Intra-genic Exon Duplications in the Human Transcriptome
Anke Busch & Peter F. Stadler
studla@bioinf.uni-leipzig.de

Supplemental File 5

List of significant GO terms characterizing exon groups depending on their maximal copy number of the same exon as well as on their general fraction of duplicated exons.

Table 1: Shared GO terms among the genes in exon groups that include at least 10 duplicates of a single exon. A) component GO terms; B) function GO terms; C) process GO terms.

A)

p-value	#occurences	percentage among all genes	GO
3.00e-51	24/79	0.15%	collagen
1.21e-48	30/79	0.48%	extracellular matrix part
7.62e-45	36/79	1.33%	proteinaceous extracellular matrix
3.58e-42	36/79	1.57%	extracellular matrix
2.93e-29	38/79	4.24%	extracellular region part
1.07e-25	46/79	9.20%	extracellular region
9.49e-23	10/79	0.05%	fibrillar collagen
2.03e-21	15/79	0.31%	basement membrane
1.01e-13	6/79	0.03%	collagen type IV
7.08e-13	6/79	0.03%	sheet-forming collagen
4.02e-09	5/79	0.04%	anchoring collagen
1.51e-07	4/79	0.03%	FACIT collagen
1.26e-06	4/79	0.04%	M band
3.13e-06	3/79	0.01%	collagen type V
3.13e-06	3/79	0.01%	collagen type IX

B)

p-value	#occurences	percentage among all genes	GO
3.66e-39	24/79	0.36%	extracellular matrix structural constituent
5.24e-29	33/79	2.74%	structural molecule activity
1.54e-11	19/79	2.68%	calcium ion binding
4.09e-11	6/79	0.05%	platelet-derived growth factor binding
1.71e-10	5/79	0.03%	extracellular matrix structural constituent conferring tensile strength
1.32e-07	7/79	0.27%	integrin binding
7.15e-06	7/79	0.47%	growth factor binding

C)

p-value	#occurences	percentage among all genes	GO
5.15e-20	16/79	0.45%	extracellular matrix organization
8.31e-20	29/79	3.48%	cell adhesion
8.60e-20	29/79	3.49%	biological adhesion
3.31e-16	16/79	0.76%	extracellular structure organization
6.95e-15	42/79	13.04%	multicellular organismal development
6.06e-14	9/79	0.11%	collagen fibril organization
2.43e-13	42/79	14.37%	developmental process
4.77e-13	48/79	19.63%	multicellular organismal process
4.22e-12	22/79	3.38%	tissue development
8.15e-12	31/79	8.08%	organ development
7.85e-11	34/79	10.78%	system development
1.14e-10	35/79	11.62%	anatomical structure development
1.20e-09	12/79	0.85%	epidermis development
2.98e-09	12/79	0.91%	ectoderm development
3.43e-08	13/79	1.42%	blood vessel development
4.60e-08	13/79	1.45%	vasculature development
4.88e-07	6/79	0.14%	skin development
8.03e-07	30/79	11.81%	cellular component organization
1.67e-06	12/79	1.59%	skeletal system development
4.06e-06	20/79	5.64%	anatomical structure morphogenesis

Table 2: Shared GO terms among the genes in exon groups that include at most 2 duplicates of the same exon. A) component GO terms; B) function GO terms; C) process GO terms.

A)

p-value	#occurences	percentage among all genes	GO
1.71e-14	925/1217	65.34%	cell part
1.75e-14	925/1217	65.35%	cell
1.03e-13	148/1217	6.10%	cytoskeleton
1.14e-13	220/1217	10.56%	non-membrane-bounded organelle
1.14e-13	220/1217	10.56%	intracellular non-membrane-bounded organelle
4.75e-13	728/1217	48.85%	intracellular
1.30e-11	110/1217	4.24%	cytoskeletal part
4.70e-11	699/1217	47.28%	intracellular part
2.88e-10	364/1217	21.49%	organelle part
1.10e-09	357/1217	21.18%	intracellular organelle part
5.32e-08	596/1217	40.30%	organelle
8.42e-08	594/1217	40.23%	intracellular organelle
1.57e-07	503/1217	33.15%	cytoplasm
3.31e-07	68/1217	2.54%	microtubule cytoskeleton
4.50e-07	35/1217	0.91%	apical part of cell
1.41e-06	30/1217	0.73%	apical plasma membrane
4.22e-06	212/1217	12.06%	protein complex
4.39e-06	167/1217	8.96%	plasma membrane part

B)

p-value	#occurences	percentage among all genes	GO
2.23e-29	853/1217	54.12%	binding
3.91e-19	584/1217	35.06%	protein binding
3.97e-12	150/1217	6.47%	adenyl ribonucleotide binding
5.30e-12	148/1217	6.37%	ATP binding
1.27e-11	72/1217	2.22%	cytoskeletal protein binding
6.58e-11	152/1217	6.82%	adenyl nucleotide binding
1.01e-10	33/1217	0.60%	motor activity
1.53e-10	154/1217	7.01%	nucleoside binding
1.91e-10	153/1217	6.97%	purine nucleoside binding
3.84e-10	87/1217	3.16%	nucleoside-triphosphatase activity
3.95e-10	52/1217	1.41%	actin binding
1.08e-09	88/1217	3.28%	pyrophosphatase activity
1.34e-09	88/1217	3.29%	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
1.77e-09	88/1217	3.31%	hydrolase activity, acting on acid anhydrides
1.16e-08	163/1217	7.99%	ribonucleotide binding
1.16e-08	163/1217	7.99%	purine ribonucleotide binding
3.80e-08	190/1217	9.89%	nucleotide binding
3.94e-08	190/1217	9.90%	hydrolase activity
4.66e-08	57/1217	1.84%	GTPase regulator activity
5.82e-08	49/1217	1.46%	ATPase activity
9.87e-08	165/1217	8.35%	purine nucleotide binding
1.04e-07	57/1217	1.88%	nucleoside-triphosphatase regulator activity
1.38e-06	42/1217	1.26%	small GTPase regulator activity
6.43e-06	69/1217	2.77%	calcium ion binding
8.67e-06	19/1217	0.34%	calcium channel activity

C)

p-value	#occurences	percentage among all genes	GO
7.48e-16	757/1217	49.94%	cellular process
4.38e-10	269/1217	14.56%	localization
1.07e-07	468/1217	30.04%	biological regulation
5.36e-07	227/1217	12.58%	establishment of localization
8.02e-07	224/1217	12.43%	transport
1.74e-06	226/1217	12.67%	cellular component organization
2.30e-06	41/1217	1.18%	regulation of small GTPase mediated signal transduction
3.87e-06	232/1217	13.21%	cellular component organization or biogenesis

Table 3: Shared GO terms among the genes in exon groups that include at most 5 copies of a single exon containing at least one exon that has three or more copies. A) component GO terms; B) function GO terms; C) process GO terms.

A)			
p-value	#occurences	percentage among all genes	GO
2.13e-07	20/236	1.58%	extracellular matrix
3.63e-06	17/236	1.33%	proteinaceous extracellular matrix

B)			
p-value	#occurences	percentage among all genes	GO
nothing found			

C)			
p-value	#occurences	percentage among all genes	GO
nothing found			

Table 4: Shared GO terms among the genes in exon groups that include at most 9 copies of a single exon containing at least one exon that has six or more copies. A) component GO terms; B) function GO terms; C) process GO terms.

A)			
p-value	#occurences	percentage among all genes	GO
2.28e-07	9/42	1.33%	proteinaceous extracellular matrix
2.78e-07	5/42	0.15%	collagen
9.70e-07	9/42	1.58%	extracellular matrix
2.92e-06	6/42	0.48%	extracellular matrix part
7.58e-06	12/42	4.28%	extracellular region part

B)			
p-value	#occurences	percentage among all genes	GO
2.95e-07	6/42	0.34%	extracellular matrix structural constituent

C)			
p-value	#occurences	percentage among all genes	GO
nothing found			

Table 5: Shared GO terms among the genes in exon groups that include more than 0 but equal or less than 25% duplicated exons. A) component GO terms; B) function GO terms; C) process GO terms.

A)

p-value	#occurences	percentage among all genes	GO
5.01e-25	943/1190	65.34%	cell part
5.17e-25	943/1190	65.35%	cell
5.15e-24	171/1190	6.10%	cytoskeleton
2.87e-23	245/1190	10.56%	non-membrane-bounded organelle
2.87e-23	245/1190	10.56%	intracellular non-membrane-bounded organelle
1.35e-22	753/1190	48.85%	intracellular
7.81e-21	728/1190	47.28%	intracellular part
3.88e-20	128/1190	4.24%	cytoskeletal part
3.75e-18	390/1190	21.49%	organelle part
1.28e-16	380/1190	21.18%	intracellular organelle part
6.54e-16	624/1190	40.30%	organelle
1.09e-15	534/1190	33.15%	cytoplasm
1.24e-15	622/1190	40.23%	intracellular organelle
1.71e-15	85/1190	2.54%	microtubule cytoskeleton
1.66e-11	231/1190	12.06%	protein complex
3.65e-10	260/1190	14.44%	macromolecular complex
1.24e-08	20/1190	0.28%	myosin complex
1.26e-08	44/1190	1.21%	actin cytoskeleton
5.33e-08	27/1190	0.54%	contractile fiber
6.09e-08	26/1190	0.50%	myofibril
9.91e-08	42/1190	1.19%	microtubule organizing center
1.05e-07	45/1190	1.33%	proteinaceous extracellular matrix
2.72e-07	25/1190	0.50%	contractile fiber part
3.44e-07	80/1190	3.29%	cell projection
3.46e-07	23/1190	0.43%	sarcomere
8.97e-07	167/1190	8.96%	plasma membrane part
2.32e-06	40/1190	1.22%	microtubule
2.58e-06	47/1190	1.58%	extracellular matrix
3.30e-06	521/1190	36.06%	membrane-bounded organelle
3.39e-06	147/1190	7.76%	nuclear part
3.55e-06	29/1190	0.73%	apical plasma membrane
4.05e-06	520/1190	36.03%	intracellular membrane-bounded organelle

B)

p-value	#occurences	percentage among all genes	GO
8.07e-43	874/1190	54.11%	binding
4.25e-31	615/1190	35.06%	protein binding
3.21e-19	85/1190	2.22%	cytoskeletal protein binding
1.29e-16	62/1190	1.41%	actin binding
4.69e-15	38/1190	0.60%	motor activity
3.14e-14	95/1190	3.16%	nucleoside-triphosphatase activity
1.12e-13	96/1190	3.28%	pyrophosphatase activity
1.44e-13	96/1190	3.29%	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides
2.01e-13	96/1190	3.31%	hydrolase activity, acting on acid anhydrides
2.70e-13	151/1190	6.47%	adenyl ribonucleotide binding
3.69e-13	149/1190	6.37%	ATP binding
1.29e-12	84/1190	2.77%	calcium ion binding
4.92e-12	153/1190	6.82%	adenyl nucleotide binding
1.17e-11	155/1190	7.01%	nucleoside binding
1.48e-11	154/1190	6.97%	purine nucleoside binding
2.15e-10	61/1190	1.84%	GTPase regulator activity
2.64e-10	166/1190	7.99%	ribonucleotide binding
2.64e-10	166/1190	7.99%	purine ribonucleotide binding
5.33e-10	61/1190	1.88%	nucleoside-triphosphatase regulator activity
7.04e-10	52/1190	1.46%	ATPase activity
8.87e-10	193/1190	9.89%	nucleotide binding
2.64e-09	168/1190	8.35%	purine nucleotide binding
5.69e-09	190/1190	9.90%	hydrolase activity
4.22e-08	279/1190	16.44%	metal ion binding
6.89e-08	281/1190	16.66%	ion binding
8.41e-08	280/1190	16.62%	cation binding
2.22e-07	43/1190	1.26%	small GTPase regulator activity
9.81e-07	20/1190	0.34%	calcium channel activity
1.06e-06	28/1190	0.64%	guanyl-nucleotide exchange factor activity
1.38e-06	49/1190	1.64%	ion channel activity
3.22e-06	49/1190	1.68%	substrate-specific channel activity

C)

p-value	#occurences	percentage among all genes	GO
3.27e-27	784/1190	49.94%	cellular process
8.11e-17	261/1190	12.67%	cellular component organization
6.98e-16	285/1190	14.50%	localization
1.35e-15	265/1190	13.21%	cellular component organization or biogenesis
3.75e-13	101/1190	3.54%	cell adhesion
4.07e-13	101/1190	3.54%	biological adhesion
4.81e-12	481/1190	30.04%	biological regulation
1.56e-11	177/1190	8.31%	cellular component organization at cellular level
7.85e-11	182/1190	8.78%	cellular component organization or biogenesis at cellular level
2.20e-10	236/1190	12.51%	establishment of localization
6.29e-10	232/1190	12.36%	transport
6.80e-08	438/1190	28.32%	regulation of biological process
1.65e-07	417/1190	26.85%	regulation of cellular process
1.76e-07	43/1190	1.20%	microtubule-based process
2.76e-07	134/1190	6.46%	organelle organization
3.48e-07	42/1190	1.18%	regulation of small GTPase mediated signal transduction
2.21e-06	62/1190	2.28%	cytoskeleton organization
6.69e-06	75/1190	3.09%	cell cycle process
7.37e-06	36/1190	1.01%	regulation of Ras protein signal transduction

Table 6: Shared GO terms among the genes in exon groups that include between 25% and 50% duplicated exons. A) component GO terms; B) function GO terms; C) process GO terms.

A)

p-value	#occurences	percentage among all genes	GO
6.99e-12	28/299	1.58%	extracellular matrix
7.82e-11	69/299	9.22%	extracellular region
3.74e-10	24/299	1.33%	proteinaceous extracellular matrix
2.91e-07	13/299	0.48%	extracellular matrix part
3.54e-06	36/299	4.28%	extracellular region part

B)

p-value	#occurences	percentage among all genes	GO
nothing found			

C)

p-value	#occurences	percentage among all genes	GO
nothing found			

Table 7: Shared GO terms among the genes in exon groups that include between 50% and 75% duplicated exons. A) component GO terms; B) function GO terms; C) process GO terms.

A)

p-value	#occurences	percentage among all genes	GO
1.78e-38	19/70	0.15%	collagen
5.24e-27	19/70	0.48%	extracellular matrix part
8.79e-23	22/70	1.33%	proteinaceous extracellular matrix
3.47e-21	22/70	1.58%	extracellular matrix
3.92e-20	9/70	0.05%	fibrillar collagen
2.41e-16	26/70	4.28%	extracellular region part
3.73e-14	6/70	0.03%	collagen type IV
2.60e-13	6/70	0.03%	sheet-forming collagen
8.66e-13	31/70	9.22%	extracellular region
3.13e-09	8/70	0.31%	basement membrane
1.69e-06	3/70	0.01%	collagen type V

B)

p-value	#occurences	percentage among all genes	GO
6.21e-30	19/70	0.34%	extracellular matrix structural constituent
7.86e-15	21/70	2.73%	structural molecule activity
6.33e-09	5/70	0.05%	platelet-derived growth factor binding
7.40e-08	4/70	0.03%	extracellular matrix structural constituent conferring tensile strength

C)

p-value	#occurences	percentage among all genes	GO
2.06e-17	14/70	0.44%	extracellular matrix organization
3.99e-14	14/70	0.74%	extracellular structure organization
2.96e-12	8/70	0.11%	collagen fibril organization
8.39e-08	17/70	3.54%	cell adhesion
8.54e-08	17/70	3.54%	biological adhesion

Table 8: Shared GO terms among the genes in exon groups that include between 75% and 100% duplicated exons. A) component GO terms; B) function GO terms; C) process GO terms.

A)

p-value	#occurences	percentage among all genes	GO
nothing found			

B)

p-value	#occurences	percentage among all genes	GO
nothing found			

C)

p-value	#occurences	percentage among all genes	GO
nothing found			