Problema 1 – em colaboração com o grupo de Diversidade Genética do Ipatimup-i3S. Co-autoria em possível publicação.

**Identificação de grupos de co-expressão entre todos os genes de proteínas mitocondriais (codificadas pelo DNA nuclear e pelo DNA mitcondrial) e todos os restantes genes (incluindo genes de RNA) nos vários tecidos humanos**

A identificação de grupos de co-expressão entre pares de genes, estando nós particularmente interessados em todos os pares em que pelo menos um dos elementos é um gene que codifica uma proteína mitocondrial, pode contribuir informação essencial acerca de redes proteicas e de como estas podem diferir ou não entre os vários tecidos humanos. Para além disso, pode também clarificar sobre a correlacção entre genes codificados pelo DNA mitocondrial e pelo DNA nuclear, uma vez estes dois genomas têm propriedades muito distintas, mas têm que de algum modo se coordenadar quanto à codificação de genes de proteínas mitocondriais. Os dados de expressão em vários tecidos humanos serão obtidos primiero da base de dados GTEx (http://www.gtexportal.org/home/) e numa segunda fase será feita uma confirmação a partir de dados da base TCGA (<http://cancergenome.nih.gov/>).

**Etapas detalhadas:**

- Download dos dados de expressão da base de dados GTEx vs6:

. usar a file All\_Tissue\_Site\_Details\_Analysis.combined.rpkm

. para a informação de que IDs pertencem a diferentes tecidos, consultar

<https://github.com/BgeeDB/expression-annotations/blob/master/annotations/gtex/uberon_GTEx_Bgee.tsv>

. dividir a file de dados de expressão geral pelos diferentes tecidos (ver lista de tecidos no final – confirmar se algum dos tecidos tem poucos indivíduos; não juntar por orgão, por exemplo, não juntar todos os “brain” uma vez que o padrão de expressão pode diferir entre as diferentes partes do cérebro)

- Usar lista de MitoCard 2 para identificação de genes de proteinas mitocondriais – a providenciar pelo grupo Diversidade Genética.

- Programar correlação entre todos os pares possíveis de genes, em que um dos membros é um gene de proteína mitocondrial (da lista MitoCard 2). Isto para cada tecido individualmente.

- Listar todos os pares significativos aplicando o cutoff >0.7 e <-0.7. Isto para cada tecido individualmente.

- Fazer uma análise de network para os pares significativos, por tecido. Discussão de métodos com o Professor e o grupo Diversidade Genética.

- Fazer uma meta-análise de networks entre vários tecidos.

- Repetir as análises para alguns tecidos cuja expressão foi obtida na base de dados TCGA. Estes dados de expressão srão providenciados pelo grupo Diversidade Genética.

**Tecidos na base de dados GTEx**

Adipose - Subcutaneous

Adipose - Visceral (Omentum)

Adrenal Gland

Artery - Aorta

Artery - Coronary

Artery - Tibial

Bladder

Brain - Amygdala

Brain - Anterior cingulate cortex (BA24)

Brain - Caudate (basal ganglia)

Brain - Cerebellar Hemisphere

Brain - Cerebellum

Brain - Cortex

Brain - Frontal Cortex (BA9)

Brain - Hippocampus

Brain - Hypothalamus

Brain - Nucleus accumbens (basal ganglia)

Brain - Putamen (basal ganglia)

Brain - Spinal cord (cervical c-1)

Brain - Substantia nigra

Breast - Mammary Tissue

Cells - EBV-transformed lymphocytes

Cells - Transformed fibroblasts

Cervix - Ectocervix

Cervix - Endocervix

Colon - Sigmoid

Colon - Transverse

Esophagus - Gastroesophageal Junction

Esophagus - Mucosa

Esophagus - Muscularis

Fallopian Tube

Heart - Atrial Appendage

Heart - Left Ventricle

Kidney - Cortex

Liver

Lung

Minor Salivary Gland

Muscle - Skeletal

Nerve - Tibial

Ovary

Pancreas

Pituitary

Prostate

Skin - Not Sun Exposed (Suprapubic)

Skin - Sun Exposed (Lower leg)

Small Intestine - Terminal Ileum

Spleen

Stomach

Testis

Thyroid

Uterus

Vagina

Whole Blood