Supplemental Data for

Characterization of Glycoproteoforms of Integrins α2 and β1 in Megakaryocytes in the Occurrence of JAK2V617F Mutation-Induced Primary Myelofibrosis

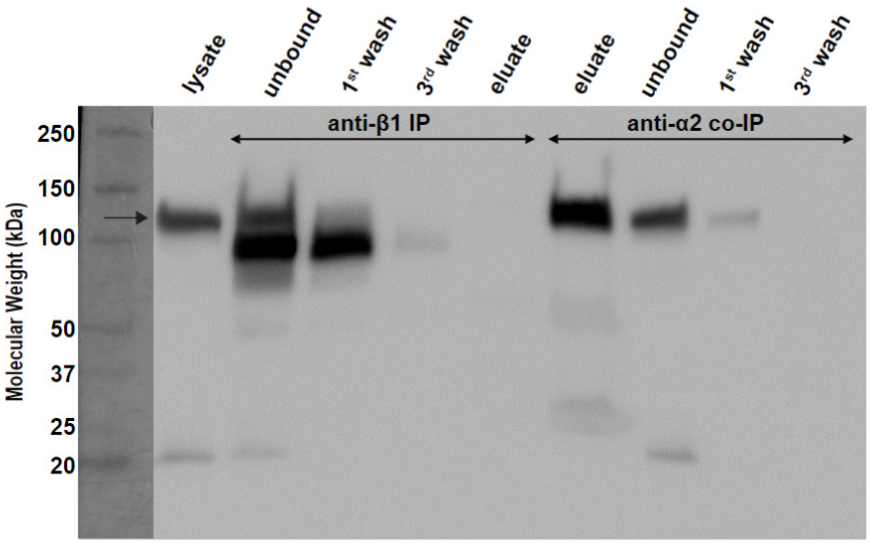
Maissa M. Gaye[1,2], Christina M. Ward[2], Andrew J. Piasecki[2], Vanessa L. Stahl[1], Aikaterini Karagianni[2,3], Catherine E. Costello[1*], Katya Ravid[2*]

[1]Department of Biochemistry, Center for Biomedical Mass Spectrometry, Boston University School of Medicine, Boston, MA, USA
[2]Department of Medicine and Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, MA, USA
[3]Department of Internal Medicine, School of Medicine, University of Crete, Heraklion, Greece

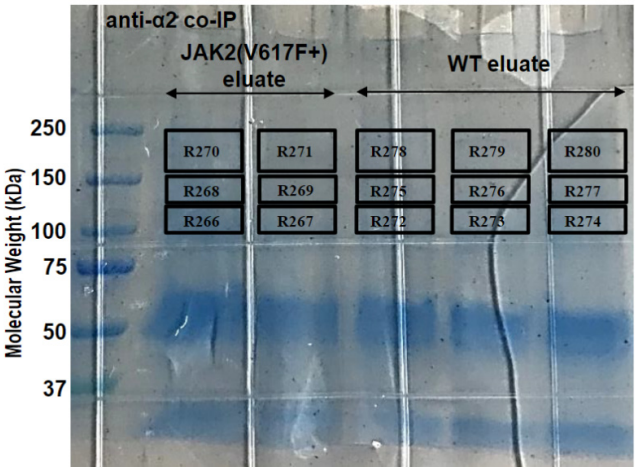
***Corresponding author information**
Catherine E. Costello
cecmsms@bu.edu
1 - (617) 358-2433

Katya Ravid
kravid@bu.edu
1 - (617) 358-8242

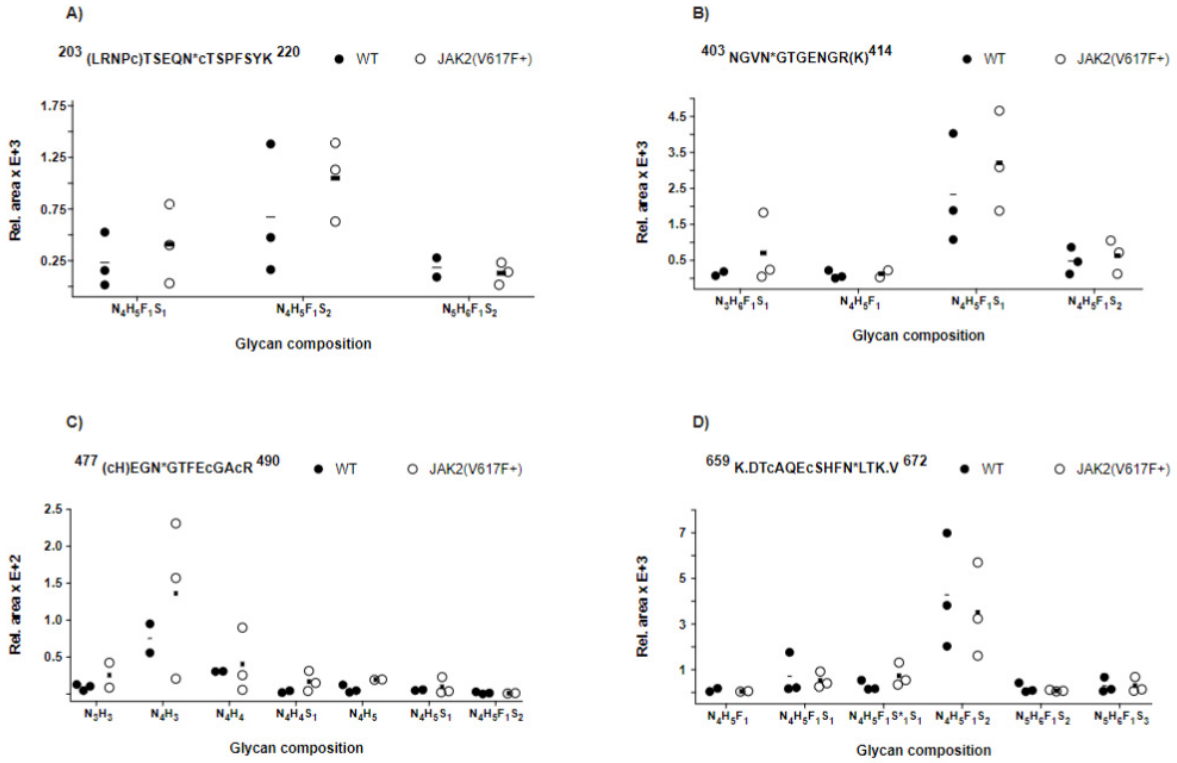
Table of Contents



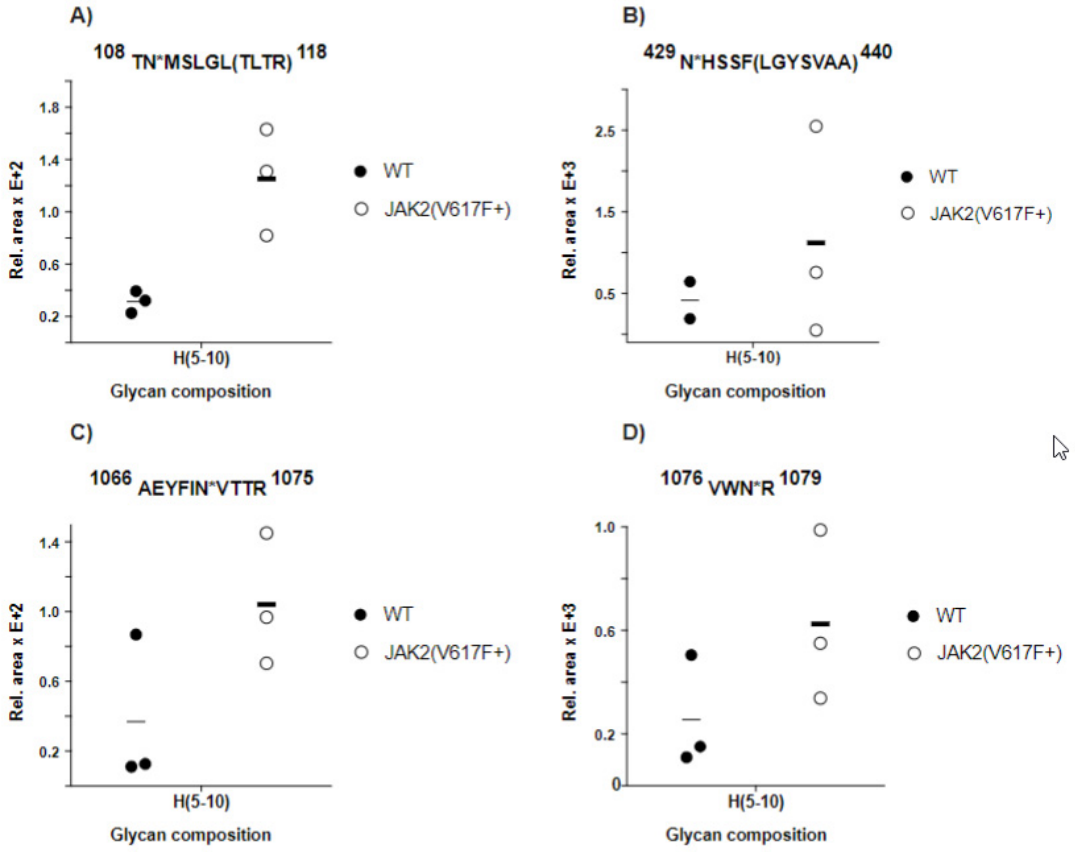
Supplemental Figure 1. Western blot analysis followed by chemiluminescent detection of integrin β1 using Immobilon Western Chemiluminescent HRP Substrate (Millipore Signa, St. Louis, MO) and ImageQuant LAS 4000 camera system (GE Healthcare, Chicago, IL). The digital image was acquired in auto-exposure mode. Purified mouse anti-CD29 Clone 18/CD29 integrin β1 (BD Biosciences, San Jose, CA) was used as the primary antibody (1:2000) and anti-mouse IgG HRP-linked (Cell Signaling Technology, Danvers, MA 01923) was used as the secondary antibody (1:4000). Integrin β1 was isolated from MKs of wild type male mice by immunoprecipitation (IP) using TS2/16 anti-β1 integrin conjugated to agarose beads or by co-immunoprecipitation using an anti-α2 integrin monoclonal antibody in solution. The lysate is represented, as well as post -IP and -co-IP unbound and wash fractions. A higher yield of integrin β1 was obtained by co-IP when the anti-α2 integrin mAb was being used and this immunoprecipitate was used for the nUPLC-MS/MS analyses. An arrow points to the molecular weight (approximately 130 kDa) of β1.



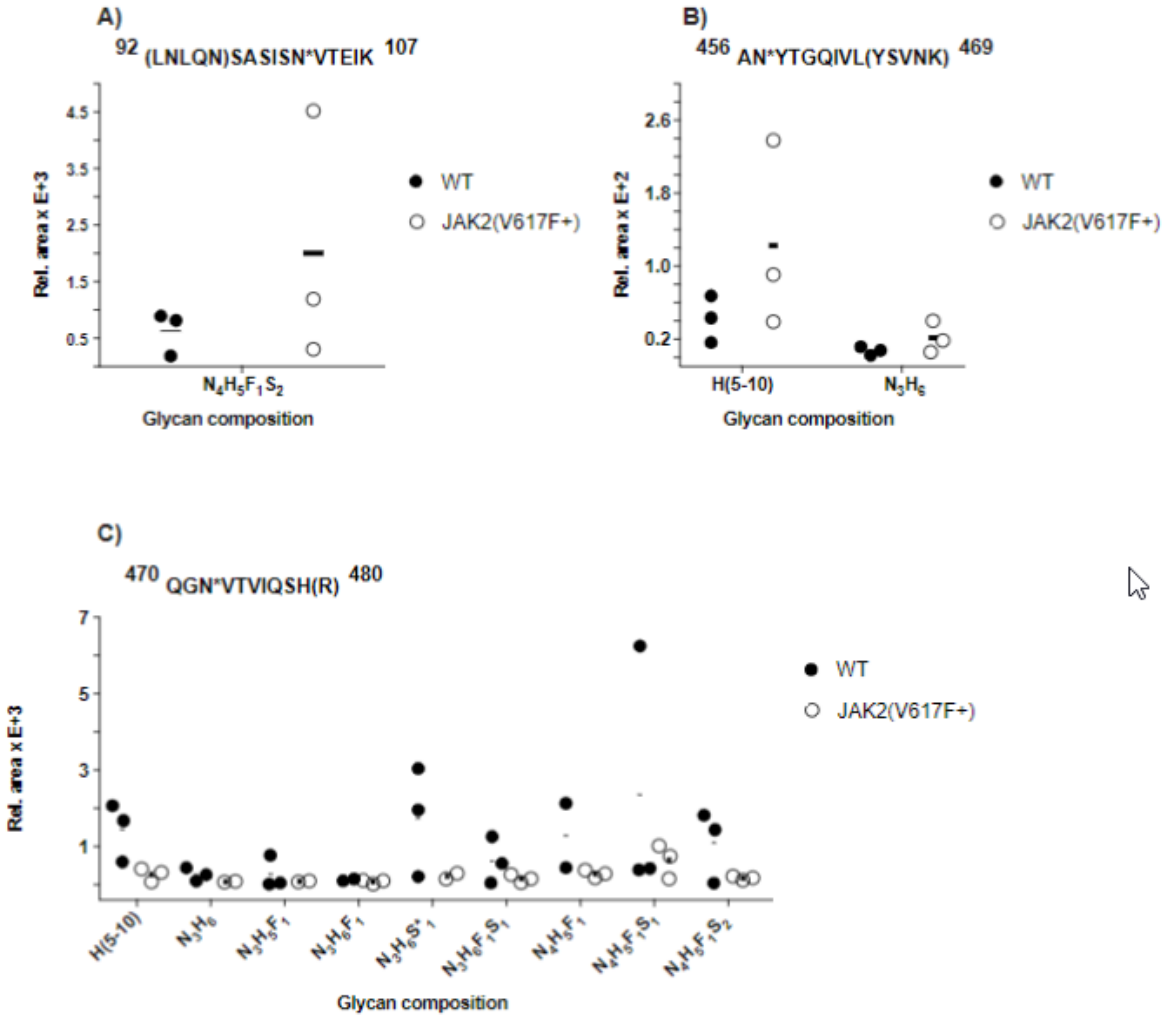
Supplemental Figure 2: SDS-PAGE gel stained with Coomassie blue showing integrin bands from proteins of wild type (WT) and JAK2(V617F+) male mice MKs isolated by co-IP using anti-α2 integrin monoclonal antibody in solution. Gel bands excised and individually submitted to proteolytic digestion followed by mass spectrometry analysis are indicated. For a given phenotype, peptides extracted from gel bands corresponding to approximately the same molecular weight regions were combined prior to further analysis, *e.g.,* R266 and R267 for WT or R272, R273 and R274 for JAK2(V617F+). Gel bands situated at ~50 kDa and < 37 kDa correspond to anti-α2 integrin monoclonal antibody heavy and light chains, respectively.



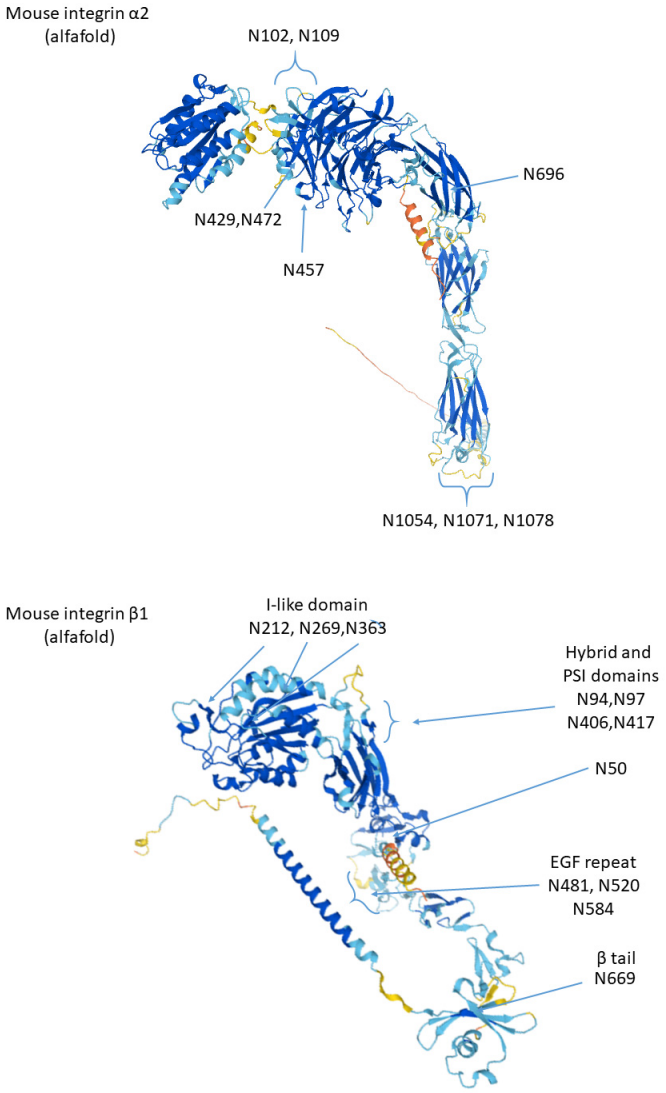
Supplemental Figure 3: *N*-glycan microheterogeneity at sites N212, N406, N481 and N669 of integrin β1 in megakaryocytes of wild type (WT) and JAK2(V617F+) 13- to 14-weeks-old male mice. The total signal corresponding to the sum of all glycoforms of the glycopeptides that span each *N*-linked site is normalized to the total ion current for each of the biological replicates. Each biological replicate is depicted as a single point, the mean is indicated by a dash (WT) or by a square [JAK2(V617F+)]. Sequence extensions for peptides resulting from alternative proteolytic cleavages detected around a given glycosite (N*) are indicated in parenthesis. A) Glycopeptide ^{203}(LRNPc)TSEQN*cTSPFSYK^{220} at N212, B) Glycopeptide ^{403}NGVN*GTGENGR(K)^{414} at N406, C) glycopeptide ^{477}(cH)EGN*GTFEcGAcR^{490} at N481 and D) glycopeptide ^{659}DTcAQEcSHFN*LTK^{672} at N669. Abbreviations used are as follows: HexNAc, N, *N*-acetylhexosamine; NeuGc, S, *N*-glycolyl neuraminic acid; NeuAc, S*, *N*-acetyl neuraminic acid; Hex, H, hexose; F, fucose (deoxyHex). Lower case c in a peptide sequence indicates a cysteine that has been modified by carboxymethylation prior to protease digestion



Supplemental Figure 4: Minimally processed (N2H5-10) glycopeptiforms at *N*-glycan sites N109, N429, N1071 and N1078 of integrin α2 in megakaryocytes of wild type (WT) and JAK2(V617F+) male mice. The total signal corresponding to the sum of all glycoforms of the glycopeptides that span each *N*-linked site is normalized to the total ion current for each of the biological replicates. Each biological replicate is depicted as a single point, the mean is indicated by a dash (WT) or by a square [JAK2(V617F+)]. Alternative proteolytic cleavages detected around a given glycosite (N*) are indicated in parenthesis. A) Glycopeptide ^{108}TN*MSLGL(TLTR)^{118} at N109, B) Glycopeptide ^{429}N*HSSF(LGYSVAA)^{440} at N429, C) glycopeptide ^{1066}AEYFIN*VTTR^{1075} at N1071 and D) glycopeptide ^{1076}VWN*R^{1079} at N1078. Abbreviations used are as follows: HexNAc, N, *N*-acetylhexosamine; NeuGc, S, *N*-glycolyl neuraminic acid; NeuAc, S*, *N*-acetyl neuraminic acid; Hex, H, hexose; F, fucose (deoxyHex). Lower case c in a peptide sequence indicates a cysteine that has been modified by carboxymethylation prior to protease digestion.



Supplemental Figure 5: More extensively processed glycopeptiforms at *N*-glycan sites N102, N457 and N472 of integrin α2 in megakaryocytes of wild type (WT) and JAK2(V617F+) male mice. The total signal corresponding to the sum of all glycoforms of the glycopeptides that span each *N*-linked site is normalized to the total ion current for each of the biological replicates. Each biological replicate is depicted as a single point, the mean is indicated by a dash (WT) or by a square [JAK2(V617F+)]. Extensions of the peptide sequence due to alternative proteolytic cleavages detected around a given glycosite (N*) are indicated in parenthesis. A) Glycopeptide ^{92}(LNLQN)SASISN*VTEIK^{107} at N102, B) Glycopeptide ^{456}AN*YTGQIVL(YSVNK)^{469} at N457 and C) glycopeptide ^{470}QGN*VTVIQSH(R)^{480} at N472 Abbreviations used are as follows: HexNAc, N, *N*-acetylhexosamine; NeuGc, S, *N*-glycolyl neuraminic acid; NeuAc, S*, *N*-acetyl neuraminic acid; Hex, H, hexose; F, fucose (deoxyHex). Lower case c in a peptide sequence indicates a cysteine that has been modified by carboxymethylation prior to protease digestion.



Supplemental Figure 6: Representation of integrin α2 (top) and integrin β1 (bottom) structures into similar projections using AlphaFold program (AlphaFold Data Copyright (2021) DeepMind Technologies Limited [46]. Predicted structures generated by AlphaFold have been modified to indicate the potential *N*-linked glycosylation sites. Twelve integrin β1 potential *N*-glycosylation sites spanning all domains are represented: I-like (N212, N269 and N363), PSI and hybrid (N50, N94, N97, N406 and N417), membrane proximal (EGF repeat and β-tail, N481, N520, N584 and N669). For integrin α2, nine potential *N*-glycosylation sites are represented: situated on or in close proximity to the β-propeller domain repeat and I-domain (in contact with the ECM, N457, N102, N109, N429, N472 and N696) and on a region close to the transmembrane (N1054, N1071 and N1078).

Supplemental Tables

Supplemental Table 1. For each phenotype, the sample was separated into three aliquots based on migration along an αSDS-PAGE gel and each of the bands was subjected to in-gel proteolytic digestion (reported as upper and lower 100-150 kDa, 150-250 kDa. The peptide digests were analyzed by LC-MS/MS and the data sets were analyzed using Byonic version 3.8-11 software (Protein Metrics, Cupertino, CA).. Listed here are the glycopeptiforms assigned to mouse integrin β1. **See the individual Excel file Table1_MK_ITB1.**

Supplemental Table 2. For each phenotype, the sample was separated into three aliquots based on migration along an SDS-PAGE gel and each of the bands was subjected to in-gel proteolytic digestion (reported as upper and lower 100-150 kDa, 150-250 kDa. The peptide digests were analyzed by LC-MS/MS and the data sets analyzed using Byonic version 3.8-11 software (Protein Metrics, Cupertino, CA). Listed here are the glycopeptiforms assigned to mouse integrin α2. **See the individual Excel file Table2_MK_ITA2_UpLow100_150gelband.**

Collection date	Biological replicate	Total number of cells		Total amount of protein in MK lysate (µg)	
		WT	JAK2(V617F)	WT	JAK2(V617F)
2/14/2020	#1	3.08E+05	5.63E+05	1951.6	2073.5
6/16/2020	#2	4.25E+05	3.33E+05	1847.2	1428.1
10/26/2020	#3	1.80E+06	1.28E+06	2872.7	2886.5

Supplemental Table 3. Cell counts and total amount of protein measured for MK lysates in WT and JAK2^{V617F} samples.

Integrin α2	Rel. area [a,b]	
	WT	JAK2(V617F+)
N109		
HexNAc(2)Hex(5-10) [H(5-10)]		
K.TN*MSLGL.T	-	3.99E-04
K.TN*mSLGLTLTR.N	3.14E-03	1.24E-02
N429		
HexNAc(2)Hex(5-10) [H(5-10)]		
R.N*HSSF.L	-	3.56E-05
R.N*HSSFLGY.S	4.18E-04	1.06E-03
R.N*HSSFLGYSVAA.I	-	1.43E-04
N1071		
HexNAc(2)Hex(5-10) [H(5-10)]		
K.AEYFIN*VTTR.V	3.69E-03	1.04E-02
N1078		
HexNAc(2)Hex(5-10) [H(5-10)]		
R.VWN*R.T	2.55E-04	6.26E-04
N102		
HexNAc(4)Hex(5)Fuc(1)NeuGc(2) [$N_4H_5F_1S_2$]		
K.LNLQNSASISN*VTEIK.T	6.06E-04	1.74E-03
N.SASISN*VTEIK.T	7.19E-05	4.01E-04
N457		
HexNAc(2)Hex(5-10) [H(5-10)]		
R.AN*YTGQIVL.Y	-	3.89E-04
R.AN*YTGQIVLY.S	9.14E-04	1.08E-03
R.AN*YTGQIVLYSVNK.Q	3.62E-03	1.14E-02
HexNAc(3)Hex(6) [N_3H_6]		
R.AN*YTGQIVLY.S	4.52E-05	4.04E-04
R.AN*YTGQIVLYSVNK.Q	7.04E-04	2.01E-03

Integrin α2	Rel. area [a,b]	
	WT	JAK2(V617F+)
N472		
HexNAc(2)Hex(5-10) [H(5-10)]		
K.QGN*VTVIQSHR.G	2.70E-04	1.45E-03
HexNAc(3)Hex(6) [N_3H_6]		
K.QGN*VTVIQSHR.G	7.79E-05	1.16E-04
HexNAc(3)Hex(5)Fuc(1) [$N_3H_5F_1$]		
K.QGN*VTVIQSH.R	-	1.61E-05
K.QGN*VTVIQSHR.G	8.77E-05	4.08E-04
HexNAc(3)Hex(6)Fuc(1) [$N_3H_6F_1$]		
K.QGN*VTVIQSHR.G	7.77E-05	1.29E-04
HexNAc(3)Hex(6)NeuAc(1) [$N_3H_6S^*_1$]		
K.QGN*VTVIQSHR.G	2.26E-04	1.74E-03
HexNAc(3)Hex(6)Fuc(1)NeuGc(1) [$N_3H_6F_1S_1$]		
K.QGN*VTVIQSHR.G	1.59E-04	6.19E-04
HexNAc(4)Hex(5)Fuc(1) [$N_4H_5F_1$]		
K.QGN*VTVIQSH.R	-	5.38E-05
K.QGN*VTVIQSHR.G	2.83E-04	1.24E-03
HexNAc(4)Hex(5)Fuc(1)NeuGc(1) [$N_4H_5F_1S_1$]		
K.QGN*VTVIQSH.R	-	7.39E-05
K.QGN*VTVIQSHR.G	6.40E-04	2.33E-03
HexNAc(4)Hex(5)Fuc(1)NeuGc(2) [$N_4H_5F_1S_2$]		
K.QGN*VTVIQSHR.G	1.76E-04	1.10E-03

[a] Peptide sequences - including sequences resulting from alternative proteolytic cleavage - covering *N*-glycosylation sites N109, N429, N1071, N1078, N102, N457 and N472 are indicated. Lower case m in peptide sequence indicates an oxidized methionine residue. Abbreviations used are as follow: HexNAc, N for *N*-acetylhexosamine; NeuGc, S for *N*-glycolyl neuraminic acid; NeuAc, S*, *N*-acetyl neuraminic acid; Hex, H for hexose; F for fucose (deoxyHex).
[b] Average relative area of characterized glycopeptidoforms characterized in at least two out of the three biological replicates.

Supplemental Table 4. Average relative area of described glycopeptidoforms characterized in at least two out of the three biological replicates that span *N*-linked glycosylation sites N109, N429, N1071, N1078, N102, N457 and N472 of integrin α2isolated from MKs derived from 13- to 14- weeks-old male wild type (WT) and JAK2(V617F+) mice. Peptide sequences - including sequences resulting from alternative proteolytic cleavage - covering *N*-glycosylation sites N109, N429, N1071, N1078, N102, N457 and N472 are indicated. Lower case m in peptide sequence indicates an oxidized methionine residue. Abbreviations used are as follows: HexNAc, N for *N*-acetylhexosamine; NeuGc, S for *N*-glycolyl neuraminic acid; NeuAc, S*, *N*-acetyl neuraminic acid; Hex, H for hexose; F for fucose (deoxyHex).

Supplemental Table 5: Protein identifications obtained using Peaks Studio X+ software (Bioinformatics Solutions Inc., Waterloo, Canada) for analysis of the data obtained for co-IP proteins in SDS-PAGE gel bands situated between 100 kDa and 250 kDa. **See individual Excel File Table5_Co-IP proteins**