Page 1: [1] Deleted MG-Sun-night-edit 8/3/15 12:16 AM

2.2. Analyzing the functional impact [2.75pg]

We have extensive experience in network studies and functional interpretation of coding mutations. Variants in genes can have a wide spectrum of ,,,,effects, which are largely governed by the diverse biological networks in which the gene participates.

Page 1: [2] Moved to page 3 (Move #1)MG-Sun-night-edit 8/3/15 12:16 AM

We have integrated multiple biological networks to investigated gene functions. We found that functionally significant and highly conserved genes tend to be more central in various networks \cite{23505346} and positioned on the top level of regulatory networks \cite{22955619}. Incorporating multiple network and evolutionary properties, we have developed a computational method - NetSNP \cite{23505346} to quantify the indispensability of each gene. This method shows its strong potential for interpretation of variants involved in Mendelian diseases and in complex disorders probed by genome-wide association studies.

Page 1: [3] Deleted MG-Sun-night-edit 8/3/15 12:16 AM

,,,2We also plan to develop approaches to quantify variant-specific effects.

Page 2: [4] Deleted MG-Sun-night-edit 8/3/15 12:16 AM

,,,,tabs

Page 3: [5] Moved to page 2 (Move #4)MG-Sun-night-edit 8/3/15 12:16 AM

We have also developed specific tools to help quantify specific types of transcripts that require special processing, particularly pseudogenes and fusion transcripts \cite{25157146, 22951037, 20964841}. We have applied our expertise in RNA-Seq analysis to analyze and compare the transcriptomes of human, worm, and fly, using ENCODE and modENCODE datasets. We found a finding striking similarity between the processes regulating transcription in these three distant organisms \cite{21177976, 25164755, 22955620}.

Page 3: [6] Moved to page 3 (Move #5)MG-Sun-night-edit 8/3/15 12:16 AM

A powerful way to integrate diverse genomic data is through representations as networks.

Page 3: [7] Deleted MG-Sun-night-edit 8/3/15 12:16 AM

Further studies showed relationships between selection and protein network topology (for instance, quantifying selection in hubs relative to proteins on the network periphery\cite{18077332,23505346}).

Page 4: [8] Deleted MG-Sun-night-edit 8/3/15 12:16 AM

,,,,,SV in coding usually log.....Weill wi do better. ,,,,whole v part

Page 5: [9] Moved to page 2 (Move #2)MG-Sun-night-edit 8/3/15 12:16 AM

To identify pseudogenes in the human genome, we developed PseudoPipe, the first large-scale pipeline for genome wide human pseudogene annotation\cite{16574694}. We also obtained the “high confidence” pseudogenes by combining computational predictions with extensive manual curation\cite{22951037,25157146}, and identified

parent gene sequence from which the pseudogene arises based on their sequence comparisons\cite{22951037

Page 5: [10] Formatted	**MG-Sun-night-edit**	**8/3/15 12:16 AM**

Line spacing: multiple 1.15 li

Page 5: [11] Deleted	**MG-Sun-night-edit**	**8/3/15 12:16 AM**

We will prioritize genetic variants impacting non-coding RNA elements in a way that is consistent ,,,invert with the approach taken for TF binding sites,

Page 5: [12] Deleted	**MG-Sun-night-edit**	**8/3/15 12:16 AM**

[[MRS2MG (31 July) cut RNA variant burden tests (not sure whether this belongs here or with hotspots section) but can add back in.]]

Page 5: [13] Deleted	**MG-Sun-night-edit**	**8/3/15 12:16 AM**

In order to define structural variations with highly impactful events, we will assess their influence on

Page 6: [14] Deleted	**MG-Sun-night-edit**	**8/3/15 12:16 AM**

effic

Our