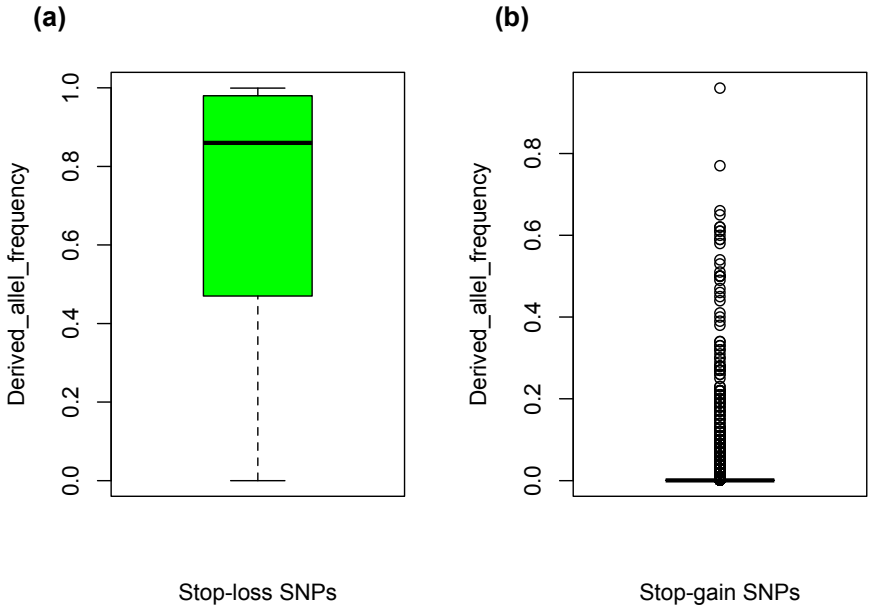
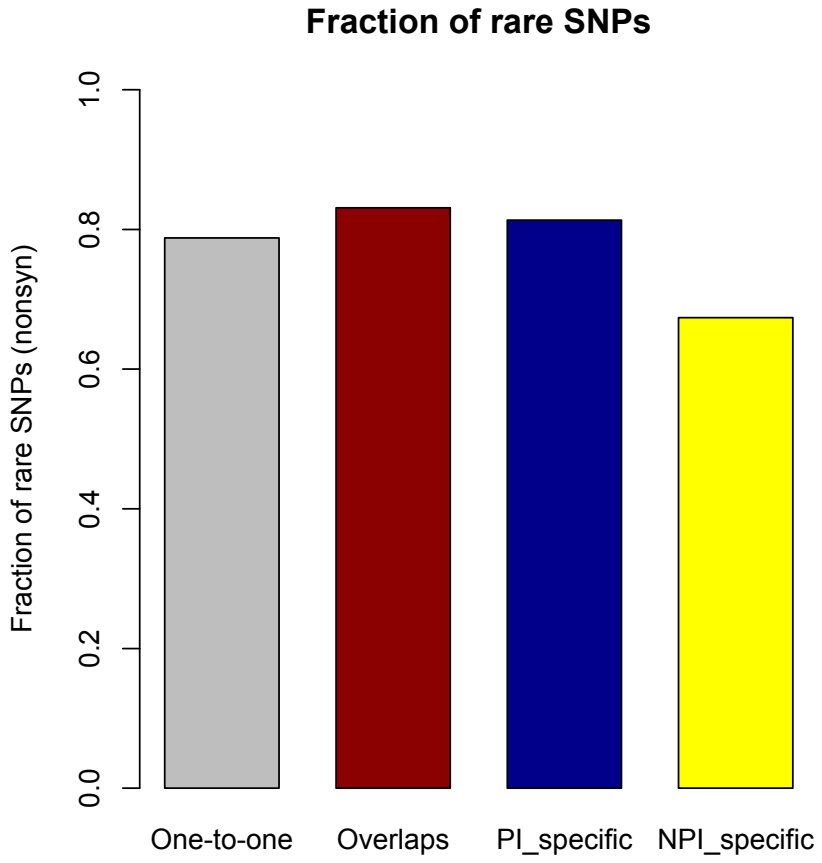
Electronic Supplementary Material (ESI) for Molecular BioSystems.
This journal is © The Royal Society of Chemistry 2015

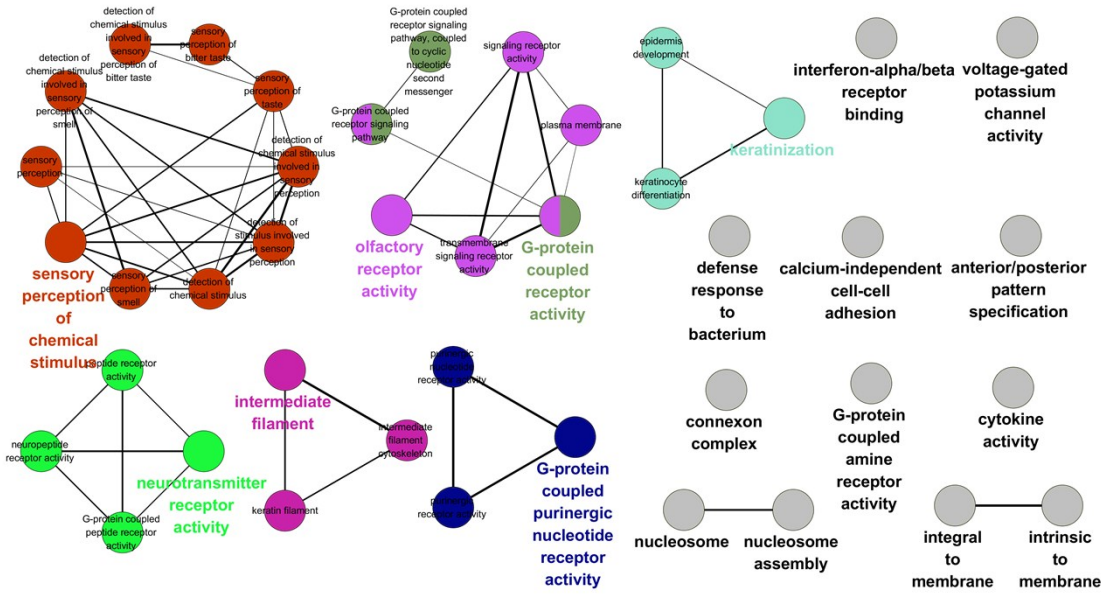


S1. Derived allele frequency of Stop-loss and Stop-gain SNPs. (a): The derived allele frequency of stop-loss SNPs in the 1000 Human Genome Project (median frequency is 0.83). (b): The derived allele frequency of stop-gain SNPs in the 1000 Human Genome Project (median frequency is 5e-04). In summary, the derived allele frequencies of stop-gain SNPs are remarkably lower than stop-loss SNPs.

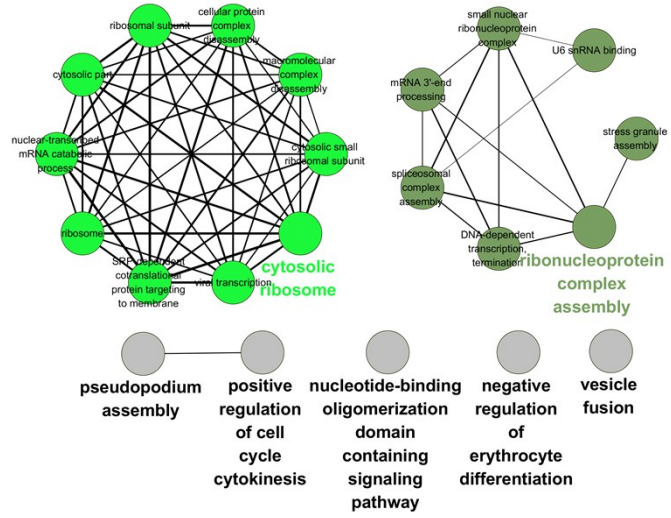


S2. Fraction of rare non-synonymous SNPs in different isoform coding categories Vertical axis represents the proportion of rare (DAF<0.5%) non-synonymous SNPs to total non-synonymous SNPs in different coding categories. The one-to-one means the protein-coding gene has only one protein-coding transcript. The proportion of rare non-synonymous in one-to-one as expected distribution, the fraction of rare non-synonymous SNPs in other coding categories are significant different (p-value<2.2e-16, binomial test).

(a)



(b)



S3. GO enrichment of two gene sets. (a): Non-synonymous SNPs that were solely present in PI-specific regions. (b): Non-synonymous SNPs that were solely present in NPI-specific regions. The enrichment result of PI-specific non-synonymous SNPs gene set was largely overlapped with the result of the gene set with stop-gain SNPs.

Table S1. The NMD-targeting prediction based on 50-55-nucleotide rule

	Positive on NMD prediction	Negative on NMD prediction	Total
PI-specific SNPs	759	171	930
NPI-specific SNPs	345	129	474

The number of stop-gain SNPs predicted as NMD positive and negative.

Table S2. The number of number of times that randomised values is less than the observed value at substitution dataset.

Number of gene sampled (percent of the gene pool)	The number of random values less than the observed value at non-synonymous dataset	The number of random values less than the observed value at stop-gain dataset
5%	702	998
10%	783	1000
15%	822	1000
20%	867	1000
25%	901	1000
30%	925	1000

It means that in 1000 random experiments, how many times the ratio of NPI-specific SNPs/all SNPs was less than the observed ratio in sampled gene set.

Table S3. The number of PI-specific, NPI-specific, and overlapping variations with different types of substitutions in multi-transcript genes.

	Frame_shift_Indel	Stop_gain/loss	Non_Synonymous	Synonymous	Length(bp)
Overlaps	289	2,786	161,160	118,522	19,314,235
PI_specific	108	937	48,876	33,449	5,654,482
NPI_specific	190	476	11,831	5,533	1,275,329

Table S4. The number of PI-specific, NPI-specific, and overlapping SNPs associated with different types of substitutions at different derived allele frequencies.

	Stop gain			Non-synonymous			Synonymous			Length
	Rare	Low	Common	Rare	Low	Common	Rare	Low	Common	(bp)
Overlaps	2,480	129	41	130,302	17,075	9,449	84,804	17,595	13,131	19,314,235
NPI-specific	320	83	42	7,646	2,174	1,531	3,534	1,018	787	1,275,329
PI-specific	812	57	22	38,179	5,502	3,261	23,282	5,118	3,848	5,654,482