

Alguns dels problemes que presenta són:

1. La difícil caracterització del perfil d'expressió de les cèl·lules amb mides o morfologies estranyes.

Les cèl·lules que tenen formes o mides estranyes resulten difícils de purificar per extreure els perfils d'expressió precisos, i per això el reconeixement d'aquests

tipus cel·lulars a les mostres mixtes no sempre és acurat, ja que es poden arribar a confondre amb altres tipus cel·lulars (Nadel, Lopez, et al., 2021).

2. La dependència de l'efectivitat depenent de la matriu de referència i les dades a analitzar.

S'ha observat en múltiples estudis, que els mètodes de deconvolució tenen una eficàcia variable depenent de les matrius de referència que s'usin i les dades que s'han d'analitzar (tecnologia, espècie, teixit, selecció de gens...), per la qual cosa no sempre pots utilitzar una matriu predefinida per tots els mètodes, sinó que s'ha de seleccionar depenent del mètode i les dades que s'utilitzin (Nadel, Lopez, et al., 2021).

3. La dificultat per comparar diferents mètodes de deconvolució en les mateixes condicions.

S'han realitzat moltes comparatives entre mètodes, però de vegades els resultats no són del tot comparables.

Això depèn de la procedència de les dades (microarrays o seqüenciació), ja que alguns mètodes treballen millor amb un tipus de dades en concret o permeten especificar el tipus de dades amb què treballen, com és el cas de QuantIseq que permet seleccionar si les dades són de microarrays o de seqüenciació.

Però per evitar aquest problema i fer que els mètodes siguin comparables, actualment, s'estan desenvolupant tècniques per poder generar les dades de manera computacional, com l'eina SimBu (Dietrich et al., 2022), per permetre comparar els mètodes en les mateixes condicions.

2. Objectius

Per tal d'adreçar els problemes mencionats, en aquest treball hem volgut fer una recerca bibliogràfica dels mètodes existents i matrius de referència disponibles. Sabent que no hi ha un mètode perfecte, hem volgut implementar diversos mètodes i proporcionar una col·lecció de matrius de referència en un paquet d'R per tal de poder provar totes les combinacions quan sigui necessari. Després, hem simulat dades d'expressió on hem controlat les proporcions dels tipus cel·lulars, per tal de donar recomanacions de quins mètodes i quines matrius funcionen millor en cada cas.

Així doncs, els objectius concrets d'aquest treball són:

1. Estudiar els actuals mètodes de deconvolució i fer una selecció a partir de dades bibliogràfiques dels millors mètodes.
2. Seleccionar les matrius de referència que millor funcionen amb els diferents mètodes de deconvolució seleccionats.
3. Desenvolupar un paquet d'R amb el que es pugui fer la deconvolució amb diversos mètodes, i generar gràfics que mostrin de manera clara els resultats de la deconvolució.
4. Provar els diferents mètodes amb matrius mixtes simulades per avaluar l'efectivitat de cada mètode i amb cada matriu de referència.

3. Mètodes

3.1. Selecció dels mètodes de deconvolució

Existeixen una gran varietat de mètodes per fer deconvolució, i tots assumeixen que l'expressió gènica global es pot descriure com una relació lineal entre les proporcions de les cèl·lules i els seus perfils d'expressió de referència. És important mencionar que per fer deconvolució les dades que s'introdueixen als mètodes han d'estar normalitzades però no log-transformades (Avila Cobos et al., 2020).

Al llarg dels anys s'han realitzat diverses comparacions entre aquests mètodes utilitzant diverses matrius de referència i diverses dades de mostres mixtes amb fraccions conegudes de tipus cel·lulars per provar la precisió dels diferents mètodes.

En aquest treball hem revisat 4 estudis comparatius de mètodes de deconvolució. A continuació resumim els mètodes inclosos en cada estudi.

Taula 2. Mètodes comparats bibliogràficament

| Mètodes testats | Comparativa 1 (Avila Cobos et al., 2020) | Comparativa 2 (Finotello & Trajanoski, 2018) | Comparativa 3 (Nadel, Oliva, et al., 2021) | Comparativa 4 (Le et al., 2021) |
|-----------------|---|---|---|------------------------------------|
| OLS | X | | | |
| nnls RLR | X | | X | |
| FARDEEP | X | | | |
| CIBERSORT | X | X | X | X |
| DWLS | X | | | |
| Music | X | | | |
| SCDC | X | | | |
| NNML | | X | | |
| EPIC | | X | X | |
| QuantIseq | | X | X | |
| PERT | | X | | |
| DCQ | | | X | |
| DeconvRNAseq | | | X | |
| dtangle | | | X | |
| GEDIT | | | X | |
| MPC counter | | | X | |
| SaVanT | | | X | |
| XCell | | | X | |
| CIBERSORT X | | | | X |
| ssGSEA DM | | | | X |
| Singscore DM | | | | X |

Tenint en compte característiques d'interès i resultats força sòlids de deconvolució amb diferents matrius de referència i diversos conjunts de dades, hem seleccionat 5 mètodes per incloure al paquet.

Taula 3. Taula introductòria als mètodes de deconvolució seleccionats per al paquet d'R

| Nom del mètode | Descripció | Referència |
|------------------|--|------------------------------|
| GEDIT | <ul style="list-style-type: none"> • Genera resultats fiables i pot treballar amb matrius de referència amb molts gens. • Compta amb matrius de referència provades, que funcionen be amb el mètode. • Dona bons resultats amb múltiples plataformes, espècies i matrius de referència. | (Nadel, Lopez, et al., 2021) |
| QuanTlseq | <ul style="list-style-type: none"> • Dona bons resultats amb mostres tumorals. • Té en compte la presència de tipus cel·lulars no presents a la matriu de referència. • Només permet utilitzar la matriu TIL10, per la qual el mètode està optimitzat. | (Plattner et al., 2020) |
| EPIC | <ul style="list-style-type: none"> • Té dues matrius per les que el mètode està optimitzat, però pot treballar amb altres matrius. • Té en compte la presència de tipus cel·lulars no especificats a la matriu de referència. | (Racle & Gfeller, 2020) |
| FARDEEP | <ul style="list-style-type: none"> • Té integrat en el seu algoritme un filtre per detectar i eliminar automàticament els outliers. | (Hao et al., 2019) |
| CIBERSORT | <ul style="list-style-type: none"> • Fa una deconvolució robusta i reproduïble amb diferents tipus de dades. • Permet l'ús no només de dades transcriptòmiques, sinó també epigenòmiques,... | (B. Chen et al., 2018) |

3.1.1. GEDIT

Gene Expression Deconvolution Interactive Tool (GEDIT) és un mètode basat en regressió lineal amb restriccions de no negativitat. Primer, fa una normalització de les dades i una selecció dels gens que presenten una major variabilitat en l'expressió de determinats tipus cel·lulars. A partir d'aquests gens, GEDIT escala l'expressió dels gens, per permetre que tots els gens seleccionats tinguin un valor que sigui comparable en el càlcul de la regressió lineal, amb la qual finalment obtenim les proporcions de cada tipus cel·lular per cada mostra. (Nadel, Lopez, et al., 2021)

3.1.2. QuanTIseq

El mètode de QuanTIseq és de la mateixa família de modes, basat en regressió lineal amb restriccions de no negativitat. Utilitza la matriu de referència TIL10, que diferencia deu tipus de cèl·lules immunes amb gran rellevància per al càncer, i també una proporció per cèl·lules no caracteritzades dins d'aquesta matriu de referència.

Al només poder utilitzar la matriu TIL10 per fer deconvolució aquest mètode només pot treballar amb dades humanes, ja que aquesta matriu deriva d'aquestes dades i el mètode no permet fer modificacions a la matriu.

El que sí permet especificar, és si la teva matriu d'entrada prové de dades tumorals i també si les dades són de microarrays o seqüenciació, per tenir-ho en compte a l'hora de fer la deconvolució. (Plattner et al., 2020)

3.1.3. EPIC

EPIC també es basa en regressió lineal amb restriccions de no negativitat. És un mètode de deconvolució que internament fa un escalat de l'expressió gènica.

Aquest mètode pot realitzar la deconvolució amb diverses matrius de referència. Igual que QuanTIseq, aquest mètode deconvoluciona les mostres de la matriu mixta entre els tipus cel·lulars presents en la matriu de referència i una proporció de cèl·lules no caracteritzades. (Racle & Gfeller, 2020)

3.1.4. FARDEEP

El mètode de deconvolució FARDEEP està dissenyat per tractar amb dades d'expressió gènica força contaminades que es troben habitualment en mostres tumorals, que presenten valors atípics d'expressió, i per això aquest mètode utilitza least trimmed square, un mètode de regressió més robust, per identificar i processar aquests valors atípics.

Aquest mètode, a diferència dels anteriors, permet fer una deconvolució relativa (sobre 1) o una deconvolució absoluta, que retorna els valors quantitius de cada fracció cel·lular (Hao et al., 2019). FARDEEP, també permet utilitzar matrius de referència personalitzades.

3.1.5. CIBERSORT

CIBERSORT és un dels mètodes de deconvolució més comunament utilitzats, que es basa en l'ús d'una tècnica d'aprenentatge automàtic, a partir de la regressió de vectors de suport, per obtenir un vector que contingui les fraccions de cada tipus de cèl·lules.

CIBERSORT permet als usuaris utilitzar matrius de referència personalitzades, i com FARDEEP, CIBERSORT pot fer una deconvolució relativa o una deconvolució absoluta de les dades d'expressió (B. Chen et al., 2018). Però la versió que s'ha utilitzat és del 2016, ja que és l'única que es pot utilitzar amb R, doncs la versió actualitzada només es troba a la web de CIBERSORT.

3.2. Matrius de referència

Hem seleccionat 9 matrius de referència (Taula 1) per incloure al paquet d'R (EPIC-BCIC, EPIC-TIC, TIL10, LM22, seqImmuCC, Tabula Muris, 10 XImmune, Donovan brain mouse i Wang 2020). Aquestes matrius s'han descarregat dels articles originals i s'han formatjat totes per poder ser utilitzades directament dins del paquet d'R.

Les matrius BCIC, TIC i Tabula Muris difereixen de la resta ja que contenen tots els gens perfilats (mes de 20000 gens per matriu). El mètode CIBERSORT no funciona amb matrius de referència que tinguin un elevat nombre de gens, per aquest motiu es va realitzar una selecció dels 500, 1000 i 2000 gens amb els majors coeficients de variació (CV) de la matriu i vam fer la deconvolució amb els tres subconjunts de cada matriu de referència que el mètode no pugues utilitzar.

3.3. Creació de matrius mixtes amb dades simulades

Per tal de provar els diferents mètodes de deconvolució seleccionats i el paquet creat amb R per fer la deconvolució, vam generar dades simulades.

Per a la creació de matrius bulk vam utilitzar el paquet d'R SimBu (Dietrich et al., 2022). Aquest paquet genera, a partir de dades de single cell RNAseq, matrius que simulen mostres amb unes fraccions determinades de diferents tipus cel·lulars.

En aquest cas es va utilitzar la funció del paquet que permet generar les dades a partir de dades de single cell obtingudes del repositori Sfaira (Fischer et al., 2021), on trobem dades de single cell RNA-seq per conjunts de dades públiques.

D'aquest repositori vam seleccionar dos conjunts de dades.

- Un conjunt de dades humanes de parènquima pulmonar, que estaven disponibles al repositori, provinents d'un estudi del 2019. (homosapiens_lungparenchyma_2019_10x3v2_madisoosn_001_10.1186/s13059-019-1906-x) (Madisoosn et al., 2019).
- Un conjunt de dades de ratolí també de pulmó, que al igual que les primeres es trobaven disponibles al repositori, procedents d'un estudi del 2020. (musmusculus_lung_2019_10x3v2_pisco_022_10.1038/s41586-020-2496-1) (Almanzar et al., 2020)

En primer lloc, vam fer dues matrius mixtes (humà i ratolí) en què cada mostra era d'un únic tipus cel·lular pur (Figura 2). Així al fer la deconvolució es pot observar el percentatge d'identificació dels tipus cel·lular amb diferents mètodes i matrius de referència. I també per veure amb quins altres tipus cel·lulars s'acostumen a confondre.

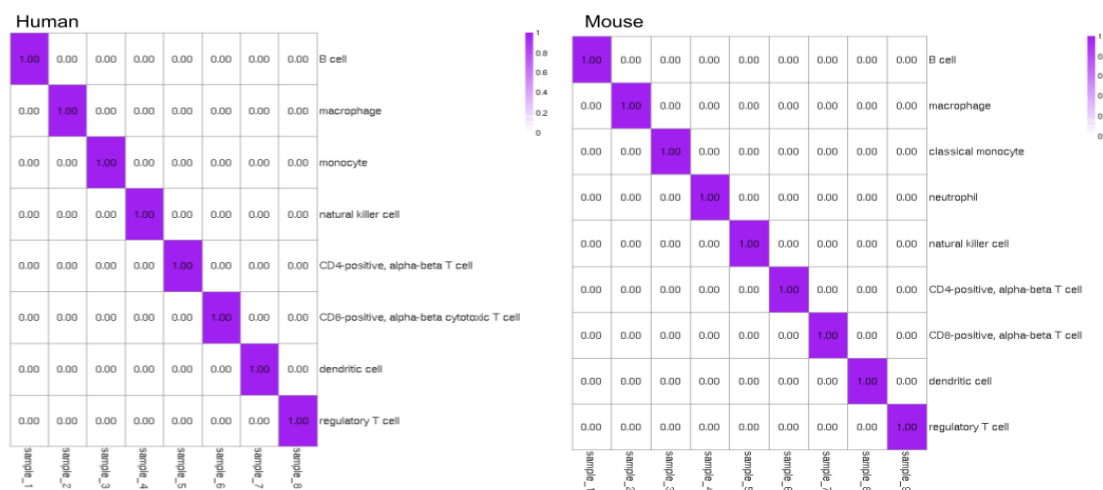


Fig 2. Fraccions reals utilitzades per crear les matrius simulades

En segon lloc, vam crear dues matrius (una amb dades humanes i l'altre amb dades de ratolí) que contenen les mateixes fraccions de cada tipus cel·lular seleccionat per veure com els diferents mètodes deconvolucionen les mostres mixtes i si els resultats s'acosten a les fraccions reals.

Totes les matrius es van normalitzar amb CPM's (Counts Per Million) amb el factor de normalització TMM, que en l'article de (Dillies et al., 2013) mostra bons resultats al normalitzar dades simulades.

Per últim, a aquestes dades no se'ls va fer una transformació logarítmica encara que és freqüent en moltes aplicacions, ja que a l'article (Avila Cobos et al., 2020) es mostra que sense log-transformar la deconvolució dona millors resultats.

4. Resultats

Com a resultats d'aquest projecte tenim dos apartats principals, en primer lloc, la creació del paquet d'R Deconv (disponible a: <https://github.com/margenomics/Deconv>), que serveix per fer deconvolució amb diferents mètodes i per generar gràfics que mostrin els resultats de la deconvolució de manera clara i visual. I, per altra banda, tenim la prova dels diferents mètodes i matrius de referència amb les dades simulades generades amb SimBu.

4.1. Paquet Deconv

4.1.1. Estructura del paquet

Per aquest projecte hem creat un paquet d'R per fer deconvolució amb els mètodes seleccionats anteriorment (taula 3).

Aquest paquet consta de set funcions principals que utilitza l'usuari. Cinc funcions són per fer la deconvolució amb els diferents mètodes seleccionats, i dues funcions extra: una funció per comparar condicions per cada tipus cel·lular amb un gràfic de boxplot i una funció que tradueix els Hugo Symbols de ratolí als seus homòlegs d'organisme humà (per poder utilitzar matrius de referència derivades de dades humanes amb matrius mixtes originàries de ratolí).

4.1.2. Funcions de deconvolució

Les funcions per fer deconvolució amb els diferents mètodes són: CIVERSORT_Deconv, FARDEEP_Deconv, EPIC_Deconv, QuanTIseq_Deconv i GEDIT_Deconv, que generen els gràfics de deconvolució i també retornen un data frame amb la matriu resultant de la deconvolució.

Les funcions que utilitzen els mètodes de CIVERSORT i FARDEEP tenen una opció per poder seleccionar entre una deconvolució relativa o absoluta depenent del que vulgui l'usuari, com es pot observar a la figura 3.

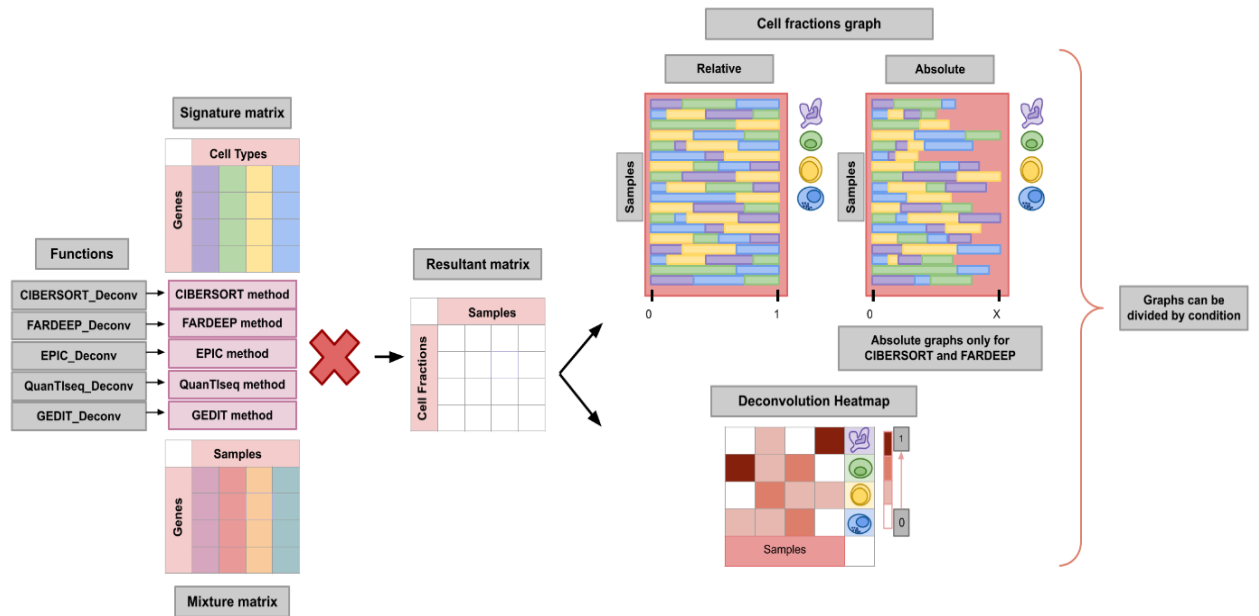


Fig 3. Estructura de les funcions per fer deconvolució

4.1.3. Funcions extra

La funció Box_deconv permet utilitzar el resultat de la deconvolució per comparar el contingut d'un tipus cel·lular entre condicions (ex. Quantitat de T CD4 en teixit normal vs en teixit tumoral). Per una altra banda, parteix de la matriu resultant de la deconvolució (resultant matrix figura 3) i d'un vector de condició. Aquesta funció agrupa les fraccions de cada tipus cel·lular de les mostres de cada condició, i genera un gràfic de caixes per cada tipus cel·lular en què es comparen les diferents condicions (Figura 4).

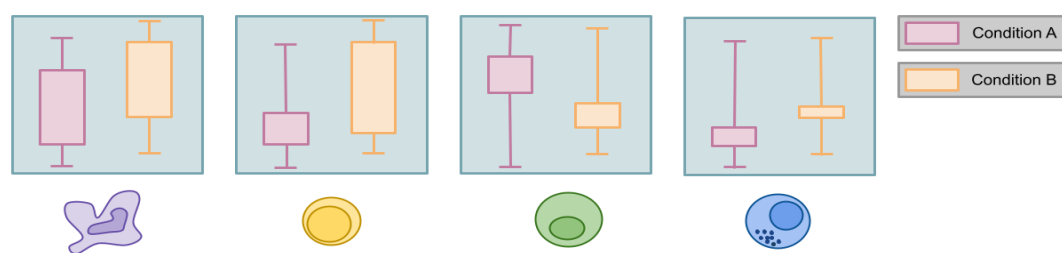


Fig 4. Esquema dels gràfics resultants de la funció Box_deconv

Per altra banda, la funció `Mouse_Human_converter` és una funció que tradueix els Hugo Symols de ratolí d'una matriu mixta als seus homòlegs humans, permetent utilitzar matrius de referència derivades de dades humanes en dades de ratolí. És important tenir en compte que, tot i que es tradueixen els noms dels gens, la diferència entre les dades pot tenir un efecte important en l'especificitat dels mètodes de deconvolució.

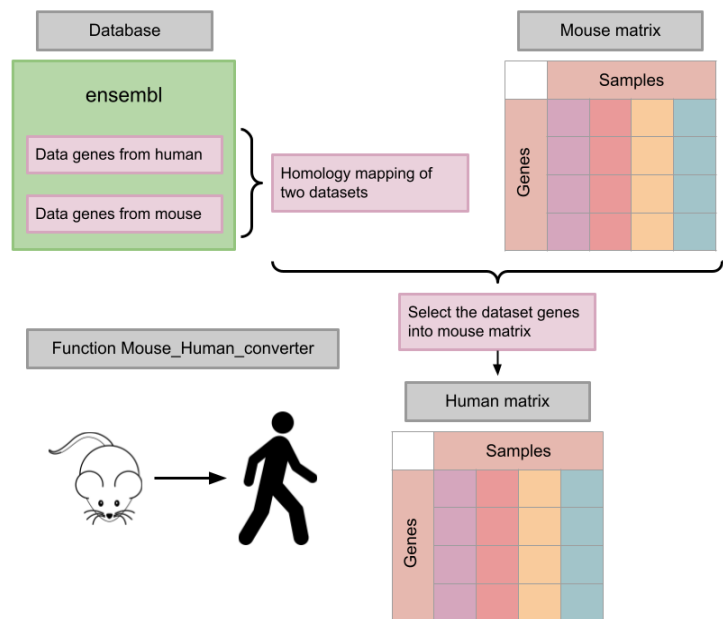


Fig 5. Esquema del funcionament de la funció `Mouse_Human_converter`

4.1.4. Funcions internes

A part de les set funcions principals el paquet també conté quatre funcions internes que són cridades per altres funcions del paquet.

- `Preprocesment_EPIC_FARDEEP`
Funció que processa les matrius d'entrada per fer la deconvolució amb els mètodes d'EPIC, FARDEEP i QuantIseq. Amb aquesta funció es treuen els gens duplicats que hi puguin haver a les matrius, tant a la de referència com a la mixta, i es fa una mitjana de l'expressió d'aquests gens duplicats.
- `GEDIT_functions`
Conjunt de les funcions necessàries per fer la deconvolució amb GEDIT.
- `Deconvolution_graph`
Funció que genera el gràfic de barres apilades que mostra els resultats de les fraccions de cada tipus cel·lulars dins de les mostres mixtes. Genera tant els gràfics amb totes les mostres com els gràfics dividits per condició.
- `Heatmap_graph`
Funció que genera el heatmap amb les dades de la matriu resultant de la deconvolució. Genera tant el heatmap amb totes les dades com amb les dades dividides per condició.

El paquet també compta amb un document RData, en què es troben algunes de les matrius de referència (LM22, BCIC, TIC, seqImmuCC, TabulaMuris, TIL10, 10Ximmune,

Donovan Brain Mouse i Wang2020) referenciades a la taula 1, amb l'estructura requerida per utilitzar-se amb les diverses funcions per fer deconvolució, exceptuant la deconvolució amb CIBERSORT, ja que aquest mètode necessita que les matrius s'introdueixin en format de document, ja sigui .txt, .csv o .tsv.

4.1.5. Paràmetres de les funcions

Les funcions tenen diferents arguments que permeten a l'usuari personalitzar alguns aspectes dels resultats que retornen les funcions.

Les funcions que fan la deconvolució amb els diferents mètodes tenen alguns arguments comuns entre totes les funcions. Aquests arguments són:

- Matrix → Matriu mixta en format Data frame amb les dades d'expressió per mostra, amb l'estructura que es mostra a la figura 6.

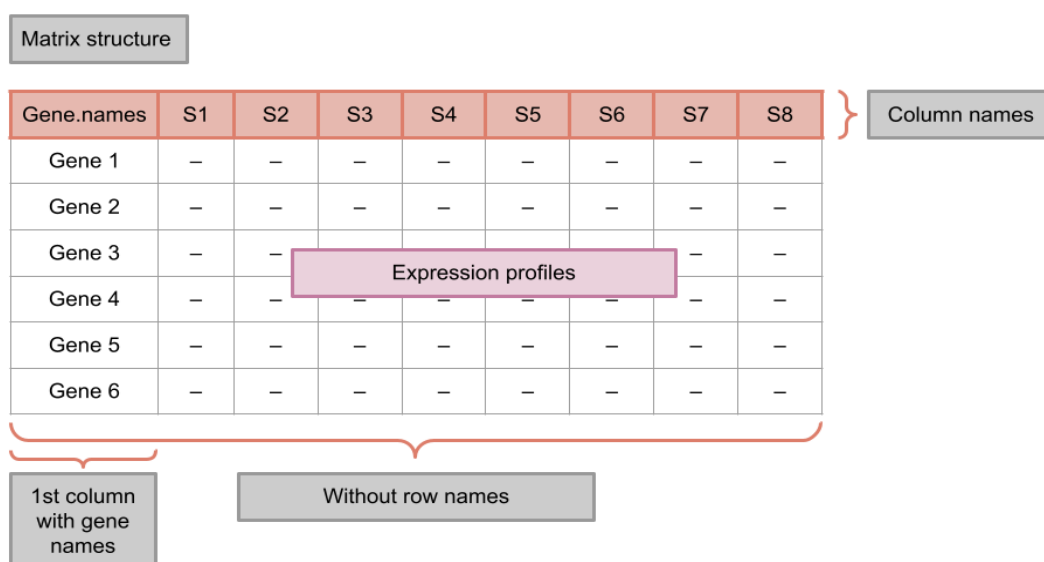


Fig 6. Estructura de les matrius mixta i de referència

- Sig.matrix → Matriu de referència per fer la deconvolució amb la mateixa estructura de la figura 6.
- Results_dir → Directori on es generaran els gràfics de deconvolució. (Ex. "/home/nidia/Deconv")
- Name → Paraula o especificació, que es vol que estigui en el nom dels fitxers .png que es generen.
(Ex. Name: LM22 → "FARDEEP_Deconv_LM22_rel_plot.png").
Per defecte el nom és ("Method_Deconv/Heatmap_rel/abs_(cond)_plot.png")
- ByCond → Paràmetre lògic que indica si s'introduirà un vector de condició, per generar els gràfics per cada condició. Per defecte és FALSE.

- Cond → Vector amb les condicions en cas d'haver introduït TRUE al paràmetre anterior.
- Data4Tyers → Objecte amb una estructura concreta data4Tyers (format intern de la unitat de bioinformàtica de l'IMIM) perquè la funció seleccioni la matriu mixta d'aquest objecte. Per defecte el paràmetre és NULL, i si s'introdueix aquest element no s'ha d'introduir la matriu mixta.
- Arguments de visualització → Height_Deconv, Width_Deconv, Height_heatmap, Width_heatmap, Number_format, X.size, Y.size i L.size. Són arguments que permeten personalitzar l'aspecte dels gràfics que mostren els resultats de la deconvolució.

A part d'aquests paràmetres que són comuns a les diferents funcions hi ha algunes funcions que tenen paràmetres extra o que no necessiten algun d'aquests paràmetres.

En el cas de la funció CIBERSORT_Deconv al necessitar que les matrius estiguin en format de fitxer no pot tenir l'argument data4Tyers, ja que necessita fer una selecció de la matriu, i al fer la selecció aquesta no està en format de fitxer.

Però, per altra banda, té un paràmetre extra que té en comú amb la funció FARDEEP_Deconv, el paràmetre method que permet seleccionar si la deconvolució és absoluta o relativa.

A part la funció FARDEEP_Deconv té dos paràmetres extra, intercept i permn_number que estan explicats dintre del help de la funció a l'R, ja que al ser un mètode més restrictiu, les dades han de tenir uns requisits mínims per fer la deconvolució i amb aquests paràmetres es poden variar aquests requisits.

En el cas de la funció Quantlseq_Deconv a l'utilitzar sempre la seva pròpia matriu de referència (TIL10), no permet introduir matrius de referència externes, i per això no té el paràmetre sig.matrix. Però té dos paràmetres extra, Array i Tumor que permeten donar més informació sobre les dades per fer una deconvolució més acurada.

Per acabar hi ha dues funcions que tenen els seus arguments propis, primerament, la funció `Mouse_Human_converter` té tres arguments.

- `Matrix` → Matriu de ratolí de la qual es vol fer la conversió a gens humans. I ha de tenir la mateixa estructura de la figura 6.
- `Resultant_file` → Nom que l'usuari vol que tingui el fitxer que genera la funció amb la matriu convertida. (Ex. "counts_mf_human.txt")
- `Results_dir` → Directori on es vol generar el fitxer amb la matriu convertida.

I en segon lloc, tenim la funció `Box_Deconv`, que té quatre arguments.

- `Data` → Matriu resultant de la deconvolució.
- `Cond` → Vector amb les condicions de cada mostra.
- `Results_dir` → Directori on es vol generar el gràfic de boxplot.
- `F_name` → Nom que l'usuari vol que tingui el gràfic resultant. Per defecte és "Deconvolution Boxplot".

4.1.6. Instal·lació i ús del paquet

Per utilitzar el paquet `Deconv`, en primer lloc, s'ha d'instal·lar el paquet que està disponible al github de `margenomics`.

```
library(devtools)
install_github("margenomics/Deconv")
```

Per seleccionar les matrius de referència disponibles dins del paquet s'ha de fer la selecció com es mostra a continuació.

```
sig.matrix→ Deconv::LM22
sig.matrix→ Deconv::EPIC_BCIC
sig.matrix→ Deconv::EPIC_TIC
sig.matrix→ Deconv::seqImmCC
sig.matrix→ Deconv::TabulaMuris
sig.matrix→ Deconv::TIL10
sig.matrix→ Deconv::Ximmune10
sig.matrix→ Deconv::Donovan_Brain.mouse
sig.matrix→ Deconv::Wang2020
```

I alguns exemples per utilitzar les diferents funcions es mostren a continuació.

```
R → CIBERSORT_Deconv(matrix = matrix, sig.matrix = sig.matrix, results_dir = results_dir,
cibersortpath = path_ciber, by_Cond = TRUE, cond = fractions, method = "abs")
```

```
R → FARDEEP_Deconv(matrix = matrix, sig.matrix = sig.matrix, results_dir = results_dir,
byCond = TRUE, cond = fractions, method = "abs")
```

```
R → EPIC_Deconv(matrix = matrix, sig.matrix = sig.matrix, results_dir = results_dir,
byCond = TRUE, cond = fractions)
```

```
R → QuanTIseq_Deconv(matrix = matrix, results_dir = results_dir, byCond = TRUE, cond =
fractions)
```

```
R → GEDIT_Deconv(matrix = matrix, sig.matrix = sig.matrix, results_dir = results_dir,
byCond = TRUE, cond = fractions)
```

```
M → Mouse_Human_converter(matrix = matrix,resultant_file = "counts_mf_human.txt",
results_dir = results_dir)
```

```
Box_Deconv(data= C_abs, cond= fractions, results_dir = "/bicoh/nidia/Deconv/", f_name=
"Box_CIBERSORT_abs")
```

4.1.7. Resultats de les funcions de deconvolució

El paquet Deconv genera a partir de la matriu resultant dos gràfics per mostrar els resultats de la deconvolució com es veu a la figura 7.

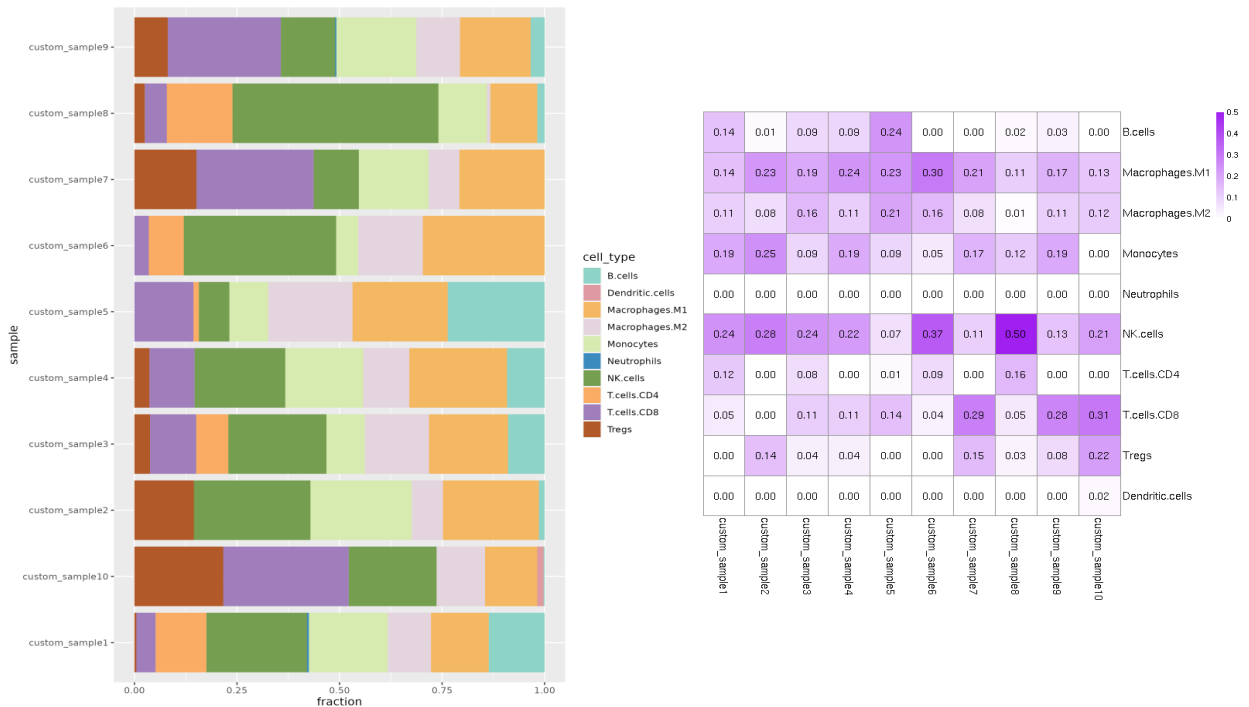


Fig 7. Figures resultants de les funcions Deconv

I la funció `Box_Deconv` genera també a partir de la matriu de resultats de la deconvolució el següent gràfic i un t test comparant les fraccions entre condicions.

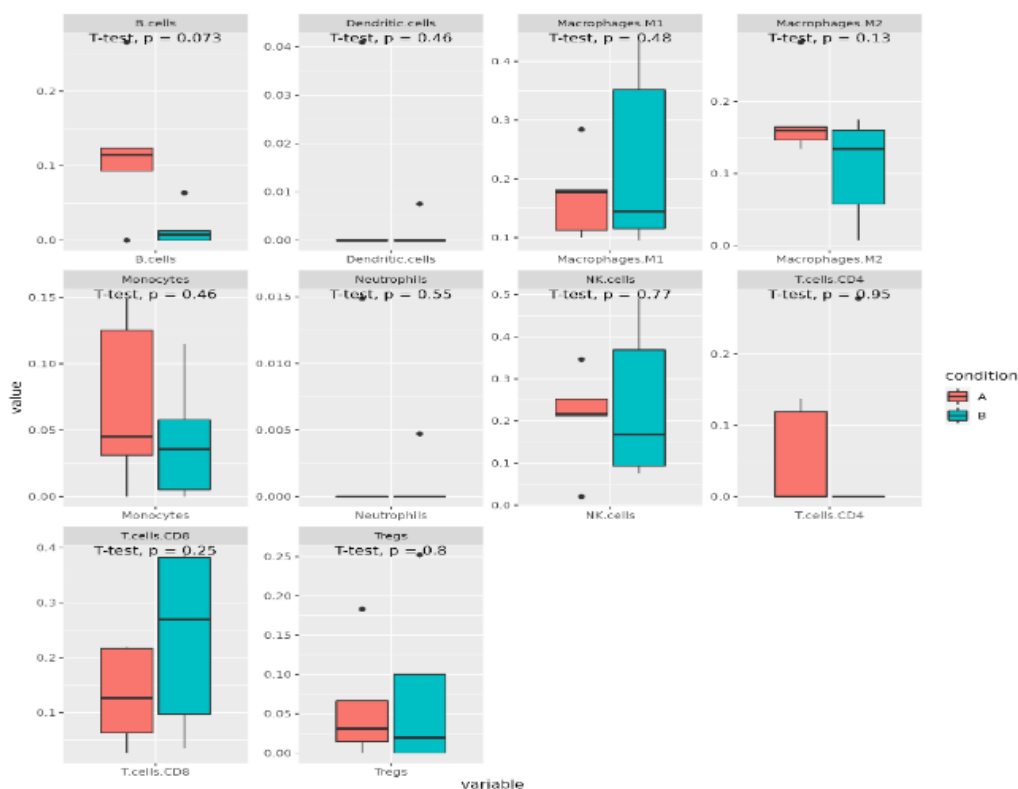


Fig 8. Gràfic resultant de la funció `Box_Deconv`

4.2. Avaluació dels diferents mètodes i matrius de referència amb dades simulades

En aquesta secció es mostren els resultats de deconvolució de dades simulades utilitzant el paquet d'R desenvolupat. L'objectiu és comparar els diferents mètodes i matrius de referència per veure el seu rendiment en diferents escenaris i fer recomanacions a l'hora de triar els mètodes més adequats. S'han simulat dos escenaris: el primer escenari és amb mostres pures (cada mostra conté un 100% d'un tipus cel·lular diferent); el segon escenari són mostres amb un percentatge igual de tots els tipus cel·lulars. Hem fet l'anàlisi amb dades humanes i dades de ratolí amb referències del mateix organisme, i també hem creuat organismes (referència i matriu mixta) per veure l'impacte en la deconvolució.

4.2.1. Resultats de dades humanes

Partint de les dades simulades ja normalitzades, es va fer la deconvolució amb els mètodes CIBERSORT, FARDEEP, EPIC, GEDIT i QuantIseq, amb les matrius de referència LM22, TIL10, EPIC_BCIC i EPIC_TIC per les dades humanes, sent

QuanTseq l'únic mètode que només es va provar amb TIL10, ja que és l'única matriu que pot utilitzar.

Taula 4. Resum dels resultats de deconvolució amb dades humanes

| | | % >= 65% | | 35% < % > 60% | | % =< 35% | | No avaluat | | | | | | | | | | | | |
|-------------------------|-------------|-----------|----------|---------------|-------|----------|------------------------|------------|--|------------------------|-------------|-----------|----------|------|-------|-------|------------------------|--------|---|--|
| Matriu LM22 | | | | | | | | | | Matriu TIL10 | | | | | | | | | | |
| Dades Humanes | Cèl·lules B | Macròfags | Monòcits | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's | | Dades Humanes | Cèl·lules B | Macròfags | Monòcits | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's | | |
| CIBERSORT | 43% | 23% | 18% | 51% | 26% | 15% | 24% | 8% | | CIBERSORT | 88% | 61% | 41% | 85% | 66% | 42% | 8% | 88% | | |
| FARDEEP | 92% | 87% | 69% | 89% | 86% | 63% | 70% | 67% | | FARDEEP | 90% | 92% | 20% | 82% | 74% | 40% | 0% | 71% | | |
| EPIC | 64% | 72% | 11% | 67% | 40% | 23% | 27% | 40% | | EPIC | 70% | 69% | 1% | 99% | 35% | 31% | 0% | 58% | | |
| QuanTseq | - | - | - | - | - | - | - | - | | QuanTseq | 53% | 59% | 0% | 96% | 44% | 49% | 0% | 58% | | |
| GEDIT | 73% | 82% | 41% | 88% | 65% | 40% | 26% | 43% | | GEDIT | 77% | 87% | 9% | 98% | 62% | 48% | 0% | 77% | | |
| Matriu EPIC_BCIC | | | | | | | | | | Matriu EPIC_TIC | | | | | | | | | | |
| Dades Humanes | Cèl·lules B | Macròfags | Monòcits | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's | | Dades Humanes | Cèl·lules B | Macròfags | Monòcits | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's | | |
| CIBERSORT | - | - | - | - | - | - | - | - | | CIBERSORT | - | - | - | - | - | - | - | - | - | |
| FARDEEP | 63% | - | 95% | 67% | 79% | 53% | - | - | | FARDEEP | 57% | 97% | - | - | 97% | 4% | - | - | - | |
| EPIC | 53% | - | 40% | 81% | 70% | 0% | - | - | | EPIC | 27% | 64% | - | - | 89% | 0% | - | - | - | |
| QuanTseq | - | - | - | - | - | - | - | - | | QuanTseq | - | - | - | - | - | - | - | - | - | |
| GEDIT | 57% | - | 81% | 90% | 73% | 0% | - | - | | GEDIT | 35% | 99% | - | - | 100% | 0% | - | - | - | |

En la taula 4 es pot observar el percentatge de reconeixement que tenen els diferents mètodes de deconvolució amb les matrius testades. En aquests resultats s'observa com els mètodes globalment presenten un reconeixement major al 50% en la majoria dels casos per les cèl·lules B, exceptuant CIBERSORT amb LM22 i EPIC amb la matriu TIC.

Els macròfags també tenen un percentatge de reconeixement que en la majoria de casos es troba al voltant o superior al 60% en tots els mètodes, exceptuant novament el mètode CIBERSORT amb LM22 que només mostra un 23% de reconeixement.

Per al cas dels monòcits aquest percentatge de reconeixement disminueix respecte els tipus cel·lulars anteriors, en aquest cas els percentatges de reconeixement es troben majoritàriament al voltant d'un 20% o menor, sent l'excepció els mètodes FARDEEP i GEDIT que amb les matrius LM22 i BCIC, que mostren uns percentatges de reconeixement superiors a la resta. També CIBERSORT amb TIL10 té un reconeixement d'un 41%.

Les cèl·lules NK's en general tenen un alt percentatge de reconeixement, en la gran majoria dels casos aquest percentatge supera el 65% de reconeixement, exceptuant CIBERSORT amb LM22 que té un reconeixement una mica inferior, del 51%.

En canvi, per les cèl·lules T CD4 el percentatge de reconeixement varia més entre els mètodes i matrius. Els millors percentatges de reconeixement els presenten els mètodes CIBERSORT, FARDEEP, GEDIT i EPIC, el primer amb la matriu TIL10, el segon amb les quatre matrius provades i el tercer també amb les quatre sent la matriu TIL10 la que presenta un menor percentatge de reconeixement del 62%, i per últim el quart mètode presenta alts percentatges de reconeixement per les matrius BCIC i TIC.

En el cas de les cèl·lules T CD8 els percentatges de reconeixement disminueixen molt respecte als observats per les cèl·lules CD4, els majors percentatges ronden el 50% i s'obtenen principalment amb la matriu TIL10 amb tots els mètodes exceptuant EPIC.

Les cèl·lules dendrítiques són les que en general tenen els percentatges de reconeixement més baixos, sobretot amb la matriu TIL10 on la majoria dels percentatges de reconeixement són del 0%, l'únic mètode amb un percentatge de reconeixement per sobre del 65% és el mètode de FARDEEP amb LM22.

Per acabar les cèl·lules T reguladores presenten amb la matriu TIL10 uns percentatges de reconeixement al voltant o per sobre del 60%, mentre que amb LM22 disminueixen una mica estant al voltant del 50%, amb l'excepció del mètode CIBERSORT que presenta un reconeixement del 8%.

Amb el mètode CIBERSORT no es van poder extreure resultats concloents amb les matrius EPIC_BCIC i EPIC_TIC, perquè les matrius contenen un nombre de gens massa elevat i el mètode no treballa bé amb dades d'aquesta mida, per la qual cosa vam fer subsets dels gens amb majors coeficients de variació (CV's). Els resultats obtinguts amb 500, 1000 o 200 gens variaven molt depenent del nombre de gens seleccionats, pel que els resultats no són fiables i no han estat inclosos en el treball.

En resum, amb dades humanes els mètodes que ha obtingut millors resultats han estat FARDEEP i GEDIT amb les quatre matrius provades, i CIBERSORT amb la matriu TIL10.

4.2.2. Resultats de dades de ratolí

Partint de les dades simulades de ratolí es va fer deconvolució amb els mètodes FARDEEP, EPIC i GEDIT amb les matrius Tabula Muris i SeqImmucC (Taula 5).

Taula 5. Resum dels resultats de deconvolució amb dades de ratolí

| Matriu TabulaMuris | | | | | | | | | | Matriu SeqImmucC | | | | | | | | | |
|--------------------|-------------|-----------|----------|------------|------|-------|-------|------------------------|--------|------------------|-------------|-----------|----------|------------|------|-------|-------|------------------------|--------|
| Dades de ratolí | Cèl·lules B | Macròfags | Monòcits | Neutròfils | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's | Dades de ratolí | Cèl·lules B | Macròfags | Monòcits | Neutròfils | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's |
| CIBERSORT | - | - | - | - | - | - | - | - | - | CIBERSORT | 78% | 96% | 10% | 68% | 92% | 41% | 98% | 87% | - |
| FARDEEP | 98% | - | 74% | - | 98% | - | - | - | - | FARDEEP | 86% | 91% | 9% | 19% | 94% | 39% | 97% | 97% | - |
| EPIC | 87% | - | 21% | - | 85% | - | - | - | - | EPIC | 94% | 9% | 0% | 37% | 62% | 18% | 22% | 87% | - |
| QuantIseq | - | - | - | - | - | - | - | - | - | QuantIseq | - | - | - | - | - | - | - | - | - |
| GEDIT | 81% | - | 15% | - | 73% | - | - | - | - | GEDIT | 93% | 82% | 0% | 14% | 26% | 31% | 48% | 97% | - |

En aquestes dades podem observar el següent. Per a les cèl·lules B el percentatge de reconeixement és elevat, en tots els casos supera el 75% de reconeixement.

En el cas dels macròfags podem observar que els mètodes CIBERSORT, FARDEEP i GEDIT tenen un alt percentatge de reconeixement superior al 80%, però EPIC presenta un reconeixement de només el 9%.

En el cas dels monòcits el percentatge de reconeixement és molt baix amb les dues matrius i els diferents mètodes, a excepció de FARDEEP amb la matriu Tabula Muris que presenta un reconeixement del 74%.

Els neutròfils igual que els monòcits també presenten percentatges de reconeixement baixos, sent l'obtingut amb CIBERSORT el més elevat amb un 68% de reconeixement, valor que varia molt dels obtinguts amb la resta de mètodes.

Les cèl·lules natural killer (NK's) presenten un alt percentatge de reconeixement amb el mètode CIBERSORT amb seqImmucC i FARDEEP per les dues matrius, mentre que EPIC i GEDIT només presenten resultats majors al 70% amb la matriu Tabula Muris.

Les cèl·lules T CD4 tenen valors relativament baixos de reconeixement, sent els majors els obtinguts amb CIBERSORT i FARDEEP.

En el cas de les T CD8 els percentatges de reconeixement augmenten respecte als de les CD4, però els mètodes amb millors resultats segueixen sent CIBERSORT i FARDEEP amb percentatges majors al 95%.

I per últim les cèl·lules dendrítiques presenten uns bons percentatges de reconeixement superiors al 80% per tots els mètodes amb la matriu SeqImmucC.

En aquests resultats no estan presents els del mètode CIBERSORT amb TabulaMuris, ja que es repetia la situació anterior amb TIC i BCIC, i per això els resultats no són fiables.

En resum, amb dades de ratolí els millors mètodes són CIBERSORT i FARDEEP, mentre que en el cas de les dades humanes els millors mètodes són FARDEEP, GEDIT i CIBERSORT.

4.2.3. Resultats de les dades de ratolí i matriu de referència humana

Taula 6. Resum dels resultats de deconvolució amb dades de ratolí traduïdes

| Matriu LM22 | | | | | | | | | | Matriu TIL10 | | | | | | | | | |
|--------------------------|-------------|-----------|----------|------------|------|-------|-------|------------------------|--------|--------------------------|-------------|-----------|----------|------------|------|-------|-------|------------------------|--------|
| Dades de ratolí - humana | Cèl·lules B | Macròfags | Monòcits | Neutròfils | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's | Dades de ratolí - humana | Cèl·lules B | Macròfags | Monòcits | Neutròfils | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's |
| CIBERSORT | 70% | 25% | 23% | 50% | 51% | 30% | 20% | 36% | 2% | CIBERSORT | 73% | 47% | 0% | 38% | 70% | 2% | 96% | 30% | 76% |
| FARDEEP | 97% | 81% | 69% | 83% | 68% | 41% | 90% | 69% | 58% | FARDEEP | 94% | 23% | 2% | 66% | 76% | 0% | 91% | 72% | 43% |
| EPIC | 80% | 28% | 4% | 54% | 50% | 12% | 79% | 48% | 20% | EPIC | 78% | 47% | 0% | 86% | 64% | 1% | 87% | 12% | 55% |
| QuanTlseq | - | - | - | - | - | - | - | - | - | QuanTlseq | 98% | 39% | 0% | 41% | 70% | 0% | 100% | 4% | 89% |
| GEDIT | 78% | 43% | 18% | 29% | 47% | 6% | 87% | 49% | 22% | GEDIT | 79% | 71% | 0% | 29% | 83% | 3% | 90% | 16% | 62% |

| Matriu EPIC_BCIC | | | | | | | | | | Matriu EPIC_TIC | | | | | | | | | |
|--------------------------|-------------|-----------|----------|------------|------|-------|-------|------------------------|--------|--------------------------|-------------|-----------|----------|------------|------|-------|-------|------------------------|--------|
| Dades de ratolí - humana | Cèl·lules B | Macròfags | Monòcits | Neutròfils | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's | Dades de ratolí - humana | Cèl·lules B | Macròfags | Monòcits | Neutròfils | NK's | CD4's | CD8's | Cèl·lules dendrítiques | Treg's |
| CIBERSORT | - | - | - | - | - | - | - | - | - | CIBERSORT | - | - | - | - | - | - | - | - | - |
| FARDEEP | 93% | - | 60% | 90% | 45% | 93% | 0% | - | - | FARDEEP | 92% | 59% | - | - | - | 68% | 3% | - | - |
| EPIC | 95% | - | 10% | 100% | 48% | 65% | 0% | - | - | EPIC | 91% | 16% | - | - | - | 38% | 0% | - | - |
| QuanTlseq | - | - | - | - | - | - | - | - | - | QuanTlseq | - | - | - | - | - | - | - | - | - |
| GEDIT | 95% | - | 38% | 100% | 57% | 77% | 0% | - | - | GEDIT | 100% | 52% | - | - | - | 43% | 0% | - | - |

Per veure l'impacte d'utilitzar una matriu de referència derivada d'un altre organisme, vam traduir els Hugo Symbols de les dues matrius mixtes de ratolí a humana i vam fer la deconvolució de les dades per tots els mètodes amb les matrius humanes. Els resultats obtinguts es mostren a la taula 6.

Pel que fa a les cèl·lules B els percentatges de reconeixement són iguals o superiors al 70% per tots els mètodes i matrius de referència.

El percentatge de reconeixement dels macròfags és més variat depenent del mètode i la matriu. Els millors resultats s'obtenen amb FARDEEP i LM22, i amb GEDIT i TIL10.

Per als monòcits el percentatge de reconeixement en la majoria dels casos és molt baix no superior al 25%, exceptuant amb el mètode de FARDEEP amb LM22 que obté un 69% de reconeixement.

Els neutròfils obtenen els millors percentatges de reconeixement amb FARDEEP per LM22, TIL10 i BCIC, amb EPIC per TIL10 i amb GEDIT per BCIC.

El reconeixement de les cèl·lules NK obté els millors percentatges de reconeixement amb la matriu TIL10 i amb el mètode FARDEEP per LM22 .

El percentatge de reconeixement de les cèl·lules T CD4 es mostra força baix amb LM22 i TIL10, però obté alts percentatges amb BCIC, superiors al 65% i amb el mètode FARDEEP per TIC.

Les CD8, en canvi, presenten els pitjors resultats per les matrius TIC i BCIC, i mostren bons resultats per la matriu TIL10 i per la matriu LM22 a excepció del mètode CIBERSORT que presenta un percentatge del 20%.

Les cèl·lules dendrítiques únicament mostren percentatges superiors al 65% per al mètode FARDEEP, però en general els percentatges són majors si s'utilitza la matriu LM22 en lloc de la TIL10.

I per acabar les cèl·lules T reguladores presenten millors resultats amb la matriu TIL10, al voltant del 60% de reconeixement, mentre que amb la matriu LM22 l'únic mètode que obté percentatges de reconeixement similars és FARDEEP.

Comparativament, als resultats obtinguts de les dades humanes i de les dades de ratolí traduïdes, es pot observar que encara que els percentatges de reconeixement disminueixen per alguns tipus cel·lulars amb les dades traduïdes, els valors de reconeixement no mostren una variació molt notòria, encara i així es pot observar que el mètode al qual menys afecta aquesta traducció és FARDEEP amb la matriu LM22, ja que és el que manté els percentatges de reconeixement més pròxims als obtinguts amb les matrius de ratolí abans de fer la traducció.

4.2.4. Resultats de les mostres amb representació homogènia de tots els tipus cel·lulars

Per estudiar el comportament dels mètodes quan tots els tipus cel·lulars estan presents en la mateixa proporció en una mostra, vam fer la deconvolució utilitzant els 5 mètodes amb la matriu TIL10, que segons els resultats anteriors té bons percentatges de reconeixement en la majoria dels tipus cel·lulars. Les fraccions estimades pels mètodes en comparació amb les fraccions reals de la mostra mixta es poden observar a la figura 9.

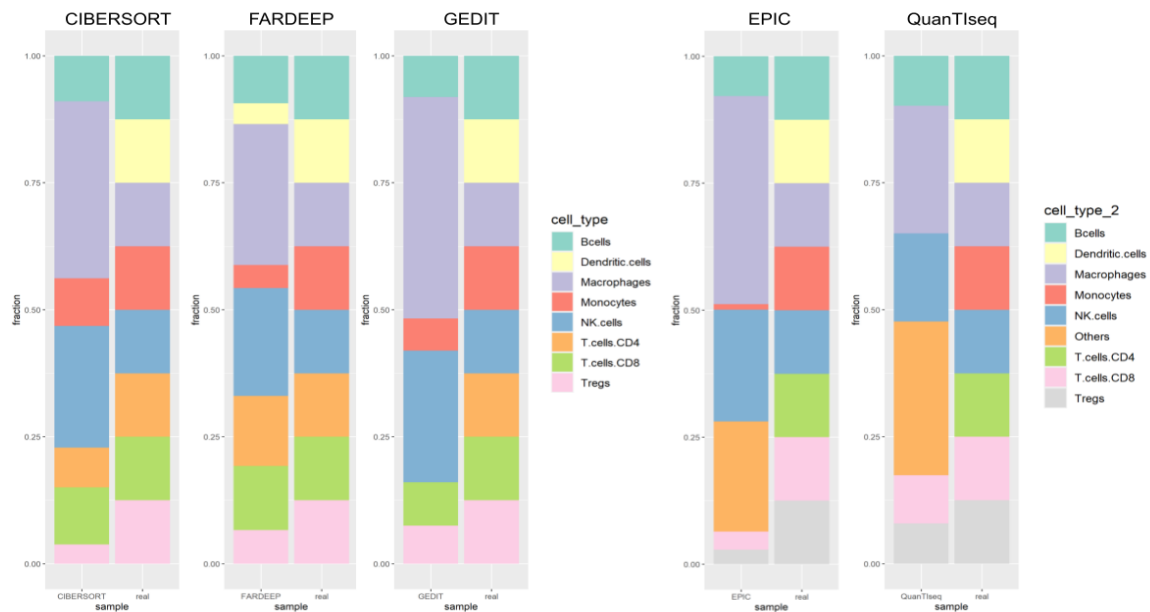


Fig 9. Comparativa de les fraccions mesurades (esquerra) i les fraccions reals (dreta) pels diferents mètodes de deconvolució

A la figura 9 es pot observar una tendència a sobreestimar els macròfags en detriment de les cèl·lules dendrítiques i també de les cèl·lules B. Sent els mètodes EPIC i GEDIT els que més sobreestimen aquesta fracció.

Un altra tendència que comparteixen els diferents mètodes, és una sobreestimació de les cèl·lules NK, en aquest cas en detriment de les cèl·lules T CD4, T reguladores i T CD8, encara que en aquestes últimes el detriment és menor.

Els resultats mostren dos patrons en la quantificació, depenent de si el mètode permet quantificar tipus cel·lulars no presents a la matriu de referència ("other"), com és el cas d'EPIC i QuantIseq, o si pel contrari es veu forçat a assignar proporcions a tots els tipus cel·lulars.

En el primer cas, la conseqüència és que hi pot haver poblacions sense representació, com és el cas de les poblacions de cèl·lules T CD4 o les cèl·lules dendrítiques amb EPIC.

I en el segon cas, la conseqüència és que hi pot haver una sobreestimació de les poblacions degut a la presència de tipus cel·lulars no presents a la matriu de referència.

4.2.5. Anàlisi detallat de tipus cel·lulars

En tots els resultats anteriors s'ha pogut observar que hi ha tipus cel·lulars que tendeixen a confondre's entre si en múltiples mètodes i amb diferents matrius de referència.

A continuació es mostren detalladament quins tipus cel·lulars s'acostumen a confondre entre ells.

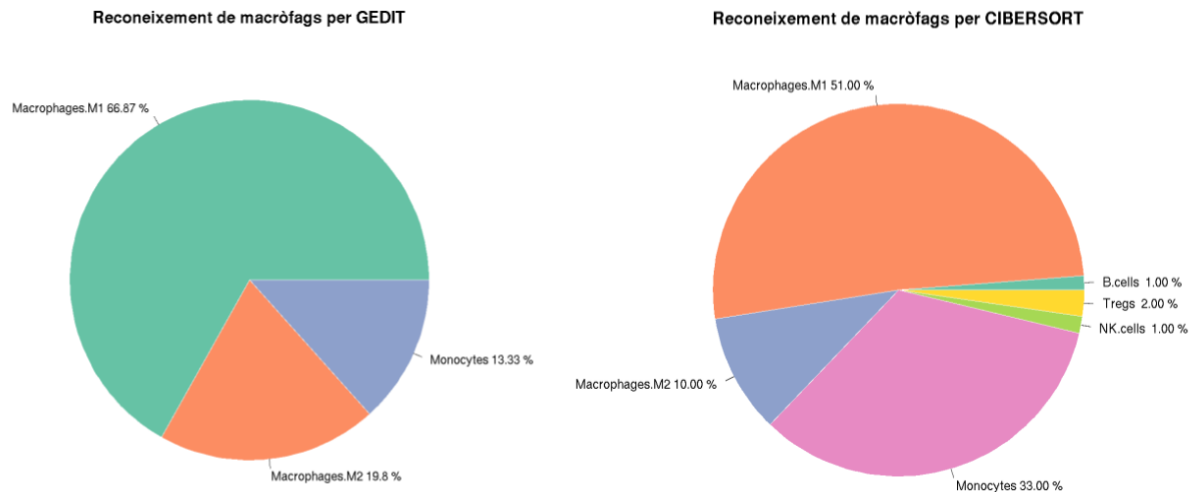


Fig 10. Exemples del reconeixement de macròfags

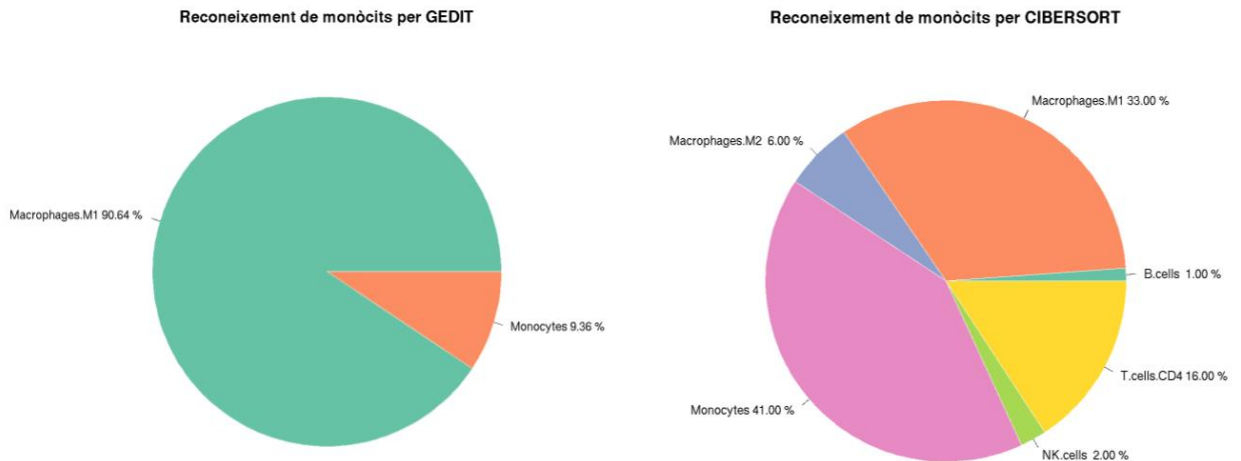


Fig 11. Exemples del reconeixement de monòcits

Els macròfags i els monòcits tendeixen a confondre's. En les figures 12 i 13 es poden observar dos exemples (GEDIT i CIBERSORT) dels errors d'assignació de tipus cel·lulars més freqüents en macròfags i monòcits. Es pot observar que els macròfags tendeixen a classificar-se bé en la gran majoria (un 60-85%) però una fracció es

classifica com a monòcits (13-33%). En canvi, els monòcits en la gran majoria es classifiquen com a macròfags (40-90%) o altres cèl·lules.

Aquests casos, com ja s'ha mencionat anteriorment, són recurrents en els diferents mètodes de deconvolució i amb diferents matrius de referència. Aquest efecte probablement és degut a la similitud en els perfils d'expressió d'aquests dos tipus cel·lulars.

Cal tenir en compte aquest resultat a l'hora d'utilitzar les fraccions d'aquests tipus cel·lulars estimades amb deconvolució, ja que tenen una alta possibilitat d'estar sobre o subestimades.

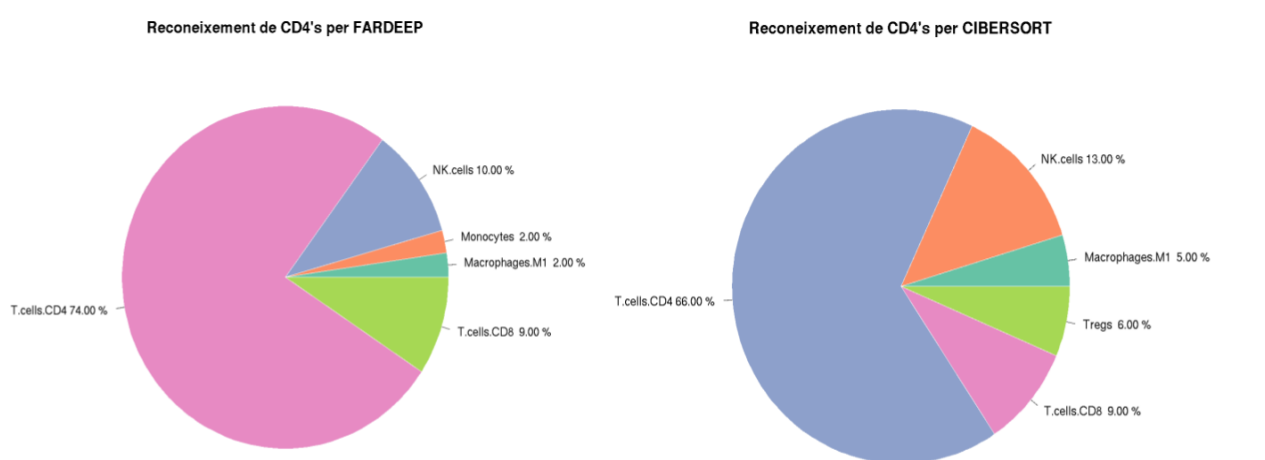


Fig 12. Exemple del reconeixement de cèl·lules CD4

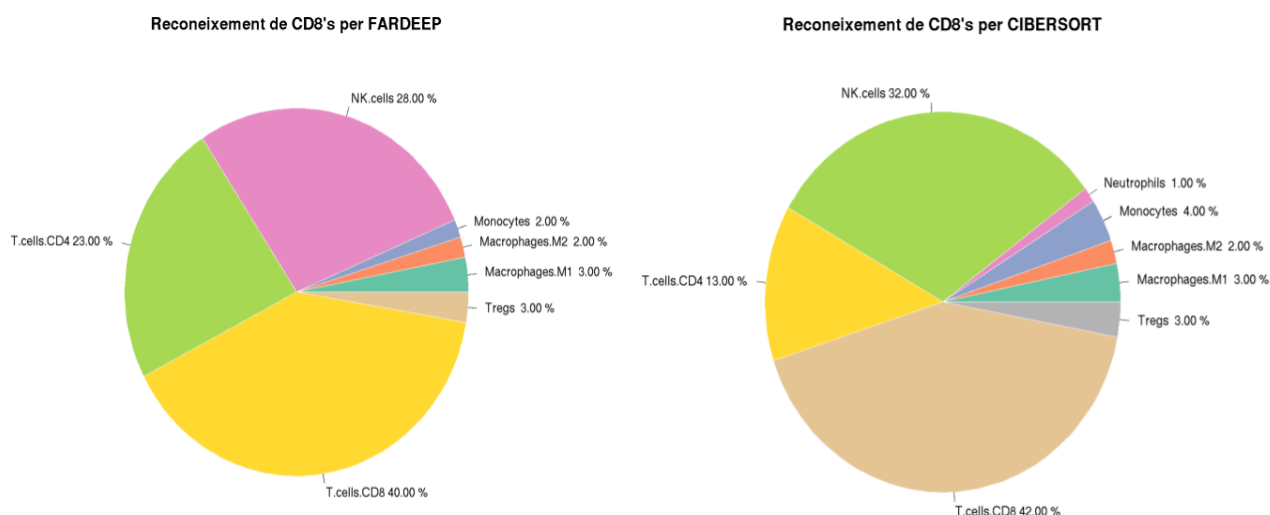


Fig 13. Exemple del reconeixement de cèl·lules CD8

Un altre cas és el de les cèl·lules T CD4 i les T CD8 que a part de confondre's entre elles també s'acostumen a confondre amb les cèl·lules natural killer.

A les figures 14 i 15 es poden observar els percentatges de reconeixement d'aquest tipus cel·lulars per als mètodes CIBERSORT i FARDEEP, i es pot veure clarament que en la majoria dels casos les cèl·lules T CD4 és subestimen a favor de les T CD8 i NK's, i en el cas de les CD8's a favor de cèl·lules CD4 i NK's.

Les CD4 i CD8 també són un exemple de cèl·lules amb perfils d'expressió molt similars entre elles i que no estan tan ben definits perquè també es confonen amb NK's.

5. Discussió

Partint de la recerca realitzada, de l'anàlisi de deconvolució, amb mostres homogènies i heterogènies de diferents tipus cel·lulars immunes, i de l'anàlisi detallat per tipus cel·lulars concrets, es poden extreure diverses conclusions.

En primer lloc, pel que respecta als diferents mètodes de deconvolució, es pot observar que presenten característiques diferencials entre ells. En el cas de FARDEEP, té un filtre més restrictiu dels gens per intentar minimitzar la presència d'outliers a les mostres, que poden empitjorar la qualitat de la deconvolució. O en el cas de QuanTlseq, que permet especificar si les dades provenen d'una mostra tumoral o normal i si provenen de seqüenciació o microarrays per adequar el mètode a les dades que s'han de deconvolucionar, i que també, juntament amb EPIC, es diferencien dels altres, ja que permeten la no classificació d'alguns tipus cel·lulars, que no encaixen amb cap dels perfils d'expressió de la matriu de referència.

Per altra banda, s'ha comprovat als anàlisis realitzats, que els mètodes presenten una gran dependència de les matrius de referència i les dades a analitzar. En concret, no hem obtingut bons resultats amb el mètode CIBERSORT amb la matriu LM22. Aquest resultat és sorprenent, ja que és un dels mètodes més àmpliament utilitzat, i a les comparatives bibliogràfiques que s'han revisat (Avila Cobos et al., 2020), (Finotello & Trajanoski, 2018), (Nadel, Oliva, et al., 2021) i (Le et al., 2021), CIBERSORT és un dels mètodes amb millor rendiment. Aquest fet pot ser degut a que nosaltres hem utilitzat dades simulades a partir de dades de scRNA-seq, que poden diferir una mica de les dades de RNAseq i de microarrays amb les que s'ha generat i provat la matriu LM22. Un altre explicació pot ser degut a la versió de CIBERSORT utilitzada, ja que el paquet funciona amb un script antic de CIBERSORT que és l'únic que és de codi obert. Actualment, la versió en desenvolupament només està disponible a la web de CIBERSORT¹.

¹ <https://cibersortx.stanford.edu/>

Els nostres resultats evidencien que, és crucial fer una selecció prèvia de les matrius i els mètodes, depenent de les dades que es volen analitzar per obtenir uns resultats òptims de la deconvolució. El paquet desenvolupat en aquest treball, pretén ser d'ajuda als investigadors a l'hora de provar diferents mètodes i matrius de referència i decidir el més adequat en cada cas, i és diferència del paquet *immunedeconv* (Sturm et al., 2019) que és el paquet més conegut que implementa diversos mètodes per fer deconvolució, perquè té mètodes diferents i no genera gràfics per mostrar els resultats de la deconvolució.

Un altre punt d'interès que s'ha trobat en aquest anàlisi ha estat l'alt percentatge de confusió entre tipus cel·lulars amb perfils d'expressió similars (ex. Macròfags i monòcits o T CD4, T CD8 i NK's). En general s'ha vist que hi havia una tendència a sobreestimar els macròfags i NK's en detriment dels monòcits i CD8/CD4, respectivament. Aquests mateixos resultats també s'havien observat a l'article (Sturm et al., 2019). Per tant, és necessari fer la deconvolució amb diferents mètodes i matrius, per aconseguir una visió global del contingut cel·lular de la mostra i detectar possibles sobreestimacions o subestimacions d'aquests tipus cel·lulars que són similars en els perfils d'expressió. S'ha de tenir en compte que l'ús de matrius de referència amb un ampli nombre de tipus cel·lulars pot no ser la millor opció, ja que si es té una matriu de referència amb molts subtipus d'un mateix llinatge, és probable que els perfils d'expressió siguin molt semblants entre si i, per tant, hi hagi més probabilitat de confusió.

Referent a les dades de ratolí, es pot observar com les matrius de referència existents per fer la deconvolució no donen en general els millors percentatges de reconeixement per alguns mètodes, i en alguns casos, les matrius provades són massa grans, el que dificulta el seu ús amb mètodes com *CIBERSORT*, que no està preparat per treballar amb matrius tan grans.

Tenint en compte que els resultats usant matrius traduïdes són força semblants als resultats de les matrius humanes, aquesta pot ser una bona opció quan no es tingui una bona matriu de referència del mateix organisme / teixit. Els millors resultats observats s'han obtingut amb *FARDEEP*, que al ser un mètode restrictiu amb els outliers, segurament minimitza l'efecte del biaix que s'introdueix a l'utilitzar homòlegs.

Segons els resultats d'aquest estudi podem concloure que els millors mètodes serien *GEDIT*, *CIBERSORT* i *FARDEEP*. El primer, dona uns bons resultats tant amb dades humanes com amb dades de ratolí, funciona de manera estable amb diferents matrius de referència i amb un temps d'execució curt. El segon dona també resultats de deconvolució força bons amb diferents matrius per dades humanes, i els millors resultats

amb dades de ratolí, encara que té l'inconvenient de que només pot treballar amb matrius de referència força petites, i tot i ser petites els temps d'execució són força llargs. El tercer és el que obté resultats més sòlids amb la gran majoria de matrius, és el mètode que millor aproxima les fraccions de la mostra mixta i és el mètode al qual menys afecta el canvi d'organisme.

La resta dels mètodes de deconvolució, encara que també funcionen força bé en la majoria dels casos, els resultats que es poden observar són més variables depenent de la matriu de referència utilitzada. A banda, tots presenten punts negatius que s'han de tenir en compte al seleccionar un mètode. QuanTIseq per la seva part, té la limitació de la matriu de referència, el fet de que només es pugui utilitzar amb la matriu TIL10 limita molt el mètode, ja que només et permet utilitzar dades humanes o traduïdes. Finalment, EPIC és un mètode que tendeix a sobreestimar tipus cel·lulars amb més facilitat que la resta de mètodes, com s'ha pogut observar a l'apartat 4.2.4, és el mètode que més sobreestima als macròfags, i amb els anàlisis de les mostres pures, té un dels percentatges de reconeixement més baixos per diversos tipus cel·lulars.

La deconvolució és un cap d'investigació en procés d'evolució constant. Actualment, amb la nova disponibilitat de les tècniques de single cell RNAseq es preveu que es podran aconseguir perfils d'expressió (matrius de referència) molt més específics que els que tenim disponibles fins ara per cada tipus cel·lular, el que permetrà millorar l'especificitat i la quantificació dels tipus cel·lulars de mostres mixtes diverses. Els experiments de single cell RNA-seq, permetran generar noves matrius de referència de teixits concrets, de subtipus cel·lulars vinculats a la localització, edat i context de múltiples organismes. D'altra banda, també s'estan desenvolupant mètodes de deconvolució basats en deep learning. Tots aquests avenços tecnològics refinaran els mètodes de deconvolució i permetran obtenir resultats més exactes i cada vegada més propers a la realitat (Kuksin et al., 2021).

6. Conclusió

Existeixen molts mètodes per fer deconvolució, però cap d'ells dona bons resultats de manera universal. A més, els mètodes depenen molt de les matrius de referència utilitzades. En aquest treball hem creat un paquet d'R que fa la deconvolució de manera automàtica i genera gràfics que mostren els resultats en un format clar, facilitant molt l'ús i avaluació de la deconvolució.

En aquest treball hem mostrar les característiques dels diferents mètodes, les seves fortaleces i limitacions, i oferim recomanacions per la selecció dels mètodes i matrius de referència.

Bibliografia

- Almanzar, N., Antony, J., Baghel, A. S., Bakerman, I., Bansal, I., Barres, B. A., Beachy, P. A., Berdnik, D., Bilen, B., Brownfield, D., Cain, C., Chan, C. K. F., Chen, M. B., Clarke, M. F., Conley, S. D., Darmanis, S., Demers, A., Demir, K., de Morree, A., ... Zou, J. (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, *583*(7817). <https://doi.org/10.1038/s41586-020-2496-1>
- Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P., & De Preter, K. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-19015-1>
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., & Alizadeh, A. A. (2018). Profiling tumor infiltrating immune cells with CIBERSORT. In *Methods in Molecular Biology* (Vol. 1711). https://doi.org/10.1007/978-1-4939-7493-1_12
- Chen, Z., Quan, L., Huang, A., Zhao, Q., Yuan, Y., Yuan, X., Shen, Q., Shang, J., Ben, Y., Qin, F. X. F., & Wu, A. (2018). seq-ImmuCC: Cell-centric view of tissue transcriptome measuring cellular compositions of immune microenvironment from mouse RNA-seq data. *Frontiers in Immunology*, *9*(JUN). <https://doi.org/10.3389/fimmu.2018.01286>
- Dietrich, A., Sturm, G., Merotto, L., Marini, F., Finotello, F., & List, M. (2022). *SimBu: Bias-aware simulation of bulk RNA-seq data with variable cell type composition*. *30*, 1–7. <https://doi.org/10.1101/2022.05.06.490889>
- Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, N. S., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., & Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, *14*(6). <https://doi.org/10.1093/bib/bbs046>
- Donovan, M. K. R., D'Antonio-Chronowska, A., D'Antonio, M., & Frazer, K. A. (2020). Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-14561-0>
- Finotello, F., & Trajanoski, Z. (2018). Quantifying tumor-infiltrating immune cells from transcriptomics data. In *Cancer Immunology, Immunotherapy* (Vol. 67, Issue 7). <https://doi.org/10.1007/s00262-018-2150-z>
- Fischer, D. S., Dony, L., König, M., Moeed, A., Zappia, L., Heumos, L., Tritschler, S., Holmberg, O., Aliee, H., & Theis, F. J. (2021). Sfaira accelerates data and model reuse in single cell genomics. *Genome Biology*, *22*(1). <https://doi.org/10.1186/s13059-021-02452-6>
- Galon, J., & Bruni, D. (2019). Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. In *Nature Reviews Drug Discovery* (Vol. 18, Issue 3). <https://doi.org/10.1038/s41573-018-0007-y>
- Hao, Y., Yan, M., Heath, B. R., Lei, Y. L., & Xie, Y. (2019). Fast and robust deconvolution of tumor infiltrating lymphocyte from expression profiles using least trimmed squares. *PLoS Computational Biology*, *15*(5). <https://doi.org/10.1371/journal.pcbi.1006976>
- Heng, T. S. P., Painter, M. W., Elpek, K., Lukacs-Kornek, V., Mauermann, N., Turley, S. J., Koller, D., Kim, F. S., Wagers, A. J., Asinovski, N., Davis, S., Fassett, M., Feuerer, M., Gray, D. H. D., Haxhinasto, S., Hill, J. A., Hyatt, G., Laplace, C.,

- Leatherbee, K., ... Kang, J. (2008). The immunological genome project: Networks of gene expression in immune cells. In *Nature Immunology* (Vol. 9, Issue 10). <https://doi.org/10.1038/ni1008-1091>
- Kuksin, M., Morel, D., Aglave, M., Danlos, F. X., Marabelle, A., Zinovyev, A., Gautheret, D., & Verlingue, L. (2021). Applications of single-cell and bulk RNA sequencing in onco-immunology. *European Journal of Cancer*, 149. <https://doi.org/10.1016/j.ejca.2021.03.005>
- Lattin, J. E., Schroder, K., Su, A. I., Walker, J. R., Zhang, J., Wiltshire, T., Saijo, K., Glass, C. K., Hume, D. A., Kellie, S., & Sweet, M. J. (2008). Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Research*, 4(1). <https://doi.org/10.1186/1745-7580-4-5>
- Le, T., Aronow, R. A., Kirshtein, A., & Shahriyari, L. (2021). A review of digital cytometry methods: Estimating the relative abundance of cell types in a bulk of cells. In *Briefings in Bioinformatics* (Vol. 22, Issue 4). <https://doi.org/10.1093/bib/bbaa219>
- Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C., & Hume, D. A. (2013). An expression atlas of human primary cells: Inference of gene function from coexpression networks. *BMC Genomics*, 14(1). <https://doi.org/10.1186/1471-2164-14-632>
- Madisson, E., Wilbrey-Clark, A., Miragaia, R. J., Saeb-Parsy, K., Mahbubani, K. T., Georgakopoulos, N., Harding, P., Polanski, K., Huang, N., Nowicki-Osuch, K., Fitzgerald, R. C., Loudon, K. W., Ferdinand, J. R., Clatworthy, M. R., Tsingene, A., Van Dongen, S., Dabrowska, M., Patel, M., Stubbington, M. J. T., ... Meyer, K. B. (2019). ScRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. *Genome Biology*, 21(1). <https://doi.org/10.1186/s13059-019-1906-x>
- Martens, J. H. A., & Stunnenberg, H. G. (2013). BLUEPRINT: Mapping human blood cell epigenomes. In *Haematologica* (Vol. 98, Issue 10). <https://doi.org/10.3324/haematol.2013.094243>
- Nadel, B. B., Lopez, D., Montoya, D. J., Ma, F., Waddel, H., Khan, M. M., Mangul, S., & Pellegrini, M. (2021). The Gene Expression Deconvolution Interactive Tool (GEDIT): accurate cell type quantification from gene expression data. *GigaScience*, 10(2). <https://doi.org/10.1093/gigascience/giab002>
- Nadel, B. B., Oliva, M., Shou, B. L., Mitchell, K., Ma, F., Montoya, D. J., Mouton, A., Kim-Hellmuth, S., Stranger, B. E., Pellegrini, M., & Mangul, S. (2021). Systematic evaluation of transcriptomics-based deconvolution methods and references using thousands of clinical samples. *Briefings in Bioinformatics*, 22(6). <https://doi.org/10.1093/bib/bbab265>
- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., Hoang, C. D., Diehn, M., & Alizadeh, A. A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*, 12(5). <https://doi.org/10.1038/nmeth.3337>
- Plattner, C., Finotello, F., & Rieder, D. (2020). Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. In *Methods in Enzymology* (Vol. 636). <https://doi.org/10.1016/bs.mie.2019.05.056>
- Racle, J., de Jonge, K., Baumgaertner, P., Speiser, D. E., & Gfeller, D. (2017). Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *ELife*, 6. <https://doi.org/10.7554/eLife.26476>
- Racle, J., & Gfeller, D. (2020). EPIC: A tool to estimate the proportions of different cell

types from bulk gene expression data. In *Methods in Molecular Biology* (Vol. 2120). https://doi.org/10.1007/978-1-0716-0327-7_17

- Schaum, N., Karkanias, J., Neff, N. F., May, A. P., Quake, S. R., Wyss-Coray, T., Darmanis, S., Batson, J., Botvinnik, O., Chen, M. B., Chen, S., Green, F., Jones, R. C., Maynard, A., Penland, L., Pisco, A. O., Sit, R. V., Stanley, G. M., Webber, J. T., ... Weissman, I. L. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, *562*(7727). <https://doi.org/10.1038/s41586-018-0590-4>
- Sturm, G., Finotello, F., Petitprez, F., Zhang, J. D., Baumbach, J., Fridman, W. H., List, M., & Aneichyk, T. (2019). Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics*, *35*(14). <https://doi.org/10.1093/bioinformatics/btz363>
- Swindell, W. R., Johnston, A., Voorhees, J. J., Elder, J. T., & Gudjonsson, J. E. (2013). Dissecting the psoriasis transcriptome: Inflammatory- and cytokine-driven gene expression in lesions from 163 patients. *BMC Genomics*, *14*(1). <https://doi.org/10.1186/1471-2164-14-527>
- Vallania, F., Tam, A., Lofgren, S., Schaffert, S., Azad, T. D., Bongen, E., Haynes, W., Alsup, M., Alonso, M., Davis, M., Engleman, E., & Khatri, P. (2018). Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. *Nature Communications*, *9*(1). <https://doi.org/10.1038/s41467-018-07242-6>
- Wang, J., Devlin, B., & Roeder, K. (2020). Using multiple measurements of tissue to estimate subject- And cell-type-specific gene expression. *Bioinformatics*, *36*(3). <https://doi.org/10.1093/bioinformatics/btz619>
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*. <https://doi.org/10.1038/ncomms14049>

Annex I: Deconvolució de dades humanes

| | | | |
|----------|---------------|----------|------------|
| % >= 65% | 35% < % > 60% | % =< 35% | No avaluat |
|----------|---------------|----------|------------|

Taula 7. Resultats de la deconvolució de dades humanes amb CIBERSORT

| CIBERSORT humà | LM22 | | TIL10 | |
|------------------------|---------|--|---------|-------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 43% | Plasma 17% NK's 13% | 88% | |
| Macròfags | 23% | Monòcits 23% NK's 20% Dendrítiques 14% | 61% | Monòcits 33% |
| Monòcits | 18% | Macròfags 40% Neutròfils 14% | 41% | Macròfags 39% CD4's 16% |
| NK's | 51% | Tcells 16% CD4's 12% CD8's 10% | 85% | |
| CD4's | 26% | NK's 27% Tcells 24% CD8's 17% | 66% | NK's 13% |
| CD8's | 15% | NK's 36% | 42% | NK's 32% CD4's 13% |
| Cèl·lules dendrítiques | 24% | Macròfags 31% Monòcits 16% | 8% | Macròfags 65% CD4's 10% |
| Treg | 8% | CD4's 35% Tcells 23% NK's 15% CD8's 13% | 88% | |

CIBERSORT no funciona be ni amb BCIC ni amb TIC de EPIC.

Taula 8. Resultats de la deconvolució de dades humanes amb FARDEEP

| FARDEEP humà | LM22 | | TIL10 | |
|--------------|---------|-------------------------------|---------|-------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 92% | | 90% | |
| Macròfags | 87% | | 92% | |
| Monòcits | 69% | | 20% | Macròfags 62% |
| NK's | 89% | | 82% | |
| CD4's | 86% | | 74% | NK's 10% |

| | | | | |
|------------------------|-----------|--|----------|--|
| CD8's | 63% | NK's 12% Tcells 11% CD4's 11% | 40% | NK's 28% CD4's 23% |
| Cèl·lules dendrítiques | 70% | Monòcits 17% | 0% | Macròfags 90% |
| Treg | 67% | CD4's 31% | 71% | CD4's 26% |
| | | | | |
| FARDEEP humà | EPIC-BCIC | | EPIC-TIC | |
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 63% | CD4's 30% | 57% | CD4's 39% |
| Macròfags | – | Monòcits 87% CD4's 13% | 97% | |
| Monòcits | 95% | | – | Macròfags 75% CD4's 24% |
| NK's | 67% | CD8's 26% | – | CD4's 60% CD8's 38% |
| CD4's | 79% | CD8's 19% | 97% | |
| CD8's | 53% | CD4's 25% NK's 19% | 4% | CD4's 95% |
| Cèl·lules dendrítiques | – | Monòcits 68% CD4's 29% | – | Macròfags 68% CD4's 17% Bcells 14% |
| Treg | – | CD8's 48% CD4's 33% Monòcits 15% | – | CD4's 98% |

Taula 9. Resultats de la deconvolució de dades humanes amb EPIC

| EPIC humà | LM22 | | TIL10 | |
|-------------|---------|-------------------------------|---------|-------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 64% | Plasma 22% | 70% | Altres 13% |
| Macròfags | 72% | | 69% | Altres 28% |
| Monòcits | 11% | Mast cells 56% | 1% | Macròfags 75% Altres 23% |
| NK's | 67% | Tcells 12% Mast cells 12% | 99% | |
| CD4's | 40% | CD8's 17% | 35% | NK's 24% |

| | | | | |
|------------------------|-----------|---|----------|--|
| | | Tcells 14% NK's 11% Eosinòfils 10% | | Macròfags 15% CD8's 11% |
| CD8's | 23% | NK's 47% | 31% | NK's 60% |
| Cèl·lules dendrítiques | 27% | Mast cells 27% Macròfags 19% NK's 11% Eosinòfils 10% | 0% | Macròfags 71% Altres 24% |
| Treg | 40% | CD4's 26% Tcells 23% | 58% | Neutròfils 13% NK's 11% |
| | | | | |
| EPIC humà | EPIC-BCIC | | EPIC-TIC | |
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 53% | CD4's 37% | 27% | CD4's 55% Altres 18% |
| Macròfags | – | Monòcits 57% NK's 26% Bcells 17% | 64% | Altres 33% |
| Monòcits | 40% | Neutròfils 27% CD4's 23% | – | CD4's 43% Altres 38% Macròfags 19% |
| NK's | 81% | Neutròfils 13% | – | CD4's 54% CD8's 24% Altres 22% |
| CD4's | 70% | NK's 17% | 89% | Altres 11% |
| CD8's | 0% | CD4's 45% NK's 42% Neutròfils 11% | 0% | CD4's 87% Altres 13% |
| Cèl·lules dendrítiques | – | Bcells 69% Monòcits 22% | – | CD4's 42% Altres 30% Macròfags 19% |
| Treg | – | NK's 41% CD4's 28% Neutròfils 21% | – | CD4's 68% CD8's 20% Altres 12% |

Taula 10. Resultats de la deconvolució de dades humanes amb QuanTlseq

| | | |
|----------------|---------|-------------------------------|
| QuanTlseq humà | TIL10 | |
| | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 53% | Altres 38% |
| Macròfags | 59% | Altres 41% |
| Monòcits | 0% | Macròfags 100% |

| | | |
|------------------------|-----|-----------------------------|
| NK's | 96% | |
| CD4's | 44% | Altres 38% |
| CD8's | 49% | NK's 28% Altres 20% |
| Cèl·lules dendrítiques | 0% | Macròfags 61% Altres 35% |
| Treg | 58% | Altres 28% |

Taula 11. Resultats de la deconvolució de dades humanes amb GEDIT

| GEDIT humà | LM22 | | TIL10 | |
|------------------------|---------|---|---------|-------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 73% | Plasma 14% | 77% | |
| Macròfags | 82% | | 87% | Monòcits 13% |
| Monòcits | 41% | Mast cells 39% | 9% | Macròfags 91% |
| NK's | 88% | Mast cells 11% | 98% | |
| CD4's | 65% | Eosinòfils 11% | 62% | NK's 13% CD8's 12% |
| CD8's | 40% | NK's 35% Mast cells 10% | 48% | NK's 46% |
| Cèl·lules dendrítiques | 26% | Mast cells 21% Macròfags 19% Eosinòfils 10% | 0% | Macròfags 83% Monòcits 15% |
| Treg | 43% | CD4's 33% Tcells 11% | 77% | |

| GEDIT humà | EPIC-BCIC | | EPIC-TIC | |
|-------------|-----------|-------------------------------|----------|-------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 57% | CD4's 39% | 35% | CD4's 65% |
| Macròfags | – | Monòcits 81% NK's 11% | 99% | |
| Monòcits | 81% | CD4's 14% | – | Macròfags 60% CD4's 40% |
| NK's | 90% | | – | CD4's 68% CD8's 32% |
| CD4's | 73% | NK's 19% | 100% | |
| CD8's | 0% | NK's 47% CD4's 46% | 0% | CD4's 100% |

| | | | | |
|------------------------|---|----------------------------|---|----------------------------|
| Cèl·lules dendrítiques | – | Monòcits 51% Bcells 46% | – | Macròfags 57% CD4's 36% |
| Treg | – | NK's 52% CD4's 33% | – | CD4's 78% CD8's 22% |

Annex II: Deconvolució de dades de ratolí

Taula 12. Resultats de la deconvolució de dades de ratolí amb CIBERSORT

| CIBERSORT ratolí | seqImmuneCC | |
|------------------------|-------------|-------------------------------|
| | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 78% | Dendrítiques 16% |
| Macròfags | 96% | |
| Monòcits | 10% | Macròfags 87% |
| Neutròfils | 68% | Macròfags 30% |
| NK's | 92% | |
| CD4's | 41% | CD8's 56% |
| CD8's | 98% | |
| Cèl·lules dendrítiques | 87% | Macròfags 12% |
| Treg's | – | |

CIBERSORT no funciona be amb Tabula Muris degut a la grandària de la matriu.

Taula 13. Resultats de la deconvolució de dades de ratolí amb FARDEEP

| FARDEEP ratolí | Tabula Muris | | seqImmuneCC | |
|----------------|--------------|-------------------------------|-------------|-------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 98% | | 86% | Dendrítiques 11% |
| Macròfags | – | Monòcits 65% Bcells 19% | 91% | |
| Monòcits | 74% | Tcells 19% | 9% | Macròfags 88% |
| Neutròfils | – | Granulòcits 99% | 19% | Macròfags 41% Monòcits 38% |
| NK's | 98% | | 94% | |
| CD4's | – | Tcells 69% NK's 24% | 39% | CD8's 61% |
| CD8's | – | Tcells 59% NK's 34% | 97% | |

| | | | | |
|------------------------|---|--------------------------------------|-----|-----------|
| Cèl·lules dendrítiques | – | Bcells 63% Monòcits 36% | 97% | |
| Treg's | – | Tcells 52% NK's 25% Bcells 21% | – | CD4's 96% |

Taula 14. Resultats de la deconvolució de dades de ratolí amb EPIC

| EPIC ratolí | Tabula Muris | | seqImmuneCC | |
|------------------------|--------------|---|-------------|---|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 87% | Epitelials 10% | 94% | |
| Macròfags | – | Bcells 37% Monòcits 20% Epitelials 17% Endotelials 10% | 9% | Altres 54% Dendrítiques 23% |
| Monòcits | 21% | Tcells 41% Epitelials 18% Altres 10% | 0% | Altres 86% |
| Neutròfils | – | Granulòcits 99% | 37% | Altres 60% |
| NK's | 85% | | 62% | Dendrítiques 38% |
| CD4's | – | Tcells 47% Bcells 20% Epitelials 15% NK's 11% | 18% | Dendrítiques 45% CD8's 23% NK's 14% |
| CD8's | – | Tcells 48% Epitelials 19% Bcells 14% | 22% | Dendrítiques 61% NK's 16% |
| Cèl·lules dendrítiques | – | Bcells 59% Monòcits 23% NK's 14% | 87% | |
| Treg's | – | Tcells 32% Bcells 25% Endotelials 16% Epitelials 13% NK's 13% | – | CD4's 52% CD8's 23% Altres 11% |

Taula 15. Resultats de la deconvolució de dades de ratolí amb GEDIT

| GEDIT ratolí | Tabula Muris | | seqImmuneCC | |
|--------------|--------------|--|-------------|-------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 81% | Epitelials 12% | 93% | |
| Macròfags | – | Bcells 38% Eritròcits 19% Epitelials 14% | 82% | Dendrítiques 16% |

| | | | | |
|------------------------|-----|---|-----|-------------------------------|
| | | Monòcits 11% | | |
| Monòcits | 15% | Tcells 45% Epitelials 20% Eritròcits 12% | 0% | Macròfags 97% |
| Neutròfils | – | Granulòcits 83% | 14% | Macròfags 83% |
| NK's | 73% | Eritròcits 10% | 26% | Dendrítiques 73% |
| CD4's | – | Tcells 51% Bcells 19% Epitelials 15% | 31% | CD8's 39% Dendrítiques 28% |
| CD8's | – | Tcells 47% Epitelials 18% Bcells 17% | 48% | Dendrítiques 48% |
| Cèl·lules dendrítiques | – | Bcells 71% | 97% | |
| Treg's | – | Tcells 38% Bcells 22% Endotelials 18% Epitelials 11% | – | CD4's 72% CD8's 24% |

Annex III: Deconvolució de dades de ratolí traduïdes

Taula 16. Resultats de la deconvolució de dades de ratolí traduïdes amb CIBERSORT

| CIBERSORT ratolí→humà | LM22 | | TIL10 | |
|--------------------------|---------|---|---------|-----------------------------------|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 70% | | 73% | |
| Macròfags | 25% | Neutròfils 20% Monòcits 17% Eosinòfiles 13% Dendrítiques 12% | 47% | Dendrítiques 36% |
| Monòcits | 23% | Neutròfils 28% Macròfags 16% Dendrítiques 10% | 0% | Macròfags 42% Dendrítiques 34% |
| Neutròfils | 50% | Macròfags 14% Monòcits 11% | 38% | Dendrítiques 23% |
| NK's | 51% | Tcells 16% CD4's 13% CD8's 10% | 70% | CD4's 19% |
| CD4's | 30% | Tcells 23% NK's 19% CD8's 18% | 2% | CD8's 66% Tregs 25% |
| CD8's | 20% | CD4's 27% NK's 23% Tcells 23% | 96% | |

| | | | | |
|--|-----|--|-----|-------------------------------|
| Cèl·lules dendrítiques | 36% | Macròfags 15% Eosinòfils 15% | 30% | Macròfags 27% Monòcits 14% |
| Treg's | 2% | CD4's 31% Tcells 24% NK's 18% CD8's 16% | 76% | CD8's 10% |
| CIBERSORT no funciona be ni amb BCIC ni amb TIC de EPIC. | | | | |

Taula 17. Resultats de la deconvolució de dades de ratolí traduïdes amb FARDEEP

| FARDEEP ratolí → humà | LM22 | | TIL10 | |
|--------------------------|-----------|-------------------------------|----------|---|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 97% | | 94% | |
| Macròfags | 81% | | 23% | Dendrítiques 62% |
| Monòcits | 69% | | 2% | Macròfags 91% |
| Neutròfils | 83% | | 66% | Monòcits 15% Macròfags 12% |
| NK's | 68% | Tcells 33% | 76% | CD4's 10% |
| CD4's | 41% | Tregs 41% CD8's 16% | 0% | CD8's 76% |
| CD8's | 90% | | 91% | |
| Cèl·lules dendrítiques | 69% | Eosinòfils 14% | 72% | Macròfags 18% |
| Treg's | 58% | CD4's 22% Tcells 12% | 43% | CD4's 31% CD8's 11% |
| | | | | |
| FARDEEP ratolí → humà | EPIC_BCIC | | EPIC-TIC | |
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 93% | | 92% | |
| Macròfags | – | Monòcits 79% Bcells 14% | 59% | Bcells 32% |
| Monòcits | 60% | CD4's 39% | – | Bcells 44% CD4's 22% Macròfags 16% CAF's 13% |
| Neutròfils | 90% | | – | Macròfags 83% CD4's 10% |

| | | | | |
|------------------------|-----|---|-----|-----------------------------|
| NK's | 45% | CD4's 33% Monòcits 19% | – | CD4's 65% Bcells 21% |
| CD4's | 93% | | 68% | Bcells 21% |
| CD8's | 0% | CD4's 85% | 3% | CD4's 55% Bcells 39% |
| Cèl·lules dendrítiques | – | Bcells 45% Monòcits 32% CD4's 23% | – | Bcells 62% Macròfags 26% |
| Treg's | – | CD4's 58% Monòcits 20% CD8's 11% | – | CD4's 67% Bcells 21% |

Taula 18. Resultats de la deconvolució de dades de ratolí traduïdes amb EPIC

| EPIC ratolí → humà | LM22 | | TIL10 | |
|------------------------|---------|--|---------|---|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 80% | | 78% | |
| Macròfags | 28% | Neutròfils 36% Tcells 17% | 47% | Altres 31% Neutròfils 11% |
| Monòcits | 4% | Macròfags 45% Altres 20% Neutròfils 14% | 0% | Macròfags 42% Neutròfils 25% Dendrítiques 13% NK's 10% |
| Neutròfils | 54% | Mast cells 21% Altres 15% | 86% | |
| NK's | 50% | Neutròfils 23% Tcells 22% | 64% | Altres 24% Macròfags 10% |
| CD4's | 12% | CD8's 44% Tcells 18% Tregs 15% | 1% | CD8's 45% Macròfags 23% Tregs 20% |
| CD8's | 79% | | 87% | Macròfags 12% |
| Cèl·lules dendrítiques | 48% | Altres 13% | 12% | Macròfags 46% Bcells 21% NK's 11% |
| Treg's | 20% | Tcells 29% CD4's 26% CD8's 12% Altres 11% | 55% | Macròfags 15% |

| EPIC ratolí → humà | EPIC_BCIC | | EPIC-TIC | |
|------------------------|-----------|--|----------|--|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 95% | | 91% | |
| Macròfags | – | Bcells 43% Neutròfils 24% Altres 17% Monòcits 16% | 16% | Bcells 37% Altres 37% CAF's 10% |
| Monòcits | 10% | CD4's 41% Altres 34% Bcells 14% | – | Bcells 29% Altres 26% CAF's 23% CD4's 15% |
| Neutròfils | 100% | | – | Altres 68% CAFs 16% Macròfags 11% |
| NK's | 48% | Altres 21% CD4's 21% | – | CD4's 41% Altres 32% Bcells 14% CAFs 14% |
| CD4's | 65% | Altres 21% Bcells 11% | 38% | Bcells 34% CAFs 15% Altres 13% |
| CD8's | 0% | CD4's 68% Bcells 17% Altres 15% | 0% | Bcells 42% CD4's 33% CAFs 16% |
| Cèl·lules dendrítiques | – | Bcells 71% Altres 19% Monòcits 10% | – | Bcells 74% Altres 17% |
| Treg's | – | Altres 35% CD4's 34% NK's 21% | – | CD4's 38% Altres 33% Bcells 15% CAFs 13% |

Taula 19. Resultats de la deconvolució de dades de ratolí traduïdes amb QuanTlseq

| QuanTlseq ratolí → humà | TIL10 | |
|----------------------------|---------|---|
| | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 98% | |
| Macròfags | 39% | Altres 23% Neutròfils 17% |
| Monòcits | 0% | Altres 45% Macròfags 23% Dendrítiques 13% Neutròfils 11% |
| Neutròfils | 41% | Altres 38% |

| | | |
|------------------------|------|---|
| NK's | 70% | Altres 15% Macròfags 12% |
| CD4's | 0% | CD8's 61% Tregs 27% Macròfags 11% |
| CD8's | 100% | |
| Cèl·lules dendrítiques | 4% | Altres 33% Monòcits 20% Macròfags 17% Bcells 17% |
| Treg's | 89% | |

Taula 20. Resultats de la deconvolució de dades de ratolí traduïdes amb GEDIT

| GEDIT ratolí → humà | LM22 | | TIL10 | |
|------------------------|-----------|--------------------------------------|----------|---|
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 78% | | 79% | |
| Macròfags | 43% | Mast cells 12% Neutròfils 12% | 71% | |
| Monòcits | 18% | Macròfags 56% | 0% | Macròfags 52% Dendrítiques 25% NK's 11% |
| Neutròfils | 29% | Mast cells 44% Macròfags 21% | 29% | Monòcits 42% Dendrítiques 16% |
| NK's | 47% | Tcells 53% | 83% | Macròfags 14% |
| CD4's | 6% | CD8's 48% Tcells 21% Tregs 18% | 3% | CD8's 52% Tregs 25% Macròfags 15% |
| CD8's | 87% | | 90% | |
| Cèl·lules dendrítiques | 49% | Macròfags 16% Eosinòfils 11% | 16% | Macròfags 42% Bcells 19% NK's 10% |
| Treg's | 22% | CD4's 34% Tcells 31% CD8's 10% | 62% | CD4's 14% |
| | | | | |
| GEDIT ratolí → humà | EPIC_BCIC | | EPIC_TIC | |
| | Detecta | Majoritàriament es confon amb | Detecta | Majoritàriament es confon amb |
| Cèl·lules B | 95% | | 100% | |
| Macròfags | – | Monòcits 50% | 52% | Bcells 38% |

| | | | | |
|------------------------|------|---------------------------------------|-----|-------------------------------------|
| | | Bcells 43% | | CAFs 10% |
| Monòcits | 38% | CD4's 45% Bcells 17% | – | Bcells 40% CAFs 27% CD4's 19% |
| Neutròfils | 100% | | – | Macròfags 64% CAFs 31% |
| NK's | 57% | CD4's 22% Monòcits 12% | – | CD4's 59% Bcells 22% CAFs 19% |
| CD4's | 77% | Bcells 14% | 43% | Bcells 41% CAFs 16% |
| CD8's | 0% | CD4's 78% Bcells 20% | 0% | Bcells 49% CD4's 35% CAFs 16% |
| Cèl·lules dendrítiques | – | Bcells 68% Monòcits 32% | – | Bcells 78% Macròfags 18% |
| Treg's | – | CD4's 44% NK's 30% Monòcits 17% | – | CD4's 55% Bcells 24% CAFs 18% |