

The Differential Rank Conservation Algorithm
(DIRAC) Reveals Deregulation of Central
Metabolic Pathways in Hepatocellular Carcinoma

Master of Science Thesis

Author: Elias Björnson
contact: eliasb@student.chalmers.se

Supervisor: Adil Mardinoglu
Examiner: Jens Nielsen

Systems & Synthetic Biology
Department of Chemical and Biological Engineering
Chalmers University of Technology
Gothenburg, Sweden

June 10, 2014



CHALMERS
UNIVERSITY OF TECHNOLOGY

Contents

1	Goal and purpose	1
2	Thesis outline	1
3	Introduction	3
3.1	Hepatocellular carcinoma	3
3.2	The Cancer Genome Atlas (TCGA) data base	4
3.3	Gene set collections	4
3.4	The DIRAC algorithm	5
3.4.1	What does deregulation mean?	8
3.4.2	Can deregulation reveal frequently mutated subnetworks?	8
3.4.3	What does differential regulation mean?	9
3.5	Reporter metabolites	10
4	Results & Discussion	11
4.1	High-entropy networks in HCC	11
4.1.1	TCA cycle	17
4.1.2	Electron transport chain	18
4.1.3	Fatty acid metabolism	18
4.1.4	Carnitine shuttle and β -oxidation	18
4.1.5	Glutamine and glutamate metabolism	19
4.1.6	Altered energy metabolism in HCC	19
4.1.7	Global deregulation	21
4.2	Low-entropy networks in HCC	22
4.2.1	Fatty acid biosynthesis	25
4.2.2	Terpenoid biosynthesis	27
4.2.3	ROS detoxification	27
4.3	Differentially regulated networks in HCC	28
4.3.1	Sphingolipid and glycosphingolipid metabolism	31
4.4	Reporter metabolites	33
4.5	Network entropy in other cancers	35
4.6	Criticism	38
5	Conclusions	38
6	Materials & Methods	40
6.1	Implementation of DIRAC	40
6.1.1	Rank conservation indices	40
6.1.2	Rank difference score	40
6.1.3	Significance testing of deregulation	41
6.2	RNA-Seq data retrieval	41
6.3	Patients and clinical data	41
6.4	Differential expression	42

Abstract

Hepatocellular carcinoma (HCC) is a deadly disease without existing cure or effective treatment. If the genetics and metabolics of this form of liver cancer would be better understood the chance of developing effective treatments would likely increase.

HCC cells are different from normal liver cells. When it comes to gene expression the traditional way is to define these differences in terms of up- or downregulation of genes or metabolic subnetworks. The purpose of this report however is to present an alternative approach regarding the definition of what makes HCC cells different from matched liver cells. This approach does not primarily focus on between-phenotype differential gene expression but rather on metabolic network regulation and can be viewed in terms of metabolic network *entropy*.

This alternative approach is made possible through implementation of the so called *Differential Rank Conservation* algorithm (DIRAC). In addition traditional differential expression analysis and detection of differentially expressed metabolites by the *reporter metabolites* algorithm was performed. The basis of the analysis was RNA-Seq data from 163 HCC patients downloaded from The Cancer Genome Atlas database.

The results indicate fundamental alterations in network regulation of central metabolic pathways in HCC. For example the TCA cycle, the electron transport chain and fatty acid metabolism show general apparent dysfunction with network deregulation and concomittant average downregulation of gene expression in HCC. In contrast fatty acid biosynthesis seems to be under tight regulatory control with average upregulation of gene expression in HCC. The DIRAC algorithm also revealed large scale average metabolic network disorganization in HCC, a phenomena referred to as *global deregulation*. In addition six other cancers were also analysed in terms of their global regulation of their metabolic networks. This analysis revealed a correlation between the degree of global deregulation and malignancy of the cancer.

In conclusion the DIRAC algorithm offers an alternative view on between-phenotype differences of gene expression which reveals important alterations of the metabolism in HCC.

1 Goal and purpose

The Differential Rank Conservation algorithm (DIRAC) offers an alternative perspective on between-phenotype differences of gene expression. This algorithm falls in the category of so called *relative expression analysis* and is designed to analyse network-level behaviour of cell metabolism. It offers a perspective beyond differential expression which would be described in terms of network regulation rather than up- or downregulation.

The purpose of this project was (i) to implement the DIRAC algorithm on RNA-Seq data from samples of HCC and matched adjacent normal liver tissue in order to define differences between HCC and normal liver in terms of network regulation and (ii) to make a comparison of these results to traditional differential expression-based approaches.

The goal of the project was to offer alternative information of the metabolic behaviour of HCC, information which traditional analytical approaches might overlook.

2 Thesis outline

The basic outline of this project is straight-forward. The Cancer Genome Atlas (TCGA) is an ongoing project with the purpose of collecting and sequencing many different cancer tissues and matched control tissues and making this data publically available. RNA-Seq gene expression data was downloaded from the TCGA data base and subsequently analysed. The analysis was two-fold. The main analysis was the implementation of the DIRAC algorithm on the TCGA RNA-Seq gene expression data of HCC. In addition differential expression-analysis was performed and the reporter metabolites algorithm was implemented on the output of the differential expression analysis (see figure 1 below). The DIRAC algorithm was also implemented on TCGA RNA-Seq gene expression data from six other cancers in order to analyse the global network behaviour of these cancers in comparison with HCC.

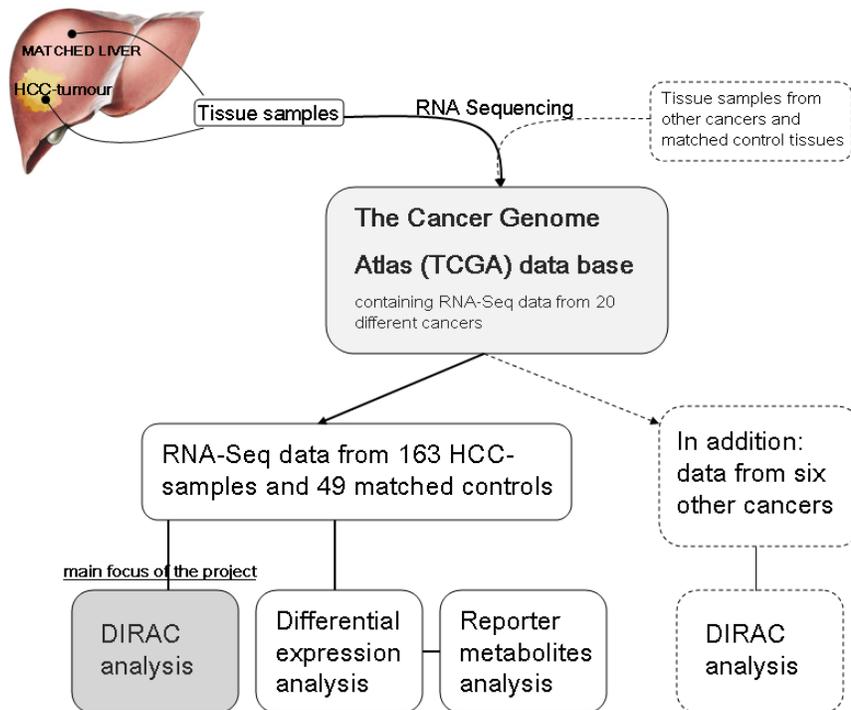


Figure 1: Basic outline of this master thesis project. The main focus of the project was the implementation of the DIRAC algorithm on the RNA-Seq data from the 163 hepatocellular carcinoma samples and the 49 matched adjacent liver control samples. Differential expression with subsequent reporter metabolites analysis was also performed. In addition the DIRAC algorithm was implemented on RNA-Seq data from six other cancers.

3 Introduction

Here information about hepatocellular carcinoma, the TCGA data base, RNA-Seq data, gene set collections, the DIRAC algorithm and the reporter metabolites algorithm will be presented. In addition some general interpretations of what the results from the DIRAC algorithm could mean in a biological context will be presented.

3.1 Hepatocellular carcinoma

Cancer in the liver can arise and manifest in several different ways. A first broad division can be made between benign and malignant liver cancers. Benign tumours do not cause severe symptoms and is usually curable [1]. Malignant liver cancer on the other hand is often life threatening. Malignant liver cancer can be of primary or secondary nature. Primary liver cancer originates in the liver whereas secondary liver cancer forms in another part of the body and later metastasize to the liver. There are different kinds of primary liver cancers. *Intrahepatic cholangiocarcinoma* arises from the cells of the bile duct canal, the canal that delivers bile to the gallbladder, and makes up approximately 10-20 percent of all liver cancers [1]. *Fibrolamellar carcinoma* is rare, making up less than one percent of all liver cancers and often have a better outlook than other malignant liver cancers [1].

The most common form of liver cancer and the form studied in this project is as mentioned called *hepatocellular carcinoma*. HCC is of primary type making up approximately 80 percent of all liver cancers. HCC can form as a single tumour and grow in size or can start as many small cancer nodules spread throughout the liver. This nodule form is often associated with cirrhosis (widespread scarring of tissue). The prognosis of HCC are poor with a one-year survival rate of less than 50 percent [2]. For men HCC is the second leading cause of cancer-related deaths in the world and for women the sixth leading cause [2]. The risk of developing HCC is highly increased with infection of hepatitis B or C. Worldwide around 80 percent of all cases of HCC are related to hepatitis B or C infection [3]. Another leading risk factor is alcohol abuse which can lead to cirrhosis and ultimately HCC. Another correlating factor is obesity, probably because obese people have a higher prevalence of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis which leads to increased risk of cirrhosis and HCC [1]. In the United States the overall prevalence of HCC has tripled from 1975 to 2005 with the largest increase in prevalence in middle-aged men [2].

There are several treatment options available for HCC. A common treatment strategy is to surgically remove the tumour tissue or replace the whole liver by transplantation. Tumour ablation aims at destroying the tumour without removing it. Radiofrequency and microwave ablation destroys the tumour by heating it up. Cryotherapy kills the cancer cells by freezing them. Another ablation technique is called percutaneous ethanol injection where, as the name suggests, ethanol is injected in order to kill the cancer cells. Tumour emboliza-

tion is another treatment option which involves cutting the blood supply to the tumour but not to nearby liver tissue thereby selectively killing the cancer. External radiation therapy is also an option to treat HCC. Chemotherapy can also be used although usually with limited success [4]. Even with all these treatment strategies the survival rate is as mentioned above very low. Acquiring HCC today almost inevitable leads to death [2]. Therefore more targeted and effective treatments are highly desirable.

3.2 The Cancer Genome Atlas (TCGA) data base

The Cancer Genome Atlas is a coordinated project between two American institutes (National Cancer Institute and the National Human Genome Research Institute) aimed at accelerating the understanding of many types of human cancers. Tumor tissue from many different patients are in this project collected and carefully catalogued. Different sequencing technologies are then applied including next generation RNA-sequencing. The transcriptome of a typical cell contains vast amount of information. Previous methods based on microarray-technologies have been shown to not entirely accurately cover and capture the complete information of the transcriptome. Next generation sequencing in the form of RNA-Seq offers superior accuracy and coverage through massively parallel direct sequencing of cDNA [5]. This method is thus the method of choice when it comes to accurate estimations of gene expression. The RNA-Seq data generated within the TCGA project is made publically available so that researchers all over the world can easily and freely access and analyse the data [6].

3.3 Gene set collections

A gene set collection is an attempt to categorize genes into gene sets representing subnetworks of the metabolic network of a cell. The gene sets are most often composed in such a way to reflect a certain cellular function. The genes and gene products which performs the cellular function in question are categorized together, constituting the gene set. There are lots of gene set collections available. In this project three gene set collections were used. Kyoto Encyclopedia of Genes and Genomes (KEGG) contains a collection of pathways of well known cellular functions and metabolic activities [7]. Biocarta is the name of another similar collection of cellular pathways. Biocarta produces gene sets in an open-source fashion, integrating genomic and proteomic data from the scientific community to produce gene set collections covering many cellular functions. The Human Metabolic Reaction (HMR) database is an attempt to collect stoichiometric information of human metabolic reactions into one large generic genome-scale metabolic model [8].

3.4 The DIRAC algorithm

When analysing gene expression between phenotypes a common basic approach is to study individually differentially expressed genes. The idea behind this is of course to identify genes that correlate with and perhaps cause the observed between-phenotype differences. However a systems approach to biology and medicine preferably puts the individual genes in a more global context in the form of metabolic networks. To enable richer understanding of human disease, network based approaches such as gene set enrichment analysis (GSEA) are successfully employed. A GSEA-type of approach identifies pre-defined subnetworks where genes of these subnetworks are collectively co-regulated [9]. Other similar approaches uses single statistics such as mean or median gene expression to compare subnetworks between phenotypes. The aim of these types of methods is to identify differentially expressed subnetworks. However these approaches do not take into account *combinatorial* differences in gene expression, which might have important impact on cellular behaviour.

The DIRAC algorithm is a network based approach like GSEA but specifically designed to take into account these combinatorial effects. For a given gene set the DIRAC algorithm calculates the *order* of the gene expressions in the gene set. The average ordering in one phenotype is called the *rank template* (T) and is simply put a ranked list of gene expressions. The calculation of the rank template is done separately for each phenotype so that each phenotype has its own rank template for each gene set. The rank template is based on that the majority of the samples for each phenotype fits the rank template (rarely the rank template fits each and every sample). Next the *rank matching score* (RMS), which is a measure of how well each sample matches the rank template, is calculated. The average of all samples RMS for each phenotype is called the *rank conservation index* (RCI) and is a measure of the so called *entropy* of the gene set in the phenotype. A highly ordered gene set in a phenotype is said to have low entropy. Conversely a highly disordered gene set in a phenotype is said to have high entropy. To put this in more clear terms a network can be seen as *tightly regulated* if the entropy is low and conversely *loosely regulated* if the entropy is high. This tightness/looseness is however somewhat of an interpretation of what the entropy means for a network, so entropy can be considered a more neutral term. These different designations will both be used throughout this report.

The general outline of the DIRAC algorithm can be seen in figure 1.

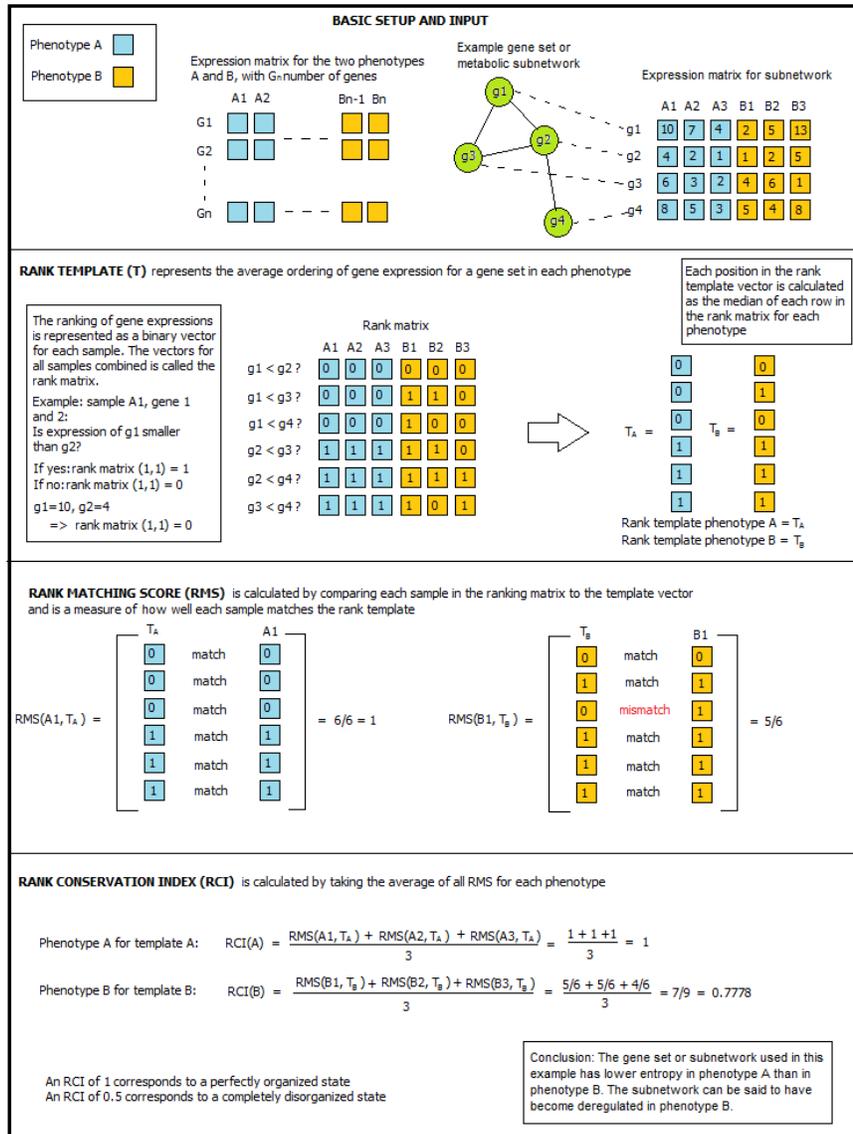


Figure 2: A basic description of the DIRAC algorithm. Rank template, rank matching score and rank conservation index is explained.

The above described procedure importantly first focuses on within-phenotype information. From this analysis however there are subsequently two types of between-phenotype differences to extract (also depicted in figure 2):

1. Identification of networks with differing entropy between phenotypes. Or

put in other words: Identification of tightly regulated networks in one phenotype and loosely regulated in another phenotype. This phenomenon is referred to as *deregulation* in this report.

2. Identification of differently ranked networks between phenotypes, that is networks with differing rank templates. This phenomenon is referred to as *differential regulation* in this report.

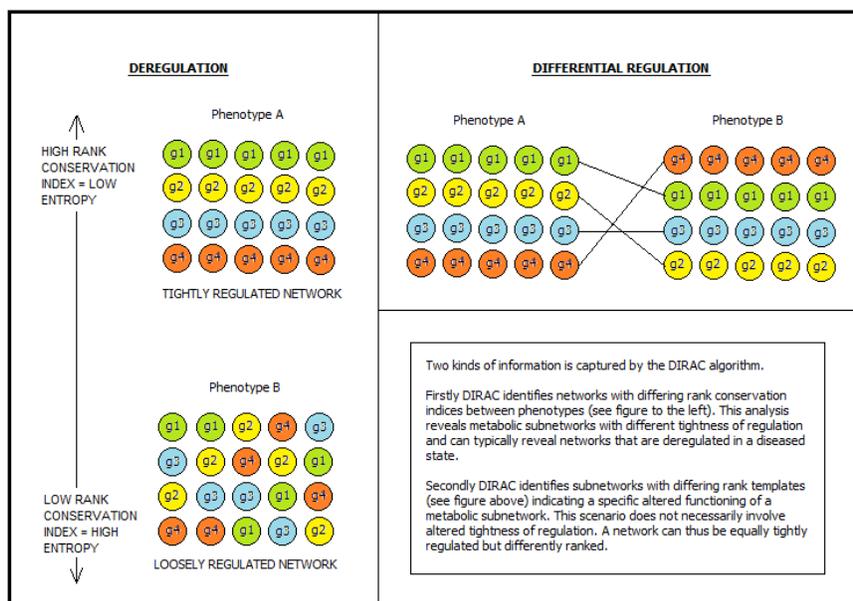


Figure 3: Two types of information can be extracted from the DIRAC algorithm. Depicted in the left hand side of the figure DIRAC can measure the entropy of a network and thus identify deregulated networks. Depicted in the right hand side DIRAC can also identify differently regulated networks as a differing ordering of gene expressions within a network between phenotypes.

The reason for identifying networks with differing entropy is to capture information of the regulation of that network. It can be speculated that dysregulation of a network in a diseased state is accompanied with high entropy of that network. If a network has low entropy it might suggest that the network is under tight regulation due to important cellular function of the network. Loss of importance or loss of function can be correlated with an increase in entropy.

The reason for identifying *differently ranked* networks between phenotypes is to capture information of *altered* cellular regulation of the network. If there is a large difference in ranking of a network between two phenotypes this provides information of what cellular functions is correlated with or even causes disease.

Since a network does not need to be up- or downregulated to have differing entropy, the DIRAC algorithm can provide useful information which could be overlooked by other GSEA-types of algorithms.

In the next subsections further elaboration of interpretations of the results of the DIRAC algorithm follows.

3.4.1 What does deregulation mean?

There are a few interpretations of what low or high rank conservation index for a network means. As defined above a network with low entropy is said to be “tightly regulated”. Conversely a network with high entropy is said to be “loosely regulated”. When a network has gone from tight regulation in one phenotype to loose regulation in another the network can be said to have become “deregulated”. Exactly what biological meaning this has is not always entirely clear and does not necessarily mean the same thing for different networks.

If a subnetwork is tightly regulated it could mean two opposing things. In one situation a subnetwork could be tightly regulated with high gene expression because the network performs some important function requiring constant activity of the pathway. Another situation with tight regulation could be when the function of the network is supposed to be repressed and closely controlled in order for its activity to not “go out of control”. This means that deregulation can possibly both reflect loss of activity and gain of activity.

Another interpretation of deregulation can be plasticity. There might be situations where varying external environments puts adaptive pressure on cells. If for example nutrients and oxygen supply varies there might be systems in the cell that respond to this by varying the activity and thus internal gene expression of certain subnetworks.

3.4.2 Can deregulation reveal frequently mutated subnetworks?

One hypothesis about what it means that a subnetwork is deregulated is related to mutation and selection. If a subnetwork is deregulated in a phenotype, speculatively this might suggest that mutations occurring in this subnetwork poses a selective advantage for the cancer cell. If the subnetwork normally has some function antagonizing cancer cell characteristics, meaning that a well functioning version of the network hinders cancer cell development, then mutations of the subnetwork arguably would be beneficial for the cancer cell. When viewed in this perspective two stereotypical scenarios would occur. In one scenario mutations of only a single specific gene results in a selective advantage. In another scenario mutation of any of the members of the gene subnetwork would result in a selective advantage.

The first scenario is traditionally well studied using established methods such as analysing single-gene somatic mutations. The second scenario would reflect a situation where mutations of different genes gives rise to the same selective advantage. For example, say there is a subnetwork containing five genes. This network performs some function that when lost or altered would

pose a selective advantage for a cancer cell. This selective advantage would emerge if *any* of the five genes would be mutated, no matter which gene is mutated. Now there are five different situations the cancer cell could acquire the selective advantage and if mutation is random then different cancer cells will have different mutations in the same subnetwork. This second scenario would likely result in internal relative variation of gene expression and thus be detected by the DIRAC algorithm as high entropy and deregulation.

Further on this hypothesis implies that the most deregulated subnetworks detected by DIRAC are the networks that when mutated in any way give rise to the most beneficial selective advantage for cancer cells. One could say that the most deregulated networks are under highest so called disruptive selective pressure. Conversely the least deregulated or even more tightly regulated subnetworks in the cancer might reflect the subnetworks with the highest stabilizing selective pressure. This means that any mutation of such a network would result in a selective disadvantage. Interestingly this situation might be revealing of therapeutic targets because the conservation of the function between different cancer cells could indicate importance for cancer cell survival and growth.

3.4.3 What does differential regulation mean?

The other scenario detected by the DIRAC algorithm is not deregulation but *differential regulation*. More specifically this reflects a situation where the rank templates are different between the phenotypes (see above). This means that the genes in a subnetwork are ranked (according to gene expression) in one particular order in one phenotype but ranked in another particular order in the other phenotype. This does not mean that the gene order is shuffled around as in the deregulated scenario.

One interpretation of differential regulation is that the subnetwork in question has acquired altered functioning. It might reflect that a pathway has become altered in a specific way to achieve some specific altered function. Usually one or several genes of the subnetwork are in this scenario differentially expressed. A key difference from deregulation is that the subnetwork is usually (but with exceptions) still tightly regulated in both phenotypes. So a subnetwork is usually not deregulated and differentially regulated at the same time.

The altered functioning can either be in terms of activation or deactivation. The whole subnetwork could be upregulated or downregulated or parts of the network can be up- or down-regulated. One example could be a signaling pathway. In healthy tissue the signaling pathway would perhaps be latent but has become activated by some perturbation of the network in a cancer cell. If this results in differential expression of some of the genes in the network then the rank templates will become different which will be detected by the algorithm.

If a scenario where mutations of specific genes (as described above) leads to a selective advantage for a cancer cell then these specific genes would change the rank templates in a specific way which would again be detected as differential regulation by the algorithm. These kinds of specific alterations might indicate importance for the cancer cell for its ability to survive and/or grow and thus

might reveal possible therapeutic targets.

3.5 Reporter metabolites

The reporter metabolites algorithm is constructed to detect significantly up- or downregulated metabolites between two conditions or phenotypes [10]. The algorithm uses information about which enzymes and metabolites are interacting. An enzyme and a metabolite are defined as interacting if the metabolite is involved in the reaction that the enzyme is catalyzing. From this information enzymes can also be defined as interacting if they share a common metabolite. The resulting network of interacting metabolites and enzymes is then combined with gene expression data of up- and downregulated genes. This combination makes it possible to, based on the gene expression data, estimate which metabolites should also be up- or downregulated. The reporter metabolites algorithm thus needs two pieces of input. As mentioned it needs a list of genes with their corresponding p-values and fold changes, statistics provided by differential expression analysis. It also needs information of which genes interact with which metabolites. This information is provided in the form of a tissue specific genome scale metabolic model (GEM) for HCC generated by the Integrative Network Inference for Tissues (INIT) algorithm [11]. A GEM is a collection of information of all the metabolic reactions connected to the underlying genes coding for the enzymes catalyzing these reactions. GEMs are constructed in a bottom-up fashion meaning gene, protein and metabolic reaction data is integrated and used to reconstruct a metabolic network. Tissue-specific GEMs are useful for interpreting high-throughput data and is being applied to systems medicine in order to better understand human complex disease [12, 13, 14, 15].

4 Results & Discussion

Here the main results from the implementation of the DIRAC algorithm on the TCGA RNA-Seq data is presented. For reasons of space only the most relevant results are presented here. For complete results see Appendix. The relevant results of high-entropy networks, low-entropy networks and differentially regulated networks will separately be presented and discussed. The results of the reporter metabolites algorithm will be presented separately but will be included in the above mentioned discussion. The DIRAC algorithm was implemented not only on HCC but also on six other cancers in order to estimate the average entropy of these cancers. These results are also presented and discussed.

4.1 High-entropy networks in HCC

Here the top 15 deregulated subnetworks from the DIRAC-analysis with three different gene set collections are presented. The three gene set collections are KEGG, Biocarta and HMR.

Table 1: The table contains the KEGG pathways with the highest degree of deregulation, that is the largest difference in entropy between HCC and matched adjacent liver tissue (set as “control”). The name of the pathway, the number of genes in the pathway, the rank conservation indices, the absolute difference in rank conservation indices between phenotypes and the p-value of a difference in rank conservation indices between phenotypes are presented.

Pathway	No. of genes	Rank conservation indices			P-value
		Control	HCC	Difference	
D-glutamine and D-glutamate metabolism	4	0.980	0.782	0.198	$< 6.7 \cdot 10^{-16}$
Caffeine metabolism	7	0.949	0.765	0.183	$< 6.7 \cdot 10^{-16}$
Vitamin b6 metabolism	5	0.918	0.774	0.144	$5.21 \cdot 10^{-6}$
Citrate cycle	27	0.945	0.805	0.140	$5.21 \cdot 10^{-6}$
3 chloroacrylic acid degradation	15	0.963	0.824	0.138	$5.21 \cdot 10^{-6}$
Pantothenate and CoA biosynthesis	16	0.957	0.825	0.131	$2.6 \cdot 10^{-5}$
Cyanoamino acid metabolism	6	0.964	0.838	0.126	$5.73 \cdot 10^{-5}$
Methionine metabolism	17	0.969	0.844	0.125	$5.73 \cdot 10^{-5}$
Glyoxylate and dicarboxylate metabolism	13	0.956	0.832	0.124	$5.73 \cdot 10^{-5}$
Fatty acid metabolism	45	0.949	0.827	0.122	$6.77 \cdot 10^{-5}$
Bile acid biosynthesis	38	0.955	0.834	0.121	$6.77 \cdot 10^{-5}$
Glycine serine and threonine metabolism	44	0.935	0.818	0.117	$7.29 \cdot 10^{-5}$
One carbon pool by folate	16	0.958	0.846	0.112	$9.9 \cdot 10^{-5}$
1 and 2 methylnaphthalene degradation	20	0.965	0.854	0.111	$1.51 \cdot 10^{-4}$

Table 2: The table contains the Biocarta pathways with the highest degree of deregulation, that is the largest difference in entropy between HCC and matched adjacent liver tissue (set as “control”). The name of the pathway, a short description of the pathway, the number of genes in the pathway, the rank conservation indices, the absolute difference in rank conservation indices between phenotypes and the p-value of a difference in rank conservation indices between phenotypes are presented.

Pathway	Short description	No. of genes	Rank conservation indices			P-value
			Control	HCC	Difference	
UREA-CYCLE	Feeding of amino acids into urea cycle	7	0.949	0.731	0.218	$< 6.7 \cdot 10^{-16}$
KREB	TCA-cycle	8	0.960	0.797	0.163	$< 6.7 \cdot 10^{-16}$
TSP1	Angiogenesis	7	0.915	0.758	0.157	$< 6.7 \cdot 10^{-16}$
MALAT-EX	Mitochondria-shuttle of acetyl-groups	8	0.955	0.799	0.156	$< 6.7 \cdot 10^{-16}$
ACET-AMINO-PHEN	Prostaglandin production and liver toxicity	5	0.974	0.820	0.154	$< 6.7 \cdot 10^{-16}$
REELIN	Signaling pathway, actin cytoskeleton	7	0.943	0.798	0.145	$< 6.7 \cdot 10^{-16}$
TERC	Telomerase activity	6	0.929	0.796	0.133	$< 6.7 \cdot 10^{-16}$
MRP	Multi drug resistance	6	0.924	0.792	0.132	$< 6.7 \cdot 10^{-16}$
FBW7	Cell cycle control	8	0.964	0.853	0.111	$< 6.7 \cdot 10^{-16}$
SRCRPTP	Cell cycle progression	9	0.973	0.863	0.109	$< 6.7 \cdot 10^{-16}$
ALTERN-ATIVE	Immune system - cell lysis	6	0.927	0.819	0.107	$< 6.7 \cdot 10^{-16}$
ETC	Electron Transport Chain	9	0.97	0.864	0.107	$< 6.7 \cdot 10^{-16}$
PKC	Activation of protein kinase C	6	0.960	0.854	0.106	$< 6.7 \cdot 10^{-16}$
ARF	Inhibition of ribosomal biogenesis	16	0.946	0.841	0.105	$< 6.7 \cdot 10^{-16}$
CDMAC	Cellular proliferation	15	0.944	0.841	0.103	$< 6.7 \cdot 10^{-16}$

Table 3: The table contains the HMR pathways with the highest degree of deregulation, that is the largest difference in entropy between HCC and matched adjacent liver tissue (set as “control”). The name of the pathway, the number of genes in the pathway, the rank conservation indices, the absolute difference in rank conservation indices between phenotypes and the p-value of a difference in rank conservation indices between phenotypes are presented.

Pathway	No. of genes	Rank conservation indices			P-value
		Control	HCC	Difference	
Biotin metabolism	4	0.9267	0.7570	0.1696	$< 6.7 \cdot 10^{-16}$
Carnitine shuttle (mitochondrial)	8	0.9821	0.8156	0.1665	$< 6.7 \cdot 10^{-16}$
Beta oxidation of di-unsaturated fatty acids (n-6) (peroxisomal)	10	0.9507	0.7989	0.1517	$< 6.7 \cdot 10^{-16}$
C5-branched dibasic acid metabolism	3	0.9867	0.8373	0.1493	$< 6.7 \cdot 10^{-16}$
Beta oxidation of unsaturated fatty acids (n-9) (peroxisomal)	8	0.9564	0.8095	0.1468	$< 6.7 \cdot 10^{-16}$
Bile acid recycling	14	0.9472	0.8113	0.1359	$< 6.7 \cdot 10^{-16}$
Pyruvate metabolism	34	0.9612	0.8287	0.1326	$8.55 \cdot 10^{-6}$
Beta oxidation of poly-unsaturated fatty acids (mitochondrial)	10	0.9533	0.8401	0.1131	$8.55 \cdot 10^{-6}$
Glycine, serine and threonine metabolism	55	0.9478	0.8380	0.1092	$2.56 \cdot 10^{-5}$
Arachidonic acid metabolism	28	0.9472	0.8384	0.1088	$2.56 \cdot 10^{-5}$
Estrogen metabolism	41	0.9402	0.8326	0.1076	$2.56 \cdot 10^{-5}$
Beta oxidation of even-chain fatty acids (peroxisomal)	6	0.9253	0.8201	0.1052	$2.56 \cdot 10^{-5}$
Sulfur metabolism	10	0.9676	0.8641	0.1034	$4.27 \cdot 10^{-5}$
Panhotenate and CoA biosynthesis	15	0.9491	0.8483	0.1008	$4.27 \cdot 10^{-5}$
Beta oxidation of branched-chain fatty acids (mitochondrial)	9	0.9561	0.8561	0.1000	$4.27 \cdot 10^{-5}$

Here results from differential expression analysis of individual genes of some of the most highly deregulated subnetworks in HCC are presented.

Table 4: The table shows results from differential expression analysis. Adjusted p-values and fold changes are presented for the genes in the TCA cycle as defined by KEGG.

TCA cycle (KEGG)		
Gene	Adjusted p-value	\log_2 FC
PCK1	1.07667E-32	-2.261662035
OGDHL	1.66665E-15	-1.355752826
PCK2	1.81166E-14	-1.296023429
ACLY	5.89657E-08	1.144451747
SUCLG2	9.8968E-08	-0.954876884
CLYBL	2.41548E-07	-1.007286189
SDHB	8.42112E-06	-0.835039333
PC	3.91037E-05	-0.809449609
SDHD	6.896E-05	-0.760153507
SDHA	0.000320178	-0.680765938
ACO1	0.000666606	-0.675664688
CS	0.002381327	0.662456888
IDH2	0.011435845	-0.545721606
IDH3G	0.057586969	0.48546214
IDH1	0.140324523	-0.356650998
DLST	0.141507146	-0.318311241
SUCLA2	0.152209105	-0.393693585
OGDH	0.2833108	0.269351421
SUCLG1	0.303284288	-0.232236239
MDH2	0.310737718	0.297669296
DLD	0.414442221	-0.205992328
ACO2	0.488729179	0.196856207
IDH3B	0.504137746	0.215998936
IDH3A	0.687148764	-0.144959285
FH	0.712071128	-0.055884232
MDH1	0.776657314	0.122204916
SDHC	0.851431229	0.078581439

Table 5: The table shows results from differential expression analysis. Adjusted p-values and fold changes are presented for the genes in fatty acid metabolism as defined by KEGG. Only the significantly differentially expressed genes are included.

Fatty acid metabolism (KEGG)		
Gene	Adjusted p-value	log ₂ FC
CYP4A11	4.02525E-41	-2.221676293
ADH4	4.19485E-28	-2.31950626
ACADSB	5.19033E-26	-1.774215507
ACSL1	5.6804E-24	-1.713170139
ACADS	5.96231E-24	-1.753279139
ALDH2	4.3937E-23	-1.621870595
ADH1A	2.06399E-21	-1.72216862
ACAA1	1.20743E-19	-1.541857315
ADH6	4.42987E-19	-1.586158232
ACAA2	8.10946E-17	-1.418732314
GCDH	7.88538E-15	-1.353528542
ADH1B	8.80644E-15	-1.652899762
ACAT1	1.14936E-14	-1.322541384
ACADL	1.207E-14	-1.517445322
ADH1C	3.86931E-14	-1.549310058
ECHS1	2.49468E-12	-1.19600793
EHHADH	3.96515E-12	-1.230875395
CYP4A22	3.02784E-11	-1.914329601
ALDH1B1	3.33338E-10	-1.261794406
ACADM	5.33911E-10	-1.124972993
CPT2	8.62112E-09	-1.034642431
ACOX1	1.16167E-06	-0.904198126
ADHFE1	1.25688E-06	-0.936300147
ALDH9A1	2.57717E-05	-0.824212504
ACSL4	3.83584E-05	2.25612074
HADH	5.63138E-05	-0.755916948
ACADVL	0.000198987	-0.664284483
ALDH1A3	0.000415123	-1.084252322
ALDH3A1	0.000815668	5.645506613
ACSL5	0.001564602	-0.786310653
HADHB	0.002307008	-0.617973862
ACAT2	0.00237431	-0.68151633
ALDH7A1	0.003966976	-0.587986156
HSD17B10	0.022996581	-0.454298314
CPT1A	0.023864082	-0.509708872
ADH5	0.027584368	-0.436588576

Table 6: The table shows results from differential expression analysis. Adjusted p-values and fold changes are presented for the genes in the electron transport chain as defined by Biocarta.

Electron transport chain (Biocarta)		
Gene	Adjusted p-value	\log_2 FC
SDHB	8.42112E-06	-0.835039333
SDHD	6.896E-05	-0.760153507
SDHA	0.000320178	-0.680765938
GPD2	0.036198419	0.566964957
NDUFA1	0.05729895	0.496614018
CYCS	0.082074042	0.429422692
ATP5A1	0.292397509	-0.249363942
UQCRC1	0.548804447	-0.153307937
SDHC	0.851431229	0.078581439
MTCO1	missing	missing

As can be seen above there are lots of highly deregulated or high-entropy networks in HCC. In fact most metabolic subnetworks in HCC are significantly deregulated (see Appendix). This general trend towards high entropy in HCC-metabolism is here called *global deregulation*. This global trend makes it hard to pinpoint specific metabolic subnetworks as being responsible for disease occurrence or progression. In other words to zone in on specific subnetworks or genes is nonsensical when it comes to deregulation in HCC. To analyse and discuss all deregulated networks would be a challenging task.

However there are some immediately noticeable results in terms of deregulation of central metabolic pathways such as the TCA-cycle, electron transport chain and fatty acid metabolism. These pathways stand out because of the central role of the pathways in normal cell metabolism and the high degree of deregulation. Therefore these pathways will be discussed in this report.

4.1.1 TCA cycle

The TCA cycle shows highly significant deregulation both in the KEGG and Biocarta analysis. The differential expression of the individual genes of the KEGG TCA cycle can be seen in Table 10. Out of the 27 genes in the TCA cycle, as defined by KEGG, 13 genes were significantly differentially expressed. Out of these 13 genes 11 were downregulated in HCC and thus only two were upregulated. Interestingly these two genes are citrate synthase (CS) and ATP-citrate lyase (ACLY). CS catalyses the formation of citrate from oxaloacetate and acetyl-CoA and is located in the mitochondrial matrix. ACLY is located in the cytosol and essentially catalyses the inverse reaction: the formation of acetyl-CoA and oxaloacetate from citrate. The net effect of these reactions is the transport of acetyl-CoA from the mitochondria to the cytosol. These reactions compensate for the fact that acetyl-CoA cannot be transported directly over the mitochondrial membrane. ACLY is an important enzyme in fatty acid

biosynthesis because it produces cytosolic acetyl-CoA which is essential for fatty acid biosynthesis.

4.1.2 Electron transport chain

The electron transport chain is another central pathway in energy metabolism which highly significantly deregulation. The pathway, defined by Biocarta, contains 10 genes. Four of these show differential expression, three of which are downregulated. These downregulated genes are three out of the four subunits of succinate dehydrogenase, SDHA, SDHB and SDHD. Succinate dehydrogenases (SDH:s) have been shown to act as tumour suppressors and mutations of both SDHB and SDHD have been linked to cancer formation [16, 17]. According to one theory of how mutation of SDH:s induce cancer, the metabolite succinate accumulate in mitochondria and leaks out to the cytosol where it affects a kind of enzymes called prolyl hydroxylases. This in turn induces resistance to apoptotic signals and can also enhance the Warburg-type of metabolism (for further explanation see subsection 4.1.6) by inducing glycolysis through hypoxia-inducible factor (HIF) [18]. The reporter metabolites analysis also reveals a trend toward accumulation of the metabolite succinate, with a p-value of upregulation for succinate of 0.109 (see Appendix).

4.1.3 Fatty acid metabolism

The KEGG-subnetwork fatty acid metabolism is constituted of 45 genes. This subnetwork is one of the most deregulated subnetworks according to the DIRAC analysis indicating possible dysfunction of the pathway. The differential expression analysis revealed 36 of the 45 genes as significantly differentially expressed. Interestingly 34 of these 36 genes were found to be downregulated in HCC and thus only two were upregulated. The two upregulated genes were found to be Acyl-CoA Synthetase Long-Chain Family Member 4 (ACSL4) and Aldehyde Dehydrogenase 3 Family Member A1 (ALDH3A1). ALDH3A1 catalyses the conversion of an aldehyde to its corresponding carboxylic acid, producing NADPH in the process. It is involved in the metabolism of corticosteroids and lipid peroxidation [19]. ACSL4 catalyses the reaction of free long-chain fatty acids into fatty acyl-CoA esters. ACSL4 plays a key role in fatty acid degradation and lipid biosynthesis [20]. The reporter metabolites analysis revealed many significantly upregulated acyl-CoA esters such as palmitoyl-CoA, propanoyl-CoA, arachidonyl-CoA, linoleoyl-CoA and several others. Interestingly palmitoyl-CoA is needed for sphingolipid biosynthesis (for further discussion see subsection 4.3.1).

4.1.4 Carnitine shuttle and β -oxidation

The HMR subnetworks carnitine shuttle and β -oxidation (of di-unsaturated, unsaturated and polyunsaturated fatty acids in peroxisomes and mitochondria respectively) show highly significant deregulation. The carnitine shuttle is a

system for transport of fatty acids into the mitochondrial matrix for subsequent breakdown into acetyl-CoA via β -oxidation. There are eight genes in the carnitine shuttle as defined by HMR, six of which are significantly differentially expressed. Four are downregulated and two upregulated. All in all there are 65 genes involved in mitochondrial β -oxidation, 42 of which are significantly differentially expressed. Only four of these are upregulated and thus 38 are downregulated in HCC.

From the DIRAC analysis, differential expression and reporter metabolites analysis there are clearly differences in the overall metabolism of fatty acids with deregulation and downregulation as a general rule. As will be discussed (in subsection 4.2.1) fatty acid biosynthesis on the other hand seems to display opposite behaviour with apparent tight regulatory control and upregulation.

4.1.5 Glutamine and glutamate metabolism

The most deregulated KEGG pathway is D-glutamine and D-glutamate metabolism. Besides glucose the amino acid glutamine has been shown to be important for proliferating cells [21]. Cancer cells display an increased uptake of glutamine and it has been known for several decades that many cancer cell lines show a dependence of glutamine for cell growth [22]. It is hypothesized that glutamine is a source of nitrogen for protein and nucleic acid synthesis. A new interesting hypothesis proposes that glutamine is used by cancer cells to neutralize the lactic acid build up from aerobic glycolysis by enzymatically cleaving glutamine to ammonia [23].

All four genes in the pathway is significantly differentially expressed with three down and one upregulated. In the reporter metabolites analysis glutamine is the single most upregulated metabolite with a p-value of 0.0396 while glutamate is significantly downregulated with a p-value of 0.00625. The three methods of analysis: DIRAC, differential expression and reporter metabolites all show abnormal glutamine and glutamate metabolism indicating that HCC cells use these metabolites differently from matched liver cells.

4.1.6 Altered energy metabolism in HCC

The fact that the TCA cycle and electron transport chain is highly significantly deregulated may point to altered energy metabolism in HCC. Under aerobic conditions normal cells break down glucose via glycolysis to pyruvate. The pyruvate then enters the TCA cycle and via the process of oxidative phosphorylation is ultimately converted to carbon dioxide and water generating large amounts of ATP in the process. Under anaerobic conditions normal cells switch to glycolysis for ATP production due to lack of intracellular oxygen. Otto Warburg discovered in the 1930:s that cancer cells use glycolysis even under aerobic conditions. This effect is called “aerobic glycolysis” or the “Warburg effect”. Why cancer cells display this behaviour has been elusive. At a first glance aerobic glycolysis seems inefficient since glycolysis has an 18-fold lower ATP production than oxidative phosphorylation. However one theory suggests that aerobic

glycolysis enables cancer cells to optimize growth. In fact this seems to be a characteristic not only for cancer cells but also for rapidly growing embryonic cells [24]. This suggests that aerobic glycolysis might be employed by all growing cells, not only malignant cells. The mechanism for this enabling of growth is suggested to be due to accumulation of glycolytic intermediates that can be shuttled into various biosynthetic pathways. Supporting this idea is the discovery that signaling pathways involved in proliferation affect metabolic pathways that incorporate nutrients into biomass [25]. It can be hypothesised that ATP production and cell growth cannot be maximized at the same time meaning there is a trade-off between biomass production and ATP production. Cancer cells maximize their biomass production and therefore switches to glycolysis, lowering their ATP production. Another supporting observation to this idea is that certain cancer related mutations enables cancer cells to utilize metabolites in a way that increases biomass production [25]. The TCA cycle in quiescent cells functions as an ATP producer but in a tumor cell the TCA cycle might function more as a source of biosynthetic precursors. Metabolites are constantly taken out of the TCA cycle, a process called cataplerosis, and therefore it needs to be refilled. Glutamine might function as a refiller of the TCA cycle, a process called anaplerosis [26]. As already discussed, glutamine metabolism indeed seems to be different in HCC cells compared to matched liver cells.

The fact that central metabolic pathways such as the TCA cycle is highly deregulated in HCC means that the individual genes of the metabolic pathways have a high degree of internal relative variation. One interpretation of this behaviour is that the pathway is dysfunctional but it might also indicate metabolic plasticity. Some cancers have been found to contain two different subpopulations of cells. One cell-type engages in Warburg-type aerobic glycolysis with breakdown of glucose and subsequent high production of lactate. Another subpopulation has been found to use the lactate produced by the first subpopulation as their main energy source. Part of the TCA cycle is used in these cells to harvest the energy from lactate [27, 28]. Oxygenation has also been shown to fluctuate both from cell to cell and over time in cancer tissue [29]. It can be hypothesized that these two observations may at least partially explain the deregulation of many of the metabolic pathways in HCC. The individual cells in the HCC tissue may have a varying type of metabolism (aerobic glycolysis or lactate metabolism) and a varying degree of nutrient and oxygen supply.

Activation of oncogenes or inactivation of tumor suppressor genes has been linked to altered energy metabolism in cancer cells. The proto-oncogene Myc was shown to activate lactate dehydrogenase A and other enzymes of the glycolytic pathway thereby upregulating glycolysis [30, 31]. Mutated Ras has also been shown to upregulate glycolysis partly due to increased activity of Myc and also increased activity of hypoxia inducible factor, HIF [32, 33]. The protein kinase Akt together with HIF-1 has been shown to increase glycolysis. Mutations of PI3K is a common event in human cancers. PI3K both activates Akt and stabilizes HIF-1 thereby affecting glycolysis. Mutation of tumor suppressor PTEN is also a common event in human cancers [34]. Since PTEN has an opposing effect of PI3K the combined activation of PI3K and loss of PTEN with subsequent

activation of Akt and stabilization of HIF-1 has a profound impact on cellular energy metabolism [35]. Altered energy metabolism seems to be widespread in cancer and might be considered as a hallmark of cancer [36]. In agreement to this the DIRAC algorithm seems to detect large and fundamental differences in energy metabolism of HCC cells. One interpretation of the above observations is to view aerobic glycolysis in cancer cells as just another consequence of the mutated oncogenes and tumor suppressors.

4.1.7 Global deregulation

Interestingly the majority of all metabolic subnetworks in HCC seems to have significantly higher entropy than matched liver tissue. This phenomenon is, as mentioned, here called global deregulation. One plausible reason for this could be cancer cell heterogeneity.

There exists a body of evidence that tumours are not simply masses of uniform cells. There is considerable genetic, epigenetic and phenotypic variations between tumours of the same origin in different patients but also between different regions of the same tumour in one patient [37, 38]. Tumours seem to be heterogenic by nature. There are two theories attempting to explain this heterogeneity. One is called the cancer stem cell model and according to this theory not all cells in a tumour are able to form new tumours, only a subset is (this ability is called tumourigenity). The originating tumourigenic cells are called cancer stem cells (CSCs). According to this model cancer arises from the CSCs and further develops into both tumorigenic and non-tumorigenic pregenitor cells making up a varying degree of the tumour. The heterogeneity of the tumours arises because of differences in originating CSCs. These differences results in a tumour being made up of different subpopulations, of both tumorigenic and non-tumorigenic cells, branching out from different CSCs.

The cancer stem cell model is however debated. Conflicting evidence regarding consistency of CSC markers and varying results from different xenograft models makes the credibility of the model questionable. Another theory suggested by Peter Nowell in 1976 is called *clonal evolution*. According to this theory cancer originates from a single cell and through the process of mutations different subpopulations of cells with different cell characteristics arises. Again the resulting tumour is heterogenic by nature with different cells perhaps evolved to be adapted to different parts of the tumour with different supplies of oxygen and nutrients.

Both theories are not completely mutually exclusive. For example it has been suggested that CSCs undergoes clonal selection thereby giving rise to subsequent different subpopulations of tumour cells.

Whatever the source, cancer cell heterogeneity could be one explanation for the global deregulation detected by the DIRAC algorithm. If this is the case it suggests an interesting use of the DIRAC algorithm, as a quantifier of tumour heterogeneity. In this project liver samples from different patients were analysed. And since the majority of the networks were highly varied and thus deregulated, in the light of the above reasoning, this might suggest that the

tumour samples from the different patients are heterogenous. This does not however mean that the individual tumours are heterogenous. No conclusions of within-tumour heterogeneity can be drawn from this analysis since several samples from the same tumour were not analysed.

Further on, to conclude that the global deregulation measured by DIRAC is solely or even partly caused by tumour heterogeneity is not certain by any means. DIRAC can measure deregulation but the reason behind it can only be speculated upon even if tumour heterogeneity is a plausible contributing factor. If it would be certain that the network-variability is solely caused by heterogeneity then DIRAC could be used to quantify tumour heterogeneity.

4.2 Low-entropy networks in HCC

Here the subnetworks with tighter regulation (higher rank conservation indices) in HCC compared to adjacent liver tissue are presented. Some are non-significantly more tightly regulated but are still included since they are at least not deregulated. These non-significant networks can be said to have equal entropy in HCC and adjacent liver tissue.

Table 7: The table contains the Biocarta pathways with equal or lower entropy in HCC compared to matched adjacent liver tissue (set as “control”). The name of the pathway, a short description of the pathway, the number of genes in the pathway, the rank conservation indices, the absolute difference in rank conservation indices between phenotypes and the p-value of a difference in rank conservation indices between phenotypes are presented.

Pathway	Short description	No. of genes	Rank conservation indices			P-value
			Control	HCC	Difference	
BB CELL	Bystander B-Cell Activation	4	0.913	0.955	-0.0415	0.00375
PLCD	Phospholipid associated cell signaling	4	0.923	0.964	-0.0405	0.00414
ASB CELL	Antigen Dependent B Cell Activation	8	0.952	0.979	-0.0270	0.0204
ION	Ion channel and phorbol esters signaling pathway	4	0.857	0.883	-0.0269	0.0206
CREM	Regulation of Spermatogenesis	7	0.958	0.980	-0.0218	0.0388
EEA1	Vesicle transport	7	0.8838	0.9013	-0.01751	0.0700 (non-significant)
FLUMAZENIL	Cardiac protection against reactive oxygen species	9	0.9072	0.9239	-0.01664	0.0791 (non-significant)

Table 8: The table contains the KEGG pathways with equal or lower entropy in HCC compared to matched adjacent liver tissue (set as “control”). The name of the pathway, the number of genes in the pathway, the rank conservation indices, the absolute difference in rank conservation indices between phenotypes and the p-value of a difference in rank conservation indices between phenotypes are presented.

Pathway	No. of genes	Rank conservation indices			P-value
		Control	HCC	Difference	
Fatty acid biosynthesis	6	0.9187	0.9386	-0.01990	0.0469
Biotin metabolism	4	0.8633	0.8815	-0.01819	0.0584 (non-significant)
Terpenoid biosynthesis	6	0.9120	0.9177	-0.00567	0.4240 (non-significant)

Table 9: The table contains the HMR pathways with equal or lower entropy in HCC compared to matched adjacent liver tissue (set as “control”). The name of the pathway, the number of genes in the pathway, the rank conservation indices, the absolute difference in rank conservation indices between phenotypes and the p-value of a difference in rank conservation indices between phenotypes are presented.

Pathway	No. of genes	Rank conservation indices			P-value
		Control	HCC	Difference	
ROS detoxification	4	0.7733	0.8263	-0.0530	0.0017
Fatty acid biosynthesis (even-chain)	6	0.8867	0.8904	-0.0037	0.65 (non-significant)

Here results from differential expression analysis of individual genes of some of the least deregulated, or even more tightly regulated, subnetworks in HCC are presented.

Table 10: The table shows results from differential expression analysis. Adjusted p-values and fold changes are presented for the genes in fatty acid biosynthesis as defined by KEGG.

Fatty acid biosynthesis (KEGG)		
Gene	Adjusted P-value	\log_2 FC
ACACB	2.72971E-11	-1.225108023
ACACA	2.30459E-05	0.931114131
FASN	0.000476437	1.099825811
OXSM	0.191528364	-0.330962651
OLAH	0.508532433	3.14741506
MCAT	0.678844489	-0.112029586

Table 11: The table shows results from differential expression analysis. Adjusted p-values and fold changes are presented for the genes in terpenoid biosynthesis as defined by KEGG.

Terpenoid biosynthesis (KEGG)		
Gene	Adjusted P-value	\log_2 FC
SQLE	6.81035E-07	1.638904424
FDPS	0.000129089	0.945182536
GGPS1	0.042640054	0.521482392
IDI1	0.175626495	0.365292724
FDFT1	0.540195981	0.178774804
IDI2	0.618654367	0.355486663

As can be seen above there are few networks with intact or lower entropy in HCC. If global deregulation exists in HCC then these few subnetworks with intact or lower entropy in HCC is plausibly informative of what parts of the metabolism is important for the cancer cells. These pathways can be seen as static and conserved in the otherwise chaotic metabolic environment of the cancer cells which is why possible targets of treatment might be found among these pathways.

4.2.1 Fatty acid biosynthesis

Fatty acid metabolism overall and β -oxidation of various fatty acids is highly significantly deregulated in HCC compared to control. Interestingly fatty acid biosynthesis is at the same time equally or even more tightly regulated in HCC compared to matched liver. This might indicate that fatty acid biosynthesis is for some reason important for HCC cells. The differential expression analysis

shows the key enzyme fatty acid synthase (FASN) as significantly upregulated in HCC as well as the important enzyme acetyl-CoA-carboxylase-alpha (ACACA or ACC- α). Acetyl-CoA-carboxylase-beta (ACACB or ACC- β) is in contrast downregulated in HCC. However ACACB may be involved in fatty acid oxidation rather than fatty acid biosynthesis [39].

There are mounting evidence to suggest that altered lipid metabolism is a common feature of cancer cells in general [40]. Human cells have two sources of fatty acids, exogenous fatty acids imported from outside the cell or endogenously produced fatty acids from so called *de novo* lipogenesis. Proliferating embryonic cells seem to be dependent on *de novo* lipogenesis whereas adult quiescent cells prefer exogenous sources for their metabolic demands of fatty acids. Interestingly, like embryonic cells cancer cells also seem to be dependent on *de novo* lipogenesis for their survival and proliferation [40]. For example, even when exogenous sources are abundant, breast cancer cells have been shown to synthesize 95 % of their metabolic demands of fatty acids [40].

Oncogenic signals seem to play a role in regulating lipid metabolism. Common oncogenes and tumour suppressors such as PI3K, MAPK, Myc, EGFR and p53 have been shown to regulate several enzymes involved in fatty acid synthesis. As discussed above oncogenic signals also seem to play a role in the shift to aerobic glycolysis. Taken together this suggests profound metabolic alterations influenced by oncogenes and tumour suppressors.

FASN plays a key role in *de novo* lipogenesis and has been reported as a possible therapeutic target in several cancers [41]. FASN inhibition has been shown to suppress cell growth and induce apoptosis of breast cancer cells in vitro and in vivo [42]. FASN inhibition by siRNAs in prostate cancer cell line LNCaP induced apoptosis of the cancer cells and was shown to not affect cell viability of nonmalignant cells [43]. RNA interference of ACC- α was also reported to induce apoptosis in LNCaP cells without affecting nonmalignant cells. The knockdown of ACC- α resulted in major reduction of cellular pools of palmitic acid. Further on, growth-media supplementation with palmitic acid rescued both FASN inhibited and ACC- α inhibited LNCaP cells from apoptosis [44]. One recent study of FASN inhibition in prostate cancer cells showed prevention of pseudopodia formation and suppressed cell adhesion, migration, and invasion [45]. Fatty acid biosynthesis produces important cellular building blocks for growing cells such as vital components of phospholipids. It is plausible to argue that FASN inhibition results in lipid starvation and membrane dysfunction in growing cancer cells ultimately resulting in reduced growth and/or apoptosis of the cancer cell. However other mechanisms such as disruption of important signaling systems might also contribute to the effect of FASN or ACC- α inhibition [45].

The tight regulation of fatty acid biosynthesis found in this analysis might reflect importance of this pathway for HCC cells. Even if tumour heterogeneity might play a role in the observed global deregulation, meaning the HCC tumours might be made up of several subpopulations of cancer cells with different cellular characteristics, the fatty acid biosynthesis pathway seems to be conserved among these possible subpopulations. It might be speculated that the fatty acid biosynthesis pathway is important for cellular growth, providing im-

portant cellular building blocks in the form of for example phospholipids for cell membrane production. These building blocks might be vital for dividing cells since fatty acids from exogenous sources might not be suitable cellular building blocks. This reasoning however remains speculative but nevertheless might be one plausible explanation for the tighter regulation of fatty acid biosynthesis in HCC.

4.2.2 Terpenoid biosynthesis

The terpenoid biosynthesis pathway has a slightly higher rank conservation index in HCC compared to adjacent liver tissue but not significantly higher. However it is not lower either which makes it a rarely tightly regulated subnetwork worth discussing. The differential expression analysis reveals three significantly differentially expressed genes out of the six genes in total in the pathway. These three genes (SQLE, FDPS and GGPS1) are all upregulated. Terpenoid biosynthesis is thus tightly regulated as well as upregulated indicating importance of this pathway for HCC cells.

The first stage of biosynthesis of cholesterol is the synthesis of the terpenoid backbone. The pathway starts with the molecule mevalonate which is further converted to IPP, FPP and squalene. Squalene is the starting molecule of cholesterol biosynthesis. The widely used cholesterol-lowering drugs *statins* target the terpenoid biosynthesis pathway by inhibiting the production of mevalonate. It has been debated whether statins prevent HCC or not. Two large (n=17380 and n=33413) case-control studies showed a correlation between statin use and decreased risk of HCC and in a dose-dependent manner in one of the studies [46, 47]. However possible confounding effects in concomitant type-2 diabetes medication arguably lowered the credibility of these studies showing a causal relationship between statin use and decreased risk of HCC.

A recent (2013) meta-analysis showed a 37 percent risk-reduction of acquiring HCC in statin-users [48]. A Danish study including the entire Danish cancer population showed a 15 percent decreased risk of overall cancer mortality in statin-users over non-users, with a similar reduction in mortality in the HCC subgroup [49]. The conclusion from these observational studies are that there is a need for randomized controlled trials on statins for the treatment or prevention of HCC.

4.2.3 ROS detoxification

Interestingly detoxification of reactive oxygen species seems to be conserved and even significantly more tightly regulated in HCC which might indicate importance of ROS detoxification for cancer cell survival and growth. Cancer cells have been observed to have increased levels of ROS [50]. In this light the tighter regulation of ROS detoxification can be seen as a response to this. ROS have the ability to damage DNA, lipids and proteins and can therefore be seen as protumorigenic but the increased levels of ROS in cancer cells could pose a disadvantage for the cancer cell and could possibly be a target of therapy [51, 52].

4.3 Differentially regulated networks in HCC

Here the top 10 most differentially regulated subnetworks from the DIRAC-analysis with three different gene set collections are presented. The three gene set collections are KEGG, Biocarta and HMR.

Table 12: The table contains the Biocarta pathways with the highest degree of differential regulation between HCC and matched adjacent liver tissue. The name of the pathway, a short description of the pathway, the number of genes in the pathway and the between-phenotype cross-validation accuracy is presented.

Pathway	Short description of function	No of genes	cross-validation accuracy
ERK	MAP kinase signaling	30	0.95
FMLP	Chemokine gene expression induction	37	0.95
PYK2	Links between pyk2 and MAP kinase	28	0.95
FCER1	Fc Epsilon receptor signaling	37	0.94
TCR	T-cell receptor signaling	42	0.94
HCMV	Cytomegalovirus and MAP kinase signaling	16	0.93
BCR	B-cell receptor signaling	34	0.93
PPARA	Peroxisome proliferator activated receptor alpha	50	0.93
GPCR	G-protein signaling	34	0.92
NFAT	Hypertrophy-like cell growth (enlargment of cells)	52	0.92

Table 13: The table contains the KEGG pathways with the highest degree of differential regulation between HCC and matched adjacent liver tissue. The name of the pathway, the number of genes in the pathway and the between-phenotype cross-validation accuracy is presented.

Pathway	No of genes	cross-validation accuracy
Arrhythmogenic Right Ventricular Cardiomyopathy	74	0.98
Dilated Cardiomyopathy	91	0.98
Hypertrophic Cardiomyopathy	85	0.97
Focal Adhesion	191	0.97
Cell Adhesion Molecules Cams	131	0.97
Glioma	62	0.96
Axon Guidance	126	0.95
Prostate Cancer	86	0.95
Hedgehog Signaling	57	0.95
Neuroactive Ligand Receptor Interaction	237	0.94

Table 14: The table contains the HMR pathways with the highest degree of differential regulation between HCC and matched adjacent liver tissue. The name of the pathway, the number of genes in the pathway and the between-phenotype cross-validation accuracy is presented.

Pathway	No of genes	cross-validation accuracy
Sphingolipid metabolism	105	0.94
Inositol phosphate metabolism	105	0.94
Protein modification	94	0.93
Purine metabolism	98	0.91
Omega-3 fatty acid metabolism	47	0.91
Glycosphingolipid metabolism	58	0.91
Lysine metabolism	48	0.90
Phenylalanine tyrosine and tryptophan biosynthesis	109	0.90
Serotonin and melatonin biosynthesis	52	0.88
Nicotinate and nicotinamide metabolism	44	0.87

Here results from differential expression analysis of individual genes of the most differentially regulated subnetwork in HCC are presented.

Table 15: The table shows results from differential expression analysis. Adjusted p-values and fold changes are presented for the genes in sphingolipid metabolism as defined by HMR. Only the significantly differentially expressed genes are included.

Sphingolipid metabolism		
Gene	PPEE	log ₂ FC
GBA3	7.66349E-32	-2.181856212
PPAP2B	2.11275E-19	-1.562249741
GBA	9.63897E-11	1.385897185
B4GALNT1	3.59245E-10	4.542742434
ALG1L	2.72001E-09	3.52689765
LPIN2	7.2724E-08	-0.953063003
LPPR2	2.70996E-07	1.242605439
GLA	4.2049E-07	1.253746226
SGMS2	1.26603E-06	-1.539511474
KDSR	1.1146E-05	-0.832416802
ST6GALNAC6	3.36003E-05	-0.76646436
ST6GALNAC4	0.000144149	1.158390229
LPPR4	0.000167888	3.070799283
CERS1	0.000168689	4.485069757
B4GALNT2	0.000252071	9.354887267
B4GALT3	0.000390308	0.807767791
B3GALNT2	0.000561763	0.862442278
PPAP2C	0.000650014	2.336046506
FUT2	0.000778422	3.053379167
GAL3ST1	0.000849577	2.618111119
B4GALT1	0.000894711	-0.671887478
A4GALT	0.000903833	1.009492141
B3GNTL1	0.001029273	1.729001144
B3GNT3	0.001908899	2.289709896
B4GALT7	0.002051555	0.745871944
SGPP1	0.002215181	-0.697335782
SMPD1	0.003686074	-0.574009297
B3GALNT1	0.006110679	1.054741943
B3GNT5	0.006464294	0.984094761
GALC	0.012621718	-0.617684678
B4GALNT4	0.013175946	4.211164881
B3GNT4	0.014543646	3.311288353
FUT1	0.019431993	0.909979401
GLB1L	0.021663587	0.810431018
SMPD4	0.022500963	0.497491413
SPTLC3	0.025581457	-0.578257103
DEGS1	0.026413508	0.540086139
ST6GALNAC5	0.028412809	2.726811672
ST6GALNAC2	0.028760463	1.245281581
SMPD3	0.032070762	-1.47761981
CERS5	0.034272562	0.573262191

As can be seen above there are lots of differentially regulated networks in HCC. There are some general conclusions to be drawn from these results. The main theme of the differentially regulated Biocarta networks is cell signaling and cell growth. For example is the well known MAP kinase signaling pathway involved in cell growth and proliferation. Other differentially regulated Biocarta networks involved in cell signaling are HCMV, PYK2, FCER1, GPCR and also TCR. Another noteworthy differentially regulated subnetwork is the peroxisome proliferator activated receptor alpha (PPARA) which is a major regulator of lipid metabolism in the liver [53].

Three of the differentially regulated KEGG subnetworks are involved in hypertrophy-like cell growth (enlargement of cells) as is the Biocarta subnetwork NFAT. The KEGG subnetworks are generally large containing many genes, which makes the interpretation of the results difficult. However different types of cell signaling and cell growth is also here a common occurrence. The three pathways involved in hypertrophy-like cell growth do have considerable overlap. These pathways show involvement of extracellular matrix receptor interactions resulting in alterations of TNF- α , insulin like growth factor-1 (IGF-1) and TGF- β signaling cascades.

As can be seen from the HMR analysis three pathways are involved in lipid signaling in the form of sphingolipid metabolism, inositol phosphate metabolism and glycosphingolipid metabolism. Yet another evidence of alterations in fatty acid metabolism in HCC can be seen in the form of differential regulation of omega-3 fatty acid metabolism. Also the metabolism of purines seems to be different in HCC. As is the metabolism of nicotinate and nicotinamide, possibly indicating altered redox status of the cancer cells. The reporter metabolites algorithm also seem to detect altered concentrations of nicotinate and nicotinamide.

4.3.1 Sphingolipid and glycosphingolipid metabolism

One HMR subnetwork with among the highest degree of differential regulation is sphingolipid metabolism. Glycosphingolipid metabolism is also among the most differentially regulated subnetworks and the two pathways have a fairly high degree of overlapping genes. The two pathways combined contain 116 unique genes, 46 of which are significantly differentially expressed. 32 of these are upregulated and 14 are downregulated.

Sphingolipids are a class of lipids with a so called sphingosine as a base. Sphingosine contains an 18-carbon chain with an amino alcohol attached as a head. Further attachment of a fatty acid makes a so called ceramide molecule. If a sugar is also attached a glycosphingolipid is formed. There are many kinds of sphingolipid and glycosphingolipid molecules.

The biosynthesis of sphingolipids starts with the addition of serine to palmitoyl-CoA. This forms the molecule dehydrosphingosine which is further converted to dihydrosphingosine and ultimately sphingosine. Sphingosine can, as mentioned, later be converted to ceramide and other more complex sphingolipids. Ceramide can also be formed from breakdown of phospholipids and also from conversion

of other sphingolipid species into ceramide, a process termed salvage.

Sphingolipids and glycosphingolipids possibly play a role in cancer occurrence and progression. It has been shown that sphingolipids influence cell cycle progression, telomerase function, cell migration and also stem-cell biology [54]. Two important sphingolipid species are ceramide, as mentioned, and sphingosine-1-phosphate (S1P). S1P is believed to promote cell growth and proliferation, angiogenesis, metastasis and resistance to apoptosis possibly by acting as a ligand for G-protein coupled receptors [55]. Ceramide on the other hand induce apoptosis, cell cycle arrest and autophagic responses and can be considered a powerful tumour suppressor [56]. Cancer cells however may have defects in ceramide metabolism making them resistant to apoptosis. It has been shown that the *de novo* generated ceramides can have opposing effects. Some ceramides have an 18-carbon base while some have a 16-carbon base. Addition of 16-carbon ceramide of head and neck squamous cell carcinoma culture has been shown indicating 16-carbon ceramide cancer promotion [55]. 18-carbon ceramide on the other hand has been shown to have strong tumour suppressor activities, as mentioned above.

The reporter metabolites analysis reveals a trend toward upregulation of the ceramide pool (p-value 0.095). Also another pool of the sphingolipid LacCer (lactosylceramide) show a trend towards upregulation with a p-value of 0.078. In addition palmitoyl-CoA also show a trend toward upregulation with a p-value of 0.07.

In conclusion, the DIRAC algorithm show clear differential regulation of sphingolipid and glycosphingolipid metabolism. The pathways displays some enrichment in the 32 significantly upregulated genes in contrast to the 14 down-regulated which is supported by the reporter metabolites analysis showing some upregulation of ceramide and lactosylceramide. The metabolism of sphingolipids seems according to these analyses to be clearly altered in HCC.

A summary of some of the above discussed results can be seen in figure 4.

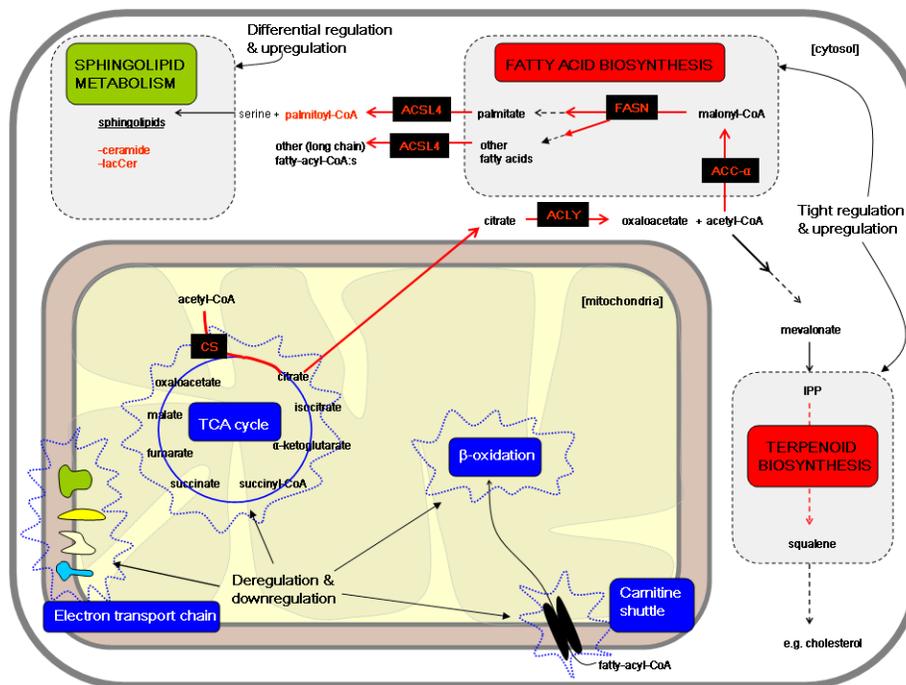


Figure 4: A summary of some of the detected alterations in the metabolic network of HCC. The TCA cycle, the electron transport chain, the carnitine shuttle and β -oxidation are highly deregulated according to the DIRAC analysis. In addition the majority of the genes in these pathways are downregulated. In contrast fatty acid biosynthesis and terpenoid biosynthesis are tightly regulated in HCC with concomittant average upregulation of gene expresison. Sphingolipid metabolism is differentially regulated in HCC compared to matched liver, with a cross-validation accuracy of 98 %. All the genes in the TCA cycle are downregulated except citrate synthase (CS) and ATP-citrate lyase (ACLY) which are responsible for effectively transporting acetyl-CoA from the mitochondrial matrix to the cytosol. The two key regulatory genes of fatty acid biosynthesis acetyl-CoA-carboxylase-alpha (ACC- α) and fatty acid synthase (FASN) are both upregulated. In addition the gene Acyl-CoA Synthetase Long-Chain Family Member 4 (ACSL4) which synthetises fatty-acyl-CoA esters from long chain fatty acids is upregulated in the KEGG pathway *fatty acid metabolism* (including genes of β -oxidation and the carnitine shuttle). This pathway contains 47 genes, 34 of which are downregulated and only two upregulated.

4.4 Reporter metabolites

Here the top 25 most upregulated and downregulated metabolites from the analysis with the reporter metabolites algorithm between HCC and matched liver tissue is presented.

Table 16: Top 25 most upregulated reporter metabolites

Metabolite	p-value
glutamine	0.039571
DNA	0.039627
DNA-5-methylcytosine	0.039627
se-adenosyl-L-selenohomocysteine	0.039627
se-adenosylselenomethionine	0.039627
propanoyl-CoA	0.041204
dUMP	0.043918
UTP	0.047238
arachidonyl-CoA	0.052907
linoleoyl-CoA	0.052907
(13Z)-eicosenoyl-CoA	0.054602
SAH	0.05606
SAM	0.05606
trans-4-hydroxy-L-proline	0.059098
cysteine	0.059332
dTDP	0.059332
glucose	0.059789
(13Z)-eicosenoyl-CoA	0.060877
linoleoyl-CoA	0.060877
arachidonyl-CoA	0.063097
choloyl-CoA	0.063852
S-(11-hydroxy-9-deoxy-delta12-PGD2)-glutathione	0.066594
GDP	0.067878
GTP	0.067878
palmitoyl-CoA	0.069959

Table 17: Top 25 most downregulated reporter metabolites

Metabolite	p-value
O ₂	1.85E-07
NADP+	9.90E-06
NADPH	1.32E-05
H ₂ O	0.00028916
aflatoxin B1	0.00052756
aflatoxin M1	0.00052756
H+	0.00088892
11,12-EET	0.00095339
14,15-EET	0.00095339
(1aalpha,2beta,3alpha,11calpha)-1a,2,3,11c-tetrahydro-6,11-dimethylbenzo[6,7]phenanthro[3,4-b]oxirene-2,3-diol	0.0010616
1,1-dichloroethylene	0.0010616
1-nitronaphthalene	0.0010616
4-[(hydroxymethyl)nitrosoamino]-1-(3-pyridinyl)-1-butanone	0.0010616
7,12-dimethylbenz[a]anthracene 5,6-oxide	0.0010616
7,12-dimethylbenz[a]anthracene	0.0010616
9-hydroxybenzo[a]pyrene	0.0010616
9-hydroxybenzo[a]pyrene-4,5-oxide	0.0010616
aflatoxin M1-8,9-epoxide	0.0010616
aflatoxin Q1	0.0010616
benzo[a]pyrene	0.0010616
benzo[a]pyrene-7,8-dihydrodiol-9,10-oxide	0.0010616
benzo[a]pyrene-7,8-oxide	0.0010616
benzo[a]pyrene-9,10-oxide	0.0010616
bromobenzene	0.0010616
chloral	0.0010616

4.5 Network entropy in other cancers

In table 18 an average rank conservation index of all 186 KEGG pathways can be seen in samples of different tissues. This average rank conservation index is a reflection of the entropy of the average pathway in the metabolic network of these tissues. Higher rank conservation index means lower entropy and tighter regulation of the metabolic network of the tissue in question.

Table 18: The table shows the average rank conservation index for all 186 KEGG pathways for a range of different non-cancerous matched tissues and tumour tissues of different kinds.

Tissue	Average rank conservation index (as a measure of entropy) of all KEGG pathways
Matched liver	0.9459
Matched colon	0.9425
Matched lung	0.9400
Matched kidney	0.9345
Matched breast	0.9248
Prostate cancer	0.9240
Matched prostate	0.9235
Colon cancer	0.9107
Kidney chromophobe cancer	0.9056
Breast cancer	0.8992
Matched bladder	0.8978
Lung cancer (small cell)	0.8928
Hepatocellular carcinoma	0.8884
Bladder cancer	0.8795

Table 19: Difference in the average network entropy for each cancer versus the corresponding tissue compared with the mortality rate of each cancer.

Cancer/matched tissue	Difference in average rank conservation index (difference in entropy)	Percentage of patients deceased within five years of diagnosis [57]
Liver vs HCC	0.0575	88.5 %
Lung vs lung cancer	0.0472	82.9 %
Colon vs colon cancer	0.0318	35.0 %
Kidney vs kidney cancer	0.0285	27.6 %
Breast vs breast cancer	0.0256	10.8 %
Bladder vs bladder cancer	0.0184	22.6 %
Prostate vs prostate cancer	-0.0005	1.4 %

As can be seen above the average network entropy varies between tissues. This average network entropy can be viewed as a measure of the global degree of entropy in the metabolism of the different tissues. In a highly ordered tissue with a static environment and tightly controlled metabolism one would expect to see low average entropy (high average rank conservation index). Interestingly

there seems to be a trend towards high entropy in cancerous tissue compared with matched healthy tissue. Only for prostate cancer the average entropy in cancerous tissue is lower than in matched tissue. For all other tissues the inverse is true. In other words there seems to be a trend towards global deregulation in other cancers besides HCC. However HCC displays the highest degree of global deregulation of all other cancer types.

Interestingly the degree of global deregulation seems to correlate with the 5-year mortality rate indicating that entropy correlates with malignancy of the cancer. If this is true for other cancers or not is unknown. Also the analysis was done only on the KEGG gene set collection, if the correlation holds true on other gene set collections or not is also unknown.

4.6 Criticism

When it comes to the DIRAC algorithm there are some considerations to have in mind in order to avoid drawing faulty conclusions. A network can have low entropy not because there is little variation in individual gene expression but because the average difference in gene expression of the members of the subnetwork is large. For example if a network contains three genes, one with average expression of 0.5, one with average expression of 100 and a third with average expression of 10000 the chance of the entropy being other than low is very small. Related to this there is also a tendency for small networks to have low entropy because the chance of any two genes having similar average expression values increases with increasing number of genes in the defined gene set. In other words large subnetworks are more prone to high entropy and small subnetworks are more prone to low entropy. This means that the DIRAC algorithm can miss true deregulation for small subnetworks and faulty classify them as tightly regulated. However if the rank conservation index is not equal to 1 but well below 1 (for example 0.8 or 0.9) in the control tissue and the rank conservation index in the diseased tissue is equal or higher then this mistake is less likely to occur. This is because the entropy in the control tissue is then clearly not as low as it could get and thus proves the existence of internal variation in gene ranking.

The HCC-patients used in this analysis varied in several aspects such as age, race, sex, concomitant conditions, stage of disease and risk-factors. Thus the group was heterogenous. To say that this is the source of variation causing the high entropy in HCC is however probably not correct for the simple reason that the entropy would then also be high in the matched control-samples. However the patients where the matched liver samples were taken from could possibly be different from the whole group in certain aspects. This would then be a confounding factor decreasing the reliability of the results of both the DIRAC analysis and the differential expression- and reporter metabolites analysis.

5 Conclusions

In conclusion the DIRAC algorithm provides evidence of high degree of deregulation of central metabolic pathways in HCC such as the TCA cycle, the electron transport chain and fatty acid metabolism (see figure 3). In addition the differential expression analysis revealed concomitant alterations in gene expression of these pathways. There seems to be a correlation between deregulation of a pathway and average downregulation of the genes in that pathway. Conversely there seems to be a correlation between tighter regulation of a pathway and average upregulation of the genes in that pathway.

The DIRAC algorithm provides evidence for so called *global deregulation* in HCC. This means that the average entropy of the metabolic pathways in HCC is higher than in matched liver tissue. In addition a positive correlation was found between the degree of global deregulation and the 5-year mortality rate

for seven different cancers.

In contrast the metabolic pathways with equal or tighter regulation in HCC are scarce. These include fatty acid biosynthesis, terpenoid biosynthesis and ROS detoxification. The fact that these pathways have intact entropy in HCC might indicate that they are important for growth and survival of the HCC cells. Thus there is some rationale behind targeting these pathways for treatment of HCC.

Some pathways also display so called *differential regulation*. These pathways seem to be altered in HCC, not in terms of entropy but in terms of altered internal ordering of gene expression of the pathways. One example of such a pathway is sphingolipid metabolism.

One speculation of the implications of deregulation regards drug targets. As mentioned there is logic behind targeting tightly regulated pathways in HCC. However targeting a deregulated subnetwork might be nonsensical exactly because it is deregulated. Speculatively the reason behind the high entropy could be differences in the so called cancer-driving mutations between different patients. If this is the case one drug might have an effect for one patient but not for another patient. However this makes a case for personalized medicine. If the driver-mutations are known, patients could hypothetically be classified into different sub-categories with different treatment strategies.

6 Materials & Methods

Here the implementation of the DIRAC algorithm is described as is the retrieval of the TCGA RNA-Seq data and the differential expression analysis. Some information about the patients and the tissue samples are included, however further information can be found in the Appendix.

6.1 Implementation of DIRAC

The implementation of the DIRAC algorithm and subsequent calculations described below was done using Matlab. The source code files are freely available at Price Lab Institute for Systems Biology’s webpage [58]. The main detailed analysis is done only on HCC but the DIRAC algorithm was also implemented on some other cancer types with corresponding matched tissue. The purpose of this was to compare the global average pathway entropy of these cancers and in the other matched tissues. This analysis puts the high and low entropy in HCC “on a scale” since a comparison with the other cancer types can be made.

6.1.1 Rank conservation indices

To calculate the tightness of regulation of a gene set the so called rank template T is constructed for each phenotype. T is constructed so that the majority of samples in the phenotype matches T . The rank matching score RMS can then be calculated. RMS is a measure of how well each sample matches the template T . The average of RMS in a phenotype is called the rank conservation index RCI. RCI is thus a measure of how well all the samples in a phenotype matches the rank template T . The values of RCI can range between 0.5 and 1.0 where 0.5 represents a completely disorganised state and 1.0 represents a completely ordered state. In other words if all the samples in a phenotype perfectly matches T the rank conservation index is 1.0. By comparing RCI between the phenotypes for a given gene set differences in entropy can be detected.

6.1.2 Rank difference score

To detect difference in regulation as opposed to difference in entropy the rank matching scores can be compared between the phenotypes. Say there are two phenotypes A and B. The rank matching score for sample n of phenotype A and gene set m is called $RMS(m, A)(n)$. By also calculating the rank matching score $RMS(m, B)(n)$ even if sample n does not belong to phenotype B one can capture differences in ranking of a gene set between the phenotypes. The rank difference score Δ is thus defined as: $\Delta = RMS(m, A)(n) - RMS(m, B)(n)$. If the gene set m is tightly regulated in both A and B but differently regulated the templates $T_{m,A}$ and $T_{m,B}$ should be different. And thus if sample n belongs to phenotype A $RMS(m, A)(n)$ should be high and $RMS(m, B)(n)$ should be low which then is captured as a high rank difference score. To get a measure of accuracy of the differentially regulated pathways 5-fold cross-validation was also performed.

6.1.3 Significance testing of deregulation

To test the significance level of any supposed deregulation the following procedure was performed. First samples of a gene set were randomly re-assigned to one of the phenotypes. Secondly the rank conservation indices were calculated for both phenotypes and compared between the phenotypes. These steps were repeated 1000 times to get a null distribution of differences in rank conservation indices between the phenotypes. The difference in rank conservation indices of the correctly labeled phenotypes could then be compared to this null distribution in order to calculate p-values of deregulation. The above was done for all gene sets using Matlab.

6.2 RNA-Seq data retrieval

The RNAseq data was downloaded from the TCGA data base using a TCGA-assembler package in R. Normalized RNA-SeqV2 data from 49 liver control samples and 163 HCC samples was downloaded. Raw count data from the same samples was also downloaded and used to perform the differential expression analysis. Normalized RNA-SeqV2 data from the other cancer-types were also downloaded from the TCGA data base using this method. For more details on how the retrieval of the TCGA RNA-Seq data was achieved see the TCGA-assembler manual [59].

6.3 Patients and clinical data

The samples were collected from HCC tumour tissue and from adjacent non-tumour tissue. The number of samples were in total 212 originating from 163 different individuals. Thus 163 tumour tissue samples and 49 adjacent non-tumour tissue samples were used in this analysis. The median and average age of the patients were 65 and 62.2 years respectively. There were 89 male patients, 52 female (and 22 where this information was not available). The race of the patients varied, with 107 white, 10 black, 30 asian, 2 native american, 5 other and 9 unknown. The patients clinical situation varied in several aspects. Some patients had liver fibrosis or even cirrhosis while others did not. 35 patients had hepatitis B or hepatitis C. Some patients had hepatic inflammation of varying degree. The patients were also in different stages of disease ranging from stage I to stage IVB. Alcohol consumption varied with 39 patients categorized as alcohol consumers in the extent that it was counted as a life style risk factor. 5 patients had non-alchololic fatty liver disease. For more detailed information on clinical data see supplementary information in Appendix.

Detailed information of the samples from the six other cancer-types is not available in this report. However the samples from all cancers are taken from tumour tissue and adjacent matched healthy tissue. The number of samples and matched controls are presented in the table below.

Table 20: The number of tumour- and matched control tissue samples of the six other cancers analysed in this project.

TCGA-code of cancer type	Trivial name	Number of matched control samples	Number of tumour tissue samples
BLCA	Bladder cancer	19	231
BRCA	Breast cancer	104	146
COAD	Colon cancer	41	209
KICH	Kidney cancer	25	66
LUSC	Lung cancer (small cell)	50	200
PRAD	Prostate cancer	50	200

6.4 Differential expression

The differential expression analysis was performed using the R-package DESeq. This is a package designed for differential expression of RNA-Seq data. As input DESeq takes a data matrix called a count table where the rows are the genes and the columns the samples/patients. The data was given in the form of raw counts of sequencing reads. This is important since normalized values will give nonsensical results. Metadata in the form of information of which samples belongs to which phenotype was also given to the DESeq algorithm. Normalization and variance estimation was performed according to the manual before subsequent differential expression analysis. For more information on the DESeq package see the manual [60].

References

- [1] American Cancer Society, last updated: 2014, <http://www.cancer.org/cancer/livercancer/detailedguide/liver-cancer-what-is-liver-cancer>
- [2] Altekruse, S.F., McGlynn, K.A., Reichman, M.E., (2009), *Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005*, J Clin Oncol., 27(9):1485.
- [3] Perz, J.F., Armstrong, G.L., Farrington, L.A., Hutin, Y.J., and Bell, B.P., (2006) *The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide*, J Hepatol, 45(4):529.
- [4] American Cancer Society, last updated: 2014, <http://www.cancer.org/cancer/livercancer/detailedguide/liver-cancer-treating-general-info>
- [5] Nagalakshmi, U., Waern, K. & Snyder, M., (2010) *RNA-Seq: A Method for Comprehensive UNIT 4.11 Transcriptome Analysis*, Molecular, Cellular, and Developmental Biology Department, Yale University, New Haven, Connecticut, Current Protocols in Molecular Biology 4.11.1-4.11.13, doi: 10.1002/0471142727.mb0411s89
- [6] The Cancer Genome Atlas, Understanding genomics to improve cancer care, National Cancer Institute & National Human Genome Research Institute, <http://cancergenome.nih.gov/abouttcga/overview>
- [7] KEGG: Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>
- [8] Human Metabolic Atlas, <http://www.metabolicatlas.com/>, Chalmers University of Technology, Department of Chemical and Biological Engineering
- [9] Subramanian, A., Tamayo, P., Mootha, K.V., Mukherjee, S., Ebert, L.B., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. & Mesirov, J.P., (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, vol. 102 no. 43, doi: 10.1073/pnas.0506580102
- [10] Patil, K.R. & Nielsen, J., (2005), *Uncovering transcriptional regulation of metabolism by using metabolic network topology*, Proceedings of the National Academy of Sciences of the United States of America, vol. 102 no. 8, doi: 10.1073/pnas.0406811102
- [11] Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I. & Nielsen, J., (2012), *Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT*, PLOS Computational Biology, doi: 10.1371

- [12] Mardinoglu, A. & Nielsen, J., (2012), *Systems medicine and metabolic modelling*. Journal of Internal Medicine, 271: 142–154. doi: 10.1111/j.1365-2796.2011.02493.x
- [13] Mardinoglu, A., Gatto, F. & Nielsen, J., (2013), *Genome-scale modeling of human metabolism – a systems biology approach*. Biotechnology Journal, 8: 985–996. doi: 10.1002/biot.201200275
- [14] Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M & Nielsen, J., (2013), *Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease*, Nature Communications 5, Article number: 3083, doi:10.1038/ncomms4083
- [15] Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Nookaew, I., Jacobson, P., Walley, A.J., Froguel, P., Carlsson, L.M., Uhlen, M. & Nielsen, J., (2013), *Integration of clinical data with a genome-scale metabolic model of the human adipocyte*, Molecular Systems Biology 9, Article number 649, doi:10.1038/msb.2013.5
- [16] GeneCards, The Human Gene Compendium, Weizmann Institute of Science, <http://www.genecards.org/cgi-bin/carddisp.pl?gene=SDHB>
- [17] GeneCards, The Human Gene Compendium, Weizmann Institute of Science, <http://www.genecards.org/cgi-bin/carddisp.pl?gene=SDHD>
- [18] King, A., Selak, M.A. & Gottlieb, E., (2006), *Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer*, Cancer Research UK, The Beatson Institute for Cancer Research, Glasgow, UK, Oncogene 25, 4675–4682. doi:10.1038/sj.onc.1209594
- [19] GeneCards, The Human Gene Compendium, Weizmann Institute of Science, <http://www.genecards.org/cgi-bin/carddisp.pl?gene=ALDH3A1&search=ALDH3A1>
- [20] GeneCards, The Human Gene Compendium, Weizmann Institute of Science, <http://www.genecards.org/cgi-bin/carddisp.pl?gene=ACSL4&search=ACSL4>
- [21] Muñoz-Pinedo, C., Mjiyad, N. El & Ricci, J-E., (2012), *Cancer metabolism: current perspectives and future directions*, Cell Death Dis. Jan 2012, Jan 12 2012. doi: 10.1038/cddis.2011.123
- [22] Eagle, H., Oyama, V.I., Levy, M., Horton, C.L. & Fleischman, R., (1956), *The growth response of mammalian cells in tissue culture to L-glutamine and L-glutamic acid*, J. Biol. Chem., 218:607-616.
- [23] Huang, W., Choi, W., Chen, Y., Zhang, Q., Deng, H., Wei, H., & Yigong, S., (2013), *A proposed role for glutamine in cancer cell growth through acid resistance*, Cell Res; 23(5): 724–727. doi: 10.1038/cr.2013.15

- [24] Krisher, R.L., Prather, R.S., (2012), *A role for the Warburg effect in preimplantation embryo development: metabolic modification to support rapid cell proliferation*, Mol Reprod Dev., doi: 10.1002/mrd.22037.
- [25] Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B., (2009), *Understanding the Warburg effect: the metabolic requirements of cell proliferation*, Science, doi: 10.1126/science.1160809.
- [26] DeBerardinis, R.J., Lum, J.J., Hatzivassiliou, G., Thompson, C.B., (2008), *The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation*, Cell Metabolism, doi: <http://dx.doi.org/10.1016/j.cmet.2007.10.002>
- [27] Kennedy, K.M., Dewhirst, M.W., (2010), *Tumor metabolism of lactate: the influence and therapeutic potential for MCT and CD147 regulation*, Future Oncol., doi: 10.2217/fon.09.145
- [28] Semenza, G.L., (2008), *Hypoxia-inducible factor 1 and cancer pathogenesis*. IUBMB Life, 60: 591–597. doi: 10.1002/iub.93
- [29] Hardee M.E, Dewhirst M.W., Agarwal N. & Sorg B.S., (2009), *Novel imaging provides new insights into mechanisms of oxygen transport in tumors*, Curr. Mol. Med. 9, 435–441, .
- [30] Shim, H., Dolde, C., Lewis, B.C., Wu, C.S., Dang, G., Jungmann, R.A., Dalla-Favera, R. & Dang, C.V., (1997), *c-Myc transactivation of LDH-A: implications for tumor metabolism and growth*. Proc. Natl. Acad. Sci. USA, 94:6658–6663. doi:10.1073/pnas.94.13.6658
- [31] Dang, C.V., O'Donnell, K.A., Zeller, K.I., Nguyen, T., Osthus, R.C. & Li, F., (2006), *The c-Myc target gene network*. Semin Cancer Biol, 16: 253–264
- [32] Sears, R., Leone, G., DeGregori, J., Nevins, J.R., (1999), *Ras enhances Myc protein stability*. Mol Cell 3: 169–179.
- [33] Semenza, G.L., (2010) *HIF-1: Upstream and downstream of cancer metabolism*. Curr Opin Genet Dev 20: 51–56.
- [34] Yamada, K.M. & Araki, M., (2001), *Tumor suppressor PTEN: modulator of cell signaling, growth, migration and apoptosis*, Journal of Cell Science 114, 2375-2382.
- [35] Elstrom, R.L., Bauer, D.E., Buzzai, M., Karnauskas, R., Harris, M.H., Plas, D.R., Zhuang, H., Cinalli, R.M., Alavi, A., Rudin, C.M., et al., (2004), *Akt stimulates aerobic glycolysis in cancer cells*. Cancer Res. 64:3892–3899
- [36] Hanahan, D. & Weinberg, R.A., (2011), *Hallmarks of Cancer: The Next Generation*, CellPress, Volume 144, Issue 5, p646–674, doi: <http://dx.doi.org/10.1016/j.cell.2011.02.013>

- [37] Meacham, C.E. & Morrison, S.J., (2013), *Tumour heterogeneity and cancer cell plasticity*, Nature 501, 328–337, doi:10.1038/nature12624
- [38] Burrell, R., McGranahan, N., Bartek, J. & Swanton, C., (2013), *The causes and consequences of genetic heterogeneity in cancer evolution*, Nature 501, 338–345, doi:10.1038/nature12625
- [39] GeneCards, The Human Gene Compendium, Weizmann Institute of Science, <http://www.genecards.org/cgi-bin/carddisp.pl?gene=ACACB&search=ACACB>
- [40] Zhang, F. & Du, G., (2012), *Dysregulated lipid metabolism in cancer*, World J Biol Chem. 26;3(8):167-74. doi: 10.4331/wjbc.v3.i8.167.
- [41] Mashima, T., Seimiya, H. & Tsuruo, T., (2009), *De novo fatty-acid synthesis and related pathways as molecular targets for cancer therapy*, British Journal of Cancer, 1369–1372. doi:10.1038/sj.bjc.6605007
- [42] Pizer, E.S., Thupari, J., Han, W.F., Pinn, M.L., Chrest, F.J., Frehywot, G.L., Townsend, C.A. & Kuhajda, F.P., (2000), *Malonyl-coenzyme-A is a potential mediator of cytotoxicity induced by fatty-acid synthase inhibition in human breast cancer cells and xenografts*, Cancer Res., 15;60(2):213-8.
- [43] De Schrijver, E., Brusselmans, K., Heyns, W., Verhoeven, G. & Swinnen, J.V., (2003), *RNA interference-mediated silencing of the fatty acid synthase gene attenuates growth and induces morphological changes and apoptosis of LNCaP prostate cancer cells*. Cancer Res 63: 3799–3804. doi: 10.1007/0-387-23761-5-33
- [44] Chajès, V., Cambot, M., Moreau, K., Lenoir, G.M., & Joulin, V., (2006), *Acetyl-CoA carboxylase alpha is essential to breast cancer cell survival*, Cancer Res., 66: 5287–5294
- [45] Yoshii, Y. et al., (2013), *Fatty Acid Synthase Is a Key Target in Multiple Essential Tumor Functions of Prostate Cancer: Uptake of Radiolabeled Acetate as a Predictor of the Targeted Therapy Outcome*, PLOS one, doi: 10.1371/journal.pone.0064570
- [46] Lai, S.W., Liao, K.F., Lai, H.C., Muo, C.H., Sung, F.C. & Chen, P.C., (2013), *Statin use and risk of hepatocellular carcinoma*, Eur J Epidemiol. 28(6):485-92. doi: 10.1007/s10654-013-9806-y
- [47] Tsan, Y.T., Lee, C.H., Wang, J.D., Chen, P.C., (2012), *Statins and the risk of hepatocellular carcinoma in patients with hepatitis B virus infection*, J Clin Oncol., 30(6):623-30. doi: 10.1200/JCO.2011.36.0917
- [48] Singh, S., Singh, P.P., Singh, A.G., Murad, M.H. & Sanchez W., (2013), *Statins are associated with a reduced risk of hepatocellular cancer: a systematic review and meta-analysis*, Gastroenterology;144(2):323-32. doi: 10.1053/j.gastro.2012.10.005

- [49] Nielsen, S.F., Nordestgaard, B.G. & Bojesen S.E., (2012), *Statin Use and Reduced Cancer-Related Mortality*, N Engl J Med 2012; 367:1792-1802 November 8, 2012
- [50] DeNicola, G.M., et al, (2011), *Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis*, Nature 475, 106–109, doi:10.1038/nature10189
- [51] Gorrini, C., Harris, I.S. & Mak, T.W., (2013), *Modulation of oxidative stress as an anticancer strategy*, Nature Reviews Drug Discovery 12, 931–947, doi:10.1038/nrd4002
- [52] Nogueira, V. & Hay, N., (2013), *Molecular Pathways: Reactive Oxygen Species Homeostasis in Cancer Cells and Implications for Cancer Therapy*, Clin Cancer Res, 19; 4309
- [53] Kersten, S., Desvergne, B. & Wahli, W., (2000), *Roles of PPARs in health and disease* Nature 405, 421-424, doi:10.1038/35013000
- [54] Oskouian, B. & Saba, J.D., (2010), *Cancer treatment strategies targeting sphingolipid metabolism*, Adv Exp Med Biol., 688:185-205
- [55] Ponnusamy, S., Meyers-Needham, M., Senkal, C.E., Saddoughi, S.A., Sentele, D., Selvam, S.P., Salas, A., Ogretmen, B., (2010), *Sphingolipids and cancer: ceramide and sphingosine-1-phosphate in the regulation of cell death and drug resistance*, Future Oncol., 6(10):1603-24. doi: 10.2217/fon.10.116.
- [56] Morad, S.A. & Cabot, M.C., (2013), *Ceramide-orchestrated signalling in cancer cells*, Nature Reviews Cancer 13, 51–65, doi:10.1038/nrc3398
- [57] National Cancer Institute, SEER Cancer Statistics Review 1975-2011, Age-Adjusted SEER Incidence and U.S. Death Rates and 5-Year Relative Survival (2004-2010), link: http://seer.cancer.gov/csr/1975_2011/results_merged/topic_survival.pdf
- [58] Price Lab, Institute for Systems Biology, last updated: 2012, <https://price.systemsbiology.net/differential-rank-conservation-dirac>
- [59] Center for Biomedical Research Informatics, NorthShore University HealthSystem, Department of Health Studies, The University of Chicago, Chicago, *TCGA-Assembler User Manual*, link: <http://health.bsd.uchicago.edu/yji/TCGA-Assembler-files/TCGA-Assembler%20User%20Manual.pdf>, Feb 4 2014
- [60] Anders, S., Huber, W., (2013), *Differential expression of RNA-Seq data at the gene level – the DESeq package*, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, link: <http://bioconductor.org/packages/release/bioc/vignettes/DESeq/inst/doc/DESeq.pdf>, Last revision: 2013-02-24