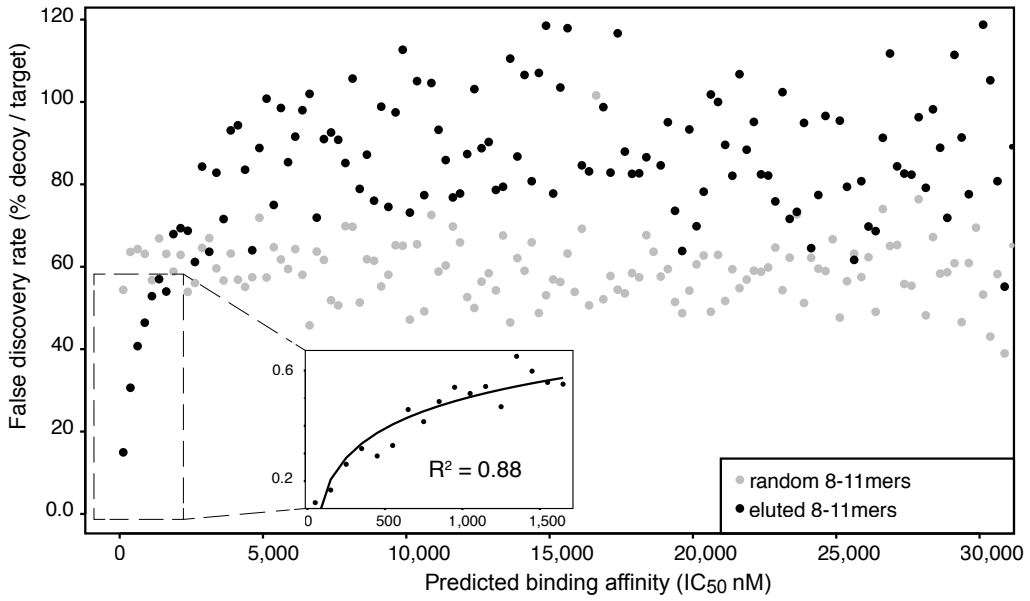
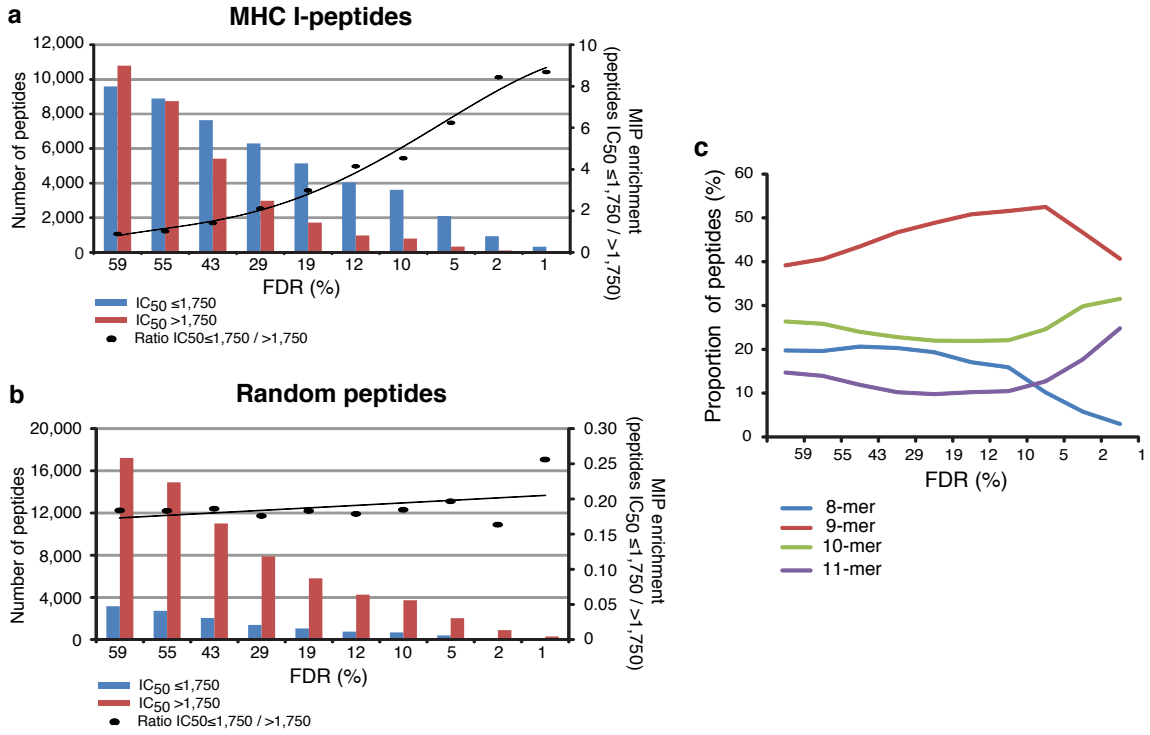
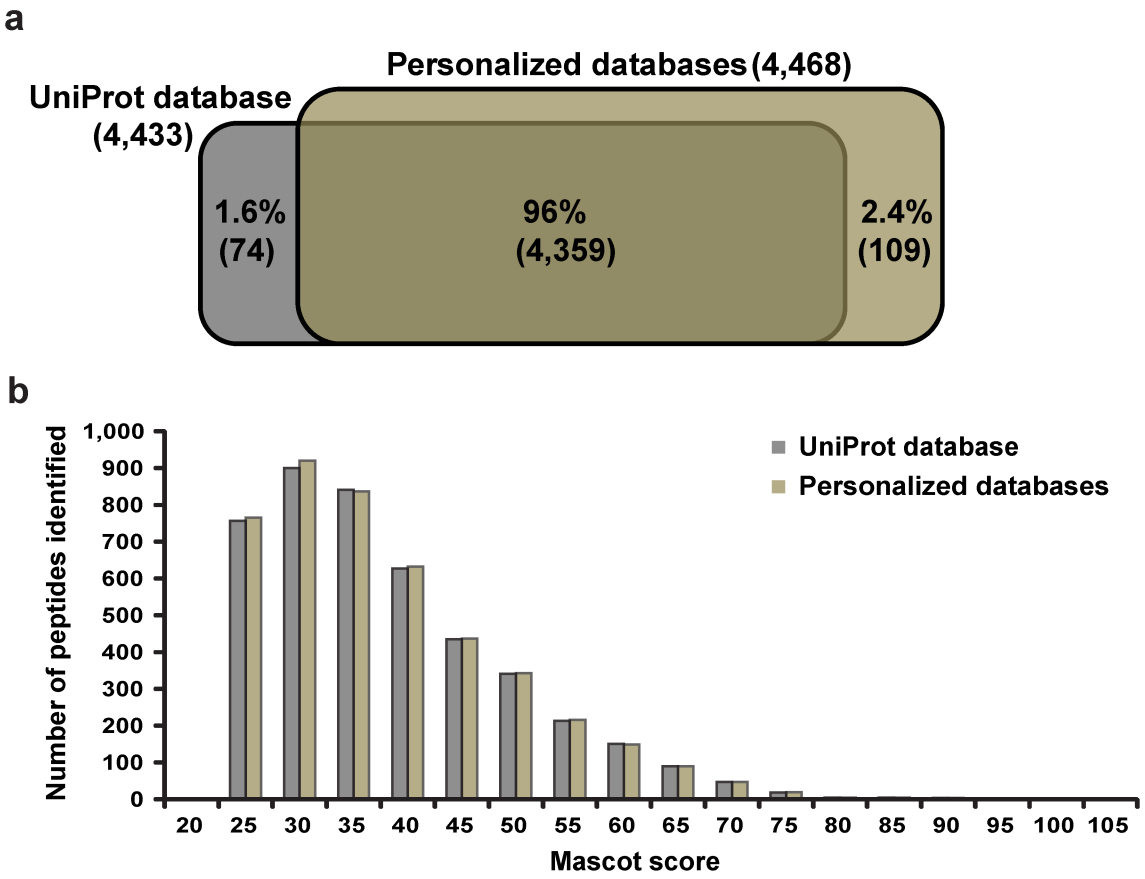


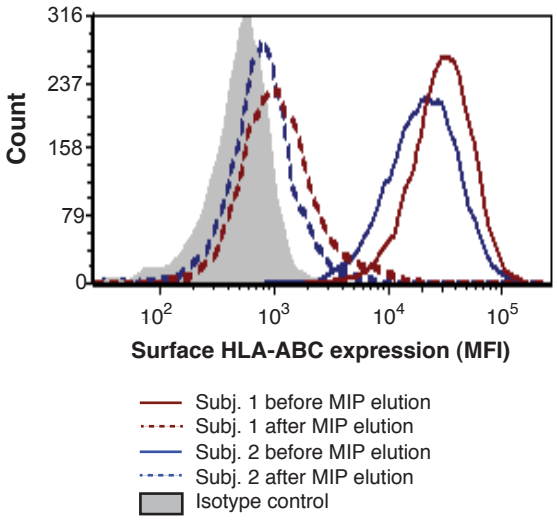
Supplementary Figure 1. Global false discovery rate (FDR) and predicted binding affinity allow discrimination between MIPs and contaminant (non-MIP) peptides. Predicted binding affinity to the relevant HLA molecules of 20,380 eluted peptides (black) and 20,380 random peptides generated from the personalized protein databases of subjects 1 and 2 (grey) was calculated using NetMHCcons. Binding score categories were generated by intervals of 500 nM (IC_{50}). Each dot represents the mean predicted binding affinity for peptides in a given bin. For each category, the level of accuracy in the peptide identification is shown as the FDR. The FDR (database of 20,319 target/12,011 decoy peptides identified by Mascot and defined by the clustering) was calculated for the eluted and the random peptides. The inset shows a high correlation between FDR values < 60% and PBA values < 1,750 nM. The distribution of random peptides shows no correlation between the predicted binding affinity and the FDR.



Supplementary Figure 2. The global false discovery rate (FDR) allows enrichment of MIPs and affects the proportion of small (8-9mers) and long peptides (10-11mers) identified. (**a**) We calculated the predicted binding affinity (IC50) for the 8-11mer peptides obtained after applying different FDR thresholds. FDR values were calculated from a dataset of 20,319 eluted peptides (target) and 12,011 reverse peptide versions (decoy). Without filtering any 8-11mer peptide identified by Mascot, the FDR value corresponds to 59% (maximal FDR). Bars show the number of peptides with a predicted binding affinity ≤ 1,750 nM (blue) or > 1,750 nM (red). The second y-axis shows the enrichment in MHC I-peptides calculated as the ratio of peptides with a predicted binding affinity ≤1,750 / >1,750 nM. (**b**) We performed the same analysis by randomly generating the same amount of peptides for each FDR threshold. The figure shows that lower FDR thresholds increase the proportion of eluted peptides with high predicted binding affinity (a) but have no impact on random peptides (b). (**c**) Proportion of 8 – 11mers identified by applying different FDR thresholds. Low FDR values favor the identification of long peptides and disfavor the identification of short peptides.

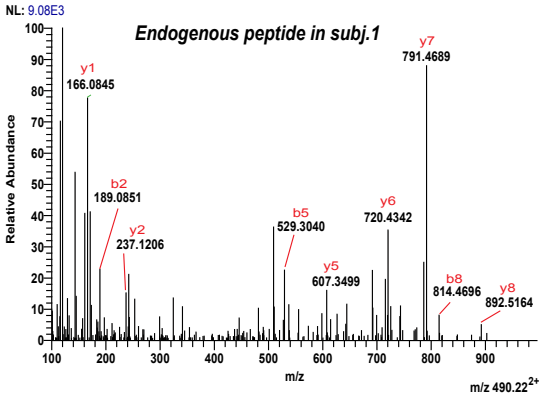


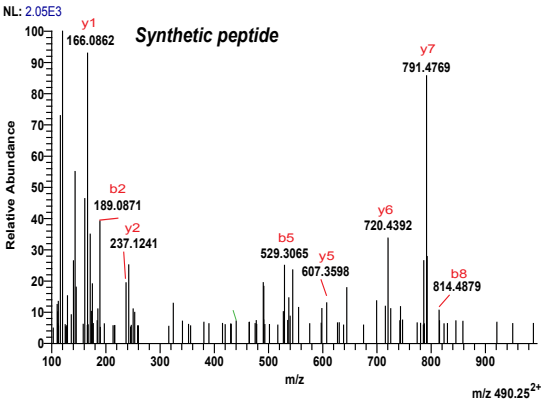
Supplementary Figure 3. Comparison of MIPs identified using UniProt vs. personalized databases built with next generation sequencing data. (**a**) Venn diagram comparing the number and percentage of unique and common MIP sequences found using UniProt vs. personalized databases. (**b**) Mascot score of MIPs identified with each database.



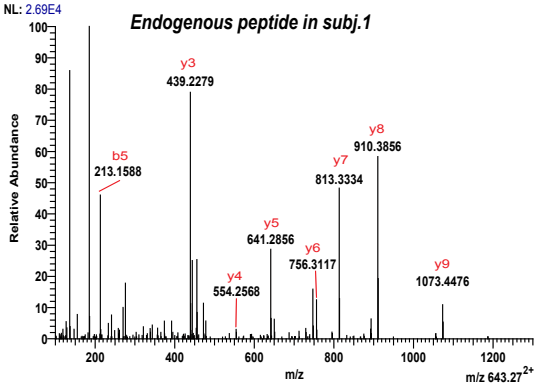
Supplementary Figure 4. Quantification of surface HLA-ABC before and after peptide elution. B-LCLs from subjects 1 and 2 were stained with PE anti human HLA-ABC monoclonal antibody (clone W6/32, Cedarlane) or the corresponding isotypic control and the mean fluorescence intensity (MFI) was analyzed by flow cytometry before or after mild acid elution of peptides. The histogram shows similar levels of HLA-ABC surface expression before and after peptide elution in representative samples of subjects 1 and 2.

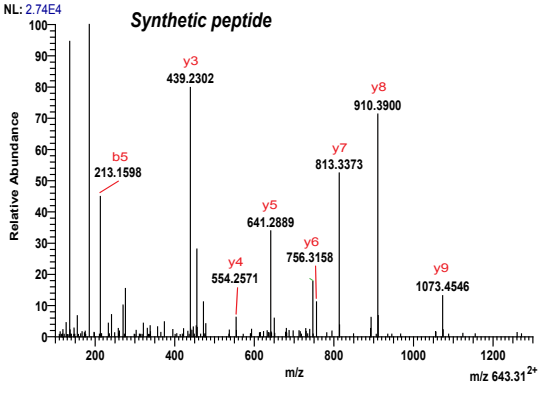
STALRLTAF_*ITGAL*



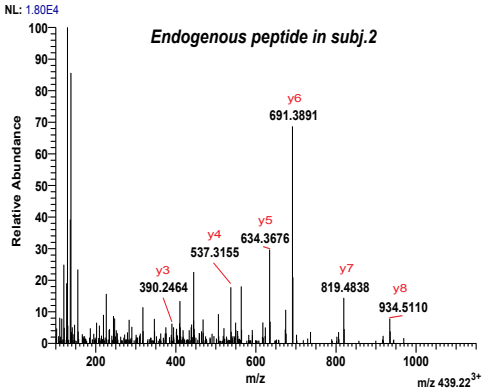


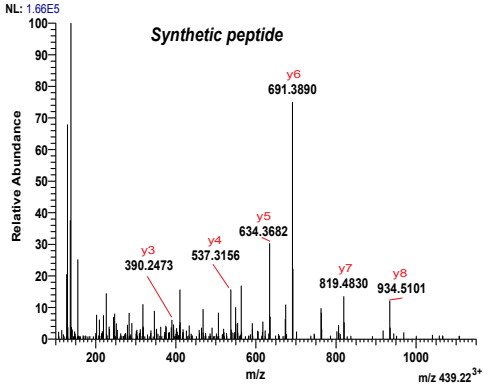
VIYPGDSDTRY_*IGHV5*



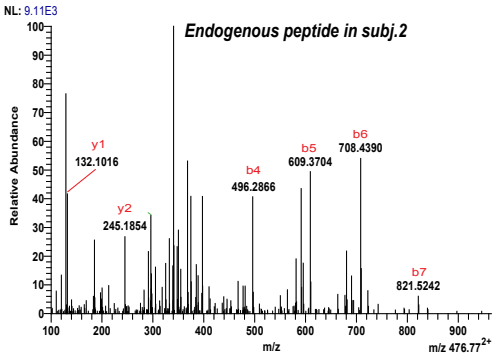


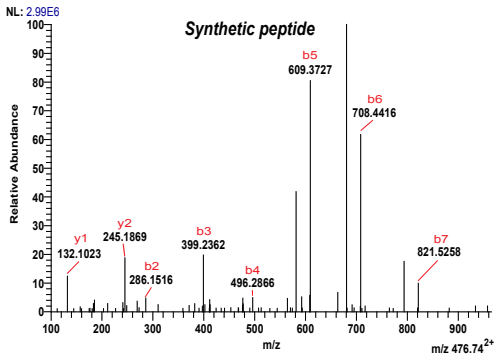
AMYDKGPFRSK_*NQ01*



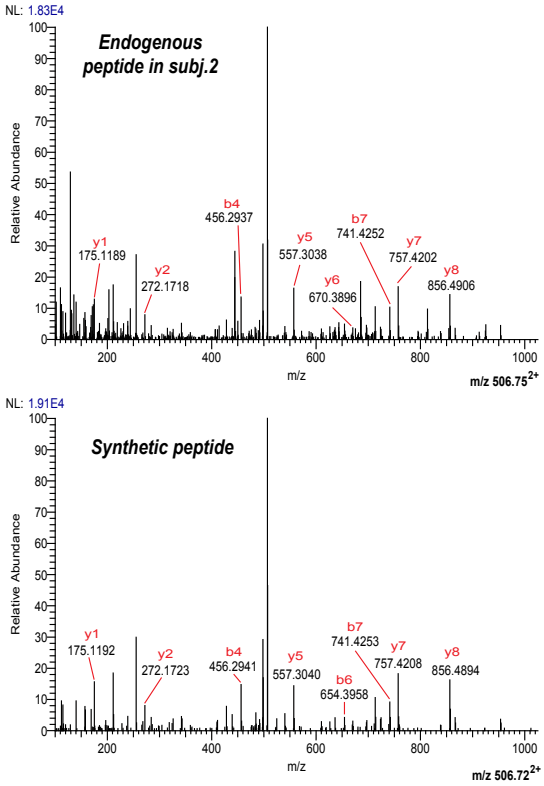


RELPLVLL_*GRP*

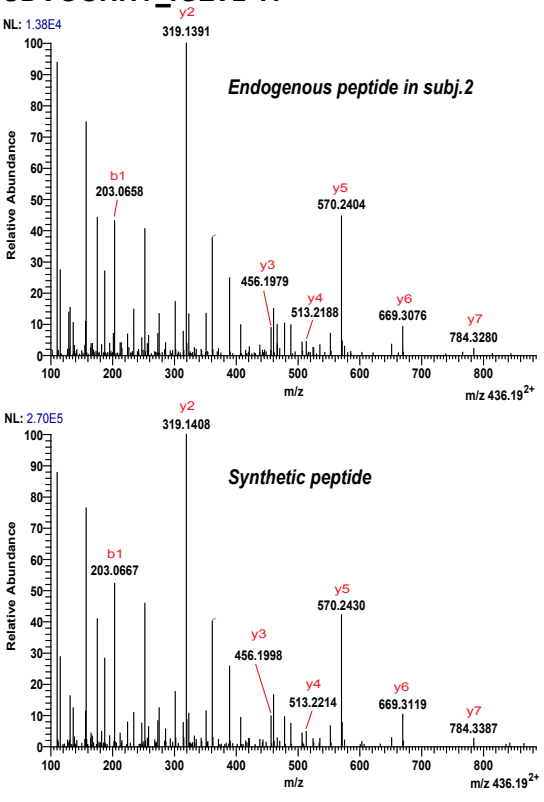




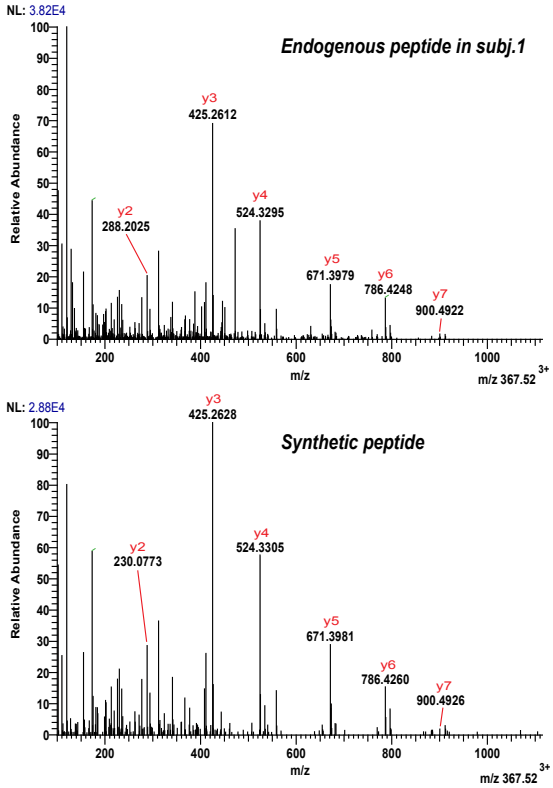
RVSLPTSPR_*C13Orf18*



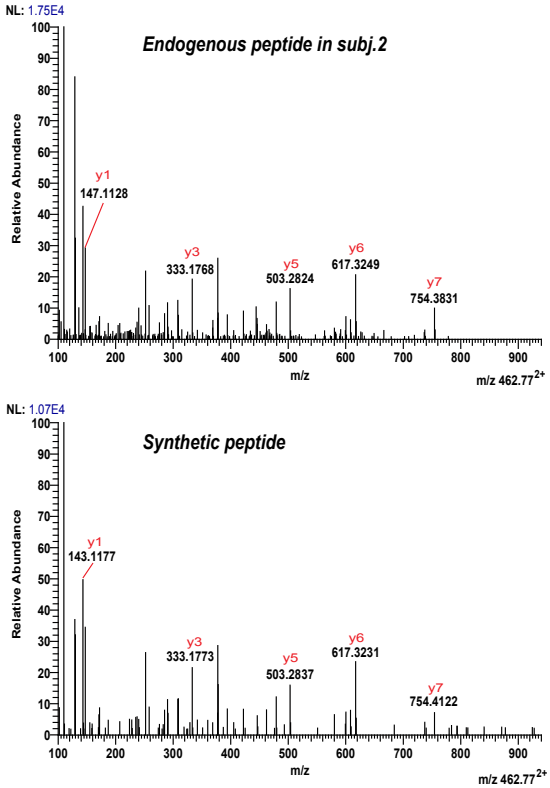
SDVGGHHY_*IGLV2-11*



AENDFVHRI_*R3HCC1*

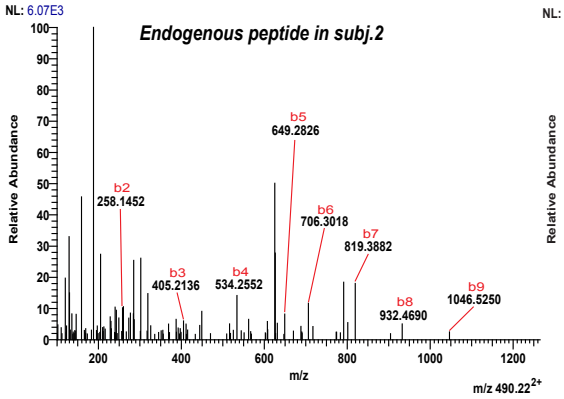


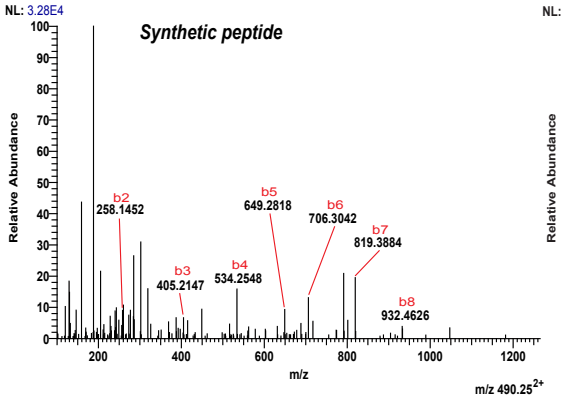
AVHNGLGEK_*NADK*



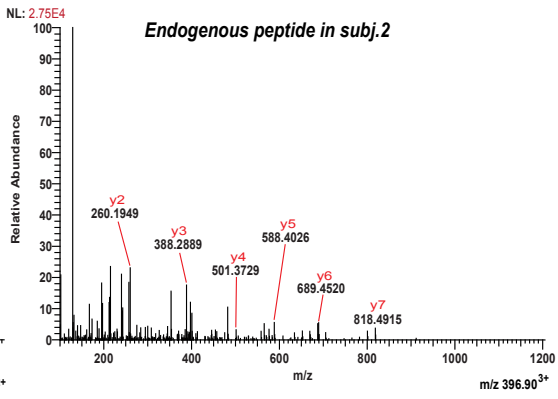
5a

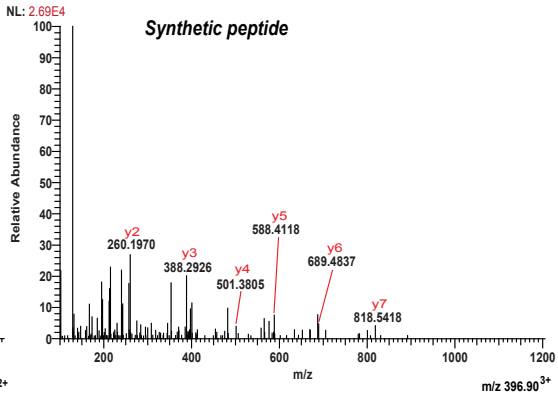
KEFEDGIINW_*BCL2A1*





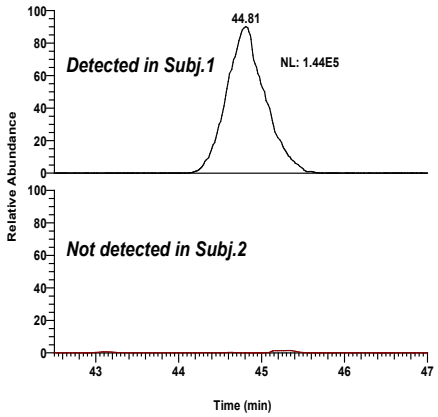
QELETSIKKI_*KIF20B*

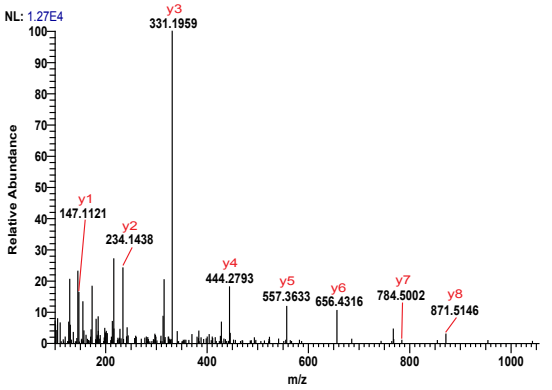




5a

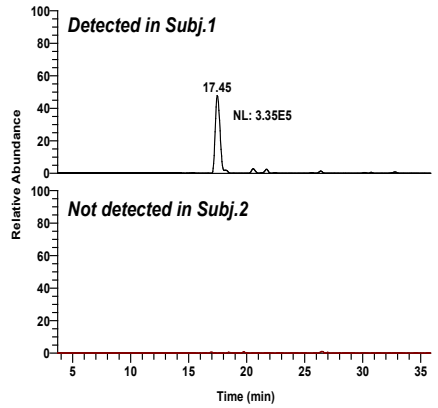
ATSQVLLPSK_*IGHM*

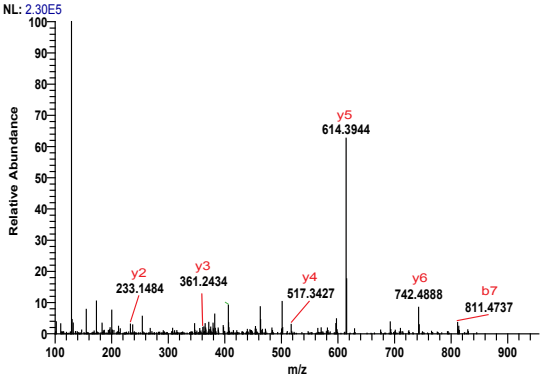




Fragment mass accurracy < 0.02Da

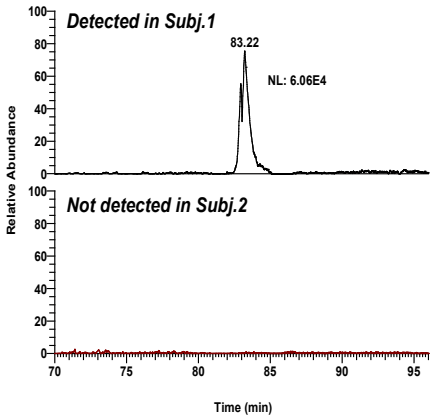
EAKPRKTL_*FAM21C*

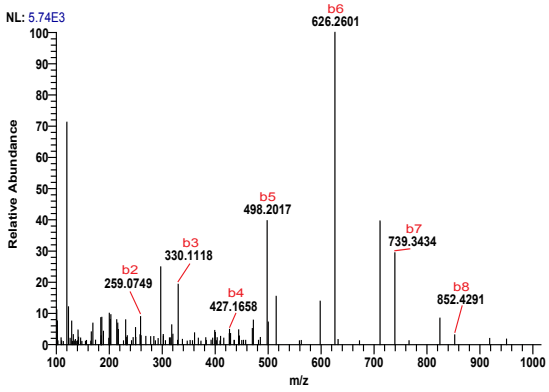




Fragment mass accurracy < 0.02Da

EEAPAQLLQ_*LRRC37B*

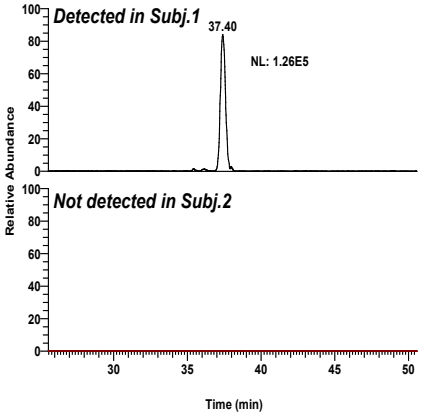


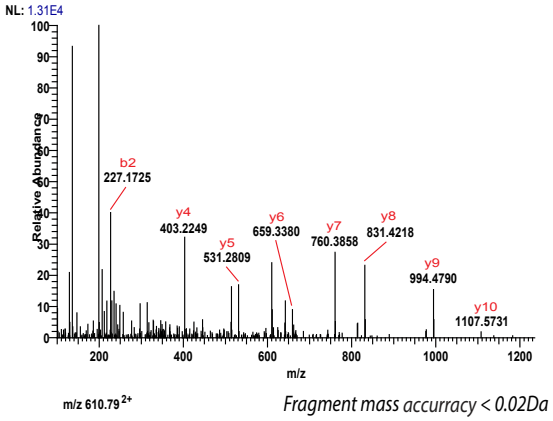


Fragment mass accurracy < 0.02Da

5b

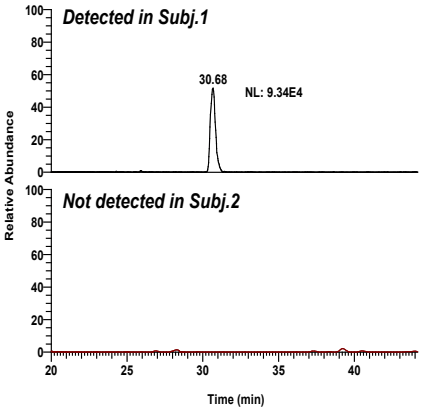
LLYATQQGQAK_*MTRR*

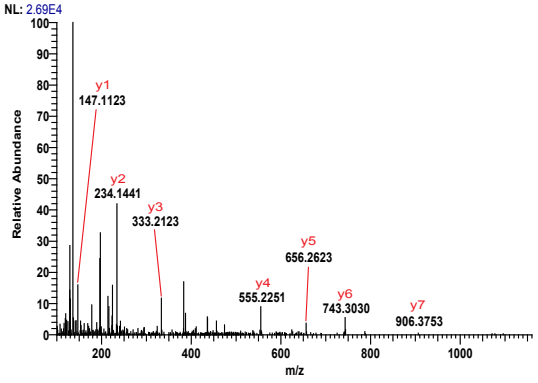




Fragment mass accurracy < 0.02Da

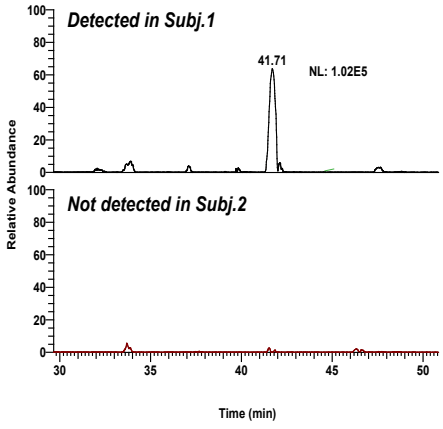
QIYSTC(cyst)VSK_*KIF21B*

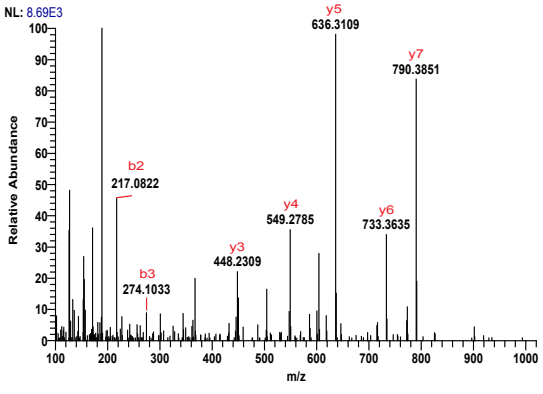




Fragment mass accurracy < 0.02Da

SEGPSTSRW_*FOXM1*

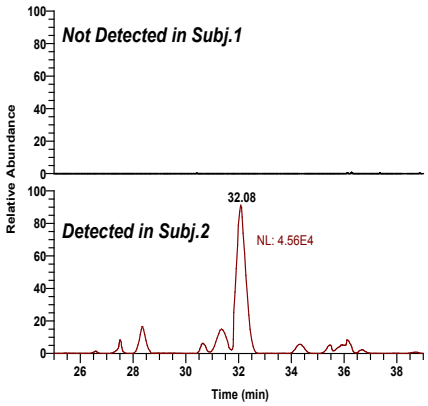


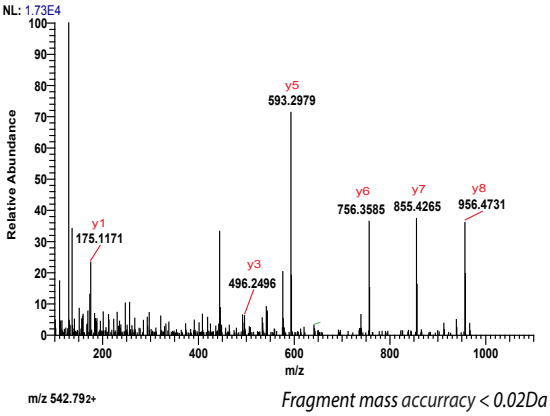


Fragment mass accurracy < 0.02Da

5b

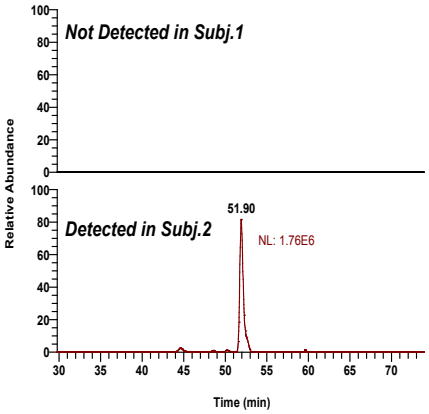
KTVYPASYR_*CDT1*

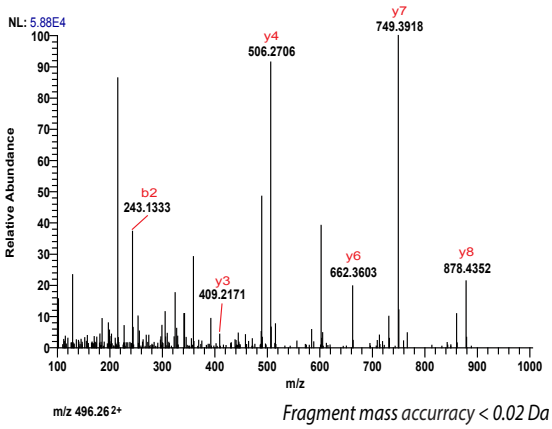




Fragment mass accurracy < 0.02Da

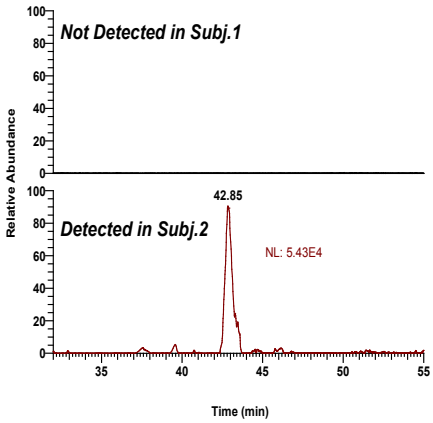
LESGVPSRF_*IGKV1D*

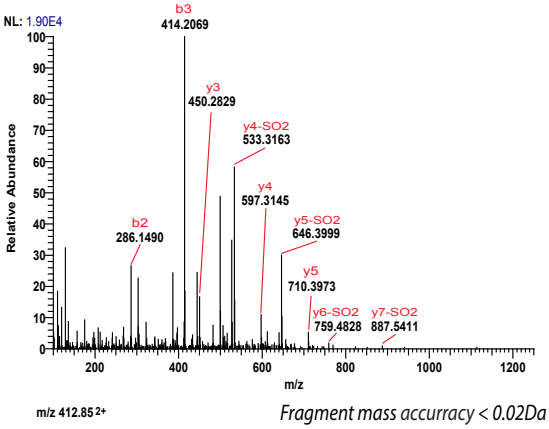




Fragment mass accurracy < 0.02 Da

REQILM(ox)KRF_*PWP2*

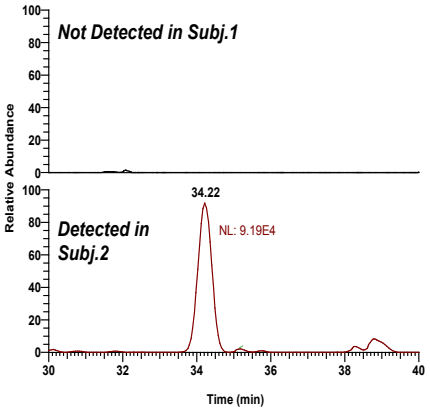


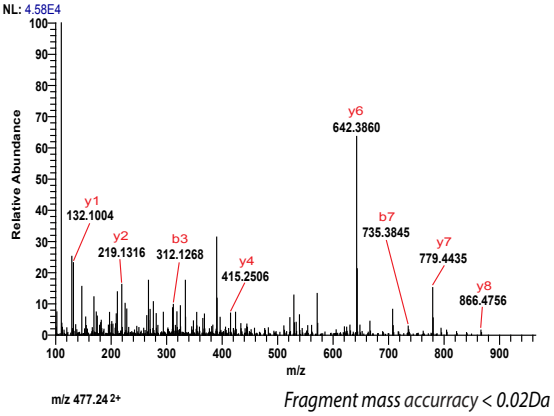


Fragment mass accurracy < 0.02Da

5b

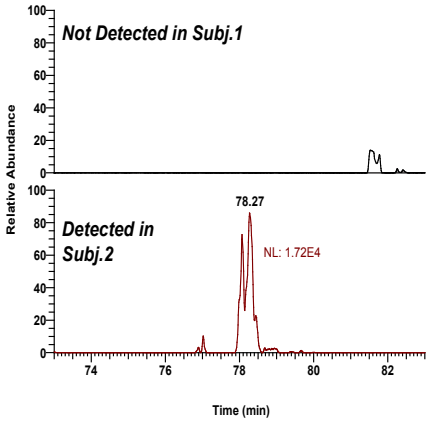
SSHQVVPSL_*PTPRZ1*

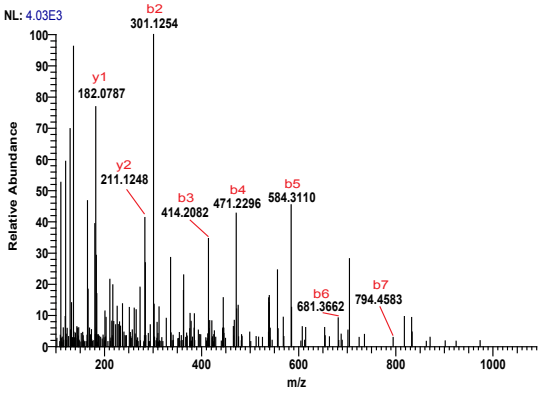




Fragment mass accurracy < 0.02Da

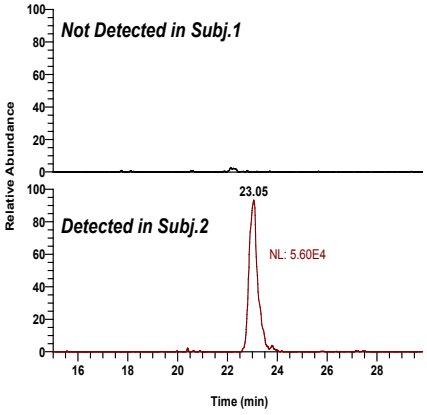
YHIGIPLTY_*VPS16*

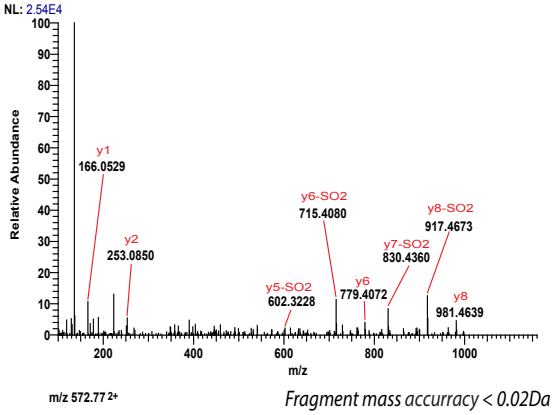




Fragment mass accurracy < 0.02Da

YSDLRKESM(ox)_*MCM2*

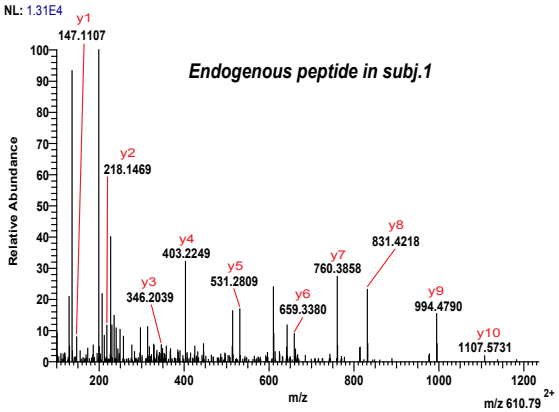


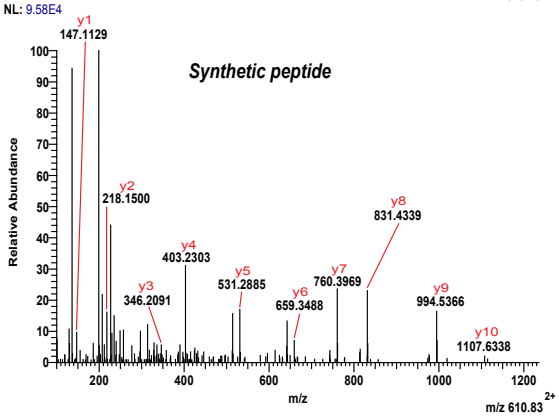


Fragment mass accurracy < 0.02Da

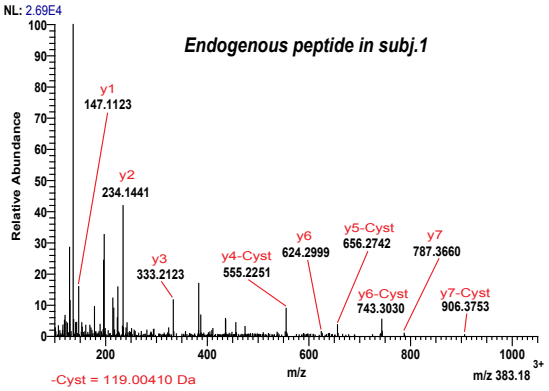
5b

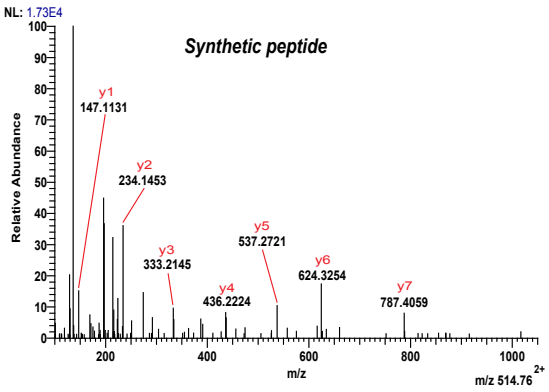
LLYATQQGQAK_MTRR



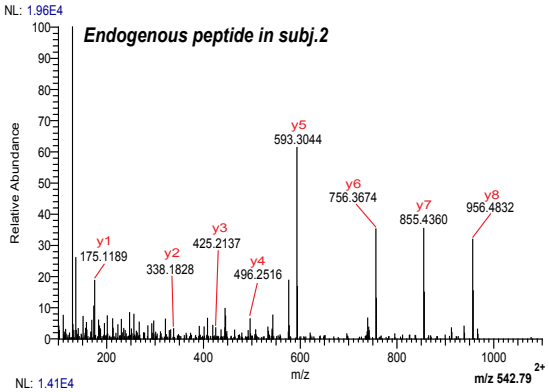


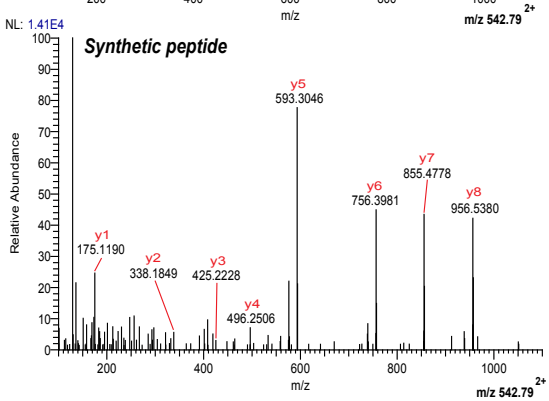
QIYSTC(cyst)VSK_KIF21B





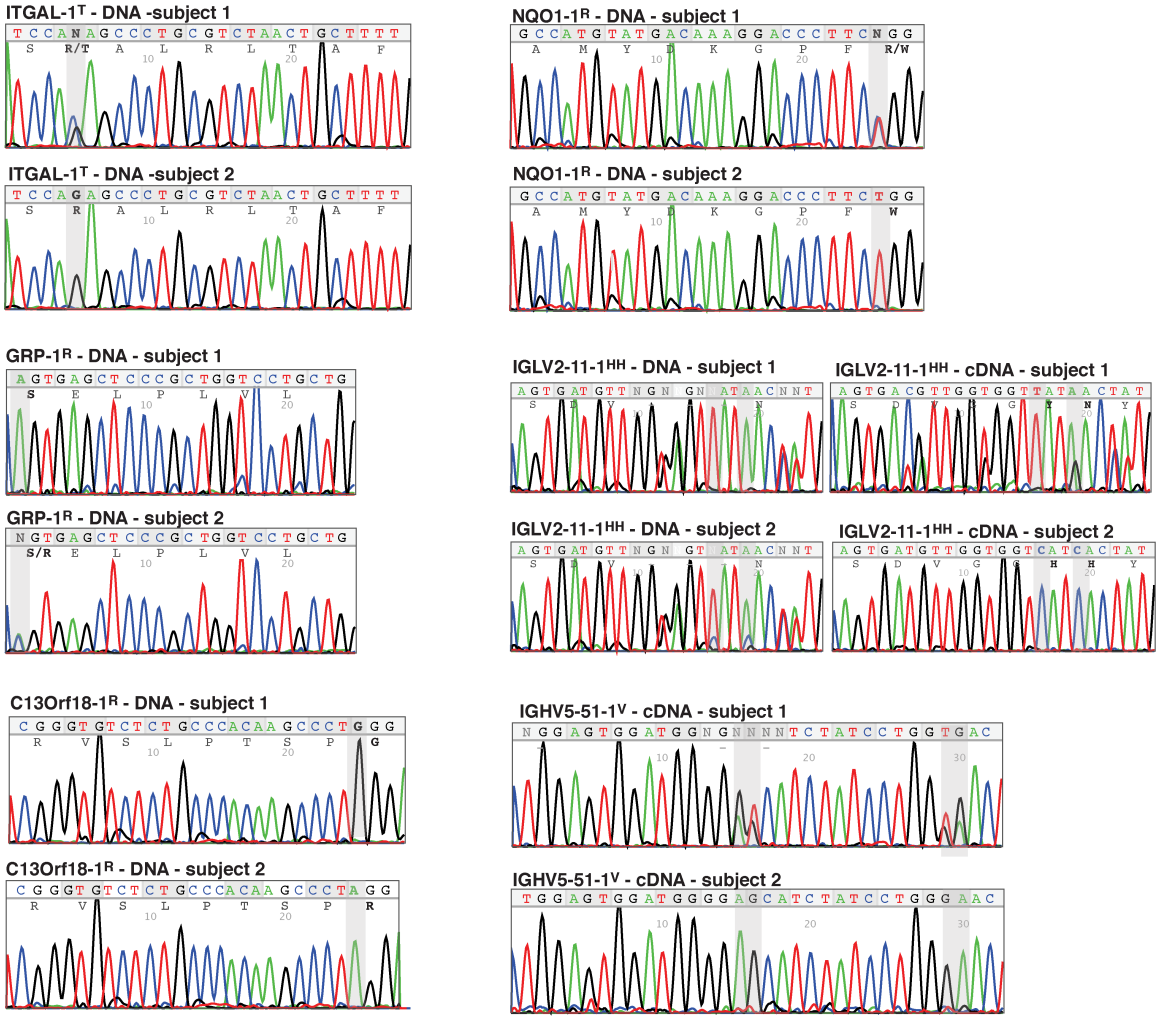
KTVYPASYR_*CDT1*



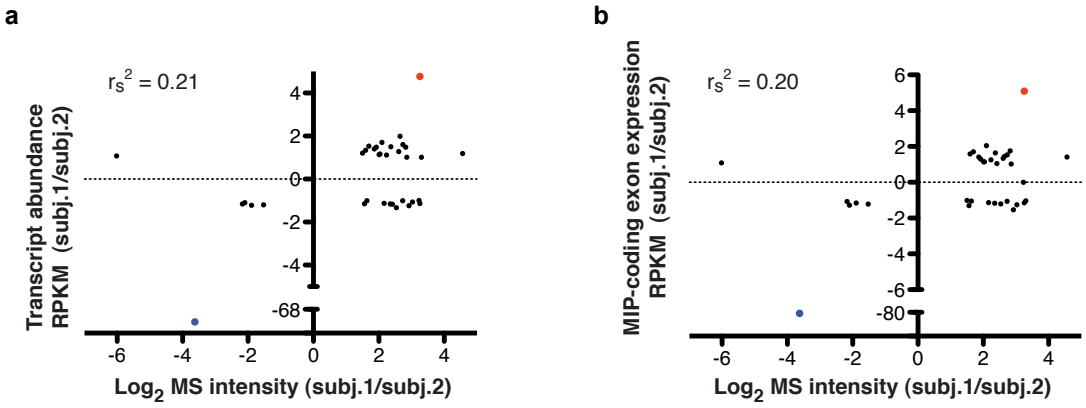


5c

Supplementary Figure 5. MS validation of polymorphic and non polymorphic peptides. (**a**) Polymorphic MiHAs resulting from ns-SNPs in the MIP-coding region, encoded by one single non-HLA gene and exclusively detected in one of the two subjects (corresponding to Table 1) were validated using the MS/MS spectra of the corresponding synthetic peptides. (**b**) Twelve peptides exclusively detected in subject 1 or 2 (6 in subject 1 and 6 in subject 2) were randomly selected among those that were present in 3 or 4 replicates. The extracted ion chromatograms of MiHA peptides for both subjects are presented along with the corresponding MS/MS spectra. Comparison of extracted ion chromatograms conrmed the selective detection of MiHA peptides in one of the two subjects. (**c**) Three non polymorphic peptides (Supplementary Data 3) used in cytotoxicity assays were also validated using MS/MS of their respective synthetic peptides.



Supplementary Figure 6. Validation of 6 MiHA-coding sequences by Sanger sequencing. Chromatograms obtained after Sanger sequencing of PCR-amplified DNA or cDNA encoding 6 MiHAs. The primers used are shown in Supplementary Table 1. Polymorphic loci are highlighted in grey. The IGLV2-11-1HH MiHA results from 2 nucleotide changes at the transcript level in subject 2.



Supplementary Figure 7. Differential expression of non-polymorphic MIPs does not correlate with differences in MIP-coding genes or exons between subjects. For 41 non-polymorphic MIPs that were exclusively detected in one sibling, we calculated the fold difference in intensity of the MIP and the fold difference in expression of the underlying MIP-coding gene (**a**) or exon (**b**) measured in Reads Per Kilobase per Million mapped Reads (RPKM). In only 2 cases, MIP abundance differences reflect MIP-coding transcript differences (dots in red and blue). The calculated Spearman r value (rs^2) shows no correlation.

Supplementary Table 1. List of PCR primers used for PCR amplification of DNA and cDNA segments containing the region coding for the MiHAs between subjects 1 and 2.

Peptide sequence (amino acids)	Gene symbol	Target nucleic acid	Forward (5' to 3')	Reverse (5' to 3')
STALRLTAF	ITGAL	DNA	TTATTTCTTTCTGGCCCACCA	AGCATCTTCTTCCAAGTTACTCAG
STALRLTAF	ITGAL	cDNA	GAAACCTGGGAGATCCCTTT	TCACATTGGCGTGCAATTC
VIYPGDSDTRY	IGHV5-51	DNA and cDNA	ACAGTAATACATGGCGGTGTC	GCTGTTCTCCAAGGTCAGTC
AMYDKGPFRSK	NQO1	DNA	AGGAATGGGAAAGGTGTGAAG	GGGAAGCTCCATCTCAAACAA
AMYDKGPFRSK	NQO1	cDNA	GCCCAGATATTGTGGCTGAA	GAAGCCACAGAAATGCAGAATG
RELPLVLL	GRP	DNA	TCTGCTCTTCCCAGCCTCTC	GCCAGGGAAACGCAAAGAAATG
RVSLPTSPR	C13orf18	DNA	TCTGAGGATACCACAGACTCC	GGTGTGAACAGAGAGGAATGAG
RVSLPTSPR	C13orf18	cDNA	CAACGTTGTCTGAGGATACCAC	GCACAAATACCTCTGGTGAGAA
SDVGGHHY	IGLV2-11	DNA and cDNA	GAGTGTGTTTCTCCCTCTTTCC	CCTGCATATGAGCAGCAGTAA
SDVGGHHY	IGLV2-11	DNA and cDNA	TGATCCTTGGTCTCCTGCT	GAAAGTGTAGCTGCCTGCATA