Relating genomic sequence variants with structural and network data provides insight about human disease

- Mutations in which genes are more likely to cause disease
- Mutations of which sites in those genes are more likely to cause disease
- We integrate sequence, structure and networks data to answer these questions
- Since now have much more data for polymorphisms in healthy humans (1000 Genomes) can answer these questions
- Also now have more structural data and larger networks (PPI, regulatory network and signaling pathway)

MUTATIONS IN WHICH GENES ARE MORE LIKELY TO CAUSE DISEASE

Fig 1 PPI

1a Disease genes tend to show higher degree centrality and betweeness centrality

1b allele freqs. of nonsyn snps are negatively correlated with degree (if remove p53 snps) so polymorphisms in hubs occur at low allele freqs.

1c allele freqs. of indels are negatively correlated with degree

1d picture of PPI network with hgmd genes and polymorphic genes

Fig 2 SIN

2a Cartoon of division of hubs into multi- and singlish- interface hubs

2b Multi- interface hubs have significantly more disease genes than singlish- hubs

2c Allele freqs of nonsyn snps are significantly lower in multi-interface hubs than singlish- hubs

Signaling pathways ?

1) Do disease genes tend to have higher degree in signaling pathways ?? No significance difference
2) allele freq positively correlated with indegree – not sure why
3) Enrichment of number of disease/ polymorphic genes in signaling pathways?

MUTATIONS AT WHICH SITES TEND TO CAUSE DISEASES

Fig 3. Sequence constraints (novel for deletions)

- 3a: Disease causing single nucleotide variants have lower dN/dS than polymorphic sites
- 3b: allele freq of snps is positively correlated with dn/ds per site
- 3c: disease deletion sites have lower gap scores
- 3d: alternate allele freq of deletions is positively correlated with gap scores

Fig 4: Structural constraints

- 4a: Disease variants tend to occur at structurally conserved sites within protein families
- 4b: structural alignments of protein structure with disease and 1KG snps