Por último tendo em vista a possível construção de uma rede metabólica, realizou-se o agrupamento de genes segundo a sua função. O primeiro passo foi a criação de um ficheiro onde cada linha representa um gene e a sua função (funções.txt), este ficheiro foi criado a partir do script funções.py, que retira essa informação do ficheiro our\_genome.gb. No ficheiro está presente a informação sobre quantos genes têm função conhecida e quantos têm função desconhecida, bem como um gene e a sua função, ou caso não tenha função conhecida 'Sem função conhecida!', um gene por linha.

De seguida tomando como partida o ficheiro criado anteriormente (funções.txt), procedeu-se a partir do script func\_excel.py à criação de um ficheiro Excel (func.xlsx) com uma tabela onde a cada gene foi atribuída uma cor conforme a sua função.

De modo a obter informação acerca de todos os genes com função metabólica e visto a quantidade deste ser bastante elevada, procedeu-se à criação de um script (KEGG.py) para a criação de uma tabela em Excel (metabolic\_proteins\_python.xlsx), os dados desta tabela foram retirados de forma automática da base de dados KEGG com recurso ao biopython. De seguida procedeu-se à formatação da tabela de forma manual e a tabela final dos genes envolvidos no metabolismo é a metabolic\_proteins\_python\_formated.xlsx. De forma a ser usado de seguida, no script KEGG.py foi também criado um ficheiro de texto (metabolic\_networks.txt) com cada gene e as vias metabólicas conhecidas onde cada um está envolvido.

Assim o ficheiro metabolic\_networks.txt foi usado pelo script networks.py para calcular o numero de genes envolvidos em cada via metabólica identificada.

As três vias onde mais genes da nossa zona do genoma atuam são:

Na via com id da KEGG lpn01100 atuam 43 dos nossos genes.

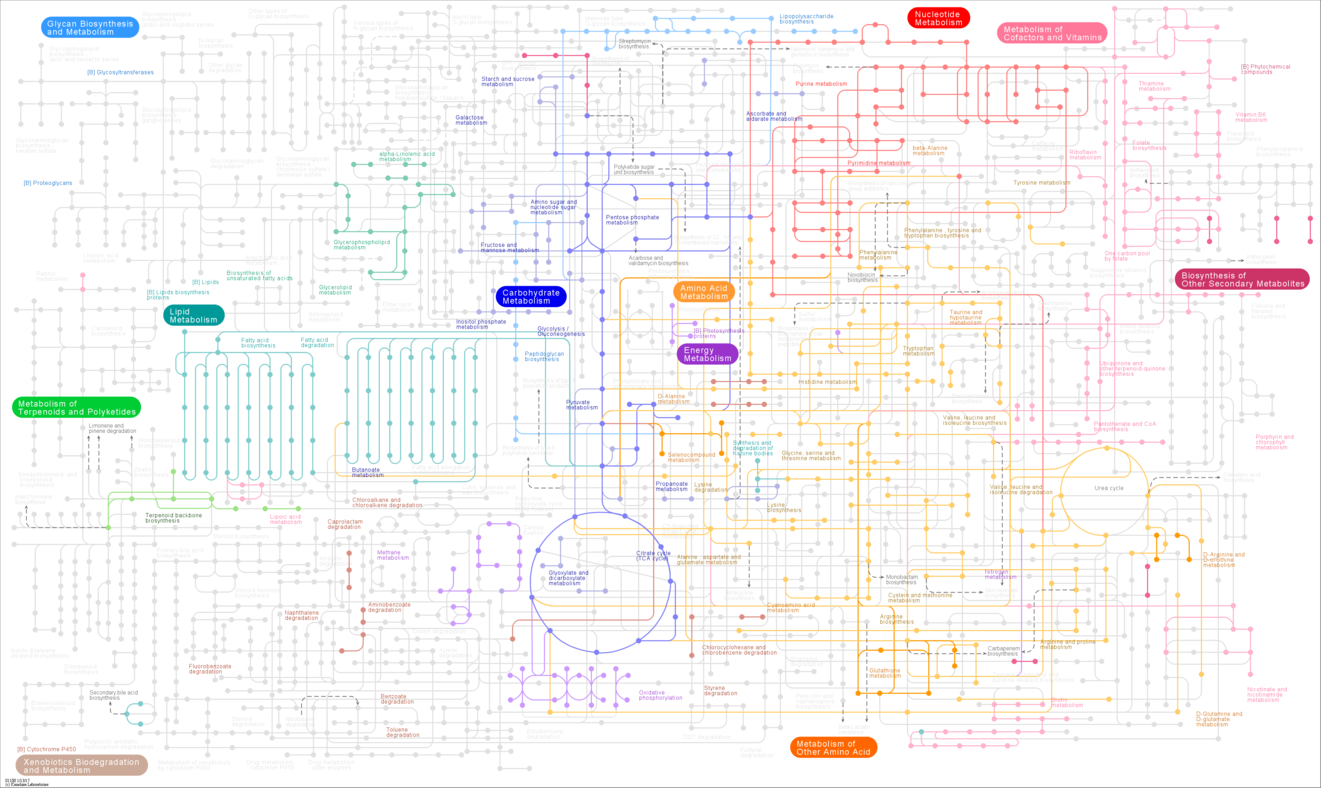


Figura 1-Via lpn01100

Na via com id da KEGG lpn01110 atuam 16 dos nossos genes.

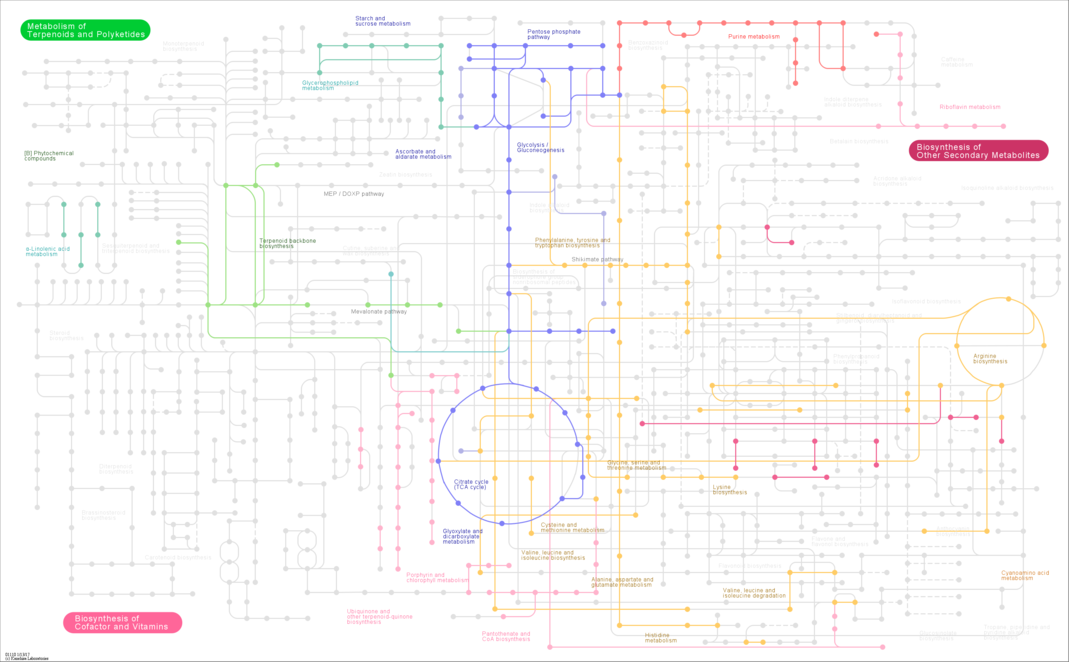


Figura 2- Via lpn01110

Na via com id da KEGG lpn01130 atuam 16 dos nossos genes.

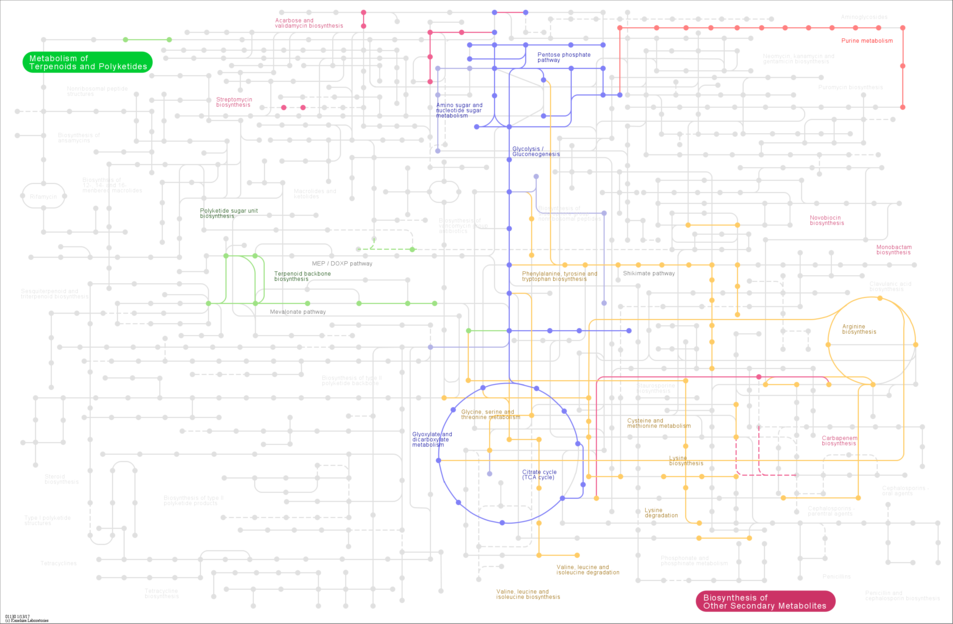


Figura 3- Via lpn01130

No que diz respeito às proteínas com função transportadora, visto serem menos (19), procedeu-se à recolha de informação de forma manual de várias bases de dados. Os resultados dessa recolha encontram-se no ficheiro transport\_proteins.xlsx.

Por ultimo a informação relativa às proteínas com função de fatores de transcrição foi também recolhida de forma manual e encontra-se no ficheiro ft.xlsx.