Transcriptome analysis reveals molecular profiles associated with evolving steps of monoclonal gammopathies

Lucía López-Corral,[1] Luis Antonio Corchete,[1] María Eugenia Sarasquete,[1] María Victoria Mateos,[1] Ramón García-Sanz,[1] Encarna Fermiñán,[2] Juan-José Lahuerta,[3] Joan Bladé,[4] Albert Oriol,[5] Ana Isabel Teruel,[6] María Luz Martino,[7] José Hernández,[8] Jesús María Hernández-Rivas,[1] Francisco Javier Burguillo,[9] Jesús F. San Miguel,[1] and Norma C. Gutiérrez[1]

[1]Servicio de Hematología del Hospital Universitario de Salamanca, IBSAL IBMCC (USAL-CSIC), Salamanca; [2]Centro de Investigación del Cáncer-IBMCC (USAL-CSIC), Salamanca; [3]Servicio de Hematología del Hospital 12 de Octubre, Madrid; [4]Servicio de Hematología del Hospital Clinic, Barcelona; [5]Servicio de Hematología del Hospital Germans Trias i Pujol, Badalona; [6]Servicio de Hematología del Hospital Universitario de Valencia; [7]Servicio de Hematología del Hospital Virgen del Rocío, Sevilla; [8]Servicio de Hematología de Hospital General de Segovia; and [9]Departamento de Química Física, Facultad de Farmacia, Universidad de Salamanca, Spain

---

doi:10.3324/haematol.2013.087809
Manuscript received on July 3, 2013. Manuscript accepted on May 6, 2014.
Correspondence: lucialopezcorral@usal.es

SUPLEMENTARY MATERIAL:

METHODS

Patients:

All the samples included in the present study were selected according to two sequential quality criteria: 1) purity > 95% for MM and SMM, and > 90% for MGUS and healthy donors; 2) at least 100 ng of high quality RNA. Since these thresholds are not easily reached in MGUS and healthy donors samples, a significant proportion of samples not fulfilling both of them were rejected. This information is provided in the table below.

Criteria	***MGUS*** *Total*	*Accepted*	*Rejected*	***Healty donors*** *Total*	*Accepted*	*Rejected*
Purity > 90%	58	26	32	28	7	21
100 ng RNA (high quality)	26	20	6	7	5	2

Comparative analysis with other gene expression datasets:

In order to validate the supervised analysis of the present gene expression data, two datasets including NPC, MGUS, SMM and MM samples were generated from three series records publicly available at Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/) belonging to two different institutions: GEO GSE5900 and GSE2658 (UAMS: NPC = 22, MGUS= 44, SMM= 12 and MM= 256) and GEO GSE6477 (Mayo Clinic: NPC = 15, MGUS = 22, SMM = 24 and MM = 73).

To reduce variations caused by different Affymetrix chip versions, those probes not present in the three arrays were subtracted from the analysis based on the " U133PlusVsHuGene_BestMatch" and "hgu133_vs_hgu133_plus_best_Match" archives.

A semi-supervised approach was used to evaluate if the significant genes selected in our study were able to identify the monoclonal gammopathy groups in the two external datasets. For this purpose, a SAM multiclass analysis was made using the R/Bioconductor package siggenes (http://www.bioconductor.org/packages/2.13/bioc/html/siggenes.html) to select differentially expressed genes (*q*-values $\leq$ 0.001) from the 4 groups (NPC, MGUS, SMM and MM). Then, the list of genes generated by the supervised learning was used for unsupervised clustering of samples including in the UAMS and Mayo Clinic datasets.

Venn diagrams were performed with significant probesets ($q \leq 0.05$) generated from the supervised analysis applied to each dataset using the equivalent "best match" with the Human Gene 1.0 probes.

Quantitative real-time PCR validation

qRT-PCR was performed by using TaqMan gene expression assay kits (Applied Biosystems Foster City, CA, USA): Hs00559441_m1 for APAF1, Hs00171642_m1 for VCAN, Hs00391048_m1 for MEGF1, Hs03454202_s1 for TERC1, Hs03806547_s1 for TPI1, Hs00185390_m1 for BAG1, Hs00605223_g1 for RPL7A, Hs00188930_m1 for BAD, Hs03296830_s1 for SNORD55, Hs-00178289_m1 for AKT1 and Hs_99999905_m1 for GAPDH as a control gene. Relative gene expression was calculated by the $2^{-\Delta Ct}$ method, $\Delta Ct = Ct_{(gene)} - Ct_{(GAPDH)}$(27). The Mann-Whitney U test was used to identify statistically significant differences between groups. The statistical analyses were conducted using SPSS (version 20.0; SPSS Inc., Chicago, IL, USA).

Supplementary Table S1:

Clinical and biological characteristics of the patients

Characteristics		***All Patients***
MGUS (n=20)		
Sex (%)	Male / Female	68 / 32
Age (years)		78 (39-85)
Ig subtype (%)	IgG / IgA	68 / 32
Bone Marrow Plasma Cells (%)		4 (2-8)
SMM (n=33)		
Sex (%)	Male / Female	52 / 48
Age (years)		65 (39-84)
Ig subtype (%)	IgG / IgA	55 / 45
Bone Marrow Plasma Cells (%)		14 (4-48)
MM (n=41)		
Sex (%)	Male / Female	47 / 53
Age (years)		67 (41-81)
Ig subtype	IgG / IgA / BJ protein	58 / 34 / 8
ISS	I / II / III	30 / 44 / 26
Bone Marrow Plasma Cells (%)		37 (10-95)

Supplementary Table S2:

Frequency of Genetic abnormalities in MGUS, SMM and MM patients

Genetic abnormalities	***MGUS (n =20)*** *Number of patients (%)*	***Smoldering MM (n=33)*** *Number of patients (%)*	***Symptomatic MM (n=41)*** *Number of patients (%)*	***P value***
IgH abnormality	4/20 (20%)	11/33 (33%)	13/41 (32%)	NS
t(11;14)	1/20 (5%)	3/33 (9%)	2/41 (4.9%)	NS
t(4;14)	1/20 (5%)	3/33 (9%)	2/41 (4.9%)	NS
t(14;16)	0/20 (0%)	2/33 (6.1%)	0/10 (0%)	NS
Other	2/20 (10%)	3/33 (9%)	9/41 (22%)	NS
13q deletion	7/20 (35%)	11/33 (33%)	18/41 (44%)	NS
1q gain	10/19 (52.6%)	16/33 (48.5%)	22/40 (55%)	NS

Supplementary Table S3: Probes identifying genes exclusive of myeloid lineage

Gene Symbol	Transcript Set ID	Gene Symbol	Transcript Set ID
S100A9	7905571	*ELANE*	8024056
MNDA	7906377	*CEACAM8*	8037222
CFHR1	7908488	*BPI*	8062444
S100A12	7920238	*LBP*	8062461
S100A8	7920244	*ICOSLG*	8070720
HBG1	7937990	*CSTA*	8082058
HBG2	7937993	*LTF*	8086607
HBB	7946033	*HCLS1*	8089911
HBB	7946041	*TLR2*	8097903
HBD	7946041	*PF4*	8100966
HBG1	7946048	*PPBP*	8100972
HBG1	7946054	*CXCL3*	8100984
CLEC6A	7953737	*CXCL2*	8100995
CLEC9A	7953925	*CXCL14*	8114249
LYZ	7957023	*CD14*	8114612
CD164	7960771	*PILRA*	8134814
CD163	7960794	*LYN*	8146500
CTSG	7978351	*DEFA4*	8149109
HBA1///2	7991762	*DEFA3*	8149116
HBA1	7991766	*DEFA3*	8149126
MPO	8016932	*FCN1*	8165011
LGALS3BP	8018976	*AZU1*	8024038

Supplementary Table S4: Genes deregulated in clonal PC compared to NPC*

			Score (d)		
Probes	**Gene Symbol**	**MGUS *vs.* NPC**	**SMM *vs.* NPC**	**MM *vs.* NPC**	**Gene Title**
7902227	**GADD45A**	6.14	8.81	6.37	growth arrest and DNA-damage-inducible, alpha
7942824	**RPS28**	6.13	4.82	8.90	ribosomal protein S28
8125835	**RPS10**	6.05	5.33	9.11	ribosomal protein S10
8043100	**TMSB10**	6.00	3.97	4.66	thymosin beta 10
7990916	**RPS17**	5.94	6.57	6.97	ribosomal protein S17
8027024	**UBA52**	5.70	5.41	6.27	ubiquitin A-52 residue ribosomal protein fusion product 1
7997702	**COX4I1**	5.64	5.78	5.57	cytochrome c oxidase subunit IV isoform 1
8009561	**RPL38**	5.45	4.48	6.89	ribosomal protein L38
8125059	**CLIC1**	5.17	5.23	4.30	chloride intracellular channel 1
8113873	**HINT1**	4.94	5.38	5.87	histidine triad nucleotide binding protein 1
8122136	**SNORA33**	4.94	5.76	4.60	nucleolar RNA, H/ACA box 33
7908407	**RPS27A**	4.76	4.35	4.81	ribosomal protein S27a
7899346	**CCDC72**	4.66	7.66	7.52	coiled-coil domain containing 72

7902448	**RPL17**	4.65	4.36	5.99	ribosomal protein L17
8048195	**RPL37A**	4.60	5.81	5.48	ribosomal protein L37a
8078984	**RPL14**	4.54	4.80	4.91	ribosomal protein L14
8136039	**ATP6V1F**	4.47	3.96	5.56	ATPase, H+ transporting, lysosomal 14kDa, V1 subunit F
8150872	**RPS20**	4.46	3.95	4.66	ribosomal protein S20
8076417	**NDUFA6**	4.44	3.87	4.65	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex,
7935810	**NDUFB8**	4.43	4.52	5.15	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 8
8086538	**LOC644714**	-10.16	-6.64	-8.38	hypothetical LOC644714
8009995	**UNK**	-7.74	-4.56	-5.22	unkempt homolog (Drosophila)
8093456	**C4orf42**	-7.71	-4.47	-4.67	chromosome 4 open reading frame 42
8020495	**CABLES1**	-7.45	-5.34	-7.96	Cdk5 and Abl enzyme substrate 1
8018708	**UBE2O**	-7.03	-3.99	-5.35	ubiquitin-conjugating enzyme E2O
7982020	**SNORD115-7**	-6.90	-5.19	-5.24	small nuclear ribonucleoprotein polypeptide N
8079334	**LIMD1**	-6.82	-4.04	-4.09	LIM domains containing 1 /// hypothetical LOC644714
7967810	**GOLGA3**	-6.77	-4.71	-5.44	golgin A3
8026533	**HSH2D**	-6.71	-4.60	-5.42	hematopoietic SH2 domain containing
8081214	**GPR15**	-6.68	-4.76	-6.99	G protein-coupled receptor 15
8138613	**OSBPL3**	-6.61	-4.06	-4.61	oxysterol binding protein-like 3
8073081	**APOBEC3F**	-6.53	-6.83	-5.36	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3F
7982052	**SNORD115-23**	-6.43	-6.45	-6.44	small nucleolar RNA, C/D box 115-23
7982046	**SNORD115-5**	-6.27	-4.32	-4.31	small nucleolar RNA, C/D box 115-5
7980523	**GTF2A1**	-6.27	-6.19	-10.06	general transcription factor IIA, 1, 19/37kDa
7982094	**SNORD115-44**	-6.26	-4.49	-4.49	small nucleolar RNA, C/D box 115-44
8016546	**ZNF652**	-6.25	-5.44	-4.80	zinc finger protein 652
8001048	**C16orf58 /// SLC5A2**	-6.00	-4.37	-4.11	chromosome 16 open reading frame 58
8072153	**CCDC117**	-5.99	-5.14	-4.65	coiled-coil domain containing 117
7982056	**SNORD115-25**	-5.97	-4.08	-4.72	small nucleolar RNA, C/D box 115-25
7965964	**SLC41A2**	-5.79	-5.49	-6.12	solute carrier family 41, member 2
8124459	**ZNF322A**	-5.76	-4.89	-5.80	zinc finger protein 322A /// zinc finger protein 322B
7982050	**SNORD115-6**	-5.73	-4.02	-3.93	small nucleolar RNA, C/D box 115-6 s
7982024	**SNORD115-24**	-5.72	-4.07	-3.97	small nucleolar RNA, C/D box 115-24
7982016	**SNORD115-4**	-5.72	-4.07	-3.97	small nucleolar RNA, C/D box 115-4
7982028	**SNORD115-3**	-5.70	-4.03	-3.92	small nucleolar RNA, C/D box 115-34
7982092	**SNORD115-9**	-5.70	-4.03	-3.92	small nucleolar RNA, C/D box 115-9
7982058	**SNORD115-11**	-5.70	-4.02	-3.91	small nucleolar RNA, C/D box 115-11
7918457	**KCNA3**	-5.69	-4.38	-6.43	potassium voltage-gated channel
8023766	**RTTN**	-5.65	-5.28	-5.38	rotatin
7985364	**IL16 ///**	-5.65	-4.13	-4.12	interleukin 16 (lymphocyte chemoattractant factor)
8043585	**CIAO1**	-5.64	-4.08	-4.33	cytosolic iron-sulfur protein assembly 1
7982070	**SNORD115-32**	-5.62	-5.02	-4.71	small nucleolar RNA, C/D box 115-32
7993478	**ABCC1**	-5.62	-4.40	-3.96	ATP-binding cassette, sub-family C (CFTR/MRP), member 1
8014551	**SYNRG**	-5.62	-6.59	-7.48	synergin, gamma
7982032	**SNORD115-10**	-5.55	-4.14	-3.91	small nucleolar RNA, C/D box 115-10
7982008	**SNORD115-12**	-5.55	-4.14	-3.91	small nucleolar RNA, C/D box 115-12
7982038	**SNORD115-**	-5.55	-4.14	-3.91	small nucleolar RNA, C/D box 115-13

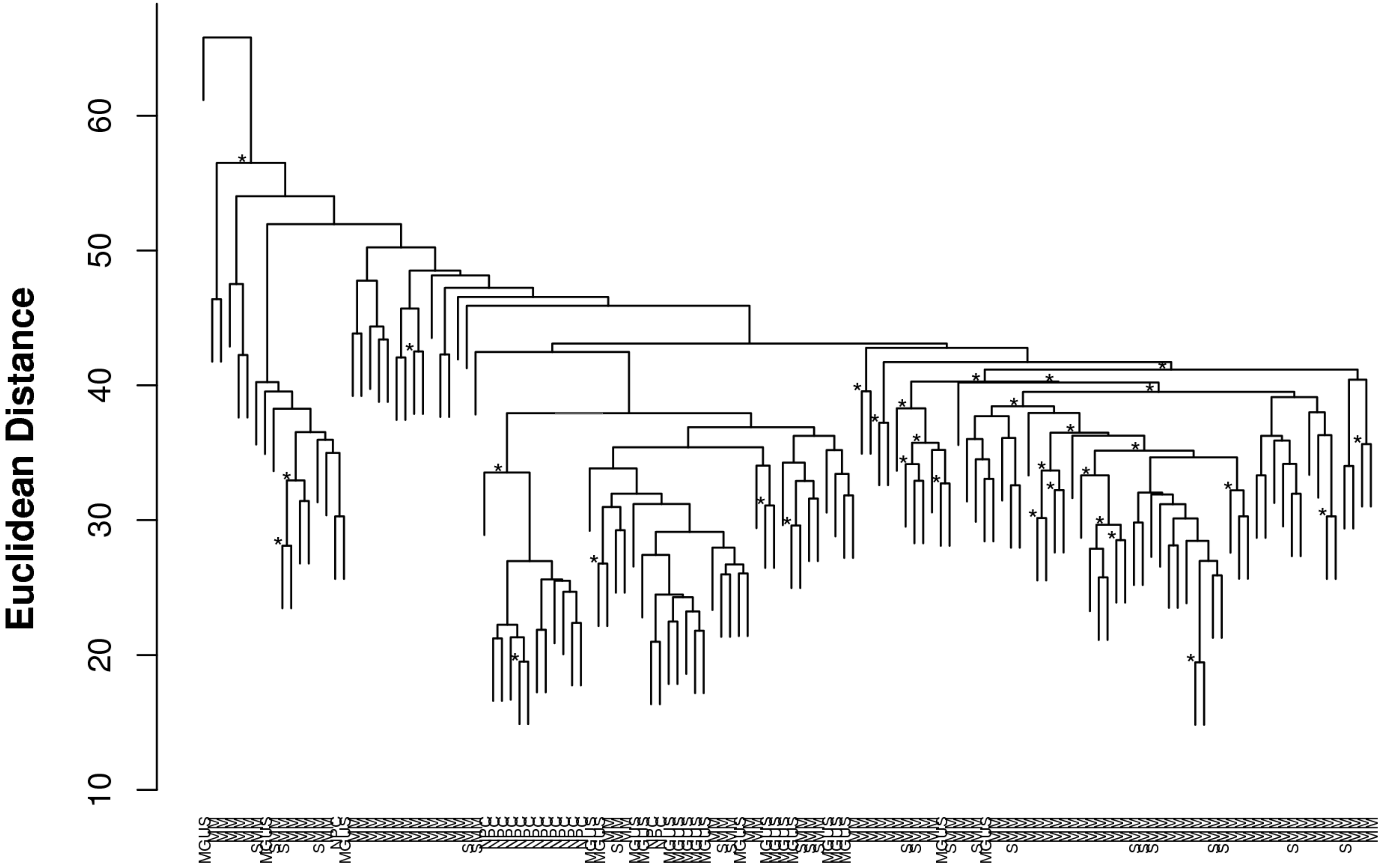
	13				
8065444	**ACSS1**	-5.55	-3.93	-4.20	acyl-CoA synthetase short-chain family member 1
8065569	**BCL2L1**	-5.46	-5.00	-4.46	BCL2-like 1
8099721	**SEL1L3**	-5.44	-4.19	-5.01	sel-1 suppressor of lin-12-like 3 (C. elegans)
8121043	**ORC3**	-5.42	-4.74	-6.95	origin recognition complex, subunit 3
7981726	**LOC100508797**	-5.40	-10.73	-7.44	hypothetical protein LOC100508797
7998666	**SNORA64**	-5.31	-5.31	-4.03	ribosomal protein S2 /// small nucleolar RNA, H/ACA box 64
8099850	**TMEM156**	-5.20	-4.15	-5.28	transmembrane protein 156
7937802	**CD81**	-5.19	-3.97	-4.71	CD81 molecule
7911870	**KIAA0562**	-5.18	-3.98	-5.10	KIAA0562
8074931	**ZNF70**	-5.17	-5.30	-5.68	zinc finger protein 70
8052834	**C2orf42**	-5.15	-4.10	-6.01	chromosome 2 open reading frame 42
7963646	**AAAS**	-5.12	-4.85	-4.48	achalasia, adrenocortical insufficiency, alacrimia
8053718	**IGKV2-40**	-5.06	-16.12	-7.35	immunoglobulin kappa variable 2-40
8046408	**PDK1**	-5.05	-4.10	-4.57	pyruvate dehydrogenase kinase, isozyme 1
8004081	**ZFP3**	-5.01	-4.06	-5.83	zinc finger protein 3 homolog (mouse)
7993167	**ATF7IP2**	-5.00	-4.29	-5.54	activating transcription factor 7 interacting protein 2
7914809	**KIAA0319L**	-4.97	-4.07	-5.30	KIAA0319-like
8073088	**APOBEC3G**	-4.87	-4.79	-3.94	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G
8022817	**KLHL14**	-4.85	-4.98	-4.60	kelch-like 14 (Drosophila)
8001746	**SNORA46**	-4.79	-4.80	-4.82	small nucleolar RNA, H/ACA box 46
8098556	**MLF1IP**	-4.75	-3.96	-5.96	MLF1 interacting protein
8073072	**APOBEC3D**	-4.75	-4.75	-4.70	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3D
8166096	**TCEANC**	-4.66	-5.04	-6.21	transcription elongation factor A (SII) N-terminal and central domain containing
7966397	**ATXN2**	-4.66	-4.31	-5.90	ataxin 2
8050875	**ASXL2**	-4.65	-5.58	-8.73	additional sex combs like 2 (Drosophila)
8098604	**ANKRD37**	-4.64	-5.56	-4.67	ankyrin repeat domain 37
8076690	**C22orf9**	-4.63	-5.53	-4.08	chromosome 22 open reading frame 9
8019563	**MGC70870**	-4.62	-4.37	-6.17	C-terminal binding protein 2 pseudogene
7900833	**KDM4A**	-4.59	-4.30	-4.61	lysine (K)-specific demethylase 4A
8012423	**C17orf68**	-4.54	-4.89	-5.72	chromosome 17 open reading frame 68
8018114	**SDK2**	-4.54	-4.45	-4.72	sidekick homolog 2 (chicken)
7994074	**SCNN1B**	-4.53	-3.95	-3.77	sodium channel, nonvoltage-gated 1, beta
7956088	**RAB5B**	-4.51	-5.18	-4.33	RAB5B, member RAS oncogene family
8005557	**EPN2**	-4.50	-4.09	-4.24	epsin 2
8153043	**ZFAT**	-4.47	-5.36	-4.67	zinc finger and AT hook domain containing
8065992	**NFS1**	-4.47	-4.18	-4.01	NFS1 nitrogen fixation 1 homolog (S. cerevisiae)
7930956	**SEC23IP**	-4.46	-4.76	-6.18	SEC23 interacting protein
8039719	**ZNF606**	-4.45	-4.35	-4.69	zinc finger protein 606 /// hypothetical LOC100128398
8166289	**CDKL5**	-4.16	-4.51	-4.72	cyclin-dependent kinase-like 5
7930614	**NHLRC2**	-4.05	-4.29	-5.80	NHL repeat containing 2
7899957	**ZMYM4**	-4.04	-4.20	-5.61	zinc finger, MYM-type 4
8086125	**TRANK1**	-4.02	-4.81	-5.00	tetratricopeptide repeat and ankyrin repeat containing 1
8050302	**ROCK2**	-4.00	-4.70	-6.42	Rho-associated, coiled-coil containing protein kinase 2
8056909	**ATF2**	-3.96	-4.94	-5.54	activating transcription factor 2
8036389	**ZNF585B**	-3.94	-4.21	-8.25	zinc finger protein 585B
7936856	**CHST15**	-3.90	-4.93	-5.13	carbohydrate (N-acetylgalactosamine 4-sulfate 6-O)
8073068	**APOBEC3C**	-3.90	-4.19	-4.12	apolipoprotein B mRNA editing enzyme
8051012	**PREB**	-3.90	-4.97	-4.20	prolactin regulatory element binding
8171896	**CXorf21**	-3.87	-4.01	-5.57	chromosome X open reading frame 21
7971422	**ZC3H13**	-3.84	-3.91	-5.70	zinc finger CCCH-type containing 13

7943162	**C11orf54**	-3.84	-4.22	-5.80	chromosome 11 open reading frame 54
8007071	**CDC6**	-3.84	-4.44	-4.18	cell division cycle 6 homolog (S. cerevisiae)
8103911	**IRF2**	-3.83	-4.70	-6.54	interferon regulatory factor 2
7930213	**TAF5**	-3.81	-4.77	-5.40	TAF5 RNA polymerase II,
7918487	**DENND2D**	-3.76	-4.34	-6.51	DENN/MADD domain containing 2D
8122807	**AKAP12**	-3.75	-4.13	-5.27	A kinase (PRKA) anchor protein 12
8095163	**EXOC1**	-3.69	-4.19	-6.29	exocyst complex component 1

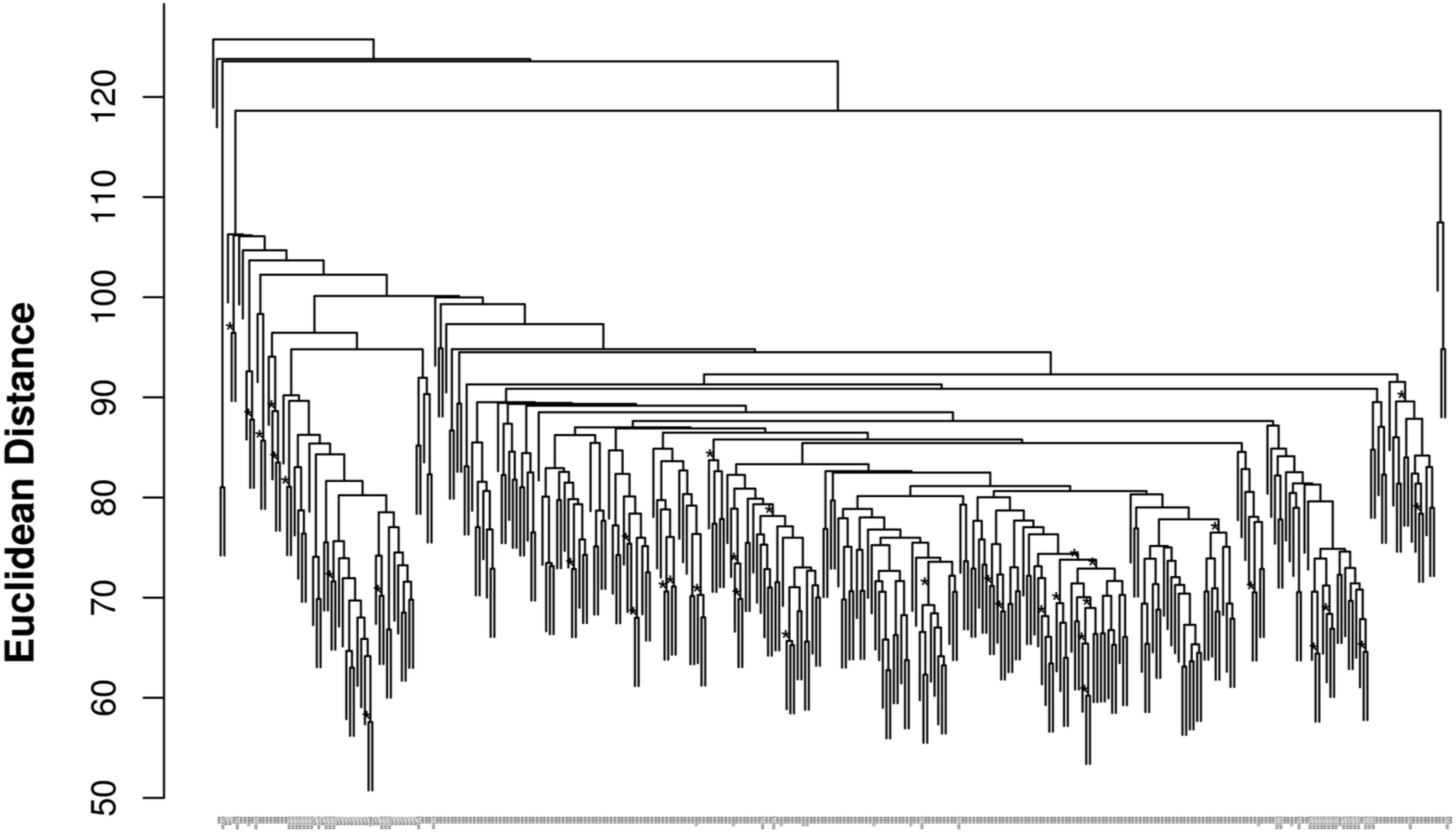
*Immunoglobulin genes physiologically upregulated in NPC were eliminated

Supplementary Figure 1:

1A) Mayo Clinic dataset: Semi-supervised analysis of the 134 samples from Mayo Clinic dataset based on the 5,596 significant probes from SAM multiclass analysis of the present study dataset **p*<0.05.

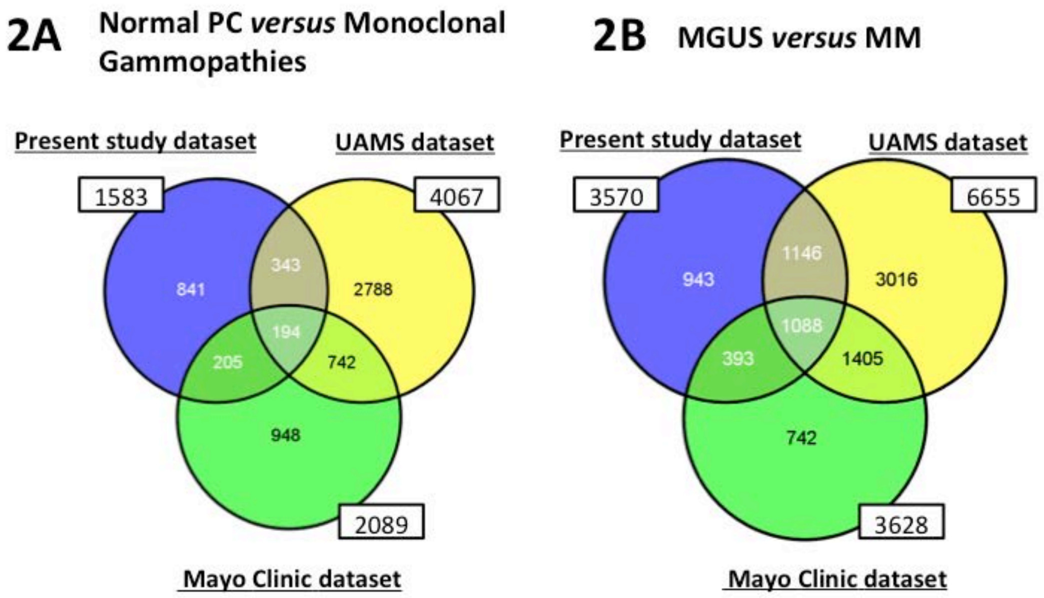


1B) UAMS dataset: Semi-supervised analysis of the 334 samples from UAMS dataset based on the 5,627significant probes from SAM multiclass analysis of the present study dataset *p<0.05.



Supplementary figure 2: Comparison of expression signatures to independent monoclonal gammopathy datasets.

Venn diagram showing the number of genes commonly and specifically deregulated in the three independent datasets (Present study, Mayo Clinic and UAMS datasets) for the normal PC *vs* monoclonal gammopathy (clonal PC) comparison (Figure 2A) and for the MGUS *vs* MM comparison (Figure 2B)



Supplementary figure 3: Expression levels of 10 genes by quantitative real-time PCR in MGUS (n = 14), SMM (n = 10) and MM (n = 10) patients. Relative values were calculated by the $2^{-\Delta Ct}$ method ($\Delta Ct = Ct_{(gene)} - Ct_{(ABL)}$). The *ABL* gene was used as control gene.