



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Estudio del Metabolismo Tumoral en un Modelo de Carcinoma

Hepatocelular en Rata

TESIS

QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias

PRESENTA:

Claudia Erika Hernandez Patiño

TUTOR PRINCIPAL:

Osbaldo Resendis Antonio

[Instituto Nacional de Medicina Genómica](#)
[Red de Apoyo a la Investigación-UNAM](#)

MIEMBROS DEL COMITÉ TUTOR

Dr. Julio Pérez Carreón

[Instituto Nacional de Medicina Genómica](#)

Dra. Victoria Chagoya

[Instituto de Fisiología Celular](#)

Ciudad Universitaria, Cd. Mx

Junio, 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

En primer lugar me gustaría agradecer a los doctores Osbaldo Resendis-Antonio y Julio Isael Pérez Carreón por permitirme participar en este proyecto y por su confianza y guía, aún en los momentos más difíciles; les debo todo lo aprendido, el resto de mi camino seguramente estará marcado por esta experiencia. A Julia, sin ella esta tesis nunca hubiese sido la que es ahora. A Christian Diener por toda la paciencia para resolver mis dudas, aunque tuviera que explicar mil veces, paciencia que también tuvo al escuchar mis interminables reflexiones del trabajo. A mis compañeros de laboratorio, que no dudaron en ayudarme a mejorar todo aspecto de esta tesis. Agradezco a la doctora Victoria Chagoya, miembro de mi comité tutor, por sus valiosos comentarios.

A Juan, que ante la mínima muestra de miedo, duda o tristeza se apura a ayudarme, pues me ha prestado su hombro, su oído y su ayuda. A Charlie que moldeó mi forma de pensar y me retaba en cada instante, sin él, yo no sería la misma. A Lucía que siempre ha sido un gran oído y se prestaba como conejillo de indias para explicarle mi proyecto. A Daniel, Jimena y Marisol, que aún sin palabras me entienden.

A mi madre que ante cada problema tenía algún consejo amigable y no me dejaba sentirme sola. A mi hermano que siempre muestra interés por lo que hago y me alienta. Y finalmente a mi padre que siempre ha apoyado mi desarrollo.

RESUMEN

El cáncer en hígado es la segunda causa de muerte por cáncer a nivel mundial. De estos, más del 85% corresponden al tipo carcinoma hepatocelular (HCC por sus siglas en inglés). Debido a la incidencia y mortalidad de esta enfermedad existe la necesidad de entender los mecanismos por los cuales progresa y se mantiene. En esta búsqueda se han observado características comunes metabólicas en tumores, como son el efecto Warburg, la adicción a la glutamina y activación de respuesta antioxidante. Características que pueden darse y mantenerse en varios niveles de regulación.

Este proyecto busca explorar las diferencias metabólicas entre HCC e hígado normal en un protocolo de hepatocarcinogénesis química en rata a partir de tres niveles: diferencias en perfiles metabólicos obtenidos por electroforesis capilar acoplado a espectrometría de masas; comportamiento transcripcional medido por microarreglos y diferencias en actividad enzimática predicho por un modelo matemático.

Los metabolitos medidos en este trabajo en efecto logran diferenciar un tejido tumoral de uno sano, característica que se conservó al analizar la transcripción enzimática. El modelo matemático indicó una mayor activación de las vías metabólicas a nivel transcripcional y en actividad catalítica. Además nos permite proponer una respuesta a un ambiente oxidante donde los siguientes procesos se producen: 1) una activación a las respuestas antioxidantes por NRF2, 2) inhibición de la actividad enzimática de la aconitasa; y 3) se activa la utilización de GABA para recuperar la poza de NADPH así como para activar el ciclo de Krebs.

ÍNDICE

1.INTRODUCCIÓN	1
1.1 Carcinoma hepatocelular	1
1.2 Fenotipo metabólico del HCC	3
1.2.1 Metabolismo proliferativo	3
1.2.2 Estrés oxidativo en cáncer	5
1.2.3 Transcripción de genes metabólicos en cáncer	7
1.3 El estudio del metaboloma	9
1.3.1 Contextualización de información metabólica	10
2. OBJETIVO E HIPÓTESIS	16
2.1 Hipótesis	16
2.2 Objetivo general	16
2.2.1 Objetivos específicos	16
3.METODOLOGÍA	17
3.1 Flujo de Trabajo y Reproducibilidad	17
3.2 Obtención de muestras	19
3.3 Exploración de datos	21
3.3.1 Exploración de transcriptoma	21
3.3.2 Exploración de los datos de metaboloma	24
3.4 Diferencias de actividad enzimática: Dycone	26
3.4.1 Reconstrucción de la red metabólica	26
3.4.2 Espacio de actividad catalítica	27
3.4.3 Diferencias de actividad	28
3.5 Expresión de genes regulados por NRF2	30
3.5.1 RT-qPCR.....	30
3.5.2 Inmunohistoquímica	30
4.RESULTADOS Y DISCUSIÓN	31
4.1 Metabolismo como un elemento diferenciador	31
4.2 Diferencias Metabólicas. Vías y componentes	34

4.3 Diferencias en actividad enzimática.....	40
5.CONCLUSIONES Y PERSPECTIVAS.....	48
6. BIBLIOGRAFÍA.....	51
Anexo.....	59
Tabla S1 Red metabólica.....	60
Tabla S2 Cambio en K y Genes.....	64

1.INTRODUCCIÓN

1.1 Carcinoma hepatocelular

El cáncer de hígado es la segunda causa de muerte por cáncer a nivel mundial (M. Ervik, F. Lam, J. Ferlay, L. Mery, I. Soerjomataram, F. Bray 2016; Ferlay et al. 2014). De los casos de cáncer de hígado entre el 85% y 90% corresponden al tipo carcinoma hepatocelular (HCC por sus siglas en inglés) (El-Serag and Lenhard Rudolph 2007).

La mayor parte de los pacientes de HCC se encuentra en un contexto de enfermedad crónica en el hígado que desarrolla cirrosis (Farazi and DePinho 2006). La cirrosis se vuelve así uno de los grandes factores de riesgo para el desarrollo de HCC. Ésta patología puede desarrollarse por una variedad factores tales como: la contaminación de alimentos con toxinas, por ejemplo las aflatoxinas; hepatitis B y C; el consumo crónico de alcohol y enfermedades metabólicas como la hemocromatosis o la esteatohepatitis no alcohólica (Badvie 2000). Estos factores producen un daño crónico que activa vías de inflamación y regeneración en el hígado ([Figura 1.1](#)). Con el tiempo se produce fibrosis y regeneración aberrante acompañada de pérdida de función y arquitectura. Si no se controla en esta etapa se desarrolla cirrosis, que puede avanzar a HCC (Pellicoro et al. 2014).

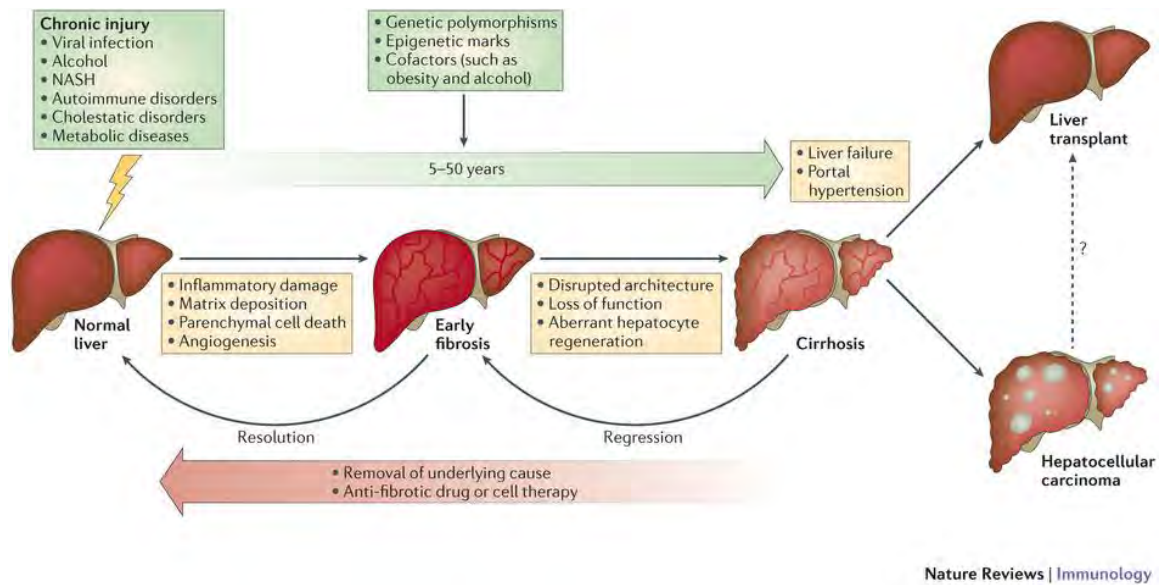


Figura 1.1 Desarrollo de HCC. (Pellicoro et al. 2014) La fibrosis hepática resulta de la respuesta a daño crónico en el hígado. Parte de esta respuesta es daño inflamatorio, depósito de matriz extracelular, muerte celular parenquimal y angiogénesis. Una vez que la fibrosis avanza a cirrosis las posibilidades de revertir el daño disminuyen y su progreso suele culminar en HCC.

Debido a la incidencia y mortalidad de esta enfermedad existe la necesidad de entender los mecanismos por los cuales progresa y se mantiene. Esta información puede guiarnos en el desarrollo de técnicas de diagnóstico, tratamiento y seguimiento. Es por ello que se han desarrollado diferentes modelos animales para simular el curso patológico del cáncer de hígado. El modelo en rata desarrollado por Schiffer (Schiffer et al. 2005) es de los pocos que logran reproducir el progreso de esta patología. El protocolo consiste en la administración del carcinógeno dietilnitrosamina (DEN) por 16 semanas, sacrificándose en la semana 18 semanas. La ventaja de este modelo radica en la obtención de muestras de HCC en un contexto similar al encontrado en pacientes, pues reproduce el proceso de inflamación, fibrosis, cirrosis y tumores multinodular.

1.2 Fenotipo metabólico del HCC

De las generalidades del cáncer al HCC.

Dentro de la variedad de elementos para entender el conjunto de enfermedades a las que se refiere el cáncer, ha llamado la atención características descritas hace una centuria por Otto Warburg. Una de sus principales contribuciones, fue la observación que células proliferativas tumorales metabolizan la glucosa a lactato aún en presencia de oxígeno; en contraste, las células normales lo oxidan por medio del ciclo de los ácidos tricarboxílicos (Warburg 1927; Vander Heiden, Cantley, and Thompson 2009). Desde entonces se han descrito más fenotipos metabólicos comunes en diferentes tipos de cáncer, principalmente en metabolismo central. Los más comunes refieren a la denominada adicción a la glutamina, efecto de Warburg (Cantor and Sabatini 2012) y el metabolismo de especies reactivas de oxígeno (ROS por sus siglas en inglés) (Gorrini, Harris, and Mak 2013). Sin embargo, y a pesar de grandes esfuerzos aún quedan varias preguntas a responder: ¿Por qué son comunes estos fenotipos metabólicos?; ¿Cuáles son los mecanismos que los promueven?; ¿Cómo podemos aprovecharlos para el desarrollo de tratamientos efectivos?; ¿Cómo y cuáles son los factores extrínsecos que los afectan?

1.2.1 Metabolismo proliferativo

Hanahan y Weinberg manejan como una característica fundamental del cáncer su habilidad para mantener una proliferación crónica (Hanahan and Weinberg 2011). Esto implica la obtención de nutrientes para suplir vías que ayuden a la síntesis de metabolitos necesarios para la duplicación celular. Por ejemplo, nucleótidos necesarios para duplicar ADN, aminoácidos para síntesis de proteínas, lípidos para la generación de otra célula, y la generación de energía necesaria para mantener estos procesos ([Figura 1.2](#)).

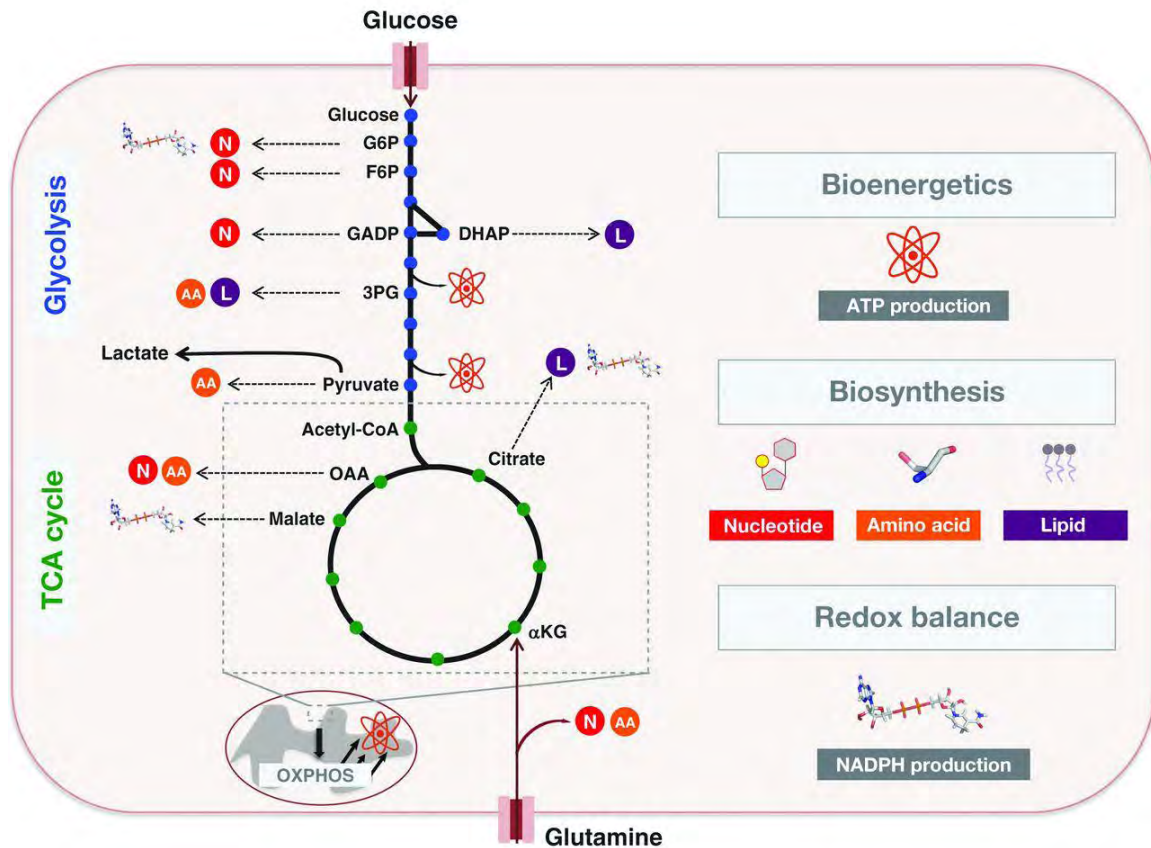


Figura 1.2 Fenotipos comunes en proliferación celular: consumo de glucosa y glutamina (Cantor and Sabatini 2012). Un fenotipo proliferativo debe satisfacer los siguientes requerimientos metabólicos: 1) la demanda energética, 2) obtención de precursores de macromoléculas, 3) mantenimiento del balance redox. El consumo de glucosa por medio de la glucólisis fomenta la obtención de metabolitos precursores para la síntesis de nucleótidos, lípidos y aminoácidos. Por otra parte la glutamina es fuente de nitrógeno para la biosíntesis de nucleótidos y aminoácidos.

En general los tejidos proliferativos, tumorales o no, coinciden en un alto consumo de glucosa, alta exportación de lactato y biosíntesis de macromoléculas como lípidos (DeBerardinis et al. 2008). Observaciones como estas han permitido crear diferentes hipótesis. Una posible explicación es que la adición a la glutamina se da para suplir de una fuente de nitrógeno, útil para la síntesis de aminoácidos y nucleótidos (Figura 1.2). Así mismo, una de las teorías más aceptadas para explicar el alto flujo glucolítico es que este genera varios precursores, como la ribosa, el glicerol y algunos aminoácidos. Particularmente la síntesis de nucleótidos y

NADPH puede favorecerse de un alto flujo en la fase de gasto de energía de la glucólisis, pues se conecta por lo menos por tres metabolitos a la vía de pentosas fosfato (glucosa 6 fosfato, fructosa 6 fosfato y gliceraldehído 3 fosfato) ([Figura 1.2](#)).

Cabe recalcar que la importancia en la síntesis de NADPH viene no solo de su papel como sustrato en vías anabólicas, si no también en el mantenimiento del balance oxido-reducción (redox).

1.2.2 Estrés oxidativo en cáncer

Otra observación común en cáncer consiste en que al comparar tejido tumoral con su contraparte normal suele observarse una mayor cantidad de especies reactivas de oxígeno (ROS) (Szatrowski and Nathan 1991; Toyokuni et al. 1995).

Aunque usualmente las ROS suelen ser asociadas con daño oxidativo y muerte celular, su concentración afecta funciones biológicas relevantes. En pequeñas y moderadas cantidades, las ROS actúan como señal promotora de procesos como diferenciación y proliferación (Boonstra and Post 2004) ([Figura 1.3](#)). Por el contrario, en grandes cantidades oxida macromoléculas como proteínas, lípidos y ADN, así como promueve procesos de muerte celular y apoptosis (Perry et al. 2000) ([Figura 1.3](#)).

En HCC podemos encontrar la importancia del estrés oxidativo en diferentes áreas de interés, desde su origen, su importancia en la reprogramación metabólica, así como posibles dianas terapéuticas. Como mencionamos anteriormente, los principales factores etiológicos del HCC coinciden en el daño crónico del hígado (Farazi and DePinho 2006), el cual usualmente induce estrés oxidativo (Urtasun, Berasain, and Avila 2015). Los marcadores de estrés oxidativo se han asociado con diferentes procesos como: el tamaño y diferenciación de tumores (Ichiba et al. 2003); incremento en neovascularización (Jo et al. 2011); activación de telomerasas (Nishikawa et al. 2009). Por estas razones, se le ha propuesto como un

predicador de recurrencia y desarrollo en HCC (Tanaka et al. 2011; Chuma et al. 2008).

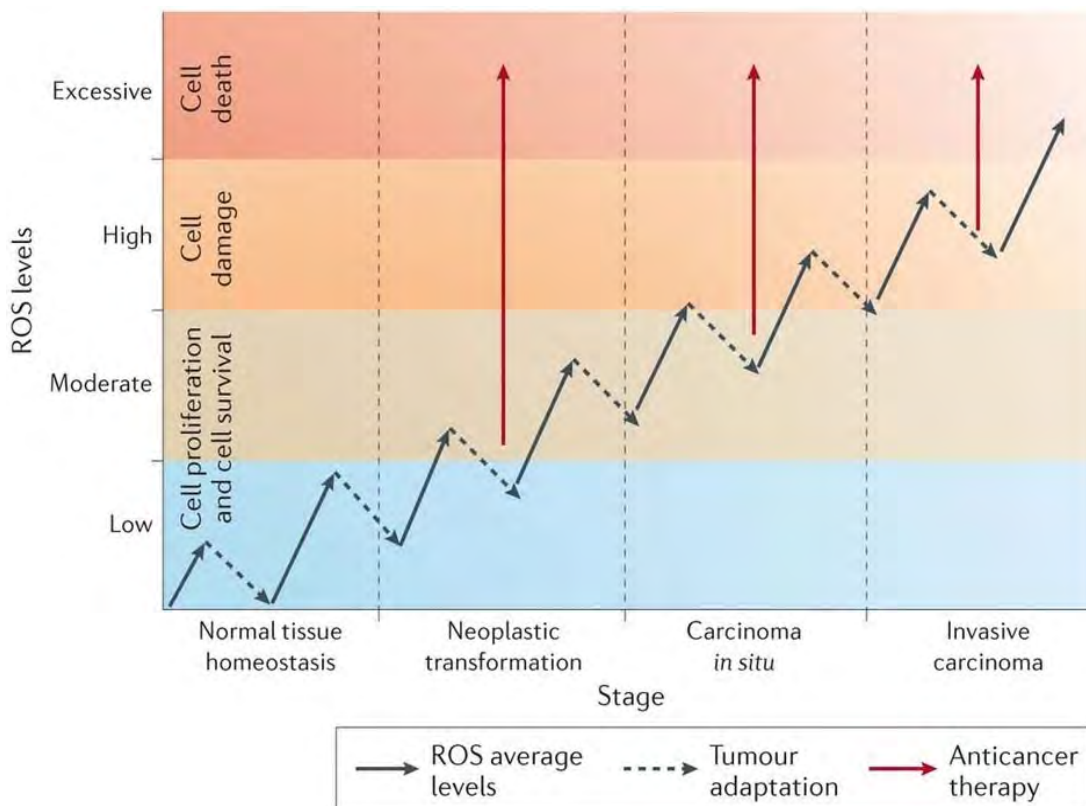


Figura 1.3 La relación de ROS y la tumorigénesis en diferentes etapas (Gorrini, Harris, and Mak 2013). En las transición de tejido normal a carcinoma invasivo las células experimentan un aumento progresivo de ROS. Las células cancerosas escapan el daño y muerte celular aumentando su defensa antioxidante, por lo que una estrategia terapéutica es aumentar las concentraciones de ROS a niveles de daño y muerte celular.

Debido a las mayores cantidades de ROS en HCC, y en cáncer en general, una de las estrategias más ocupadas en terapia es aumentar de moderadas a altas concentraciones de ROS para favorecer procesos de muerte celular y apoptosis (Gorrini, Harris, and Mak 2013) (Figura 1.3). Sin embargo, las células tumorales más agresivas suelen adaptar parte de su metabolismo para sobrevivir diferentes estreses, incluido el oxidativo, lo que las vuelve resistente a terapias como la quimioterapia y radiación (Jones and Thompson 2009; Diehn et al. 2009; Gorrini, Harris, and Mak 2013). Para ello, activan vías antioxidantes y llegan a reprimir vías

de generación de ROS. Es por estas razones que ha tomado relevancia estudiar el balance entre ROS y antioxidantes en cáncer.

Sin embargo no sólo es importante entender los requerimientos metabólicos del cáncer y cómo estos influyen el metabolismo, igualmente es necesario entender mejor los procesos de regulación que mantienen estos fenotipos.

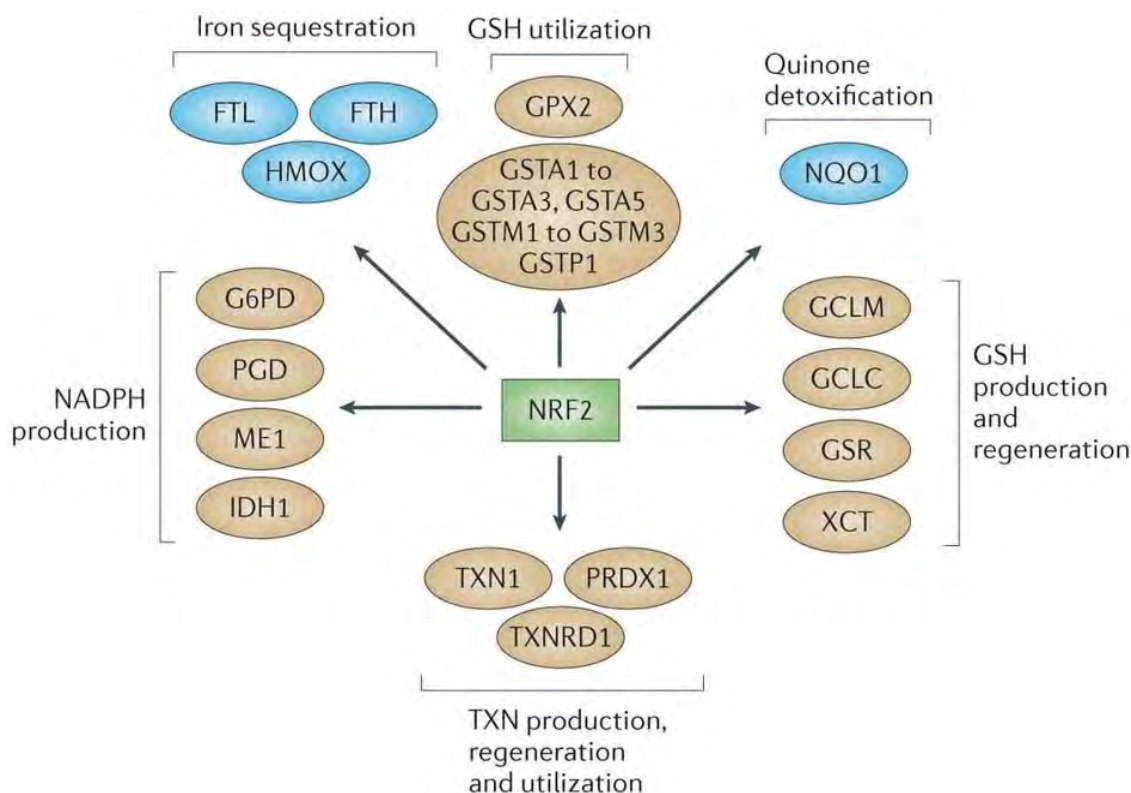
1.2.3 Transcripción de genes metabólicos en cáncer

La importancia de la reprogramación metabólica en cáncer es igualmente visible al ver las consecuencias de oncogenes y supresores de tumores (Dang et al. 2008; Wei, HaoRan, and Ping 2017). Por ejemplo, el oncogen MYC colabora con HIF1 para activar la expresión de genes glicolíticos y transportadores de glucosa (Dang et al. 2008); a su vez, es un importante regulador del metabolismo de la glutamina (Li et al. 2005).

Otro factor de transcripción frecuentemente activo en cáncer, sobre todo en tumores resistentes es el factor de transcripción conocido como NRF2 proveniente del gen NFE2L2 (Sporn and Liby 2012). Su importancia radica en su papel en la respuesta a estrés oxidativo, aunque recientemente se ha investigado su papel en la proliferación (Mitsuishi et al. 2012). En condiciones normales este factor de transcripción se mantiene en citoplasma por la proteína conocida como KEAP1. En contraste, en condiciones de estrés oxidativo se inhibe esta asociación lo que permite su translocación al núcleo (Taguchi, Motohashi, and Yamamoto 2011). Una vez en el núcleo activa genes encargados de: producción y utilización del glutatión (GSH); control de abundancia de la cisteína; síntesis de NADPH, así como la transcripción de enzimas antioxidantes, como tiorredoxina ([Figura 1.4](#)) (Gorrini, Harris, and Mak 2013). Es su papel en regulación de genes de producción de NADPH la que parece tener importancia en procesos proliferativos (Mitsuishi et al. 2012).

A pesar de que se han descrito distintas temáticas en las diferencias metabólicas de células neoplásicas, la regulación transcripcional de la red

metabólica es heterogénea (Hu et al. 2013). En general estas diferencias podrían ser asociadas por el tejido de origen del cáncer y diferencias en sus capacidades metabólicas, como su capacidad proliferativas o de resistencia a estrés oxidativo. Trabajos anteriores han descrito una menor expresión de vías metabólicas en tumores en tejidos del aparato digestivo (Hu et al. 2013), sin embargo el alcance de estas diferencias y su influencia en el metabolismo es todavía un tema de estudio.



Nature Reviews | Drug Discovery

Figura 1.4 Dianas de NRF2 (Gorrini, Harris, and Mak 2013). NRF2 activa diferentes vías asociadas con la respuesta antioxidante y producción de NADPH. Los genes relacionados con la producción, utilización y regeneración de glutatión y de tiorredoxina; producción de NADPH y antioxidantes como la NAD(P)H quinona oxidorreductasa 1.

1.3 El estudio del metaboloma

Con cada vez más información del metabolismo y su importancia en diferentes enfermedades, incluida el cáncer, surge la necesidad de estudiar el metabolismo de manera holística. La metabolómica es una estrategia cuyo propósito más ambicioso refiere a la medición total de metabolitos en una condición específica; sin embargo, considerando la diversidad de propiedades de estos, en la actualidad no es posible medirlos en su totalidad ocupando una sola tecnología (Goodacre 2004). En general las técnicas de medición del metaboloma se dividen en resonancia magnética nuclear y espectrometría de masas. La espectrometría de masas se acopla a diferentes sistemas de separación, cromatografía líquida, gases, o por sistemas capilares (Goodacre 2004; Villas-Bôas et al. 2005). Dependiendo de la especificidad de la metodología se pueden medir de 50 a 500 metabolitos; dependiendo de la sensibilidad, estos pueden tener rangos de concentración entre picomoles a mili moles (Dunn, Bailey, and Johnson 2005). Estos rangos son sólo algunos de los retos al manejar tecnología metabolómica; sin embargo, también hay retos importantes en la interpretación de cambios en la abundancia; por ejemplo, en la importancia de cambios de concentración y los mecanismos por los cuales surgen. Aunque la estadística para manejar grandes cantidades de datos biológicos ha avanzado desde el desarrollo de la transcriptómica, las características del metaboloma son muy diferentes a datos transcriptómicos o proteómicos. La problemática general es que la concentración de un metabolito es producto de todas aquellas enzimas que lo producen o lo ocupan, y esta relación con las enzimas va a afectar la concentración de otros metabolitos. En otras palabras, una disminución en un metabolito podría deberse a una menor síntesis, o una mayor degradación. Distinguir entre estas dos diferentes hipótesis no es sencillo por lo que se suele ocupar más información que contextualice las concentraciones.

1.3.1 Contextualización de información metabólica

Hay en general dos tipos de datos que se suelen ocupar para contextualizar los cambios de concentración en el metaboloma: la red metabólica y datos sobre la expresión de enzimas. La estrategia más sencilla y más ocupada es mapear en una red los metabolitos y enzimas diferenciales para poder dilucidar reacciones activadas o reprimidas (Jha et al. 2015; Mitsuishi et al. 2012; Eckel-Mahan et al. 2012). El problema de esta estrategia es que no siempre se obtiene una gran cantidad de enzimas diferenciadas, además ignora otros tipos de regulación, como cambios de actividad catalítica dado por efectores alostéricos o señalización. Sin embargo cambios generales en vías pueden dar pistas de cuales son los procesos regulados.

Otra estrategia es ocupar redes de correlación, que toman en cuenta no solo enzimas, también proteínas de señalización (Akiyama et al. 2008). Sin embargo estas redes suelen perder la relación causal existente entre diferentes metabolitos, reduciéndolo a correlaciones que podrían ser relaciones indirectas.

Sin embargo recientemente se ha ocupado un tercer dato para contextualizar concentraciones metabólicas, el fenotipo metabólico (Resendis-Antonio et al. 2015; Diener et al. 2016). Este tipo de trabajos se ha hecho a partir de modelos matemáticos, principalmente modificaciones a modelos de balance de flujos. La ventaja de estos modelos es que permite explorar funciones o fenotipos metabólicos complejos (Nielsen and Oliver 2005), como la proliferación; sin embargo, las modificaciones para explorar datos metabolómicos no son directas, pues estos modelos no manejan de forma explícita la concentración de metabolitos.

1.3.1.1 Modelos de balance flujos: exploración de funciones metabólicas e integración de expresión génica.

Dada la necesidad de integrar distintos niveles de información biológica, se han desarrollado los modelos de balance de flujo (FBA por sus siglas en inglés). Una de las grandes ventajas de este tipo de modelos es que logra integrar

información genómica y bioquímica. La idea básica es que si conocemos las reacciones bioquímicas que pueden darse en una célula y sabemos cuales son los nutrientes que mantienen el sistema de reacciones, podemos explorar y predecir todos los posibles flujos metabólicos posibles en un cultivo celular o tejido. En otras palabras, a partir de una red metabólica y flujos de entrada podemos obtener el espacio de flujos de un sistema biológico. Una vez obtenido este espacio de flujos se pueden explorar diferentes tipo de preguntas, por ejemplo: ¿Qué sucede con la actividad metabólica al truncar un flujo?, como la obtención de algún nutriente o el mismo oxígeno; ¿Qué sucede cuando mutamos una enzima?

Este tipo de técnicas también ha sido ocupado para encontrar los flujos óptimos para la generación de biomasa, o de productos industriales (Orth, Thiele, and Palsson 2010).

Los modelos de flujo tienen también la ventaja de requerir muy poca información cinética, pues depende básicamente de la arquitectura de la red metabólica. Estos modelos se llevan a cabo a partir de la descripción del conjunto de reacciones de la célula en una matriz (S o matriz estequiométrica) de dimensiones $M \times N$, donde N reacciones bioquímicas se llevan a cabo con M metabolitos ([Figura 1.5 B](#)). Cada columna de esa matriz cuenta con los coeficientes de las reacciones bioquímicas de la siguiente manera: cero si el metabolito no participa, con el coeficiente con signo positivo si el metabolito es producto y con el coeficiente con el signo negativo si es sustrato. Estos coeficientes representan la primera restricción, un metabolito sólo puede surgir de flujos de entrada o la interconversión con otros metabolitos. La otra gran suposición del modelo consiste en hacer la exploración en un estado estacionario. Es decir que la cantidad total producida de un metabolito será igual a la total consumida ([Figura 1.5C](#)). Los flujos de todas las reacciones bioquímicas se representan en un vector de tamaño N que denominaremos V . Estos flujos pueden ser restringidos a máximos (b) y mínimos (a) con significado biológico. Quedando entonces las representaciones matemática como:

$$(1) \quad SV = 0,$$
$$a_i < v_i < b_i.$$

En los sistemas biológicos hay más reacciones que metabolitos, por lo que el resultado será un espacio de posibilidades. De esta forma, explorar fenotipos complejos como la formación de biomasas es únicamente cuestión de definir una reacción (Z) con los metabolitos necesarios ([Figura 1.5 B,D](#)), si queremos ver los flujos óptimos para su creación o disminución es cuestión de encontrar los puntos máximos y mínimos de ese flujo ([Figura 1.5 E](#)).

Este tipo de modelos permite la exploración de actividades metabólicas a nivel sistémico, siendo una parte fundamental en la reconstrucción funcional de vías metabólicas. En la actualidad contamos con reconstrucciones metabólicas a escala genómica (GEMs por sus siglas en inglés) de varios organismos, incluyendo rata (Blais et al. 2017), ratón (Mardinoglu et al. 2015; Sigurdsson et al. 2010) y humano (Duarte et al. 2007; Thiele et al. 2013).

Una vez creada la redes de organismos, se pueden integran con información transcriptómica y/o proteómica (Wang, Eddy, and Price 2012; Agren et al. 2012). Este concepto ha ayudado a la reconstrucción semi-automatizada de redes tejido específico, entre las cuales destacan aquellas redes funcionales del hígado (Gille et al. 2010; Jerby, Shlomi, and Ruppin 2010; Mardinoglu et al. 2014; Agren et al. 2012). La construcción de estas redes son especialmente complejas debido a que deben de ser capaces de reproducir funciones del hígado, las cuales son variadas. Únicamente para probar la funcionalidad de las diferentes redes de hígado se suelen explorar alrededor de 250 funciones diferentes. Entre ellas se encuentran la capacidad de síntesis de ácidos grasos, aminoácidos, colesterol, ácido biliares; así como la activación de vías como la gluconeogénesis, detoxificación de amonio, entre otros.

Otra gran aplicación proviene de la formulación de redes patológicas, que al comparar con su contraparte sana pueden proveer información valiosa de las

diferentes capacidades metabólicas. Esta clase de modelos ha permitido una mejor exploración de las consecuencias metabólicas ante infección de hepatitis C, logrando proponer terapias que ayuden a prevenir HCC (Elseman et al. 2016).

La integración de datos de metaboloma en esta clase de modelos queda relegada a comprobar las funcionalidades de la red, pues se puede comprobar si estas son capaces de producir y consumir los metabolitos medidos (Agren et al. 2012). Esto es debido a que esta clase de modelo no maneja propiedades cinéticas de forma explícita, por lo que no es posible explorar cómo la concentración de un metabolito influencia el flujo, información vital para una mejor interpretación de diferencias en concentraciones.

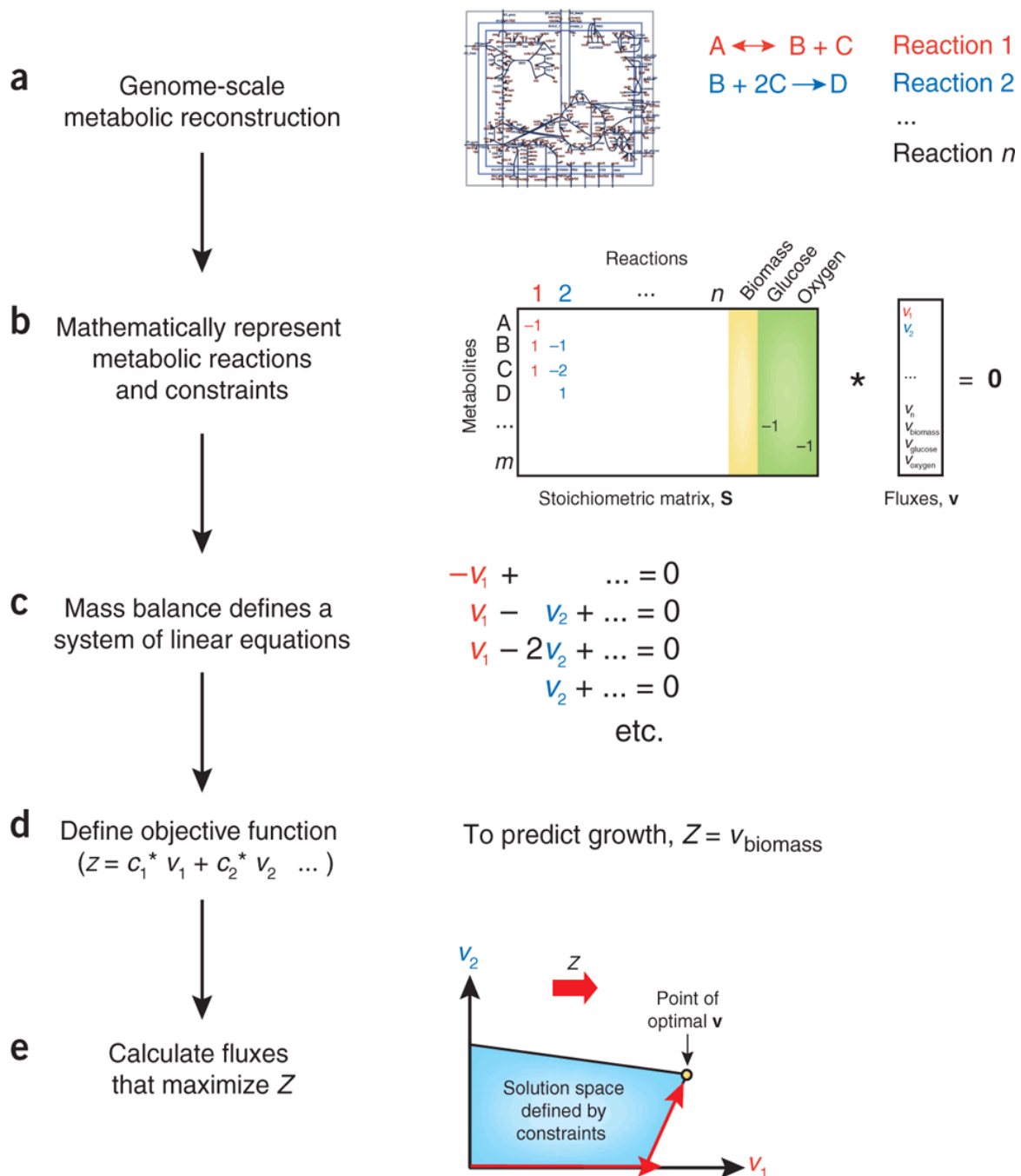


Figura 1.5 Análisis de balance de flujos (Orth, Thiele, and Palsson 2010). Este tipo de análisis depende de una red metabólica que se describe en matriz estequiométrica S . Esta matriz se ocupa para resolver los flujos que permitan un estado estacionario. Se pueden analizar fenotipos metabólicos definiéndolos como una reacción que se agrega a la matriz estequiométrica y así explorar los flujos que permiten el máximo o el mínimo de esta función.

1.3.1.2 Modificaciones a FBA, integración de datos metabólicos

Recientemente y para sobrepasar las limitaciones que tienen los modelos de balance de flujo para integrar diferencias de concentraciones, se han realizado formulaciones alternativas basadas más en constantes cinéticas que en flujos. Es decir, se transforma el espacio flujos para poder explorar el de constantes cinéticas que denominaremos el “cono-k” o (K) (Famili, Mahadevan, and Palsson 2005; Resendis-Antonio 2009). Esta transformación es basada en la concentración de metabolitos, por lo que por cada muestra de metaboloma podríamos obtener un cono-k diferente. La comparación entre estos espacios nos provee predicciones de aquellas reacciones más afectadas por las diferencias en concentración (Diener et al. 2016). Este formalismo conserva la capacidad de explorar funciones metabólicas, y tiene la ventaja de analizar diferencias en concentración de metabolitos en el contexto de las reacciones en las que participan.

Cambios en constantes cinéticas pueden darse por una gran variedad de razones, desde cantidad de enzima, competencia entre sustratos, la presencia o ausencia de reguladores de las enzimas responsables, entre otros.

En este sentido la complementación de este análisis, con un análisis de la regulación transcripcional puede proveer pistas del contexto regulatorio de la célula.

2. OBJETIVO E HIPÓTESIS

2.1 Hipótesis

Cambios en el estado metabólico entre HCC e hígado sano implicarán una inhibición de procesos comúnmente activos en hígado, y una activación de fenotipos característicos del metabolismo tumoral. Estas diferencias están reflejadas en mecanismos de regulación transcripcional y un perfil de concentración de metabolitos.

2.2 Objetivo general

Caracterizar el estado metabólico y su regulación en células cancerosas de un protocolo de hepatocarcinogenesis química en rata respecto a su contraparte normal.

2.2.1 Objetivos específicos

La caracterización general del estado metabólico en HCC se hizo a partir de tres niveles.

- Encontrar diferencias metabólicas entre tejido tumoral e hígado sano.
- Explorar el comportamiento transcripcional de vías metabólicas
- Identificar enzimas con actividad diferencial.

3.METODOLOGÍA

3.1 Flujo de Trabajo y Reproducibilidad

Los datos, el código y la implementación de herramientas ocupadas en este proyecto se encuentra disponible en un repositorio github: https://github.com/cerikahp/Hepatocelular_carcinoma y en el anexo de esta tesis. El análisis computacional se realizó principalmente en R versión 3.3.1.

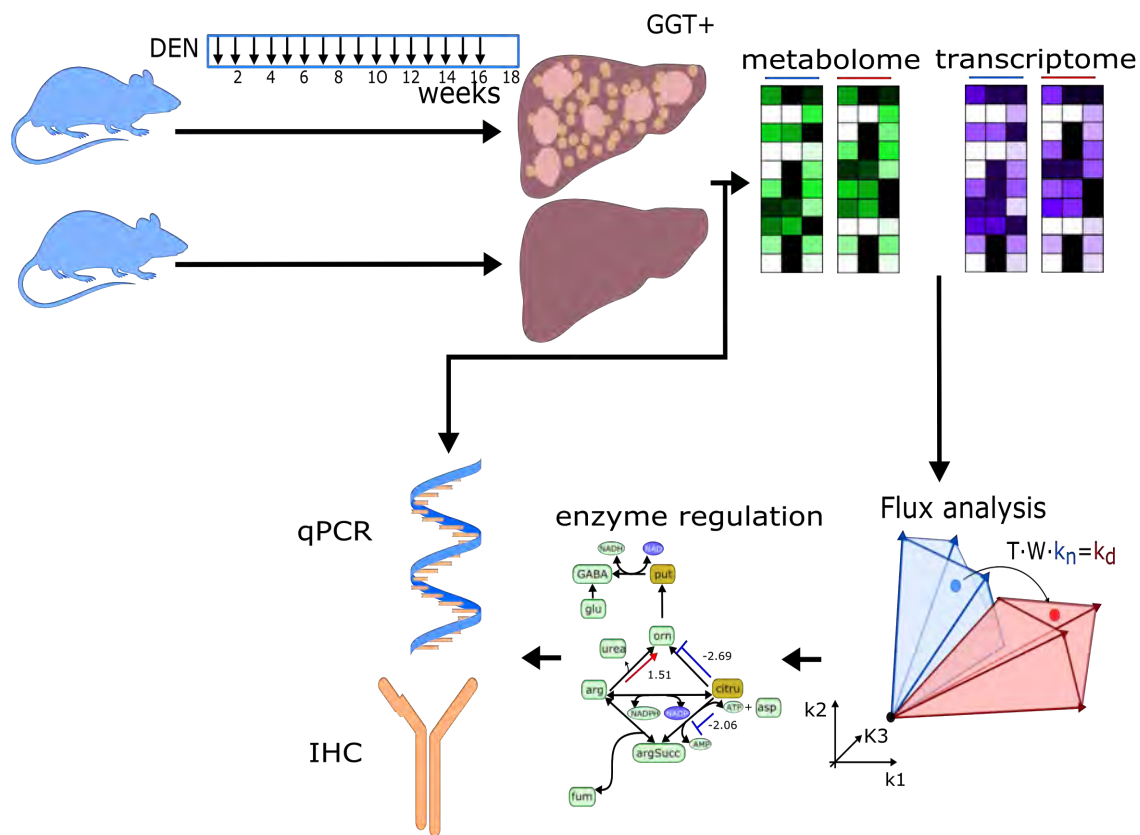


Figura 3.1 Esquema general de la metodología. Se ocupó un modelo de hepatocarcinogenesis química en rata para obtener muestras de metaboloma y transcriptoma tanto de hígado normal como de tumores. Estos datos se ocuparon para crear un modelo de flujos metabólicos. La validación del modelo se hizo por inmunohistoquímica (IHC) y qPCR.

Para estudiar el metabolismo tumoral se tomaron muestras de transcriptoma y metaboloma de hígado sano ([figura 3.1](#)). Los datos de metaboloma y transcriptoma se ocuparon para desarrollar un modelo matemático, cuyos resultados fueron comparados con datos de transcriptoma. Finalmente parte del proceso de validación se hizo por medio de qPCR e Inmunohistoquímica.

3.2 Obtención de muestras

Las muestras se obtuvieron en colaboración con el Doctor Julio Isael Pérez Carreón y la Doctora Julia Torres Mena. Estas muestras provienen de un modelo de hepatocarcinogénesis química en rata denominado modelo de Schiffer (Schiffer et al. 2005). De acuerdo a este modelo a 7 ratas se les administró por 16 semanas el carcinógeno DEN, a una dosis de 50 mg por kg de peso semanal. En la semana 18 las ratas fueron sacrificadas ([Figura 3.1 A](#)). Las muestras de tejido tumoral se identificaron a nivel histológico por medio de la actividad enzimática de gamma-glutamyl transferasa (GGT) ([Figura 3.1 B](#)). La importancia de este marcador radica en que alrededor de 80 %-85% los focos tumorales son positivos en este marcador tanto en modelos animales como en pacientes (Hanigan and Pitot 1985). En contraste, en hígado sano este marcador no se encuentra ([Anexo Figura S1](#)).

Tanto del tejido hepático congelado que contiene tumores inducidos experimentalmente, como del tejido de hígado de ratas normal se extrajo RNA mensajero. Otra porción de estos mismos tejidos fue seleccionada y procesada para el análisis del metaboloma por medio de espectrometría de masas de fracciones hidrosolubles. El RNA se procesó por tecnología de microarreglos *affymetrix* de genoma completo, RaGene2.0ST, mientras que el perfil metabolómico se obtuvo a través de una unidad de servicio en Japón (*Human Metabolome Technology*) que ocupó un sistema de electroforesis capilar acoplado a espectrometría de masas (CE-TOF/MS) que cuantificó alrededor de 98 metabolitos tanto en tejido tumoral como tejido normal.

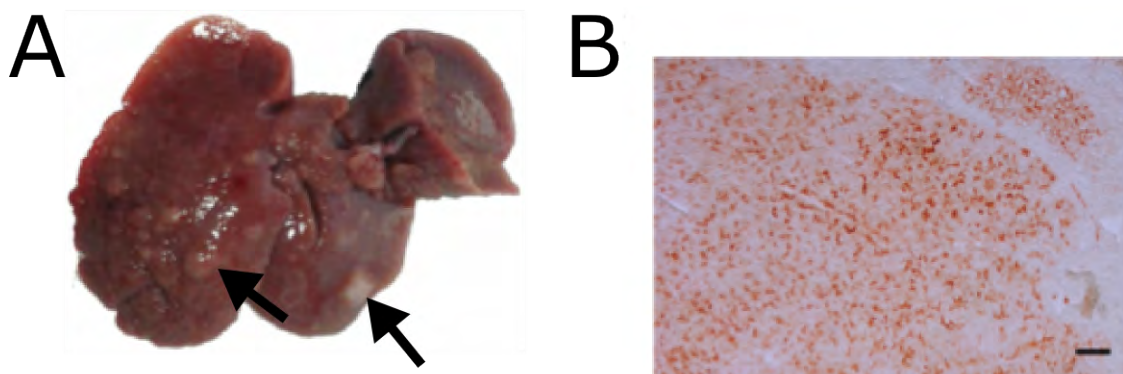


Figura 3.2 Obtención de muestras. Cortesía del Dr. Julio Pérez Carreón. A) Hígado de rata después de las 18 semanas, marcado se encuentran posibles muestras tumorales. B) Sección histológica del hígado marcado con GGT a una magnitud 10x, la línea de escala equivale a 100 μm

En total procesamos RNA mensajero de 14 muestras de 7 ratas sanas, y 13 muestras de 7 ratas a las que se les indujo HCC. Para el análisis de metaboloma se analizaron 4 muestras por cada condición ([Tabla 3.1](#)).

Tabla 3.1 Descripción de datos. Entre paréntesis está el número de ratas que se ocuparon.		
Tipo de datos	Hígado normal	Cáncer (GGT+)
CE-TOF/MS	4(4)	4(4)
Affymetrix rat gene 2.0 st array	14(7)	13(7)

3.3 Exploración de datos

3.3.1 Exploración de transcriptoma

La anotación de sondas que se ocupó es aquella basada en la base de datos *entrez*, implementada por *brain array* (ENTREZG versión 20) (Dai et al. 2005). La ventaja de esta anotación radica en que se define un conjunto de sondas por cada gen definido en la base de datos ENTREZG. Posteriormente fueron procesados por *Robust Multiple-array Average* (RMA)(Irizarry et al. 2003). Este algoritmo de preprocesamiento quita ruido de fondo, normaliza las distribuciones de intensidades para hacerlas comparables y obtiene la intensidad de genes según una anotación de sondas. Su fortaleza radica en la normalización a nivel de sondas, en vez de genes, minimizando la pérdida de información y regresando intensidades en una distribución ideal para subsecuentes pasos estadísticos.

Una vez preprocesados se ocuparon dos análisis exploratorios diferentes para verificar que los datos se separaran por condición. El primero fue un análisis de componentes principales que consiste en transformar un conjunto de variables posiblemente correlacionadas a un conjunto de variables linealmente no correlacionadas llamados componentes principales (PC por sus siglas en inglés). Esta transformación se define de tal forma que el primer PC tiene la mayor varianza posible. Este análisis vio que en efecto el transcriptoma diferenciaba ambas condiciones (Codigo Anexo).

Para enfocar el análisis a metabolismo se ocupó la definición de vías de la base datos *Small Metabolite Pathway Database* (SMPDB). Para poder convertir los identificadores de enzimas de SMPDB y obtener la relación vía - gen se ocuparon los grupos de ortología del KEGG (Minoru Kanehisa et al. 2015; M. Kanehisa 2000). Una vez hecho el filtro de genes metabólicos se ocuparon dos métodos: nuevamente un PCA y un algoritmo de aglomeramiento no supervisado Ward con distancia euclidiana. Este último análisis busca encontrar una jerarquía de grupos donde se minimice la varianza en cada grupo. El resultado final es un

dendograma con la relación de cada grupo. Este mismo algoritmo se ocupó más adelante para ver el comportamiento de enzimas que ocupan como cofactor NADP/H.

Para correr el análisis de enriquecimiento se ocupó el algoritmo *Gene Set Enrichment Analysis* (GSEA) (Subramanian et al. 2005; Mootha et al. 2003) ocupando la base datos (SMPDB) (Frolkis et al. 2010; Jewison et al. 2014). Este algoritmo tiene la característica de tomar todos los datos de expresión para encontrar vías diferencialmente expresadas. Para mantener el poder estadístico y tomar en cuenta el mayor número de vías, se tomo como 15 el número mínimo de genes para considerar una vía metabólica (Subramanian et al. 2005; Mootha et al. 2003). Tomando en cuenta que tenemos más de 7 muestras por condición se ocupo el método de permutación de muestras para calcular el valor de la tasa de falsos positivos (FDR). El valor de FDR en este algoritmo se calcula tomando en cuenta dos distribuciones diferentes 1) la calificación de enriquecimiento en comparación con todas la calificaciones de vías en todas las permutaciones; 2) la calificación del enriquecimiento de la vía en comparación con el resto de las vías. Esto hace el calculo de FDR mucho más exigente que la permutación de genes, por lo que se consideró 0.25 de valor de corte, como recomendado por los autores (Subramanian et al. 2005; Mootha et al. 2003).

Dado que los datos presentan un comportamiento cercano a una distribución normal ([Figura 3.3](#)), los genes diferencialmente expresados se obtuvieron usando el paquete de R, *limma*; en particular su método de Bayes empírico que tiene la ventaja de considerar que una parte del ruido se comparte en todas las sondas (Ritchie et al. 2015).

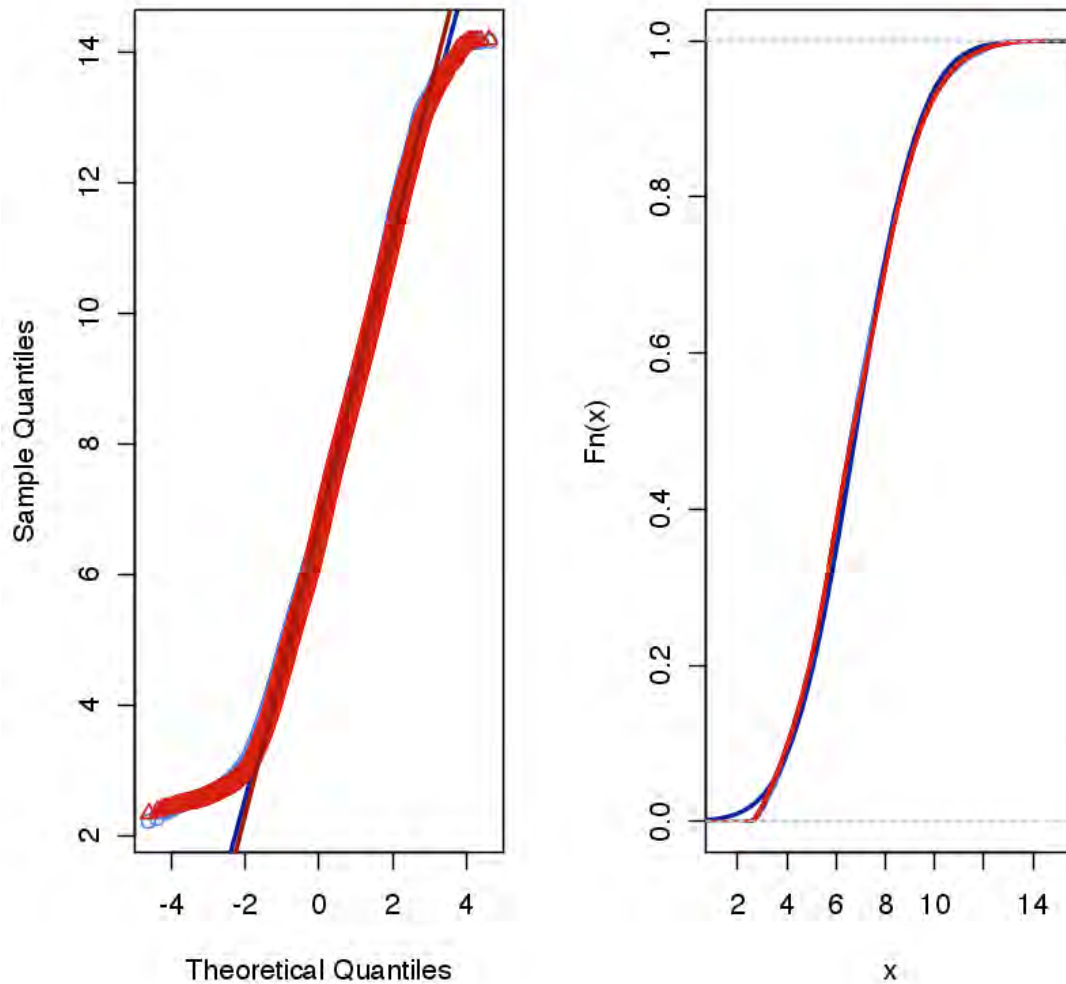


Figura 3.3 Gráfica cuantil-cuantil de transcriptoma. Los puntos en la gráfica en la izquierda son los cuantiles de las muestras graficado contra los cuantiles esperados en una distribución normal. Rojo si son HCC y azul si son normales. Las líneas representan datos teóricos con distribución normal que pasan por los cuantiles de las muestras, estas líneas están en una tonalidad más oscuras que su equivalente real. La gráfica de la derecha representa la función empírica de la distribución con el mismo código de colores que en la gráfica de la derecha.

3.3.2 Exploración de los datos de metaboloma

Se transformó los datos de metaboloma a un espacio logarítmico. Este espacio tiene una distribución similar a la normal ([Figura 3.4](#)) por lo que se podían ocupar y comparar la mayor parte de los algoritmos ocupados en el análisis de transcriptoma.

Para hacer la exploración inicial también se ocuparon los algoritmos de PCA, y un algoritmo de agrupamiento jerárquico.

El análisis de vías metabólicas se hizo ocupando *Metabolite Set Enrichment Analysis* (MSEA) en la página de *metaboAnalyst* (Jianguo Xia et al. 2015; J. Xia et al. 2012, 2009; Jianguo Xia and Wishart 2010), en específico su versión de enriquecimiento cuantitativo. La idea detrás de este algoritmo es ocupar un estadístico Q, que es proporcional a un agregado del cuadrado de la covarianza entre la concentración y el fenotipo. Es decir, este algoritmo es especialmente útil para ver si hay en su conjunto diferencias significativas, o suficientes elementos muy significativos sin tomar en cuenta la dirección del cambio.

El análisis diferencial también se analizó por estadística Bayesiana empírica del paquete *limma*.

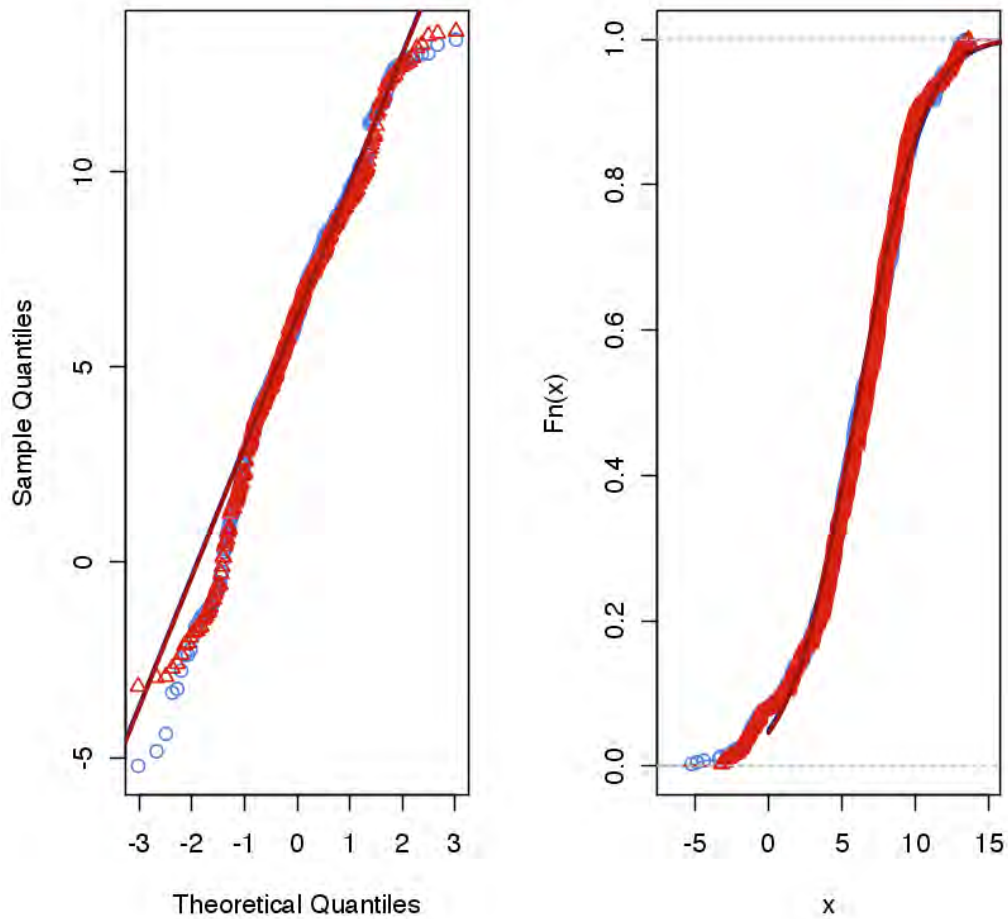


Figura 3.4 Gráfica cuantil-cuantil de metaboloma. Los puntos en la gráfica en la izquierda son los cuantiles de las muestras graficado contra los cuantiles esperados en una distribución normal. Rojo si son HCC y azul si son normales. Las líneas representan datos teóricos con distribución normal que pasan por los cuantiles de las muestras, esta líneas están en una tonalidad más oscura que su equivalente real. La gráfica de la derecha representa la función empírica de la distribución con el mismo código de colores que en la gráfica de la derecha.

3.4 Diferencias de actividad enzimática:

Dycone

Dycone es un algoritmo que en base a una red metabólica y datos de metaboloma en dos condiciones, predice cambios en la actividad catalítica de las reacciones (Diener et al. 2016). El algoritmo ocupa los datos de metaboloma para limitar el espacio de actividad catalítica posible, de tal forma que podamos identificar aquellas reacciones con mayor diferencia de actividad entre dos condiciones. Esto lo logra asumiendo una cinética de ley de acción de masas y estado estacionario. Por lo que se puede decir que existen tres grandes pasos para ocupar dycone: 1) la reconstrucción de la red; 2) obtención de espacio de actividad catalítica; 3) la obtención de reacciones diferenciales por medio de una transformación.

3.4.1 Reconstrucción de la red metabólica

Para elegir la red metabólica que analizar se tiene que considerar que dycone es muy sensible a la ausencia de metabolitos. Por lo que la reconstrucción de la red debía de tener los siguientes criterios: 1) una red que nos ayude a analizar los efectos identificados en el análisis de vías; 2) una red funcional con el menor número posible de metabolitos faltantes.

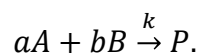
Para lograr esto ocupamos como base la red publicada por (Diener et al. 2016) y añadimos reacciones de acuerdo a los siguientes pasos: 1) de las vías enriquecidas en el metaboloma tomamos reacciones de las que tuvieramos los sustratos medidos; 2) incluimos las reacciones necesarias para que todas las reacciones fueran capaces de crear flujos; 3) checamos el balance en la red metabólica de rata (Blais et al. 2017); 4) de ser posibles colapsamos reacciones que tuvieran intermediarios no medidos.

De esta manera adicionamos diferentes vías de metabolismo nitrogenado, como la ocupación de ácido gamma aminobutírico (GABA), metabolismo de alanina, ciclo de la urea, y metabolismo de aspartato, así como la adquisición de metabolitos adecuados para que las reacciones pudiesen presentar flujo. Dada la importancia que tiene el pentosa fosfato tanto en procesos de proliferación, estrés oxidativo y producción de NADPH y considerando que habíamos medido la mayor parte de sus metabolitos decidimos mantener también esta vía en la reconstrucción. Se puede ver una representación de la red ocupada en la [Figura 4.4](#).

Las reacciones fueron anotadas de acuerdo a su *enzyme commission number* (E.C) para hacer la relación reacciones – genes de acuerdo a la base de datos de KEGG.

3.4.2 Espacio de actividad catalítica.

La idea detrás de este formalismo consiste en tomar en consideración que el flujo sigue la ley de acción de masas, es decir es proporcional al término de acción de masas. Por lo que dado una reacción:



Su flujo (v) en un determinado tiempo, se puede definir en una cinética siguiendo ley de acción de masas como:

$$v = k \cdot A^a \cdot B^b.$$

Donde las concentraciones de los reactantes (A y B) provienen de los datos de metaboloma asumiendo estado estacionario (homeostasis). En un sistema de reacciones el flujo de una reacción va a limitar el flujo de las demás reacciones. Por lo que es posible delimitar un espacio de flujos ([1.3.1.1 Modelos de balance de flujos](#)). Siguiendo ley de acción de masas este espacio de flujos se puede generalizar para obtener un espacio de actividad catalítica. Para ello se generaliza la fórmula general

a un sistema de matrices donde dado la concentración de un metabolito x , los flujos v son descritos por la ecuación:

$$(2) \quad v_j(x) = k_j \cdot m_j(x).$$

Donde m_j denota los términos metabólicos tal que $m_j = \prod_i x_i^{s_{ij}}$, siendo s_{ij} igual al coeficiente del metabolito x_i en la reacción j . Para encontrar el estado estacionario podemos reescribir el sistema definiendo una matriz diagonal $M := \text{diag}(m_j)$ y resolvemos para:

$$(3) \quad SMk = 0.$$

Ocupando la definición de flujos de la ecuación 2 tenemos que

$$(4) \quad K = M^{-1}V.$$

Siguiendo estos pasos para cada muestra podemos obtener un espacio de actividad catalítica.

3.4.3 Diferencias de actividad

El último paso consiste en identificar las reacciones con diferencias en el espacio de la siguiente manera. Si tenemos los espacio K para dos condiciones diferentes, normal (n) y enfermedad (d) en una misma red metabólica, tenemos entonces que los espacios son definidos de la siguiente manera.

$$(5) \quad K_d = M_d^{-1}V.$$

$$(6) \quad K_n = M_n^{-1}V.$$

Si despejamos V en (6) y sustituimos en (5) tenemos entonces:

$$(7) \quad K_d = M_n M_d^{-1} K_n.$$

Por lo que encontramos una transformación diagonal T , tal que $T = M_n M_d^{-1}$ por lo que obtenemos la ecuación

$$(8) \quad K_d = T K_n.$$

Donde T es una matriz diagonal que describe los cambios entre el espacio K_d y K_n , únicamente dependiente de los cambios en concentraciones metabólicas. Es importante denotar que inclusive si el espacio de flujos no es el mismo, puede hacerse una transformación dependiente de T que describa los cambios metabólicos. Entre más similares son los flujos más significativa es la matriz T .

Esta matriz da la información necesaria para una interpretación en donde se infieren las reacciones que tienen suficiente cambio en sus constante cinética considerando concentraciones metabólicas. Este cambio se calcula haciendo todas las combinaciones posibles de los conos de la condición control versus la condición enfermedad. En caso de que alguna muestra le faltara un metabolito en específico se ocupó la media de ese metabolito en la condición correspondiente. Una vez que se encuentran todos los cambios de K se calcula su significancia estadística ocupando una prueba t de Welch, la estimación de varianza se hizo por el método de Bayes empírico como es implementado en el paquete `limma`.

3.5 Expresión de genes regulados por NRF2

3.5.1 RT-qPCR

Se hizo un análisis cuantitativo de expresión de genes cuantitativos por medio de RT-qPCR usando las sondas TaqMan y sondas etiquetadas por FAM (límites exon-exon) para rata. El cDNA fue sintetizado usando "High-Capacity RNA to cDNA Master Mix" (Applied Biosystems) (Biosystems 2010) y la cuantificación de PCR se realizó ocupando el ensayo de expresión génica con Taqman de acuerdo a instrucciones del fabricante (Biosystems 2010) . La expresión génica relativa fue calculado con el método Ct delta-delta ocupando 18s rRNA como el control endógeno y el hígado normal como la referencia control.

3.5.2 Inmunohistoquímica

Se hicieron cortes histológicos seriados de tejido embebido en parafina con 5 μm de espesor e rehidratado. La recuperación del antígeno fue realizado a PH 9 e incubado una noche con el anticuerpo policlonal a 1:100 de dilución anti-ME1(abcam) de conejo.

4.RESULTADOS Y DISCUSIÓN

4.1 Metabolismo como un elemento diferenciador

Como producto de la colaboración con el Dr. Julio Pérez Carreón, se obtuvieron perfiles de transcriptoma y metaboloma de hígado de rata sano y de tumores inducidos por el modelo de Schiffer (Schiffer et al. 2005) ([métodos: 3.2](#)) ([Tabla 3.1](#)).

La primera de las preguntas que buscamos responder fue conocer si en efecto los 98 metabolitos medidos (Código anexo) podrían diferenciar un fenotipo tumoral en comparación con el sano ([métodos 3.3.2](#)). Para responder esta pregunta de forma global se ocuparon dos estrategias. Un algoritmo de aglomeración y un análisis de componentes principales (PCA por sus siglas en inglés). De forma consistente, los datos de metaboloma podían separar ambas condiciones por ambos algoritmos ([Figura 4.1 A,B](#)).

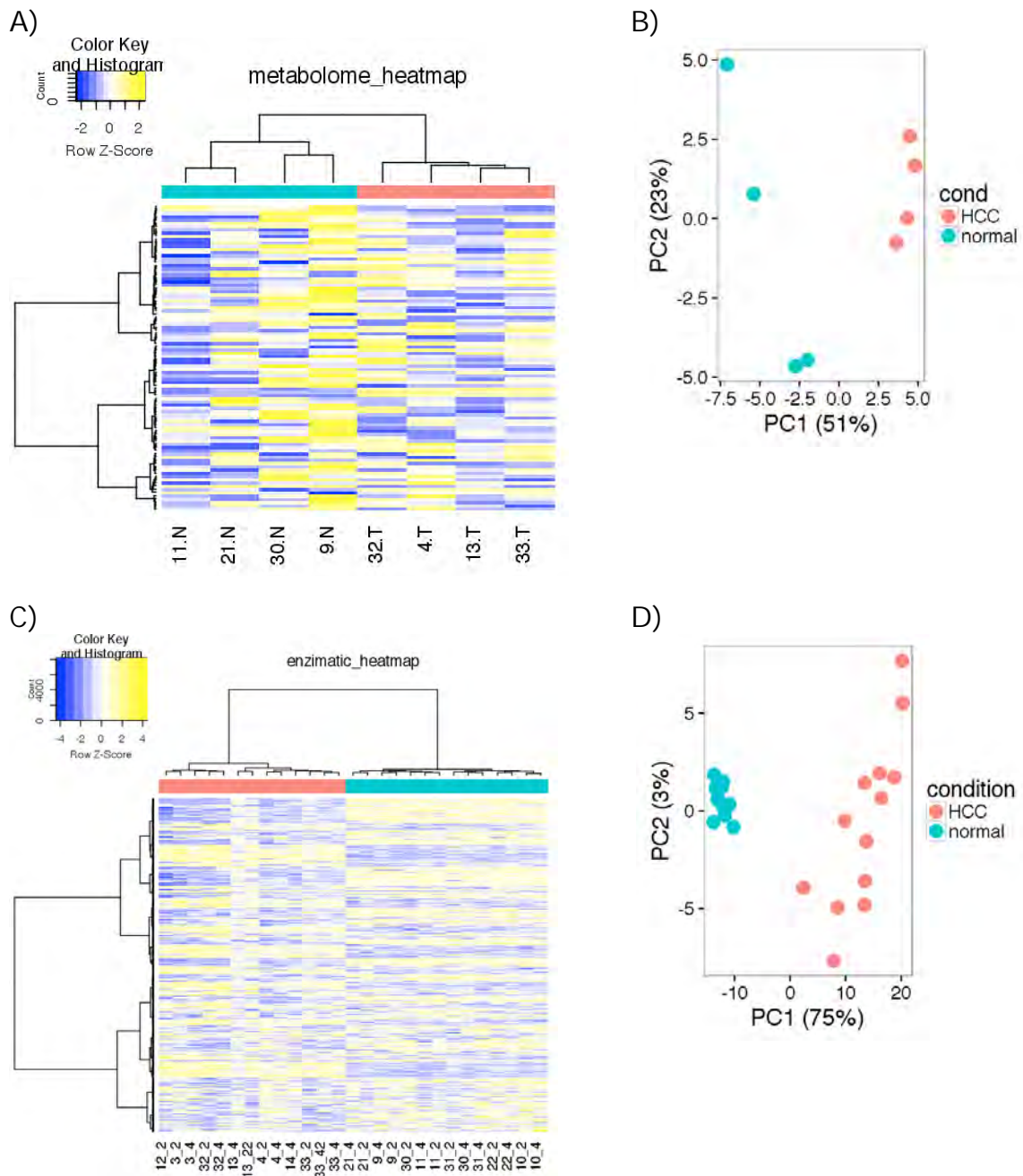


Figura 4.1. Separación de muestras por metabolismo. A) y C) son el mapa de calor de las muestras de metaboloma y transcriptoma de enzimas, respectivamente. En columnas están las muestras y en la fila son la medición (metabolito, transcrito). Abajo del dendrograma están coloreadas en rojo aquellas que pertenecen a HCC, y en Azul las normales. B) y D) contienen el PCA de los datos de metaboloma y transcriptoma, respectivamente.

Quedaba entonces la pregunta de si la expresión de genes enzimáticos mantiene estas características. Para explorar esta pregunta hicimos el mismo análisis pero enriquecido con enzimas descritas en la base de datos SMPDB ([métodos 3.3.1](#)) (Frolkis et al. 2010; Jewison et al. 2014), para obtener los ortólogos de rata ocupamos la base de datos KEGG (M. Kanehisa 2000; Minoru Kanehisa et al. 2015), apoyando la misma observación ([Figura 4.1 C, D](#)). Este análisis concuerda con una reprogramación transcripcional en enzimas.

4.2 Diferencias Metabólicas. Vías y componentes

Una vez que hemos visto que el estado metabólico logra diferenciar HCC del tejido sano, la siguiente pregunta a contestar fue ¿cuáles genes y metabolitos son los que logran tal diferencia? Por lo cual hicimos un análisis diferencial por medio de una prueba t de Welch corregido por bayes ([métodos 3.3.2](#)).

En el caso de metaboloma seleccionamos como diferenciados aquellos con al menos 0.5 del cambio de concentración en logaritmo base dos, lo cual implicaría un incremento de al menos el 50%. La corrección de pruebas múltiples del significativo estadístico fue por Benjamin Hochberg (Benjamini and Hochberg 1995). Esta corrección se ocupa para disminuir la cantidad de falsos positivos, por lo que también es conocido como FDR por sus siglas en inglés (*False Discovery Rate*). El punto de corte que ocupamos fue de 0.05, que implica la espera de un 5% de falsos positivos. Ambas consideraciones encontraron 10 de 98 metabolitos diferenciales ([Tabla 4.1](#)), algunos de los cuales son comúnmente estudiados al analizar el efecto tipo warburg, como: fructosa 1 fosfato, acetil coenzima A y ácido pirúvico. Adicionalmente los datos parecen apuntar a un desbalance en los cofactores típicos de reacciones redox, con una disminución en su forma oxidada.

Tabla 4.1 Metabolitos diferencialmente concentrados. Están con valor positivos aquellos que están acumulados en cáncer y negativo aquellos disminuidos.

Metabolito	logFC
fructosa 1,6-difosfato	3.2034617
citrulina	2.6938798
cisteina	3.9051437
acetil CoA	-2.2022600
putrescina	3.1333392
NADP+	-1.0423461
Ala	-1.0505876
NAD+	-0.7762601
fumarato	-0.9115285
piruvato	-1.8035716

Cuando hacemos un análisis de clustering vemos que en efecto estos metabolitos diferencian entre ambos fenotipos ([Figura 4.2](#)). El comportamiento de estos metabolitos diferenciales parece seguir tres grupos, uno con con metabolitos del ciclo de krebs junto con alanina, un tercero de metabolismo de aminoácidos, y un tercer grupo con putrescina y acetil coenzima A.

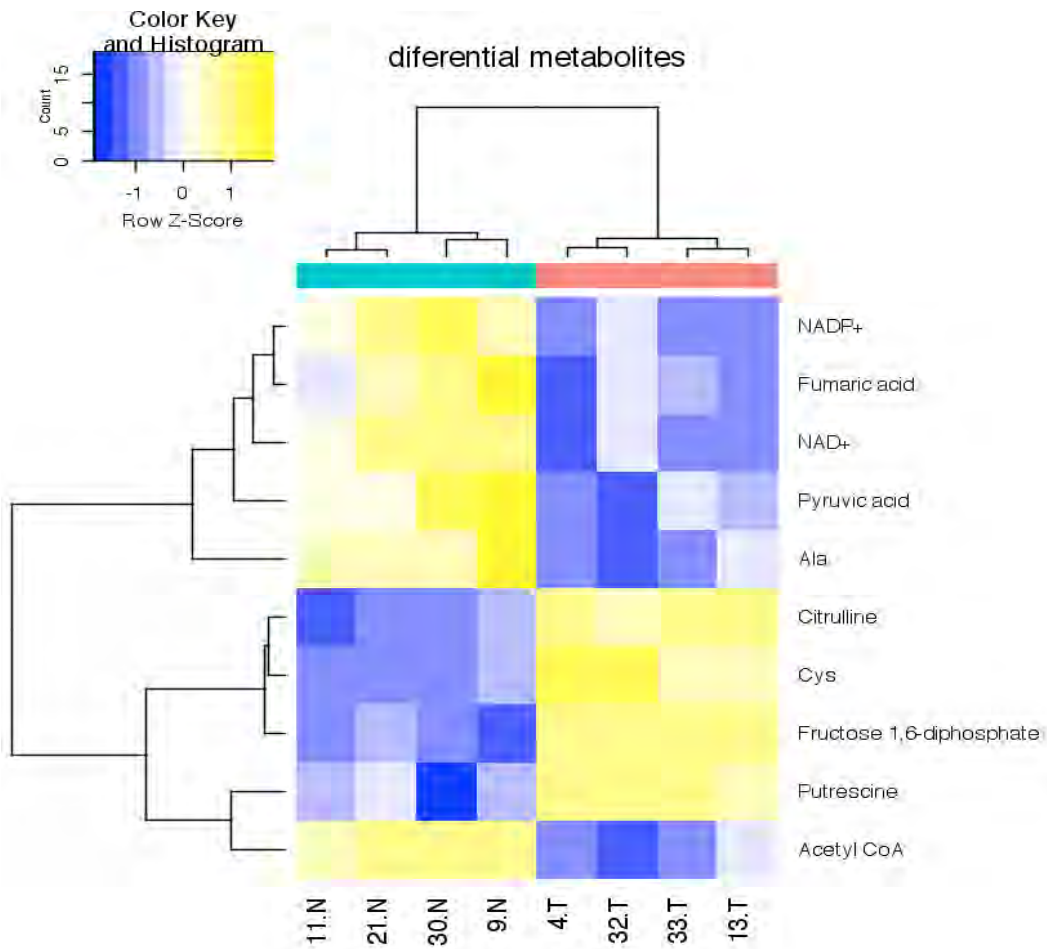


Figura 4.2 Separación de muestras por metabolitos diferenciales. En azul se encuentran las muestras tumorales y en azul se encuentran las muestras normales.

Sin embargo los cambios en metabolitos describen poco cuáles vías podrían ser las más afectadas, por lo que ocupamos algoritmos de enriquecimiento que sirven para obtener aquellas vías que dado los cambios en sus metabolitos tienen una actividad diferencial, *Metabolite Set Enrichment Analysis* (MSEA) (Jianguo Xia et al. 2015; J. Xia et al. 2012, 2009; Jianguo Xia and Wishart 2010). Se consideró como una vía diferenciada todas aquellas con un FDR menor 0.05 (Tabla 4.2). Los resultados de este análisis confirman cambios en vías concernientes al efecto de Warburg como glucólisis, gluconeogénesis y ciclo de Krebs. Además

encontramos vías en respuesta a estrés oxidante, en específico glutatión, así como cambios en el metabolismo de aminoácidos. Esto resultados concuerdan con una reprogramación metabólica tipo Warburg.

Tabla 4.2 Vías metabólicas enriquecidas en los datos metabolómicos
Ciclo de la Urea
Metabolismo del aspartato
Metabolismo de Glutatión
Metabolismo de cuerpos cetónicos
Metabolismo de metionina
Metabolismo cisteína
Metabolismo de glicina, serina y treonina
Metabolismo de ácidos grasos
Degradación de valina, leucina e isoleucina
Degradación de lisina
Metabolismo de glutamato
Ciclo del ácido cítrico
Metabolismo de piruvato
Metabolismo de fenilalanina y tirosina
Metabolismo de tirosina
Metabolismo de seleno aminoácidos
Metabolismo de cisteína
Gluconeogenesis
Glucólisis

Para complementar este análisis en los datos de microarreglos decidimos ocupar un método similar de análisis de vías, en específico GSEA (Subramanian et al. 2005; Mootha et al. 2003), el cual toma los cambios en todos los genes en una

vía y ve si en conjunto tienen una actividad diferencial. Ocupamos este algoritmo en vías definidas por SMPDB ([Tabla 4.3](#)). Todas las vías diferenciales estaban sobre reguladas en cáncer. Entre los resultados podemos encontrar sobrerrepresentadas vías comúnmente asociadas a respuesta a estrés oxidante como metabolismo de glutatión. Este resultado está en concordancia con alteraciones en balance redox anteriormente encontradas en cáncer (Lane and Fan 2015; Chaiswing, Luksana, and Oberley 2010). Curiosamente la vía de pentosas fosfato, fuente de NADPH, se encuentra enriquecida por datos de transcriptoma pero no en metaboloma. Es interesante también encontrar vías de síntesis de pirimidinas que concuerda con un metabolismo proliferativo (Lane and Fan 2015).

En general podemos decir que los datos del transcriptoma proporciona evidencia para decir que hay una activación transcripcional de vías, principalmente en aquellas relacionadas con metabolismo de aminoácidos. Los datos de metaboloma muestran una clara diferencia en el metabolismo de glucosa, ciclo de Krebs y metabolismo de aminoácidos específicos como metionina. Es también de recalcar que en ambos estudios encontraron cambios en vías relacionadas a la respuesta al estrés oxidante, en específico metabolismo de glutatión.

Tabla 4.3 Vías enriquecidas en datos transcriptómicos	
Vía	Puntuación de enriquecimiento normalizado
Efecto Warburg	1.454
Metabolismo de glutatión	1.423
Metabolismo de tirosinas	1.437
Metabolismo de Purinas	1.405
Metabolismo de pirimidinas	1.391
Metabolismo de piruvato	1.379
Metabolismo de esfingolípidos	1.349
Degradación de fructosa	1.461
Ciclo de ácido cítrico	1.355
Vía de pentosas fosfato	1.362
Metabolismo de galactosa	1.328
Metabolismo de almidón	1.312
Metabolismo de glutamato	1.462

4.3 Diferencias en actividad enzimática

Uno de los primeros objetivos de este trabajo fue darle contexto a las diferencias metabólicas de los perfiles sanos en comparación con los enfermos. Esto se exploró por medio de Dycone (Diener et al. 2016), un algoritmo que permite inferir cambios en la regulación enzimática entre dos estados fisiológicos. El objetivo de este algoritmo consiste en a partir de cambios en los datos de metaboloma inferir cambios en la actividad catalítica (K) en reacciones descritas en una red metabólica ([métodos: 3.4](#)).

La red metabólica que ocupamos se enfocó en aquellas vías encontradas en el análisis estadístico. Principalmente, en aquellas con un mayor porcentaje de metabolitos medidos ([métodos: 3.4.1](#)). Así, se obtuvo una red de 122 reacciones con 58 metabolitos ([Tabla S1](#)) de los cuales se midieron 53. Esta red consiste en ciclo de Krebs, estrés oxidativo, metabolismo de nitrógeno (ciclo de la urea, GABA, y alanina). Se agregó la vía de pentosas fosfato dada la importancia que tiene en procesos de proliferación, estrés oxidativo y producción de NADPH.

De las 122 reacciones, 27 parecían presentar cambios de actividad catalítica (K). La [Figura 4.3](#) describe cuál fue el comportamiento general de las vías estudiadas. Estos cambios en general parecen indicar que la mayor parte de los cambios en K tienden a ser positivos lo que se interpreta como una activación de la reacción. Esto está en concordancia con el enriquecimiento por GSEA, donde todas las vías diferenciales se encontraron sobre reguladas. En particular, el ciclo de Krebs está más activo en células de HCC, en concordancia con el análisis de GSEA ([Tabla 4.2](#) y [4.3](#)).

Una visualización de la red y las reacciones enzimáticas alteradas entre ambas condiciones puede verse en la [Figura 4.4](#).

Es también de recalcar que aunque pocas, la magnitud de las K inhibidas es considerable. Además algunas de estas inhibiciones ya han sido reportadas en la literatura. Por ejemplo, la inhibición predicha en la enzima argininosuccinato

sintetasa [Figura 4.4](#). ha sido reportada en pacientes con HCC (Wu et al. 2013), además de ser común en diversas líneas celulares de cáncer (Dillon et al. 2004; Delage et al. 2010). Su inhibición podría explicar la acumulación de citrulina. Otra inhibición importante es la de la fructosa-1,6-bifosfatasa, que se ha asociado con menor actividad gluconeogénica, una actividad importante del hígado y un peor pronóstico en pacientes de HCC (Hirata et al. 2016).

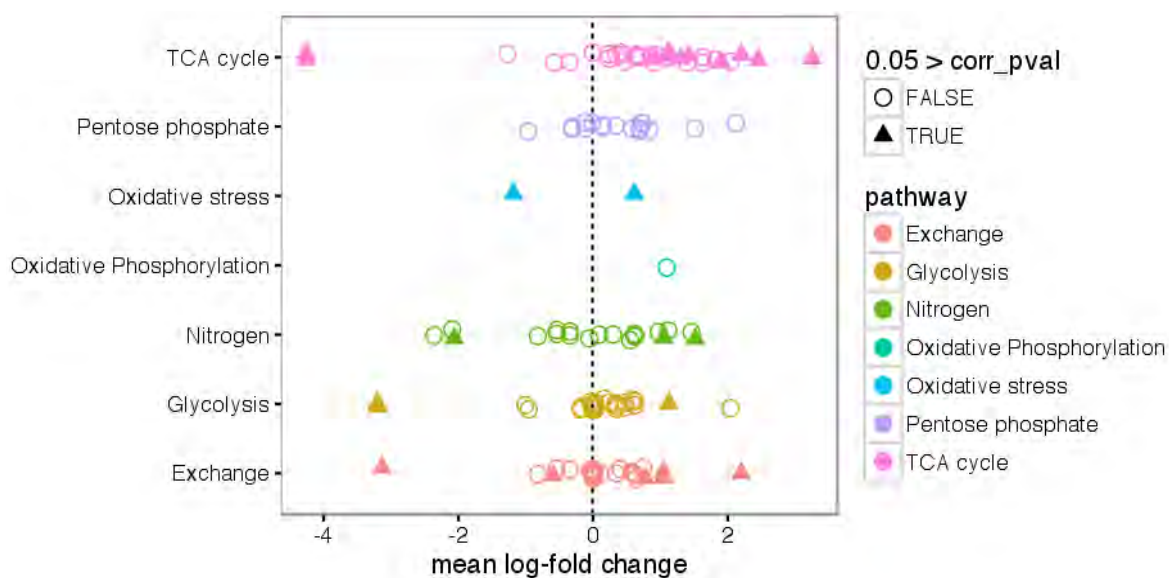


Figura 4.3 Cambio en los coeficientes de actividad catalítica k. Cada punto equivale a una reacción bioquímica y la forma de triángulo implica que la observación es estadísticamente significativa. En el eje de las x esta el cambio predicho de actividad catalítica (K), donde un cambio positivo implica mayor actividad en cáncer; en contraste un cambio con signo negativo implica una menor actividad en cáncer.

Otra de las predicciones apoyada por la literatura es la enzima málica, que ha sido relacionada a metabolismo de glutamina en cáncer, así como fuente de NADPH citosólico para síntesis de ácidos grasos. (Armstrong et al. 2004; Moreadith and Lehninger 1984).

Para ver si las predicciones de cambios de actividad podrían darse por métodos transcripcionales lo que hicimos fue analizar la expresión diferencial de los genes codificantes de las enzimas estudiadas en la red ([Tabla S2](#)). De las 27 reacciones significativas, 4 tienen al menos una isoenzima o subunidad con un cambio de expresión significativo (al menos 1 de cambio de expresión logarítmico y significativo estadístico de 0.05), [Figura 4.4](#).

Curiosamente la enzima mítica fue encontrada sobre expresada en su isoforma citosólica (Me1) y reprimida en la mitocondrial (Me3) ([Figura 4.4](#)). En concordancia la expresión proteica de Me1, se encuentra de forma homogénea en tumores ([Figura 4.6 A](#)) mientras que en tejido normal se encuentra principalmente cerca de la vena central. Esto podría indicar una mayor reducción de piruvato a malato en el citosol, mediado por NADPH. De igual modo, encontramos actividad diferencial en enzimas que utilizan NADP⁺/NADPH como la glutatión reductasa y la glutatión peroxidasa. Por lo que las concentraciones de NADP⁺/NADPH están teniendo un alto impacto en nuestro modelo. Sin embargo, este es un metabolito altamente conectado tanto en metabolismo anabólico como en balance redox, por lo que decidimos explorar mediante un mapa de calor el comportamiento transcripcional de las enzimas diferenciales que ocupan NADP⁺ o NADPH ([Figura 4.5](#)).

En las filas se denota la agrupación de genes por un algoritmo de aglomeración jerárquico basado en su expresión. Para el caso de enzimas dependientes de NADP/NADPH se pueden observar dos grandes agrupaciones. El primer grupo consiste en genes (Cyp) pertenecientes al citocromo p450, localizado en el retículo endoplásmico. Este complejo tiene la característica que al desacoplarse produce superóxido (Hrycay and Bandiera 2015; Penning et al. 1996; Krueger and Williams 2005). En los datos de transcriptoma estos genes (Cyp) están reprimidos en HCC.

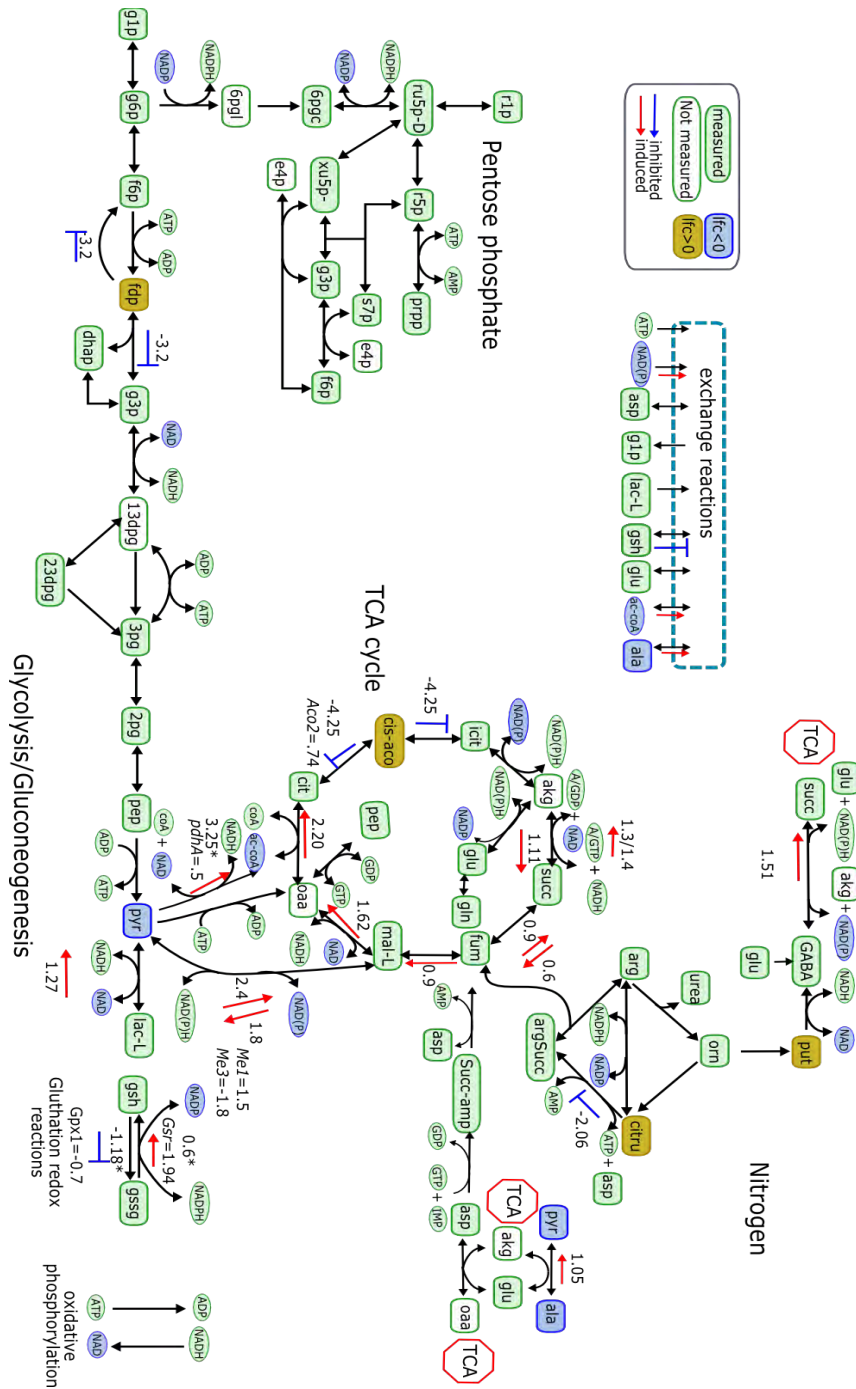


Figura 4.4 Red metabólica y sus cambios. En verde se encuentran todos los metabolitos medidos. Azul se encuentran en menor concentraciones en cáncer, amarillo están en mayor concentración. La flechas rojas implican la dirección en la que esta activada la reacción, y una flecha azul roma implica la dirección en donde se encuentra inhibida. Si la expresión diferencial de algún gen codificante de enzima tiene un p.valor de 0.05 se incluye el cambio de expresión del gen.

El segundo grupo se encontro sobre expresado en HCC y consiste en genes regulados por NRF2 (Gorrini, Harris, and Mak 2013). Este factor de transcripción se caracteriza por activar enzimas antioxidantes.

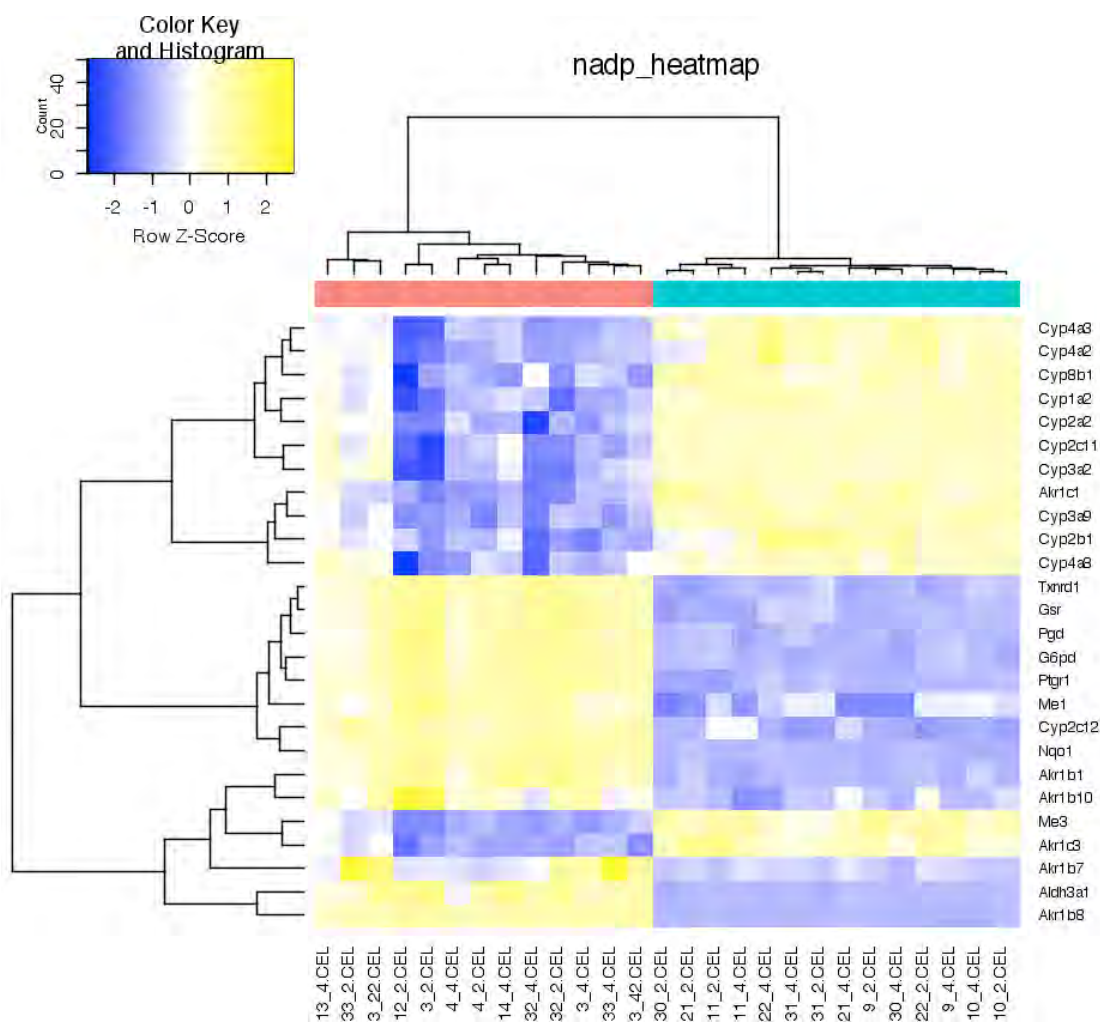


Figura 4.5 mapa de calor de comportamiento de enzimas relacionadas con NADP+ o NADPH. En rojo están las muestras de HCC y en azul aquellas de tejido normal. Se forman tres grandes grupos. El primer grupo es de genes del citocromo P450. El segundo grupo coincide con dianas de Nrf2.

Para verificar que efectivamente genes controlados por NRF2 están sobre expresados se comprobó la expresión de genes blancos de NRF2 como la glutation

s-transferasa (*Gstp1*), la gama-glutamil-transpeptidasa (*Ggt1*), la reductasa de prostaglandina (*Ptgr1*) y NADPH quinona dehidrogenasa (*Nqo1*) por RT-qPCR [Figura 4.6B](#).

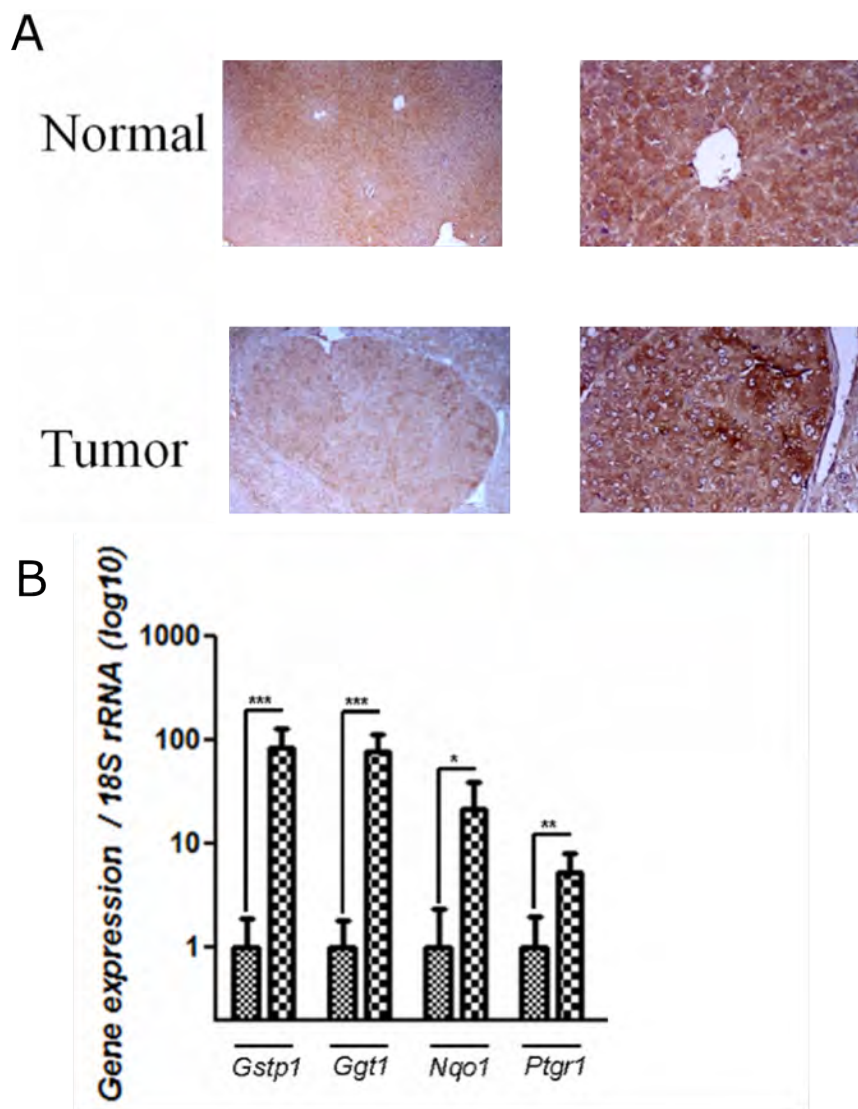


Figura 4.6 Comprobación de predicciones A) Inmunohistoquímica de la enzima málica está más expresadas en nódulos tumorales. B)qRT-PCR de genes diana de NRF2

Estos resultados parecen apuntar a una adaptación metabólica en respuesta al estrés oxidativo, inhibiendo reacciones que producen superóxido y activando la respuesta a estrés oxidativo por medio de NRF2. En concordancia, se ha reportado que pacientes de HCC tienen más marcadores de daño por estrés oxidativo que controles normales (Ma-On et al. 2017), su presencia se han asociado con el tamaño y diferenciación de tumores (Ichiba et al. 2003) y se ha considerado como predictor de recurrencia, desarrollo y pronóstico (Tanaka et al. 2011; Chuma et al. 2008). NRF2 ha sido relacionado con proliferación, invasión y quimio resistencia (Mitsuishi et al. 2012; Gao et al. 2013). su papel ha sido descrito ya sea aumentando la cantidad de enzimas antioxidantes (Mitsuishi et al. 2012), incrementando el metabolismo de glutathione (Rocha et al. 2016) y/o por el papel que tienen sobre la vía de pentosas fosfato (Mitsuishi et al. 2012).

Esta hipótesis podría explicar la disminución de actividad en aconitasa predicha por Dycone ([Figura 4.4](#)), pues la aconitasa mitocondrial tiene un centro hierro azufre (Fe-S) que es inactivado por la oxidación con superóxido (Armstrong et al. 2004). Cabe recalcar que la expresión transcripcional de la aconitasa mitocondrial es frecuentemente inhibida en cáncer (Hu et al. 2013).

En este sentido es interesante explorar los cambios predichos en el metabolismo del ácido gamma aminobutírico (GABA) [Figura 4.4](#). El catabolismo de GABA en hígado no ha sido tan estudiado como su función como neurotransmisor, sin embargo existen antecedentes en plantas (Coleman et al. 2001; Ling et al. 2015) y levadura (Coleman et al. 2001) de que este metabolismo puede llegar a ser utilizado para amortiguar los efectos de estrés oxidante.

El paso final del catabolismo de GABA, catalizado por la enzima succínico semialdehído deshidrogenasa (SS-ADH), da como resultado final succinato, metabolito que entra en el ciclo de Krebs. La entrada por succinato podría saltar el paso inhibido por la actividad de aconitasa, manteniendo así activo el ciclo de Krebs. Se ha propuesto que SS-ADH tiene un papel en el balance redox en humanos, ya que pacientes con deficiencia en esta enzima tienen marcadores de

estrés oxidativo: como menores niveles de glutatión en sangre, y mayores niveles de ácido dicarboxílico en orina (Niemi et al. 2014). La importancia de esta enzima en cáncer ha aumentado en años recientes, la inhibición de SS-ADH inhibe la proliferación de glioblastoma (El-Habr et al. 2017) y cultivos celulares de cáncer de mama (Kaur et al. 2012). Es posible que la actividad de SS-ADH provee de NADPH para mantener el balance redox, y permite el uso del ciclo de Krebs a pesar de la inhibición de aconitasa ante el estrés oxidativo. Esto está soportado en el hecho de que la actividad predicha de aconitasa y SS-ADH está anti-correlacionada con un coeficiente de -0.71 .

5.CONCLUSIONES Y PERSPECTIVAS

Dada la incidencia y mortalidad del HCC, este se ha vuelto un problema de salud pública nacional y mundial (Misael et al. 2010), por lo que se ha buscado un mejor entendimiento de su desarrollo, progreso y mantenimiento. Un elemento importante en el estudio del cáncer ha sido el estudio de las adaptaciones que desarrolla para mantener funciones metabólicas, tales como la proliferación y el balance redox.

En este contexto, este trabajo analiza tres niveles de descripción importantes de un modelo de hepatocarcinogénesis química en rata, diferencias metabólicas, comportamiento transcripcional y diferencias en actividad enzimática. Ambos datos logran diferenciar nuestros dos tipos de muestra, hígado sano y tumor de HCC.

Un análisis de vías en los datos de metaboloma apuntó a cambios significativos en vías relacionadas con metabolismo de aminoácidos, así como del metabolismo central y balance redox. Acorde con estos resultados, los datos de transcripción indican una activación transcripcional con genes relacionados con enzimas del ciclo de krebs, metabolismo de azúcares y balance redox.

Este análisis nos guió para reconstruir una red metabólica que, acoplada a un modelo matemático, nos permitió inferir cambios en actividad enzimática dado las concentraciones de metabolitos. Ocupando este análisis encontramos que la regulación de las vías metabólicas estudiadas tiende a la activación. Esta observación está apoyada por el análisis de enriquecimiento de vías en transcriptomas.

Adicionalmente se pudieron recuperar diferencias en la actividad de reacciones ya descritas en diferentes tipos de cáncer, como la inhibición de argininosuccinato sintasa y la glucosa 1,6 bifosfatasa.

A su vez encontramos cambios de mayor actividad en reacciones asociadas con NADP/NADPH. Al observar el comportamiento de enzimas que ocupan este cofactor observamos que el comportamiento de aquellas enzimas con mayor diferencias estaban asociadas con estrés oxidativo. Aquellas enzimas que favorecen la aparición de ROS, como citocromo p450, están inhibidas. En contraste, aquellas enzimas relacionadas con respuesta antioxidante, específicamente por NRF2, están activadas.

Estos resultados nos permiten proponer que en respuesta a un ambiente oxidante los siguientes procesos se producen (Figura 5.1): 1) Activación a las respuestas antioxidantes por NRF2; 2) Inhibición de la actividad enzimática de la aconitasa; 3) Utilización de GABA para recuperar la poza de NADPH así como para activar el ciclo de Krebs.

Aunque el modelo en rata nos permite generar hipótesis sobre el estado metabólico en HCC, queda por saber el impacto que estas enzimas podrían tener en pacientes humanos, sobre todo aquellas involucradas en el balance redox. A su vez, es importante considerar que aunque los datos de tumores son una fuente importante de información, es difícil identificar los cambios metabólicos que permiten la progresión de tejido normal a tejido neoplásico. Aun cuando es posible que el ambiente cirrótico influye en el metabolismo de HCC, sería importante investigar qué impacto tiene el insulto inicial que provoca esta cirrosis.

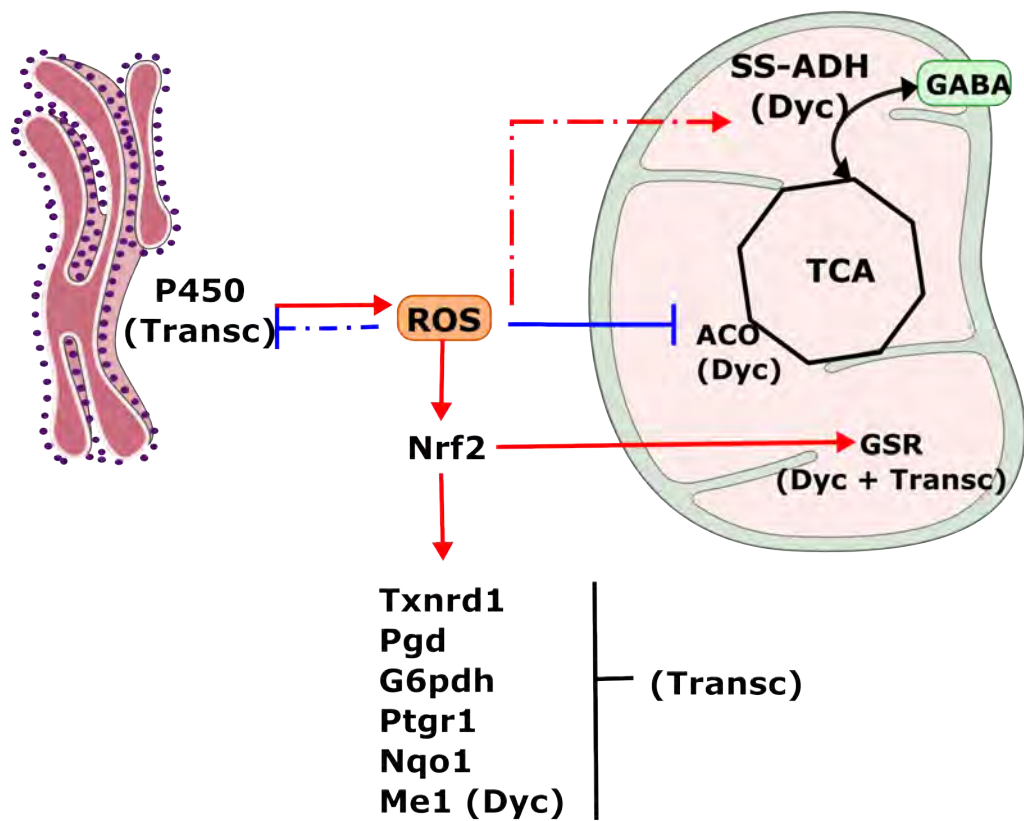


Figura 5.1. Estrés oxidativo en modelo de hepatocarcinogénesis química. Dado los resultados del modelo computacional y los datos transcriptómicos proponemos consecuencias del estrés oxidante. Entre paréntesis se encuentra el tipo de dato que soporta este modelo. Dyc para Dycone y Transc para transcripción. Líneas punteadas son interacciones aun no descritas en cáncer mientras que las líneas continuas ya han sido reportadas.

6. BIBLIOGRAFÍA

- Agren, Rasmus, Sergio Bordel, Adil Mardinoglu, Natapol Pornputtpong, Intawat Nookaew, and Jens Nielsen. 2012. "Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT." *PLoS Computational Biology* 8 (5):e1002518.
- Akiyama, Kenji, Eisuke Chikayama, Hiroaki Yuasa, Yukihisa Shimada, Takayuki Tohge, Kazuo Shinozaki, Masami Yokota Hirai, Tetsuya Sakurai, Jun Kikuchi, and Kazuki Saito. 2008. "PRIME: A Web Site That Assembles Tools for Metabolomics and Transcriptomics." *In Silico Biology* 8 (3-4):339–45.
- Armstrong, Jeffrey S., Matthew Whiteman, Hongyuan Yang, and Dean P. Jones. 2004. "The Redox Regulation of Intermediary Metabolism by a Superoxide-Aconitase Rheostat." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 26 (8):894–900.
- Badvie, S. 2000. "Hepatocellular Carcinoma." *Postgraduate Medical Journal* 76 (891):4–11.
- Benjamini, Yoav, and Yoel Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B, Statistical Methodology*. JSTOR, 289–300.
- Biosystems", "applied. 2010. "High Capacity RNA-to-cDNA Master Mix Protocol." Life Technologies Corporation.
- Blais, Edik M., Kristopher D. Rawls, Bonnie V. Dougherty, Zhuo I. Li, Glynis L. Kolling, Ping Ye, Anders Wallqvist, and Jason A. Papin. 2017. "Reconciled Rat and Human Metabolic Networks for Comparative Toxicogenomics and Biomarker Predictions." *Nature Communications* 8 (February):14250.
- Boonstra, Johannes, and Jan Andries Post. 2004. "Molecular Events Associated with Reactive Oxygen Species and Cell Cycle Progression in Mammalian Cells." *Gene* 337 (August):1–13.
- Cantor, Jason R., and David M. Sabatini. 2012. "Cancer Cell Metabolism: One Hallmark, Many Faces." *Cancer Discovery* 2 (10):881–98.
- Chaiswing, Luksana, Chaiswing Luksana, and Terry D. Oberley. 2010. "Extracellular/Microenvironmental Redox State." *Antioxidants & Redox Signaling* 13 (4):449–65.
- Chuma, Makoto, Shuhei Hige, Mitsuru Nakanishi, Koji Ogawa, Mitsuteru Natsuizaka, Yoichi Yamamoto, and Masahiro Asaka. 2008. "8-Hydroxy-2'-Deoxy-Guanosine Is a Risk Factor for Development of Hepatocellular Carcinoma in Patients with Chronic Hepatitis C Virus Infection." *Journal of Gastroenterology and Hepatology* 23 (9):1431–36.
- Coleman, S. T., T. K. Fang, S. A. Rovinsky, F. J. Turano, and W. S. Moye-Rowley. 2001. "Expression of a Glutamate Decarboxylase Homologue Is Required for Normal Oxidative Stress Tolerance in *Saccharomyces Cerevisiae*." *The Journal of Biological Chemistry* 276 (1):244–50.
- Dai, Manhong, Pinglang Wang, Andrew D. Boyd, Georgi Kostov, Brian Athey, Edward G. Jones, William E. Bunney, et al. 2005. "Evolving Gene/transcript

- Definitions Significantly Alter the Interpretation of GeneChip Data." *Nucleic Acids Research* 33 (20):e175.
- Dang, Chi V., Jung-Whan Kim, Ping Gao, and Jason Yustein. 2008. "The Interplay between MYC and HIF in Cancer." *Nature Reviews. Cancer* 8 (1):51–56.
- DeBerardinis, Ralph J., Julian J. Lum, Georgia Hatzivassiliou, and Craig B. Thompson. 2008. "The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation." *Cell Metabolism* 7 (1):11–20.
- Delage, Barbara, Dean A. Fennell, Linda Nicholson, Iain McNeish, Nicholas R. Lemoine, Tim Crook, and Peter W. Szlosarek. 2010. "Arginine Deprivation and Argininosuccinate Synthetase Expression in the Treatment of Cancer." *International Journal of Cancer. Journal International Du Cancer* 126 (12):2762–72.
- Diehn, Maximilian, Robert W. Cho, Neethan A. Lobo, Tomer Kalisky, Mary Jo Dorie, Angela N. Kulp, Dalong Qian, et al. 2009. "Association of Reactive Oxygen Species Levels and Radioresistance in Cancer Stem Cells." *Nature* 458 (7239):780–83.
- Diener, Christian, Felipe Muñoz-Gonzalez, Sergio Encarnación, and Osbaldo Resendis-Antonio. 2016. "The Space of Enzyme Regulation in HeLa Cells Can Be Inferred from Its Intracellular Metabolome." *Scientific Reports* 6 (June):28415.
- Dillon, Brian J., Victor G. Prieto, Steven A. Curley, C. Mark Ensor, Frederick W. Holtsberg, John S. Bomalaski, and Mike A. Clark. 2004. "Incidence and Distribution of Argininosuccinate Synthetase Deficiency in Human Cancers: A Method for Identifying Cancers Sensitive to Arginine Deprivation." *Cancer* 100 (4):826–33.
- Duarte, Natalie C., Scott A. Becker, Neema Jamshidi, Ines Thiele, Monica L. Mo, Thuy D. Vo, Rohith Srivas, and Bernhard Ø. Palsson. 2007. "Global Reconstruction of the Human Metabolic Network Based on Genomic and Bibliomic Data." *Proceedings of the National Academy of Sciences of the United States of America* 104 (6):1777–82.
- Dunn, Warwick B., Nigel J. C. Bailey, and Helen E. Johnson. 2005. "Measuring the Metabolome: Current Analytical Technologies." *The Analyst* 130 (5):606–25.
- Eckel-Mahan, Kristin L., Vishal R. Patel, Robert P. Mohny, Katie S. Vignola, Pierre Baldi, and Paolo Sassone-Corsi. 2012. "Coordination of the Transcriptome and Metabolome by the Circadian Clock." *Proceedings of the National Academy of Sciences of the United States of America* 109 (14):5541–46.
- El-Habr, Elias A., Luiz G. Dubois, Fanny Burel-Vandenbos, Alexandra Bogeas, Joanna Lipecka, Laurent Turchi, François-Xavier Lejeune, et al. 2017. "A Driver Role for GABA Metabolism in Controlling Stem and Proliferative Cell State through GHB Production in Glioma." *Acta Neuropathologica* 133 (4):645–60.
- Elseman, Ibrahim E., Adil Mardinoglu, Saeed Shoaie, Taysir H. Soliman, and Jens Nielsen. 2016. "Systems Biology Analysis of Hepatitis C Virus Infection Reveals the Role of Copy Number Increases in Regions of Chromosome 1q in Hepatocellular Carcinoma Metabolism." *Molecular bioSystems* 12 (5):1496–1506.

- El-Serag, Hashem B., and K. Lenhard Rudolph. 2007. "Hepatocellular Carcinoma: Epidemiology and Molecular Carcinogenesis." *Gastroenterology* 132 (7):2557–76.
- Famili, Iman, Radhakrishnan Mahadevan, and Bernhard O. Palsson. 2005. "K-Cone Analysis: Determining All Candidate Values for Kinetic Parameters on a Network Scale." *Biophysical Journal* 88 (3):1616–25.
- Farazi, Paraskevi A., and Ronald A. DePinho. 2006. "Hepatocellular Carcinoma Pathogenesis: From Genes to Environment." *Nature Reviews. Cancer* 6 (9):674–87.
- Ferlay, Jacques, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. 2014. "Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012." *International Journal of Cancer* 136 (5):E359–86.
- Frolkis, Alex, Craig Knox, Emilia Lim, Timothy Jewison, Vivian Law, David D. Hau, Phillip Liu, et al. 2010. "SMPDB: The Small Molecule Pathway Database." *Nucleic Acids Research* 38 (Database issue):D480–87.
- Gao, Ai-Mei, Zun-Ping Ke, Fang Shi, Guang-Chun Sun, and Hui Chen. 2013. "Chrysin Enhances Sensitivity of BEL-7402/ADM Cells to Doxorubicin by Suppressing PI3K/Akt/Nrf2 and ERK/Nrf2 Pathway." *Chemico-Biological Interactions* 206 (1):100–108.
- Gille, Christoph, Christian Bölling, Andreas Hoppe, Sascha Bulik, Sabrina Hoffmann, Katrin Hübner, Anja Karlstädt, et al. 2010. "HepatoNet1: A Comprehensive Metabolic Reconstruction of the Human Hepatocyte for the Analysis of Liver Physiology." *Molecular Systems Biology* 6 (September):411.
- Goodacre, R. 2004. "Metabolomics by Numbers: Acquiring and Understanding Global Metabolite Data." *Trends in Biotechnology* 22 (5):245–52.
- Gorrini, Chiara, Isaac S. Harris, and Tak W. Mak. 2013. "Modulation of Oxidative Stress as an Anticancer Strategy." *Nature Reviews. Drug Discovery* 12 (12):931–47.
- Hanahan, Douglas, and Robert A. Weinberg. 2011. "Hallmarks of Cancer: The Next Generation." *Cell* 144 (5):646–74.
- Hanigan, Marie H., and Henry C. Pitot. 1985. "Gamma-Glutamyl Transpeptidase – Its Role in Hepatocarcinogenesis." *Carcinogenesis* 6 (2):165–72.
- Hirata, Hidenari, Keishi Sugimachi, Hisateru Komatsu, Masami Ueda, Takaaki Masuda, Ryutaro Uchi, Shotaro Sakimura, et al. 2016. "Decreased Expression of Fructose-1,6-Bisphosphatase Associates with Glucose Metabolism and Tumor Progression in Hepatocellular Carcinoma." *Cancer Research* 76 (11):3265–76.
- Hrycay, Eugene G., and Stelvio M. Bandiera. 2015. "Involvement of Cytochrome P450 in Reactive Oxygen Species Formation and Cancer." *Advances in Pharmacology* 74 (May):35–84.
- Hu, Jie, Jason W. Locasale, Jason H. Bielas, Jacintha O'Sullivan, Kieran Sheahan, Lewis C. Cantley, Matthew G. Vander Heiden, and Dennis Vitkup. 2013. "Heterogeneity of Tumor-Induced Gene Expression Changes in the Human Metabolic Network." *Nature Biotechnology* 31 (6):522–29.

- Ichiba, Miho, Yoshiko Maeta, Tomoyuki Mukoyama, Toshiya Saeki, Sakiko Yasui, Takamasa Kanbe, Jun-Ichi Okano, et al. 2003. "Expression of 8-Hydroxy-2'-Deoxyguanosine in Chronic Liver Disease and Hepatocellular Carcinoma." *Liver International: Official Journal of the International Association for the Study of the Liver* 23 (5):338–45.
- Irizarry, Rafael A., Benjamin M. Bolstad, Francois Collin, Leslie M. Cope, Bridget Hobbs, and Terence P. Speed. 2003. "Summaries of Affymetrix GeneChip Probe Level Data." *Nucleic Acids Research* 31 (4):e15.
- Jerby, Livnat, Tomer Shlomi, and Eytan Ruppin. 2010. "Computational Reconstruction of Tissue-Specific Metabolic Models: Application to Human Liver Metabolism." *Molecular Systems Biology* 6 (September):401.
- Jewison, Timothy, Yilu Su, Fatemeh Miri Disfany, Yongjie Liang, Craig Knox, Adam Maciejewski, Jenna Poelzer, et al. 2014. "SMPDB 2.0: Big Improvements to the Small Molecule Pathway Database." *Nucleic Acids Research* 42 (Database issue):D478–84.
- Jha, Abhishek K., Stanley Ching-Cheng Huang, Alexey Sergushichev, Vicky Lampropoulou, Yulia Ivanova, Ekaterina Loginicheva, Karina Chmielewski, et al. 2015. "Network Integration of Parallel Metabolic and Transcriptional Data Reveals Metabolic Modules That Regulate Macrophage Polarization." *Immunity* 42 (3):419–30.
- Jo, Masayasu, Taichiro Nishikawa, Tomoki Nakajima, Yoshihisa Okada, Kanji Yamaguchi, Hironori Mitsuyoshi, Kohichiroh Yasui, et al. 2011. "Oxidative Stress Is Closely Associated with Tumor Angiogenesis of Hepatocellular Carcinoma." *Journal of Gastroenterology* 46 (6):809–21.
- Jones, Russell G., and Craig B. Thompson. 2009. "Tumor Suppressors and Cell Metabolism: A Recipe for Cancer Growth." *Genes & Development* 23 (5):537–48.
- Kanehisa, M. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28 (1):27–30.
- Kanehisa, Minoru, Kanehisa Minoru, Sato Yoko, Kawashima Masayuki, Furumichi Miho, and Tanabe Mao. 2015. "KEGG as a Reference Resource for Gene and Protein Annotation." *Nucleic Acids Research* 44 (D1):D457–62.
- Kaur, Hitchintan, Shihong Mao, Quanwen Li, Mansoureh Sameni, Stephen A. Krawetz, Bonnie F. Sloane, and Raymond R. Mattingly. 2012. "RNA-Seq of Human Breast Ductal Carcinoma in Situ Models Reveals Aldehyde Dehydrogenase Isoform 5A1 as a Novel Potential Target." *PloS One* 7 (12):e50249.
- Krueger, Sharon K., and David E. Williams. 2005. "Mammalian Flavin-Containing Monooxygenases: Structure/function, Genetic Polymorphisms and Role in Drug Metabolism." *Pharmacology & Therapeutics* 106 (3):357–87.
- Lane, A. N., and T. W. -. M. Fan. 2015. "Regulation of Mammalian Nucleotide Metabolism and Biosynthesis." *Nucleic Acids Research* 43 (4):2466–85.
- Li, Feng, Yunyue Wang, Karen I. Zeller, James J. Potter, Diane R. Wonsey, Kathryn A. O'Donnell, Jung-Whan Kim, Jason T. Yustein, Linda A. Lee, and Chi V. Dang. 2005. "Myc Stimulates Nuclearly Encoded Mitochondrial Genes and Mitochondrial Biogenesis." *Molecular and Cellular Biology* 25 (14):6225–34.

- Ling, Li, Li Jiangang, Shen Minchong, Zhang Chunlei, and Dong Yuanhua. 2015. "Cold Plasma Treatment Enhances Oilseed Rape Seed Germination under Drought Stress." *Scientific Reports* 5 (1). <https://doi.org/10.1038/srep13033>.
- Ma-On, Chakriwong, Anapat Sanpavat, Patcharawalai Whongsiri, Surasit Suwannasin, Nattiya Hirankarn, Pisit Tangkijvanich, and Chanchai Boonla. 2017. "Oxidative Stress Indicated by Elevated Expression of Nrf2 and 8-OHdG Promotes Hepatocellular Carcinoma Progression." *Medical Oncology* 34 (4):57.
- Mardinoglu, Adil, Rasmus Agren, Caroline Kampf, Anna Asplund, Mathias Uhlen, and Jens Nielsen. 2014. "Genome-Scale Metabolic Modelling of Hepatocytes Reveals Serine Deficiency in Patients with Non-Alcoholic Fatty Liver Disease." *Nature Communications* 5:3083.
- Mardinoglu, Adil, Saeed Shoaie, Mattias Bergentall, Pouyan Ghaffari, Cheng Zhang, Erik Larsson, Fredrik Bäckhed, and Jens Nielsen. 2015. "The Gut Microbiota Modulates Host Amino Acid and Glutathione Metabolism in Mice." *Molecular Systems Biology* 11 (10):834.
- M. Ervik, F. Lam, J. Ferlay, L. Mery, I. Soerjomataram, F. Bray. 2016. "Cancer Today." *Cancer Today*. <http://gco.iarc.fr/today>.
- Misael, Uribe-Esquivel, Mauricio García-Sáenz-de-Sicilia, Norberto Chávez-Tapia, and José de Jesús Román-Sandoval. 2010. "Carcinoma Hepatocelular." *Revista de Gastroenterología de México* 75 (S2):76–168.
- Mitsuishi, Yoichiro, Keiko Taguchi, Yukie Kawatani, Tatsuhiko Shibata, Toshihiro Nukiwa, Hiroyuki Aburatani, Masayuki Yamamoto, and Hozumi Motohashi. 2012. "Nrf2 Redirects Glucose and Glutamine into Anabolic Pathways in Metabolic Reprogramming." *Cancer Cell* 22 (1):66–79.
- Mootha, Vamsi K., Cecilia M. Lindgren, Karl-Fredrik Eriksson, Aravind Subramanian, Smita Sihag, Joseph Lehar, Pere Puigserver, et al. 2003. "PGC-1alpha-Responsive Genes Involved in Oxidative Phosphorylation Are Coordinately Downregulated in Human Diabetes." *Nature Genetics* 34 (3):267–73.
- Moreadith, R. W., and A. L. Lehninger. 1984. "The Pathways of Glutamate and Glutamine Oxidation by Tumor Cell Mitochondria. Role of Mitochondrial NAD(P)⁺-Dependent Malic Enzyme." *The Journal of Biological Chemistry* 259 (10):6215–21.
- Nielsen, Jens, and Stephen Oliver. 2005. "The next Wave in Metabolome Analysis." *Trends in Biotechnology* 23 (11):544–46.
- Niemi, Anna-Kaisa, Candida Brown, Tereza Moore, Gregory M. Enns, and Tina M. Cowan. 2014. "Evidence of Redox Imbalance in a Patient with Succinic Semialdehyde Dehydrogenase Deficiency." *Molecular Genetics and Metabolism Reports* 1:129–32.
- Nishikawa, Taichiro, Tomoki Nakajima, Tatsuo Katagishi, Yoshihisa Okada, Masayasu Jo, Keizo Kagawa, Takeshi Okanoue, Yoshito Itoh, and Toshikazu Yoshikawa. 2009. "Oxidative Stress May Enhance the Malignant Potential of Human Hepatocellular Carcinoma by Telomerase Activation." *Liver International: Official Journal of the International Association for the Study of the Liver* 29 (6):846–56.

- Orth, Jeffrey D., Ines Thiele, and Bernhard Ø. Palsson. 2010. "What Is Flux Balance Analysis?" *Nature Biotechnology* 28 (3):245–48.
- Pellicoro, Antonella, Prakash Ramachandran, John P. Iredale, and Jonathan A. Fallowfield. 2014. "Liver Fibrosis and Repair: Immune Regulation of Wound Healing in a Solid Organ." *Nature Reviews. Immunology* 14 (3):181–94.
- Penning, T. M., S. T. Ohnishi, T. Ohnishi, and R. G. Harvey. 1996. "Generation of Reactive Oxygen Species during the Enzymatic Oxidation of Polycyclic Aromatic Hydrocarbon Trans-Dihydrodiols Catalyzed by Dihydrodiol Dehydrogenase." *Chemical Research in Toxicology* 9 (1):84–92.
- Perry, G., A. K. Raina, A. Nunomura, T. Wataya, L. M. Sayre, and M. A. Smith. 2000. "How Important Is Oxidative Damage? Lessons from Alzheimer's Disease." *Free Radical Biology & Medicine* 28 (5):831–34.
- Resendis-Antonio, Osbaldo. 2009. "Filling Kinetic Gaps: Dynamic Modeling of Metabolism Where Detailed Kinetic Information Is Lacking." *PLoS One* 4 (3):e4967.
- Resendis-Antonio, Osbaldo, Carolina González-Torres, Gustavo Jaime-Muñoz, Claudia Erika Hernandez-Patiño, and Carlos Felipe Salgado-Muñoz. 2015. "Modeling Metabolism: A Window toward a Comprehensive Interpretation of Networks in Cancer." *Seminars in Cancer Biology* 30:79–87.
- Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7):e47.
- Rocha, Clarissa Ribeiro Reily, Gustavo Satoru Kajitani, Annabel Quinet, Rodrigo Soares Fortunato, and Carlos Frederico Martins Menck. 2016. "NRF2 and Glutathione Are Key Resistance Mediators to Temozolomide in Glioma and Melanoma Cells." *Oncotarget* 7 (30):48081–92.
- Schiffer, Eduardo, Chantal Housset, Wulfran Cacheux, Dominique Wendum, Christèle Desbois-Mouthon, Colette Rey, François Clergue, Raoul Poupon, Véronique Barbu, and Olivier Rosmorduc. 2005. "Gefitinib, an EGFR Inhibitor, Prevents Hepatocellular Carcinoma Development in the Rat Liver with Cirrhosis." *Hepatology* 41 (2):307–14.
- Sigurdsson, Martin I., Neema Jamshidi, Eiríkur Steingrímsson, Ines Thiele, and Bernhard Ø. Palsson. 2010. "A Detailed Genome-Wide Reconstruction of Mouse Metabolism Based on Human Recon 1." *BMC Systems Biology* 4 (October):140.
- Sporn, Michael B., and Karen T. Liby. 2012. "NRF2 and Cancer: The Good, the Bad and the Importance of Context." *Nature Reviews. Cancer* 12 (8):564–71.
- Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences of the United States of America* 102 (43):15545–50.
- Szatrowski, T. P., and C. F. Nathan. 1991. "Production of Large Amounts of Hydrogen Peroxide by Human Tumor Cells." *Cancer Research* 51 (3):794–98.
- Taguchi, Keiko, Hozumi Motohashi, and Masayuki Yamamoto. 2011. "Molecular

- Mechanisms of the Keap1–Nrf2 Pathway in Stress Response and Cancer Evolution." *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 16 (2):123–40.
- Tanaka, Shinji, Kaoru Mogushi, Mahmut Yasen, Daisuke Ban, Norio Noguchi, Takumi Irie, Atsushi Kudo, et al. 2011. "Oxidative Stress Pathways in Noncancerous Human Liver Tissue to Predict Hepatocellular Carcinoma Recurrence: A Prospective, Multicenter Study." *Hepatology* 54 (4):1273–81.
- Thiele, Ines, Neil Swainston, Ronan M. T. Fleming, Andreas Hoppe, Swagatika Sahoo, Maike K. Aurich, Hulda Haraldsdottir, et al. 2013. "A Community-Driven Global Reconstruction of Human Metabolism." *Nature Biotechnology* 31 (5):419–25.
- Toyokuni, Shinya, Keisei Okamoto, Junji Yodoi, and Hiroshi Hiai. 1995. "Persistent Oxidative Stress in Cancer." *FEBS Letters* 358 (1):1–3.
- Urtasun, Raquel, Carmen Berasain, and Matías A. Avila. 2015. "Oxidative Stress Mechanisms in Hepatocarcinogenesis." In *Oxidative Stress in Applied Basic Research and Clinical Practice*, 449–77.
- Vander Heiden, M. G., L. C. Cantley, and C. B. Thompson. 2009. "Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation." *Science* 324 (5930):1029–33.
- Villas-Bôas, Silas G., Sandrine Mas, Mats Åkesson, Jørn Smedsgaard, and Jens Nielsen. 2005. "Mass Spectrometry in Metabolome Analysis." *Mass Spectrometry Reviews* 24 (5):613–46.
- Wang, Yuliang, James A. Eddy, and Nathan D. Price. 2012. "Reconstruction of Genome-Scale Metabolic Models for 126 Human Tissues Using mCADRE." *BMC Systems Biology* 6 (December):153.
- Warburg, O. 1927. "THE METABOLISM OF TUMORS IN THE BODY." *The Journal of General Physiology* 8 (6):519–30.
- Wei, Haoran, Wei HaoRan, and Gao Ping. 2017. "Regulation of Cancer Cell Metabolism." *SCIENTIA SINICA Vitae* 47 (1):132–39.
- Wu, Lina, Li Li, Shucong Meng, Ruizhao Qi, Zebin Mao, and Ming Lin. 2013. "Expression of Argininosuccinate Synthetase in Patients with Hepatocellular Carcinoma." *Journal of Gastroenterology and Hepatology* 28 (2):365–68.
- Xia, Jianguo, Xia Jianguo, Igor V. Sinelnikov, Han Beomsoo, and David S. Wishart. 2015. "MetaboAnalyst 3.0—making Metabolomics More Meaningful." *Nucleic Acids Research* 43 (W1):W251–57.
- Xia, Jianguo, and David S. Wishart. 2010. "MSEA: A Web-Based Tool to Identify Biologically Meaningful Patterns in Quantitative Metabolomic Data." *Nucleic Acids Research* 38 (Web Server issue):W71–77.
- Xia, J., R. Mandal, I. V. Sinelnikov, D. Broadhurst, and D. S. Wishart. 2012. "MetaboAnalyst 2.0—a Comprehensive Server for Metabolomic Data Analysis." *Nucleic Acids Research* 40 (W1):W127–33.
- Xia, J., N. Psychogios, N. Young, and D. S. Wishart. 2009. "MetaboAnalyst: A Web Server for Metabolomic Data Analysis and Interpretation." *Nucleic Acids Research* 37 (Web Server):W652–60.

Anexo

A



B



Figura S1 Tejido normal. a) tejido de una rata sana sin tratamiento con DEN. b) marcaje histológico de actividad GGT en tejido normal

Tabla S1 Red metabólica

rxn	abbreviation	pathway	KEGG_enzyme
13dpg \rightleftharpoons 23dpg	BPGM	Glycolysis	5.4.2.4, 5.4.2.11
23dpg \rightarrow 3pg	BPGM	Glycolysis	3.1.3.13
2pg \rightleftharpoons pep	ENO	Glycolysis	4.2.1.11
fdp \rightleftharpoons dhap + g3p	ALDO	Glycolysis	4.1.2.13
fdp \rightarrow f6p	FBP	Glycolysis	3.1.3.11
g3p + nad + \rightleftharpoons 13dpg + nadh	GAPDH	Glycolysis	1.2.1.12, 1.2.1.59
atp + f6p \rightarrow adp + fdp	PFK	Glycolysis	2.7.1.11
g6p \rightleftharpoons f6p	G6I	Glycolysis	5.3.1.9
13dpg \rightarrow + 3pg	ACYP	Glycolysis	3.6.1.7
3pg + atp \rightleftharpoons 13dpg + adp	PGK	Glycolysis	2.7.2.3
2pg \rightleftharpoons 3pg	PGAM	Glycolysis	5.4.2.11, 5.4.2.12
g1p \rightleftharpoons g6p	PGM	Glycolysis	5.4.2.2, 5.4.2.5
adp + pep \rightarrow atp + pyr	PKM	Glycolysis	2.7.1.40
dhap \rightleftharpoons g3p	TPI	Glycolysis	5.3.1.1
nad + lac-L \rightleftharpoons nadh + pyr	LDH	Glycolysis	1.1.1.27
g6p + nadp \rightarrow 6pgc + nadph	G6PD	Pentose phosphate	1.1.1.49, 1.1.1.363
nadp + 6pgc \rightleftharpoons nadph + ru5p-D	PGD	Pentose phosphate	1.1.1.44, 1.1.1.351, 3.1.1.31
r1p \rightleftharpoons r5p	PGM	Pentose phosphate	5.4.2.2, 5.4.2.7
atp + r5p \rightleftharpoons amp + prpp	PRPS	Pentose phosphate	2.7.6.1
ru5p-D \rightleftharpoons xu5p-D	RPE	Pentose phosphate	5.1.3.1
r5p \rightleftharpoons ru5p-D	RPI	Pentose phosphate	5.3.1.6
g3p + s7p \rightleftharpoons f6p + e4p	TALDO	Pentose phosphate	2.2.1.2
r5p + xu5p-D \rightleftharpoons g3p + s7p	TKT	Pentose phosphate	2.2.1.1

xu5p-D + e4p \rightleftharpoons f6p + g3p	TKT	Pentose phosphate	2.2.1.1
pyr + nad+coa->acoA+nadh	PDH	TCA cycle	1.2.4.1, 1.8.1.4, 2.3.1.12
cit +atp + coa->oaa+acoA+adp	ACLY	TCA cycle	2.3.3.8
oaa +acoA -> cit + coa	CS	TCA cycle	2.3.3.1, 2.3.3.3, 2.3.3.16
cit \rightleftharpoons cis-aco	ACO	TCA cycle	4.2.1.3
cis-aco \rightleftharpoons icit	ACO	TCA cycle	4.2.1.4
nad + icit -> nadh + akG	IDH1	TCA cycle	1.1.1.41, 1.1.1.286
icit + nadp \rightleftharpoons akG + nadph	IDH2	TCA cycle	1.1.1.42
succ + atp + nadh \rightleftharpoons akG + nad + adp	SUCL+OGDH	TCA cycle	6.2.1.4, 1.2.4.2, 1.8.1.4, 2.3.1.61
succ + gtp + nadh \rightleftharpoons akG + nad + gdp	SUCL+OGDH	TCA cycle	6.2.1.5, 1.2.4.2, 1.8.1.4, 2.3.1.61
fum \rightleftharpoons mal-L	FH	TCA cycle	4.2.1.2
nad + mal-L \rightleftharpoons nadh + oaa	MDH	TCA cycle	1.1.1.37, 1.1.1.299
succ +1.5*adp \rightleftharpoons fum+1.5*atp	SDH	TCA cycle	1.3.5.1, 1.3.5.4
mal-L + nad \rightleftharpoons pyr + nadh	ME1	TCA cycle	1.1.1.38
mal-L + nadp \rightleftharpoons pyr + nadph	ME2	TCA cycle	1.1.1.40
oaa + gtp \rightleftharpoons pep + gdp	PCK	TCA cycle	4.1.1.32
pyr + atp -> oaa + adp	PC	TCA cycle	6.4.1.1
gln \rightleftharpoons glu	GLS	TCA cycle	3.5.1.2
glu + nad \rightleftharpoons akG + nadh	GLUD	TCA cycle	1.4.1.3
glu + nadp \rightleftharpoons akG + nadph	GLUD	TCA cycle	1.4.1.3
2.5* adp + nadh -> 2.5*atp + nad	HA	Oxidative Phosphorylation	3.6.3.14,1.6.5.3, 1.6.5.11,1.6.5.3
2*gsh -> gssg	GPX	Oxidative stress	1.11.1.9
gssg + nadph -> 2*gsh + nadp	GSR	Oxidative stress	1.8.1.7

ala+akg<=>pyr+glu	ALT	Nitrogen	2.6.1.2
asp+atp+gln=>amp+asn+glu	ASNS	Nitrogen	6.3.5.4
asp=>oaa	DDO	Nitrogen	1.4.3.1
asp+akg<=>glu+oaa	AAT	Nitrogen	2.6.1.1
asp+imp+gtp->succ-amp+gdp	ADSS	Nitrogen	6.3.4.4
asp+succ-amp->fum+amp	ADSL	Nitrogen	4.3.2.2
atp +citru+ asp->amp +argsucc	ASS	Nitrogen	6.3.4.5
argsucc<=>fum+arg	ASL	Nitrogen	4.3.2.1
arg->orn+urea	ARG1	Nitrogen	3.5.3.1
2*arg + 3*nadph -> 2*citru + 3*nadp	NOS	Nitrogen	1.14.13.39_____
orn->citru	OTC	Nitrogen	2.1.3.3
orn -> put	ODC	Nitrogen	4.1.1.17
put+nad->gaba+nadh	AOC+ALDH7a1	Nitrogen	1.2.1.3,1.4.3.22
put+nadp->gaba+nadph	AOC+ALDH7a1	Nitrogen	1.2.1.3,1.4.3.22
glu->gaba	GAD	Nitrogen	4.1.1.15
gaba+akg+nad<=>glu+nadh+s ucc	ABAT+ALDH5A1	Nitrogen	1.2.1.24 , 2.6.1.19
gaba+akg+nadp<=>glu+nadph +succ	ABAT+ALDH5A1	Nitrogen	1.2.1.24 , 2.6.1.19
<=>glu	EX_glu	Exchange	
<=>g1p	EX_g1p	Exchange	
<=> gln	EX_gln	Exchange	
nadp ->	EX_nadp	Exchange	
-> nadph	EX_nadph	Exchange	
-> adp	EX_adp	Exchange	
<=>imp	ex_imp	Exchange	
<=> amp	EX_amp	Exchange	
nad ->	EX_nad	Exchange	
-> nadh	EX_nadh	Exchange	
<=> gdp	EX_gdp	Exchange	
<=> gtp	EX_gtp	Exchange	
<=> gsh	EX_gsh	Exchange	
atp ->	EX_atp	Exchange	
lac-L <=>	EX_lac	Exchange	
<=>acoA	EX_coA	Exchange	
urea<=>	EX_urea	Exchange	

<=>asp	EX_asp	Exchange	
orn<=>	EX_orn	Exchange	
ala<=>	EX_ala	Exchange	
arg<=>	EX_arg	Exchange	
gaba<=>	EX_gaba	Exchange	
put<=>	EX_put	Exchange	

Tabla S2 Cambio en K y Genes

r_id x	ENZYME	met_lfc	met_pval	pathway	ge_lfc	ge_pval	SYMBOL
1	5.4.2.4	0.00	1.00	Glycolysis	0.14	0.02	Bpgm
1	5.4.2.11	0.00	1.00	Glycolysis	0.14	0.02	Bpgm
1	5.4.2.11	0.00	1.00	Glycolysis	0.72	0.00	Pgam1
1	5.4.2.11	0.00	1.00	Glycolysis	-0.12	0.12	Pgam2
2	5.4.2.4	0.62	0.37	Glycolysis	0.14	0.02	Bpgm
2	5.4.2.11	0.62	0.37	Glycolysis	0.14	0.02	Bpgm
2	5.4.2.11	0.62	0.37	Glycolysis	0.72	0.00	Pgam1
2	5.4.2.11	0.62	0.37	Glycolysis	-0.12	0.12	Pgam2
4	4.2.1.11	0.04	1.00	Glycolysis	0.05	0.63	Eno1
4	4.2.1.11	0.04	1.00	Glycolysis	-0.08	0.39	Eno3
4	4.2.1.11	0.04	1.00	Glycolysis	0.17	0.04	Eno4
5	4.2.1.11	-0.18	1.00	Glycolysis	0.05	0.63	Eno1
5	4.2.1.11	-0.18	1.00	Glycolysis	-0.08	0.39	Eno3
5	4.2.1.11	-0.18	1.00	Glycolysis	0.17	0.04	Eno4
6	4.1.2.13	-3.20	0.00	Glycolysis	0.01	0.93	Aldoa
6	4.1.2.13	-3.20	0.00	Glycolysis	-0.22	0.01	Aldoc
6	4.1.2.13	-3.20	0.00	Glycolysis	-0.05	0.20	Aldob
6	4.1.2.13	-3.20	0.00	Glycolysis	-0.23	0.00	Aldoart2
7	4.1.2.13	-1.00	0.80	Glycolysis	0.01	0.93	Aldoa
7	4.1.2.13	-1.00	0.80	Glycolysis	-0.22	0.01	Aldoc
7	4.1.2.13	-1.00	0.80	Glycolysis	-0.05	0.20	Aldob
7	4.1.2.13	-1.00	0.80	Glycolysis	-0.23	0.00	Aldoart2
8	3.1.3.11	-3.20	0.00	Glycolysis	-0.45	0.00	Fbp1
8	3.1.3.11	-3.20	0.00	Glycolysis	-0.09	0.40	Fbp2
9	1.2.1.12	-0.18	1.00	Glycolysis	0.01	0.89	Gapdhs
9	1.2.1.12	-0.18	1.00	Glycolysis	-0.06	0.39	LOC303448
10	1.2.1.12	0.24	0.37	Glycolysis	0.01	0.89	Gapdhs
10	1.2.1.12	0.24	0.37	Glycolysis	-0.06	0.39	LOC303448
11	2.7.1.11	0.01	1.00	Glycolysis	1.05	0.00	Pfkm
11	2.7.1.11	0.01	1.00	Glycolysis	0.21	0.00	Pfkl
11	2.7.1.11	0.01	1.00	Glycolysis	0.89	0.00	Pfkp
12	5.3.1.9	0.47	0.51	Glycolysis	0.51	0.00	Gpi
13	5.3.1.9	-0.01	1.00	Glycolysis	0.51	0.00	Gpi
14	3.6.1.7	0.00	1.00	Glycolysis	0.33	0.00	Acyp1
14	3.6.1.7	0.00	1.00	Glycolysis	0.02	0.77	Acyp2

15	2.7.2.3	0.36	0.74	Glycolysis	0.27	0.00	Pgk1
15	2.7.2.3	0.36	0.74	Glycolysis	-0.13	0.09	Pgk2
16	2.7.2.3	0.34	0.06	Glycolysis	0.27	0.00	Pgk1
16	2.7.2.3	0.34	0.06	Glycolysis	-0.13	0.09	Pgk2
17	5.4.2.11	0.04	1.00	Glycolysis	0.14	0.02	Bpgm
17	5.4.2.11	0.04	1.00	Glycolysis	0.72	0.00	Pgam1
17	5.4.2.11	0.04	1.00	Glycolysis	-0.12	0.12	Pgam2
18	5.4.2.11	0.34	0.81	Glycolysis	0.14	0.02	Bpgm
18	5.4.2.11	0.34	0.81	Glycolysis	0.72	0.00	Pgam1
18	5.4.2.11	0.34	0.81	Glycolysis	-0.12	0.12	Pgam2
19	5.4.2.2	0.59	0.57	Glycolysis	0.14	0.03	Pgm1
20	5.4.2.2	0.47	0.51	Glycolysis	0.14	0.03	Pgm1
21	2.7.1.40	0.16	1.00	Glycolysis	-0.39	0.00	Pkm
21	2.7.1.40	0.16	1.00	Glycolysis	-1.07	0.00	Pklr
22	5.3.1.1	-0.05	1.00	Glycolysis	0.13	0.11	LOC500959
23	5.3.1.1	-0.95	0.80	Glycolysis	0.13	0.11	LOC500959
24	1.1.1.27	1.13	0.03	Glycolysis	0.48	0.00	Ldhb
24	1.1.1.27	1.13	0.03	Glycolysis	-0.11	0.12	Ldhc
24	1.1.1.27	1.13	0.03	Glycolysis	-0.16	0.05	Ldhal6b
25	1.1.1.27	2.04	0.05	Glycolysis	0.48	0.00	Ldhb
25	1.1.1.27	2.04	0.05	Glycolysis	-0.11	0.12	Ldhc
25	1.1.1.27	2.04	0.05	Glycolysis	-0.16	0.05	Ldhal6b
26	1.1.1.49	1.51	0.08	Pentose phosphate	2.95	0.00	G6pd
26	1.1.1.363	1.51	0.08	Pentose phosphate	2.95	0.00	G6pd
27	3.1.1.31	2.12	0.06	Pentose phosphate	-0.15	0.03	Pgls
27	3.1.1.31	2.12	0.06	Pentose phosphate	-0.39	0.00	H6pd
27	1.1.1.44	2.12	0.06	Pentose phosphate	2.13	0.00	Pgd
28	3.1.1.31	0.34	1.00	Pentose phosphate	-0.15	0.03	Pgls
28	3.1.1.31	0.34	1.00	Pentose phosphate	-0.39	0.00	H6pd
28	1.1.1.44	0.34	1.00	Pentose phosphate	2.13	0.00	Pgd
29	5.4.2.2	0.58	0.18	Pentose phosphate	0.14	0.03	Pgm1
30	5.4.2.2	0.69	0.13	Pentose phosphate	0.14	0.03	Pgm1

31	2.7.6.1	0.71	0.36	Pentose phosphate	0.49	0.00	Prps2
31	2.7.6.1	0.71	0.36	Pentose phosphate	0.19	0.03	Prps1
31	2.7.6.1	0.71	0.36	Pentose phosphate	-0.12	0.10	LOC314140
31	2.7.6.1	0.71	0.36	Pentose phosphate	-0.18	0.08	Prps1I1
32	2.7.6.1	0.73	0.32	Pentose phosphate	0.49	0.00	Prps2
32	2.7.6.1	0.73	0.32	Pentose phosphate	0.19	0.03	Prps1
32	2.7.6.1	0.73	0.32	Pentose phosphate	-0.12	0.10	LOC314140
32	2.7.6.1	0.73	0.32	Pentose phosphate	-0.18	0.08	Prps1I1
33	5.1.3.1	-0.31	1.00	Pentose phosphate	0.31	0.00	Rpe
34	5.1.3.1	0.15	1.00	Pentose phosphate	0.31	0.00	Rpe
35	5.3.1.6	0.69	0.13	Pentose phosphate	0.03	0.73	Rpia
36	5.3.1.6	-0.31	1.00	Pentose phosphate	0.03	0.73	Rpia
37	2.2.1.2	-0.12	1.00	Pentose phosphate	1.85	0.00	Taldo1
38	2.2.1.2	-0.01	1.00	Pentose phosphate	1.85	0.00	Taldo1
39	2.2.1.1	0.84	0.65	Pentose phosphate	1.59	0.00	Tkt
39	2.2.1.1	0.84	0.65	Pentose phosphate	-0.25	0.01	TktI2
39	2.2.1.1	0.84	0.65	Pentose phosphate	-0.09	0.57	TktI1
40	2.2.1.1	-0.12	1.00	Pentose phosphate	1.59	0.00	Tkt
40	2.2.1.1	-0.12	1.00	Pentose phosphate	-0.25	0.01	TktI2
40	2.2.1.1	-0.12	1.00	Pentose phosphate	-0.09	0.57	TktI1
41	2.2.1.1	0.15	1.00	Pentose phosphate	1.59	0.00	Tkt
41	2.2.1.1	0.15	1.00	Pentose phosphate	-0.25	0.01	TktI2

41	2.2.1.1	0.15	1.00	Pentose phosphate	-0.09	0.57	Tktl1
42	2.2.1.1	-0.97	0.71	Pentose phosphate	1.59	0.00	Tkt
42	2.2.1.1	-0.97	0.71	Pentose phosphate	-0.25	0.01	Tktl2
42	2.2.1.1	-0.97	0.71	Pentose phosphate	-0.09	0.57	Tktl1
43	1.2.4.1	3.25	0.02	TCA cycle	-0.19	0.01	Pdha2
43	2.3.1.12	3.25	0.02	TCA cycle	0.34	0.00	Dlat
43	1.2.4.1	3.25	0.02	TCA cycle	0.51	0.00	Pdha1
43	1.8.1.4	3.25	0.02	TCA cycle	0.25	0.00	Dld
43	1.2.4.1	3.25	0.02	TCA cycle	-0.44	0.00	Pdhb
44	2.3.3.8	-0.58	0.82	TCA cycle	-0.64	0.00	Acly
45	2.3.3.1	2.20	0.01	TCA cycle	0.05	0.51	Cs
46	4.2.1.3	-1.27	0.11	TCA cycle	0.73	0.00	Aco2
46	4.2.1.3	-1.27	0.11	TCA cycle	-0.05	0.34	Aco1
47	4.2.1.3	-4.26	0.02	TCA cycle	0.73	0.00	Aco2
47	4.2.1.3	-4.26	0.02	TCA cycle	-0.05	0.34	Aco1
50	1.1.1.41	1.13	0.26	TCA cycle	1.08	0.00	ldh3a
50	1.1.1.41	1.13	0.26	TCA cycle	0.30	0.00	ldh3B
50	1.1.1.41	1.13	0.26	TCA cycle	0.10	0.05	ldh3g
51	1.1.1.42	1.39	0.18	TCA cycle	0.24	0.00	ldh1
51	1.1.1.42	1.39	0.18	TCA cycle	0.71	0.00	ldh2
52	1.1.1.42	0.65	0.07	TCA cycle	0.24	0.00	ldh1
52	1.1.1.42	0.65	0.07	TCA cycle	0.71	0.00	ldh2
53	6.2.1.4	0.91	0.10	TCA cycle	-0.15	0.05	Suc1g1
53	1.2.4.2	0.91	0.10	TCA cycle	0.13	0.10	Ogdh1
53	1.2.4.2	0.91	0.10	TCA cycle	0.36	0.00	Ogdh
53	1.8.1.4	0.91	0.10	TCA cycle	0.25	0.00	Dld
53	1.2.4.2	0.91	0.10	TCA cycle	-0.85	0.00	Dhtkd1
53	2.3.1.61	0.91	0.10	TCA cycle	0.37	0.00	Dlst
53	6.2.1.4	0.91	0.10	TCA cycle	-0.15	0.04	Suc1g2
53	6.2.1.4	0.91	0.10	TCA cycle	-0.15	0.08	Suc1a2
54	6.2.1.4	1.12	0.01	TCA cycle	-0.15	0.05	Suc1g1
54	1.2.4.2	1.12	0.01	TCA cycle	0.13	0.10	Ogdh1
54	1.2.4.2	1.12	0.01	TCA cycle	0.36	0.00	Ogdh
54	1.8.1.4	1.12	0.01	TCA cycle	0.25	0.00	Dld
54	1.2.4.2	1.12	0.01	TCA cycle	-0.85	0.00	Dhtkd1
54	2.3.1.61	1.12	0.01	TCA cycle	0.37	0.00	Dlst
54	6.2.1.4	1.12	0.01	TCA cycle	-0.15	0.04	Suc1g2

54	6.2.1.4	1.12	0.01	TCA cycle	-0.15	0.08	Sucla2
55	6.2.1.5	1.30	0.04	TCA cycle	-0.15	0.05	Sucgl1
55	1.2.4.2	1.30	0.04	TCA cycle	0.13	0.10	Ogdhl
55	1.2.4.2	1.30	0.04	TCA cycle	0.36	0.00	Ogdh
55	1.8.1.4	1.30	0.04	TCA cycle	0.25	0.00	Dld
55	1.2.4.2	1.30	0.04	TCA cycle	-0.85	0.00	Dhtkd1
55	2.3.1.61	1.30	0.04	TCA cycle	0.37	0.00	Dlst
55	6.2.1.5	1.30	0.04	TCA cycle	-0.15	0.04	Sucgl2
55	6.2.1.5	1.30	0.04	TCA cycle	-0.15	0.08	Sucla2
56	6.2.1.5	1.42	0.02	TCA cycle	-0.15	0.05	Sucgl1
56	1.2.4.2	1.42	0.02	TCA cycle	0.13	0.10	Ogdhl
56	1.2.4.2	1.42	0.02	TCA cycle	0.36	0.00	Ogdh
56	1.8.1.4	1.42	0.02	TCA cycle	0.25	0.00	Dld
56	1.2.4.2	1.42	0.02	TCA cycle	-0.85	0.00	Dhtkd1
56	2.3.1.61	1.42	0.02	TCA cycle	0.37	0.00	Dlst
56	6.2.1.5	1.42	0.02	TCA cycle	-0.15	0.04	Sucgl2
56	6.2.1.5	1.42	0.02	TCA cycle	-0.15	0.08	Sucla2
57	4.2.1.2	0.91	0.03	TCA cycle	-0.05	0.43	Fh
58	4.2.1.2	0.85	0.11	TCA cycle	-0.05	0.43	Fh
59	1.1.1.37	1.63	0.05	TCA cycle	0.04	0.50	Mdh2
59	1.1.1.37	1.63	0.05	TCA cycle	0.16	0.01	Mdh1
60	1.1.1.37	0.24	0.37	TCA cycle	0.04	0.50	Mdh2
60	1.1.1.37	0.24	0.37	TCA cycle	0.16	0.01	Mdh1
61	1.3.5.1	1.17	0.01	TCA cycle	0.13	0.01	Sdha
61	1.3.5.1	1.17	0.01	TCA cycle	-0.10	0.05	Sdhb
62	1.3.5.1	0.94	0.21	TCA cycle	0.13	0.01	Sdha
62	1.3.5.1	0.94	0.21	TCA cycle	-0.10	0.05	Sdhb
63	1.1.1.38	1.63	0.05	TCA cycle	1.34	0.00	Me2
64	1.1.1.38	2.04	0.05	TCA cycle	1.34	0.00	Me2
65	1.1.1.40	1.90	0.03	TCA cycle	1.58	0.00	Me1
65	1.1.1.40	1.90	0.03	TCA cycle	-1.78	0.00	Me3
66	1.1.1.40	2.45	0.04	TCA cycle	1.58	0.00	Me1
66	1.1.1.40	2.45	0.04	TCA cycle	-1.78	0.00	Me3
67	4.1.1.32	0.41	0.37	TCA cycle	-0.95	0.00	Pck1
67	4.1.1.32	0.41	0.37	TCA cycle	0.51	0.00	Pck2
68	4.1.1.32	0.47	0.57	TCA cycle	-0.95	0.00	Pck1
68	4.1.1.32	0.47	0.57	TCA cycle	0.51	0.00	Pck2
69	6.4.1.1	1.82	0.07	TCA cycle	-0.22	0.01	Pc
70	3.5.1.2	0.00	1.00	TCA cycle	0.59	0.00	Gls
70	3.5.1.2	0.00	1.00	TCA cycle	-0.99	0.00	Gls2

71	3.5.1.2	-0.34	0.06	TCA cycle	0.59	0.00	Gls
71	3.5.1.2	-0.34	0.06	TCA cycle	-0.99	0.00	Gls2
72	1.4.1.3	0.43	0.16	TCA cycle	-0.32	0.00	Glud1
73	1.4.1.3	0.24	0.37	TCA cycle	-0.32	0.00	Glud1
74	1.4.1.3	0.70	0.07	TCA cycle	-0.32	0.00	Glud1
75	1.4.1.3	0.65	0.07	TCA cycle	-0.32	0.00	Glud1
76	3.6.3.14	1.09	0.05	Oxidative Phosphorylation	0.25	0.00	Atp5b
76	1.6.5.3	1.09	0.05	Oxidative Phosphorylation	0.20	0.00	Ndufv1
76	1.6.5.3	1.09	0.05	Oxidative Phosphorylation	-0.10	0.13	Ndufv2
76	1.6.5.3	1.09	0.05	Oxidative Phosphorylation	0.04	0.55	Ndufs2
76	1.6.5.3	1.09	0.05	Oxidative Phosphorylation	-0.06	0.28	Ndufs8
76	1.6.5.3	1.09	0.05	Oxidative Phosphorylation	0.32	0.00	Ndufs1
76	1.6.5.3	1.09	0.05	Oxidative Phosphorylation	0.02	0.78	Ndufs3
76	1.6.5.3	1.09	0.05	Oxidative Phosphorylation	-0.41	0.00	Ndufs7
76	3.6.3.14	1.09	0.05	Oxidative Phosphorylation	0.22	0.01	Atp6v1a
76	3.6.3.14	1.09	0.05	Oxidative Phosphorylation	-0.89	0.00	Ddx58
76	3.6.3.14	1.09	0.05	Oxidative Phosphorylation	-0.30	0.00	Dhx58
77	1.11.1.9	-1.18	0.02	Oxidative stress	-0.75	0.00	Gpx1
77	1.11.1.9	-1.18	0.02	Oxidative stress	1.26	0.00	Gpx3
77	1.11.1.9	-1.18	0.02	Oxidative stress	-0.32	0.00	Gpx6
77	1.11.1.9	-1.18	0.02	Oxidative stress	1.60	0.00	Gpx2
77	1.11.1.9	-1.18	0.02	Oxidative stress	0.02	0.91	Gpx5
77	1.11.1.9	-1.18	0.02	Oxidative stress	0.07	0.39	Gpx7
77	1.11.1.9	-1.18	0.02	Oxidative stress	0.11	0.15	Gpx8
78	1.8.1.7	0.61	0.01	Oxidative stress	1.95	0.00	Gsr
79	2.6.1.2	1.05	0.02	Nitrogen	-0.69	0.00	Gpt
79	2.6.1.2	1.05	0.02	Nitrogen	-0.03	0.83	Gpt2
80	2.6.1.2	1.46	0.10	Nitrogen	-0.69	0.00	Gpt
80	2.6.1.2	1.46	0.10	Nitrogen	-0.03	0.83	Gpt2
81	6.3.5.4	0.63	0.37	Nitrogen	0.88	0.00	Asns
82	1.4.3.1	0.61	0.16	Nitrogen	-0.46	0.00	Ddo

83	2.6.1.1	0.61	0.16	Nitrogen	-0.73	0.00	Got1
83	2.6.1.1	0.61	0.16	Nitrogen	-0.05	0.29	Got2
84	2.6.1.1	-0.34	0.06	Nitrogen	-0.73	0.00	Got1
84	2.6.1.1	-0.34	0.06	Nitrogen	-0.05	0.29	Got2
85	6.3.4.4	0.97	0.37	Nitrogen	0.13	0.03	Adss
85	6.3.4.4	0.97	0.37	Nitrogen	0.04	0.63	Adssl1
86	4.3.2.2	1.12	0.31	Nitrogen	0.53	0.00	Adsl
87	6.3.4.5	-2.06	0.03	Nitrogen	-0.36	0.00	Ass1
88	4.3.2.1	-0.06	1.00	Nitrogen	-0.45	0.00	Asl
89	4.3.2.1	0.09	1.00	Nitrogen	-0.45	0.00	Asl
90	3.5.3.1	-0.82	0.10	Nitrogen	0.42	0.00	Arg2
90	3.5.3.1	-0.82	0.10	Nitrogen	-0.19	0.02	Arg1
92	2.1.3.3	-0.54	0.13	Nitrogen	-0.38	0.00	Otc
93	4.1.1.17	-0.54	0.13	Nitrogen	0.99	0.00	Odc1
93	4.1.1.17	-0.54	0.13	Nitrogen	-0.13	0.04	RGD1562036
94	1.2.1.3	-2.36	0.05	Nitrogen	-0.10	0.02	Aldh2
94	1.2.1.3	-2.36	0.05	Nitrogen	0.17	0.00	Aldh3a2
94	1.4.3.22	-2.36	0.05	Nitrogen	-0.07	0.45	Aoc1
94	1.2.1.3	-2.36	0.05	Nitrogen	0.16	0.29	Aldh1b1
94	1.2.1.3	-2.36	0.05	Nitrogen	-0.09	0.25	Aldh7a1
94	1.2.1.3	-2.36	0.05	Nitrogen	-0.25	0.00	Aldh9a1
95	1.2.1.3	-2.09	0.05	Nitrogen	-0.10	0.02	Aldh2
95	1.2.1.3	-2.09	0.05	Nitrogen	0.17	0.00	Aldh3a2
95	1.4.3.22	-2.09	0.05	Nitrogen	-0.07	0.45	Aoc1
95	1.2.1.3	-2.09	0.05	Nitrogen	0.16	0.29	Aldh1b1
95	1.2.1.3	-2.09	0.05	Nitrogen	-0.09	0.25	Aldh7a1
95	1.2.1.3	-2.09	0.05	Nitrogen	-0.25	0.00	Aldh9a1
96	4.1.1.15	-0.34	0.06	Nitrogen	-0.07	0.24	Gad1
96	4.1.1.15	-0.34	0.06	Nitrogen	-0.13	0.06	Gad2
97	1.2.1.24	1.52	0.01	Nitrogen	0.02	0.80	Aldh5a1
97	2.6.1.19	1.52	0.01	Nitrogen	-0.51	0.00	Abat
98	1.2.1.24	0.55	0.16	Nitrogen	0.02	0.80	Aldh5a1
98	2.6.1.19	0.55	0.16	Nitrogen	-0.51	0.00	Abat
99	1.2.1.24	1.79	0.01	Nitrogen	0.02	0.80	Aldh5a1
99	2.6.1.19	1.79	0.01	Nitrogen	-0.51	0.00	Abat
100	1.2.1.24	0.96	0.08	Nitrogen	0.02	0.80	Aldh5a1
100	2.6.1.19	0.96	0.08	Nitrogen	-0.51	0.00	Abat

Dycone on tissue

May 21, 2018

1 Protocol

1.1 Installation

All of the analysis is performed in R. As such the first thing you will need is to install R. For installation instructions see <http://r-project.org>. In Ubuntu R can be installed via the Terminal using

```
sudo apt-get install r-base r-base-dev
```

Additionally some of the dependencies of dycone require development versions of some libraries for web security and scraping. In Ubuntu and Debian those can be installed via

```
sudo apt-get install libxml2-dev libcurl4-openssl-dev libssl-dev libgmp-dev
```

Most of the actual analysis is implemented in the dycone R package. It can be installed using devtools in the following manner. We will also install all optional dependencies so we can build this document. In a Terminal type R to start R, then use the following commands:

```
install.packages(c("devtools","ff","doMC","qplots"))
source("http://bioconductor.org/biocLite.R")
biocLite(c("Biobase", "IRanges", "AnnotationDbi", "affy", "frma", "genefilter",
          "GEOquery", "limma","pdInfoBuilder", "org.Rn.eg.db"))
devtools::install_github("cdiener/dycone", dependencies = TRUE)
```

This will install dycone and all additional dependencies. You will see a lot of messages from the compiler and the whole process might take a few minutes. After that the dycone library can be loaded with

```
library(dycone)
```

It is important to set the directory where we will work and have the figures

1.2 Statistical analysis to select pathways

1.2.1 Microarrays

The microarray data can be found in https://www.dropbox.com/s/a4e4ufbh8y7ccec/HCC_Microarrays.zip?dl=0. The first thing we did was to see which pathways are differentiated. Either from

metabolome or other transcriptome. For this we identified differentially expressed genes and pathways from each type of data metabolome or transcriptome. To analyse the transcriptome we used the annotation ENTREZ version 20 from brain array (http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/genomic_curated_CDF.asp) In order to use it in combination with oligo we had to build the package

```
library(pdInfoBuilder)
library(ff)
library(doMC)
registerDoMC(10)

download.file("http://mbni.org/customcdf/20.0.0/entrezg.download/ragene20st_Rn_ENTREZG_20.0.0.zip")
unzip("tmp.zip")
dir()
z <- cdf2table("ragene20st_Rn_ENTREZG.cdf")
seed <- new("GenericPDInfoPkgSeed", table=z, author = "me", email = "me@mine.org", species = "Rattus norvegicus")
makePdInfoPackage(seed)
install.packages("pd.ragene20st/", repos = NULL)
```

This package will map the probe sets so that "286921_at" map to the entrez gene identified by "286921"

```
In [1]: library(oligo,quietly=T, warn.conflicts=F)
sample_info<-read.table("MicroarraysRats/phenodata2.txt",header=T)
sample_info<-sample_info[!sample_info$FileName%in%c("13_2.CEL", "14_42.CEL", "32_42.CEL"),]
condition <- rep.int("HCC", length(sample_info$FileName))
condition[sample_info$Target == "Normal"] <- "normal"
condition <- factor(condition)
names(condition)<-sample_info$FileName
raw_data <-read.celfiles(paste("MicroarraysRats",sample_info$FileName,sep="/"),pkgname = "pd.ragene20st")
pData(raw_data)$cell_line <- sample_info$Target
pData(raw_data)$condition <- condition
```

Attaching package: ‘BiocGenerics’

The following objects are masked from ‘package:parallel’:

```
clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
clusterExport, clusterMap, parApply, parCapply, parLapply,
parLapplyLB, parRapply, parSapply, parSapplyLB
```

The following objects are masked from ‘package:stats’:

```
IQR, mad, xtabs
```

The following objects are masked from ‘package:base’:

```
anyDuplicated, append, as.data.frame, cbind, colnames, do.call,
```

duplicate, eval, evalq, Filter, Find, get, grep, grepl, intersect,
is.unsorted, lapply, lengths, Map, mapply, match, mget, order,
paste, pmax, pmax.int, pmin, pmin.int, Position, rank, rbind,
Reduce, rownames, sapply, setdiff, sort, table, tapply, union,
unique, unsplit

No methods found in "RSQLite" for requests: dbGetQuery
Welcome to oligoClasses version 1.34.0
Welcome to Bioconductor

Vignettes contain introductory material; view with
'browseVignettes()'. To cite Bioconductor, see
'citation("Biobase")', and for packages 'citation("pkgname")'.

Attaching package: 'S4Vectors'

The following objects are masked from 'package:base':

colMeans, colSums, expand.grid, rowMeans, rowSums

```
=====
Welcome to oligo version 1.36.1
=====
Loading required package: pd.ragene20st
Loading required package: RSQLite
Warning message:
“package ‘RSQLite’ was built under R version 3.3.2”Loading required package: DBI
Warning message:
“package ‘DBI’ was built under R version 3.3.2”Platform design info loaded.
```

Reading in : MicroarraysRats/10_2.CEL
Reading in : MicroarraysRats/10_4.CEL
Reading in : MicroarraysRats/11_2.CEL
Reading in : MicroarraysRats/11_4.CEL
Reading in : MicroarraysRats/12_2.CEL
Reading in : MicroarraysRats/13_22.CEL
Reading in : MicroarraysRats/13_4.CEL
Reading in : MicroarraysRats/14_4.CEL
Reading in : MicroarraysRats/21_2.CEL
Reading in : MicroarraysRats/21_4.CEL
Reading in : MicroarraysRats/22_2.CEL
Reading in : MicroarraysRats/22_4.CEL
Reading in : MicroarraysRats/30_2.CEL
Reading in : MicroarraysRats/30_4.CEL
Reading in : MicroarraysRats/31_2.CEL
Reading in : MicroarraysRats/31_4.CEL

```

Reading in : MicroaraysRats/32_2.CEL
Reading in : MicroaraysRats/32_4.CEL
Reading in : MicroaraysRats/3_2.CEL
Reading in : MicroaraysRats/33_2.CEL
Reading in : MicroaraysRats/33_42.CEL
Reading in : MicroaraysRats/33_4.CEL
Reading in : MicroaraysRats/3_4.CEL
Reading in : MicroaraysRats/4_2.CEL
Reading in : MicroaraysRats/4_4.CEL
Reading in : MicroaraysRats/9_2.CEL
Reading in : MicroaraysRats/9_4.CEL

```

In order to make the data comparable we will normalize the data using RMA

```
In [2]: eset <- rma(raw_data)
```

```
Background correcting... OK
Normalizing... OK
```

Calculating Expression

As a quick quality control we can see if the data can separate the two groups we are analyzing

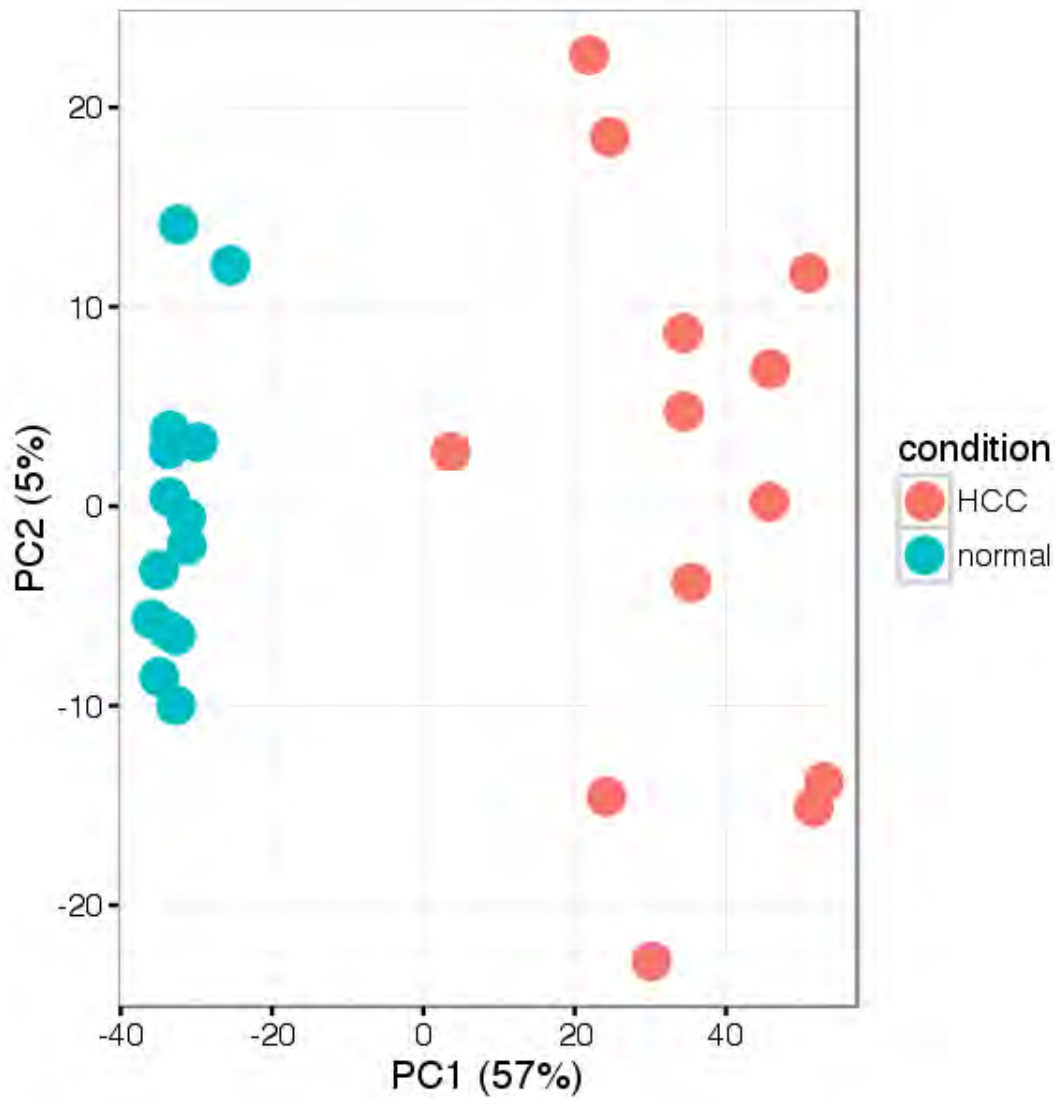
```
In [3]: library(ggplot2)
       pca <- prcomp(t(exprs(eset)))
       sum(pca$sdev[1:2])/sum(pca$sdev)
       summary(pca)
```

```
0.252816899405613
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	35.8409	10.72187	8.93751	8.48615	8.08800	7.53009	7.23690
Proportion of Variance	0.5747	0.05143	0.03574	0.03222	0.02927	0.02537	0.02343
Cumulative Proportion	0.5747	0.62616	0.66190	0.69412	0.72339	0.74876	0.77219
	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Standard deviation	6.63349	6.32259	6.11255	5.91181	5.64722	5.56982	5.44216
Proportion of Variance	0.01969	0.01789	0.01672	0.01564	0.01427	0.01388	0.01325
Cumulative Proportion	0.79188	0.80976	0.82648	0.84211	0.85638	0.87026	0.88351
	PC15	PC16	PC17	PC18	PC19	PC20	PC21
Standard deviation	5.31310	5.25791	5.05424	4.93184	4.73280	4.58984	4.48866
Proportion of Variance	0.01263	0.01237	0.01143	0.01088	0.01002	0.00943	0.00901
Cumulative Proportion	0.89614	0.90851	0.91994	0.93082	0.94084	0.95027	0.95928
	PC22	PC23	PC24	PC25	PC26	PC27	
Standard deviation	4.42764	4.26269	4.25124	4.22023	4.16472	5.811e-14	
Proportion of Variance	0.00877	0.00813	0.00809	0.00797	0.00776	0.000e+00	
Cumulative Proportion	0.96806	0.97619	0.98427	0.99224	1.00000	1.000e+00	


```
In [4]: options(repr.plot.width=5, repr.plot.height=5)
ggplot(data.frame(pca$x), aes(x=PC1, y=PC2, z=PC3)) + theme_bw() +
  geom_point(aes(col=condition), size=5) + labs(x = "PC1 (57%)", y="PC2 (5%)")
```



```
In [5]: gg_color_hue <- function(n) {
  hues = seq(15, 375, length = n + 1)
  hcl(h = hues, l = 65, c = 100)[1:n]
}
n = 2
colors_hex = gg_color_hue(n)
```

We can also do this by clustering the samples by their gene expression pattern in two clusters and seeing how well they correspond to their phenotype

```
In [6]: cl <- kmeans(t(exprs(eset)), 2)
normal_cluster <- cl$cluster[1] # first condition is normal
cl_err <- sum(cl$cluster[condition == "normal"] != normal_cluster)/length(cl$cluster)
cat(sprintf("Clustering error between normal/disease: %f%%\n", cl_err*100))
```

Clustering error between normal/disease: 0.000000%

1.2.2 Getting GMT for gsea

In order to run de gsea geneset were downloaded Small metabolite pathway and converted to a gmt format <http://smpdb.ca/downloads>, ids were also converted from human to rat using their E.C numbers

```
In [7]: library(AnnotationDbi)
library(org.Hs.eg.db)
library(org.Rn.eg.db)

genes<- read.csv("proteins.csv")
SMPDB<-read.csv("metabolites.csv")
SMPDB<-SMPDB[SMPDB$Pathway.Subject=="Metabolic",c("Pathway.Name" ,"Pathway.Subject" , "K

genes<-genes[genes$Pathway.Subject%in%SMPDB$Pathway.Subject,]
ko <- (merge(read.table("http://rest.kegg.jp/link/ko/rno"), read.table("http://rest.kegg
, by.x=2, by.y=2))
colnames(ko)<-c("ko", "rno", "hsa")
ko$rno<-gsub("rno:", "", ko$rno)
con_up<-read.table("http://rest.kegg.jp/conv/hsa/uniprot")
colnames(con_up)<-c("up", "hsa")
con_up$up<-gsub("up:", "", con_up$up)
#head(ko)
#head(con_up)
head(genes)
ensG<-merge(ko, con_up, by='hsa')
#head(ensG)
paths<-paste( unique(genes$Pathway.Name), unique(genes$SMPDB.ID), sep="\t")
mergin<-function(x){
  (paste(unique(ensG[ensG$up%in%genes[genes$SMPDB.ID==x, "Uniprot.ID"], "rno"]), collapse =
)}
gmt<-sapply(unique(genes$SMPDB.ID),mergin)
gmt<-cbind(paths,data.frame(gmt))
head(gmt)
```

SMPDB.ID	Pathway.Name	Pathway.Subject	Uniprot.ID	Protein.Name
SMP00055	Alanine Metabolism	Metabolic	P49588	Alanine-tRNA ligase, cytoplasmic
SMP00055	Alanine Metabolism	Metabolic	P24298	Alanine aminotransferase 1
SMP00055	Alanine Metabolism	Metabolic	P11498	Pyruvate carboxylase, mitochondrial
SMP00055	Alanine Metabolism	Metabolic	P21549	Serine-pyruvate aminotransferase
SMP00055	Alanine Metabolism	Metabolic	Q5JTZ9	Alanine-tRNA ligase, mitochondrial
SMP00067	Aspartate Metabolism	Metabolic	P80404	4-aminobutyrate aminotransferase, m

	paths	gmt
Alanine Metabolism SMP00055	301254 292023 24792 307759 81670 25104	
Aspartate Metabolism SMP00067	684425 289276 315150 116483 304919 81632	
Glutamate Metabolism SMP00072	361641 497840 81632 24379 24380 297417 3	
Glutathione Metabolism SMP00015	24422 686922 103690044 24421 171341 244	
Citric Acid Cycle SMP00057	170587 81654 298942 299201 24368 94173 2	
Alpha Linolenic Acid and Linoleic Acid Metabolism SMP00018	691810 296091 691905 24653 311341 69190	

```
In [8]: library(RColorBrewer)
library(gplots)
options(repr.plot.width=8, repr.plot.height=8)
exp<-exprs(eset)

col<-as.character(condition)
col[col=="HCC"]<-colors_hex[1]
col[col=="normal"]<-colors_hex[2]
library(gplots)
met_transc<-exprs(eset)[rownames(eset)%in%paste(ensG$rho[ensG$up%in%genes$Uniprot.ID], "_",
#write.csv((ensG[paste(ensG$rho[ensG$up%in%genes$Uniprot.ID], "_at", sep="")%in%rownames(e
```

Attaching package: ‘gplots’

The following object is masked from ‘package:IRanges’:

space

The following object is masked from ‘package:S4Vectors’:

space

The following object is masked from ‘package:stats’:

lowess

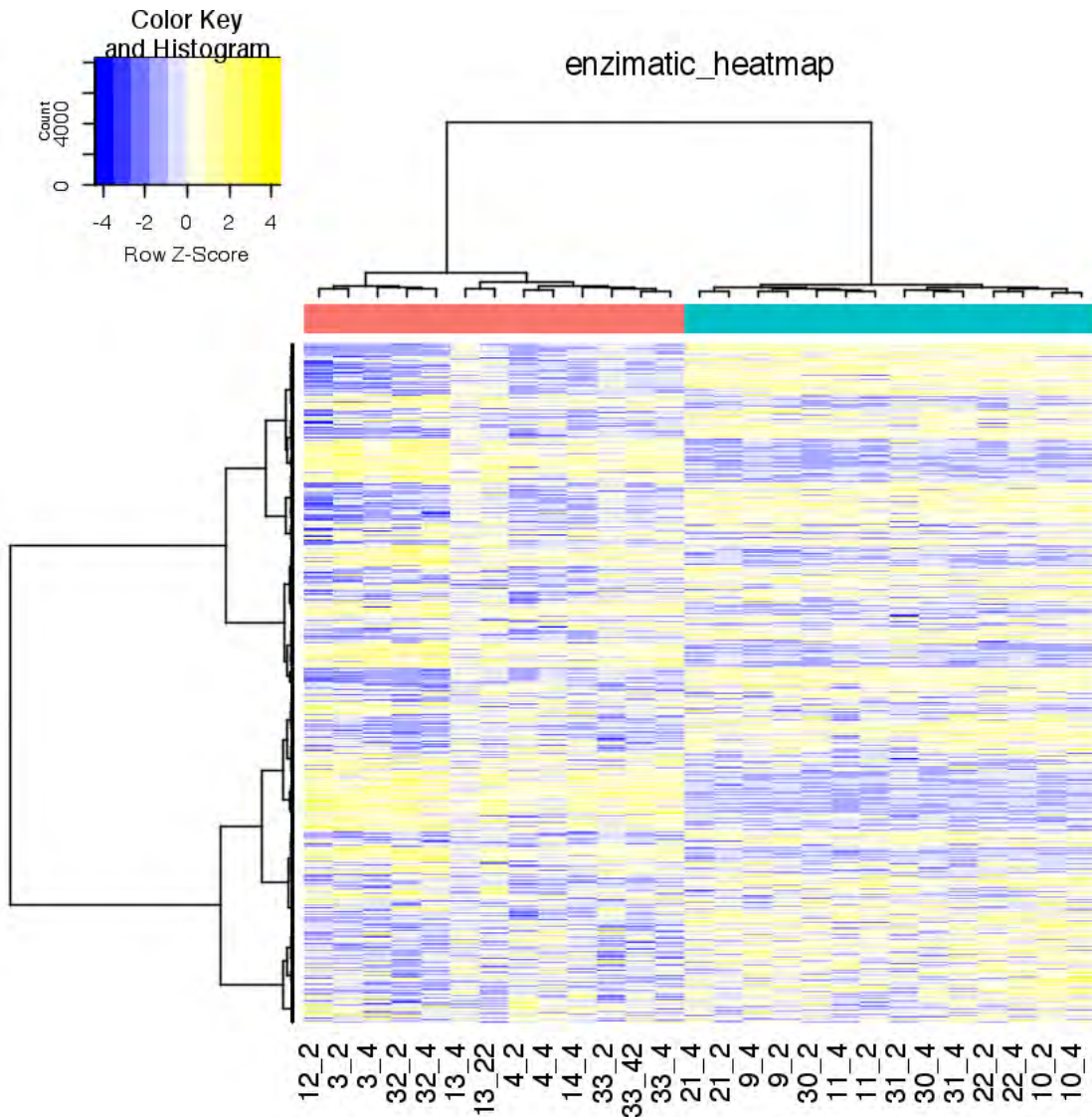
1.2.3 Transcriptional behaviour of metabolic genes

Once we have identified enzymes in our data set we can begin to explore the general behaviour

```

In [9]: colnames(met_transc)<-sub(".CEL","",colnames(met_transc))
#svg("metabolic_exp.svg")
options(repr.plot.width=7, repr.plot.height=7)
heatmap.2(met_transc
           ,ColSideColors=col,scale="row",denscol = F,hclustfun=function(x)hclust(x,method="
           ,trace='none',main="enzimatic_heatmap",col=colorRampPalette(c("blue","wh
           ,cexCol = 1.5,margins = c(5,1),labRow = NULL)
#dev.off()

```



```

In [10]: sum(pca$sdev[1:2])/sum(pca$sdev)
summary(pca)
pca <- prcomp(t(exprs(eset)[rownames(eset)%in%paste(ensG$rno[ensG$up%in%genes$Uniprot.I

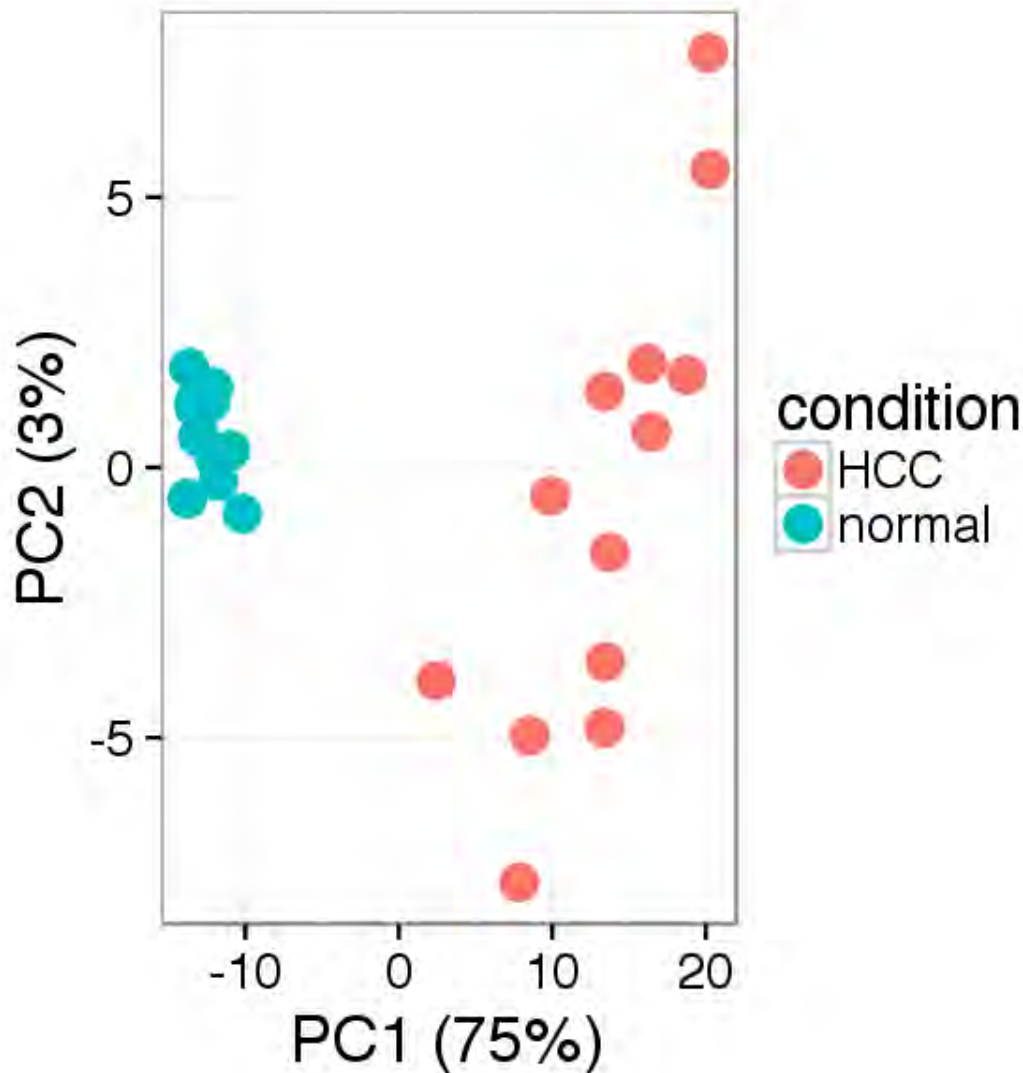
```

0.252816899405613

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	35.8409	10.72187	8.93751	8.48615	8.08800	7.53009	7.23690
Proportion of Variance	0.5747	0.05143	0.03574	0.03222	0.02927	0.02537	0.02343
Cumulative Proportion	0.5747	0.62616	0.66190	0.69412	0.72339	0.74876	0.77219
	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Standard deviation	6.63349	6.32259	6.11255	5.91181	5.64722	5.56982	5.44216
Proportion of Variance	0.01969	0.01789	0.01672	0.01564	0.01427	0.01388	0.01325
Cumulative Proportion	0.79188	0.80976	0.82648	0.84211	0.85638	0.87026	0.88351
	PC15	PC16	PC17	PC18	PC19	PC20	PC21
Standard deviation	5.31310	5.25791	5.05424	4.93184	4.73280	4.58984	4.48866
Proportion of Variance	0.01263	0.01237	0.01143	0.01088	0.01002	0.00943	0.00901
Cumulative Proportion	0.89614	0.90851	0.91994	0.93082	0.94084	0.95027	0.95928
	PC22	PC23	PC24	PC25	PC26	PC27	
Standard deviation	4.42764	4.26269	4.25124	4.22023	4.16472	5.811e-14	
Proportion of Variance	0.00877	0.00813	0.00809	0.00797	0.00776	0.000e+00	
Cumulative Proportion	0.96806	0.97619	0.98427	0.99224	1.00000	1.000e+00	

```
In [11]: options(repr.plot.width=5, repr.plot.height=5)
         #svg("pcaTranscMetb.svg")
         ggplot(data.frame(pca$x), aes(x=PC1, y=PC2)) + theme_bw(base_size=20) +
           geom_point(aes(col=condition), size=5) + labs(x = "PC1 (75%)", y="PC2 (3%)")
         #dev.off()
```



Differential expression of genes To see if we can use a t-test we need to see if the data behave in a normal-like.

In [12]: `options(repr.plot.width=7, repr.plot.height=7)`

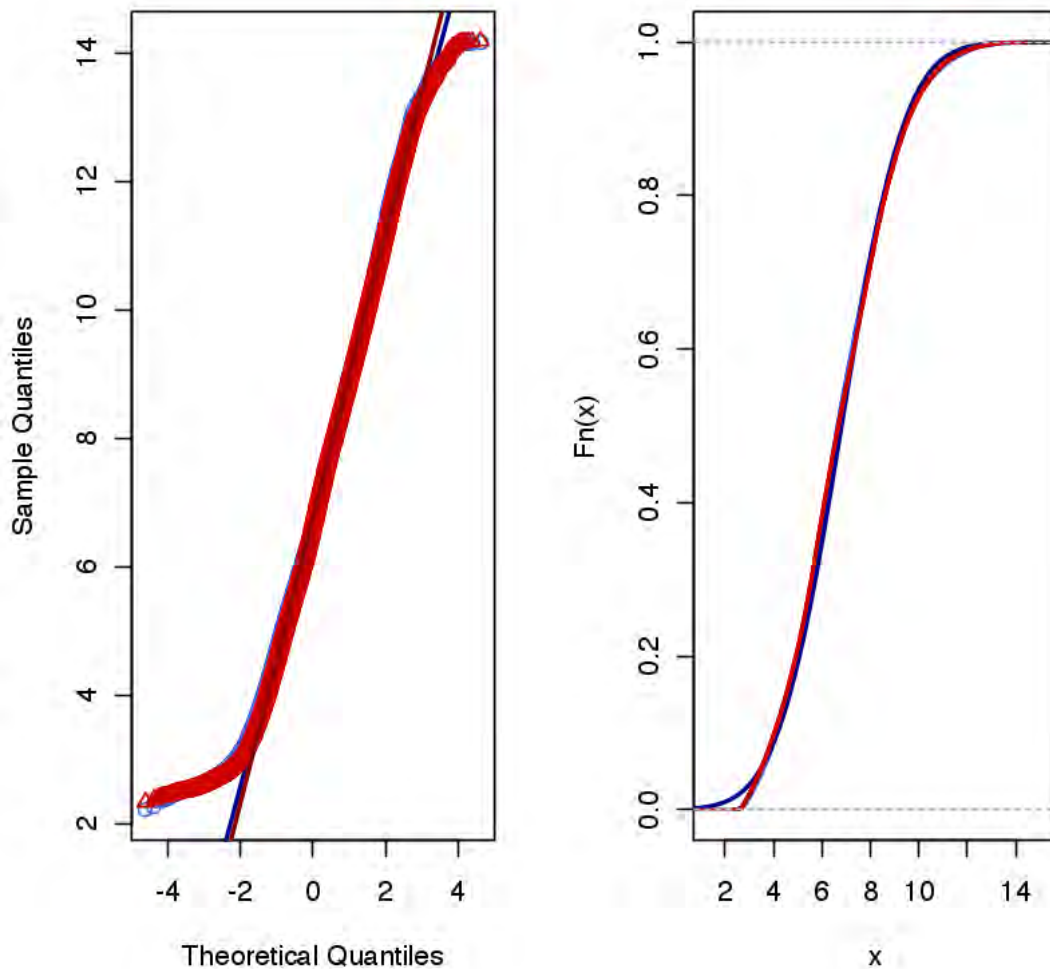
```
par(mfrow = 1:2)
qqnorm(exprs(eset[,condition=="normal"]), pch = 1, col = "royalblue", main = "")
qqline(exprs(eset[,condition=="normal"]), lwd = 2, col = "darkblue")
points(qqnorm(exprs(eset[,condition=="HCC"]), plot.it = F), pch = 2, col = "red3")
qqline(exprs(eset[,condition=="HCC"]), lwd = 2, col = "darkred")

x <- seq(0, 16, length.out = 256)
plot(ecdf(exprs(eset[,condition=="normal"])), pch = 1, col = "royalblue", main = "")
```

```

lines(x, pnorm(x, mean(exprs(eset[,condition=="normal"]), na.rm = T)
      , sd(exprs(eset[,condition=="normal"]), na.rm = T)), lwd = 2,col = "darkred")
plot(ecdf(exprs(eset[,condition=="HCC"])), pch = 2, add = T, col = "red3")
lines(x, pnorm(x, mean(exprs(eset[,condition=="HCC"]), na.rm = T), sd(exprs(eset[,condition=="HCC"], na.rm = T)), lwd = 2,col = "darkred")

```



Now that we know they behave similar to a normal distribution, we can use limma, so genes will be judged by a t-test where the sample variances are estimated using the empirical Bayes method from the limma package.

```

In [13]: library(limma)
design <- model.matrix(~ 0 + condition)
colnames(design) <- levels(condition)
fit <- lmFit(eset, design)

```

```

contrast.matrix <- makeContrasts(HCC - normal, levels=design)
cfit <- contrasts.fit(fit, contrast.matrix)
ebfit <- eBayes(cfit)

ma_lfcs <- topTable(ebfit, number=Inf)
ma_sig <- topTable(ebfit, p.value = 0.05, number=Inf)
ma_sig <- ma_sig[abs(ma_sig$logFC) > 1,]

save(ma_lfcs, file="gene_expression.Rd")
dim(ma_sig)[1]/dim(ma_lfcs)[1]
dim(ma_sig)[1]

```

Attaching package: ‘limma’

The following object is masked from ‘package:oligo’:

backgroundCorrect

The following object is masked from ‘package:BiocGenerics’:

plotMA

```

0.041931823787677
850

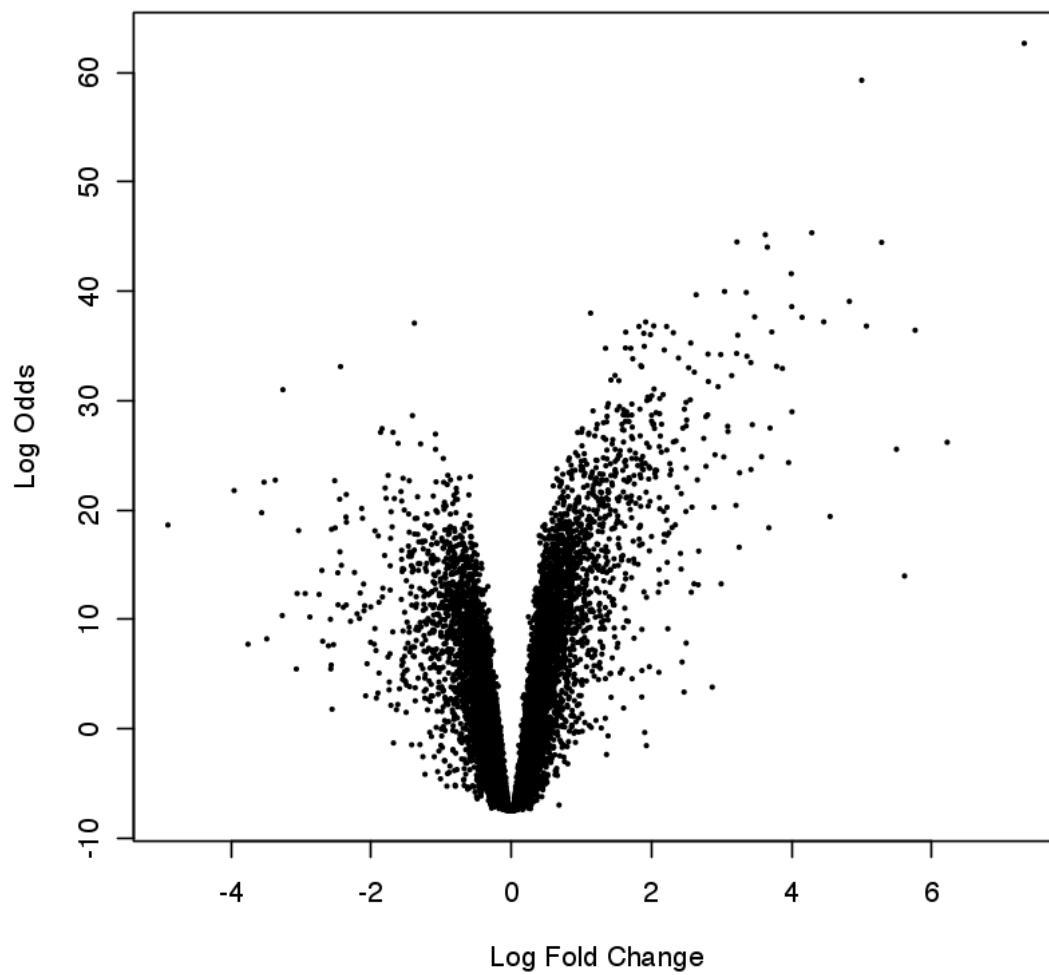
```

In [14]: `head(ma_lfcs)`

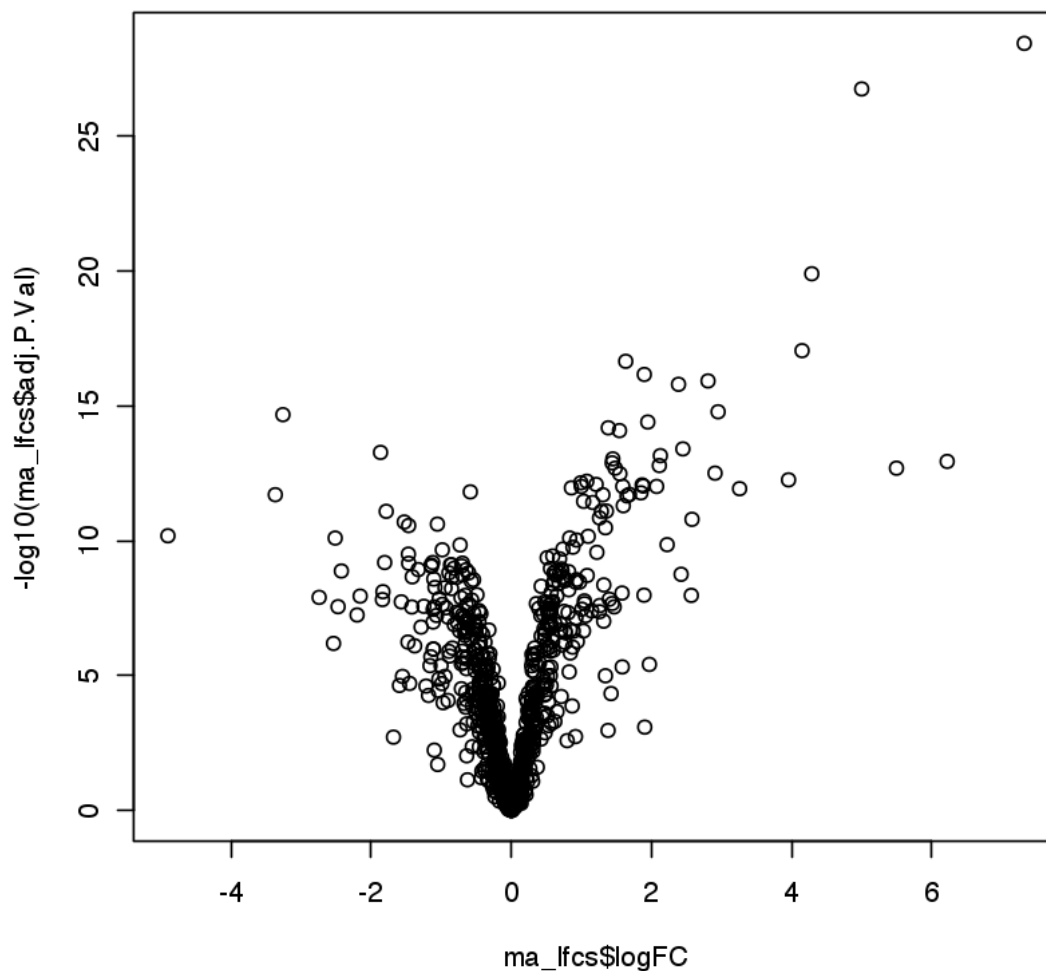
	logFC	AveExpr	t	P.Value	adj.P.Val	B
286921_at	7.321047	7.113584	65.12546	1.744396e-33	3.536064e-29	62.67885
24426_at	5.000867	9.725349	55.69383	1.715248e-31	1.738489e-27	59.30223
24314_at	4.288272	9.598571	31.81713	2.004249e-24	1.235635e-20	45.36074
56611_at	3.625186	8.996332	31.60090	2.438232e-24	1.235635e-20	45.17906
291946_at	3.220498	7.531939	30.82986	4.956296e-24	1.745671e-20	44.51949
100036765_at	5.285365	7.994760	30.78516	5.166999e-24	1.745671e-20	44.48068

We can also do a volcano plot from both, the complete data set and our metabolic genes.

In [15]: `volcanoplot(ebfit)`



```
In [16]: ma_lfcs<-ma_lfcs[rownames(met_transc),]  
plot(ma_lfcs$logFC,-log10(ma_lfcs$adj.P.Val))
```



1.3 Metabolome analysis

Most of the analysis was run on metaboanalsys (<http://www.metaboanalyst.ca/>) The log-fold changes for all combinations between HCC and liver cells can be obtained by

```
In [17]: library(tidyr)
library(ggplot2)

metab <- read.table("Metabolome/R_Ready_MetabolomeAnalysisData.csv", sep="\t", header=T)
comb <- expand.grid(a = 1:4, b = 5:8)
mlfc <- apply(metab[, 6:13], 1, function(m) log(m[comb$b], 2) - log(m[comb$a], 2))
colnames(mlfc) <- as.character(metab$Compound.name)
mlfc <- gather(data.frame(mlfc, check.names = F), name, lfc)
```

Attaching package: 'tidyr'

The following object is masked from 'package:S4Vectors':

expand

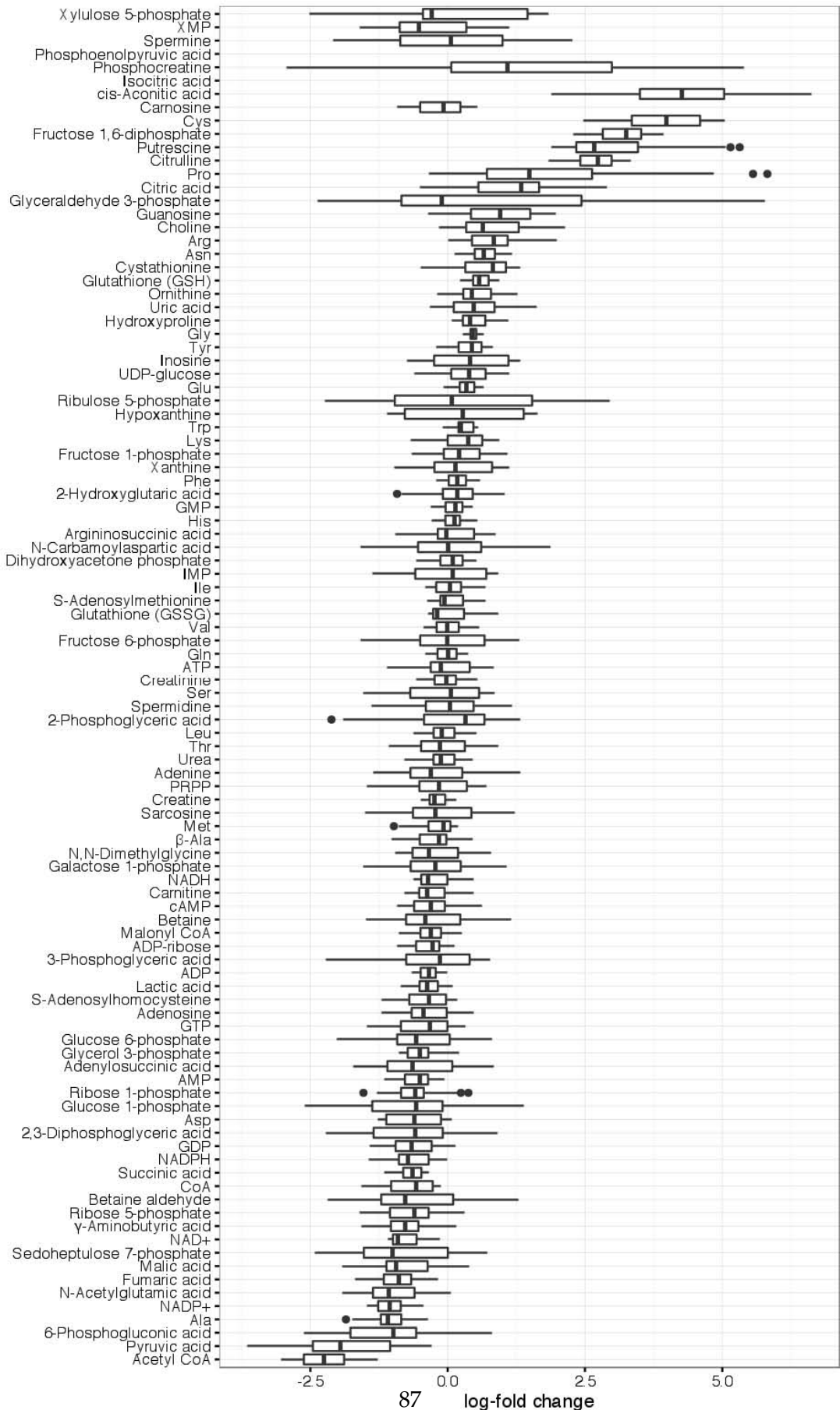
Warning message in data.row.names(row.names, row_si, i):

“some row.names duplicated: 2,3,4,6,7,8,10,11,12,14,15,16 --> row.names NOT used”

```
In [18]: #png("Metabolome/boxplot.png",height = 900)
options(repr.plot.width=8, repr.plot.height=13)
ggplot(mlfc, aes(x = reorder(name, lfc, function(x) mean(x)), y = lfc))+
  geom_boxplot(position=position_dodge(width=0.8)) + theme_bw() +
  xlab("") + ylab("log-fold change") + coord_flip()
#dev.off()
```

Warning message:

“Removed 69 rows containing non-finite values (stat_boxplot).”Warning message in grid.Call(L_text
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <ce>”Warning m
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <b3>”Warning m
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <ce>”Warning message in grid
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <b2>”Warning message in grid
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <ce>”Warning m
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <b3>”Warning m
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <ce>”Warning message in grid
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <b2>”Warning message in grid
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <ce>”Warning m
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <b3>”Warning m
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <ce>”Warning message in grid
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <b2>”Warning message in grid
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <ce>”Warning m
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <b3>”Warning m
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <ce>”Warning message in grid
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <b2>”Warning message in grid
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <ce>”Warning m
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <b3>”Warning m
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <ce>”Warning message in grid
“conversion failure on ' β -Ala' in 'mbscsToSbscs': dot substituted for <b2>”Warning message in grid
“conversion failure on ' γ -Aminobutyric acid' in 'mbscsToSbscs': dot substituted for <ce>”Warning m



```

In [19]: library(gplots)
options(repr.plot.width=8, repr.plot.height=10)

met<-(as.matrix(metab[,6:13]))
rownames(met)<-metab$Compound.name
colnames(met)<-sub('X','',colnames(met))
met<-met[!apply(met,1,function(x){sum(is.na(x))})>0,]
met<-log2(met)

In [20]: multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (
    is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                     layout.pos.col = matchidx$col))
    }
  }
}

```

```
In [21]: library(tidyr)
long_abund <- cbind(Compound=rownames(met),data.frame(met))%>% gather(sample, abundance)

cond<-gsub("X[0-9]*.", "", long_abund$sample, perl=T)
long_abund<-cbind(long_abund,cond)
#long_abund$sample<-gsub("[A-z]+", "", long_abund$sample, perl=T)
head(long_abund)
ylim1 = boxplot.stats(long_abund$abundance)$stats[c(1, 5)]
ylim1
library(ggplot2)
PBrawmet<-ggplot(long_abund, aes(y=abundance,x=sample,fill=cond))+
  geom_boxplot(position=position_dodge(width=0.8)) + theme_bw() +
  # scale_y_continuous(limits = ylim1+c(-.05,+.05))+
  xlab("") + ylab("abund")+coord_flip()

PDrawmet<-ggplot(long_abund, aes(abundance, col=cond,group=sample)) +geom_density()
```

Compound	sample	abundance	cond
NAD+	X9.N	9.608640	N
cAMP	X9.N	-0.746050	N
NADH	X9.N	3.474743	N
Xanthine	X9.N	7.623999	N
ADP-ribose	X9.N	5.017300	N
UDP-glucose	X9.N	8.132176	N

1. -1.91756155021496 2. 13.6063122627569

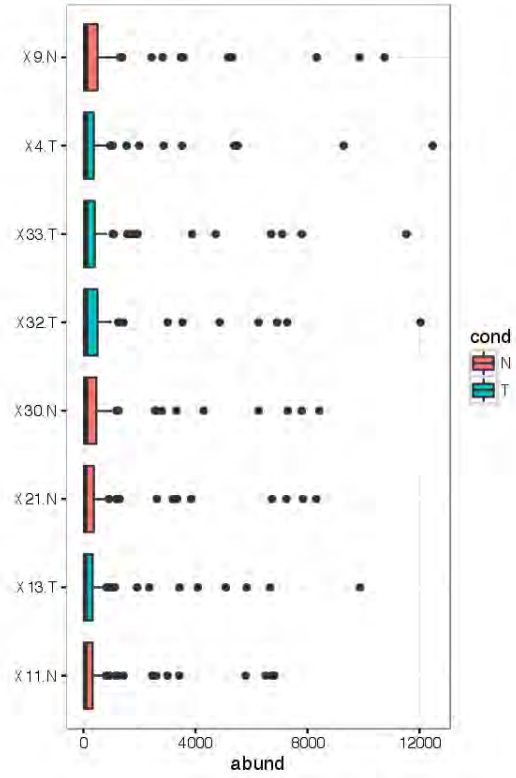
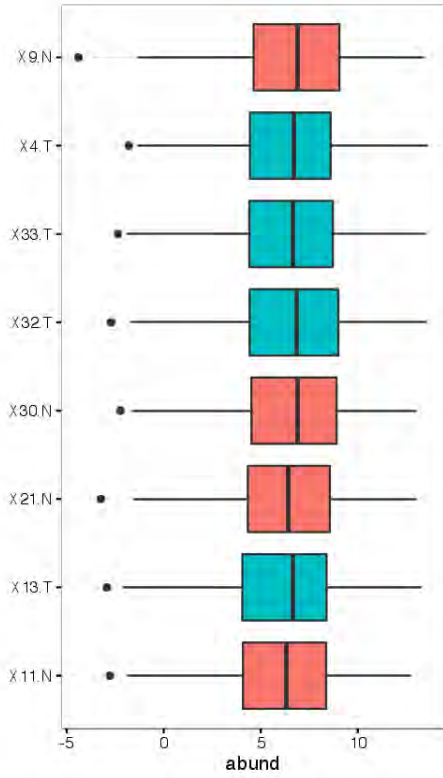
```
In [22]: long_abund <- cbind(Compound=metab$Compound,data.frame(metab)) %>% gather(sample, abundance)
cond<-gsub("X[0-9]*.", "", long_abund$sample, perl=T)
#long_abund$sample<-gsub("[A-z]+", "", long_abund$sample, perl=T)
long_abund<-cbind(long_abund,cond)
ylim1 = boxplot.stats(long_abund$abundance)$stats[c(1, 5)]
ylim1
PBnorm<-ggplot(long_abund, aes(y=abundance,x=sample,fill=cond))+
  geom_boxplot(position=position_dodge(width=0.8)) + theme_bw() +
  # scale_y_continuous(limits = ylim1+c(.05,+.05))+
  xlab("") + ylab("abund")+coord_flip()
PDnorm<-ggplot(long_abund, aes(abundance, col=cond,group=sample)) +geom_density()

1. 0.026925221 2. 952.3727874
```

```
In [23]: options(repr.plot.width=10,repr.plot.height=7)
multiplot(PBrawmet,PBnorm, cols=2)
```

Warning message:

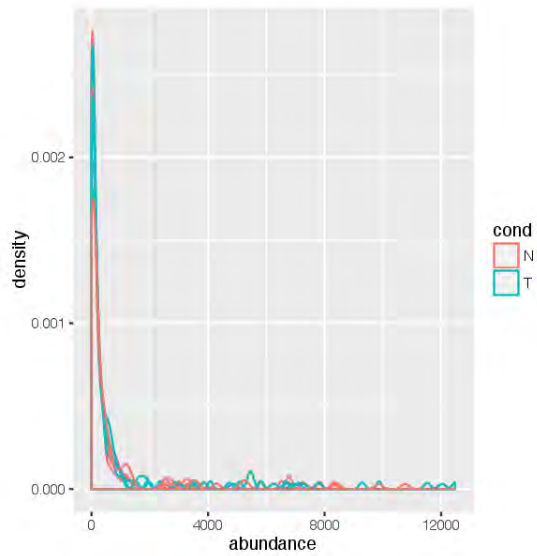
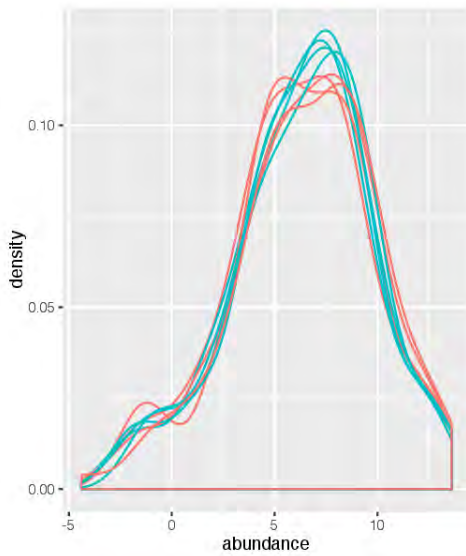
“Removed 22 rows containing non-finite values (stat_boxplot).”



```
In [24]: options( repr.plot.width=10,repr.plot.height=5)
          multiplot(PDrawmet,PDnorm, cols=2)
```

Warning message:

“Removed 22 rows containing non-finite values (stat_density).”




```

In [25]: #svg("metabo.svg")
options( repr.plot.width=5,repr.plot.height=5)

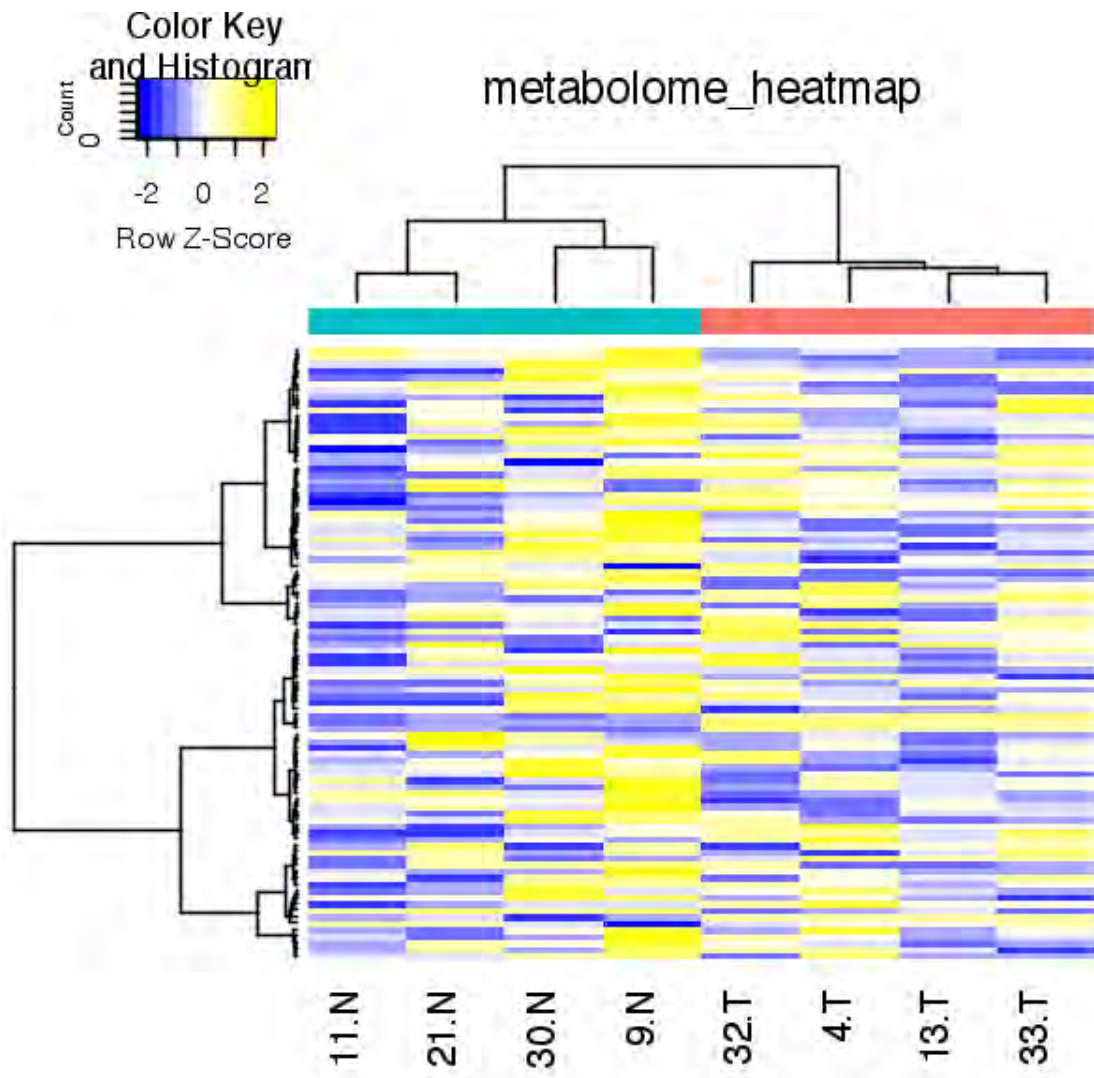
heatmap.2(met,ColSideColors=rep(rev(colors_hex),each=4),scale="row",denscol = F
          ,hclustfun=function(x)hclust(x,method="ward.D")
          ,trace='none',main="metabolome_heatmap",col=colorRampPalette(c("blue", "
          ,margins = c(6,1),cexCol = 1.5,cexRow = 1,labRow = NULL)
#dev.off()

```

```

Warning message in axis(4, iy, labels = labRow, las = 2, line = -0.5 + offsetRow, :
“conversion failure on 'γ-Aminobutyric acid' in 'mbscsToSbcs': dot substituted for <ce>”Warning m
“conversion failure on 'γ-Aminobutyric acid' in 'mbscsToSbcs': dot substituted for <b3>”Warning m
“conversion failure on 'γ-Aminobutyric acid' in 'mbscsToSbcs': dot substituted for <ce>”Warning m
“conversion failure on 'γ-Aminobutyric acid' in 'mbscsToSbcs': dot substituted for <b3>”Warning m
“conversion failure on 'β-Ala' in 'mbscsToSbcs': dot substituted for <ce>”Warning message in axis
“conversion failure on 'β-Ala' in 'mbscsToSbcs': dot substituted for <b2>”Warning message in axis
“conversion failure on 'β-Ala' in 'mbscsToSbcs': dot substituted for <ce>”Warning message in axis
“conversion failure on 'β-Ala' in 'mbscsToSbcs': dot substituted for <b2>”

```

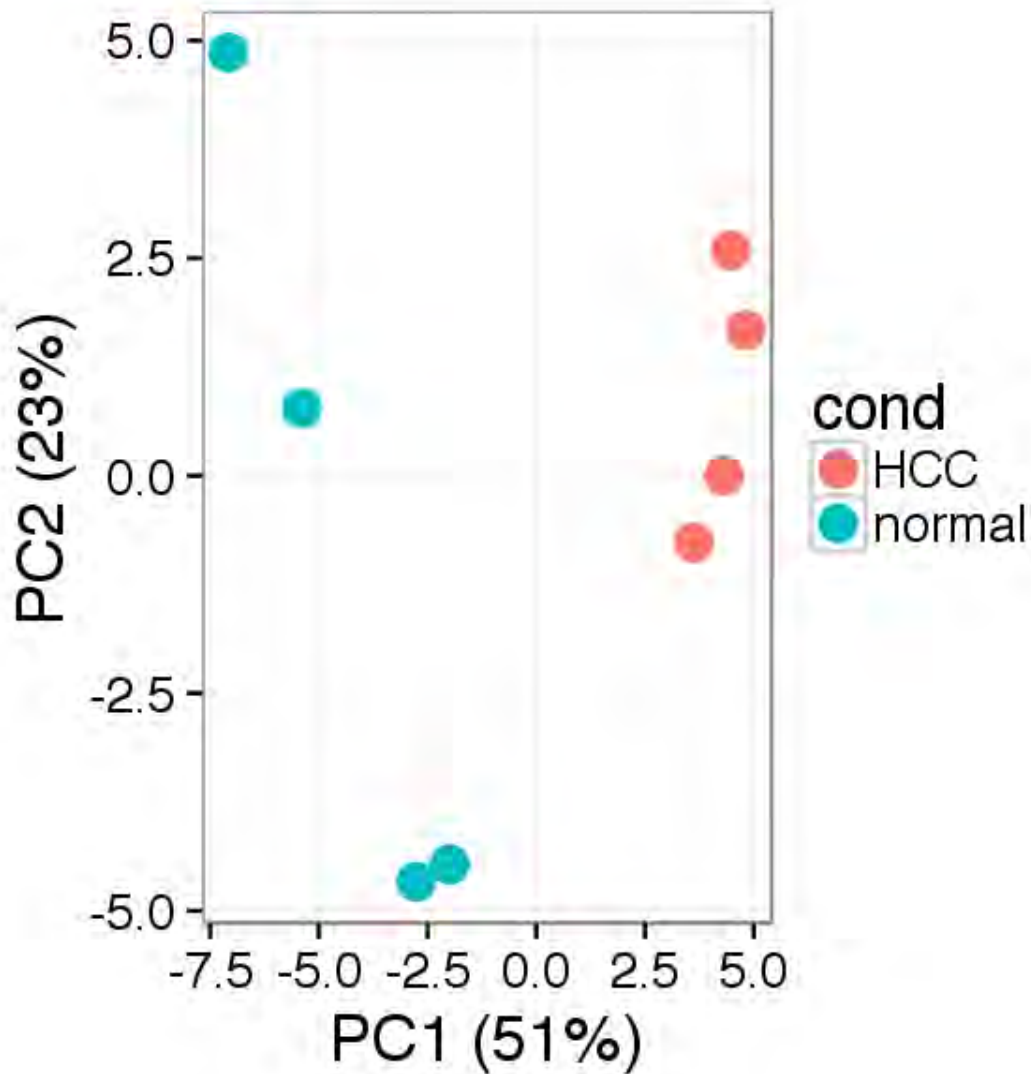


```
In [26]: pca<-prcomp(t(met))
         summary(pca)
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	4.8674	3.2961	2.3013	1.48389	1.34243	1.20460	1.03307
Proportion of Variance	0.5109	0.2343	0.1142	0.04748	0.03886	0.03129	0.02301
Cumulative Proportion	0.5109	0.7451	0.8594	0.90684	0.94570	0.97699	1.00000
	PC8						
Standard deviation	4.072e-15						
Proportion of Variance	0.000e+00						
Cumulative Proportion	1.000e+00						

```
In [27]: cond<-rep(c("normal","HCC"),each=4)
#svg("pca_met.svg")
ggplot(data.frame(pca$x), aes(x=PC1, y=PC2),size=5) + theme_bw(base_size = 20) +
  geom_point(aes(col=cond),size=5) + labs(x="PC1 (51%)",y="PC2 (23%)")
#dev.off()
```



```
In [28]: options( repr.plot.width=7,repr.plot.height=5)
```

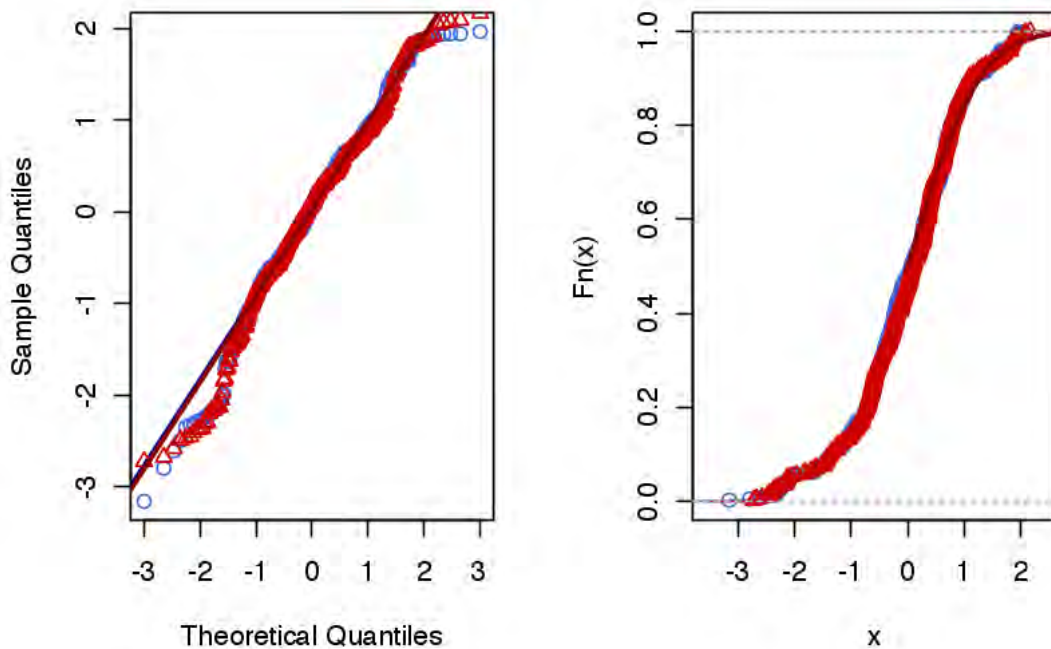
```
log_all <-met
log_all<-scale(log_all,center = T,scale=T)
colnames(log_all) <- c(paste0("N_", 1:4), paste0("HCC_", 1:4))
log_N <- unlist(log_all[, 1:4])
log_HCC <- unlist(unlist(log_all[, 5:8]))
```

```

#svg(paste(net, "norm.svg", sep=""))
par(mfrow = 1:2)
qqnorm(log_N, pch = 1, col = "royalblue", main = "")
qqline(log_N, lwd = 2, col = "darkblue")
points(qqnorm(log_HCC, plot.it = F), pch = 2, col = "red3")
qqline(log_HCC, lwd = 2, col = "darkred")

x <- seq(0, 16, length.out = 256)
plot(ecdf(log_N), pch = 1, col = "royalblue", main = "")
lines(x, pnorm(x, mean(log_N, na.rm = T), sd(log_N, na.rm = T)), lwd = 2,
      col = "darkblue")
plot(ecdf(log_HCC), pch = 2, add = T, col = "red3")
lines(x, pnorm(x, mean(log_HCC, na.rm = T), sd(log_HCC, na.rm = T)), lwd = 2,
      col = "darkred")
#dev.off()

```



```

In [29]: library(limma)
fenotipe<-factor(c(rep("Normal",4),rep("HCC",4)))####fenotipo
design <- model.matrix(~0+fenotipe)
contrast.matrix <- makeContrasts(TvN=fenotipeHCC-fenotipeNormal
                                ,levels=design)

```

```

fit <- lmFit(met,design)
fit<-contrasts.fit(fit,contrast.matrix)
fit <- eBayes(fit,robust =T)

```

```

In [30]: sigmet<-topTable(fit,coef=1,p.value=.05,lfc=0.5,adjust.method = "fdr")
sigmet

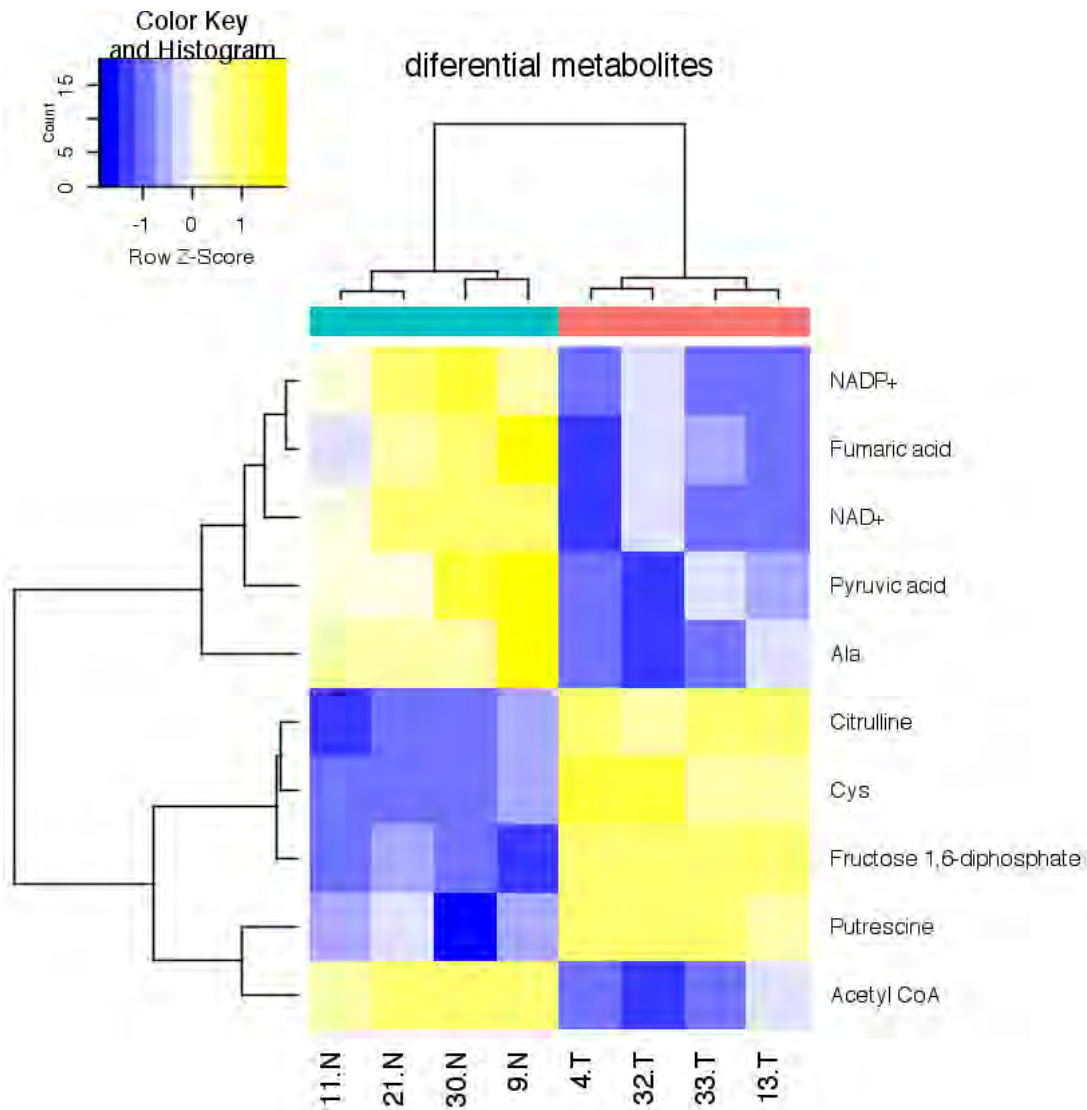
```

	logFC	AveExpr	t	P.Value	adj.P.Val	B
Fructose 1,6-diphosphate	3.2034617	4.997026	12.117303	5.813349e-07	5.464548e-05	6.7675085
Citrulline	2.6938798	5.855073	11.039801	1.301997e-06	6.119385e-05	5.9320172
Cys	3.9051437	5.675837	9.860578	4.010958e-06	1.256767e-04	4.7815871
Acetyl CoA	-2.2022600	1.866947	-7.569467	3.044429e-05	7.154409e-04	2.6114845
Putrescine	3.1333392	2.145449	5.452596	4.038106e-04	7.591639e-03	-0.1164250
NADP+	-1.0423461	7.994880	-5.196642	5.281895e-04	8.274969e-03	-0.4216809
Ala	-1.0505876	11.297497	-4.333992	1.799577e-03	2.416574e-02	-1.7168822
NAD+	-0.7762601	9.115463	-3.962764	3.151554e-03	3.703076e-02	-2.3046274
Fumaric acid	-0.9115285	7.470164	-3.765476	4.277325e-03	4.467428e-02	-2.6233946
Pyruvic acid	-1.8035716	6.234831	-3.721154	4.759225e-03	4.473671e-02	-2.7205262

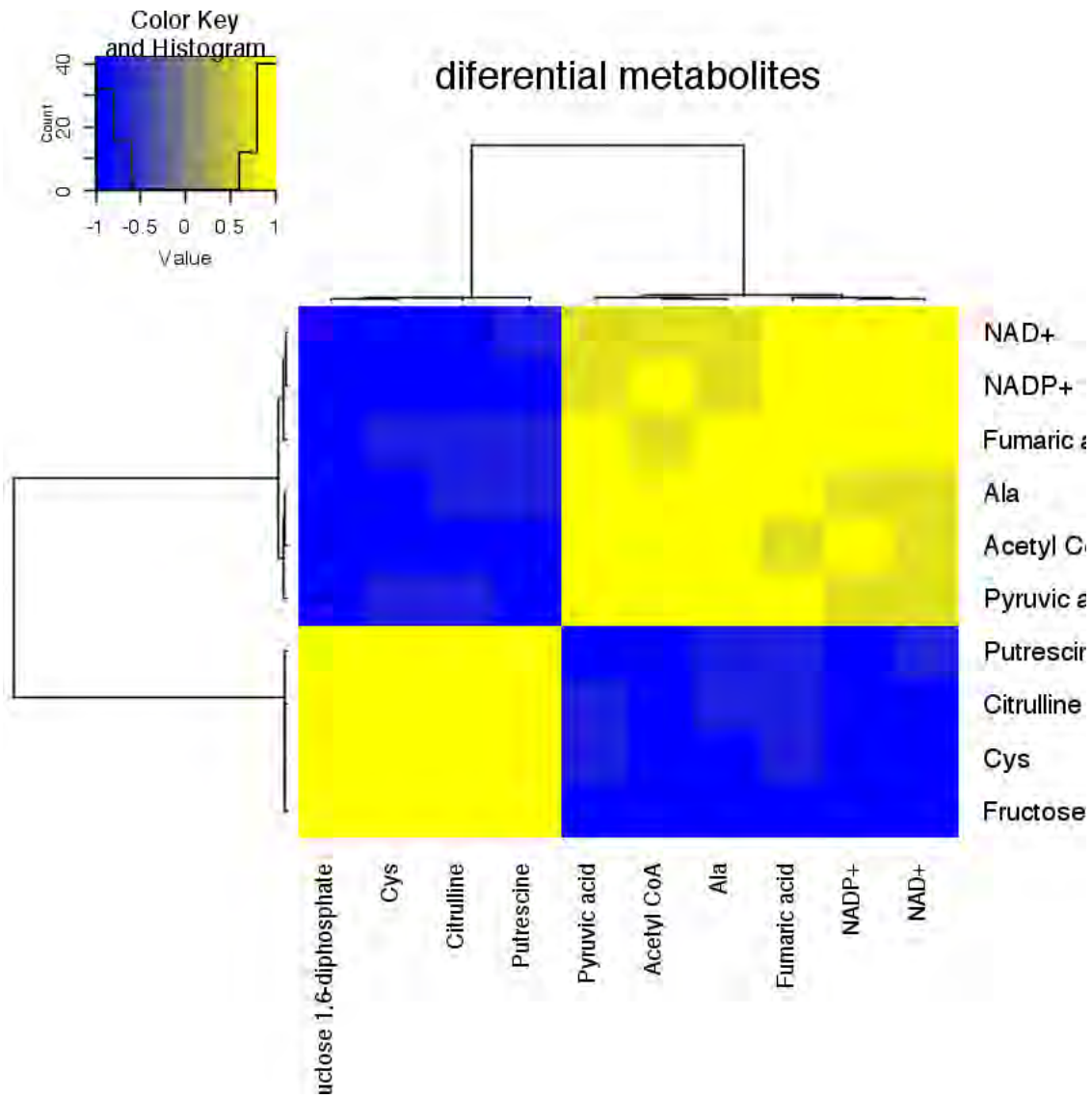
```

In [31]: options( repr.plot.width=7,repr.plot.height=7)
heatmap.2(met[rownames(sigmet),],,ColSideColors=rep(rev(colors_hex),each=4)
, scale="row",denscol = F
, hclustfun=function(x)hclust(x,method="ward.D")
, trace='none',main="diferential metabolites"
, col=colorRampPalette(c("blue","white","yellow"))(10)
, cexCol = 1.5 ,margins = c(5,15))

```



```
In [32]: heatmap.2(cor(t(met[rownames(sigmet),]))
, scale="non", denscol = T
, hclustfun=function(x)hclust(x,method="ward.D")
, trace='none', main="diferential metabolites"
, col=colorRampPalette(c("blue","yellow"))(10)
, cexCol = 1 , margins = c(10,5))
```



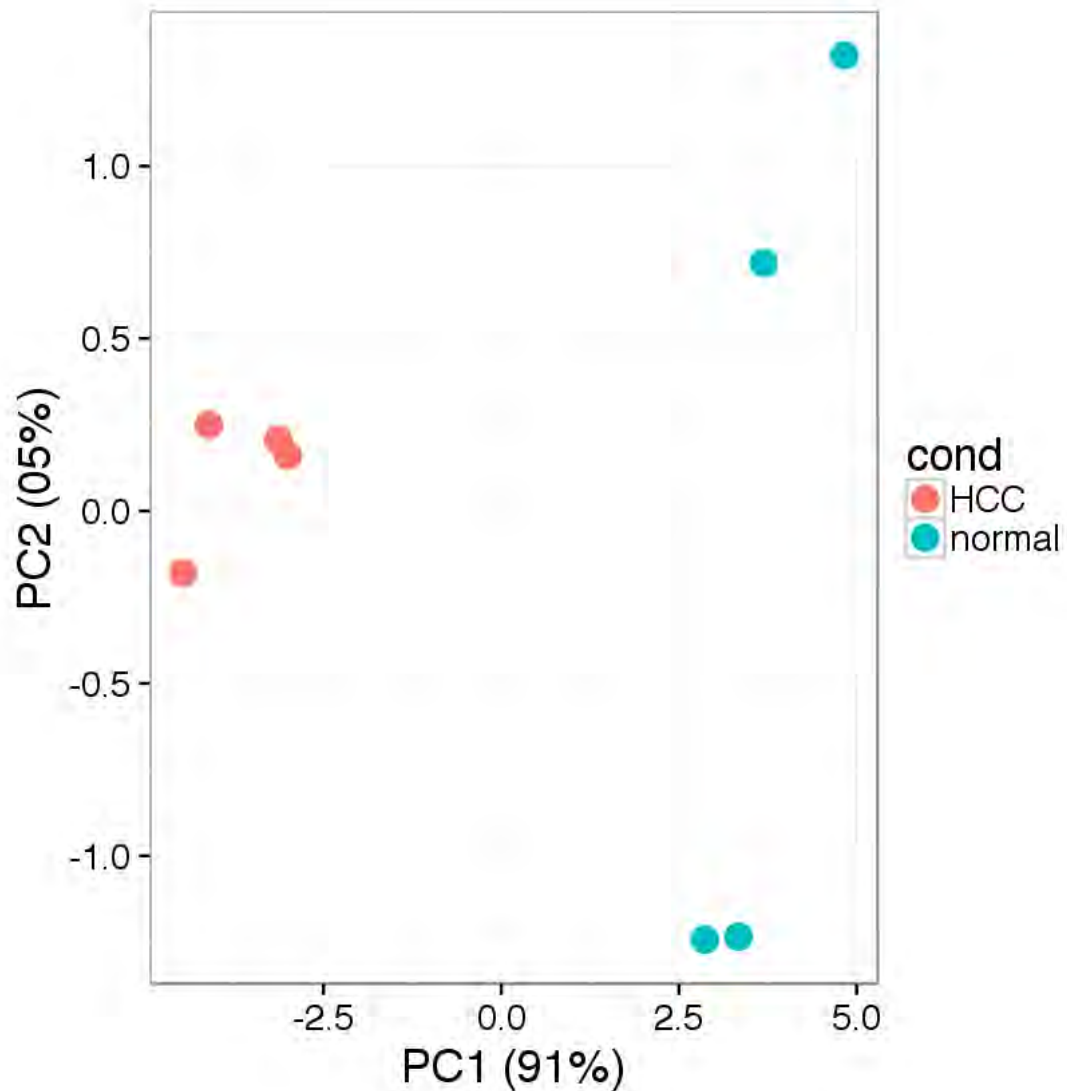
```
In [33]: pca<-prcomp(t(met[rownames(sigmet),]))
summary(pca)
cond<-rep(c("normal", "HCC"), each=4)
#svg("pca_met.svg")
ggplot(data.frame(pca$x), aes(x=PC1, y=PC2), size=5) + theme_bw(base_size = 20) +
  geom_point(aes(col=cond), size=5) + labs(x="PC1 (91%)", y="PC2 (05%)")
```

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	4.003	0.88604	0.74618	0.51291	0.37600	0.27791	0.13256
Proportion of Variance	0.897	0.04394	0.03116	0.01472	0.00791	0.00432	0.00098
Cumulative Proportion	0.897	0.94089	0.97206	0.98678	0.99469	0.99902	1.00000

PC8

Standard deviation 9.879e-16
Proportion of Variance 0.000e+00
Cumulative Proportion 1.000e+00



1.3.1 Dycone

First we will read the network that expands the one used at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4917846/> with metabolism of aminoacids in specific urea cycle and alanine metabolism

```
In [34]: net<- "myIDH_PDH_noPI_UREA_ALA_rev"  
library(dycone)
```



```
r <- read_reactions(paste("reactions",net,".csv",sep=""))
length(r)
```

Attaching package: ‘dycone’

The following object is masked from ‘package:AnnotationDbi’:

```
species
```

The following object is masked from ‘package:BiocGenerics’:

```
species
```

86

At first we will try to see which metabolites in the model are missing in the measurements. Matching will be done based on KEGG IDs.

```
In [35]: id_map <- read.csv("id_map.csv", stringsAsFactors=FALSE)
matches <- sapply(id_map$kegg, grep_id, x = metab$KEGG.ID)
miss <- is.na(matches)
table(miss)
```

```
miss
FALSE TRUE
  53    5
```

This means we are missing 10 percent of the metabolites used in the model. Given that there is no database for rat metabolome and the missing metabolites measure in liver could not be found in open access data we used human metabolome database to complete it. This data will only be used in the stability analysis and not in the differential analysis. HMDB assigns a single KEGG ID to each metabolite in the database. However, there are cases where a single metabolite can be identified by several IDs in the HMDB. This may happen if we have “Glucose” which may for instance map to “alpha-D-Glucose” or “beta-D-Glucose”. This is the reason why our id_map includes several HMDB and KEGG IDs for some metabolites.

We will now scrape the concentrations for all metabolites in the model from HMDB. This might take some time so we will use the caching operator %c% from dycone which will take an R expression and will cache all assigned variables in that expression to a cache file, so that rerunning the script will read the results from the cache and not rerun the analysis. To run the analysis again simply delete the cache file without changing any of the code.

```
In [36]: full <- id_map[, 1:2]
d_idx <- 6:13
m <- matrix(NA, nrow = nrow(full), ncol = length(d_idx))
colnames(m) <- names(metab)[d_idx]
full <- cbind(full, m)
```

```

matched_idx <- !is.na(matches)
full[matched_idx, 3:10] <- metab[matches[matched_idx], d_idx]
{
  concs <- hmdb_concentration(id_map$hmdb, add = id_map[, 1:2])
} %c% "scraped_concs.Rd"

```

In order to fill the gaps in the full data set we will use the patch function from dycone. patch will first attempt to fill any hole with measurements from the own data set, either by using the mean values of the same cell line or, if not available, by the mean value of the other cell line. Thus, giving priority to the local data before the scraped ones. Only measurements with missing entries in both cell lines are filled with the scraped mean concentrations from HMDB.

```

In [37]: m_concs <- as.vector(by(concs, concs$name, priority_mean))
names(m_concs) <- levels(factor(concs$name))
scraped <- data.frame(kegg = names(m_concs), normal = m_concs)
rownames(scraped) <- NULL
patched <- patch(full, id = 1, normal = 3:6, treatment = 7:10, ref_data = scraped)

```

The statistical tests used in dycone assume a log-normal distribution of the metabolic terms. It is sufficient to show an approximate normal distribution for the log-transformed metabolite measurements since the log-transform of the metabolic terms is a weighted sum of the logarithmic metabolite measurements for most of the common kinetics (such as mass-action, Michalis-Menten and Hill). So the comparison made in the previous section is enough

We already calculated the log-fold changes before, but in order to also assign some statistics to that we can also use the function hyp from dycone which does that for us. It will generate all log-fold changes between disease and normal measurements and perform an empirical Bayes version of t-test. To visualize the results those results, and also to compare the obtained regulations on a pathway level we will mark those reaction with a significant p.value as a triangle.

```

In [38]: options( repr.plot.width=8,repr.plot.height=7)
S <- stoichiometry(r)
mats <- ma_terms(S, patched[, c(1, 3:10)])
samples <- rep(c("normal", "disease"), each = 4)
h <- hyp(r, samples, mats,full = T,type="bias")
pw <- rp(make_irreversible(r), "pathway")[,2]
r_ids <- rp(make_irreversible(r), "KEGG_reaction")[,2]
h$hyp <- cbind(h$hyp, pathway = pw[h$hyp$idx])
h$hyp <- cbind(h$hyp, reaction_id = r_ids[h$hyp$idx])
write.csv(h$hyp, file = paste(net,"transform.csv",sep="_"), quote = F, row.names = F)
pw <- rp(make_irreversible(r), "pathway")[,2]
r_ids <- rp(make_irreversible(r), "KEGG_reaction")[,2]
h$hyp <- cbind(h$hyp, pathway = pw[h$hyp$idx])
h$hyp <- cbind(h$hyp, reaction_id = r_ids[h$hyp$idx])
#png("paths.png")
ggplot(h$hyp, aes(y = pathway, x = k_lfc, col = pathway)) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  geom_point(aes(shape = .05 > corr_pval),
             position = position_jitter(height = 0.2), size = 3) + scale_shape_manual(v
             coord_fixed(ratio = 1)+

```

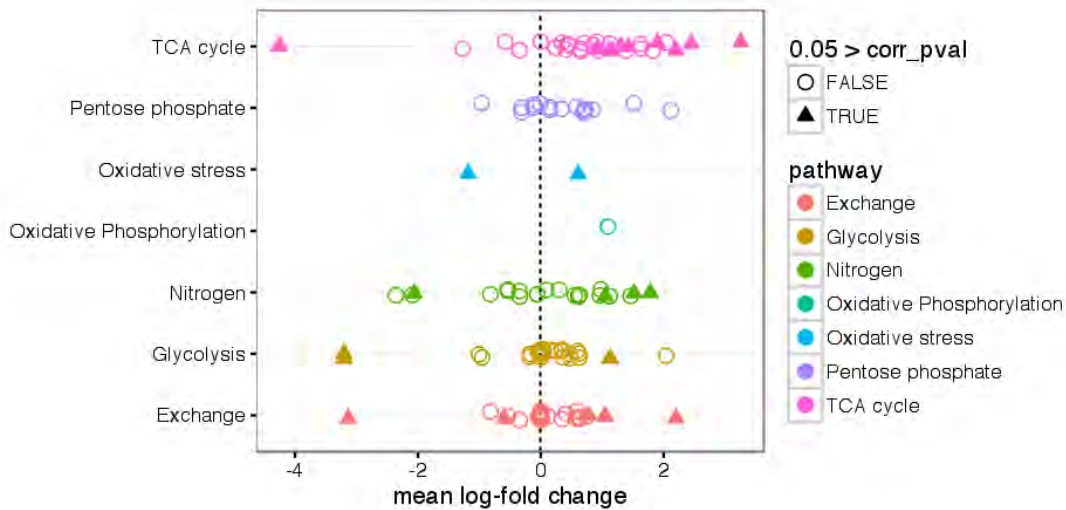
```

theme_bw() + xlab("mean log-fold change") + ylab("")
#dev.off()

```

Warning message:

“Zero sample variances detected, have been offset”



Comparison with gene expression data

Gene expression analysis We will start by mapping all the EC numbers of the reactions used in the model to its respective ENTREZ gene ids and names and saving that information into the info data frame. Note that a single reaction might be associated to several EC numbers (isoenzymes for example) and every EC number might be associated with several genes. We will only use those EC numbers for which we could find a corresponding gene on the arrays.

```

In [39]: library(AnnotationDbi, quietly=TRUE, warn.conflicts=F)
         library(org.Rn.eg.db, quietly=TRUE, warn.conflicts=F)

```

```

library(dplyr, quietly=TRUE, warn.conflicts=F)

ecs <- rp(make_irreversible(r), "KEGG_enzyme")
ens <- read.table("http://rest.kegg.jp/link/ec/rno")
ens[,1]<-gsub("rno:", "", ens[,1])
ens[,2]<-gsub("ec:", "", ens[,2])
colnames(ens)<-c("ENTREZID", "ENZYME")
load("gene_expression.Rd")
info <- ecs %>% group_by(r_idx) %>% do(ens[ens$ENZYME %in% .$KEGG_enzyme,])
good <- sapply(info$ENTREZID, function(eid) !is.na(eid) &
              (eid %in% gsub("_at", "", rownames(ma_lfcs))))
cat(sprintf("%f%% of enzymes found on array.\n", sum(good)/length(good)*100))
info <- info[good, ]
info <- info %>% group_by(r_idx) %>% mutate(met_lfc =
                                          h$hyp$k_lfc[h$hyp$idx %in% r_idx], met_pval =
                                          h$hyp$corr_pval[h$hyp$idx %in% r_idx], pathway =
                                          h$hyp$pathway[h$hyp$idx %in% r_idx])

head(info)
get_ma_lfc <- function(eid) {
found <- which(eid[1] == gsub("_at", "", rownames(ma_lfcs)))

  return(data.frame(ge_lfc = ma_lfcs$logFC[found],
                    ge_pval = ma_lfcs$adj.P.Val[found]))
}

ge <- lapply(info$ENTREZID, get_ma_lfc)
info <- cbind(as.data.frame(info), do.call(rbind, (ge)))
info$significant <- info$ge_pval<0.05 & info$met_pval<0.05 & abs(info$ge_lfc)>1
symbol<-AnnotationDbi::select(org.Rn.eg.db, keys = info$ENTREZID, keytype = "ENTREZID",
                              columns = c("SYMBOL"))

```

87.313433% of enzymes found on array.

r_idx	ENTREZID	ENZYME	met_lfc	met_pval	pathway
1	296973	5.4.2.4	2.220446e-16	1.0000000	Glycolysis
1	296973	5.4.2.11	2.220446e-16	1.0000000	Glycolysis
1	24642	5.4.2.11	2.220446e-16	1.0000000	Glycolysis
1	24959	5.4.2.11	2.220446e-16	1.0000000	Glycolysis
2	296973	5.4.2.4	6.220281e-01	0.3691219	Glycolysis
2	296973	5.4.2.11	6.220281e-01	0.3691219	Glycolysis

'select()' returned many:1 mapping between keys and columns

```

In [40]: info<-cbind(info,SYMBOL=symbol$SYMBOL)
info[info$significant,]
info[info$met_pval<0.05,]
write.table(unique(info),paste(net,"genDyc2.txt",sep="_"),sep="\t",row.names = F,col.names = F)

```

	r_idx	ENTREZID	ENZYME	met_lfc	met_pval	pathway	ge_lfc	ge_pval
156	65	24552	1.1.1.40	1.8952834	0.02761741	TCA cycle	1.580866	8.544120e-0
157	65	361602	1.1.1.40	1.8952834	0.02761741	TCA cycle	-1.784294	7.988058e-1
158	66	24552	1.1.1.40	2.4503906	0.04040969	TCA cycle	1.580866	8.544120e-0
159	66	361602	1.1.1.40	2.4503906	0.04040969	TCA cycle	-1.784294	7.988058e-1
185	77	64317	1.11.1.9	-1.1832261	0.01624760	Oxidative stress	1.258622	4.224979e-0
187	77	29326	1.11.1.9	-1.1832261	0.01624760	Oxidative stress	1.598727	4.939836e-1
191	78	116686	1.8.1.7	0.6123639	0.01030744	Oxidative stress	1.948235	3.906375e-1

	r_idx	ENTREZID	ENZYME	met_lfc	met_pval	pathway	ge_lfc	ge_pval
15	6	24189	4.1.2.13	-3.203462	0.0006876987	Glycolysis	0.01126604	9.2799
16	6	24191	4.1.2.13	-3.203462	0.0006876987	Glycolysis	-0.21713224	7.9922
17	6	24190	4.1.2.13	-3.203462	0.0006876987	Glycolysis	-0.04733125	2.0160
18	6	299052	4.1.2.13	-3.203462	0.0006876987	Glycolysis	-0.22546997	1.2421
23	8	24362	3.1.3.11	-3.203462	0.0006876987	Glycolysis	-0.45170663	5.9150
24	8	114508	3.1.3.11	-3.203462	0.0006876987	Glycolysis	-0.09045153	3.9873
52	24	24534	1.1.1.27	1.127217	0.0344430488	Glycolysis	0.47801135	2.4375
53	24	29634	1.1.1.27	1.127217	0.0344430488	Glycolysis	-0.11232324	1.2330
54	24	369018	1.1.1.27	1.127217	0.0344430488	Glycolysis	-0.16459827	4.7714
94	43	117098	1.2.4.1	3.254589	0.0162476048	TCA cycle	-0.19003090	7.6153
95	43	81654	2.3.1.12	3.254589	0.0162476048	TCA cycle	0.34058119	4.9803
96	43	29554	1.2.4.1	3.254589	0.0162476048	TCA cycle	0.51326836	1.1736
97	43	298942	1.8.1.4	3.254589	0.0162476048	TCA cycle	0.25444714	1.9478
98	43	289950	1.2.4.1	3.254589	0.0162476048	TCA cycle	-0.44343594	2.6792
100	45	170587	2.3.3.1	2.202260	0.0061096921	TCA cycle	0.05029997	5.0689
103	47	79250	4.2.1.3	-4.256198	0.0176008276	TCA cycle	0.72969809	2.1822
104	47	50655	4.2.1.3	-4.256198	0.0176008276	TCA cycle	-0.05483686	3.4262
120	54	114597	6.2.1.4	1.118495	0.0092437755	TCA cycle	-0.14566412	4.7089
121	54	290566	1.2.4.2	1.118495	0.0092437755	TCA cycle	0.13255724	1.0091
122	54	360975	1.2.4.2	1.118495	0.0092437755	TCA cycle	0.35842843	2.1098
123	54	298942	1.8.1.4	1.118495	0.0092437755	TCA cycle	0.25444714	1.9478
124	54	361272	1.2.4.2	1.118495	0.0092437755	TCA cycle	-0.84769299	2.1885
125	54	299201	2.3.1.61	1.118495	0.0092437755	TCA cycle	0.36880699	1.4221
126	54	362404	6.2.1.4	1.118495	0.0092437755	TCA cycle	-0.15235565	3.6413
127	54	361071	6.2.1.4	1.118495	0.0092437755	TCA cycle	-0.15115172	7.7057
128	55	114597	6.2.1.5	1.303046	0.0408319508	TCA cycle	-0.14566412	4.7089
129	55	290566	1.2.4.2	1.303046	0.0408319508	TCA cycle	0.13255724	1.0091
130	55	360975	1.2.4.2	1.303046	0.0408319508	TCA cycle	0.35842843	2.1098
131	55	298942	1.8.1.4	1.303046	0.0408319508	TCA cycle	0.25444714	1.9478
132	55	361272	1.2.4.2	1.303046	0.0408319508	TCA cycle	-0.84769299	2.1885
:	:	:	:	:	:	:	:	:
136	56	114597	6.2.1.5	1.4228994	0.017600828	TCA cycle	-0.14566412	4.7089
137	56	290566	1.2.4.2	1.4228994	0.017600828	TCA cycle	0.13255724	1.0091
138	56	360975	1.2.4.2	1.4228994	0.017600828	TCA cycle	0.35842843	2.1098
139	56	298942	1.8.1.4	1.4228994	0.017600828	TCA cycle	0.25444714	1.9478
140	56	361272	1.2.4.2	1.4228994	0.017600828	TCA cycle	-0.84769299	2.1885
141	56	299201	2.3.1.61	1.4228994	0.017600828	TCA cycle	0.36880699	1.4221
142	56	362404	6.2.1.5	1.4228994	0.017600828	TCA cycle	-0.15235565	3.6413
143	56	361071	6.2.1.5	1.4228994	0.017600828	TCA cycle	-0.15115172	7.7057
144	57	24368	4.2.1.2	0.9115285	0.034443049	TCA cycle	-0.05440262	4.2751
150	61	157074	1.3.5.1	1.1698395	0.008253373	TCA cycle	0.13433221	6.9137
151	61	298596	1.3.5.1	1.1698395	0.008253373	TCA cycle	-0.09545681	4.7671
156	65	24552	1.1.1.40	1.8952834	0.027617406	TCA cycle	1.58086599	8.5441
157	65	361602	1.1.1.40	1.8952834	0.027617406	TCA cycle	-1.78429408	7.9880
158	66	24552	1.1.1.40	2.4503906	0.040409686	TCA cycle	1.58086599	8.5441
159	66	361602	1.1.1.40	2.4503906	0.040409686	TCA cycle	-1.78429408	7.9880
184	77	24404	1.11.1.9	-1.1832261	0.016247605	Oxidative stress	-0.74546655	4.8653
185	77	64317	1.11.1.9	-1.1832261	0.016247605	Oxidative stress	1.25862199	4.2249
186	77	259233	1.11.1.9	-1.1832261	0.016247605	Oxidative stress	-0.31943852	8.2548
187	77	29326	1.11.1.9	-1.1832261	0.016247605	Oxidative stress	1.59872708	4.9398
188	77	113919	1.11.1.9	-1.1832261	0.016247605	Oxidative stress	0.01908708	9.1411
189	77	298376	1.11.1.9	-1.1832261	0.016247605	Oxidative stress	0.06864310	3.9176

```
In [41]: h <- hyp(r, samples, mats, full = T,type="bias",sorted=F)
pw <- rp(make_irreversible(r), "pathway")[,2]
r_ids <- rp(make_irreversible(r), "KEGG_reaction")[,2]
h$hyp <- cbind(h$hyp, pathway = pw[h$hyp$idx])
h$hyp <- cbind(h$hyp, reaction_id = r_ids[h$hyp$idx])
```

Warning message:

```
“Zero sample variances detected, have been offset”
```

1.4 Exploration of NADP Enzymes

Given that the NADP/NADPH balance is affecting many of the reactions we explore how enzymes dependent of NADPH are expressing. This inhibition could be generating reactive oxygen species and activating enzymes NADPH dependent to regenerate antioxidants. So we are going to retrieve from kegg all the enzymes that use NADPH and which are differential express

```
In [42]: NADp_reactions<-c(as.matrix(read.table("http://rest.kegg.jp/link/ec/cpd:C00005")[-1] )
,as.matrix(read.table("http://rest.kegg.jp/link/ec/cpd:C00006")[-1]))
RNO_Genes<-read.table("http://rest.kegg.jp/link/ec/rno")##genes
colnames(RNO_Genes)<-c("Gene", "E.C")
nadp_RNO<-RNO_Genes[RNO_Genes$E.C%in%NADp_reactions,]
nadp_RNO$Gene<-gsub("rno:", "", nadp_RNO$Gene)
head(nadp_RNO)
```

	Gene	E.C
9	24192	ec:1.1.1.21
29	78959	ec:1.1.1.2
40	81656	ec:1.3.1.2
48	81708	ec:1.2.1.18
80	84493	ec:1.14.13.8
107	24377	ec:1.1.1.49

```
In [43]: nadp<-exprs(eset[rownames(eset)%in%paste(unique(nadp_RNO$Gene), "_at", sep="")&
rownames(eset)%in%rownames(ma_1fcs[abs(ma_1fcs$logFC)>1.5 & ma_1fcs$ad
```

In order to see if these enzymes can differentiate between the control group and the drug group we are going to use a heatmap

```
In [44]: options(repr.plot.width=8,repr.plot.height=7)

library(RColorBrewer)

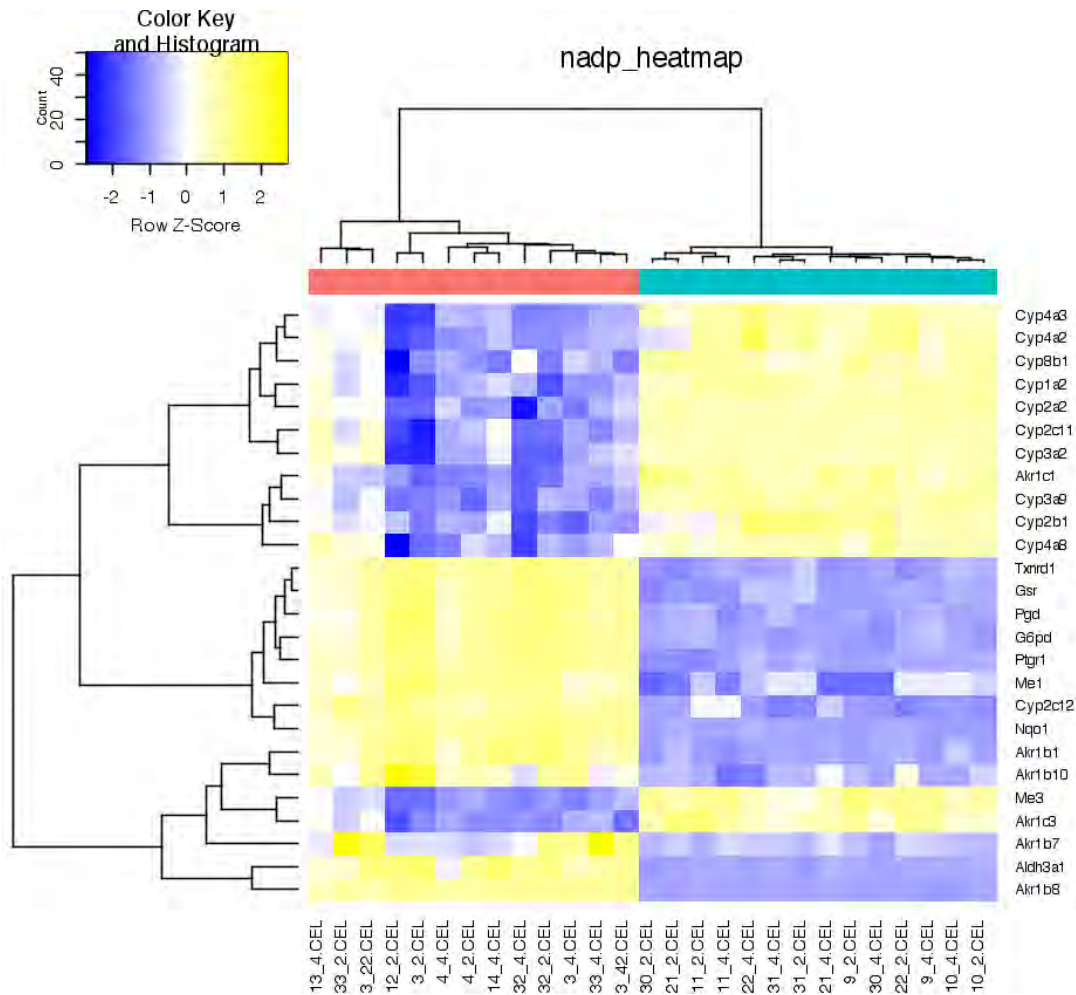
col<-as.character(condition)
col[condition=="normal"]<-colors_hex[2]
col[condition=="HCC"]<-colors_hex[1]
library(gplots)
library(AnnotationDbi, quietly=TRUE, warn.conflicts=F)
library(org.Rn.eg.db, quietly=TRUE, warn.conflicts=F)
```

```
ens <- AnnotationDbi::select(org.Rn.eg.db, keys = gsub("_at","",rownames(nadp)), keytype = "SYMBOL", columns = c("SYMBOL", "ENTREZID"))
```

```
dim(ens)
heatmap.2(nadp,ColSideColors=col,scale="row",denscol = F,labRow = ens$SYMBOL,margins = c(10,10),trace='non',main="nadp_heatmap",col=colorRampPalette(c("blue","white","yellow")),hclustfun=function(x)hclust(x,method="ward.D2"))
```

'select()' returned 1:1 mapping between keys and columns

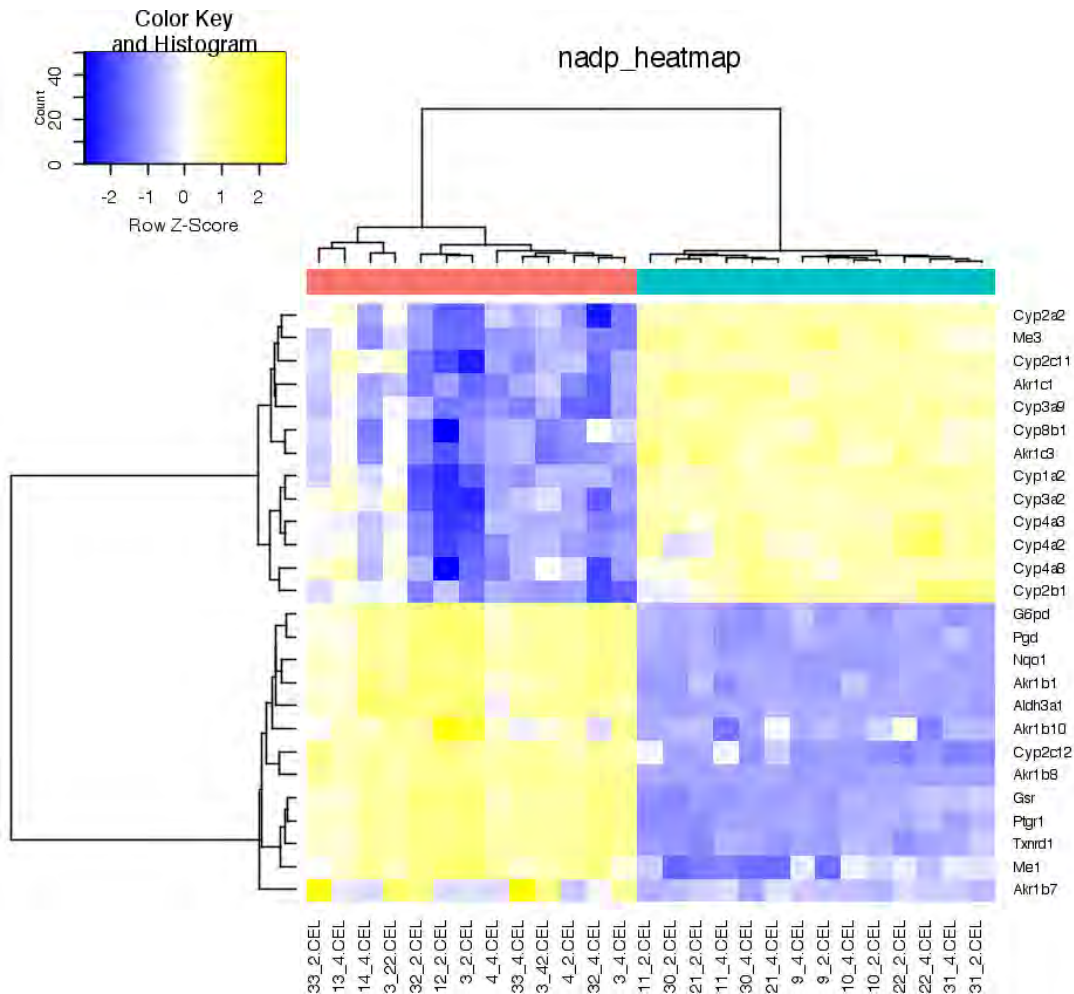
1.26.2.2



```
In [45]: heatmap.2(nadp,ColSideColors=col,scale="row",denscol = F,labRow = ens$SYMBOL,margins = c(10,10),trace='none',main="nadp_heatmap",col=colorRampPalette(c("blue","white","yellow")))
```



```
,hclustfun=function(x)hclust(x,method="ward.D2")
,distfun=function(x)as.dist(1-(cor(t(x),method="kendall")))
)
```



```
In [46]: d <- cor(t(h$lfc_disease))
colnames(d)<-h$hyp$idx
rownames(d)<-h$hyp$idx
(d[c(47,48),order(abs(d[48,]),decreasing=T))][,1:15]
(d[c(99),order(abs(d[99,]),decreasing=T)][1:15]
d <- cor(t(h$lfc_normal))
colnames(d)<-h$hyp$idx
rownames(d)<-h$hyp$idx
(d[c(47,48),order(abs(d[48,]),decreasing=T))][,1:15]
```

Warning message in cor(t(h\$lfc_disease)):
 "the standard deviation is zero"

	47	48	16	46	99	84	71	96	102	97
47	1	1	0.8674616	0.8285037	-0.7760347	0.7517033	0.7517033	0.7517033	0.7517033	-0.71337
48	1	1	0.8674616	0.8285037	-0.7760347	0.7517033	0.7517033	0.7517033	0.7517033	-0.71337
99	1	97	0.915080803559941	133	0.878488650108689	79	0.878488650108689	137		
0.795891502129328	47		-0.776034695985081	48	-0.776034695985081	29	0.748014767166488	45		
0.677313954813114	126		0.677313954813114	16	-0.611908462025232	71	-0.608953391597263	96		
-0.608953391597263	102				-0.608953391597263	84			-0.608953391597263	

Warning message in cor(t(h\$lfc_normal)):

“the standard deviation is zero”

	48	47	22	95	94	139	6	8	88	12
47	1	1	-0.9846091	-0.9830447	-0.9768927	-0.9458789	-0.9016433	-0.9016433	-0.8852346	0.8
48	1	1	-0.9846091	-0.9830447	-0.9768927	-0.9458789	-0.9016433	-0.9016433	-0.8852346	0.8