Analysis of the network component of the pathogenicity of mutations in cancer patients

CAMDA
July 12, 2014. Boston

Luz Garcia-Alonso
luz.g.alonso@gmail.com

PRINCIPE FELIPE
CENTRO DE INVESTIGACION
Systems Genomics Laboratory

SCENARIO

Identification of mutational cancer driver genes

Tumors evolve from benign to malignant lesions by acquiring a series of **mutations over time**

Studies based on individual gene/mutation recurrence have identified about 140 genes containing intragenic mutations **initiating** tumorigenesis

Most of tumors have only one or two driver gene mutations, but tumor development and **progression** require multiple sequential genetic alterations.

Cancer genomes exhibit extensive mutational **heterogeneity among the tumors of different patients**, is probably related to the differences somatic mutations within tumors

(Vogelstein et al., 2013)

SCENARIO

Identification of mutational cancer driver genes

Tumors evolve from benign to malignant lesions by acquiring a series of **mutations over time**

Studies based on individual gene/mutation recurrence have identified about 140 genes containing intragenic mutations **initiating** tumorigenesis

Most of tumors have only one or two driver gene mutations, but tumor development and **progression** require multiple sequential genetic alterations.

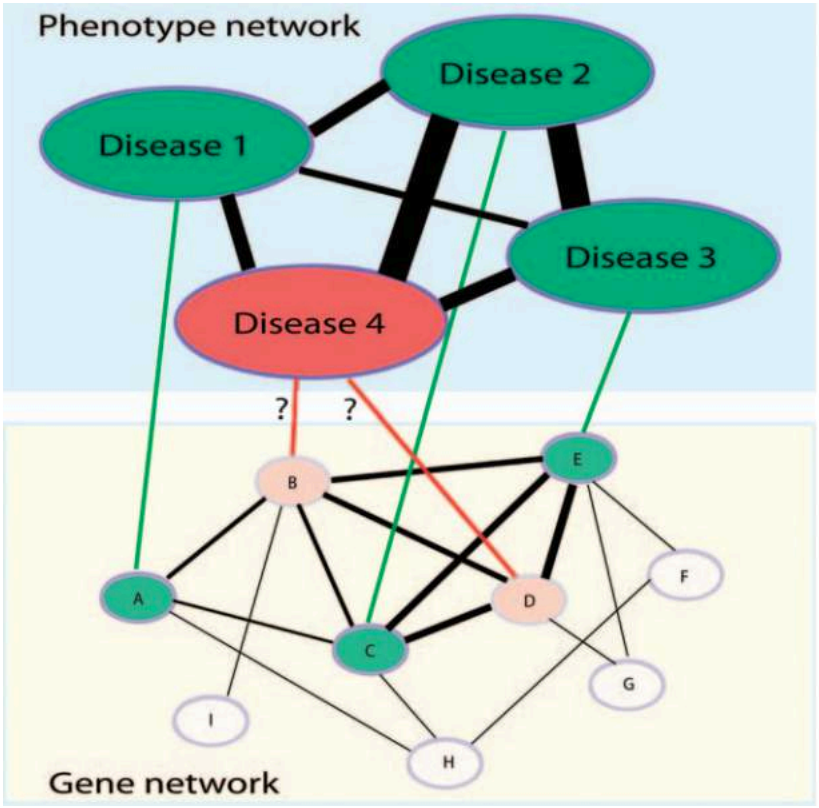
Cancer genomes exhibit extensive mutational **heterogeneity among the tumors of different patients**, is probably related to the differences somatic mutations within tumors

(Vogelstein et al., 2013)

This heterogeneity is clearly underestimated in the current driver versus passenger model.

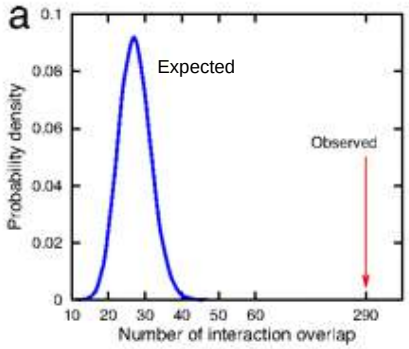
THE CELL AS A SYSTEM

The modular nature of genetic diseases

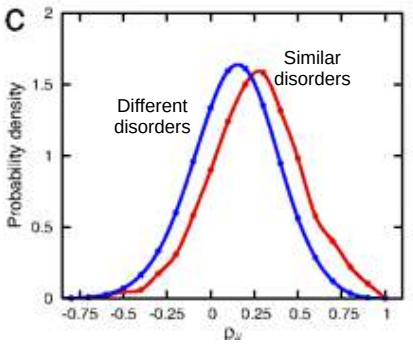


Diseases can be understood as failures of functional modules

(Oti & Brunner. Clinical genetics, 2007)



Disease genes tend to be connected to other disease genes



Genes associated with similar disorders show higher expression profiling similarity for their transcripts

(Goh et al. PNAS, 2007)

THE CELL AS A SYSTEM

The modular nature of genetic diseases

A modular view of disease genes would help the **identification process of additional disease genes** for multifactorial diseases

THE CELL AS A SYSTEM

The modular nature of genetic diseases

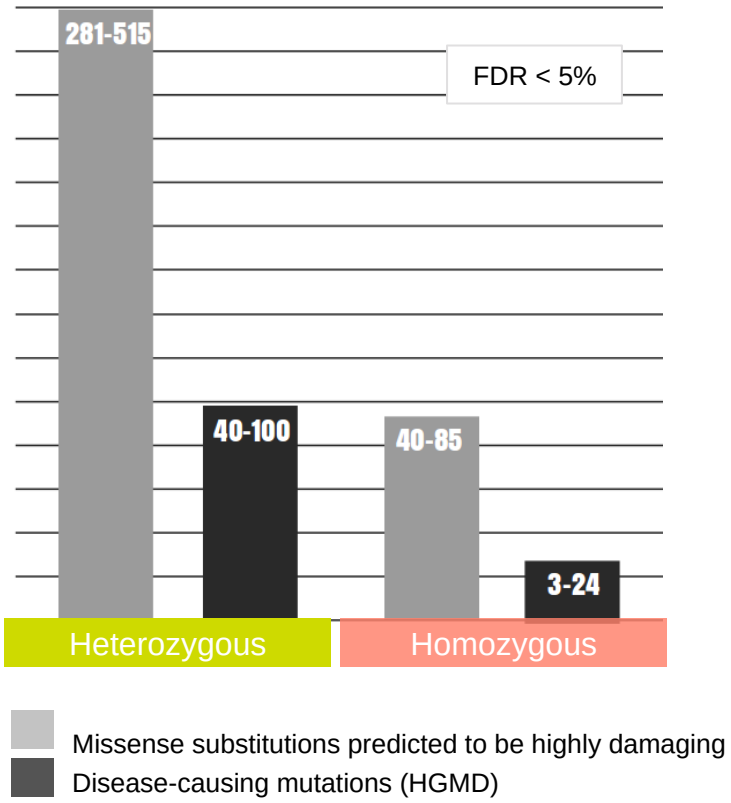
A modular view of disease genes would help the **identification process of additional disease genes** for multifactorial diseases

Would the study of healthy genomes help us to improve also the process?

HUMAN GENOMES

Mutational load

Recent human genomic projects have revealed the existence of an unexpectedly high amount of **deleterious variability** in apparently normal, healthy individuals.



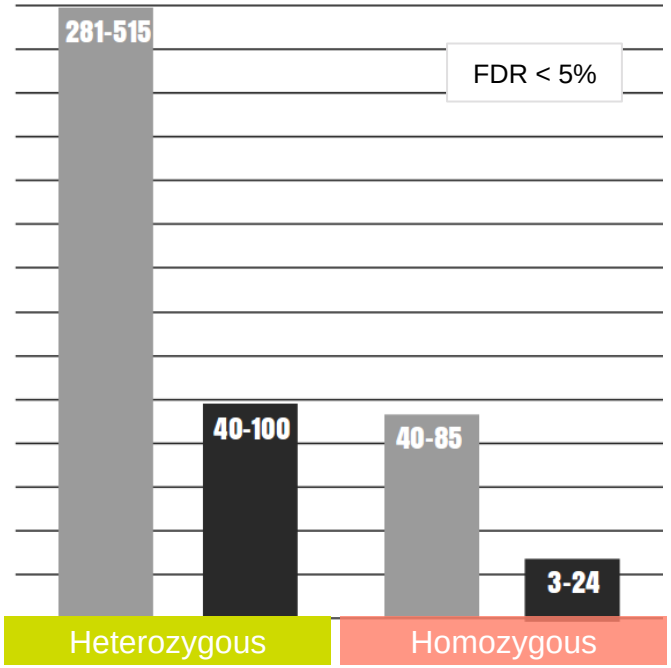
(Xue et al. Am J Hum Genet. 2012)

HUMAN GENOMES

Mutational load

Recent human genomic projects have revealed the existence of an unexpectedly high amount of **deleterious variability** in apparently normal, healthy individuals.

Important implications for the clinical interpretation of human genome–sequencing data



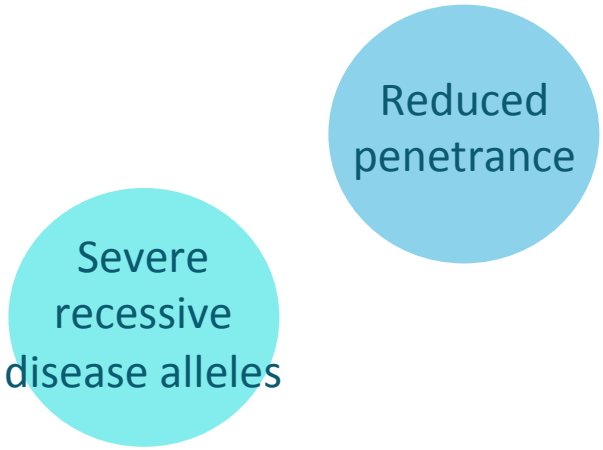
(Xue et al. Am J Hum Genet. 2012)

HUMAN GENOMES

Mechanisms for the maintenance of deleterious variants

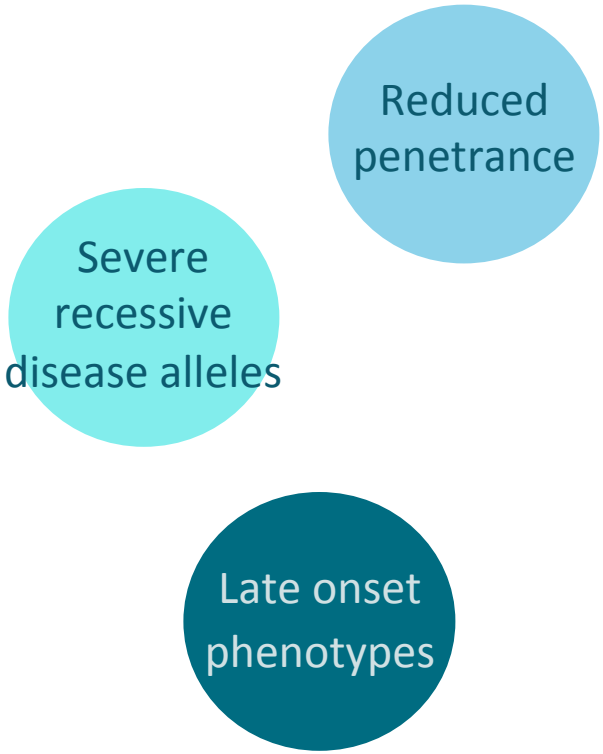
HUMAN GENOMES

Mechanisms for the maintenance of deleterious variants



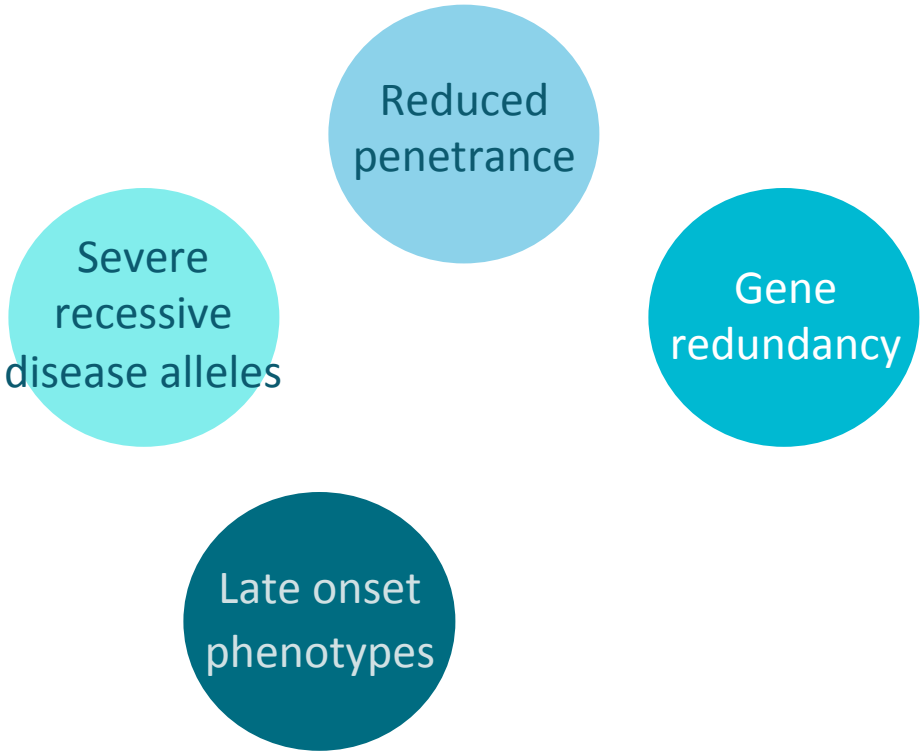
HUMAN GENOMES

Mechanisms for the maintenance of deleterious variants



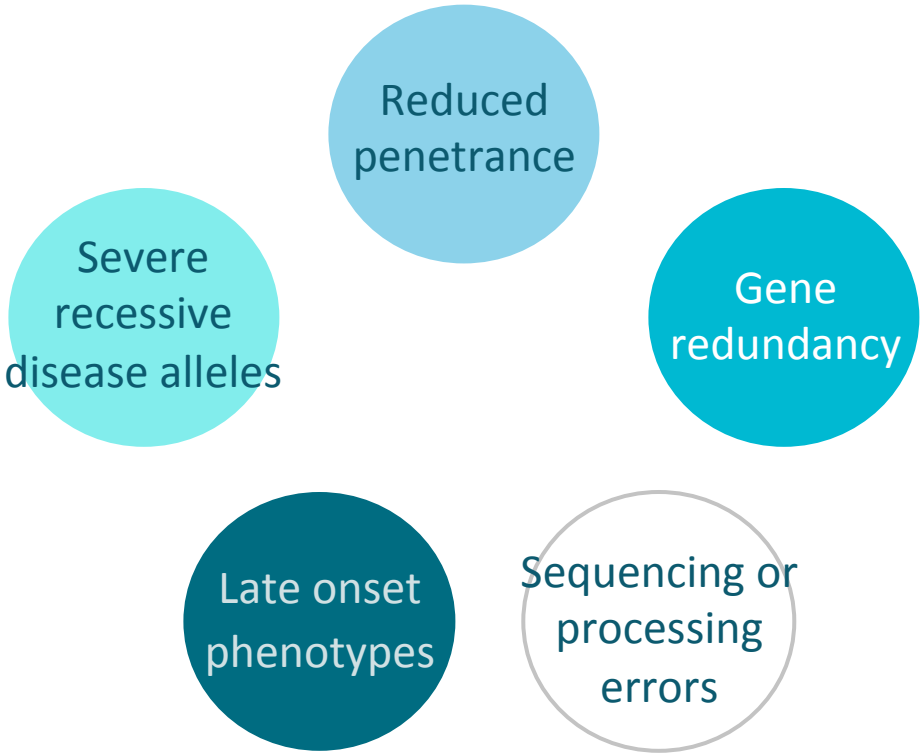
HUMAN GENOMES

Mechanisms for the maintenance of deleterious variants



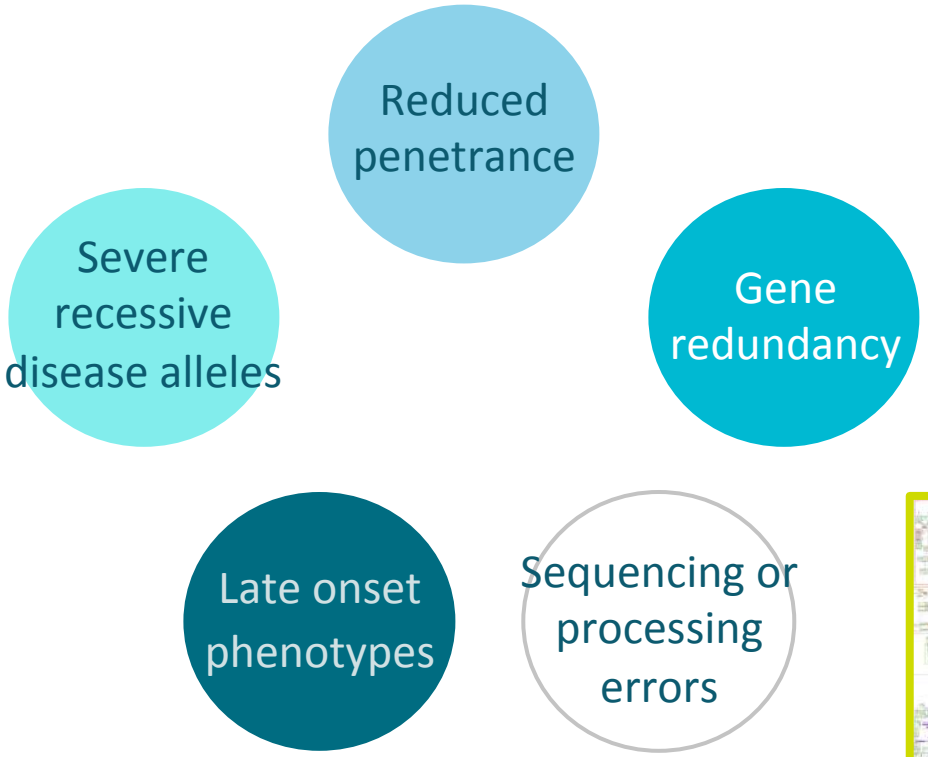
HUMAN GENOMES

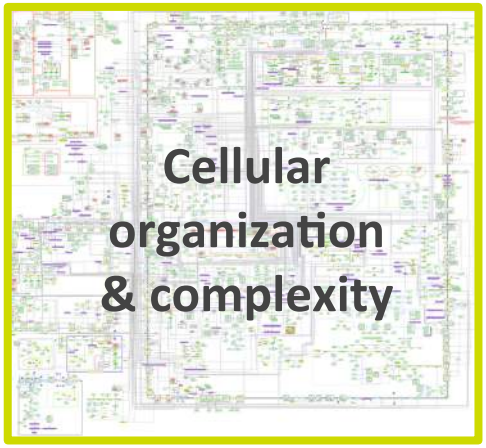
Mechanisms for the maintenance of deleterious variants



HUMAN GENOMES

Mechanisms for the maintenance of deleterious variants





THE CELL AS A SYSTEM

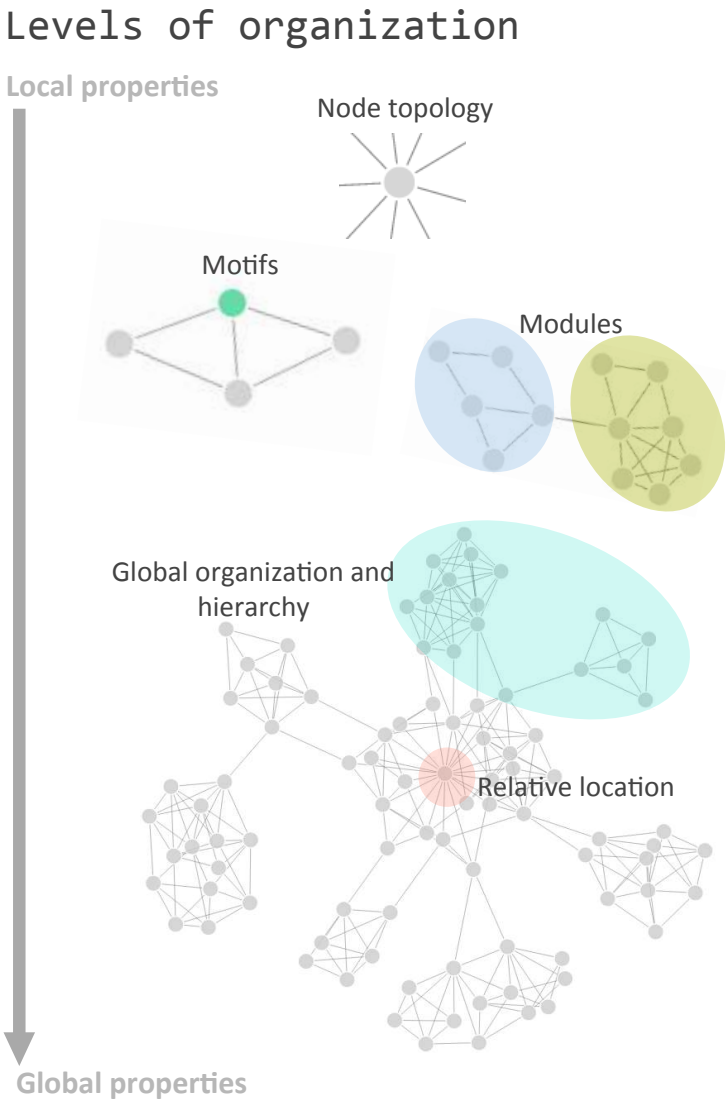
Protein interactome

Theoretical scaffold that relates proteins among them

Modelled as a **graph**, in which proteins are the nodes and the interactions the edges.

Organized into **modules**, carring specific biological functions

It's **organization** is not random but follows a series of principles



THE CELL AS A SYSTEM

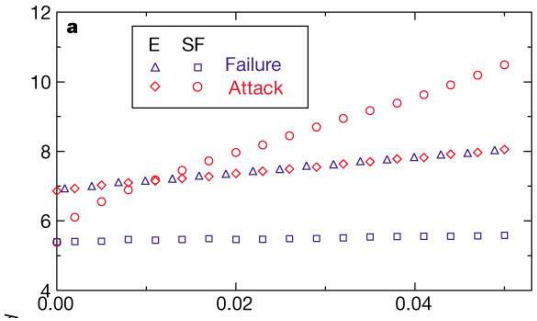
Robustness against perturbations

Robustness: the ability of the biological networks to buffer a phenotype in the face of genetic and environmental perturbations

Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong & Albert-László Barabási

NATURE | VOL 406 | 27 JULY 2000 | www.nature.com



This robustness of scale-free networks is rooted in their extremely inhomogeneous connectivity distribution

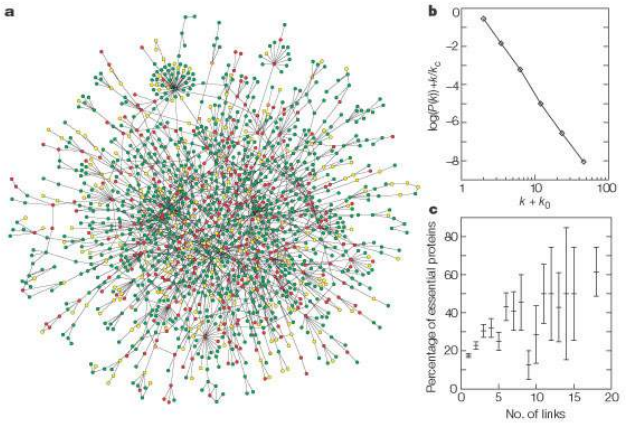
brief communications

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

H. Jeong*, S. P. Mason†, A.-L. Barabási*, Z. N. Oltvai†

NATURE | VOL 411 | 3 MAY 2001 |



HYPOTHESIS

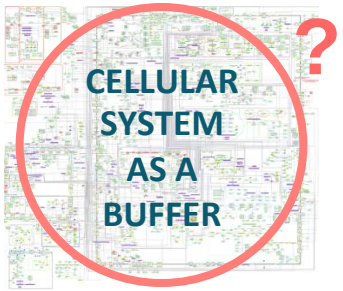
Reduced penetrance

Severe recessive disease alleles

Mechanisms for the maintenance of deleterious variants

Gene redundancy

Late onset phenotypes



Luz Garcia-Alonso et al., 2014 (Under revision)

DATA PROCESSING

Human populations

13 from the **1000 Genomes Proj**
- European (TSI, FIN, GBR & CEU)
- Asian (CHB, CHS & JPT)
- American (MXL, PUR & CLM)
- African (YRI, LWK & ASW)

(Durbin et al, 2010)

1 newly sequenced from **Spain**

1 population of paired samples of 41 **Chronic Lymphocytic Leukemia** (CLL) patients

(Quesada et al, 2012)

Luz Garcia-Alonso et al., 2014 (Under revision)

DATA PROCESSING

Variants were annontated using VARIANT software. The variants were classified as **potentially delterious** as follows:

Variant type	Filter
Stop gain, loss and splicing disrrupting sites	Conserved
Nonsynonymous	Conserved + SIFT / Polyphen

Luz Garcia-Alonso et al., 2014 (Under revision)

DATA PROCESSING

Variants were annontated using VARIANT software. The variants were classified as **potentially delterious** as follows:

Variant type	Filter
Stop gain, loss and splicing disrrupting sites	Conserved
Nonsynonymous	Conserved + SIFT / Polyphen

COMPUTATIONAL VALIDATION

In silico modeling of the mutations in the protein using its previously solved crystal structure

Luz Garcia-Alonso et al., 2014 (Under revision)

DATA PROCESSING

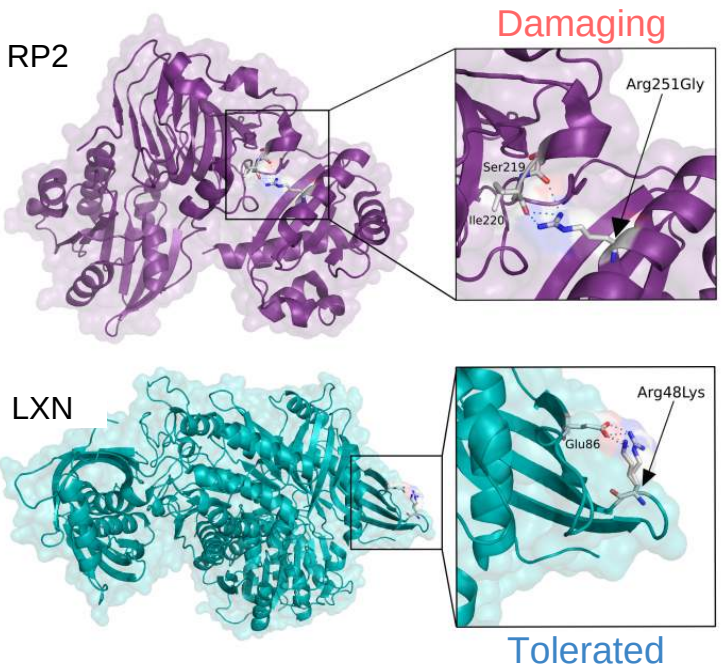
Human populations → **Variants selection**

Variants were annontated using VARIANT software. The variants were classified as **potentially delterious** as follows:

Variant type	Filter
Stop gain, loss and splicing disrrupting sites	Conserved
Nonsynonymous	Conserved + SIFT / Polyphen

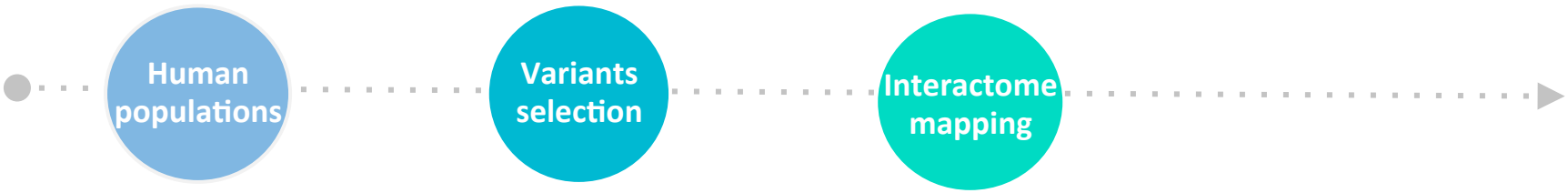
COMPUTATIONAL VALIDATION

In silico modeling of the mutations in the protein using its previously solved crystal structure



Luz Garcia-Alonso et al., 2014 (Under revision)

DATA PROCESSING



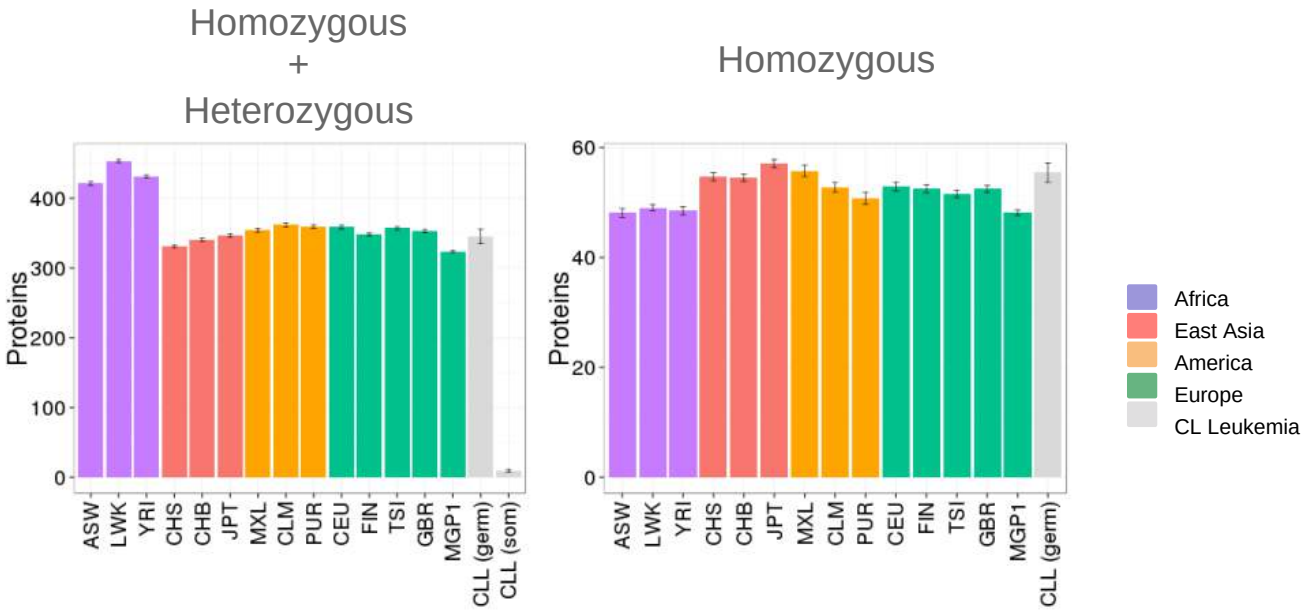
The model of the human interactome was built using data on **binary protein-protein interactions** from BioGRID, IntAct and MINT databases.

To avoid false positives only interactions detected by at least two different detection methods were used.

Luz Garcia-Alonso et al., 2014 (Under revision)

RESULTS

Deleterious variants in proteins of the interactome: Population overview



Proteins affected by potentially deleterious variants that configure the human interactome among all the populations analyzed

Luz Garcia-Alonso et al., 2014 (Under revision)

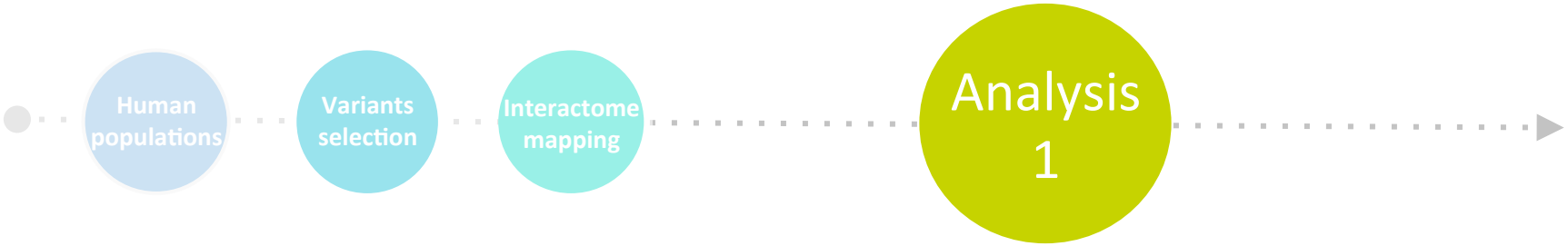
OBJECTIVE

Compare the topological role of proteins affected by deleterious variants observed in normal population, in monogenic diseases and in cancer patients

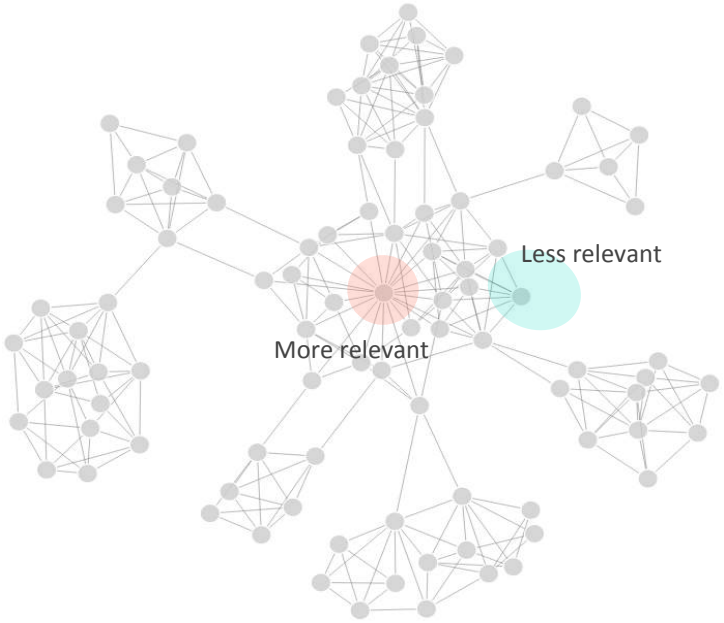
Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

Compare the topological role of proteins affected by deleterious variants observed in normal population, in monogenic diseases and in cancer patients



Properties describing the **relevance of the protein for the network integrity**
- the direct number of partners
- the betweenness (measure of the extent to which a protein lies on the paths between others)
- the closeness centrality (a measure of centrality in the interactome)

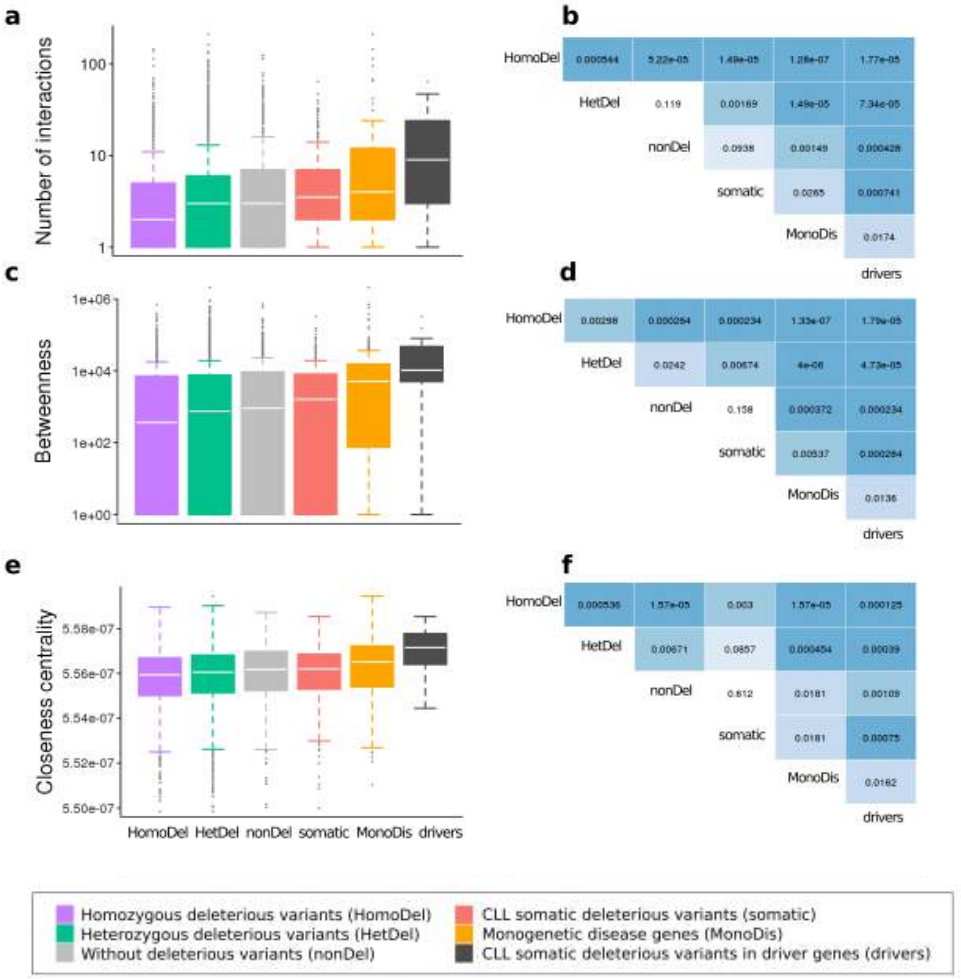


Luz Garcia-Alonso et al., 2014 (Under revision)

RESULTS

Proteins affected by deleterious variants observed in normal population, in monogenic diseases and in cancer patients have different topological roles

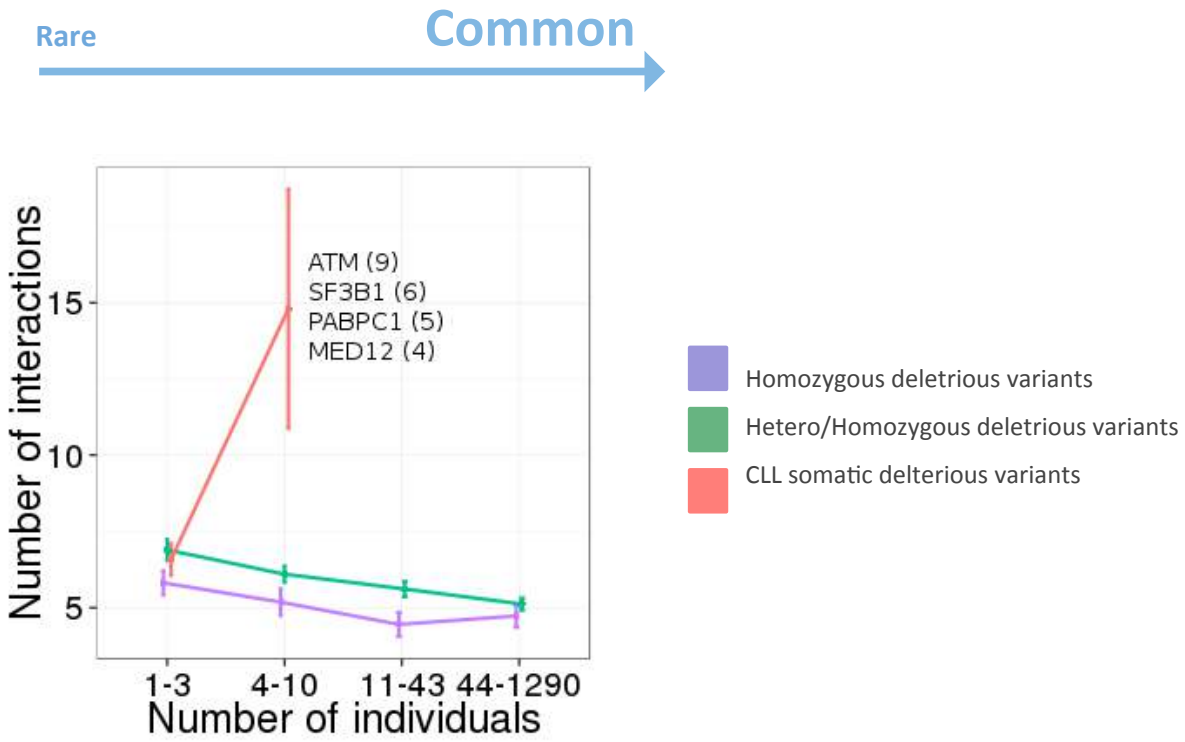
Relevance of the protein for the network integrity



Luz Garcia-Alonso et al., 2014 (Under revision)

RESULTS

Proteins affected by deleterious variants observed in normal population, in monogenic diseases and in cancer patients have different topological roles



Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

Effect of deleterious variants observed in normal individuals over the interactome structure

Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

Effect of deleterious variants observed in normal individuals over the interactome structure

Human populations → Variants selection → Interactome mapping → Analysis 2

Real individuals

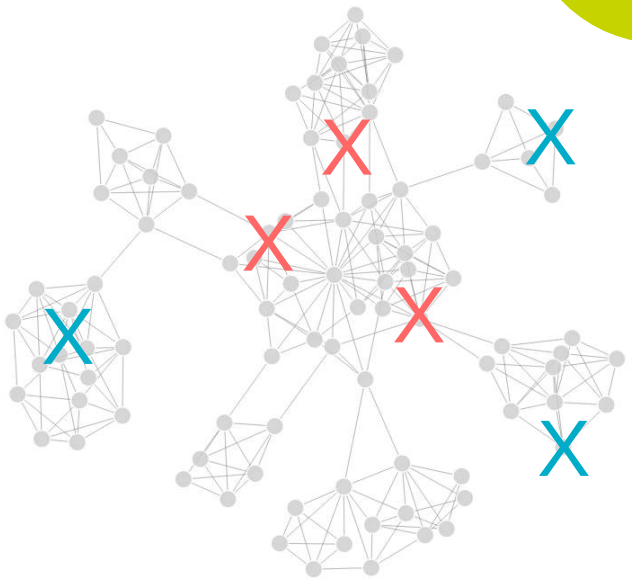
Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

Effect of deleterious variants observed in normal individuals over the interactome structure

Human populations → Variants selection → Interactome mapping → Analysis 2

Real individuals



homozygote deleterious variants = protein lost from the system

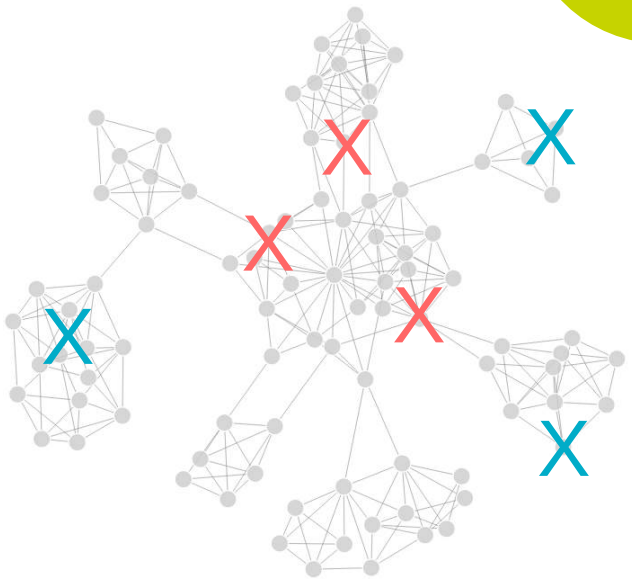
Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

Effect of deleterious variants observed in normal individuals over the interactome structure

Human populations → Variants selection → Interactome mapping → Analysis 2

Real individuals



homozygote deleterious variants = **protein lost from the system**

Global estructural parameters that describe the **interconnectedness** and **integrity** the interactomes:

1. the number of connections
2. the average length of shortest paths
3. the number of components

Luz Garcia-Alonso et al., 2014 (Under revision)

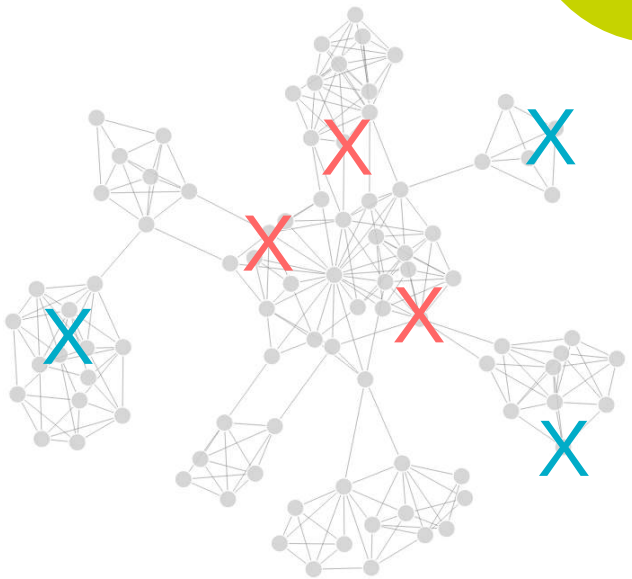
OBJECTIVE

Effect of deleterious variants observed in normal individuals over the interactome structure

Human populations → Variants selection → Interactome mapping → Analysis 2

Real individuals

Simulated individuals
(**uniform probability**)



homozygote deleterious variants = **protein lost from the system**

Global estructural parameters that describe the **interconnectedness** and **integrity** the interactomes:

1 the number of connections
2 the average length of shortest paths
3 the number of components

Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

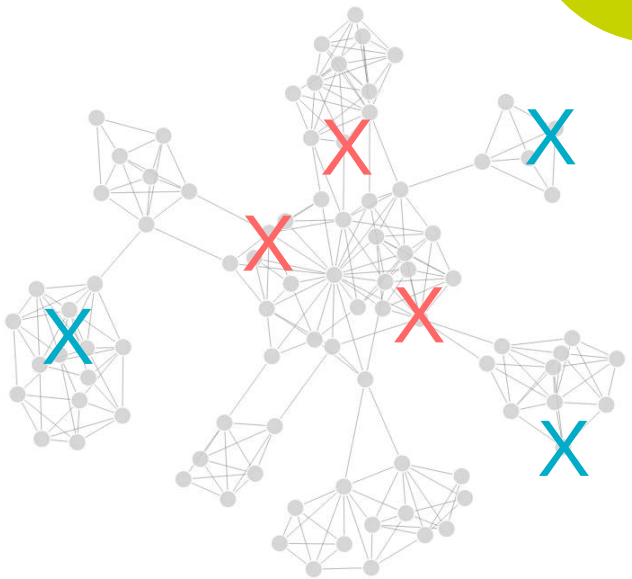
Effect of deleterious variants observed in normal individuals over the interactome structure

Human populations → Variants selection → Interactome mapping → Analysis 2

Real individuals

Simulated individuals (**uniform probability**)

Simulated individuals (**observed populational frequencies**)



homozygote deleterious variants = **protein lost from the system**

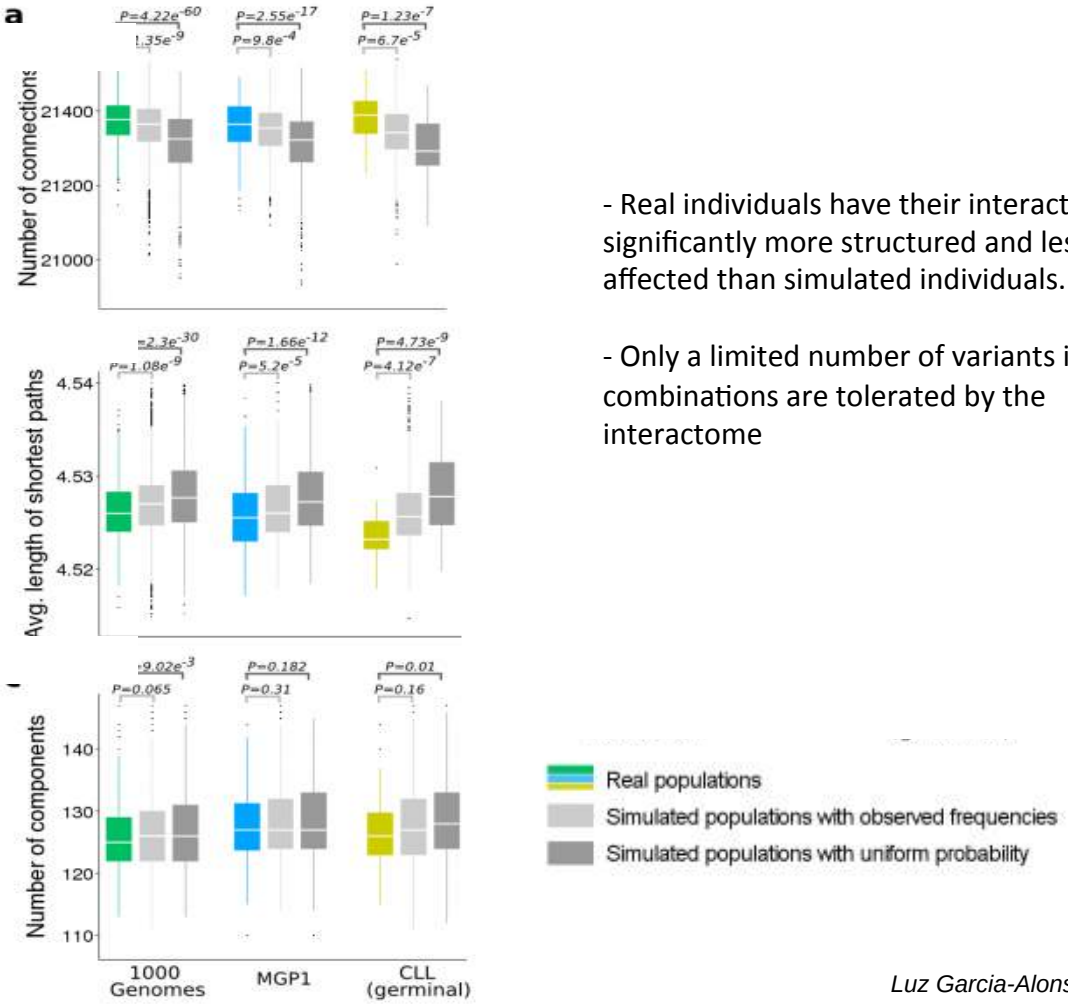
Global estructural parameters that describe the **interconnectedness** and **integrity** the interactomes:

1. the number of connections
2. the average length of shortest paths
3. the number of components

Luz Garcia-Alonso et al., 2014 (Under revision)

RESULTS

Effect of deleterious variants observed in normal individuals over the interactome structure



- Real individuals have their interactomes significantly more structured and less affected than simulated individuals.

- Only a limited number of variants in specific combinations are tolerated by the interactome

Luz Garcia-Alonso et al., 2014 (Under revision)

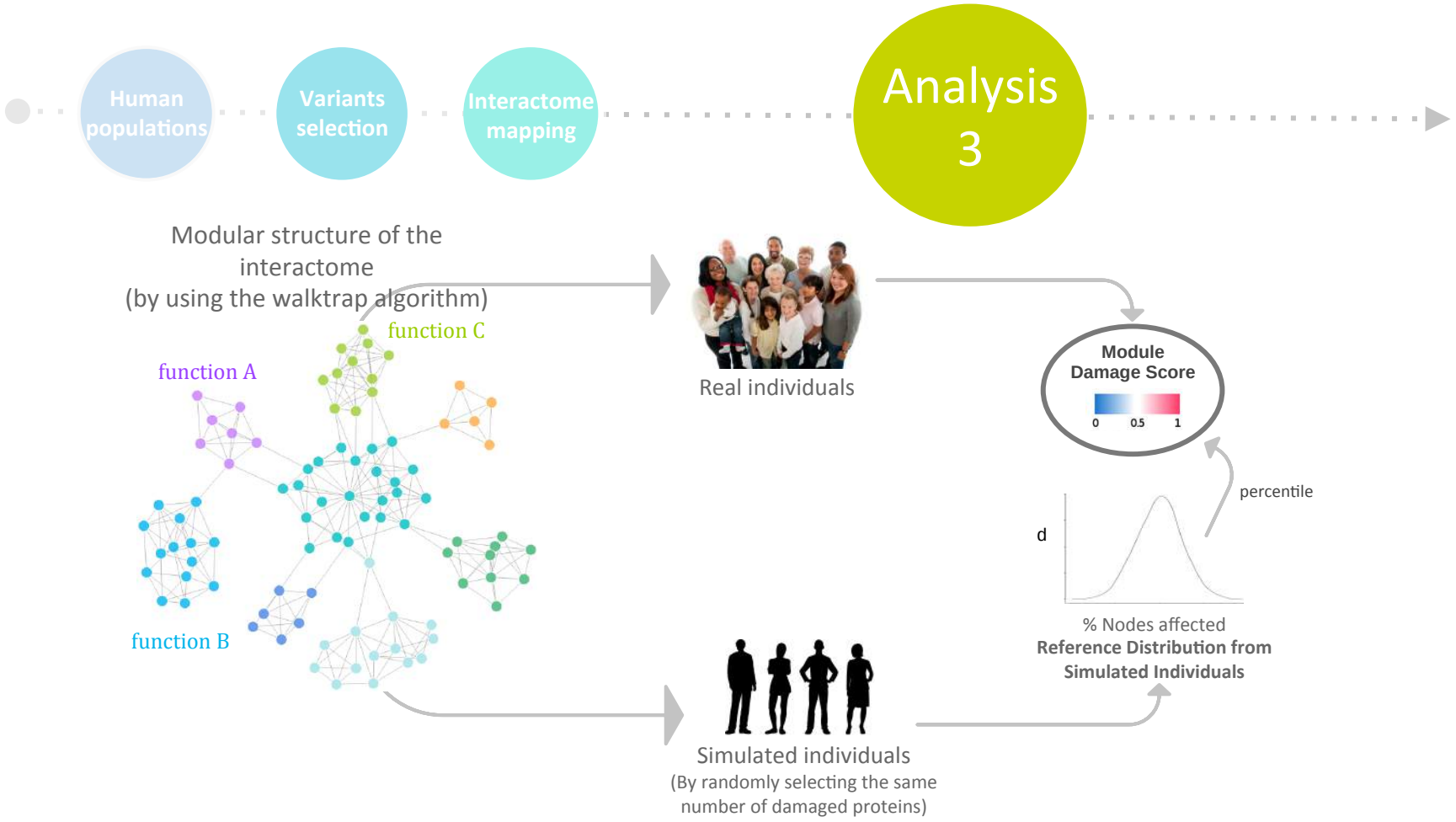
OBJECTIVE

Deleterious variants distribution across the interactome modules

Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

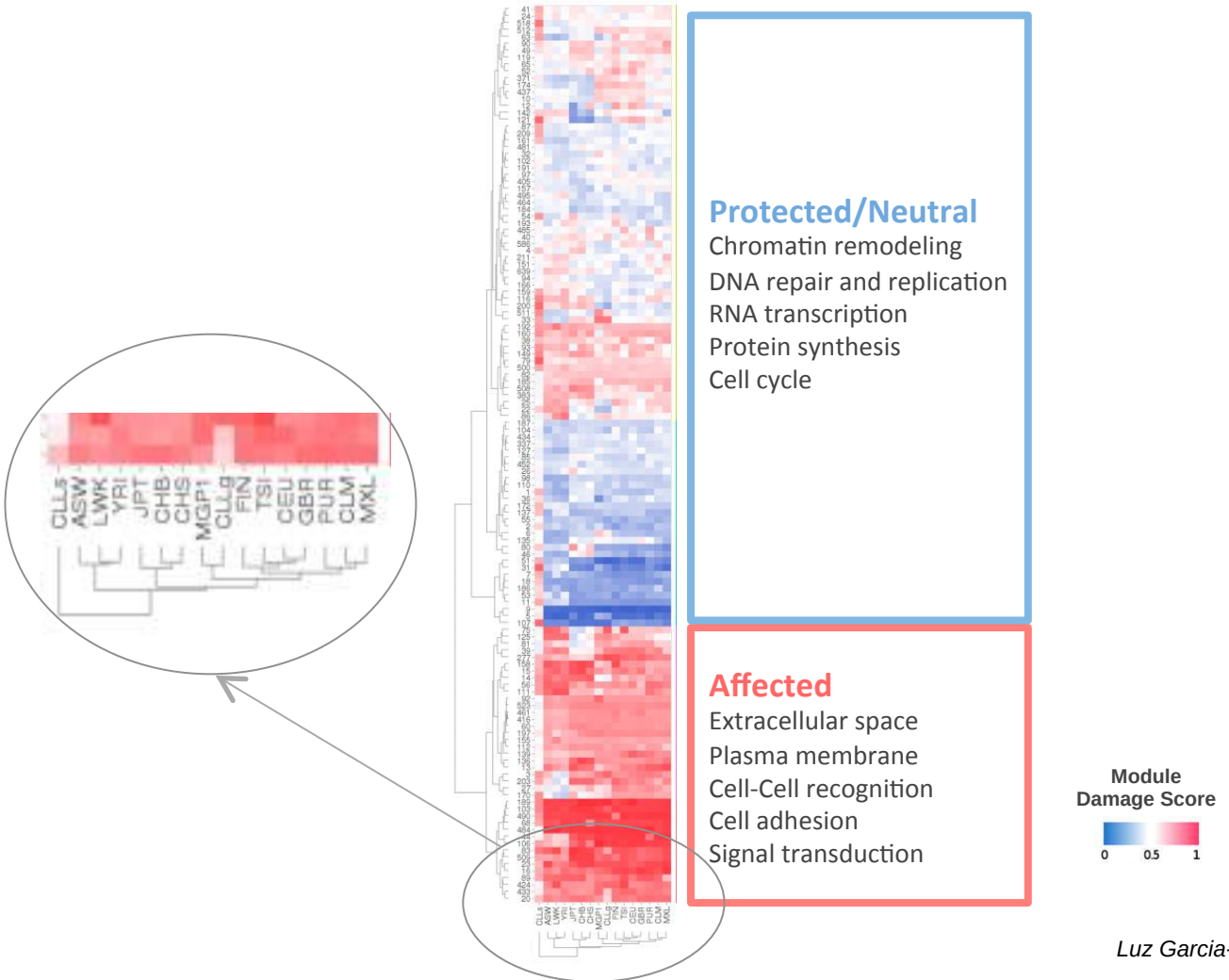
Deleterious variants distribution across the interactome modules



Luz Garcia-Alonso et al., 2014 (Under revision)

RESULTS

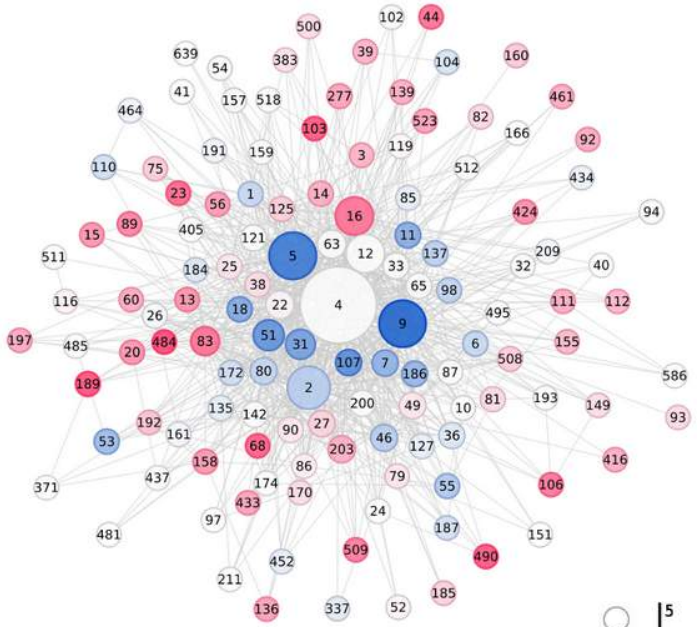
Deleterious variants observed in normal individuals follow a characteristic pattern across interactome modules

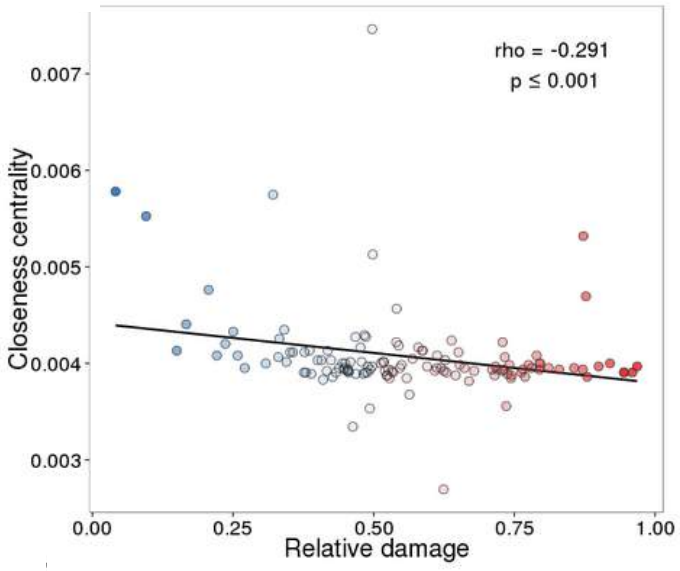


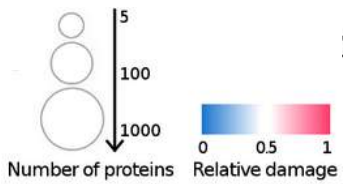
Luz Garcia-Alonso et al., 2014 (Under revision)

RESULTS

Deleterious variants observed in normal individuals follow a characteristic pattern across interactome modules



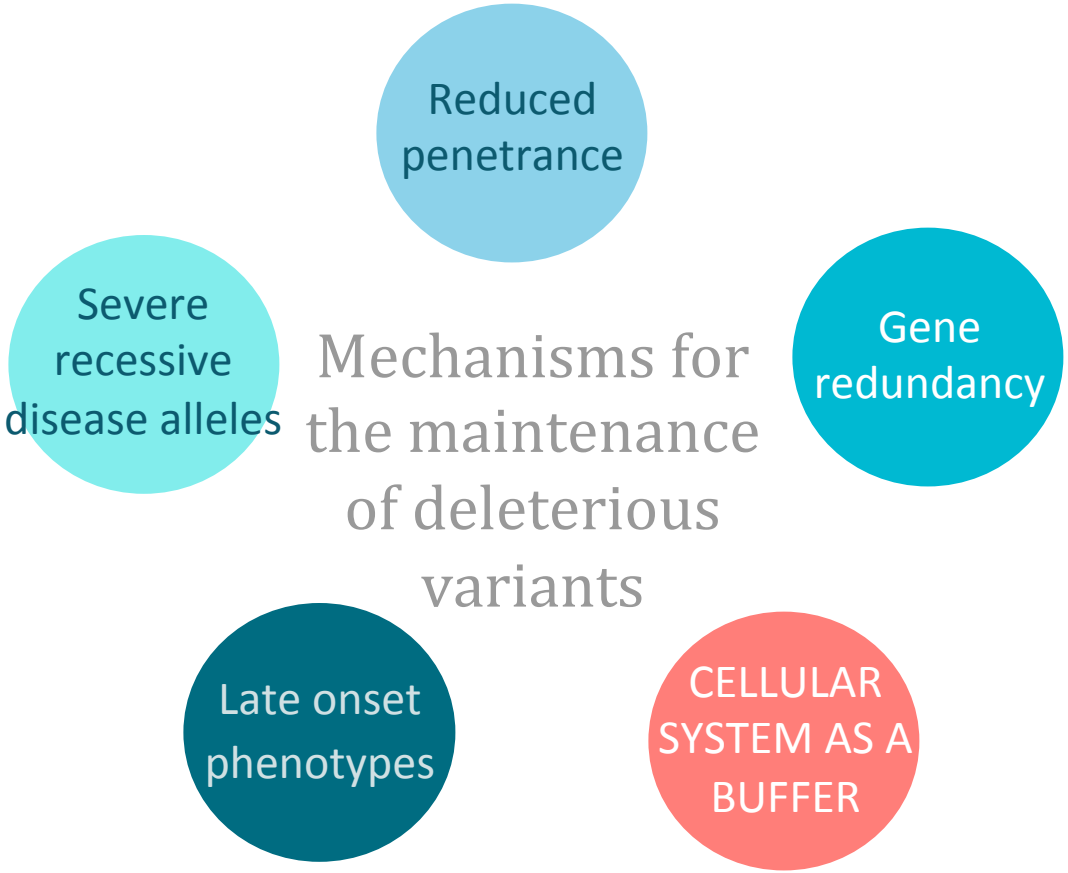




Reflect of the periphery ability to adapt under changing conditions

Luz Garcia-Alonso et al., 2014 (Under revision)

CONCLUSION



The **deleterious character of a variant** NOT ONLY depends on the damage that causes to the protein BUT ultimately is **a system's property**

SCENARIO

Identification of mutational cancer driver genes

Tumors evolve from benign to malignant lesions by acquiring a series of **mutations over time**

Studies based on individual gene/mutation recurrence have identified about 140 genes containing intragenic mutations driving, **initiating** tumorigenesis

Most of tumors have only one or two driver gene mutations, but tumor development and **progression** require multiple sequential genetic alterations.

Cancer genomes exhibit extensive mutational **heterogeneity among the tumors of different patients**, is probably related to the differences somatic mutations within tumors

(Vogelstein et al., 2013)

This heterogeneity is underestimated in the current driver versus passenger model.

An alternative is to examine the somatic mutations in the context of the protein interactome

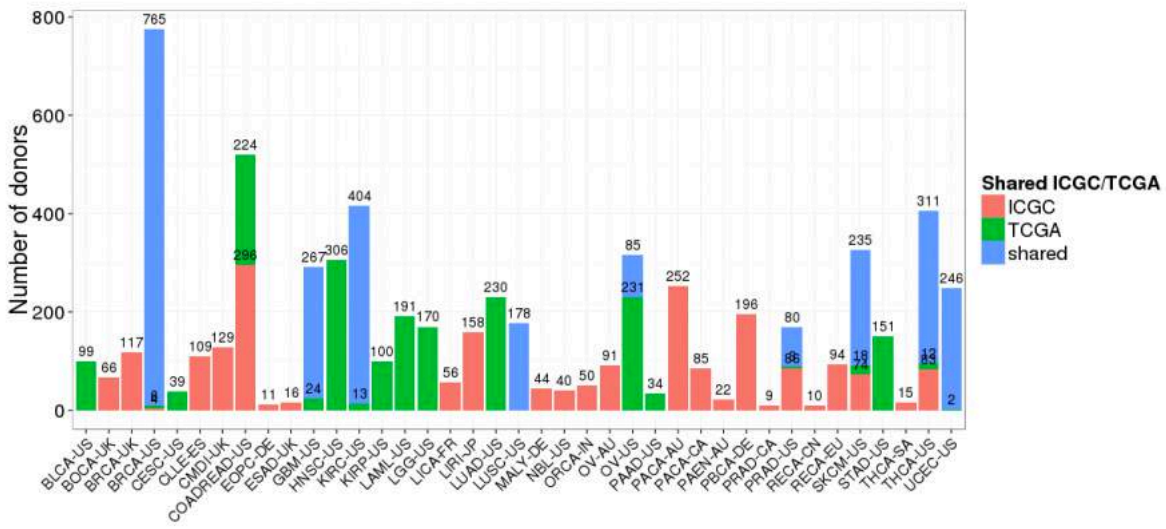
DATA PROCESSING

Cancer genome somatic variants

25 cancers from the **ICGC** (release 15.1)

19 cancers from the **TCGA Pan-Cancer** downloaded from syn1729383 (Kandoth et al., 2013)

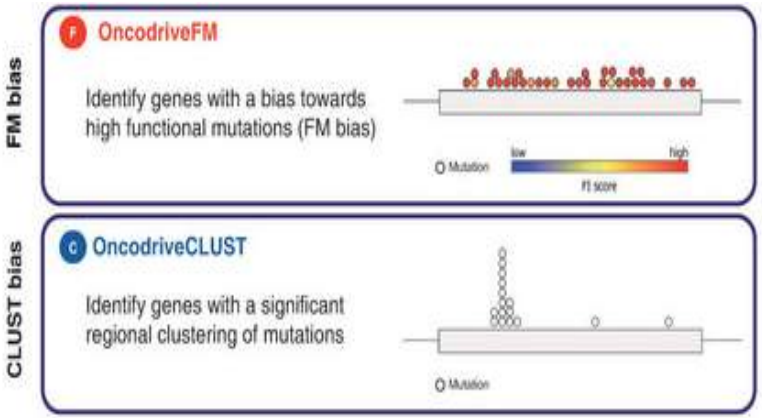
A total of 6573 unique donors in the raw dataset



DATA PROCESSING

Cancer genome somatic variants → **Variants selection**

Search for **somatic variants under positive selection** during the development of cancer



(*Gonzalez-Perez et al., 2003*)

DATA PROCESSING

The model of the human interactome was built using data on **binary protein-protein interactions** from BioGRID, IntAct and MINT databases.

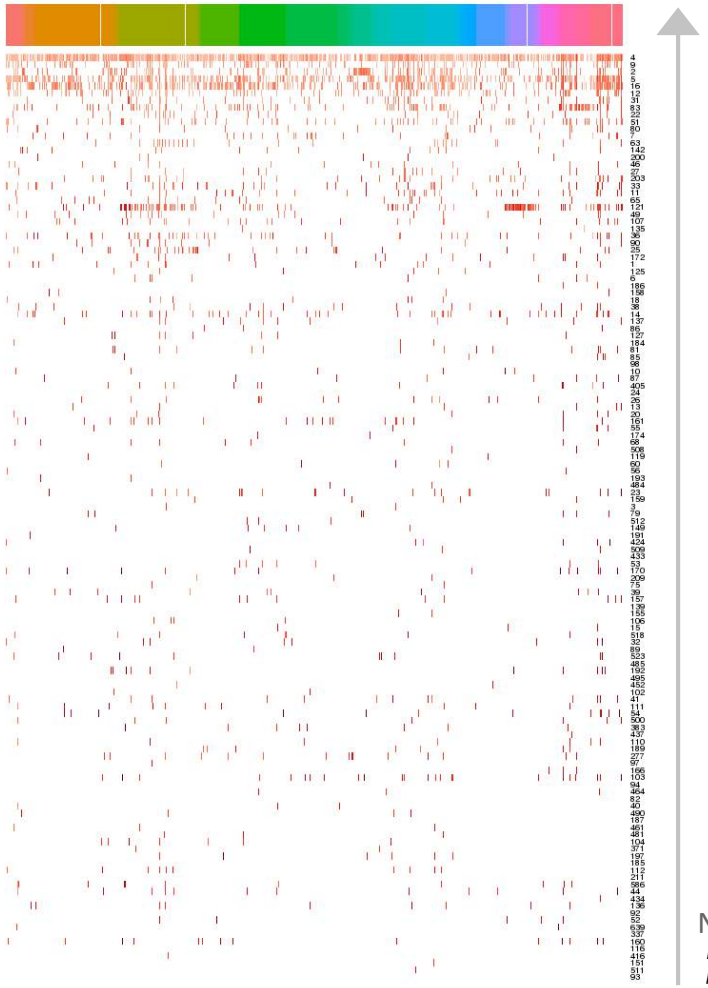
To avoid false positives only interactions detected by at least two different detection methods were used.

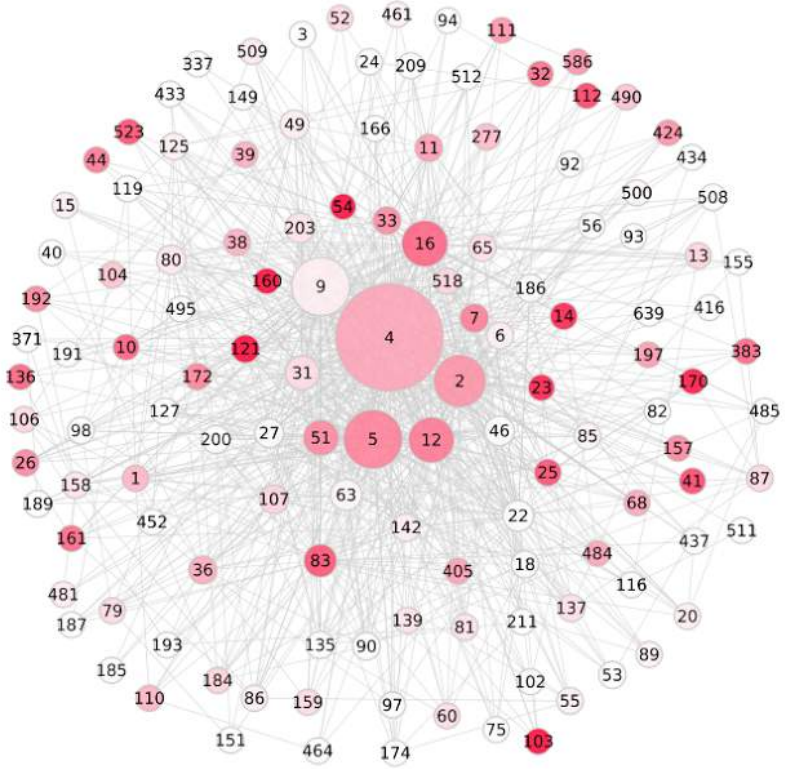
UPDATE *April 2014*

RESULTS

Cancer-specific mutations in the PAN-CANCER

Cancer types





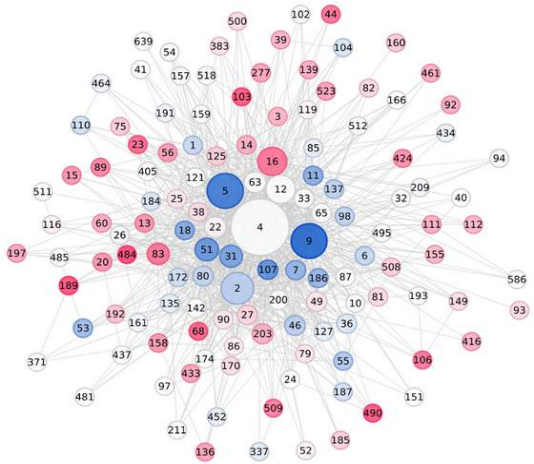
Network centralty *

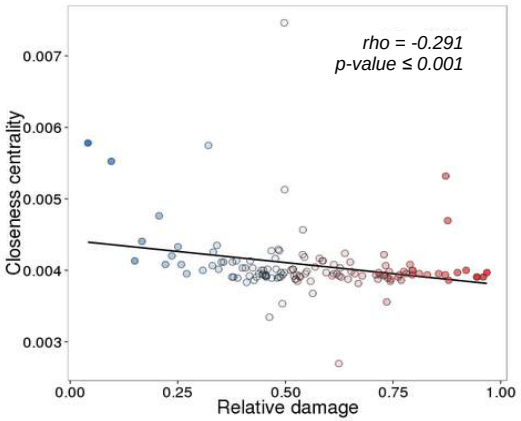
rho=0.197
p-value=0.024

RESULTS

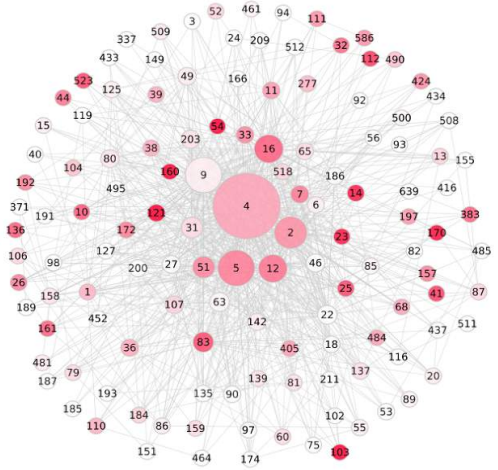
Population versus cancer-specific mutations

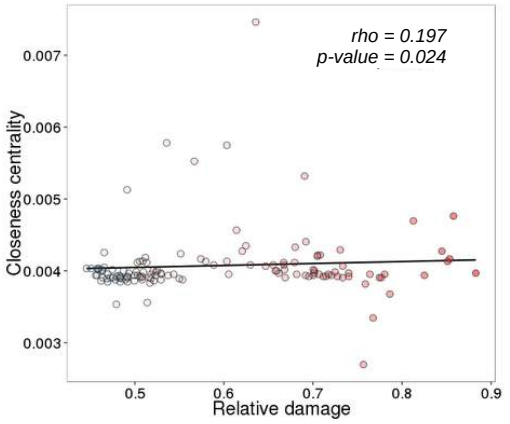
Healthy controls





Pancancer





Luz Garcia-Alonso et al., 2014 (Under revision)

OBJECTIVE

Zoom in into the interactome modules concentrating cancer-specific mutations

OBJECTIVE

Zoom in into the interactome modules concentrating cancer-specific mutations

Human populations

Variants selection

Interactome mapping

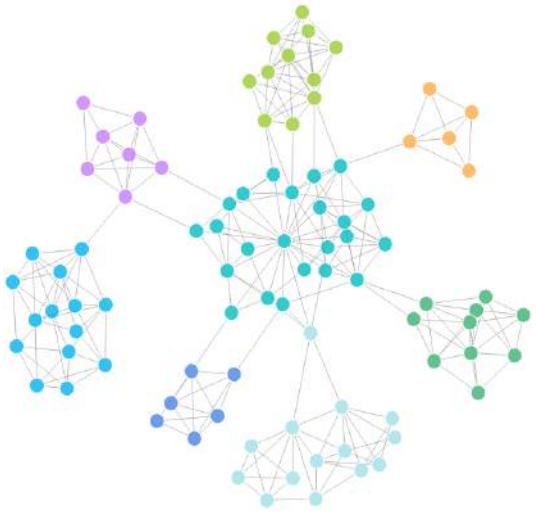
Analysis 5

Foreach module, we calculated the probability of observing X affected proteins out of the total of affected proteins, taking into account the module size bias by calculating the **binomial cumulative distribution function.**

RESULTS

Zoom in into the interactome modules concentrating cancer-specific mutations

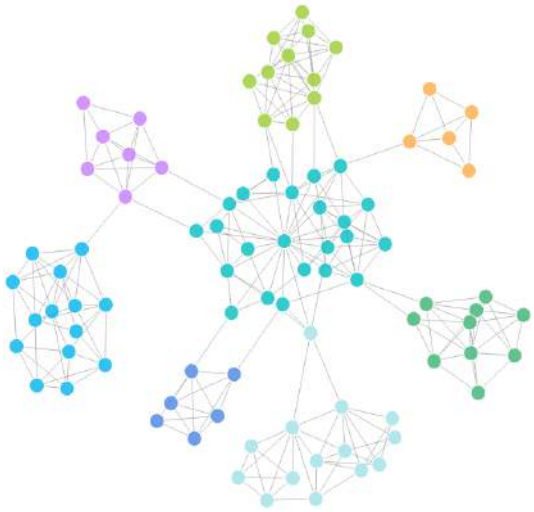
93 / 532 modules identified



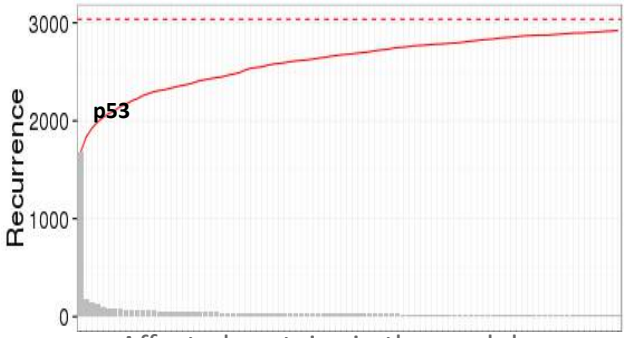
RESULTS

Zoom in into the interactome modules concentrating cancer-specific mutations

93 / 532 modules identified

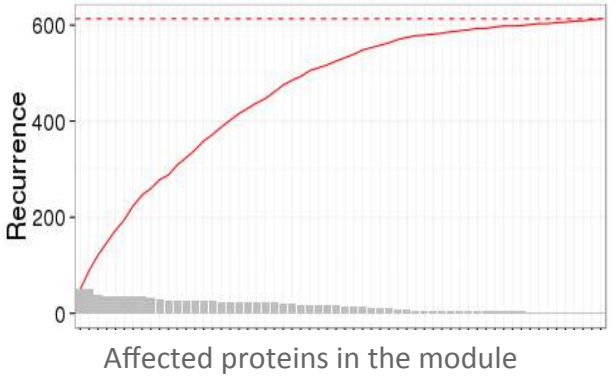


Distribution of the recurrence of the affected proteins



Exponential like shape

were most of the recurrency is explined by the top recurrent genes



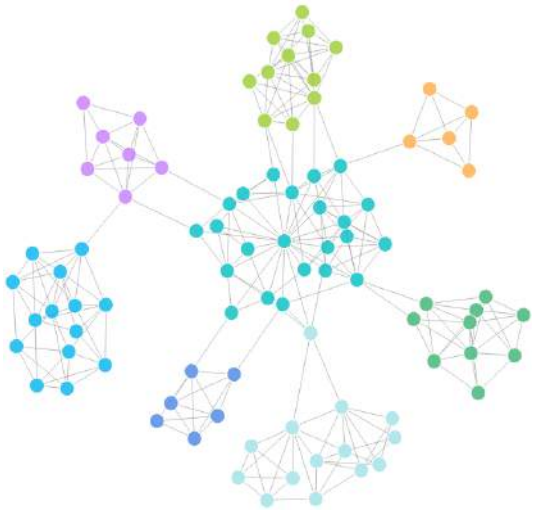
Lineal like shape

were the recurrency requires a large proportion of the module proteins to be explained

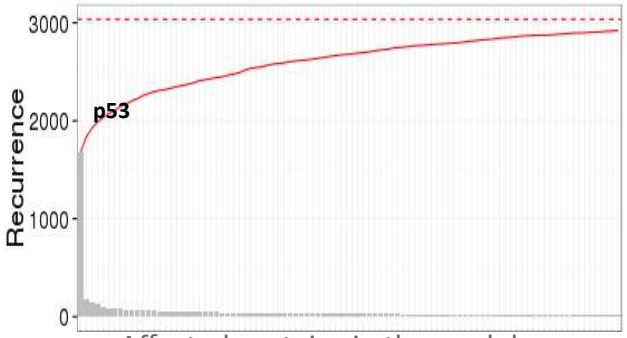
RESULTS

Zoom in into the interactome modules concentrating cancer-specific mutations

93 / 532 modules identified

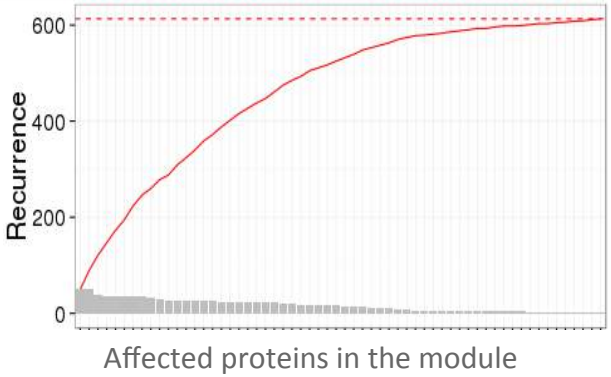


Distribution of the recurrence of the affected proteins



Exponential like shape

were most of the recurrency is explined by the top recurrent genes



Lineal like shape

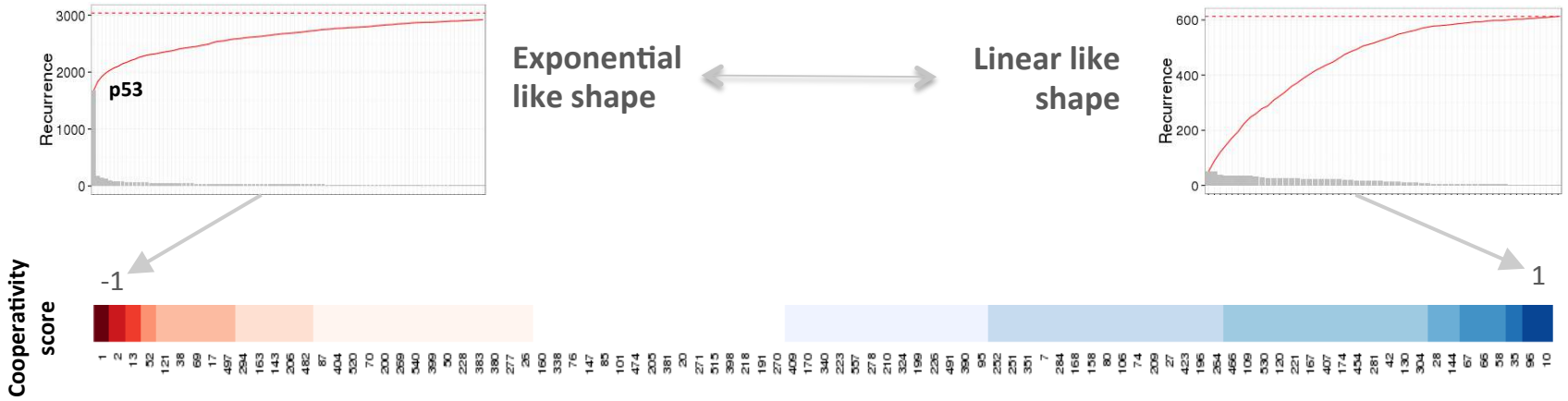
were the recurrency requires a large proportion of the module proteins to be explained

RESULTS

Zoom in into the interactome modules concentrating cancer-specific mutations

We calculated whether the observed distribution of the recurrency correlates with an exponential or a linear pattern

multigenicity score

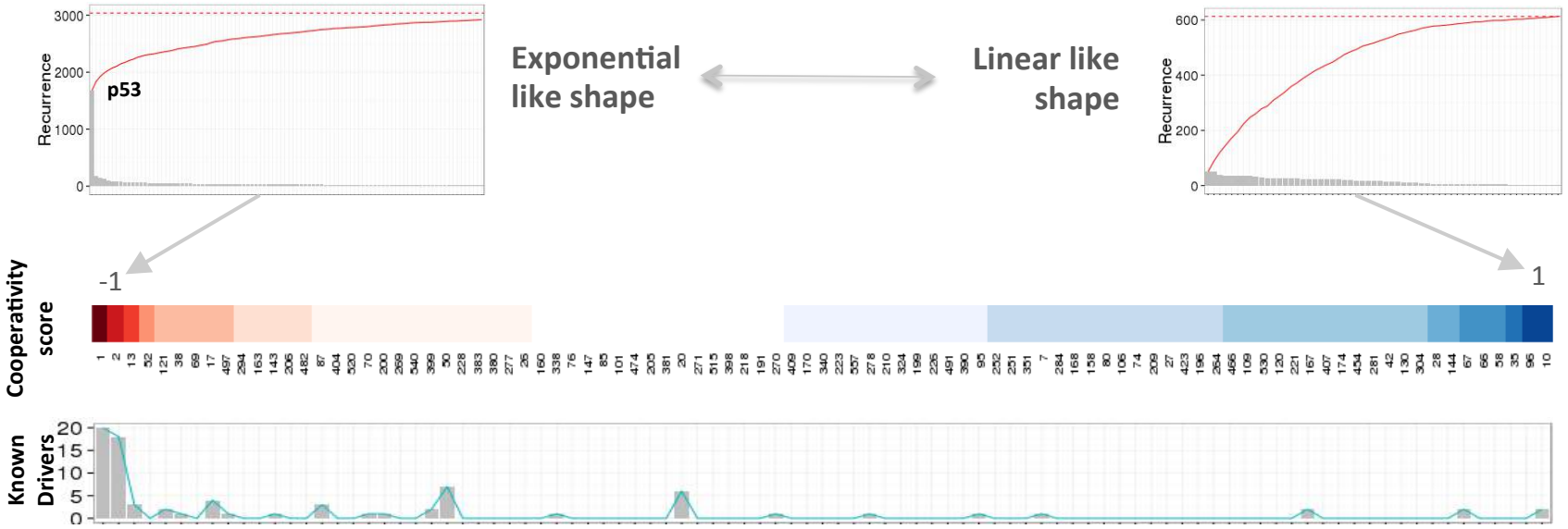


RESULTS

Zoom in into the interactome modules concentrating cancer-specific mutations

We calculated whether the observed distribution of the recurrency correlates with an exponential or a linear pattern

multigenicity score



OBJECTIVE

Temporal component of the multigenicity

OBJECTIVE

Temporal component of the multigenicity

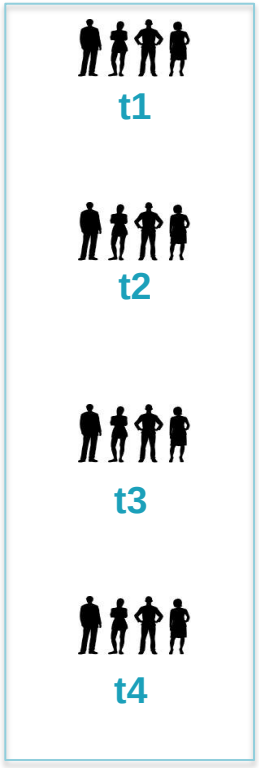
Human populations

Variants selection

Interactome mapping

Analysis 6

Primary tumour pathologic spread

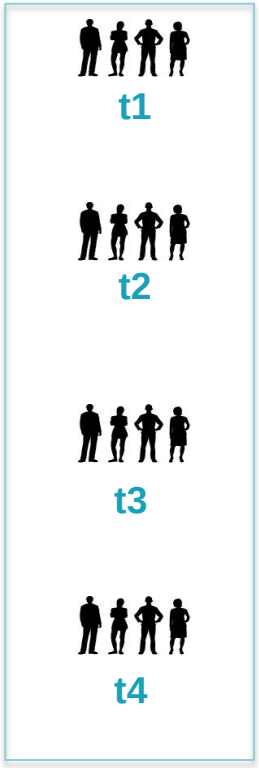


OBJECTIVE

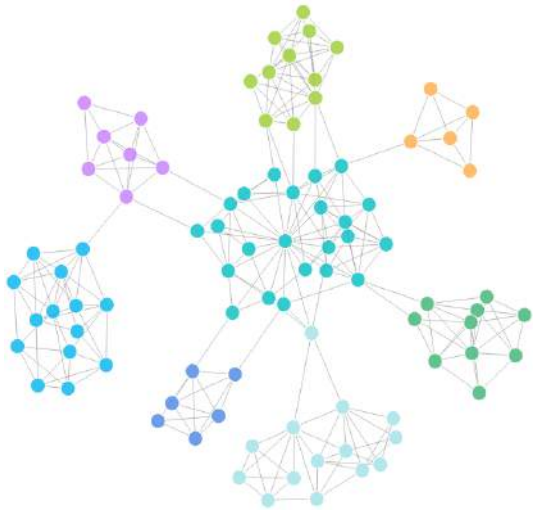
Temporal component of the multigenicity

Human populations → Variants selection → Interactome mapping → Analysis 6

Primary tumour pathologic spread



We identified those interactome modules concentrating cancer-specific mutations



OBJECTIVE

Temporal component of the multigenicity

Human populations → Variants selection → Interactome mapping → Analysis 6

Primary tumour pathologic spread

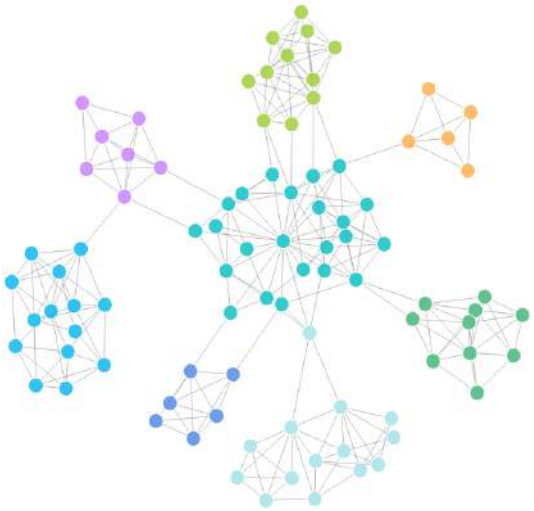
t1

t2

t3

t4

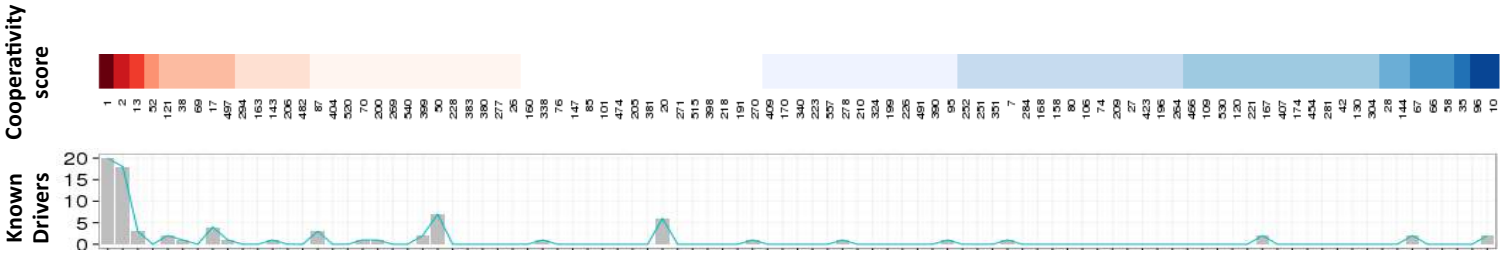
We identified those interactome modules concentrating cancer-specific mutations

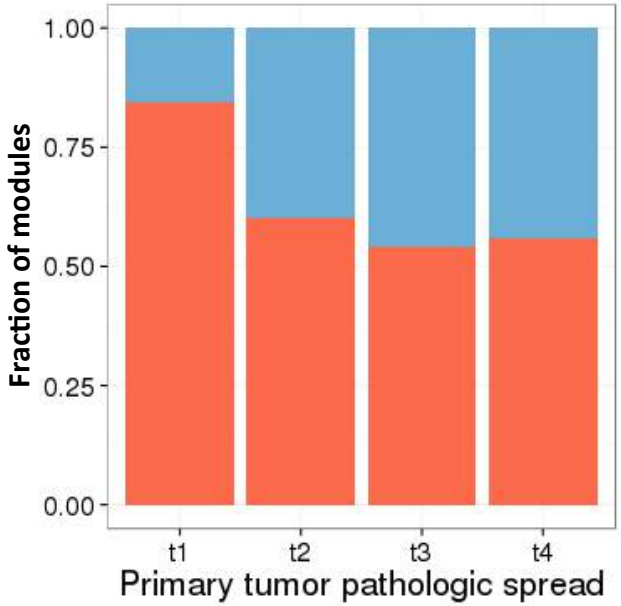


multigenecity score

RESULTS

Temporal component of the multigenicity

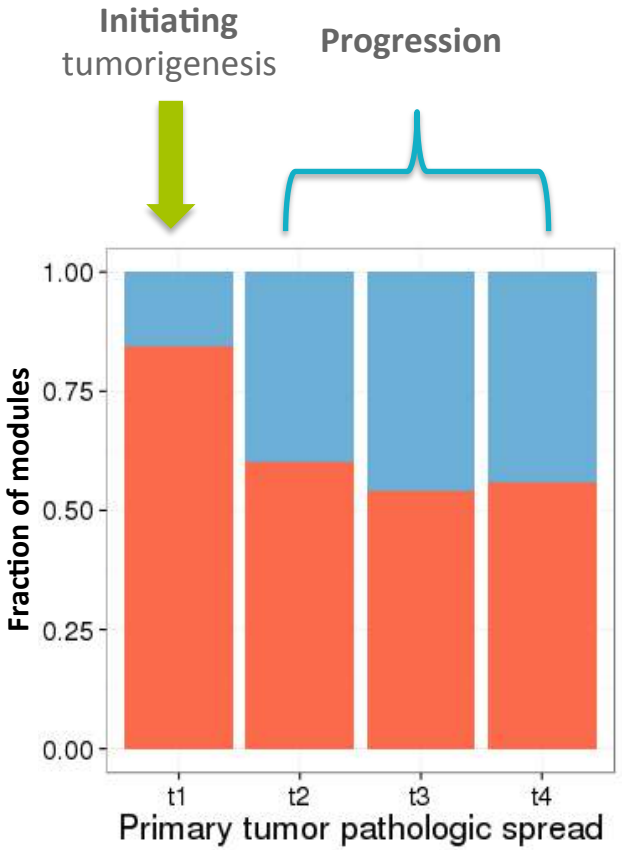




TNM system

RESULTS

Temporal component of the multigenicity

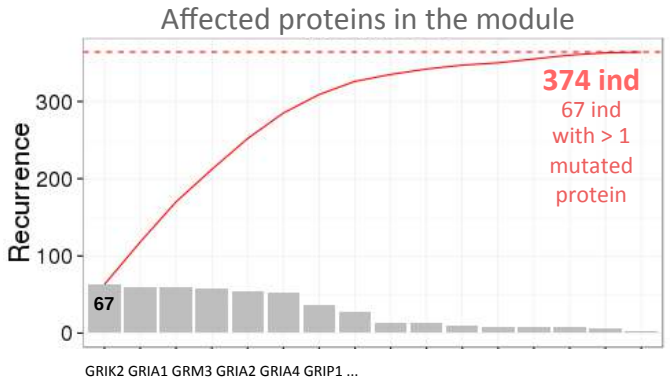


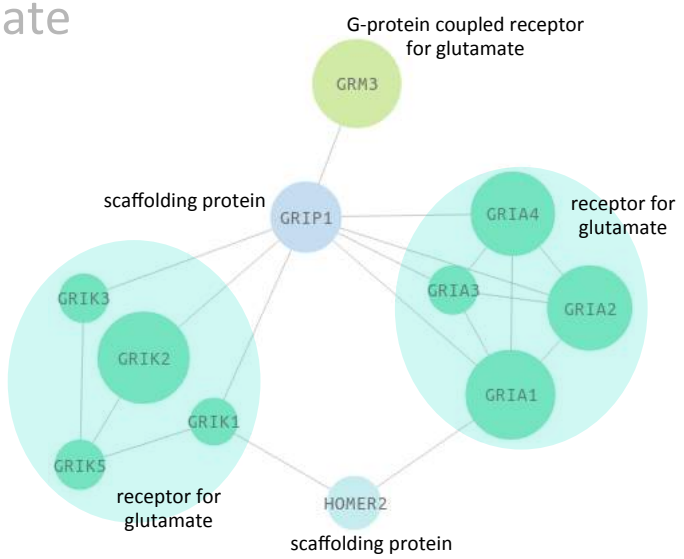
TNM system

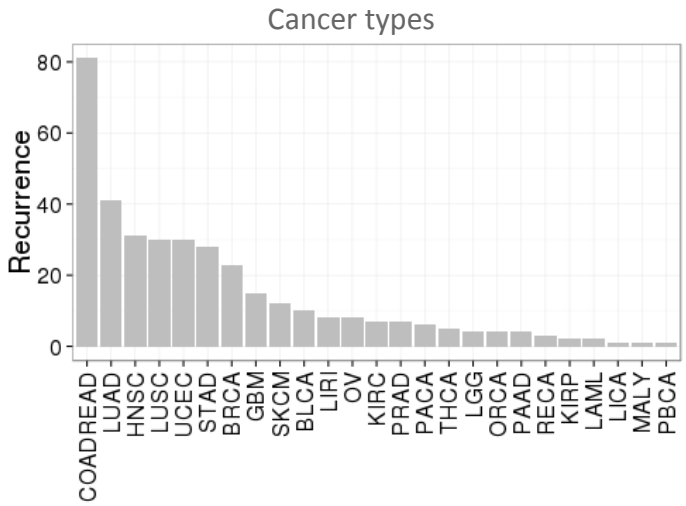
Those mutations acquired after initiation would be occupying multigenic modules and therefore, would be contributing to the tumour heterogeneity

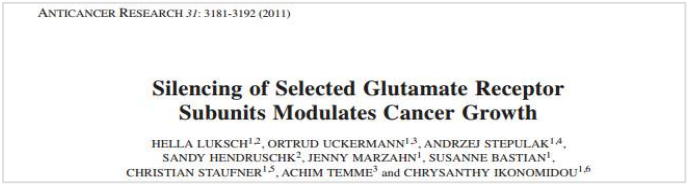
RESULTS

An example: The receptor for the glutamate








ANTICANCER RESEARCH *31*: 3181-3192 (2011)

Silencing of Selected Glutamate Receptor Subunits Modulates Cancer Growth

HELLA LUKSCH[1,2], ORTRUD UCKERMANN[1,3], ANDRZEJ STEPULAK[1,4], SANDY HENDRUSCHK[2], JENNY MARZAHN[1], SUSANNE BASTIAN[1], CHRISTIAN STAUFNER[1,5], ACHIM TEMME[3] and CHRYSANTHY IKONOMIDOU[1,6]


Molecular Pathways

Clin Cancer Res; 18(16) August 15, 2012

Clinical Cancer Research

Molecular Pathways: Dysregulated Glutamatergic Signaling Pathways in Cancer

Todd D. Prickett and Yardena Samuels

CONCLUSIONS

In healthy individuals, the deleterious character of a variant NOT ONLY depends on the damage that causes to the protein BUT ultimately is a system's property

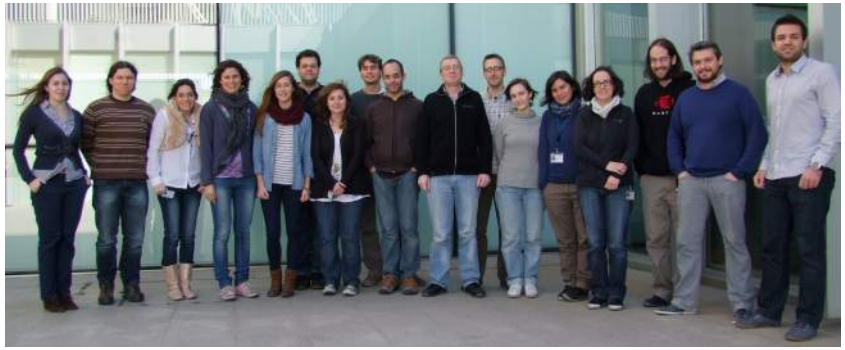
In cancer patients, it is known that the cancer initiation requires a driver

However, driver/modulator genes for cancer progression are not as clearly identifiable

Here, using a syetems oriented approach, we have identified driver/helper modules within the protein interactome related to the tumor pathologic spread and the cancer progression

Thank you!

Systems Genomics Lab



Collaborations

Medical Genome Project, Sevilla

JUNTA DE ANDALUCIA
CONSEJERÍA DE IGUALDAD, SALUD Y POLÍTICAS SOCIALES

MGP Medical Genome Project

Financial support

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

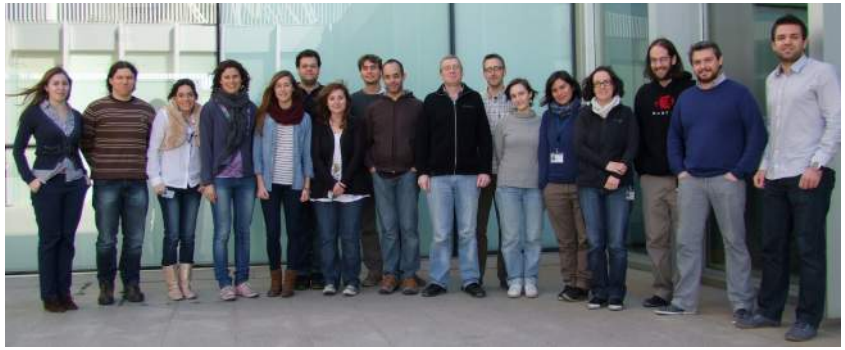
Instituto de Salud Carlos III

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

ciberer
Plataforma Bioinformática de Enfermedades Raras

Thank you!

Systems Genomics Lab



Collaborations

Medical Genome Project, Sevilla

JUNTA DE ANDALUCIA
CONSEJERÍA DE IGUALDAD, SALUD Y POLÍTICAS SOCIALES

MGP Medical Genome Project

Financial support

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Instituto de Salud Carlos III

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

ciberer
Plataforma Bioinformática de Enfermedades Raras

Pathogenic: contributes mechanistically to disease, but is not necessarily fully penetrant (i.e., may not be sufficient in isolation to cause disease).

Implicated: possesses evidence consistent with a pathogenic role, with a defined level of confidence.

Associated: significantly enriched in disease cases compared to matched controls.

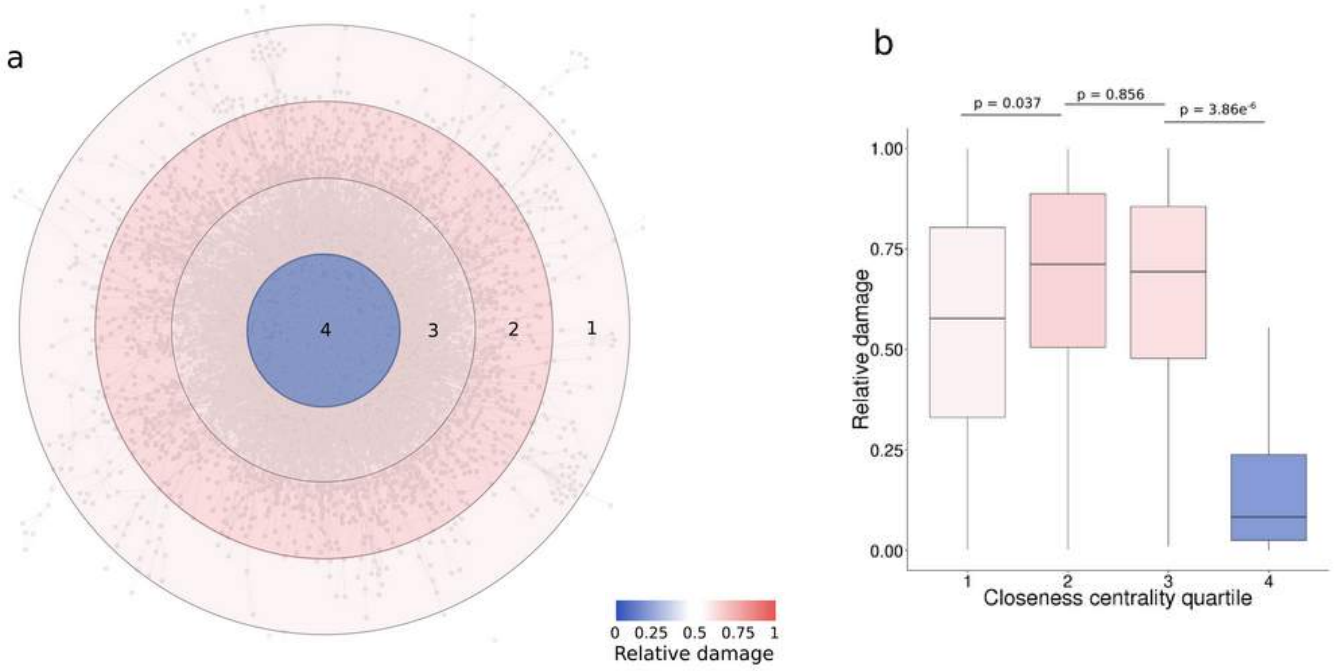
Damaging: alters the normal levels or biochemical function of a gene or gene product.

Deleterious: reduces the reproductive fitness of carriers, and would thus be targeted by purifying natural selection.

D. G. MacArthur et al., 2014

RESULTS

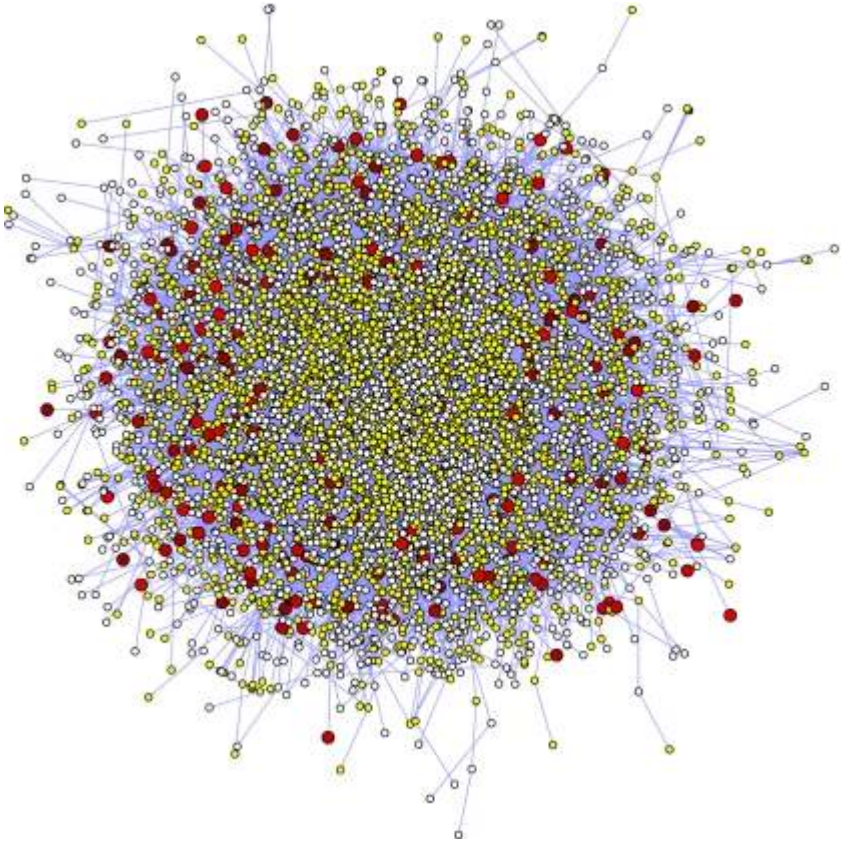
Deleterious variants observed in normal individuals tend to occur at the periphery of the interactome



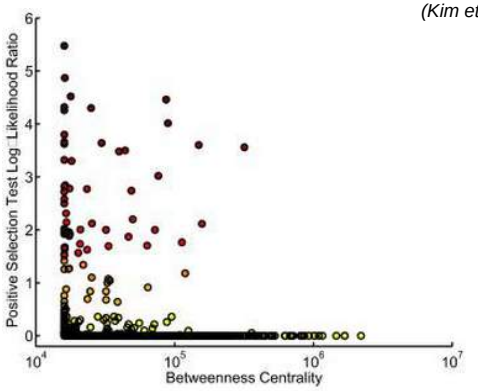
Luz Garcia-Alonso et al., 2014 (Under revision)

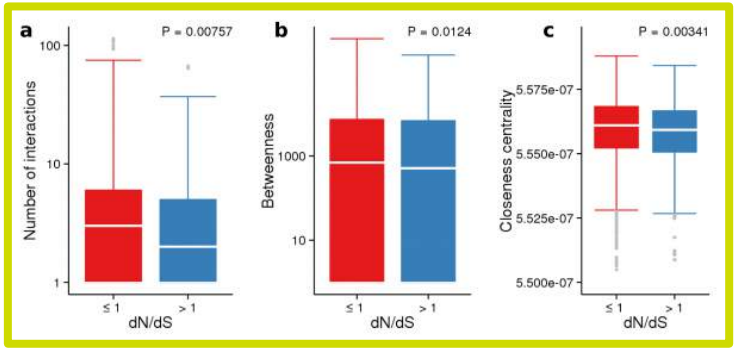
RESULTS

Positive selection at the protein network periphery



(Kim et al. PNAS, 2007)

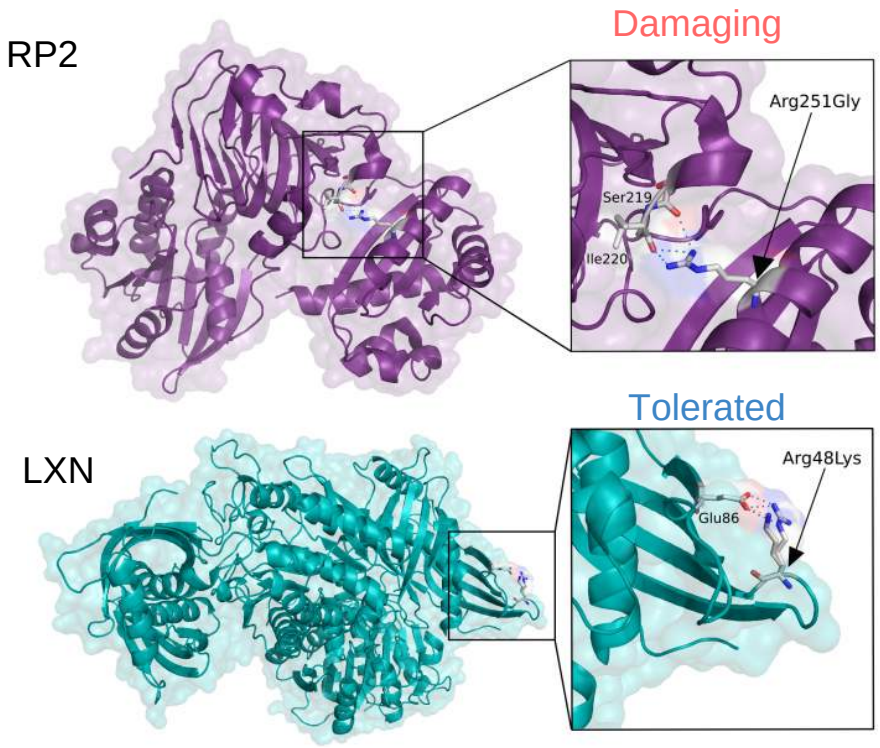




Luz Garcia-Alonso et al., 2014 (Under revision)

RESULTS

Deleterious variants in proteins of the interactome: Validation the predictions



In silico modeling of the mutations in the protein using its previously solved crystal structure

Molecular models of the human RP2 (a) and LXN (b) proteins and detailed view of the altered amino acids (Arg251Gly and Arg48Lys, respectively)

Luz Garcia-Alonso et al., 2014 (Under revision)