

Supplemental information for A.Samsonova *et al.*
Prediction of gene expression in embryonic structures of *Drosophila melanogaster*.

<i>Functional unit</i>	Training Set		Confirmed <i>de novo</i> Predictions	
<i>mat2pep</i>	(72)		(23)	
	\emptyset	\emptyset	gastrulation	1.218E-7
<i>tma2smusclep</i>	(35)		(75)	
	myogenesis	1.78404E-4	e & m interaction	2.445E-6
			ectoderm development	0.00223
			mesoderm development	2.5152E-8
<i>aep2egut</i>	(54)		(73)	
	\emptyset	\emptyset	\emptyset	\emptyset
<i>pep2ebrian</i>	(40)		(46)	
	‡‡ CNS development	1.59837E-5	nervous system development	2.4105E-7
	†† NB fate determination	3.30863E-4	** GMC fate determination	8.6352E-4
	ectoderm development	0.04427	ectoderm development	0.0024
<i>vna2lcard</i>	(34)		(202)	
	CNS development	4.61455E-6	CNS development	9.3945E-25
	NB fate determination	1.35891E-4	ectoderm development	3.0268E-20
	ventral cord development	0.01054	ventral cord development	6.3021E-15
			nervous system development	1.1861E-11
		NB fate determination	8.8756E-9	
		axon guidance	5.012E-8	
‡‡ Central nervous system		** ganglion mother cell fate determination		
ectoderm & mesoderm interaction		†† neuroblast fate determination		

Table 1: GO over-representation scores for the groups of genes that have been used to train the SVM classifier, i.e. where expression pattern is documented in the BDGP database (Training set) and for those genes where *in situ* expression pattern is annotated in FlyBase. Each cell in the table contains the number of genes in every functional unit and GO:Biological Process terms that are detected to be over-represented in it. The over-representation scores (e-scores) are calculated with GeneMerge software. \emptyset indicates that no over-represented GO terms are found for a given functional unit.