# Supplemental Methods

## CpG selection

CpGs used for CLL epitype discrimination were determined from CLL genome-wide methylation data sets and analysis performed in Kulis and colleagues[1] and Oakes and colleagues[2]. Briefly, DNA methylation subgroups were defined by selecting the most-variable probes on illumina arrays in the discovery cohort. The CLL maturation signature that underlies epitype is the most prevalent signature in genome-wide data (1st principal component in most-variable probe lists) and encompasses a large proportion of variable CpGs among CLL samples.[2] CpGs were ranked by P-value and fold difference between groups using Qlucore Omics Explorer software. The panel from Oakes and colleagues included the top 18 ranked CpGs (top 9 showing hypermethylation programming and top 9 hypomethylation programming). The discriminatory potential of the top candidate CpGs was validated in the Kulis and colleagues cohort. From this panel, we further selected the best 6 CpGs (3 hyper- and 3 hypomethylated) based on the most favorable Me-iPLEX primer design parameters (see below) and technical performance in preliminary experiments. We further included the *ZAP70* promoter region in the panel as *ZAP70* promoter methylation retains strong independent prognostic significance in CLL[3], but lacked representation on Illumina 450K arrays. CpGs used in assessing sample purity and composition were selected from Illumina 450K profiles of purified healthy hematopoietic subsets.[4] Here we performed supervised clustering analyses individually separating CLL versus normal PBMCs, T and NK cells and myeloid cells (granulocytes and monocytes). Probes that showed the most consistent differences ranked by p-value and fold change were selected for Me-iPLEX assay design. The high degree of redundancy within the CLL maturation signature affords many CpGs to be selected with similar discriminatory power leading to only a single overlapping CpG (within the TNF locus) between those reported here and by Queiros and colleagues.[5]

## Me-iPLEX method

DNA was isolated using DNeasy column purification (Qiagen) or using the QIAmp FFPE kit (Qiagen) and quantified using Qubit (ThermoFisher). DNA was bisulfite-converted (500ng) using the EZ DNA Methylation kit (Zymo Research) as recommended. For the Me-iPLEX workflow, we followed the standard iPLEX method (Agena Biosciences). Briefly, regions around CpGs of interest are amplified in a multiplexed capture PCR reaction followed by treatment with Shrimp alkaline phosphatase to dephosphorylate unincorporated dNTPs. A multiplexed single base extension PCR is then performed

using mass modified terminator nucleotides and primers that anneal immediately 5' to the CpGs of interest. Following desalting, samples are dispensed onto the SpectroCHIP array for analysis by MADLI-TOF. The ratio of the extension products is reported directly as the percent methylation. Samples were dispensed using the RS1000 nanodispenser and analyzed using the MassARRAY Analyzer4 system in 384 sample chip format. For data quality control, single CpG measurements displaying less than 70% primer extension or combined extended peak areas (methylated+unmethylated) <10 intensity units in the mass spectrum were censored and imputed using the missForest package in R/Bioconductor. Samples missing >6/20 data points were excluded, which ranged from 2.1% – 7.4% per cohort.

### Me-iPLEX primer design

CpGs interrogated by Illumina probes were targeted for Me-iPLEX primer design when possible, and/or adjacent CpGs. Methylation values were averaged per loci. MassARRAY iPLEX capture and extension primers were designed to amplify and target CpGs from bisulfite-converted DNA sequences using the Typer4.0 software (Agena Biosciences). As bisulfite conversion creates unique forward and reverse strand sequences, assays were designed to original top and bottom strands independently. Target capture primer length was increased to 25 bp and product size decreased to 100 bp to better function with bisulfite-converted DNA. Capture primers were designed to overlap >3 non-CpG cytosines to ensure amplification of bisulfite-converted DNA only. As primers were designed to hybridize to thymines at the positions of non-CpG cytosines in the native sequence, primer binding to unconverted DNA was prevented. For extension primer design, CpGs of interest were treated as a C/T SNP. Capture and extension primers were separately combined into a working multiplex according to the Typer software design. To compensate for the decrease in peak intensities as mass increases in MALDI-TOF MS, extension primers were combined targeting a working concentration of $\ln_{(\text{primer mass})}$-7.82.

### Additional Me-iPLEX quality control assessments

To compare the accuracy of Me-iPLEX, we performed MassARRAY EpiTYPER (Agena Biosciences) on the same DNA samples using standard conditions and primers.[2] Due to technical differences in primer design between methods, methylation values were only compared for the 12 CpGs that were able to be individually interrogated by both methods. Me-iPLEX accuracy was also compared to Illumina 450K beta values on the same DNA samples. Methylation values were compared for the 3 CpGs interrogated individually by both methods. To evaluate purity limits, CLL PBMC samples were purified by immunomagnetic CD19-positive selection (Miltenyi Biotech) prior to mixture with non-CLL or normal

PBMC samples. CLL purity level was determined by immunostaining of CD19+/CD5+ cells using a FC500 flow cytometer (BD Biosciences).

### Determination of epitypes using random forest classification

To classify samples into epitypes we used a random forest algorithm modified from Capper et al.[6] For training, we used Me-iPLEX values from 20 genomic loci generated on the training set, fitting 500 decision trees using the RandomForest package in R. To determine class fit of individual samples, we generated a calibrated probability score using class score distribution. Calibration results in similar probability distributions across the methylation classes allowing for cross-class comparison. Calibrated probability scores were generated by recalibrating random forest scores by fitting an L2-penalized, multinomial logistic regression model with the class (HP-CLL, IP-CLL, LP-CLL, and normal PBMC) as the response variable and the score as the explanatory using the glmnet R package. Samples with a calibrated class call lower than 0.9 cannot confidently be assigned to a subtype and are labeled as either insufficient purity if the purity estimate is below 60% (see supplemental Figure 3C) or ambiguous. For validation of the calibrated scores, we generated scores within each of the nested cross-validation loops used to validate the random forest model. We used internal cross-validation and an external validation sample set to evaluate accuracy. For internal cross-validation, known epitype calls were obtained from Illumina 450K classification[2]; the data set was split into three roughly equal groups and the classifier trained on 2/3 and tested with the remaining third and repeated until each sample had served as both training and test samples. Using the full training sample set, the classifier was additionally compared to an independent validation sample set with known epitypes also from Illumina 450K analysis.[7] Data were visualized by t-distributed stochastic neighbor embedding (t-SNE) plots using the Rtsne R package directly to beta values from the Me-iPLEX without dimensionality reduction. R code for the random forest classifier is available in the supplemental text file.

### Biological features of CLL samples

IGHV gene usage and mutations were determined by the source institution. We performed resequencing of IGHV for samples showing a discordance for the common epitype/IGHV correlations (LP-CLL/IGHV-U and HP-CLL/IGHV-M) as performed previously.[2] Immunoglobulin light chain usage was determined from available RNA sequencing data using an established method.[8] Single nucleotide mutations in recurrently mutated genes in CLL were obtained from the source institution for the respective cohort. Samples with data unavailable for recurrent hotspot mutations in *MYD88* (L265), *XPO1* (E571) and *EGR2* (E356, H384, D411, and E412) were interrogated using the traditional

MassARRAY iPLEX assay with standard conditions (Agena Biosciences). Flow cytometry (ZAP70% and CD38%), cytogenetic (FISH), serum protein level, and blood cell count data were obtained from the source institution.

### Statistical analyses

Descriptive statistics were used to summarize patient characteristics. Fisher's exact tests or Kruskal-Wallis tests were used to compare characteristics among epitype groups as appropriate. TTFT was calculated from the date of diagnosis until the date of first treatment or last follow-up. OS was calculated from the date of diagnosis to death, censoring those alive at last follow up. TTP was calculated from the date of first treatment to disease progression, censoring those who have not progressed at last follow up or expired. Kaplan-Meier curves estimated TTFT, TTP, and OS probability and the log-rank tests were used to test for the difference across epitypes. Cox proportional hazard models were used to examine the association between epitype groups and TTFT, TTP, or OS. For the OSU-ibrutinib cohort, the cumulative incidence of discontinuation of treatment was measured from the first date of treatment until the date off-study for CLL progression or Richter's transformation. Gray's test was used to compare differences in the cumulative incidence rates between epitype groups, and Fine and Gray regression models accounting for competing risks were used to examine the association between epitype and risk of discontinuation. Stratified analyses were conducted to examine the associations between epitype and IGHV mutational status or ZAP70 expression. Final multivariable model focused on epitype and its prognostic importance relative to other common clinically important variables in CLL that included age, sex, Rai stage, and presence of del(17p). Due to incomplete del(17p) data at the time of diagnosis in CRC cohort, further multivariable modeling was not performed. The analyses were performed using Stata 14, and the statistical tests were 2-sided with statistical significance defined as $p<0.05$.

## References:


1. Kulis M, Heath S, Bibikova M, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. *Nature Genetics*. 2012;44(11):1236-1242.
2. Oakes CC, Seifert M, Assenov Y, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet*. 2016.
3. Claus R, Lucas DM, Stilgenbauer S, et al. Quantitative DNA Methylation Analysis Identifies a Single CpG Dinucleotide Important for ZAP-70 Expression and Predictive of Prognosis in Chronic Lymphocytic Leukemia. *Journal of Clinical Oncology*. 2012;30(20):2483-2491.
4. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *Plos One*. 2012;7(7):e41361.
5. Queiros AC, Villamor N, Clot G, et al. A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. *Leukemia*. 2014.
6. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature*. 2018;555(7697):469-474.
7. Dietrich S, Oles M, Lu J, et al. Drug-perturbation-based stratification of blood cancer. *J Clin Invest*. 2018;128(1):427-445.
8. Blachly JS, Ruppert AS, Zhao W, et al. Immunoglobulin transcript sequence and somatic hypermutation computation from unselected RNA-seq reads in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A*. 2015;112(14):4322-4327.

# Figure Legends:

**Supplemental Figure 1**: Evaluation of the accuracy and reproducibility of the Me-iPLEX compared to other established methods for measuring DNA methylation at single CpG resolution using Bland-Altman analysis. **(A)** Comparison of 12 CpGs in 192 samples to the EpiTYPER assay. Methylation values were highly correlated ($r^2$=0.90) and varied less than ±20%, with differences enriched at extreme methylation ranges as expected from known methylation bias in the EpiTYPER assay. The number of CpGs compared was limited by the ability to assess the identical CpG in both assays. **(B)** Correlation of Me-iPLEX to beta values obtained from Illumina 450K arrays across 3 CpGs in 153 samples ($r^2$=0.89). As in (A), only identical CpGs assessed in both assays were compared. **(C)** Reproducibility of the Me-iPLEX assay as determined by running the same set of 192 DNA samples (8 CpGs) on the same sample on separate occasions ($r^2$=0.95).

**Supplemental Figure 2**: Epitype classification of training and validation cohorts using Me-iPLEX. **(A)** Comparison of raw forest scores and calibrated probabilities in the training set. While the raw random forest score was 100% accurate, the calibrated probabilities improve the Brier score and log loss of the classifier (lower values represent higher accuracy). **(B)** t-SNE plot showing unsupervised clustering of the combined training and validation sets. Samples from both sets are equally distributed across and within clusters. **(C)** Confusion matrix showing the accuracy of the combined reference cohort (n=305) using calibrated probabilities. **(D)** A similarly high level of classification accuracy is maintained when combining the training and validation sets into the combined reference cohort.

**Supplemental Figure 3**: Technical evaluation of the CLL epitype Me-iPLEX classifier. **(A)** Relationship between CLL sample purity and the calibrated epitype probability scores for each epitype determined by mixing purified CLL and normal PBMC DNA at fixed ratios. Dotted line represents the mean probability of the normal PBMC epitype call across the three subtypes. **(B)** Calibrated epitype probability scores of CLL samples with known purity as determined by flow cytometry. CLL samples were MACS-separated into purified and CLL-depleted fractions or remixed at fixed ratios. **(C)** Generation of a CLL purity estimator by relating the average methylation of the CLL-specific CpGs as determined by Me-iPLEX to known DNA sample. A quadratic curve (dotted line) was fit to all observations and was used to estimate purity in test

samples. **(D)** The epitype probability score with decreasing DNA input in eight representative samples from each epitype. The solid colored line represents the average probability score of 8 samples, surrounding colored region represents the 95% confidence interval, and the black dashed horizontal line marks the 0.9 probability cut off. Samples that failed to meet data quality control thresholds were assigned zero probability.

**Supplemental Figure 4**: Heatmaps of Me-iPLEX methylation values in each cohort. **(A-D)** Epitype and CLL-specific probes are shown in rows and samples in columns. Samples are arranged by Me-iPLEX call. Dark blue; methylated, white; unmethylated.

**Supplemental Figure 5**: Sample quality metrics of epitype classification and relationship to CLL immunophenotype and atypical methylation patterns. **(A)** Distribution of the estimated purity across all samples, determined as shown in supplemental Figure 3C. Increasing levels of impurity were observed with closer proximity to the normal PBMC cluster (indicated in Figure 3B). **(B)** Distribution of calibrated probability scores among CLL samples. Lower probability scores were found at the periphery of the t-SNE clusters. A small cluster of samples with simultaneous low CLL purity and low identity to normal PBMCs is indicated as an 'atypical' cluster. **(C)** Matutes score grading the similarity to a consensus CLL immunophenotypic pattern (5 being most typical) in the MDACC cohort. High-confidence epitypes and ambiguous samples show high Matutes scores, whereas atypical methylation patterns are associated with lower scores. **(D)** Clustering of n=49 non-CLL samples with samples from the four independent test cohorts. Non-CLL samples are indicated as cyan triangles. **(E)** Epitype classification breakdown of the non-CLL samples. The majority (68%) of samples are not classified with a CLL epitype. **(F)** Classification breakdown by non-CLL malignancy type revealing that mantle cell lymphoma commonly display an IP-CLL-like methylation pattern.

**Supplemental Figure 6:** Clinical impact of CLL epitypes combining the cohorts sampled prior to treatment (Mayo, MDACC and CRC). **(A)** Kaplan-Meier analysis of CLL patients separated by epitype for TTFT and **(B)** OS from diagnosis. In both comparisons, all epitypes are universally distinct with the IP-CLL group being statistically different from LP- and HP-CLL epitypes. P-values assessed by log-rank test.

**Supplemental Figure 7:** Relative clinical impact of epitype on OS after separating patients from the OSU-Ibrutinib cohort by IGHV mutation status. **(A)** IGHV-M cases are relatively rare in this cohort (n=35/192 with epitype), however patients classified with the IP-CLL epitype were observed to exhibit a shorter OS than HP-CLL cases. **(B)** Only two HP-CLL patients were found in the IGHV-U subgroup, both patients were alive at last follow-up and have experienced a relatively favorable overall survival for this cohort (approx. 4 and 5 years).

**Supplemental Table 1:** Summary table of CLL patient features within each cohort separated by epitype classification.

**Supplemental Table 2:** Illumina 450K probes and Me-iPLEX primer sequences used to classify CLL samples into epitype subgroups. Each row represents an interrogated CpG and are grouped together by locus if more than one CpG or Me-iPLEX assay was interrogated per locus. Left columns represent Illumina 450K probe annotation of CpGs of interest. Middle columns represent the sequence for the capture and extension primers, rows with two extension primers interrogate the same CpG. The right columns represents average methylation values determined by Me-iPLEX in the training set samples per subtype.

**Supplemental Table 3:** Assessment of Me-iPLEX epityping on DNA derived from FFPE samples. DNA was extracted from sections from bone marrow clots fixed in formaldehyde and embedded in paraffin from the MDACC cohort. DNA was extracted from purified, viably-frozen PBMC samples from the same patient and compared. Table includes the Me-iPLEX calls along with the epitype probability and estimated purity from Me-iPLEX. In samples where confident epitypes were obtained, 7/7 samples resulted in identical epitype calls between the DNA source types (3x HP-CLL 1x IP-CLL, and 3x LP-CLL). Two other samples (classified with high confidence as IP-CLL and LP-CLL with fresh frozen-derived DNA) were unclassifiable in FFPE-derived DNA. One IP-CLL FFPE sample had very low purity (30%), the other had sufficient purity but fell below the confidence threshold (0.80 probability). Me-iPLEX methylation

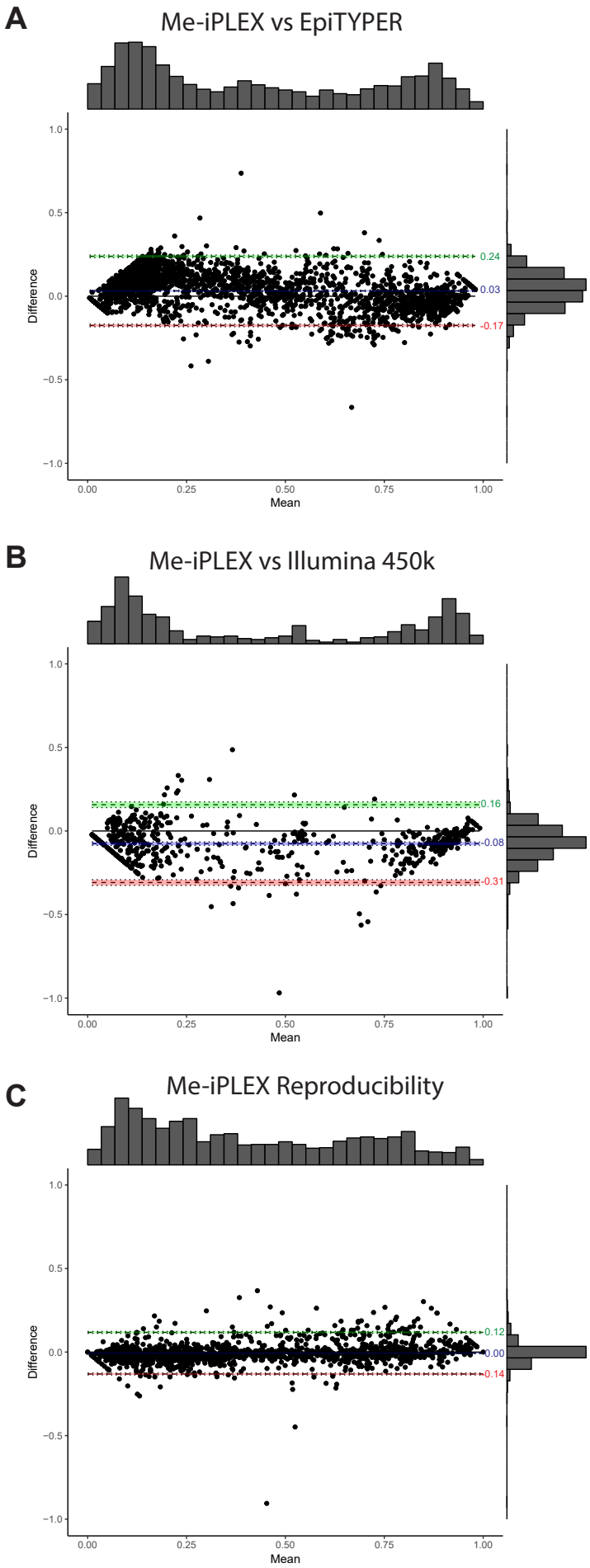
beta values (0=unmethylated and 1=methylated) for the two DNA sample types are shown for all patients.

**Supplemental Table 4:** Table of biological features by epitype combining discovery, validation and test cohorts. Features were not available for all cohorts, interrogated cohorts for each feature are indicated

**Supplemental Table 5:** Univariable and multivariable analyses comparing epitype, IGHV mutation status and ZAP70 positivity for TTFT and OS in the Mayo, CRC and MDACC cohorts, and ibrutinib discontinuation and OS in the OSU-ibrutinib cohort.

**Supplemental Table 6:** Multivariable analyses comparing epitype, for TTFT and OS in the Mayo and MDACC cohorts, and ibrutinib discontinuation and OS in the OSU-ibrutinib cohort.

# Supplemental Figure 1:

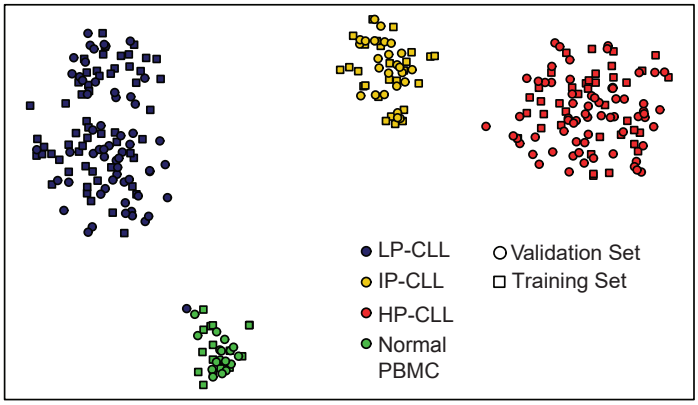


# Supplemental Figure 2:

**A**

| Random Forest Classifier Performance in the Training Set | | | | |
|---|---|---|---|---|
| | AUC | misclassification | Brier score | log loss |
| Raw random forest scores | 1 | 0.000 | 0.064 | 0.178 |
| Calibrated probabilities | 1 | 0.000 | 0.011 | 0.062 |

**B**



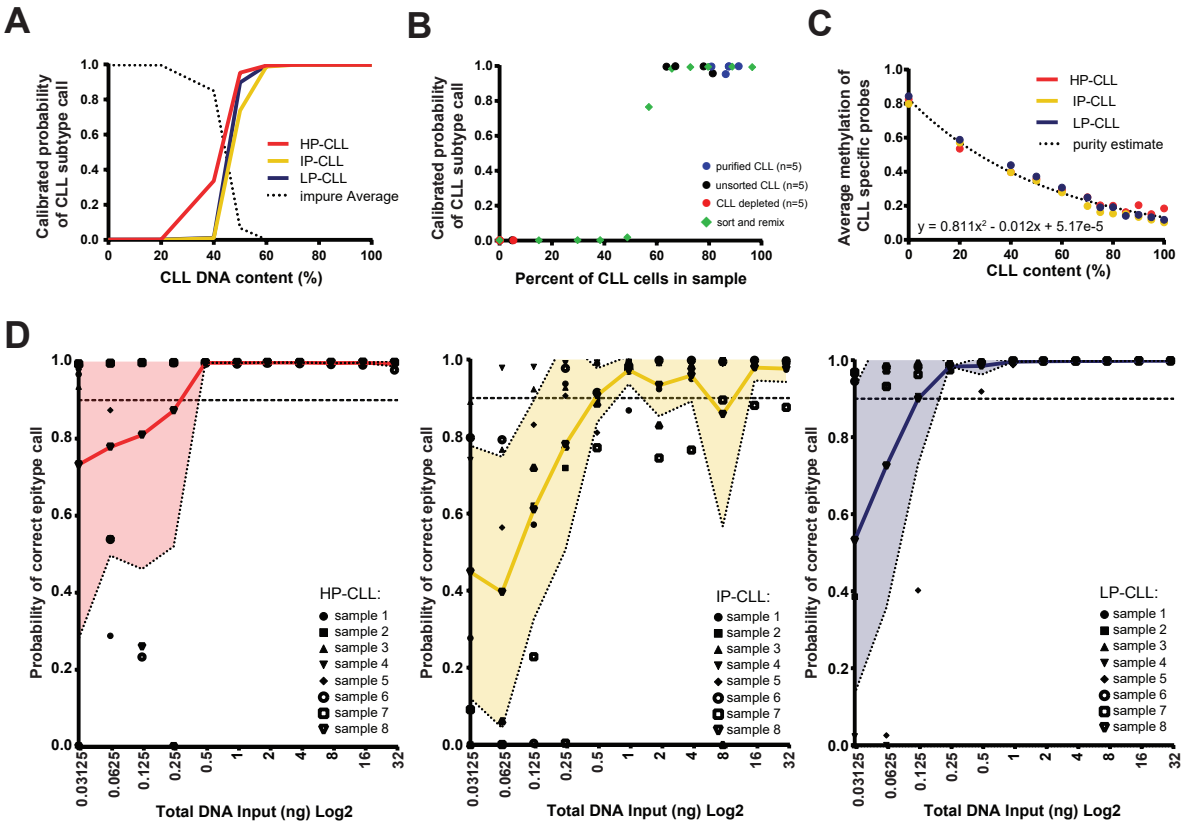

**C**

| Combined Set | | Me-iPLEX Subtype Call | | | |
|---|---|---|---|---|---|
| | | HP-CLL | IP-CLL | LP-CLL | PBMC |
| True Subtype (450K) | HP-CLL | 105 | 0 | 0 | 0 |
| | IP-CLL | 0 | 50 | 1 | 0 |
| | LP-CLL | 0 | 0 | 115 | 0 |
| | PBMC | 0 | 0 | 0 | 34 |

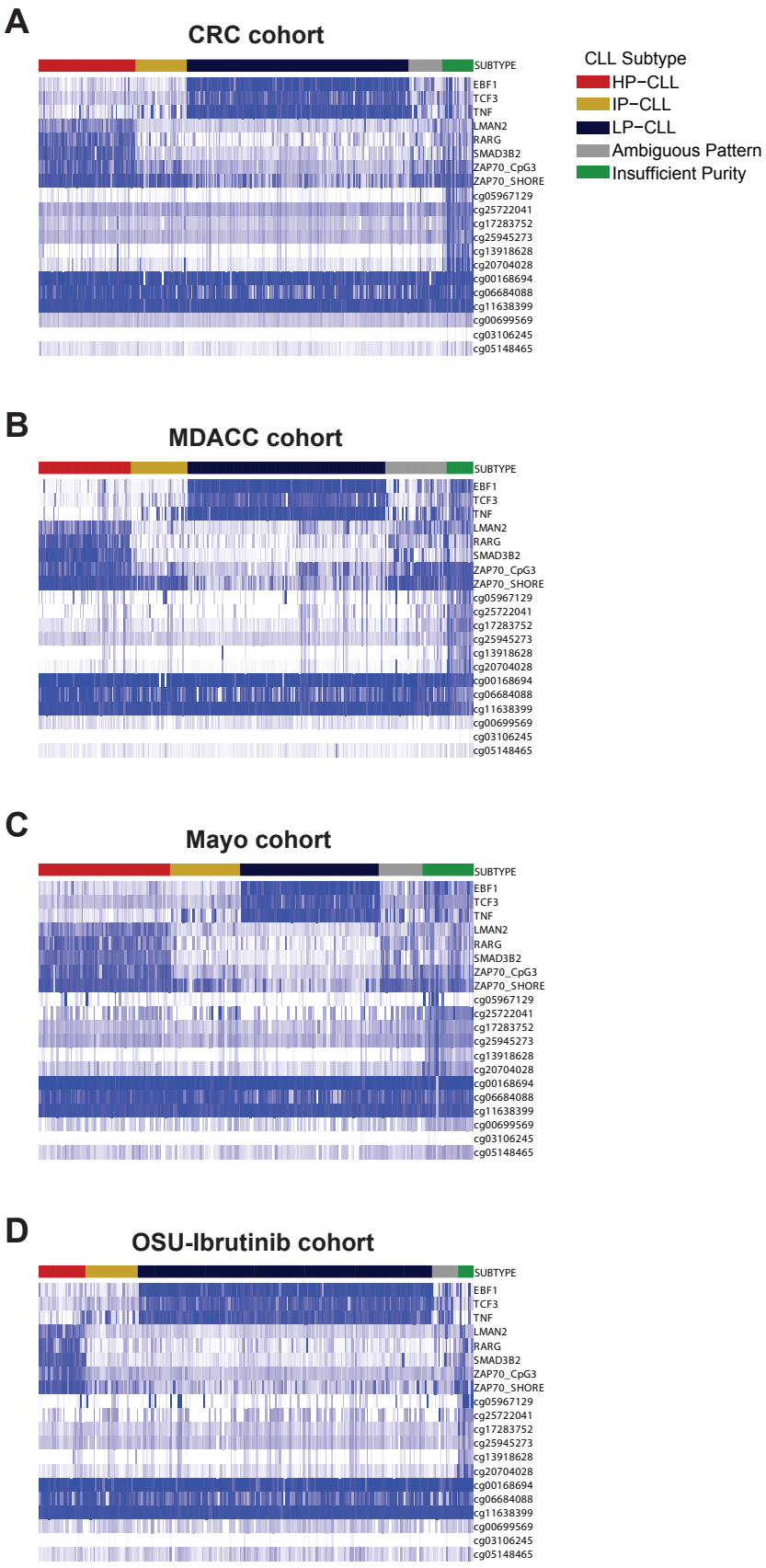
**D**

| Random Forest Classifier Performance in the Combined Cohort | | | | |
|---|---|---|---|---|
| | AUC | misclassification | Brier score | log loss |
| Raw random forest scores | 1 | 0.000 | 0.050 | 0.139 |
| Calibrated probabilities | 1 | 0.003 | 0.013 | 0.055 |

# Supplemental Figure 3:

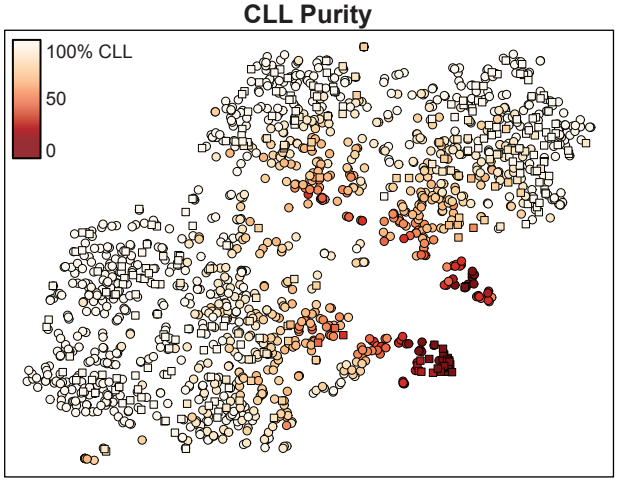


# Supplemental Figure 4:

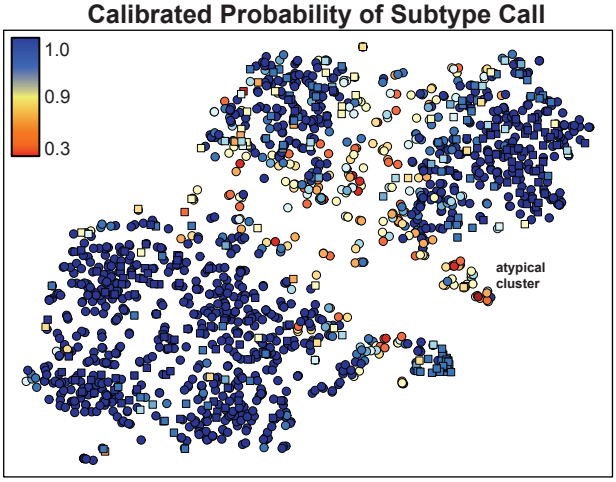


# Suppmemental Figure 5:

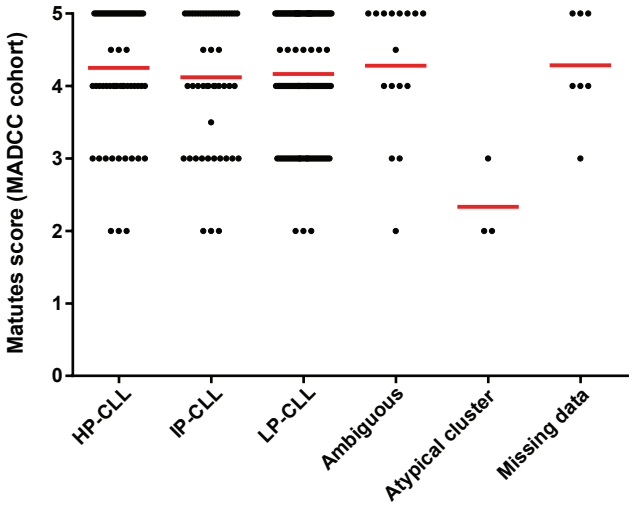
A



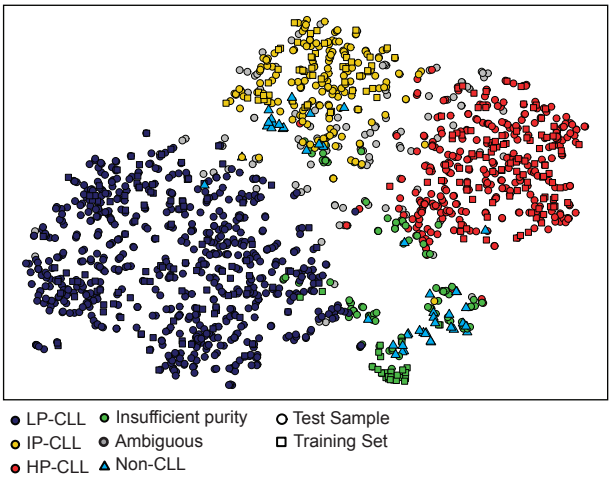

B



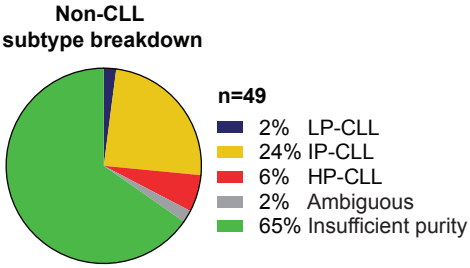

C




D



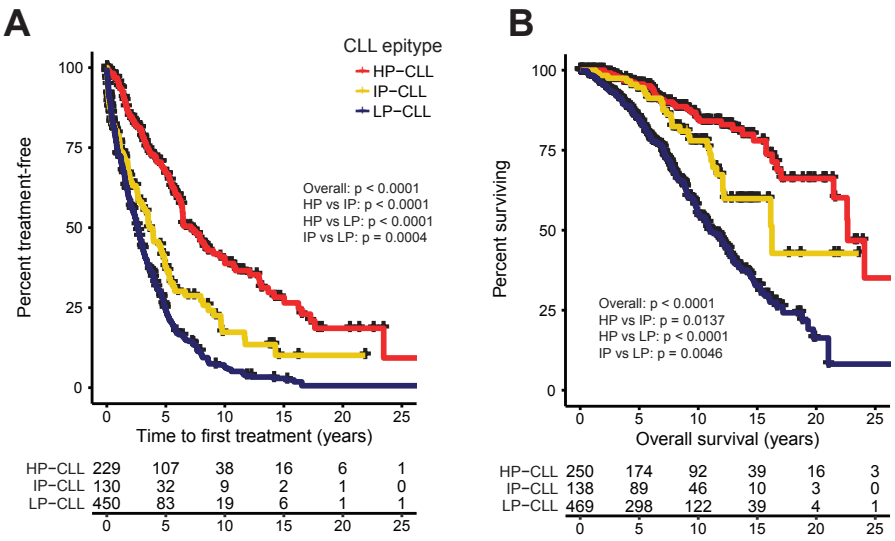

E



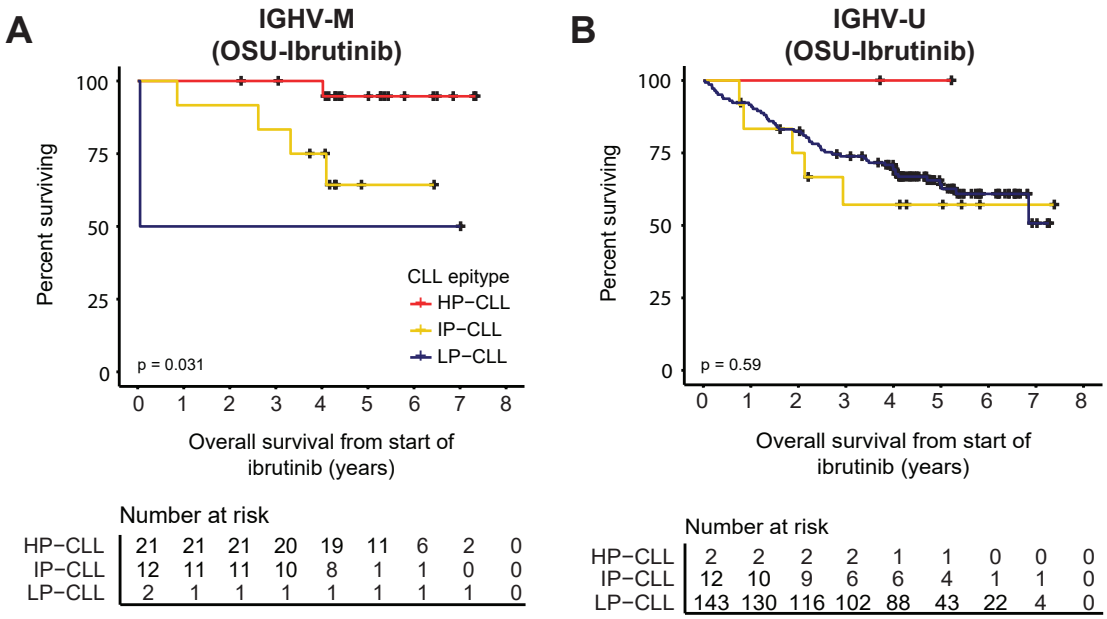

F

| Epitype classification | FL | LPL | MCL | MZL | B-PLL | HCL | B-NHL |
|---|---|---|---|---|---|---|---|
| LP-CLL | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| IP-CLL | 0 | 0 | 9 | 0 | 1 | 0 | 2 |
| HP-CLL | 0 | 0 | 2 | 1 | 0 | 0 | 0 |
| Ambiguous | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Insufficient purity | 2 | 2 | 12 | 5 | 3 | 4 | 4 |

# Supplemental Figure 6:



# Supplemental Figure 7:



**Supplemental Table 1: Patient characteristics per cohort**

| Cohort | | Total (% or range) | LP-CLL (% or range) | IP-CLL (% or range) | HP-CLL (% or range) | Ambiguous (% or range) | *P* -value* |
|---|---|---|---|---|---|---|---|
| **Mayo** | **Sample n** | 248 (100) | 79 (32) | 40 (16) | 75 (30) | 25 (10) | |
| | **Gender (Male)** | 179 (72) | 62 (78) | 30 (75) | 53 (71) | 16 (64) | *0.53* |
| | **Rai (>1)** | 28 (11) | 10 (13) | 11 (28) | 3 (4) | 0 (0) | ***0.0014*** |
| | ***IGHV* (≥98%)** | 110 (44) | 77 (97) | 15 (38) | 7 (9) | 3 (12) | ***<0.0001*** |
| | **ZAP70 (≥20%)** | 95 (39) | 59 (77) | 13 (33) | 6 (8) | 8 (32) | ***<0.0001*** |
| | **del(17p)** | 9 (4) | 8 (10) | 0 (0) | 0 (0) | 1 (4) | ***0.0021*** |
| | **Age at Dx. (median yr)** | 62.5 (35.9 - 90.9) | 62.3 (38.8 - 86.5) | 61.5 (37.4 - 83.1) | 63.0 (42.8 - 90.9) | 66.2 (35.9 - 87) | *0.41* |
| | **Prob. score (median)** | 0.90 (0.31 - 1.0) | 1.0 (0.90 - 1.0) | 0.95 (0.90 - 1.0) | 0.96 (0.91 - 0.99) | 0.68 (0.47 - 0.80) | ***<0.0001*** |
| | **Est. Purity (median %)** | 79 (3.8 - 99.9) | 88 (57.7 - 99.7) | 84 (43.4 - 99.6) | 81 (33.8 - 99.9) | 75.8 (60.1 - 95.4) | ***0.0021*** |
| **MDACC** | **Sample n** | 367 (100) | 167 (46) | 48 (13) | 78 (21) | 52 (14) | |
| | **Gender (Male)** | 259 (71) | 128 (77) | 33 (69) | 44 (56) | 37 (71) | ***0.0055*** |
| | **Rai (>1)** | 224 (61) | 94 (56) | 30 (64) | 45 (58) | 36 (69) | *0.74* |
| | ***IGHV* (≥98%)** | 192 (56) | 156 (97) | 11 (23) | 3 (4) | 13 (29) | ***<0.0001*** |
| | **ZAP70 (≥20%)** | 164 (50) | 115 (76) | 17 (39) | 6 (9) | 17 (39) | ***<0.0001*** |
| | **del(17p)** | 13 (4) | 9 (7) | 1 (2) | 0 (0) | 2 (5) | *0.064* |
| | **Age at Dx. (median yr)** | 56.4 (24 - 85) | 56.9 (27 - 85) | 56.3 (35.7 - 75.5) | 56 (26.7 - 82.4) | 56 (36 - 76) | *0.56* |
| | **Prob. score (median)** | 0.94 (0.38 - 1.0) | 1.0 (0.90 - 1.0) | 0.98 (0.91 - 1.0) | 0.98 (0.92 - 1.0) | 0.75 (0.38 - 0.90) | ***0.0011*** |
| | **Est. Purity (median %)** | 88.0 (2.6 - 100) | 91.5 (47.4 - 99.9) | 93.2 (61.8 - 100) | 91.0 (48.7 - 99.6) | 88.5 (20.6 - 99.9) | *0.55* |
| **OSU-** | **Sample n** | 232 (100) | 157 (68) | 28 (12) | 25 (11) | 14 (6) | |
| **Ibrutinib** | **Gender (Male)** | 162 (69) | 111 (71) | 17 (61) | 14 (56) | 12 (86) | *0.25* |
| | **Rai (>1)** | 180 (76) | 117 (65) | 19 (11) | 20 (11) | 12 (86) | *0.58* |
| | ***IGHV* (≥98%)** | 173 (80) | 143 (99) | 12 (50) | 2 (9) | 9 (64) | ***<0.0001*** |
| | **del(17p)** | 95 (41) | 70 (45) | 8 (29) | 11 (44) | 6 (43) | ***0.28*** |
| | **Age at Dx. (median yr)** | 65.4 (37.3-88.9) | 64.2 (37.3-85.1) | 68.4 (52.1-85.8) | 70.5 (50.9-88.9) | 61.9 (41.3-75.6) | ***0.032*** |
| | **Prob. score (median)** | 0.95 (0.47 - 1.0) | 1.0 (0.90 - 1.0) | 0.97 (0.90 - 0.99) | 0.97 (0.90 - 0.99) | 0.67 (0.49 - 0.85) | *0.072* |
| | **Est. Purity (median %)** | 83.8 (2.3 - 100) | 85.0 (36.6 - 100) | 86.9 (52.7 - 99.6) | 89.9 (48.1 - 99.6) | 85.9 (64.7 - 99.8) | ***0.018*** |
| **CRC** | **Sample n** | 439 (100) | 224 (51) | 52 (12) | 98 (22) | 34 (8) | |
| | **Gender (Male)** | 295 (70) | 154 (70) | 43 (84) | 55 (58) | 26 (79) | ***0.0034*** |
| | **Rai (>1)** | 38 (16) | 27 (23) | 4 (15) | 4 (7) | 0 (0) | *0.071* |
| | ***IGHV* (≥98%)** | 274 (62) | 214 (96) | 13 (25) | 14 (14) | 16 (47) | ***<0.0001*** |
| | **ZAP70 (≥20%)** | 227 (52) | 130 (58) | 27 (52) | 46 (47) | 11 (32) | *0.17* |
| | **del(17p)** | 20 (13) | 12 (15) | 1 (7) | 3 (8) | 1 (9) | *0.61* |
| | **Age at Dx. (median yr)** | 56.4 (24.7 - 83.8) | 55.5 (24.7 - 83.6) | 56.6 (28.8 - 81) | 56.5 (33.5 - 83.8) | 58.7 (31.2 - 82.9) | *0.69* |
| | **Prob. score (median)** | 0.93 (0.40 - 1.0) | 1.0 (0.90 - 1.0) | 0.95 (0.90 - 0.99) | 0.96 (0.90 - 0.99) | 0.67 (0.49 - 0.80) | ***<0.0001*** |
| | **Est. Purity (median %)** | 80.0 (3.5 - 99.9) | 86.1 (52.5 - 99.9) | 81.6 (43.1 - 98.7) | 80.7 (35.3 - 99.8) | 80.3 (30.3 - 99.8) | ***<0.0001*** |

***P* -values calculated across LP-, IP- and HP-CLL epitypes only using 3x2 Fisher's exact test or 3x2 Kruskal-Wallis test where appropriate**

**Dx.; diagnosis**

**Supplemental Table 2: Illumina 450K probes and Me-iPLEX primer sequences**

| | Target CpG Annotation | | | | Me-iPLEX Probes | | | | Epitype CpG Methylation (Avg %Me) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Gene / Probe | Chr | Position (hg19) | RefGene Group | Capture primer Forward | Capture Primer Reverse | Extension Probes (Forward) | Extension Probes (Reverse) | HP | IP | LP | PBMC |
| Epitype Assignment | **EBF1 (cg11181763)** | 5 | 158379078 | Intronic | ACGTTGGATGGTTTGGAGTTTAGGTTGTTTTGTG | ACGTTGGATGCCCACCAACTTTTATTAACAAAAAC | TTTAGGTTGTTTTGTGAATATTAT<br>TAAAAATCATATCTCCTAACCAC<br>GGATGAAAAATTTTTGAGTGTT | | 14% | 11% | 84% | 85% |
| | **TCF3 (cg26615224)** | 19 | 1621124 | Intronic | ACGTTGGATGACTCTCCCTAACACAACAATATC | ACGTTGGATGTGTTTTATTGAGGTTAGTGTTTTG | ACCACAACTAACAACTCC<br>GAGTTGTTAGTTGTGGTTTTT<br>TTCACCCACAAACTCCC<br>TGAAGAGAGGTGAGGTTTA | CCTATAAATAAAAAAAAAATAAAACCCAC<br>GGTTTTATTTTTTTTTATTTATAGGTTTT<br>ACTACCAACTATAATCCCTC<br>AGGGATTATAGTTGGTAGTTT | 19% | 22% | 63% | 71% |
| | **TNF (cg09637172)** | 6 | 31545252 | Gene Body | ACGTTGGATGGAGGGGTTTTTTAGTTGGAGAAG | ACGTTGGATGCCTCCTCACAAAACAATAATCCC | AGTTGGAGAAGGGTGAT<br>CAAAATAAACCTACCCAAACTC | | 10% | 31% | 79% | 52% |
| | **LMAN2 (cg25662041)** | 5 | 176759170 | Gene Body | ACGTTGGATGAGGAGGGTAGTGGGGATGGAATA | ACGTTGGATGTAAACTTCACCTAACCTCTACCC | GGGGATGGAATAGGATAGT<br>AACCTCTACCCCTCCCC | gAACCCAAAAACAAACCCC<br>TCATCTTATTATTCTTTTATAATCCC | 50% | 23% | 21% | 38% |
| | **RARG (cg13940444)** | 12 | 53617383 | Intronic | ACGTTGGATGTGAGAGTGGGGTTTAAAGTAGGG | ACGTTGGATGAAACTACACCCCCCTCTCCTAAC | AAAGTAGGGATGTGAGG<br>CCCCCATCCTACCCAAC | | 61% | 13% | 13% | 30% |
| | **SMAD3B2 (cg19193595)** | 15 | 67396488 | Intronic | ACGTTGGATGTGGTGGGGAAAAGTTTTTGGTTG | ACGTTGGATGCCCTATTTTTAAAAATACCTACCTC | TTTGGTTGTTAAAAGTTGGGGG<br>TCAAAAACAAAAAAACAAATAAAATC<br>ATTTTATTTGTTTTTTTGTTTTTGAG | | 71% | 13% | 11% | 26% |
| | **ZAP70** | 2 | 98330254 | Intronic | ACGTTGGATGAAGTTTTTAGTTTGAGAGGGAGG | ACGTTGGATGCCCCCCCATCCATAAATAAAAAAC | TGAGAGGGAGGTTTTGT<br>GGGGGAGGATGTTATATTT<br>ATCCATAAATAAAAAACCCTAAC | | 60% | 37% | 31% | 40% |
| | | 2 | 98330415 | Gene Body | ACGTTGGATGGGATGTATAGGTGGATGTTTAGG | ACGTTGGATGTTATAAACCCCAATTCCTCACCC | ATGTTTAGGTAAATAGGTTATTTG<br>gCATTAAAACCAAAAACCCC<br>aaCAATTCCTCACCCAAAACC | | 74% | 52% | 31% | 43% |
| Purity Estimation | **SND1 (cg05967129)** | 7 | 127544717 | Intronic | ACGTTGGATGATAATGGGAAGGGGTAGGAG | ACGTTGGATGTCATACATCCCCACTTTAAC | GGTGTAAAAATGAAGAGAGAAG | | 5% | 3% | 3% | 80% |
| | **RERE (cg25722041)** | 1 | 8623473 | Intronic | ACGTTGGATGCCTACAAACCAAACAAAAC | ACGTTGGATGATTAGTTGGTGATTAGTGGG | AAAACAACTATAAATTCACCTAC | | 17% | 18% | 16% | 70% |
| | **CDYL (cg17283752)** | 6 | 4909143 | Intronic | ACGTTGGATGCCTCTCCTATACATAATCAC | ACGTTGGATGTTTTTTGTGGTTTAATTAGG | cTCTACCTATCTATTAAAAAACTCC | | 18% | 16% | 16% | 70% |
| | **CLEC17A (cg25945273)** | 19 | 14692863 | TSS | ACGTTGGATGCAATCTCACTACTAAAACCC | ACGTTGGATGGTAGAGGTTGGAAATAGTAAG | AACCATACTTCTTTAAAAAACTACAAC | | 24% | 22% | 22% | 69% |
| | **CD72 (cg13918628)** | 9 | 35610380 | 3'UTR (Intronic) | ACGTTGGATGTTTATGGGAGGTTGGTGGTG | ACGTTGGATGACTAAACCCTCATCCCACTC | GTTGGTGGTGTGGTTGT | | 3% | 3% | 1% | 70% |
| | **INPP5F (cg20704028 )** | 10 | 121493706 | Intronic | ACGTTGGATGCAACAACAACCTTCTCTACAC | ACGTTGGATGGGGATTTATAGATATTTTTG | ACCTTCTCTACACAAAAACC | | 10% | 11% | 7% | 74% |
| | **ETS2 (cg00168694)** | 21 | 40193056 | Intronic | ACGTTGGATGAACCAAACTCACATTAAAC | ACGTTGGATGGGGGTGTTTTATTATAGA | CAAAAAAATAACATAAAATCCCAC | | 92% | 89% | 90% | 75% |
| | **cg06684088** | 7 | 119131 | Intergenic | ACGTTGGATGCCCCTCCCTACCACTAAAAA | ACGTTGGATGGTTATTATGTGGTAATTGGG | TCTTTAAACTCCCTACTACAAC | | 76% | 76% | 58% | 56% |
| | **cg11638399** | 8 | 29441416 | Intergenic | ACGTTGGATGACCCAAACAAACCCTCATTC | ACGTTGGATGGTAAGATATTTTTTTTAAGG | TCATTCACATCAAAACAAAAC | | 88% | 89% | 89% | 67% |
| | **cg00699569** | 1 | 157536296 | Intergenic | ACGTTGGATGCCTTTAATAACACCTTTTTTC | ACGTTGGATGTGGTTATATGATTTGTGGG | ACAAAAACCTTACATTTTCTTTTTAC | | 17% | 17% | 18% | 41% |
| | **SPI1 (cg03106245)** | 11 | 47399980 | Gene Body | ACGTTGGATGGTAGAGTTTTTTAGGATGGG | ACGTTGGATGCAAACCTAAACCCTACACCC | TTTAGGATGGGGTGTTT | | 0% | 0% | 0% | 23% |
| | **cg05148465** | 17 | 7034129 | Intergenic | ACGTTGGATGCCAAATTCCCAACCAAAAAC | ACGTTGGATGTTTGAGTTTTAGAGGAAGG | AATACTTCTACTCCCATCC | | 12% | 10% | 11% | 47% |

**Supplemental Table 3: Me-iPLEX classification and methylation beta values using FFPE- and PBMC-derived DNA**

| | | | | | Me-iPLEX probe methylation values | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Patient ID | DNA type | Epitype call | Call Probability | Purity (%) | EBF1-cg11181763 | TCF3-cg26615224 | TNF-cg09637172 | LMAN2-cg25662041 | RARG-cg13940444 | SMAD3B2-cg19193595 | ZAP70-CpG3 | ZAP70 | SND1-cg05967129 | RERE-cg25722041 | CDYL-cg17283752 | CLEC17A-cg25945273 | CD72-cg13918628 | INPP5F-cg20704028 | ETS2-cg00168694 | cg06684088 | cg11638399 | cg00699569 | SPI1-cg03106245 | cg05148465 |
| FCR098 | FFPE | HP-CLL | 1.00 | 85.7 | 0.19 | 0.22 | 0.16 | 0.65 | 0.49 | 0.70 | 0.73 | 0.84 | 0.25 | 0.12 | 0.22 | 0.25 | 0.07 | 0.13 | 0.72 | 0.74 | 0.78 | 0.09 | 0.00 | 0.10 |
| FCR098 | PBMC | HP-CLL | 0.99 | 99.6 | 0.04 | 0.05 | 0.00 | 0.55 | 0.54 | 0.75 | 0.80 | 0.84 | 0.00 | 0.00 | 0.08 | 0.15 | 0.21 | 0.03 | 1.00 | 0.44 | 0.92 | 0.11 | 0.00 | 0.05 |
| FCR099 | FFPE | insufficient purity | 0.42 | 31.3 | 0.35 | 0.41 | 0.71 | 0.57 | 0.28 | 0.13 | 0.65 | 0.78 | 0.37 | 0.37 | 0.40 | 0.36 | 0.72 | 0.54 | 0.70 | 0.48 | 0.63 | 0.06 | 0.00 | 0.04 |
| FCR099 | PBMC | IP-CLL | 0.97 | 97.5 | 0.07 | 0.04 | 0.71 | 0.30 | 0.33 | 0.09 | 0.50 | 0.86 | 0.00 | 0.00 | 0.10 | 0.16 | 0.00 | 0.02 | 1.00 | 0.76 | 0.91 | 0.08 | 0.00 | 0.06 |
| FCR110 | FFPE | LP-CLL | 0.93 | 98.6 | 0.89 | 0.48 | 0.53 | 0.59 | 0.00 | 0.02 | 0.60 | 0.51 | 0.00 | 0.00 | 0.03 | 0.16 | 0.00 | 0.01 | 0.87 | 0.64 | 0.86 | 0.08 | 0.00 | 0.04 |
| FCR110 | PBMC | LP-CLL | 0.97 | 77.2 | 1.00 | 0.72 | 0.57 | 0.16 | 0.00 | 0.00 | 0.11 | 0.36 | 0.00 | 0.00 | 0.11 | 0.26 | 0.00 | 0.00 | 1.00 | 0.78 | 0.85 | 0.00 | 0.00 | 0.21 |
| FCR135 | FFPE | HP-CLL | 0.99 | 72.7 | 0.47 | 0.29 | 0.23 | 0.38 | 0.56 | 0.48 | 0.67 | 0.76 | 0.16 | 0.34 | 0.33 | 0.32 | 0.14 | 0.15 | 0.66 | 0.34 | 0.66 | 0.00 | 0.00 | 0.06 |
| FCR135 | PBMC | HP-CLL | 0.99 | 97.6 | 0.05 | 0.05 | 0.00 | 0.51 | 0.71 | 0.72 | 0.76 | 0.82 | 0.00 | 0.00 | 0.07 | 0.25 | 0.00 | 0.00 | 0.80 | 0.19 | 0.89 | 0.04 | 0.00 | 0.09 |
| FCR136 | FFPE | IP-CLL | 0.91 | 95.5 | 0.22 | 0.12 | 0.31 | 0.79 | 0.13 | 0.00 | 0.77 | 0.79 | 0.00 | 0.20 | 0.30 | 0.21 | 0.00 | 0.10 | 0.85 | 0.56 | 0.93 | 0.00 | 0.00 | 0.02 |
| FCR136 | PBMC | IP-CLL | 0.99 | 99.5 | 0.03 | 0.02 | 0.21 | 0.26 | 0.03 | 0.05 | 0.31 | 0.86 | 0.00 | 0.00 | 0.12 | 0.23 | 0.00 | 0.02 | 1.00 | 0.85 | 0.87 | 0.22 | 0.00 | 0.07 |
| FCR148 | FFPE | LP-CLL | 1.00 | 78.3 | 0.63 | 0.63 | 0.80 | 0.23 | 0.07 | 0.12 | 0.51 | 0.39 | 0.22 | 0.26 | 0.29 | 0.25 | 0.12 | 0.18 | 0.80 | 0.40 | 0.68 | 0.10 | 0.00 | 0.09 |
| FCR148 | PBMC | LP-CLL | 0.97 | 96.2 | 0.46 | 0.66 | 0.92 | 0.13 | 0.01 | 0.05 | 0.23 | 0.20 | 0.00 | 0.00 | 0.10 | 0.19 | 0.00 | 0.03 | 1.00 | 0.64 | 0.89 | 0.15 | 0.00 | 0.05 |
| FCR207 | FFPE | HP-CLL | 1.00 | 86.5 | 0.15 | 0.30 | 0.30 | 0.53 | 0.51 | 0.63 | 0.76 | 0.74 | 0.22 | 0.24 | 0.18 | 0.24 | 0.00 | 0.17 | 0.84 | 0.75 | 0.74 | 0.00 | 0.00 | 0.12 |
| FCR207 | PBMC | HP-CLL | 0.99 | 91.3 | 0.05 | 0.04 | 0.13 | 0.54 | 0.55 | 0.65 | 0.83 | 0.90 | 0.10 | 0.00 | 0.06 | 0.18 | 0.00 | 0.04 | 0.86 | 0.56 | 0.91 | 0.04 | 0.00 | 0.09 |
| FCR211 | FFPE | ambiguous | 0.80 | 99.3 | 0.84 | 0.16 | 0.78 | 0.50 | 0.08 | 0.37 | 0.64 | 0.55 | 0.00 | 0.00 | 0.24 | 0.27 | 0.10 | 0.13 | 0.40 | 0.10 | 0.73 | 0.00 | 0.00 | 0.02 |
| FCR211 | PBMC | LP-CLL | 0.95 | 98.0 | 0.91 | 0.09 | 0.88 | 0.29 | 0.02 | 0.29 | 0.31 | 0.38 | 0.00 | 0.00 | 0.04 | 0.17 | 0.00 | 0.04 | 1.00 | 0.81 | 0.91 | 0.00 | 0.00 | 0.04 |
| FCR240 | FFPE | LP-CLL | 0.98 | 64.7 | 0.92 | 0.74 | 0.88 | 0.77 | 0.12 | 0.12 | 0.68 | 0.47 | 0.37 | 0.41 | 0.22 | 0.38 | 0.21 | 0.10 | 0.62 | 0.21 | 0.50 | 0.00 | 0.00 | 0.12 |
| FCR240 | PBMC | LP-CLL | 0.99 | 93.6 | 0.93 | 0.79 | 0.89 | 0.21 | 0.02 | 0.08 | 0.23 | 0.20 | 0.00 | 0.00 | 0.06 | 0.11 | 0.00 | 0.03 | 1.00 | 0.42 | 0.93 | 0.07 | 0.00 | 0.04 |

**Supplemental Table 4: Biological characteristics of epitypes using a combined analysis of all cohorts**

| Feature | Cohorts | Total (n) | LP-CLL (% or range) | IP-CLL (% or range) | HP-CLL (% or range) | *P* -value* |
|---|---|---|---|---|---|---|
| **IGHV% mut. (median)** | Ma,Md,O,C,T,V | 1095 | 100.0 (81.5 - 100) | 96.2 (72.7 - 100) | 92.7 (73.6 - 100) | ***<0.0001*** |
| **IGHV usage:** | | | | | | |
| **VH3 (all)** | Ma,Md,O,C,T,V | 1272 | 201 (29) | 141 (65) | 176 (49) | ***<0.0001*** |
| **VH1-2** | | | 26 (4) | 3 (1) | 7 (2) | ***n.s.*** |
| **VH1-69** | | | 254 (36) | 5 (2) | 11 (3) | ***<0.0001*** |
| **VH2-5** | | | 8 (1) | 3 (2) | 15 (5) | ***0.049*** |
| **VH3-11** | | | 34 (5) | 5 (3) | 1 (0) | ***0.0031*** |
| **VH3-15** | | | 3 (0) | 7 (4) | 20 (5) | ***<0.0001*** |
| **VH3-21** | | | 18 (2) | 33 (14) | 4 (1) | ***<0.0001*** |
| **VH3-23** | | | 8 (1) | 34 (18) | 32 (10) | ***<0.0001*** |
| **VH3-30** | | | 48 (7) | 12 (8) | 25 (8) | ***<0.0001*** |
| **VH3-33** | | | 15 (2) | 3 (2) | 8 (2) | ***n.s.*** |
| **VH3-48** | | | 18 (3) | 11 (3) | 7 (2) | ***n.s.*** |
| **VH3-7** | | | 3 (0) | 6 (3) | 22 (9) | ***<0.0001*** |
| **VH4-34** | | | 28 (5) | 11 (5) | 45 (13) | ***<0.0001*** |
| **VH4-39** | | | 39 (5) | 4 (1) | 13 (4) | ***n.s.*** |
| **VH4-4** | | | 9 (2) | 2 (1) | 8 (2) | ***n.s.*** |
| **VH4-59** | | | 13 (2) | 14 (5) | 19 (6) | ***0.014*** |
| **VH5-51** | | | 19 (3) | 4 (2) | 8 (2) | ***n.s.*** |
| **IGVK/L (Lambda)** | Md,O,C,V | 887 | 151 (28) | 93 (67) | 78 (36) | ***<0.0001*** |
| **VL3-21** | O,V | 286 | 5 (3) | 24 (52) | 3 (4) | ***<0.0001*** |
| **Cytometry:** | | | | | | |
| **ZAP70% (median)** | Md,C | 563 | 27.9 (0 - 97) | 12.2 (0.4 - 79.5) | 5.0 (0.1 - 90.7) | ***<0.0001*** |
| **CD38% (median)** | Md,C | 588 | 23.6 (0 - 99.7) | 8.4 (0.1 - 99.6) | 3.8 (0 - 99.1) | ***<0.0001*** |
| **Cytogenetics (FISH):** | | | | | | |
| **del17p** | Ma,Md,O,C,T,V | 858 | 56 (14) | 8 (5) | 9 (3) | ***0.0019*** |
| **del11q** | Md,O,C,T,V | 657 | 79 (24) | 17 (14) | 3 (1) | ***<0.0001*** |
| **trisomy12** | Md,O,C,T,V | 662 | 60 (18) | 14 (11) | 22 (11) | ***n.s.*** |
| **del13q** | Md,O,C,T,V | 664 | 150 (45) | 86 (70) | 124 (60) | ***<0.0001*** |
| **Mutations:** | | | | | | |
| ***NOTCH1*** | Ma,Md,O,C,T,V | 1256 | 102 (15) | 13 (6) | 12 (3) | ***<0.0001*** |
| ***SF3B1*** | O,C,T,V | 484 | 46 (17) | 25 (29) | 6 (5) | ***<0.0001*** |
| ***TP53*** | O,C,T,V | 651 | 90 (27) | 20 (16) | 19 (10) | ***<0.0001*** |
| ***XPO1*** | Ma,Md,O,C,T,V | 1176 | 72 (11) | 2 (1) | 2 (1) | ***<0.0001*** |
| ***EGR2*** | Ma,Md,O,C,T,V | 1248 | 39 (6) | 9 (4) | 6 (2) | ***<0.0001*** |
| ***MYD88*** | Ma,Md,O,C,T,V | 1270 | 1 (0) | 21 (10) | 14 (4) | ***<0.0001*** |
| **Blood & serum:** | | | | | | |
| **B2M mg/L (median)** | Md,C | 327 | 3.3 (1.5 - 11.1) | 2.9 (1.1 - 8.3) | 2.7 (1.3 - 9.9) | ***<0.0001*** |
| **Hgb g/dL (median)** | Md,C | 379 | 13.0 (6.1 - 17.7) | 13.4 (7.5 - 16.2) | 13.4 (4.2 - 17) | ***n.s.*** |
| **LDH U/L (median)** | Md,C | 308 | 528.5 (120 - 1818) | 476.0 (130 - 1421) | 459.0 (110 - 912) | ***0.00011*** |
| **PLT $10^9$/L (median)** | Md,C | 385 | 172.0 (6 - 476) | 160.0 (60 - 363) | 167.0 (47 - 338) | ***n.s.*** |
| **WBC $10^9$/L (median)** | Md,C | 382 | 79.4 (9.4 - 537) | 44.0 (10 - 333.9) | 43.7 (9.5 - 257.2) | ***<0.0001*** |

******P* **-values calculated using 3x2 Fisher's exact test or 3x2 Kruskal-Wallis test where appropriate**

**Cohorts: Ma; Mayo, Md; MDACC, O; OSU-ibrutinib, C; CRC, T; Training, V; Validation**

**n.s.; non-siginificant**

**Supplemental Table 5: Univariate and multivariate analyses comparing *IGHV* mutation status, ZAP70 positivity and Epitype**

| Cohort | | HR (95% CI) | P-value | | HR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Mayo | Time to First Treatment | | | Overall Survival | | |
| | Univariable modeling | | | Univariable modeling | | |
| | Epitypes | | <0.001 | Epitypes | | 0.030 |
| | HP-CLL vs. IP-CLL | 0.28 (0.14-0.57) | <0.001 | HP-CLL vs. IP-CLL | 1.01 (0.26-3.92) | 0.991 |
| | HP-CLL vs. LP-CLL | 0.20 (0.11-0.36) | <0.001 | HP-CLL vs. LP-CLL | 0.35 (0.14-0.85) | 0.021 |
| | IP-CLL vs. LP-CLL | 0.70 (0.40-1.24) | 0.226 | IP-CLL vs. LP-CLL | 0.34 (0.10-1.19) | 0.091 |
| | IGHV (mutated vs. unmutated) | 0.38 (0.23-0.61) | <0.001 | IGHV (mutated vs. unmutated) | 0.71 (0.33-1.57) | 0.401 |
| | Zap70 (positive vs. negative) | 2.25 (1.41-3.59) | 0.001 | Zap70 (positive vs. negative) | 1.64 (0.75-3.60) | 0.218 |
| | Multivariable modeling | | | Multivariable modeling | | |
| | Epitypes | | <0.001 | Epitypes | | 0.096 |
| | HP-CLL vs. IP-CLL | 0.32 (0.15-0.68) | 0.003 | HP-CLL vs. IP-CLL | 0.93 (0.23-3.81) | 0.917 |
| | HP-CLL vs. LP-CLL | 0.23 (0.11-0.47) | <0.001 | HP-CLL vs. LP-CLL | 0.31 (0.09-1.03) | 0.056 |
| | IP-CLL vs. LP-CLL | 0.70 (0.38-1.29) | 0.257 | IP-CLL vs. LP-CLL | 0.34 (0.09-1.27) | 0.108 |
| | IGHV (mutated vs. unmutated) | 0.63 (0.36-1.09) | 0.102 | IGHV (mutated vs. unmutated) | 1.04 (0.42-2.59) | 0.931 |
| | Zap70 (positive vs. negative) | 0.86 (0.48-1.53) | 0.603 | Zap70 (Positive vs. Negative) | 0.82 (0.29-2.32) | 0.715 |
| CRC | Time to First Treatment | | | Overall Survival | | |
| | Univariable modeling | | | Univariable modeling | | |
| | Epitypes | | <0.001 | Epitypes | | <0.001 |
| | HP-CLL vs. IP-CLL | 0.51 (0.31-0.86) | 0.011 | HP-CLL vs. IP-CLL | 0.60 (0.24-1.47) | 0.265 |
| | HP-CLL vs. LP-CLL | 0.28 (0.20-0.41) | <0.001 | HP-CLL vs. LP-CLL | 0.24 (0.13-0.43) | <0.001 |
| | IP-CLL vs. LP-CLL | 0.56 (0.36-0.85) | 0.007 | IP-CLL vs. LP-CLL | 0.39 (0.19-0.81) | 0.011 |
| | IGHV (mutated vs. unmutated) | 0.42 (0.31-0.57) | <0.001 | IGHV (mutated vs. unmutated) | 0.35 (0.21-0.57) | <0.001 |
| | Zap70 (Positive vs. Negative) | 1.43 (1.10-1.86) | 0.007 | Zap70 (Positive vs. Negative) | 1.54 (1.04-2.28) | 0.033 |
| | Multivariable modeling | | | Multivariable modeling | | |
| | Epitypes | | <0.001 | Epitypes | | 0.015 |
| | HP-CLL vs. IP-CLL | 0.52 (0.31-0.88) | 0.015 | HP-CLL vs. IP-CLL | 0.61 (0.24-1.56) | 0.301 |
| | HP-CLL vs. LP-CLL | 0.37 (0.22-0.62) | <0.001 | HP-CLL vs. LP-CLL | 0.31 (0.13-0.73) | 0.008 |
| | IP-CLL vs. LP-CLL | 0.71 (0.43-1.18) | 0.184 | IP-CLL vs. LP-CLL | 0.51 (0.23-1.16) | 0.107 |
| | IGHV (mutated vs. unmutated) | 0.73 (0.46-1.15) | 0.176 | IGHV (mutated vs. unmutated) | 0.67 (0.32-1.41) | 0.296 |
| | Zap70 (Positive vs. Negative) | 1.41 (1.07-1.86) | 0.014 | Zap70 (Positive vs. Negative) | 1.40 (0.93-2.10) | 0.106 |
| MDACC | Time to Progression | | | Overall Survival | | |
| | Univariable modeling | | | Univariable modeling | | |
| | Epitypes | | 0.002 | Epitypes | | <0.001 |
| | HP-CLL vs. IP-CLL | 0.71 (0.26-1.94) | 0.500 | HP-CLL vs. IP-CLL | 0.44 (0.22-0.88) | 0.021 |
| | HP-CLL vs. LP-CLL | 0.33 (0.14-0.76) | 0.010 | HP-CLL vs. LP-CLL | 0.25 (0.14-0.43) | <0.001 |
| | IP-CLL vs. LP-CLL | 0.47 (0.24-0.91) | 0.025 | IP-CLL vs. LP-CLL | 0.55 (0.33-0.94) | 0.028 |
| | IGHV (mutated vs. unmutated) | 0.40 (0.23-0.71) | 0.002 | IGHV (mutated vs. unmutated) | 0.34 (0.22-0.52) | <0.001 |
| | Zap70 (Positive vs. Negative) | 1.81 (1.10-2.97) | 0.019 | Zap70 (Positive vs. Negative) | 2.56 (1.69-3.88) | <0.001 |
| | Multivariable modeling | | | Multivariable modeling | | |
| | Epitypes | | 0.621 | Epitypes | | 0.239 |
| | HP-CLL vs. IP-CLL | 0.88 (0.29-2.64) | 0.813 | HP-CLL vs. IP-CLL | 0.50 (0.23-1.12) | 0.092 |
| | HP-CLL vs. LP-CLL | 0.56 (0.14-2.17) | 0.402 | HP-CLL vs. LP-CLL | 0.50 (0.17-4.16) | 0.205 |
| | IP-CLL vs. LP-CLL | 0.64 (0.24-1.73) | 0.379 | IP-CLL vs. LP-CLL | 0.99 (0.44-2.23) | 0.983 |
| | IGHV (mutated vs. unmutated) | 0.74 (0.26-2.11) | 0.572 | IGHV (mutated vs. unmutated) | 0.60 (0.25-1.44) | 0.255 |
| | Zap70 (Positive vs. Negative) | 1.24 (0.70-2.19) | 0.467 | Zap70 (Positive vs. Negative) | 1.43 (0.86-2.38) | 0.165 |
| OSU-Ibrutinib | Time to Ibrutinib Discontinuation Due to Disease | | | Overall Survival | | |
| | Univariable modeling | | | Univariable modeling | | |
| | Epitypes | | 0.035 | Epitypes | | <0.001 |
| | HP-CLL vs. IP-CLL | 0.22 (0.05-1.01) | 0.052 | HP-CLL vs. IP-CLL | 0.09 (0.01-0.69) | 0.021 |
| | HP-CLL vs. LP-CLL | 0.17 (0.04-0.67) | 0.011 | HP-CLL vs. LP-CLL | 0.09 (0.01-0.62) | 0.015 |
| | IP-CLL vs. LP-CLL | 0.76 (0.36-1.60) | 0.474 | IP-CLL vs. LP-CLL | 0.96 (0.49-1.88) | 0.911 |
| | IGHV (mutated vs. unmutated) | 0.33 (0.13-0.84) | 0.020 | IGHV (mutated vs. unmutated) | 0.41 (0.18-0.96) | 0.041 |
| | Zap70 (Positive vs. Negative) | NA | | Zap70 (Positive vs. Negative) | NA | |
| | Multivariable modeling | | | Multivariable modeling | | |
| | Epitypes | | 0.023 | Epitypes | | 0.023 |
| | HP-CLL vs. IP-CLL | 0.29 (0.05-1.76) | 0.179 | HP-CLL vs. IP-CLL | 0.09 (0.01-0.79) | 0.029 |
| | HP-CLL vs. LP-CLL | 0.27 (0.03-2.03) | 0.202 | HP-CLL vs. LP-CLL | 0.11 (0.01-0.99) | 0.049 |
| | IP-CLL vs. LP-CLL | 0.91 (0.36-2.33) | 0.851 | IP-CLL vs. LP-CLL | 1.12 (0.45-2.78) | 0.802 |
| | IGHV (mutated vs. unmutated) | 0.66 (0.15-2.83) | 0.578 | IGHV (mutated vs. unmutated) | 0.91 (0.29-2.88) | 0.872 |
| | Zap70 (Positive vs. Negative) | NA | | Zap70 (Positive vs. Negative) | NA | |

**Supplemental Table 6: Multivariable analyses comparing epitype adjusting for age, gender, Rai-stage and del(17p).**

| Cohort | | HR (95% CI) | P-value | | HR (95% CI) | P-value |
|---|---|---|---|---|---|---|
| Mayo | Time to First Treatment | | | Overall Survival | | |
| | Epitypes | | <0.001 | Epitypes | | 0.484 |
| | HP-CLL vs. IP-CLL | 0.38 (0.19-0.76) | 0.006 | HP-CLL vs. IP-CLL | 0.84 (0.25-2.87) | 0.787 |
| | HP-CLL vs. LP-CLL | 0.21 (0.12-0.39) | <0.001 | HP-CLL vs. LP-CLL | 0.59 (0.25-1.41) | 0.236 |
| | IP-CLL vs. LP-CLL | 0.57 (0.31-1.04) | 0.065 | IP-CLL vs. LP-CLL | 0.70 (0.21-2.29) | 0.555 |
| | Age | 1.00 (0.98-1.02) | 0.898 | Age | 1.06 ( 1.03-1.1) | <0.001 |
| | Male vs Female | 1.19 (0.68-2.05) | 0.545 | Male vs Female | 1.56 ( 0.58-4.21) | 0.377 |
| | Rai stage (3/4 vs. 0-2) | 9.01 (4.09-19.88) | <0.001 | Rai stage (3/4 vs. 0-2) | 0.86 ( 0.16-4.75) | 0.865 |
| | Del17p (positive vs. negative) | 1.77 (0.63-4.95) | 0.276 | Del17p (positive vs. negative) | 6.8 ( 1.45-31.81) | 0.015 |
| MDACC | Time to Progression | | | Overall Survival | | |
| | Epitypes | | <0.001 | Epitypes | | <0.001 |
| | HP-CLL vs. IP-CLL | 0.62 (0.24-1.62) | 0.331 | HP-CLL vs. IP-CLL | 0.44 (0.22-0.89) | 0.023 |
| | HP-CLL vs. LP-CLL | 0.23 (0.11-0.51) | <0.001 | HP-CLL vs. LP-CLL | 0.25 ( 0.15-0.44) | <0.001 |
| | IP-CLL vs. LP-CLL | 0.37 (0.19-0.71) | 0.003 | IP-CLL vs. LP-CLL | 0.57 (0.33-0.98) | 0.041 |
| | Age | 0.98 ( 0.96-1.00) | 0.068 | Age | 1.05 ( 1.03-1.07) | <0.001 |
| | Male vs Female | 1.42 ( 0.84-2.42) | 0.194 | Male vs Female | 1.45 ( 0.96-2.18) | 0.076 |
| | Rai stage (3/4 vs. 0-2) | 1.5 ( 0.94-2.4) | 0.088 | Rai stage (3/4 vs. 0-2) | 1.52 ( 1.01-2.28) | 0.042 |
| | Del17p (positive vs. negative) | 6.34 ( 2.39-16.78) | <0.001 | Del17p (positive vs. negative) | 4.96 ( 2.54-9.71) | <0.001 |
| OSU-Ibrutinib | Time to Ibrutinib Discontinuation Due to Disease | | | Overall Survival | | |
| | Epitypes | | 0.009 | Epitypes | | <0.001 |
| | HP-CLL vs. IP-CLL | 0.18 (0.04-0.88) | 0.034 | HP-CLL vs. IP-CLL | 0.04 (0.01-0.34) | 0.003 |
| | HP-CLL vs. LP-CLL | 0.18 (0.04-0.75) | 0.019 | HP-CLL vs. LP-CLL | 0.06 (0.01-0.44) | 0.006 |
| | IP-CLL vs. LP-CLL | 1.03 (0.48-2.20) | 0.941 | IP-CLL vs. LP-CLL | 1.42 (0.69-2.90) | 0.337 |
| | Age | 0.98 (0.95-1.01) | 0.164 | Age | 1.04 ( 1.01-1.07) | 0.01 |
| | Male vs Female | 1.06 (0.62-1.81) | 0.841 | Male vs Female | 1.56 ( 0.91-2.67) | 0.109 |
| | Rai stage (3/4 vs. 0-2) | 1.26 (0.73-2.18) | 0.404 | Rai stage (3/4 vs. 0-2) | 2.6 ( 1.42-4.76) | 0.002 |
| | Del17p (positive vs. negative) | 2.69 (1.61-4.47) | <0.001 | Del17p (positive vs. negative) | 2.4 ( 1.46-3.94) | 0.001 |