

Additional Data File 2

Comparison between profiles from

Chaussabel D, Sher A: **Mining microarray expression data by literature proling.** *Genome Biol* 2002, 3:1-16.

We re-clustered the 69 genes present in the clustergram in Figure 6 from Chaussabel *et al.* For each of the resulting clusters we constructed a text profile using our GO-based vocabulary. Complementary to the manuscript, we include profiles of clusters *f-g* below. Although profiles of clusters *a-e* are present in the manuscript, we reprint all info here, *along with the LocusLink ids used*, for the reader's convenience.

In the first two columns the HUGO names and corresponding LocusLink identifiers are provided. The terms extracted by Chaussabel *et al.* are shown in column three and contrasted to the TXTGate terms in column four. The TXTGate profiles are descriptive and match the terms from the initial analysis well.

HUGO Names	LocusLink ids	Terms Chaussabel	TXTGate terms (GO vocabulary)
CLUSTER (a)			
LPL	4023	Lipoprotein	lipoprotein
CD36L1	949	Density	lipas
LDLR	3949	Cholesterol	ldl
		Lipid	ldl_receptor
		Adipose	cholesterol
			hdl
			scaveng_receptor
			high_densiti_lipoprotein
			low_densiti_lipoprotein_receptor
			low_densiti_lipoprotein
			plasma
			lipid
			...
			densiti
CLUSTER (b)			
UPA	118471	Invasive	metalloproteinas
PLAUR	5329	Invasion	matrix
SERPIN	5265	Metastasis	metalloendopeptidas
MMP1	4312	UPAR	collagenas
MMP10	4319	UPA	extracellular_matrix
MMP14	4323	Plasminogen	alpha
SPARC	6678	Urokinase-type	upar
		Urokinase	plasminogen_activ
		Plasmin	interstiti
		Activator	invasion
		Collagenase	stromelysin_2
		Collagen	tumor

Matrix	gelatinas
MMP	stromelysin_1
Metalloproteinase	mmp
Molecule-1	urokinas
Adhesion	inhibitor
Vascular	increas
Endothelial	cell_surfac
	invas
	...
	matrix
	extracellular_matrix
	activ
	collagen
	cell_invasion
	vascular
	cell_adhesion
	adhesion
	endotheli
	plasmin
	plasminogen

CLUSTER (c)			
AMPD3	272	Adenosine	purinerg
ADA	100	A2A	adenosin
ADORA2A	135	A1	deaminas
ADORA3	140	Antagonist	p2
P2RX		Agonist	p2x
P2RX1	5023	NM	p1
P2RX7	5027		agonist
			receptor
			adenosin_receptor
			ada
			nucleosid
			a2a
			atp
			a3_adenosin_receptor
			triphosph
			a3
			a2a_adenosin_receptor
			extracellular
			erythrocyt
			antagonist
			select
			ion_channel
			amp
			a1
CLUSTER (d)			
IP10	3627	Interferon	tumor_necrosi_factor
MIP1A	6348	IFN-alpha	cytokin
MIP1B	6351	IFN	induc
IL8	3576	Interferon-gamma	interferon
STAT4	6775	IFN-gamma	inflammatori

IL12B	3593	Inducible	antigen
TNFRSF9	3604		lymphocyt_activ
TNFSF9	8744		stimul
SLAM	6504		chemokin
TNFRSF5	958		monocyt
CD83	9308		induct
FGL2	10875		...
TNFRSF1A	7132		ifn
TNF	7124		interferon_gamma
CD14	929		interferon_alpha
PTX3	5806		
IL6	3569		
MX1	4599		
GBP1	2633		
ISG15	9636		
IRF1	3659		
INDO	3620		

CLUSTER (e)			
--------------------	--	--	--

CKB	1152	Population	purinerg
AMPD3	272	Frequency	adenosin
ADA	100	Allele	deaminas
ADORA2A	135	Unrelated	lipoprotein
ADORA3	140	Families	determin
P2RX		Recessive	chromosom
P2RX1	5023	Autosomal	exon
P2RX7	5027	Disorder	p2
GEM	2669	Severe	pair
ARHH	399	Patient	dna
LPL	4023	Deficiency	receptor
CD36L1	949		region
LDLR	3949		diseas
BF	629		map
GALC	2581		line
LAMB3	3914		result
GJB2	2706		
TGFBI	7045		
JAG1	182		
DSCR1	1827		

CLUSTER (f)			
--------------------	--	--	--

CCR7	1236	Chemokine	chemokin
MCP1	6347	Chemotactic	inflammatori
RANTES	6352	Chemoattractant	monocyt
IP10	3627	Infiltration	cytokin
MIP1A	6348	Migration	macrophag
MIP1B	6351	CC	il_8
IL8	3576	Chemotaxis	chemokin_receptor
MGSA	2919	Neutrophil	chemoattract
MIP2A	2920	IL-8	induc
MIP2B	2921	Protein-1	stimul
PBEF	10135	MCP-1	phorbol
		Monocyte	blood

cc
 leukocyt
 lymphocyt
 lipopolysaccharid
 chemotaxi
 ...
 neutrophil
 migrat
 neutrophil_activ
 cell_migrat
 neutrophil_chemotaxi

CLUSTER (g)			
BIRC3	330	Interleukin	tumor_necrosi_factor
IER3	8870	IL	inflammatori
CFLAR	8837	IL-1beta	cytokin
TRAF1	7185	IL-1	monocyt
TNFAIP3	7128	IL-6	antigen
NFKBIA	4792	Inflammation	chemokin
IL15RA	3601	Cytokine	induc
STAT5A	6776	LPS	receptor
CCR7	1236	Lipopolysaccharide	lymphocyt
MCP1	6347	TNF-alpha	macrophag
RANTES	6352	TNF	...
IP10	3627	Bacterial	interleukin_1
MIP1A	6348		lipopolysaccharid
MIP1B	6351		interleukin
IL8	3576		bacteri
STAT4	6775		inflammatori_cell
IL12B	3593		
TNFRSF9	3604		
TNFSF9	8744		
SLAM	6504		
TNFRSF5	958		
CD83	9308		
FGL2	10875		
TNFRSF1A	7132		
TNF	7124		
CD14	929		
PTX3	5806		
IL6	3569		
IL1RN	3557		
CD163	9332		
MGSA	2919		
MIP2A	2920		
MIP2B	2921		
PBEF	10135		
CD44	960		
PECAM1	5175		
ICAM1	3383		