Integrating genomic alterations in diffuse large B-cell lymphoma identifies new relevant pathways and potential therapeutic targets

Supplemental Material

Supplemental Methods

1. Gene Selection
2. Target next generation sequencing
3. Variant Calling Algorithms
4. NGS Verification. Mutation analysis of *CDKN2A* and 3'UTR region of *NOTCH1*
5. Selection of potential driver mutations
6. Copy number variant and structural alterations analysis
7. Cell of origin determination
8. Gene expression analysis

Supplemental Tables

- Table S1. A) Clinical characteristics of the patients in the study vs patients not included B) COO of the 150 patients included in the initial series.
- Table S2. Genes sequenced using target NGS and mean coverage by gene and amplicon.
- Table S3. All called mutations in the 150 DLBCL of the initial series.
- Table S4. Potential driver mutations identified in the 150 DLBCL of the initial series.
- Table S5. Distribution of frequent mutations in COO subgroups.
- Table S6. Gene sets /pathways.
- Table S7. Copy number and copy number neutral loss of heterozygosity alterations in 119 DLBCL cases of the initial series.
- Table S8. Distribution of frequent CNV alterations in COO subgroups.
- Table S9. Prognostic impact of COO, R-IPI and other initial characteristics on CR, PFS and OS.
- Table S10. Baseline clinical and biological features of the patients according to the mutational status of the pathways.
- Table S11. Frequency of mutations of selected genes in the initial and validation

cohorts.

- Table S12. Baseline clinical features of the patients according to mutational status of the NOTCH pathway, TP53/CDKN2A, and JAK/STAT pathway in the validation series.
- Table S13. *In silico* drug prescription of DLBCL patients.
- Table S14. Primers used for Sanger sequencing, Access Array Fluidigm and Nextera XT approaches.

Supplemental Figure legends

- Figure S1. PFS and OS in the initial and validation series.
- Figure S2. Pipeline followed to filter potential driver mutations in 150 DLBCL samples. *SIFT was only used for mutations in which a definitive score was not provided by Mutation Assessor (MA).
- Figure S3. Genetic alterations characterized in 150 DLBCL patients.
- Figure S4. Alterations of *TMEM30A, PRDM1, SGK1, TNFAIP3* in the common deleted 6q14-q23 region.
- Figure S5. Chromothripsis-like patterns detected in DLBCL cases. (a) Amplifications of *miR17-92*-13q31.3 in 3 cases of DLBCL with a chromothripsis-like pattern on chromosome 13. (b) Amplifications of *REL* and *BCL11A* genes in one DLBCL case with a chromothripsis-like pattern in chromosome 2.
- Figure S6. Gene set enrichment analysis (GSEA) comparing *SGK1* and NOTCH pathway mutated *vs* unmutated cases.

Supplemental Methods

Gene Selection

We selected 106 genes from previous whole exome sequencing studies of diffuse large B-cell lymphomas (DLBCL) and a large number of studies that analyzed the mutations of individual or small sets of genes in these lymphomas (Supplemental Table S2).[1-32] In total, 3951 different genes were initially reviewed mainly from whole exome sequencing studies.[13, 16, 24, 32] Genes were selected according to the following criteria: Recurrent mutated genes in more than 10% of DLBCL in any of the studies, mutated

genes in which a somatic mutation was confirmed in at least 2 cases or annotated in COSMIC as somatic, and genes located on frequently altered regions in DLBCL. Three additional criteria were: functional evidence of the pathogenic role of the mutated gene in DLBCL, genes involved in pathways relevant for DLBCL, and drug target genes (45 genes were selected because they had been considered as drug targets in clinical trials according to the public database (http://www.cancer.gov/clinicaltrials). Some additional genes were added because they were reported in other lymphomas or could complement some of the pathways of interest (*BLM, BRCA2, CCNH, FBXW7, HIST1H2BD, IDH1, ID3, MEF2C, MKI67, MPL, POU2F2, TCF3, SEMA5A, and WHSC1).* We excluded several genes initially selected according to previous criteria for different reasons such as difficulties for primer design (e.g *P2RY8*) or because they were known to be late replication genes, or because of their large size (*PCLO, LRP1B, MUC16* and *UNC5D*) and the fact that they have been reported to accumulate large numbers of passenger mutations.[33]

Target next generation sequencing

Two-hundred twenty-five nanograms of genomic DNA extracted from frozen tumor tissues were used to generate NGS libraries. Five 10 μm-thick sections per sample were used to extract DNA using the QIAamp DNA Mini Kit according to the manufacturer's instructions (Qiagen). In the initial series, libraries were generated using HaloPlex technology (Agilent technologies, Santa Clara, CA; following the manufacturer's protocol). The customed HaloPlex kit included all exons and their flanking regions (Supplemental Table S2). Libraries were sequenced in a MiSeq instrument (Illumina, San Diego, CA) in a paired-end run of 150 bp. The average sequencing coverage across regions was 600x and a coverage >20x was obtained in >98.7% of the target regions.

Libraries of the validation cohort were generated using Access-Array technology (Fluidigm) and Nextera XT procedure (Illumina). Briefly, primers to amplify *TP53* (exons 4-10), *MYD88* (exons 2-8), *NOTCH2* (exon 34), *CCND3* (exon 5), *SGK1* (all exons), *STAT3* (exons 20-21), *STAT6* (exons 11-17), *PIM1* (exons 1-14), *FBXW7* (all exons) and *TMEM30A* (exons 1-7) regions were designed with the D3-Assay Design web-based tool (Fluidigm) (Supplemental Table S14). Libraries were generated using

50 ng of total DNA in the Access-Array system with a BioMark thermal cycler (Fluidigm) and sequenced in a MiSeq instrument with a paired-end run of 210 bp. Specific primers to sequence *NOTCH1* (exon 34), *SOCS1* (all exons) and *FOXO1* (all exons) were designed using the Primer3 program (Supplemental Table S14). Long-PCR amplifications were performed using the KAPA HiFi DNA Polymerase HotStart ReadyMix (Kapa Biosystems) and normalized with the SequalPrep Normalization Plate kit (Invitrogen). Libraries were generated with the Nextera XT DNA Library Preparation Kit (Illumina) and sequenced with a 2x150 bp MiSeq run. The median sequencing coverage across region was 555x (range 131-933) and a coverage >20x was obtained in >98.7% of the target regions.

Variant calling algorithms and verification assessment

Two different bioinformatics pipelines were used for the alignment and variant calling, HD Genome One Research Edition software (DREAMgenics; http://www.dreamgenics.com/) and Agilent SureCall tool (http://www.genomics.agilent.com). FASTQ files generated by MiSeq control software were processed using the above mentioned algorithms and both were compared.

Genome One Research Edition software

A new algorithm was developed in cooperation with DREAMgenics.Inc to perform the alignment, calling and annotation of the variants obtained with Haloplex Libraries. The allelic frequency cut-off for considering mutations was 5%. All variants were confirmed by visual inspection. Low coverage calls (total read depth < 10, or mutated allele count <5 calls) and 17 low quality calls were excluded. We excluded for further analysis all synonymous and intron variants and known polymorphisms included in dbSNP database (dbSNP138), ESP6500 (http://evs.gs.washington.edu/EVS) with more than 1% frequency in European population or in our own database of polymorphisms in Spanish population.[34]

Agilent SureCall

We used SureCall tool (ver1.1) with all default settings to analyze sequencing results and to call the variants. All variants were confirmed by visual inspections. All

synonymous variants, intron variants and known polymorphisms included in dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) or in our own database of polymorphisms in Spanish population were excluded.[34]

Variant calls of both algorithms were integrated and, finally, 1331calls were selected in the 150 cases (Supplemental Figure S2 and Supplemental Table S3). To determine the accuracy of the sequencing method and analytical algorithms, we selected 152 (11%) variants from these 1331 called variants, and we verified them using Sanger sequencing. One hundred fifty-one (99%) of these variants were confirmed, indicating the accuracy of the analysis.

For the validation series, the complete bioinformatic analysis, alignment and variant calling were performed with the MiSeq Reporter Software (MSR, version 2.4.60). All variants detected by any of these two algorithms were combined and annotated using ANNOVAR[35] as well as custom scripts. Like in the initial series analysis, we excluded all synonymous and intron variants and known polymorphisms included in dbSNP database (dbSNP138), ESP6500 (http://evs.gs.washington.edu/EVS/) with more than 1% frequency in European population or in our own database of polymorphisms in Spanish population.[34]

Sanger sequencing. Verification of NGS results and mutational analysis of *CDKN2A* and 3'UTR region of *NOTCH1*

Sanger sequencing was used to verify the results of the NGS described above and to analyze Exon 1alfa, 1beta and 2 of *CDKN2A* and 3'UTR region of *NOTCH*1.[34] PCR primers were designed using Primer 3 (http://frodo.wi.mit.edu/primer3) and purchased from Sigma-Aldrich (St. Louis, MO). Amplification by PCR was performed using AmpliTaq Gold DNA Polymerase (Life technologies, Grand Island, NY) or QIAGEN Multiplex PCR Kit (Qiagen, Madrid, Spain) with 50ng of DNA and 200µM dNTP mix (Life technologies) following the manufacturer's recommendations. All PCR products were run in a capillary electrophoresis gel (QIAxcel Advanced System, Qiagen) with the QIAxcel DNA screening kit (Qiagen). The multiband PCR products were purified using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Bethlehem, PA).

Regarding Sanger sequencing, PCR products were cleaned using ExoSAP-IT (Affymetrix, Santa Clara, CA) and sequenced using ABI Prism BigDye terminator v3.1 (Life technologies) with 5 pmol of each primer. Sequencing reactions were run on an ABI-3730 automated sequencer (Life Technologies). All sequences were examined with the Mutation Surveyor DNA Variant Analysis Software (SoftGenetics, State College, PA). Sequences of the used primers are listed in Supplemental Table S14.

Selection of potential driver mutations and verification in germline DNA

Potential driver mutations were selected according to the following criteria (Supplementary Fig S2): 1) We initially selected 216 "relevant" mutations by manual curation based on previous reports and COSMIC, including somatic and functional mutations and mutations clustering in known functional domains [1, 3, 5, 7, 11, 12, 15, 17, 19, 22, 25,27,29-31,36-54] . 2) All truncating mutations (n= 274), except two found in known oncogenes (*MYD88*, M1I and *CD79A*, Q222*), were also considered as potential driver events. 3) The potential drivers of the remaining missense and in-frame mutations were selected based on the functional prediction established by the OncodriveCLUST, Mutation Assessor (MA) and SIFT algorithm.[55-57] MA and SIFT algorithms were selected after comparing several methods including these two and CHASM[58], CONDEL[59], FATHMM[60], Mutation Tester[61], and Polyphen2 (PPH2)[62]. To select the most appropriate algorithm we initially explored the performance of each of these algorithms when distinguishing the variants found in our study that were known polymorphisms or known somatic recurrent mutations described in COSMIC. In this particular data set, the scores that better predicted the expected characteristics of the variant were MA and SIFT, followed by CHASM. We did not observe any benefit in combining these scores. We selected MA because it showed a narrower score range for polymorphisms and larger differences between recurrent and non-recurrent COSMIC entries as compared to SIFT. For those cases in which the MA score could not be retrieved, we used a SIFT score. Using these two algorithms, 271 out of the 841 missense or in-frame mutations were selected as driver mutations.

To test the accuracy of our "functional prediction" algorithm for missense mutations, we selected 92 variants in 32 patients who had germline DNA available. We observed that

90% of the mutations classified as functional were somatic (28/31) while 89% of the germline mutations were classified as non-functional (24/27) (Supplemental Methods and Supplemental Table S15). The 34 somatic variants predicted as non-functional by the algorithm were not considered drivers. Taking these three criteria together we selected 761 potential driver mutations (58% of the total) for the clinicopathological analysis (Supplemental Figure S2 and Supplemental Table S4).

Copy Number and Structural Alteration Analysis

Samples were analyzed using CytoScan HD Array (Affymetrix) according to the manufacturer's instructions. Scanned data from CytoScan HD were processed by Chromosome Analysis Suite (Affymetrix) for subsequent analyses. The analytical programs of "Nexus CN 7.5 Discovery edition" (Biodiscovery, Hawthorne, CA), SNP-FASST2 (Biodiscovery) and ASCAT (http://heim.ifi.uio.no/bioinf/Projects/ASCAT/) were used to analyze genomic alterations. Minimal common regions (MCRs) of gains and losses were picked up using an R custom script. The most frequently altered regions (>=20%) were extracted and selected by visual inspection of two different observers (K.K. and I.S.). A total number of 34 MCRs were selected. Then ABC and GCB type DLBCL cases were analyzed separately and MCRs specific for each molecular subtype were identified.

Genomic alterations satisfying the following criteria were regarded as "deep losses" or "high gains": (1) An amplitude of an alteration was more than 1 or less than -1 in log2 ratio. (2) Deeper and shorter gains or losses were identified in longer and shallower alterations (gain in gain or loss in loss). "Homozygous loss" and "amplification" were defined as "deep loss" and "high gain" regions less than 5Mb, respectively. 6q14.1 and 10q23.31 regions, in which *TMEM30A* and *PTEN* were located respectively, were further added to the alteration list because they showed recurrent homozygous losses. Additionally, 17 regions previously reported were included.[3, 63-70] In total, 62 chromosomal regions were selected for further analysis including clinical correlation. Loss of heterozygosity (LOH) without genomic alteration (CNN-LOH) was considered when the size of the altered region was >5Mb. Cases were regarded to have

chromothripsis when at least seven switches between two or more copy number states were detected on an individual chromosome in which LOH was retained.[71]

The copy number alterations of *CDKN2A* were investigated in the validations series using TaqMan® Genotyping Master Mix and TaqMan® Copy Number Assay Hs02738179_cn for *CDKN2A*. DNA was analyzed using duplicates in a StepOnePlus™ Real-Time PCR System (Thermo Fisher Scientific). Relative quantification of the gene locus was analyzed with the $2^{-\Delta\Delta Ct}$ method using TaqMan® Copy Number Reference Assay RNaseP as the endogenous control and JVM-2 cell line (wild type for *CDKN2A*) as mathematical calibrator. We included the cell line MAVER-1 as a control of homozygous deletion.

Interphase fluorescence *in situ* hybridization (FISH) analysis was carried out on FFPE sections to detect *MYC*, *BCL2* and *BCL6* breaks using specific probes supplied by Abbott Molecular (Des Plaines, IL, USA) following the manufacturer's specifications. The FISH probe used for the BCL6 FISH analysis was the LSI *BCL6* (ABR) Break Apart Rearrangement Probe, (Abbot Molecular Des Plaines, USA) that identifies both breakpoints located in the major and alternative breakpoint region.

Cell of origin determination

The molecular cell of origin (COO) of the tumors was established using the Gene Chip Human Genome U133 Plus 2.0 arrays (Affymetrix, Santa Clara, CA) and/or the Lymph2Cx assay (Nanostring technologies, Seattle, WA).[72] To determine the COO in the training series, total RNA was extracted from frozen tissues using RNeasy Kit (Qiagen) following the manufacturer's instructions. For the validation series, total RNA was obtained from formalin-fixed paraffin-embedded material (FFPE-M). Five 10 μm-thick sections per sample were used to extract RNA using the RNeasy FFPE Kit according to the manufacturer's instructions (Qiagen). RNA integrity from frozen specimens was examined with the Agilent 2100 Bioanalyzer (Agilent Technologies) and only high-quality RNA samples were hybridized to Affymetrix Human Genome Array U219 array plates according to Affymetrix standard protocols. Summarized expression values were computed using the robust multichip average approach

implemented in the Expression Console Software (Affymetrix Inc.). COO was determined as previously reported.[73] For RNA extracted from FFPE tissues we used a digital multiplexed gene expression analysis with the nCounter/Nanostring technology following the established protocol.[72] Samples were classified as GCB, ABC and Unclassified (UC) subtypes using the algorithm previously described.[72]

Gene expression analysis

To verify the biological relevance of NOTCH pathway activation in DLBCL, we compared the gene expression profiles of 12 cases with NOTCH pathway mutations (5 *NOTCH2*, 4 *SGK1*, 2 *NOTCH1*, 1 *FBWX7*) and 27 with wild-type genes of this pathway using Affymetrix® Human Genome U219. A gene set enrichment analysis (GSEA) was performed comparing *SGK1* mutated and unmutated cases. We tested the KEGG NOTCH signaling pathway, two lists of genes upregulated by NOTCH signaling and two other gene-sets downregulated by NOTCH.[74, 75] In addition, *HES1* mRNA expression was analyzed in 14 cases with NOTCH pathway mutations and 13 with wild-type genes of this pathway by qRT-PCR using a designed human Taqman® Gene expression Assay for *HES1* (Hs00172878_m1; Applied Biosystems, Foster City, CA). Gene expression was quantified by the comparative cycle threshold (Ct) method (ΔΔCt) using GUS as endogenous control. All real-time PCR reactions for the individual samples were performed in triplicate. Results were expressed as relative gene expression (versus *GUS* gene expression) using arbitrary units.

Supplemental References

1. Abubaker J, Bavi PP, Al-Harbi S, Siraj AK, Al-Dayel F, Uddin S, *et al.* PIK3CA mutations are mutually exclusive with PTEN loss in diffuse large B-cell lymphoma. *Leukemia* 2007; **21:** 2368-2370.

2. Asmar F, Punj V, Christensen J, Pedersen MT, Pedersen A, Nielsen AB, *et al.* Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. *Haematologica* 2013; **98:** 1912-1920.

3. Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty VV, *et al.* Combined genetic inactivation of beta2-Microglobulin and CD58 reveals

frequent escape from immune recognition in diffuse large B cell lymphoma. *Cancer cell* 2011; **20:** 728-740.

4. Compagno M, Lim WK, Grunn A, Nandula SV, Brahmachary M, Shen Q, *et al.* Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* 2009; **459:** 717-721.

5. Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* 2010; **463:** 88-92.

6. Dong G, Chanudet E, Zeng N, Appert A, Chen YW, Au WY, *et al.* A20, ABIN-1/2, and CARD11 mutations and their prognostic value in gastrointestinal diffuse large B-cell lymphoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2011; **17:** 1440-1451.

7. Du MQ, Peng H, Liu H, Hamoudi RA, Diss TC, Willis TG, *et al.* BCL10 gene mutation in lymphoma. *Blood* 2000; **95:** 3885-3890.

8. Honma K, Tsuzuki S, Nakagawa M, Tagawa H, Nakamura S, Morishima Y, *et al.* TNFAIP3/A20 functions as a novel tumor suppressor gene in several subtypes of non-Hodgkin lymphomas. *Blood* 2009; **114:** 2467-2475.

9. Kato M, Sanada M, Kato I, Sato Y, Takita J, Takeuchi K, *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* 2009; **459:** 712-716.

10. Lee SY, Kumano K, Nakazaki K, Sanada M, Matsumoto A, Yamamoto G, *et al.* Gain-of-function mutations and copy number increases of Notch2 in diffuse large B-cell lymphoma. *Cancer science* 2009; **100:** 920-926.

11. Lenz G, Davis RE, Ngo VN, Lam L, George TC, Wright GW, *et al.* Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* 2008; **319:** 1676-1679.

12. Li H, Kaminski MS, Li Y, Yildiz M, Ouillette P, Jones S, *et al.* Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A underlying the pathogenesis of follicular lymphoma. *Blood* 2014; **123:** 1487-1498.

13. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, *et al.* Discovery and prioritization of somatic mutations in diffuse large B-cell

lymphoma (DLBCL) by whole-exome sequencing. *Proceedings of the National Academy of Sciences of the United States of America* 2012; **109:** 3879-3884.

14. Love C, Sun Z, Jima D, Li G, Zhang J, Miles R, *et al.* The genetic landscape of mutations in Burkitt lymphoma. *Nature genetics* 2012; **44:** 1321-1325.

15. Mandelbaum J, Bhagat G, Tang H, Mo T, Brahmachary M, Shen Q, *et al.* BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer cell* 2010; **18:** 568-579.

16. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 2011; **476:** 298-303.

17. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, *et al.* Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature genetics* 2010; **42:** 181-185.

18. Muppidi JR, Schmitz R, Green JA, Xiao W, Larsen AB, Braun SE, *et al.* Loss of signalling via Galpha13 in germinal centre B-cell-derived lymphoma. *Nature* 2014; **516:** 254-258.

19. Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim KH, *et al.* Oncogenically active MYD88 mutations in human lymphoma. *Nature* 2011; **470:** 115-119.

20. Pasqualucci L, Migliazza A, Basso K, Houldsworth J, Chaganti RS, Dalla-Favera R. Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma. *Blood* 2003; **101:** 2914-2923.

21. Pasqualucci L, Compagno M, Houldsworth J, Monti S, Grunn A, Nandula SV, *et al.* Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *The Journal of experimental medicine* 2006; **203:** 311-317.

22. Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* 2011; **471:** 189-195.

23. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Kuppers R, *et al.* Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 2001; **412:** 341-346.

24. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, *et al.* Analysis of the coding genome of diffuse large B-cell lymphoma. *Nature genetics* 2011; **43:** 830-837.

25. Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B, *et al.* Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nature genetics* 2012; **44:** 1316-1320.

26. Schif B, Lennerz JK, Kohler CW, Bentink S, Kreuz M, Melzner I, *et al.* SOCS1 mutation subtypes predict divergent outcomes in diffuse large B-Cell lymphoma (DLBCL) patients. *Oncotarget* 2013; **4:** 35-47.

27. Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M, *et al.* Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature* 2012; **490:** 116-120.

28. Schuetz JM, Johnson NA, Morin RD, Scott DW, Tan K, Ben-Nierah S, *et al.* BCL2 mutations in diffuse large B-cell lymphoma. *Leukemia* 2012; **26:** 1383-1390.

29. Trinh DL, Scott DW, Morin RD, Mendez-Lago M, An J, Jones SJ, *et al.* Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. *Blood* 2013; **121:** 3666-3674.

30. Xu-Monette ZY, Wu L, Visco C, Tai YC, Tzankov A, Liu WM, *et al.* Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study. *Blood* 2012; **120:** 3986-3996.

31. Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, *et al.* MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nature immunology* 2013; **14:** 1084-1092.

32. Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, *et al.* Genetic heterogeneity of diffuse large B-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* 2013; **110:** 1398-1403.

33. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, *et al.* Mutational heterogeneity in cancer and the search for new cancer-

associated genes. *Nature* 2013; **499:** 214-218.

34. Puente XS, Bea S, Valdes-Mas R, Villamor N, Gutierrez-Abril J, Martin-Subero JI, *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2015; **526:** 519-524.

35. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 2010; **38:** e164.

36. Minde DP, Anvarian Z, Rudiger SG, Maurice MM. Messing up disorder: how do missense mutations in the tumor suppressor protein APC lead to cancer? *Molecular cancer* 2011; **10:** 101.

37. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, *et al.* Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* 2013; **122:** 1256-1265.

38. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, *et al.* Mutations of the BRAF gene in human cancer. *Nature* 2002; **417:** 949-954.

39. Bohers E, Mareschal S, Bouzelfen A, Marchand V, Ruminy P, Maingonnat C, *et al.* Targetable activating mutations are very frequent in GCB and ABC diffuse large B-cell lymphoma. *Genes, chromosomes & cancer* 2014; **53:** 144-153.

40. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA, *et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* 2011; **471:** 235-239.

41. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature genetics* 2012; **44:** 1104-1110.

42. Hambleton S, Salem S, Bustamante J, Bigley V, Boisson-Dupuis S, Azevedo J, *et al.* IRF8 mutations and human dendritic-cell immunodeficiency. *The New England journal of medicine* 2011; **365:** 127-138.

43. Heinrich MC, Owzar K, Corless CL, Hollis D, Borden EC, Fletcher CD, *et al.* Correlation of kinase genotype and clinical outcome in the North American Intergroup Phase III Trial of imatinib mesylate for treatment of advanced gastrointestinal stromal tumor: CALGB 150105 Study by Cancer and Leukemia

Group B and Southwest Oncology Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008; **26:** 5360-5367.

44. Fan HY, Richards JS. Minireview: physiological and pathological actions of RAS in the ovary. *Molecular endocrinology* 2010; **24:** 286-298.

45. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 2011; **475:** 101-105.

46. Kiel MJ, Velusamy T, Betz BL, Zhao L, Weigelin HG, Chiang MY, *et al.* Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *The Journal of experimental medicine* 2012; **209:** 1553-1565.

47. Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S, Spina V, *et al.* The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *The Journal of experimental medicine* 2012; **209:** 1537-1551.

48. Urick ME, Rudd ML, Godwin AK, Sgroi D, Merino M, Bell DW. PIK3R1 (p85alpha) is somatically mutated at high frequency in primary endometrial cancer. *Cancer research* 2011; **71:** 4061-4067.

49. Koskela HL, Eldfors S, Ellonen P, van Adrichem AJ, Kuusanmaki H, Andersson EI, *et al.* Somatic STAT3 mutations in large granular lymphocytic leukemia. *The New England journal of medicine* 2012; **366:** 1905-1913.

50. Fasan A, Kern W, Grossmann V, Haferlach C, Haferlach T, Schnittger S. STAT3 mutations are highly specific for large granular lymphocytic leukemia. *Leukemia* 2013; **27:** 1598-1600.

51. Ohgami RS, Ma L, Merker JD, Martinez B, Zehnder JL, Arber DA. STAT3 mutations are frequent in CD30+ T-cell lymphomas and T-cell large granular lymphocytic leukemia. *Leukemia* 2013; **27:** 2244-2247.

52. Ritz O, Guiter C, Castellano F, Dorsch K, Melzner J, Jais JP, *et al.* Recurrent mutations of the STAT6 DNA binding domain in primary mediastinal B-cell lymphoma. *Blood* 2009; **114:** 1236-1242.

53. Zhao S, Choi M, Overton JD, Bellone S, Roque DM, Cocco E, *et al.* Landscape

of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 2013; **110:** 2916-2921.

54. Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A, *et al.* Mutation in TET2 in myeloid cancers. *The New England journal of medicine* 2009; **360:** 2289-2301.

55. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research* 2011; **39:** e118.

56. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols* 2009; **4:** 1073-1081.

57. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics* 2013; **29:** 2238-2244.

58. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, *et al.* Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer research* 2009; **69:** 6660-6667.

59. Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *American journal of human genetics* 2011; **88:** 440-449.

60. Shihab HA, Gough J, Cooper DN, Day IN, Gaunt TR. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics* 2013; **29:** 1504-1510.

61. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* 2010; **7:** 575-576.

62. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, *et al.* A method and server for predicting damaging missense mutations. *Nature methods* 2010; **7:** 248-249.

63. Scholtysik R, Kreuz M, Hummel M, Rosolowski M, Szczepanowski M, Klapper

W, *et al.* Characterization of genomic imbalances in diffuse large B-cell lymphoma by detailed SNP-chip analysis. *International journal of cancer Journal international du cancer* 2015; **136:** 1033-1042.

64. Bea S, Zettl A, Wright G, Salaverria I, Jehn P, Moreno V, *et al.* Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood* 2005; **106:** 3183-3190.

65. Monti S, Chapuy B, Takeyama K, Rodig SJ, Hao Y, Yeda KT, *et al.* Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large B cell lymphoma. *Cancer cell* 2012; **22:** 359-372.

66. Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, *et al.* Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proceedings of the National Academy of Sciences of the United States of America* 2008; **105:** 13520-13525.

67. Kameoka Y, Tagawa H, Tsuzuki S, Karnan S, Ota A, Suguro M, *et al.* Contig array CGH at 3p14.2 points to the FRA3B/FHIT common fragile region as the target gene in diffuse large B-cell lymphoma. *Oncogene* 2004; **23:** 9148-9154.

68. Takeyama K, Monti S, Manis JP, Dal Cin P, Getz G, Beroukhim R, *et al.* Integrative analysis reveals 53BP1 copy loss and decreased expression in a subset of human diffuse large B-cell lymphomas. *Oncogene* 2008; **27:** 318-322.

69. Gonzalez-Aguilar A, Idbaih A, Boisselier B, Habbita N, Rossetto M, Laurenge A, *et al.* Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2012; **18:** 5203-5211.

70. Bertrand P, Maingonnat C, Penther D, Guney S, Ruminy P, Picquenot JM, *et al.* The costimulatory molecule CD70 is regulated by distinct molecular mechanisms and is associated with overall survival in diffuse large B-cell lymphoma. *Genes, chromosomes & cancer* 2013; **52:** 764-774.

71. Salaverria I, Martin-Garcia D, Lopez C, Clot G, Garcia-Aragones M, Navarro A, *et al.* Detection of chromothripsis-like patterns with a custom array platform for chronic lymphocytic leukemia. *Genes, chromosomes & cancer* 2015; **54:** 668-680.

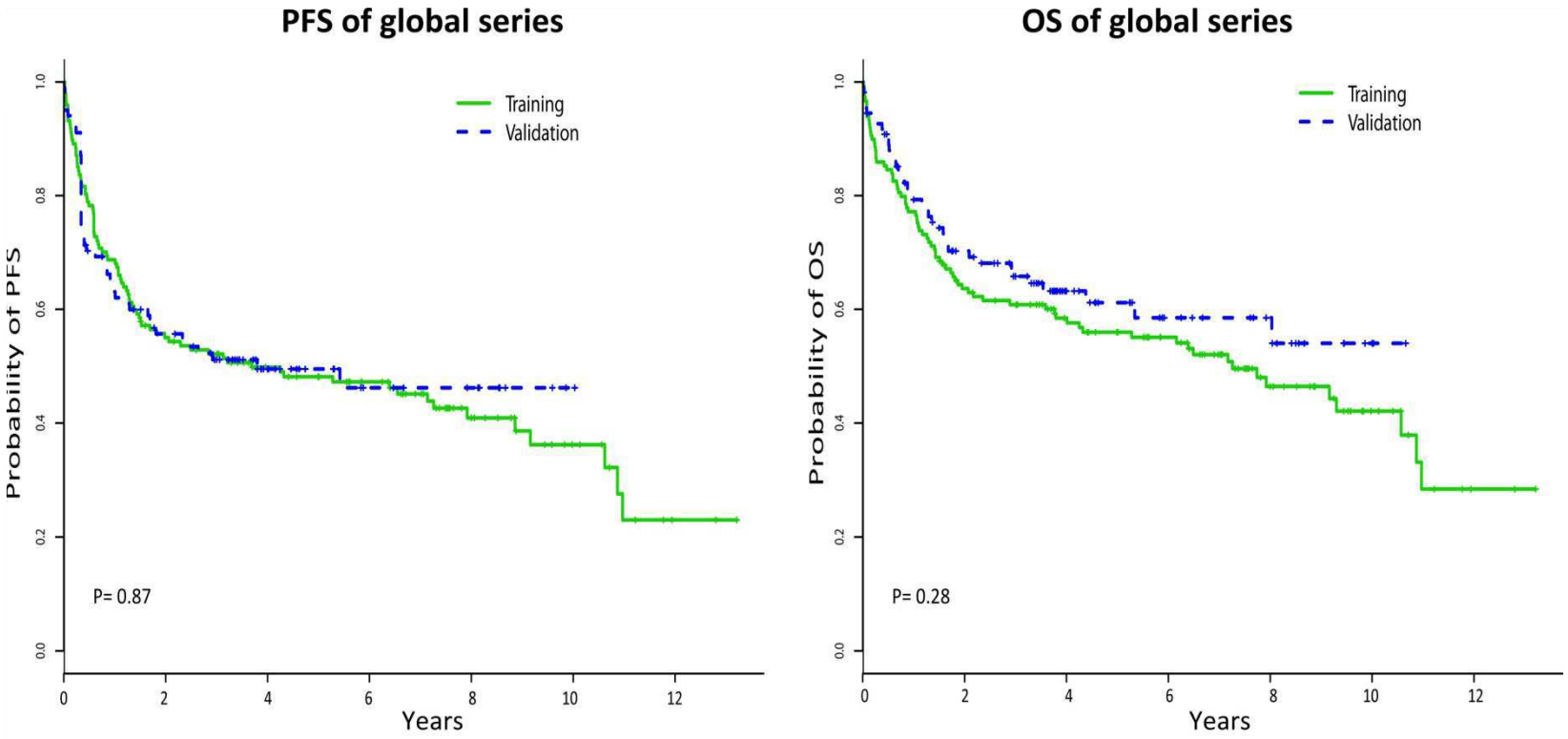
72. Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, *et al.* Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* 2014; **123:** 1214-1217.

73. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* 2003; **100:** 9991-9996.

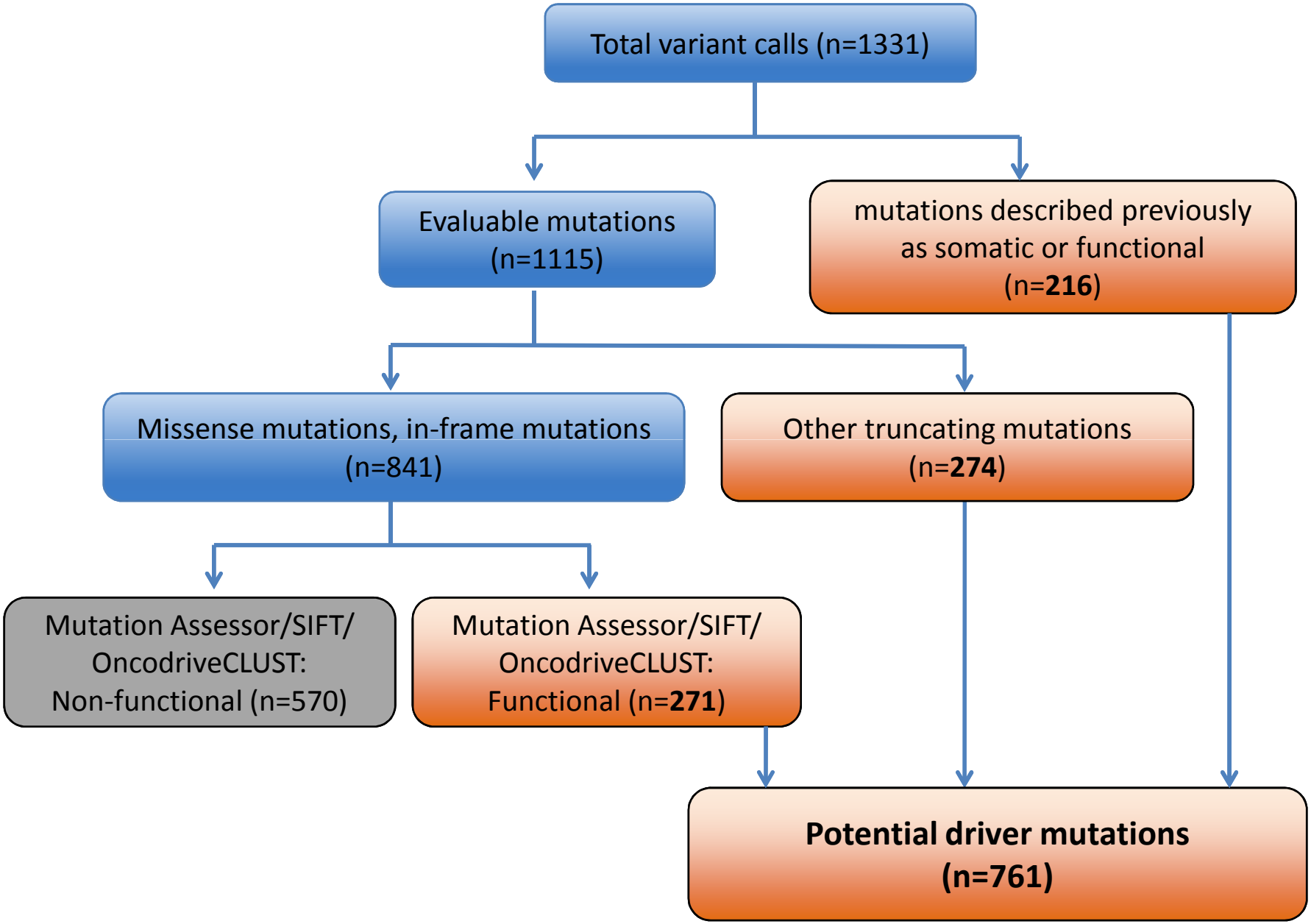
74. Sharma VM, Calvo JA, Draheim KM, Cunningham LA, Hermance N, Beverly L, *et al.* Notch1 contributes to mouse T-cell leukemia by directly inducing the expression of c-myc. *Molecular and cellular biology* 2006; **26:** 8022-8031.

75. Weng AP, Millholland JM, Yashiro-Ohtani Y, Arcangeli ML, Lau A, Wai C, *et al.* c-Myc is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. *Genes & development* 2006; **20:** 2096-2109.

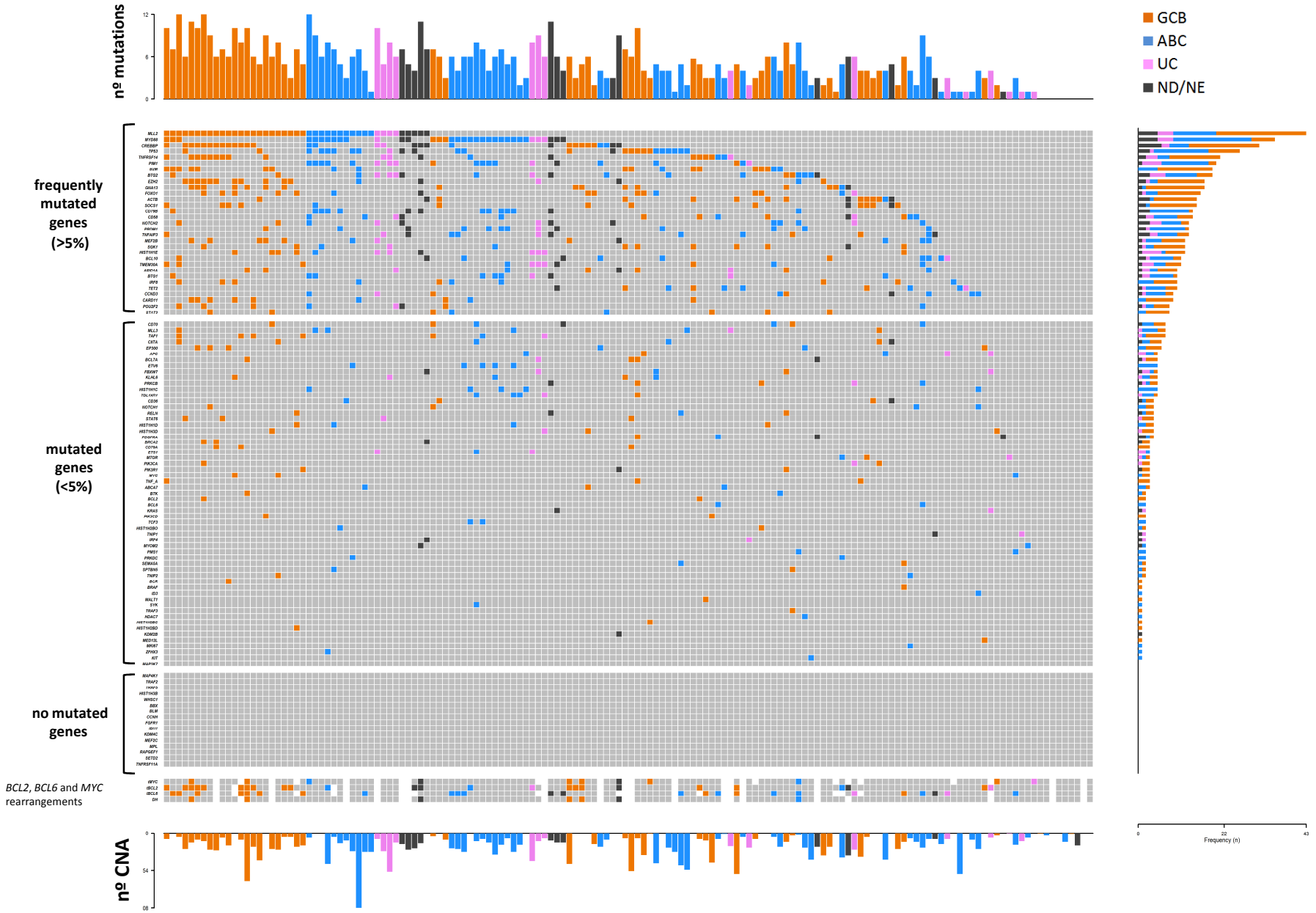
Supplemental Figure 1. PFS and OS in training and validation series.



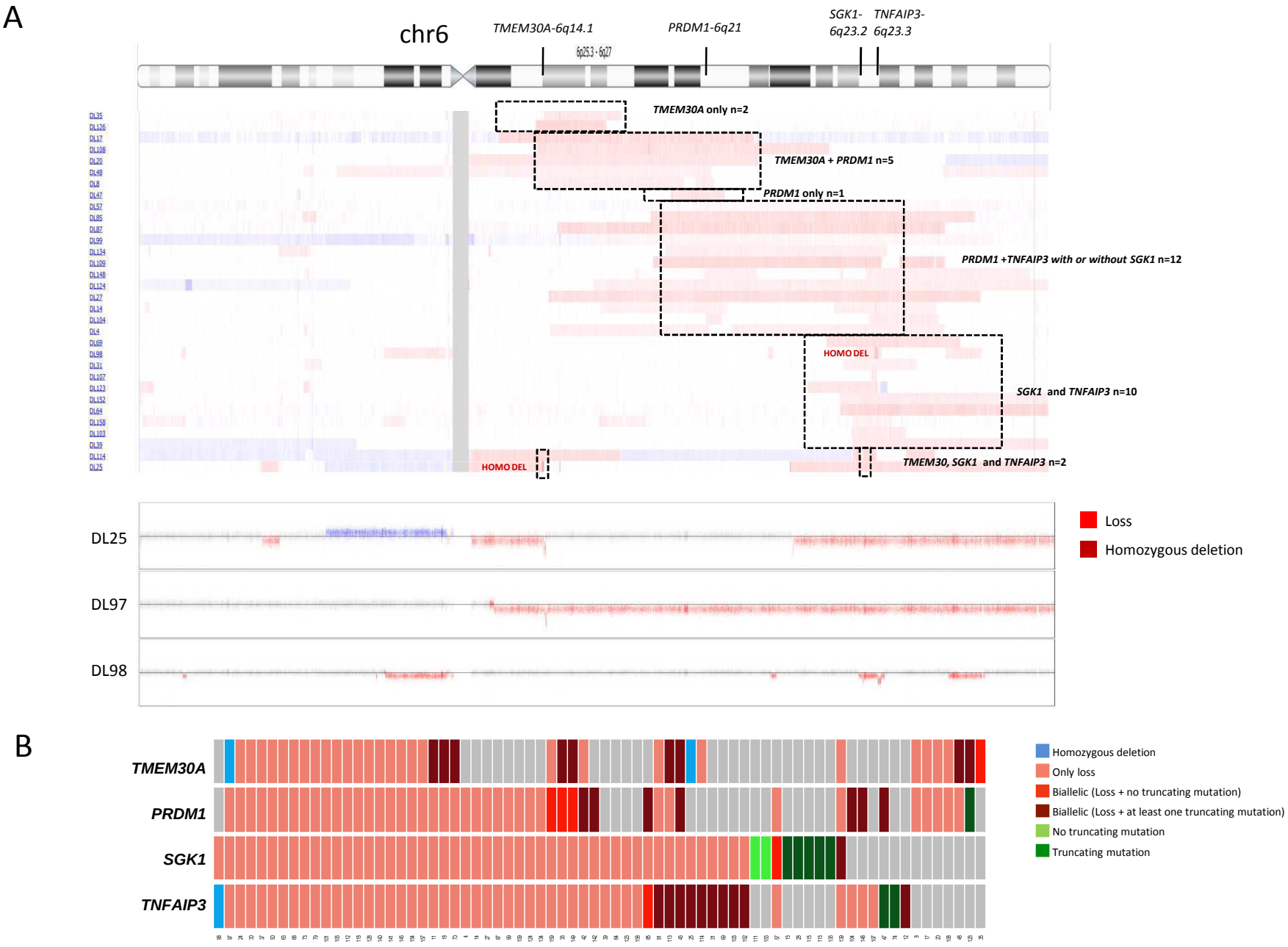
Supplemental Figure 2: Pipeline followed to filter potential driver mutations in 150 DLBCL samples



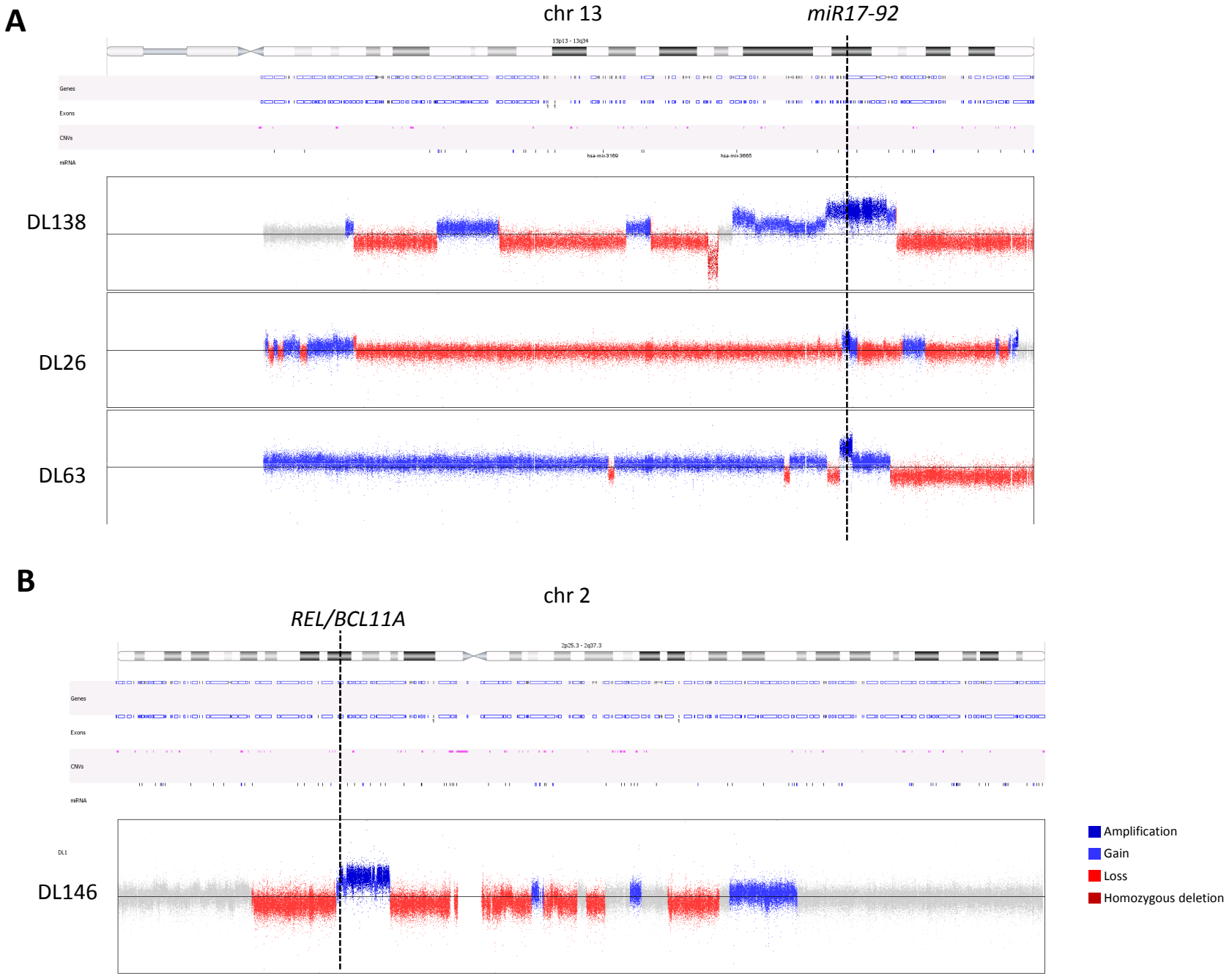
Supplemental Figure 3. Genetic alterations characterized in 150 DLBCL patients.



Supplemental Figure 4. Alterations of four genes in the common deleted 6q14-q23 region.



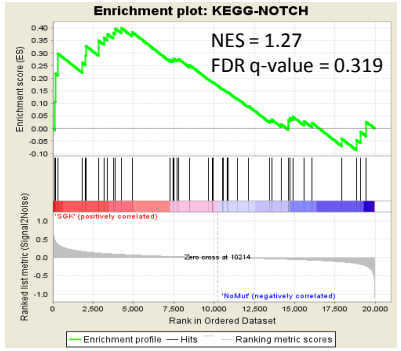
Supplemental Figure 5. Chromothripsis-like patterns detected in DLBCL cases.



Supplemental Figure 6. *GSEAs of SGK1* and NOTCH pathway mutated cases vs not mutated.

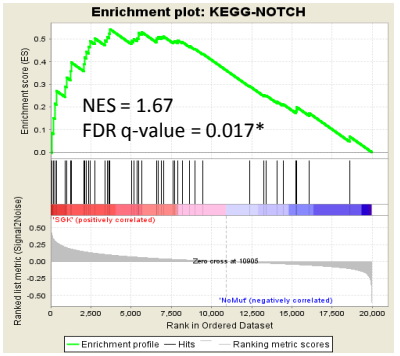
A

Mutated in *SGK*1 (4) vs
Not mutated in Notch Pathway (27)

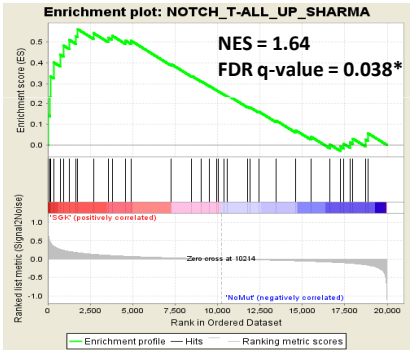


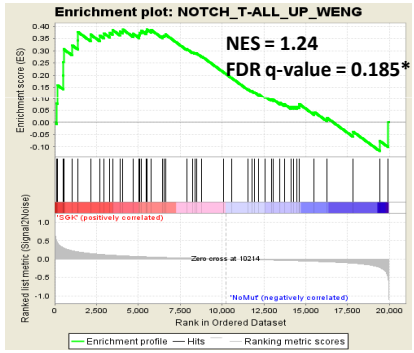
B

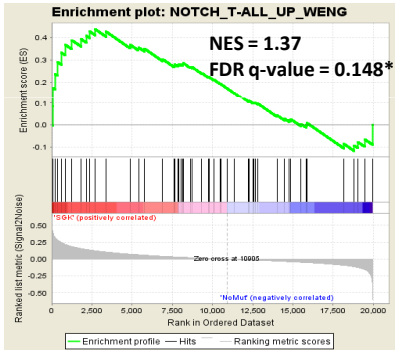
Mutated in *NOTCH pathway* (12) vs
Not mutated in Notch Pathway (27)

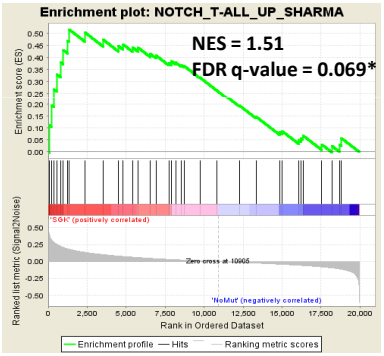


Genes up-regulated by Notch signalling pathway









Genes down-regulated by Notch signalling pathway