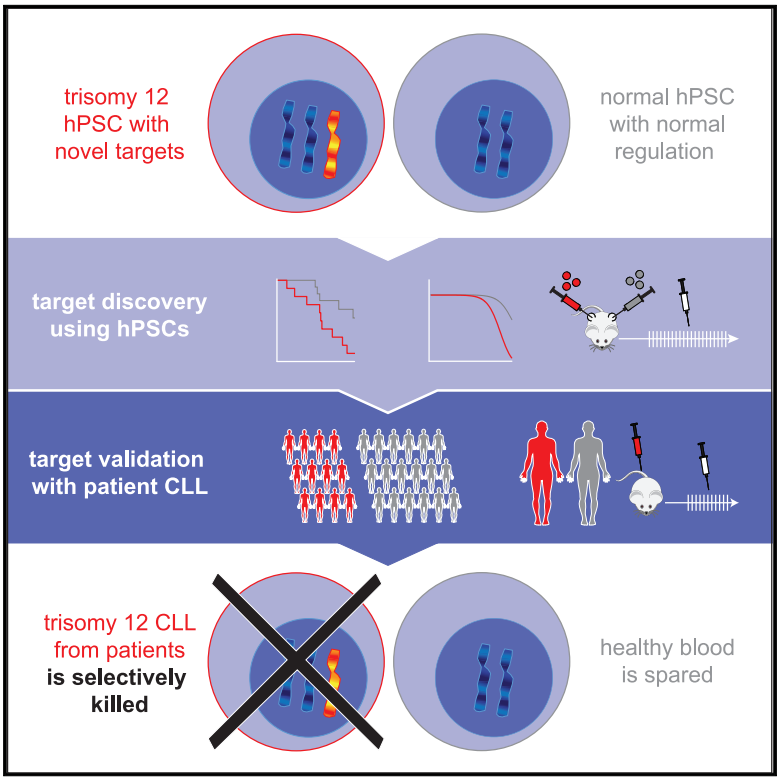
Cell Reports

Article

Human pluripotent stem cells identify molecular targets of trisomy 12 in chronic lymphocytic leukemia patients

Graphical Abstract



Highlights

- Trisomy 12 (tri12) human PSCs and tri12 CLL share similar transcriptional perturbations
- Drug testing of 44 CLL patient samples identifies EDNRB and IRAK4 as tri12 CLL targets
- IRAK4 inhibition selectively targets tri12 CLL in patient-derived xenografted mice
- IRAK4 inhibition does not impact healthy blood in patient-derived xenografted mice

Authors

Jennifer C. Reid, Diana Golubeva, Allison L. Boyd, ..., Antonino Neri, Brian Leber, Mickie Bhatia

Correspondence

mbhatia@mcmaster.ca

In brief

Reid et al. develop a trisomy 12 human pluripotent stem cell model to discover targetable features of trisomy 12 chronic lymphocytic leukemia in patient samples. This model enables controlled study of large genetic aberrations and uncovers inhibitors that target unique features, which are both currently unavailable for this disease population.

Reid et al., 2021, Cell Reports *34*, 108845
March 16, 2021
https://doi.org/10.1016/j.celrep.2021.108845

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Human pluripotent stem cells identify molecular targets of trisomy 12 in chronic lymphocytic leukemia patients

Jennifer C. Reid,[1] Diana Golubeva,[1] Allison L. Boyd,[1] Cameron G. Hollands,[1] Charisa Henly,[1] Luca Orlando,[1] Andrew Leber,[1] Josée Hébert,[2,3] Fortunato Morabito,[4,5] Giovanna Cutrona,[6] Luca Agnelli,[7,8,9] Massimo Gentile,[4] Manlio Ferrarini,[10] Antonino Neri,[7,8] Brian Leber,[11] and Mickie Bhatia[1,12,*]

[1]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada
[2]Department of Medicine, Université de Montréal, Montreal, QC, Canada
[3]Division of Hematology-Oncology, Maisonneuve-Rosemont Hospital, Montreal, QC, Canada
[4]Department of Onco-Hematology, Biotechnology Research Unit, AO of Cosenza, Cosenza, Italy
[5]Hematology and Bone Marrow Transplant Unit, Augusta Victoria Hospital, Jerusalem, Israel
[6]Molecular Pathology Unit, IRCCS Ospedale Policlinico San Martino, Genoa, Italy
[7]Department of Oncology and Hemato-oncology, University of Milan, Milan, Italy
[8]Department of Onco-Hematology, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy
[9]Pathobiology Unit 2, IRCCS National Cancer Institute, Milan, Italy
[10]Department of Experimental Medicine, University of Genoa, Genoa, Italy
[11]Department of Medicine, McMaster University, Juravinski Hospital, Hamilton, ON, Canada
[12]Lead contact
*Correspondence: mbhatia@mcmaster.ca
https://doi.org/10.1016/j.celrep.2021.108845

SUMMARY

Identifying precise targets of individual cancers remains challenging. Chronic lymphocytic leukemia (CLL) represents the most common adult hematologic malignancy, and trisomy 12 (tri12) represents a quarter of CLL patients. We report that tri12 human pluripotent stem cells (hPSCs) allow for the identification of gene networks and targets specific to tri12, which are controlled by comparative normal PSCs. Identified targets are upregulated in tri12 leukemic cells from a cohort of 159 patients with monoclonal B cell lymphocytosis and CLL. tri12 signaling patterns significantly influence progression-free survival. Actionable targets are identified using high-content drug testing and functionally validated in an additional 44 CLL patient samples. Using xenograft models, interleukin-1 receptor-associated kinase 4 (IRAK4) inhibitor is potent and selective against human tri12 CLL versus healthy patient-derived xenografts. Our study uses hPSCs to uncover targets from genetic aberrations and apply them to cancer. These findings provide immediate translational potential as biomarkers and targets for therapeutic intervention.

INTRODUCTION

Chronic lymphocytic leukemia (CLL) is the most common adult hematologic malignancy and leads to acquired immune dysfunction and related complications in most patients (Parikh, 2018; Stewart and Wild, 2014). CLL is classically defined as monoclonal expansion of B cells, which suggests genetic lesions may drive disease manifestation. Cytogenetic analysis of CLL patient samples shows a relatively small number of recurrent cytogenetic aberrations that have been incorporated into standard of care using fluorescence *in situ* hybridization (FISH) (Döhner et al., 2000; Parikh et al., 2016). Chromosomal abnormalities strongly correlate to overall survival and can be segregated into five groups: deletion of chromosome 17 p-arm (del(17p)), del(11q), trisomy 12 (tri12), normal karyotype ($FISH^{NEG}$), and del(13q) (in order from shortest to longest survival) (Döhner et al., 2000). Detection of mutation status of immunoglobulin heavy-chain variable region (*IGHV*) is another clinical biomarker devoid of targeting capacity (Hallek et al., 2018). All FISH-based cytogenetic abnormalities are observed in *IGHV*-mutated (*IGHV*-M) and *IGHV*-unmutated (*IGHV*-U) cases, with the latter experiencing a worse prognosis (Hu et al., 2019; Morabito et al., 2013).

Originally, risk stratification for CLL was based on tumor size and degree of healthy hematopoietic suppression (Binet et al., 1981; Rai et al., 1975), but more recently risk stratification efforts for CLL patients include *IGHV* mutation status combined with patient age (known as CLL-IPI) (Bahlo et al., 2016; Muñoz-Novas et al., 2018). However, clinical studies have called for further stratification based on signaling pathways and genetics (Hu et al., 2019), especially because CLL-IPI risk stratification does not assess patient genetics beyond *TP53* and del(17p) abnormalities. Current and modern clinical investigations recommend enrolling CLL patients in clinical trials whenever possible to

evaluate new targeted therapies (Hallek, 2019). Several targeted therapies recently have been developed, including ibrutinib and venetoclax. Ibrutinib is a chemical inhibitor of Bruton's tyrosine kinase (BTK), a receptor important in B cell development (Honigberg et al., 2010). Venetoclax is a chemical inhibitor of B cell leukemia 2 (BCL2) that reactivates apoptosis (Roberts et al., 2016). These drugs have shown promising efficacy separately and together for CLL patients in clinical trials (Fischer et al., 2019; Fraser et al., 2019; Hillmen et al., 2019; Roberts et al., 2016). However, ibrutinib and venetoclax are not without side effects (Brown, 2018; Fischer et al., 2019; Roberts et al., 2016), do not target genetic lesions, and often achieve single-digit complete remission rates and low rates of undetectable minimal residual disease (Hallek, 2019). However, their partnering with other drugs remains of future interest (Kater and Brown, 2019). Targeted therapies based on cytogenetic aberrations have not yet been developed for CLL, nor are they available for tri12 in other diseases. Accomplishing this would represent a significant advance toward precision therapy and understanding disease biology.

The frequent occurrence of tri12 in a quarter of CLL patients (Döhner et al., 2000) and its association with shorter time to first treatment (Hu et al., 2019) strongly suggest that genes located on chromosome 12 are involved in CLL pathogenesis. Additional genetic abnormalities (*KRAS*, *NRAS*, *BRAF*, *NOTCH1*, or del(14q)) may be enriched in tri12 patients compared with CLL patients with other cytogenetic abnormalities (Del Giudice et al., 2012; Roos-Weil et al., 2018; Vendramini et al., 2019). However, none of these mutations have been found to co-occur in more than 24% of tri12 patient cases, indicating that no particular genetic abnormality can systematically explain the biological differences between tri12 versus non-tri12 disease. Furthermore, modern clinical perspectives of CLL therapy have noted that "the genes involved in the pathogenesis of CLL carrying tri12 are largely unknown. Furthermore, the prognostic relevance of tri12 remains a matter of debate" (Hallek, 2019). This may be in part to due to inadequate model systems useful for initial screening endeavors. Genetic models are not available for tri12 CLL beyond the use of patient samples (summarized in Table S1). However, pluripotent models of disease offer the potential to study patient-specific genetic aberrations, such as acute myeloid leukemia (AML) (Chao et al., 2017; Kotini et al., 2017) and trisomy 21 (Down syndrome) (Banno et al., 2016), but have not been applied to CLL.

Here we report the use of tri12 human pluripotent stem cells (hPSCs) that provide the ability to overcome previous barriers in studying tri12 transformation by using a normal PSC control. This approach allowed the use of unbiased machine learning and transcriptomics, followed by functional validation using *in vitro* and *in vivo* drug testing, e.g., healthy versus CLL patient-derived xenograft (PDX) mice, and over 200 deeply characterized primary patient samples. We identify transcriptional features of tri12 able to predict disease progression and identify druggable tri12-selective targets that were translatable to primary CLL patient samples cohorts independently. This study offers an application of human stem cell models to deconvolute the pathology of a recurrent cytogenetic abnormality and provide more tailored forms of targeted genetic therapies.

RESULTS

Characterization of human tri12 PSC as a genetic model for tri12 CLL disease

Spontaneous acquisition of cytogenetic aberrations upon prolonged culture of hPSCs has been associated with detrimental consequences to properties of normal PSCs (Ben-David et al., 2014; Draper et al., 2004; Herszfeld et al., 2006; International Stem Cell Initiative et al., 2011; Werbowetski-Ogilvie et al., 2009). We identified and isolated a subline of CA2 human embryonic stem cells that spontaneously acquired non-mosaic tri12 (in all cells; Figure 1A) with no other detected chromosomal aberrations (Figures S1A and S1B), and we termed this subline "tri12 PSC." In parallel, we identified a second CA2 PSC line devoid of tri12, which maintained a normal karyotype over the same time and cultures conditions, thus providing a near-isogenic normal comparative line to tri12 PSCs, which we have termed "normal PSCs" (Figures S1A and S1B; Table S2). Genetically aberrant hPSCs have been previously reported to acquire hallmarks of cancer, known as transformation (Herszfeld et al., 2006; Werbowetski-Ogilvie et al., 2009) and enhanced proliferation and tumorigenicity (Ben-David et al., 2014). These and additional features were investigated for tri12 and normal PSCs to determine if cancerous transformation was associated with tri12 in the pluripotent state. tri12 and normal PSCs share typical PSC colony morphology (Figures S1C and S1D). Using a battery of biological assays for PSCs, tri12 PSCs compared with normal PSCs had not significantly acquired phenotypic markers associated with enhanced proliferation (Figures S1E and S1F), expression of the Oct4 pluripotency marker (Figure S1G), or resistance to apoptosis (Figure S1H). Furthermore, differentiation capacity to all three embryonic germ layers (mesoderm, ectoderm, and endoderm) using teratoma formation was maintained in individual teratomas derived from tri12 and normal PSCs (Figure S1I). However, when we evaluated colony-initiating capacity (CIC) as a functional test of proliferation and stem cell self-renewal (Bendall et al., 2007), we observed significantly higher CIC frequency in tri12 PSCs (Figure 1B). This suggests a possible basis for the favorable acquisition of tri12 emergence coupled with *in vitro* selection, similar to a previous report (Ben-David et al., 2014). Gene set enrichment analysis (GSEA) showed enrichment of cancer-related pathways in tri12 PSCs compared with normal PSCs (Figure 1C), which is also similar to a previous report (Ben-David et al., 2014). Gene expression analysis of tri12 and normal PSCs showed that significantly differentially regulated genes could be identified on all chromosomes, while the proportional majority of upregulated genes were predictably located on chromosome 12 (Figure 1D; Table S3). Mapping genotype to phenotype can be difficult with large aberrations. A recent genome-wide association study demonstrated the powerful effects and cooperativity of *trans*-activating genes (Liu et al., 2019). Because chromosome 12 contains at least 1,000 genes (Consensus Coding Sequence [CCDS] database, NIH), tri12-based gene dosage likely alters the expression (both up and down) of transcriptional activators and repressors located elsewhere in the genome (Figure 1D). It is likely, therefore, that genes exclusively resident on chromosome 12 may not be the only genes involved in phenotypes associated with tri12.

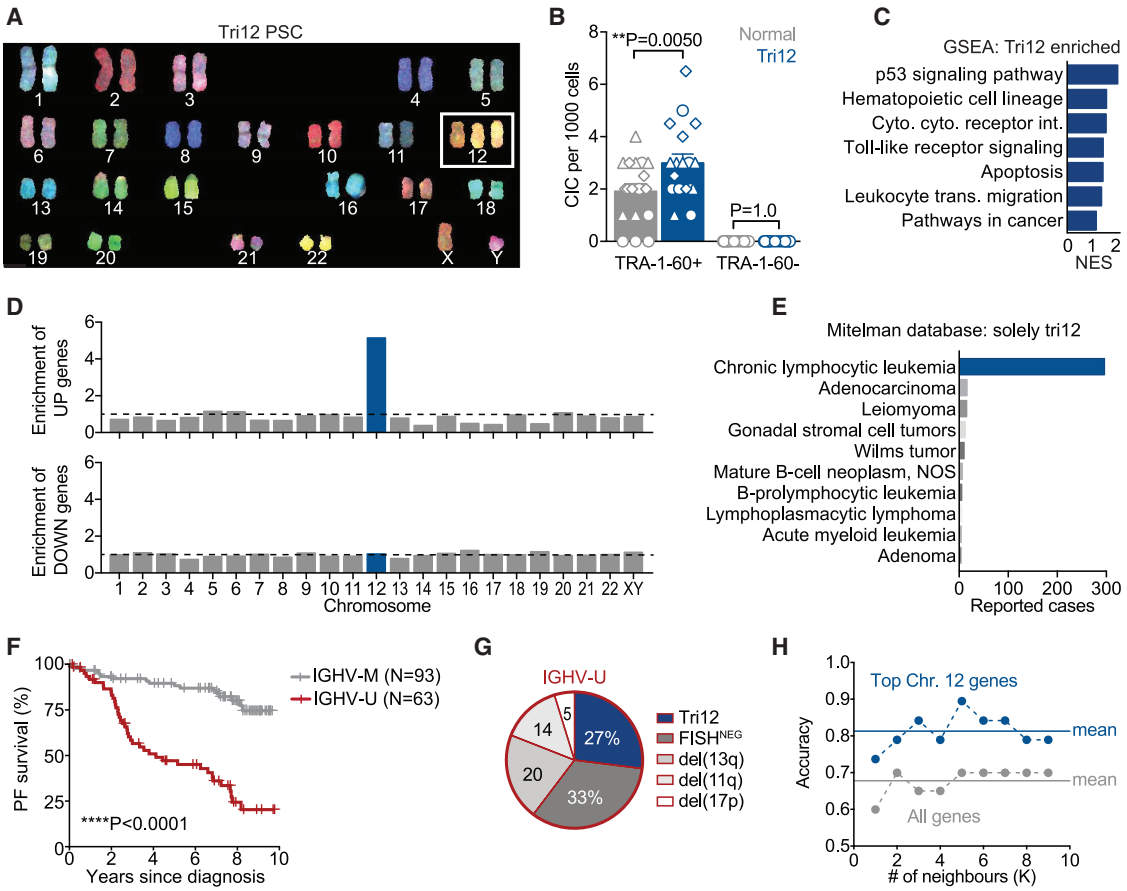


Figure 1. Discovery of a tri12 human PSC line with shared features of tri12 CLL patients

(A) Spectral karyotype shows hPSCs with tri12 (50/50 metaphases).

(B) Number of Oct4^{+} colony-initiating cells (CICs) quantified per 1,000 sorted pluripotent (TRA-1-60^{+}) or differentiated (TRA-1-60^{-}) cells. Mean ± SD. 3 independent experiments and n = 18 biological replicates. One-way ANOVA with Sidak's multiple comparisons test.

(C) Result of gene set enrichment analysis (GSEA; 1,000 permutations) of Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets significantly enriched in tri12 versus normal hPSCs. False detection rate-corrected (FDR) $p < 0.2$. Normalized enrichment score (NES).

(D) Frequency distribution of significantly differentially regulated genes per chromosome of tri12 versus normal hPSC (fold change > |1.25| and FDR $p < 0.05$). Raw data are presented in Table S3, where "Ratio of DOWN or UP genes; Percent of Total Count" is divided by "Total genes; Percent of Total Count," calculated per chromosome. Chromosome 12 is indicated in blue.

(E) The Mitelman database lists 404 reported cases for tri12 as the sole chromosomal abnormality. The top 10 diseases are shown by descending number of reported cases. CLL represents 73.3% of all reported cases.

(F) Kaplan-Meier estimator of progression-free (PF) survival (PFS) since diagnosis of 159 Rai stage 0 MBL and CLL patients (all patient annotation in Table S4).

(G) Pie chart shows the prevalence of cytogenetic abnormalities in n = 63 IGHV-U patients shown in (F).

(H) Accuracy of correctly predicting cytogenetic class (tri12 versus not tri12) of n = 63 IGHV-U patients, using nine iterations of K-nearest neighbors (KNNs; 70% train, 30% test) machine learning algorithm.

KNNs trained and tested independently using data from PCA of top chromosome 12 (Top Chr. 12) genes (H), and data from PCA of all genes in transcriptome (Figure S1M). NES, normalized enrichment score.

See also Figure S1 and Tables S2–S5.

To more widely investigate diseases where tri12 is prevalent, we accessed the Mitelman database (Mitelman et al., 2003) because the tri12 and normal PSCs potentially offered a discovery platform based on genetic perturbations. CLL was the most prevalent report of sole acquisition of tri12 (73% of reported cases), with nine other diseases representing less than 20% of reported cases combined (Figure 1E). We therefore analyzed the number of people newly diagnosed with tri12 CLL every year in Europe (Figure S1J). The total number of patients with CLL is higher than AML. Despite the rarity of MLL-AF9, which affects ~3% of AML patients (Papaemmanuil et al., 2016), it is clinically relevant and described with multiple pluripotent disease models (Chao et al., 2017; Lee et al., 2017). Meanwhile, there are over 10× more tri12 CLL patients than AML patients with an MLL genetic perturbation, but there are no tri12-specific disease models. Based on this comparison, the patient impact of tri12 CLL cannot be understated and required attention.

Due to the high incidence of tri12 in CLL (Figure S1J) and the lack of tri12 CLL disease models in general (Table S1), we also tested the capacity of tri12 and normal PSCs to be differentiated into B lymphoid cells (Figure S1K), which is the primary affected cell type in CLL disease. We observed a reproducible enhancement of $CD45^{+}CD19^{+}$ B lymphoid potential from tri12 PSCs as compared with normal PSCs tested in parallel (Figures S1L and S1M). This suggests tri12 PSCs may share features with primary CLL disease. However, expansive testing of tri12 PSC-derived B cells with normal control was precluded by the low efficiency of normal PSC potential for B cells required as an essential control and is consistent with previous reports of B cell derivation from hPSCs (French et al., 2015; Vodyanik et al., 2005). Further direct comparisons between additional tri12 and isogenic/normal PSCs would help strengthen these observations and determine if they are more widely applicable. Furthermore, it is difficult to prospectively separate normal control B cells and leukemic B cells from a CLL patient's tissue sample. Therefore, we explored undifferentiated tri12 and normal PSCs as a platform to understand the transcriptional impact of tri12 in order to apply these findings to human CLL disease.

Initially reported at 3.8 years for CLL disease monitoring (Morabito et al., 2013), we accrued 159 patients who were diagnosed with Rai stage 0 clinical monoclonal B cell lymphocytosis (MBL) and CLL, and recorded their progression-free survival (PFS) for up to 10 years. Patient characteristics, including sex, age, and white blood cell (WBC) count, were also collected (Table S4). Purified peripheral blood (PB) leukemic B cells from these 159 patients were also profiled for genetic lesions, cytogenetic abnormalities, *IGHV* mutation status, and transcriptome (Table S4). Approximately half of these MBL and CLL patients are classified as *IGHV*-U, which has significantly worse PFS than *IGHV*-M patients (****$p < 0.0001$, Figure 1F) (Parikh, 2018) but is not an oncogenic driver. Of the four major FISH-based cytogenetic groups of CLL (Döhner et al., 2000), tri12 was the most common genetic abnormality in the *IGHV*-U cohort (Figure 1G). Our cohort is also highly enriched in del(13q) and FISH-negative samples, in addition to tri12, while very few patients carried del(11q) and del(17p) alterations. This is consistent with reports suggesting del(11q) and del(17p) may be selected by therapeutic pressure over time (Landau et al., 2015), and therefore less represented in a cohort of patients at diagnosis such as ours. Therefore, our results likely underestimate the frequency of disease progression in the general CLL population. However, given the enrichment in tri12 patients, we proceeded with our analysis given there are no current validated gene expression signatures that predict PFS in tri12 CLL patients. In order to develop this, we compared the transcriptomes of *IGVH*-U tri12 (n = 17) with all other *IGHV*-U (n = 46) patients. Using these 63 *IGHV*-U patients, we performed principal-component analysis (PCA) of all genes in the transcriptome. However, no tri12-specific clustering trends could be observed (Figure S1N), consistent with approaches used in previous studies (Abruzzo et al., 2018; Beekman et al., 2018; Morabito et al., 2013). Therefore, we computed the top differentially regulated genes determined using tri12 and normal PSC gene expression analysis (Table S5). We filtered the transcriptome from the 63 *IGHV*-U patients using genes from this list located on chromosome 12 and then performed PCA a second time (Figure S1N). We observed partial segregation of tri12 patients from all other *IGHV*-U patients. In order to quantify this segregation, we computed the accuracy of the K-nearest neighbor (KNN) machine learning algorithm. Indeed, the KNN mean accuracy was 10% higher when using top chromosome 12 genes than all genes of the transcriptome (Figure 1H). These results revealed the opportunity that commonly activated genes in tri12 CLL patients may be shared in tri12 hPSCs and could be used for further investigation of putative targets of disease.

Unique classifier for CLL patient disease derived from tri12 hPSC modeling

To identify whether there are shared perturbed gene networks reflected in tri12 CLL, we performed GSEA using our "tri12 PSC gene signature" based on normal PSC control (Table S5) removing genes on sex chromosomes. GSEA determines whether statistically significant concordant differences between two biological states may exist based on a defined set of genes and, accordingly, compared two cytogenetic groups at a time: tri12 versus cytogenetically normal $FISH^{NEG}$, as well as tri12 versus del(13q) patients. Both comparisons were essential, because tri12 and cytogenetically normal $FISH^{NEG}$ both have an intermediate prognosis (Van Dyke et al., 2016), and del(13q) potentially provides a more homogeneous genetic landscape than $FISH^{NEG}$ patients. GSEA analysis demonstrated genes in the tri12 signature to be highly enriched in tri12 CLL and MBL compared with $FISH^{NEG}$ (Figure 2A) and del(13q) (Figure S2A) CLL and MBL patients. Furthermore, this analysis was repeated using the complete tri12 PSC gene signature, with genes on sex chromosomes included when comparing GSEA of male-only tri12 and $FISH^{NEG}$ patients to validate the entire signature, which was again highly enriched in tri12 patients (Figure S2B). These results indicate that key genes perturbed by tri12 aberration are shared between hPSCs and CLL patients and can be identified based on genetically controlled tri12 PSCs.

To understand the relationship of tri12-specific gene expression and patient outcome, we identified six genes common to tri12 CLL and MBL patients from the GSEA leading edge (LE; shown with green shading in Figure 2A; Figure S2A) (Subramanian et al., 2005), where four were located on chromosome 12 (Figure S2C). We studied disease progression in *IGHV*-U subtype tri12 and $FISH^{NEG}$ patients from our cohort who underwent at minimum 3 years of disease monitoring. The rate of disease progression is highest within the 4 years following diagnosis (Figure 1F), presenting a more targeted analysis with higher clinical relevance. The LE six-gene expression of these patients (n = 9 tri12, n = 13 $FISH^{NEG}$ cytogenetics, all *IGHV*-U) were scaled using minimum/maximum (min/max) normalization so each gene would range from 0 to 1 and therefore be equally weighted (Figure S2D). The sum of the six LE genes appeared to be more highly expressed for tri12 patients who also experienced disease progression (Figure S2E). We computed survival based on the segregation of these patients into two strata: above and below mean LE six-gene score (Figure 2B). PFS was significantly shorter when the LE six genes were more highly expressed (*$p = 0.036$, Figure 2D), but not significantly affected when segregated by disease or cytogenetic group (Figure S2F). The differences in survival were significantly attributable to LE six-gene expression using multivariate Cox proportional hazards analysis

CellPress
OPEN ACCESS

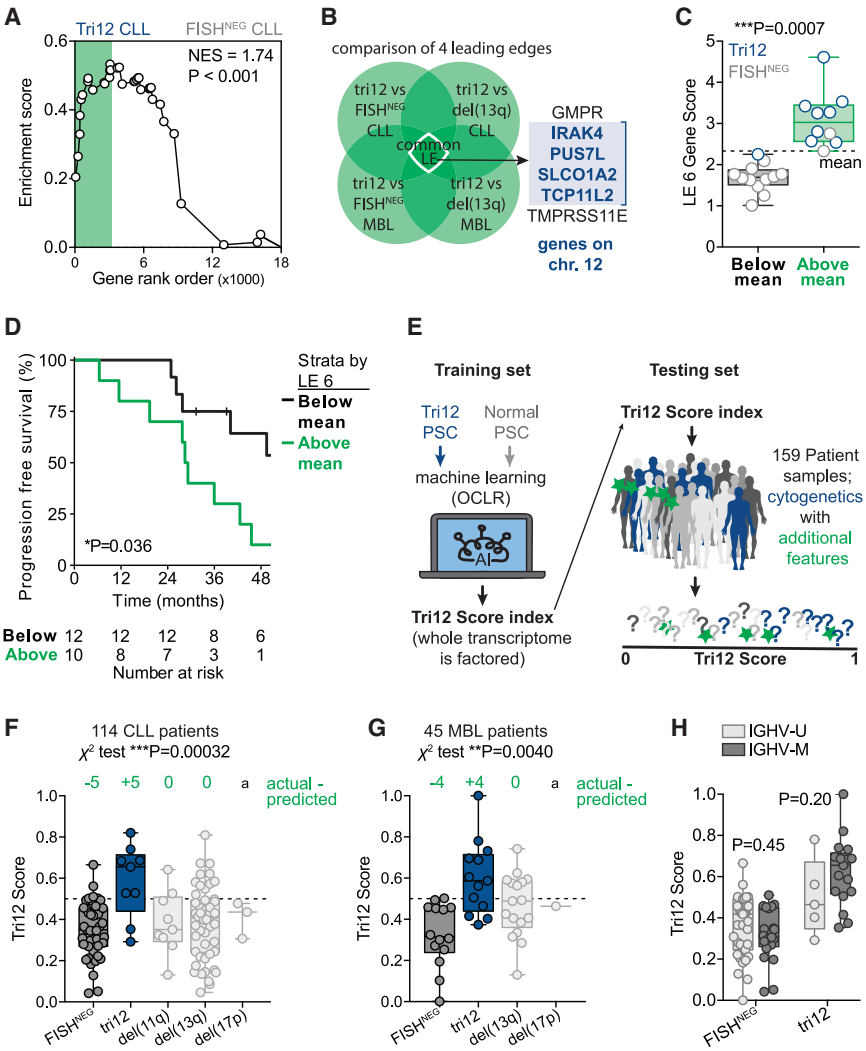


Figure 2. Tri12 gene networks underlie disease progression in patients

(A and B) (A) GSEA enrichment plots of leukemic B cells from Rai stage 0 diagnosis patients; tri12 CLL (N = 9) versus FISH^{NEG} CLL (N = 43). GSEA used 1,000 permutations; the gene list is in Table S5; and genes on sex chromosomes were removed. Leading edge (LE) subset is shown in green shading. (B) Four leading edge (LE) subsets (shown in Figures 2A and S2A–S2B) were compared for the common genes found in all subsets. These genes are shown to the right, and the genes located on chromosome 12 are indicated in blue.

(C and D) LE six-gene score represented as box and whisker plots with errors bars representing min/max (C) and Kaplan-Meier estimator of PFS (D) for n = 22 IGHV-U and Rai stage 0 diagnosis patients with >3 years of disease monitoring, where disease progression occurred within 4 years (if progression occurred). Patients stratified into two portions by LE six-gene score; further described in Figures S2C–S2E. Statistical assessment was performed using Fisher's exact test (C) and log rank test (D).

(E) Schematic of machine learning using global transcriptomes. The tri12 score was developed with machine learning using tri12 and normal hPSCs, which represent the training set. This was then applied to the gene expression data of leukemic B cells from n = 159 Rai stage 0 diagnosis CLL and MBL patients, which represent the testing set.

(F and G) tri12 score of CLL patients (F) and MBL patients (G) represented as box and whisker plots with errors bars representing min/max. Chi-square tests for relationship of tri12 score > 0.5 (indicated by dotted line) and cytogenetic group. The difference between actual and chi-square predicted values is shown in green text above each cytogenetic group. [a]Insufficient del(17p) patient sample size precluded chi-square analysis.

(H) tri12 score of CLL and MBL patients (n = 159), segregated by IGHV mutation status (U, unmutated; M, mutated), and represented as box and whisker plots with errors bars representing min/max. One-way ANOVA with Sidak's multiple comparisons test. All patient annotation is in Table S4.

See also Figure S2 and Tables S3–S6.

(*p = 0.025, Figure S2G). This multivariate analysis also found equivalent risk for attributes such as cytogenetics, consistent with previous reports (Döhner et al., 2000), and disease, which presents a higher risk for CLL when comparing larger cohorts (Morabito et al., 2013). We previously reported slight but significantly worse PFS based on WBC count in CLL (but not MBL), when using a threshold of 10 × 10^3 WBCs/mm^3 (Morabito et al., 2013). We stratified our LE six-gene score patients (Figure 2D) based on this WBC count threshold and did not observe a significantly worse PFS (Figure S2H), likely because of the use of MBL samples and a lower power compared with our previous analysis. Importantly, WBC count was not correlated to LE six-gene score in this cohort (Figure S2I), suggesting it is unlikely that the prognostic value of these genes is a product of cell number and, alternatively, more likely a product of altered pathways to drive leukemogenesis supported by tri12. Our findings indicate that tri12 molecular signatures from hPSCs provide a significant approach to clinical segregation and stratification of CLL patients.

To further apply our tri hPSC and normal cell line system, we compared all cytogenetic groups of CLL and MBL in an unbiased global-transcriptome approach, using a machine learning algorithm known as one-class logistic regression (OCLR; Figure 2D). The OCLR machine learning algorithm trains a model based on a desired population of samples, incorporating a regularization scheme emphasizing mechanistically interlinked genes (Sokolov et al., 2016) that can be used to classify seemingly heterogenous tumor samples based on their gene expression patterns (Malta et al., 2018). OCLR was used to train a model using tri12 PSCs compared with normal PSCs, to develop a tri12 scoring index (Table S6). Our tri12 scoring index was applied to a testing dataset composed of leukemic cells from the 159 individual MBL and CLL patients. Patients were individually classified with a tri12 score ranging from 0 (lowest) to 1 (highest; Table S4). Pearson's

chi-square test determined a statistically significant difference existed between the expected and observed frequency of a higher tri12 score (above 0.5; midway point) across the cytogenetic groups in CLL patients. A higher tri12 score was significantly more frequent for tri12 patients than the chi-square test predicted, which was not observed for any other cytogenetic group (***p = 0.00032, Figure 2E). This same relationship was observed in MBL, where tri12 patients were again more frequently associated with a higher tri12 score than the chi-square test predicted (**p = 0.0040, Figure 2F). Important to the current interest in the field, there were no significant differences of tri12 score between *IGHV*-U and *IGHV*-M patients within cytogenetic groups (Figure 2H), nor was tri12 score significantly affected by the presence of CD38 expression, ZAP-70 expression, *NOTCH1* mutation, or *SF3B1* mutation (Figure S2H). These observed distinctions are especially important due to our previously reported effects of *IGHV* mutational status on patient transcriptional profiles (Morabito et al., 2013) and support the specificity of this index signature to tri12 CLL patients that was derived from a model using genetically controlled tri12 hPSC transcriptomics.

Identification of chemical targeting of tri12 gene dosage with selective properties

Despite representing a potential adjuvant for therapy in CLL patients and other diseases (Figure 1E), tri12-specific chemical inhibitors have yet to be developed. To determine whether these features of tri12 could be targeted using cell identity-independent pathways, we designed and optimized a chemical screening platform for both genetically controlled tri12 hPSCs and relapse/refractory sex- and age-matched CLL patient B cell samples with tri12 and normal cytogenetics (Table 1; Figures S3A–3C). We tested the potency of ibrutinib in both screening systems as a negative control for tri12 versus normal selectivity. Testing ibrutinib potency mitigates *in vitro* artifacts of our study to assure robustness of our screening response and any related interpretations. Thus, ibrutinib could be used as a functional benchmark for the relative potency of a clinically useful drug compared with the other compounds being tested.

Our goal was to identify highly potent chemical inhibitors of gene targets from genetically controlled tri12 PSC gene signature, as well as the recently reported reference epigenome of CLL (Beekman et al., 2018) (Tables S5 and S7). Two potential functional protein-protein networks within the tri12 PSC signature using STRING analysis suggested potential targets could affect tri12 CLL and PSC biology through gene expression networks enriched in genes located on chromosome 12 (Figure 3A). Selectivity toward cells harboring tri12 was tested using a 10-point 2-fold dilution series and by calculating the growth inhibition of 50% based on the concentration-response curves (GI50; Figure S3D). The lead compounds in both the PSC- and CLL-based screens were inhibitors of endothelin receptor type B (EDNRB-i, known as bosentan) and interleukin-1 receptor-associated kinase 4 (IRAK4-i) (Figures 3B–3E; Figures S3D and S3E; Table S8). We also tested inhibitors of KIT, as Kit ligand (*KITLG*) was in the tri12 gene signature. One of three inhibitors of KIT, axitinib, was also significantly selective toward tri12 cells in both screens (Table S8). But axitinib was less selective toward tri12 cells than bosentan (EDNRB-i) or IRAK4-i (Figures 3B and 3C, shown by dotted lines). The mean fluorescence intensity of each of these three proteins was also significantly higher at the single-cell protein level in PSCs (Figures 3F and 3G), which is consistent with RNA expression and drug sensitivity (Tables S5 and S8). Finally, we did not observe evidence of synergistic effects between the two lead compounds (Figure S3G) and so investigated their respective pathways separately. Mechanistically, neither EDNRB nor IRAK4 has been linked to tri12 CLL to date. Elevated endothelin ligand has been observed to be elevated in CLL patient blood, although it is not clear which CLL genetic subtypes contributed to this subpopulation (Martinelli et al., 2017). The endothelin receptor pathway is involved in cancerous properties, such as cell proliferation, invasion, metastasis, and angiogenesis (Rosanò et al., 2013), while the IRAK4 receptor pathway mediates downstream targets in the Toll-like receptor and nuclear factor κB (NF-κB) pathways, which are known to mediate cancer initiation and progression through cellular processes, such as inflammation, apoptosis, and differentiation (Rhyasen and Starczynowski, 2015). These are especially of interest due to their effective chemical targeting and tri12 selectivity. This motivated us to test the effectiveness of our two lead compounds in a larger CLL patient cohort.

Validation of selective targeting in tri12 CLL patient cells

To estimate the therapeutic potential of the two lead compounds and their targets, we collected leukemic PB samples from 44 CLL patients spanning three independent clinical sites as an independent cohort from our previous transcriptional analyses (Table 1). These patients were composed of n = 12 tri12 (no other cytogenetic abnormality), n = 18 $FISH^{NEG}$, and n = 14 CLL patients with one to three deletions in chromosomes 11q, 13q, and 17p (Figure 4A). Moving beyond FISH detection of the main genetic drivers of CLL, the clinical standard-of-care cytogenetic test, whole-genome screening of five randomly selected $FISH^{NEG}$ CLL patients showed there were no cytogenetic abnormalities present (Table 1). This larger cohort provided a more diverse population for further drug testing, while avoiding unbalanced clinical characteristics such as age, sex ratio, disease state, and *IGHV* mutation status across patients (Figure 4B). We first examined each of the 44 leukemic patient samples for protein expression of EDRNB and IRAK4 and found elevated expression of both targets in tri12 CLL patient B cell samples, compared with all other CLL patient cytogenetic groups (Figures 4C and 4D). B cells from four healthy adult volunteer donors were used as non-diseased controls for protein expression, where single-cell protein-level expressions of EDRNB and IRAK4 were found to be extremely low (Figure S4A). Interestingly, EDNRB and IRAK4 protein expressions were significantly and positively correlated for tri12 CLL patients (**p = 0.0013), but not for $FISH^{NEG}$ CLL patients (Figure S4B). Furthermore, differences in EDRNB and IRAK4 protein expressions between *IGHV*-U and *IGHV*-M patients were not observed (Figure S4C), consistent with tri12 specificity. GI50 drug responses to EDNRB-i (bosentan) and IRAK4-i were individually tested for all 44 CLL patient samples *in vitro*. Results indicated tri12 CLL was selectively targeted by IRAK4-i and EDNRB-i, as compared with CLL patient samples that were $FISH^{NEG}$ and carried

Table 1. Forty-four CLL patient samples used for drug testing

CLL Pt[a]	Site[b]	Sex	Age	Status (years)[c]	WBC[d]	Rai stage	CD38 (%)	Cytogenetics Subtype	+12	−11q	−13q	−17p	Assay	Tx[e]	*IGHV*
160	ON	Ma	57	Pr. (9)	177	II	1	tri12	+	–	–	–	CHD	5	N/A
161	ON	Ma	55	Pr. (17)	35	IV	3	normal	–	–	–	–	CHD	7	N/A
162	MB	Ma	73	(4)	27	III	N/A	tri12	+	–	–	–	FISH	1	M
163	MB	F	71	(2)	36	II	61	tri12	+	–	–	–	FISH	1	M
164	MB	Ma	65	(4)	156	I	8	tri12	+	–	–	–	FISH	1	M
165	MB	Ma	94	(4)	51	0	12	tri12	+	–	–	–	FISH	1	U
166	MB	Ma	68	(2)	60	I	60	tri12	+	–	–	–	FISH	1	U
167	QC	Ma	63	Dx	20	N/A	3	tri12	+	–	–	–	FISH	0	N/A
168	QC	F	57	Dx	20	II	40	tri12	+	–	–	–	FISH	0	N/A
169[a]	QC	Ma	69	Dx	52	I	2	tri12	+	–	–	–	FISH	0	N/A
170	QC	Ma	73	Dx	293	0	2	tri12	+	–	–	–	FISH	0	N/A
171	QC	F	62	Pr.	58	3	3	tri12	+	–	–	–	FISH	0	N/A
172[a]	QC	Ma	76	Pr.	130	N/A	6	tri12	+	–	–	–	FISH	7	N/A
173	MB	F	66	(5)	232	I	0	normal	–	–	–	–	FISH	3	U
174	MB	Ma	71	(13)	19	I	42	normal	–	–	–	–	FISH	2	U
175	MB	F	63	(8)	19	I	14	normal	–	–	–	–	FISH	2	M
176	MB	Ma	72	(9)	80	0	0	normal	–	–	–	–	FISH	0	M
177	MB	Ma	61	(15)	111	0	0	normal	–	–	–	–	FISH	1	U
178	QC	Ma	46	Dx	29	0	1	normal	–	–	–	–	FISH	0	N/A
179	QC	Ma	44	Pr.	33	0	3	normal	–	–	–	–	FISH	0	N/A
180	QC	Ma	48	Dx	36	I	42	normal	–	–	–	–	FISH	0	N/A
181	QC	Ma	48	Pr.	108	I	0	normal	–	–	–	–	FISH	0	N/A
182	QC	Ma	56	Pr.	52	0	51	normal	–	–	–	–	FISH	0	N/A
183	QC	Ma	50	Dx	362	IV	0	normal	–	–	–	–	FISH	0	N/A
184	QC	F	72	Pr.	53	0	2	normal	–	–	–	–	FISH	4	N/A
185	QC	Ma	52	Dx	335	III	51	normal	–	–	–	–	FISH	0	N/A
186	QC	Ma	72	Pr.	21	0	2	normal	–	–	–	–	CHD	1	N/A
187	QC	Ma	67	Pr.	270	N/A	40	normal	–	–	–	–	CHD	2	N/A
188	QC	Ma	53	Dx	222	II	1	normal	–	–	–	–	CHD	0	N/A
189	QC	Ma	64	Dx	184	IV	1	normal	–	–	–	–	CHD	0	N/A
190	QC	Ma	42	Dx	80	I	16	-11q	–	+	–	–	FISH	0	N/A
191	QC	Ma	50	Pr.	49	0	2	-13q	–	–	+	–	FISH	1	N/A
192	QC	Ma	54	Pr.	121	N/A	1	-13q	–	–	+	–	FISH	0	N/A
193	QC	Ma	51	Dx	50	I	2	-13q	–	–	+	–	FISH	0	N/A
194	QC	F	74	Pr.	77	0	5	-13q	–	–	+	–	FISH	0	N/A
195	QC	Ma	41	Dx	231	II	2	-13q	–	–	+	–	FISH	1	N/A
196	QC	F	66	Pr.	125	N/A	1	-13q	–	–	+	–	FISH	0	N/A
197	QC	Ma	56	Pr.	68	0	25	-17p	–	–	–	+	FISH	1	N/A
198	QC	F	46	Dx	66	I	1	-17p	–	–	–	+	FISH	0	N/A
199	QC	Ma	45	Dx	59	II	44	-11q-13q	–	+	+	–	FISH	1	N/A
200	QC	Ma	59	Pr.	17	II	17	-11q-13q	–	+	+	–	FISH	3	N/A
201	QC	Ma	58	Pr.	25	N/A	83	-11q-13q	–	+	+	–	FISH	7	N/A

(*Continued on next page*)

Table 1. *Continued*

CLL Pt[a]	Site[b]	Sex	Age	Status (years)[c]	WBC[d]	Rai stage	CD38 (%)	Cytogenetics Subtype	+12	−11q	−13q	−17p	Assay	Tx[e]	*IGHV*
202	QC	F	66	Pr.	68	II	3	-13q-17p	–	–	+	+	FISH	6	N/A
203	QC	F	65	Pr.	135	0	32	-11q-13q-17p	–	+	+	+	FISH	11	N/A

CHD, Cytoscan HD microarray; Dx, diagnosis; F, female; Ma, male; M, mutated; MB, Manitoba; N/A, not available; ON, Ontario; Pr., progression; Pt, patient; QC, Quebec; Tx, treatment; U, unmutated; WBC, white blood cell. See also Figures 3, 4, 5, and S3–S5.

[a]Each row is a different CLL patient, except for two paired samples from one patient separated by 7 years.

[b]Source of CLL patient samples: Juravinski Hospital, Hamilton, ON, Canada; Manitoba Tumor Bank, Winnipeg, MB, Canada; Quebec Leukemia Cell Bank, Montreal, QC, Canada.

[c]Disease status (Dx or Pr.) and/or years from Dx until sample donation.

[d]WBC counts ($\times 10^9$ cells/L); normal WBC range is 1.3–4.4 $\times$ 10^9 cells/L.

[e]Number of treatment regimens received prior to sample donation.

deletions (11q, 13q,17p; ****$p < 0.0001$; Figures 4E and 4F). Consistent with protein expression, we did not observe differences in GI50 between *IGHV*-U and *IGHV*-M patients (Figure S4C). EDNRB-i enabled a larger selectivity index between $FISH^{NEG}$ and tri12 CLL patients than IRAK4-i, while IRAK4-i was the more potent compound overall (Figures 4E and 4F). Linear regression comparing both drug responses on a per-tri12 CLL patient basis demonstrated significant correlation and consistency (****$p < 0.0001$, Pearson correlation $r = 0.92$; Figure 4G), which is supported by a similar correlation observed from protein target expression (Figure S4B). Furthermore, there was no significant linear relationship comparing both drug responses on a per-$FISH^{NEG}$ patient basis, supported by a lack of correlated protein targets (Figures S4B and S4E).

We also attempted to expand the genetic diversity of our tri12 PSC platform through pluripotent reprogramming of the leukemic B cells from patient CLL#160, who tested positive for tri12. Unfortunately, zero pluripotent colonies could be derived from leukemic B cells (Figures S4F and S4G). This was in stark contrast with the relative efficacy of deriving pluripotent colonies from the T cells of patient CLL#160, which all tested negative for tri12 (Figure S4H). This difficulty in reprogramming genetically aberrant leukemic primary patient cells is not new (Lee et al., 2017; Muñoz-López et al., 2016). These data firmly support that tri12 selectivity is achievable *in vitro* across a diverse CLL patient cohort, which could not be assessed using tri12 CLL-derived iPSCs.

In vivo selectivity and efficacy of chemical targeting of human tri12 CLL

Based on the selectivity of EDRNB and IRAK4 inhibitors *in vitro* (Figures 3 and 4), we evaluated the effectiveness of targeting human tri12 cells transplanted into human-mouse xenografts. PDXs of human cells transplanted into immunocompromised mouse models have become a gold standard for demonstrating preclinical therapeutic potential in several hematological malignancies (Boyd et al., 2018), with the conspicuous exception of CLL (Bertilaccio et al., 2013), where PDX models are seldomly applied or applied in a limited fashion for preclinical drug development (Arruga et al., 2020; Herman et al., 2010; Honigberg et al., 2010; Ljungars et al., 2018; Matas-Céspedes et al., 2017; Souers et al., 2013). Although hematopoietic stem cells purified from CLL patients can sustain grafts beyond 4 weeks, these grafts are consistently devoid of the genetic aberrations observed and used to stratify CLL patients and raises concerns of the clinical surrogacy of this approach (Kikushige et al., 2011). However, CLL PDX models have permitted short-term (3–8 days *in vivo*) study of drug interactions in physiological microenvironments (Arruga et al., 2020; Ljungars et al., 2018; Matas-Céspedes et al., 2017). Based on these reports, we designed and tested three different *in vivo* drug treatment regimens for our lead chemical inhibitors of EDNRB and IRAK4, EDNRB-i (bosentan), and IRAK4-I (PF06650833).

Subcutaneous (s.c.) injection of transformed cells for drug development is an accepted standard for many cancer studies, including leading new drugs for CLL (Souers et al., 2013). To examine the tolerability and selectivity of drug treatment *in vivo* with as little genetic variation as possible between tri12 and a control sample, we performed teratoma studies in mice using normal and tri12 PSCs transplanted s.c. into separate flanks of the same recipient mouse (Figure 5A). This approach was chosen because palpable tumors robustly form within 4 weeks (Figure S5A) (Hentze et al., 2009). Teratomas were established for 24 days and confirmed through palpation prior to drug administration, which was given daily by oral gavage for 21 days at 15 mg/kg for both EDNRB-i and IRAK4-i. This dose was selected as a low-to-moderate dose based on previous animal studies (Chilloux et al., 2018; Serafim et al., 2015) and clinical trial of arthritis patients (ClinicalTrials.gov: NCT02996500). This treatment regimen was sublethal and did not interfere with recipient mouse health or overall body weight (Figure S5B). Caliper measurements of teratoma volume were measured prior to and upon completion of drug treatment, allowing for the growth kinetics of individual teratomas to be assessed (Figures S5C and S5D). At the completion of the 45-day experiment, teratomas were measured and excised (Figure S5E). Consistent with previous reports (Ben-David et al., 2014), tri12 PSC-derived teratomas grew larger-than-normal PSC-derived teratomas (Figure S5F). Strikingly, we observed a 50% reduction of tri12-derived teratomas growth rate in the EDRNB-i and IRAK4-i treatment groups, as compared with vehicle-treated mice (Figure 5B). We did not observe any differences of normal-derived teratoma growth rate across treatment groups (Figure 5C). Contribution to all three germ layers was observed from both tri12 and normal PSC-derived teratomas, and there were no histological

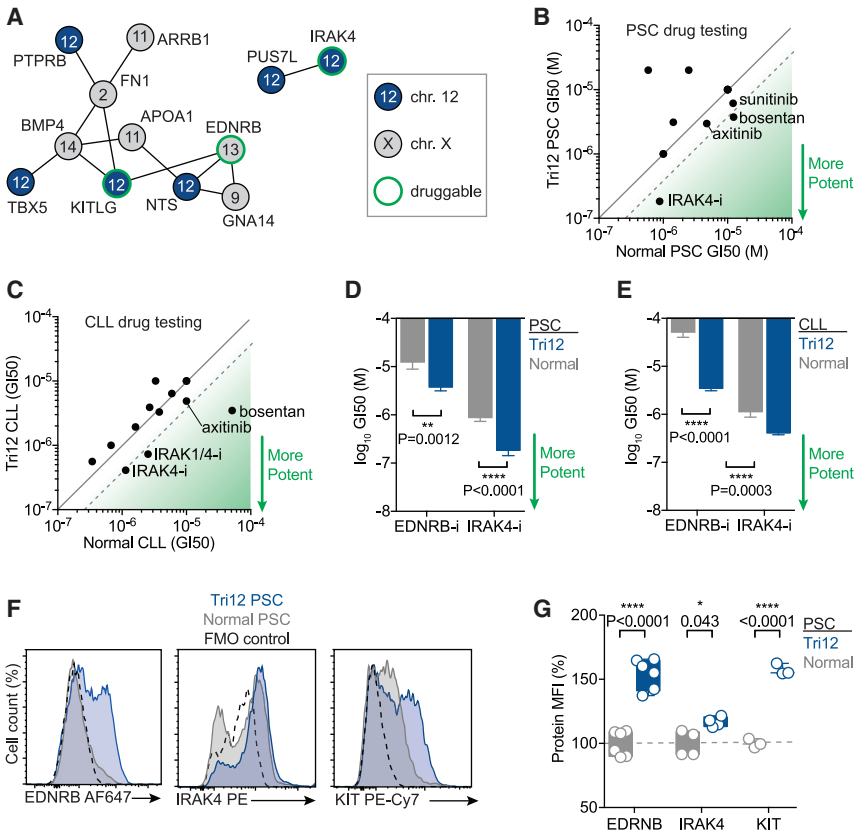


Figure 3. Identification of tri12-selective chemical inhibitors

(A) STRING plot showing tri12 upregulated genes identified from tri12 PSC transcriptome (Table S5). Disconnected nodes were removed. Chromosome (chr.) location is indicated in the center of each node, and genes located on chr. 12 are indicated in blue. Druggable genes are edged in green.

(B) Identification of tri12-selective chemical inhibitors in hPSC drug screen (green; tri12 GI50 < normal GI50 value). Source data and statistics are in Table S8.

(C) Identification of tri12-sensitive chemical inhibitors in relapse/refractory CLL patient sample screen (green; tri12 GI50 < normal karyotype GI50 value). Source data and statistics are in Table S8.

(D) Comparison of $\log_{10}$GI50 calculations for the top two selective compounds tested in both hPSC lines. Two-tailed t tests at 99% confidence and source data and statistics are in Table S8. A lower $\log_{10}$GI50 number indicates a more potent compound. Error bars represent standard error of the $\log_{10}$GI50.

(E) Comparison of $\log_{10}$GI50 calculations for the top two selective compounds tested using CLL patient samples: CLL#160 (tri12) and CLL#161 (normal karyotype). Patient annotation is in Table 1. Two-tailed t tests at 99% confidence and source data and statistics are in Table S8. A lower $\log_{10}$GI50 number indicates a more potent compound. Error bars represent standard error of the $\log_{10}$GI50.

(F) Flow cytometry plot of top 3 protein targets, tested on tri12 and normal hPSCs. Fluorescence minus one (FMO) shown with black dotted line.

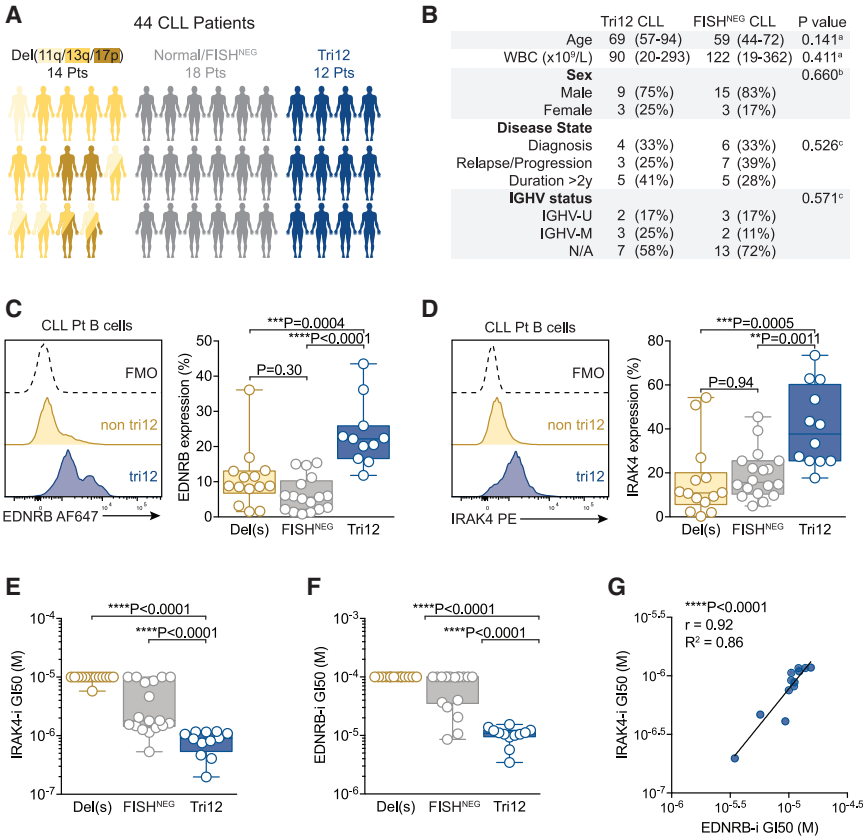
(G) Mean fluorescence intensity (MFI) of total live cells assessed by flow cytometry of tri12 and normal hPSCs, which were tested in parallel, represented as box and whisker plots with errors bars representing min/max. MFI is relative to mean of normal hPSCs (100%). Each dot is one biological replicate. One-way ANOVA with Sidak's multiple comparisons test.

See also Figure S3 and Tables S7 and S8.

differences (Figure S5G). By co-transplanting both tri12 and normal PSCs into the same mouse (in separate locations, using opposite flanks), we were able to control for inter-mouse variability and determine tri12-based selectivity using *in vivo* drug treatment. We also suggest that use of this co-transplantation approach to test drug responses may be more clinically relevant, because it reflects patients possessing both normal and cancerous cells during drug treatment.

Expanding on these initial observations, we used human CLL PDX mice for a short-term regimen consistent with previous reports (Arruga et al., 2020; Ljungars et al., 2018; Matas-Céspedes et al., 2017). CLL PDX models were established from patient CLL#160, whose leukemic cells harbor tri12 (Figure 5D). This patient had persistent active disease in the absence of infection but had not received any treatment in the 3 months prior to sample donation. 100 × 10^6 PB MNCs were transplanted intravenously into preconditioned immunocompromised mice and allowed to home to lymphoid organs for 4 days prior to initiation of chemical treatment. EDNRB-i and IRAK4-i were administered daily by oral gavage (100 mg/kg p.o.) (Bien et al., 2007; Imhof et al., 2011) for the following 4 days, and mice were harvested on the 8th day. Human CLL engraftment ($CD45^+CD19^+$) of the spleen was significantly reduced upon treatment with EDNRB-i (**p = 0.0057), as well as IRAK4-i (**p = 0.0031; Figure 5E). Furthermore, significant reduction of human CLL cells in the PB was detected after only 4 days of treatment with EDNRB-i (**p = 0.0006) or IRAK4-i (*p = 0.0206; Figure 5F).

Encouraged by these results, we completed a longer *in vivo* PDX experiment for 2 weeks and included a healthy mobilized PB (MPB) sample alongside the primary CLL#160 tri12 sample (Figure 4G). We used the same dose of EDNRB-i and IRAK4-i, which was again administered daily by oral gavage (100 mg/kg p.o.). After two weeks *in vivo*, spleens in the vehicle group were enlarged compared with a control mouse that did not receive a transplant but was otherwise treated the same (shown by dotted line, Figure 5H). We observed striking differences in spleen weight across treatment groups, with the mean IRAK4-i group's spleen weight approximating that of the un-transplanted control, which was significantly lower than the vehicle group for the CLL PDX mice (***p < 0.0007). We did not observe any significant differences across groups for spleen weight in healthy PDX mice (Figure S5H). Upon flow cytometric analysis, we observed consistent and high engraftment in both CLL and healthy PDX mice, where the healthy mice showed a more balanced composition between B and T cells in the spleen (Figure 5I). We did not observe any significant differences in human B cell engraftment ($CD45^+CD19^+$) across healthy PDX treatment groups, while human CLL engraftment of the spleen was significantly reduced



	Tri12 CLL	FISH^{NEG} CLL	P value
Age	69 (57-94)	59 (44-72)	0.141[a]
WBC (x10^9/L)	90 (20-293)	122 (19-362)	0.411[a]
Sex			0.660[b]
Male	9 (75%)	15 (83%)	
Female	3 (25%)	3 (17%)	
Disease State			
Diagnosis	4 (33%)	6 (33%)	0.526[c]
Relapse/Progression	3 (25%)	7 (39%)	
Duration >2y	5 (41%)	5 (28%)	
IGHV status			0.571[c]
IGHV-U	2 (17%)	3 (17%)	
IGHV-M	3 (25%)	2 (11%)	
N/A	7 (58%)	13 (72%)	

Figure 4. Selective targeting of tri12 CLL patient cells

(A) Schematic of 44 CLL patient samples tested in this figure. Patient annotation is described in Table 1.

(B) Statistical assessment of the composition of 30 CLL patients was performed using: [a]two-tailed t test, [b]Fisher's exact test, and [c]Fisher-Freeman-Halton exact test.

(C) Flow cytometric plots (left) and quantitation represented as box and whisker plots with errors bars representing min/max (right) of EDNRB expression within CD19$^+$ B cells from peripheral blood (PB) mononuclear cells (MNCs) of 44 CLL patients. One-way ANOVA with Sidak's multiple comparisons test. Fluorescence minus one (FMO) shown with black dotted line.

(D) Flow cytometric plots (left) and quantitation represented as box and whisker plots with errors bars representing min/max (right) of IRAK4 expression within CD19$^+$ B cells from PB MNCs of 44 CLL patients. One-way ANOVA with Sidak's multiple comparisons test. Fluorescence minus one (FMO) shown with black dotted line.

(E) IRAK4-i GI50 (molar values) of 44 CLL patient samples represented as box and whisker plots with errors bars representing min/max. One-way ANOVA with Sidak's multiple comparisons test.

(F) EDNRB-i (bosentan) GI50 (molar values) of 44 CLL patient samples represented as box and whisker plots with errors bars representing min/max. One-way ANOVA with Sidak's multiple comparisons test.

(G) Pearson correlation (p value and Pearson correlation coefficient r) and linear regression analysis (R^2) of GI50 values shown in (E) and (F) per tri12 CLL patient sample.

See also Figure S4.

upon treatment with IRAK4-i (***p = 0.003; Figure 5J). Importantly, human CLL also co-expresses CD5 and CD23 antigens, which we used to further demonstrate specific targeting of leukemic cells (Figure 5K). We did not observe these markers in healthy PDX mice (Figure S5I), which is consistent with their use as diagnostic criteria for CLL. Finally, features of healthy hematopoiesis used in the clinic, such as platelet and red blood cell (RBC) counts, were not affected in the CLL PDX mice across treatment groups (Figure 5L). These results indicate the *in vivo* drug treatment of hPSC-derived teratomas, and healthy and CLL PDX modeling permits the study of selectivity and efficacy of chemical targeting for tri12. Collectively, our results demonstrate IRAK4-i is the most potent and bioavailable suppressor of tri12 CLL disease, thereby providing demonstrated *in vivo* efficacy in inhibiting a target biologically linked to tri12 CLL.

DISCUSSION

Our study reveals transcriptional and signaling alterations impacted by tri12 in human cells using genetically controlled hPSCs. This has allowed subsequent identification of unique targets for therapeutic development in CLL patients harboring the tri12 chromosomal abnormality. tri12 CLL has previously been shown to possess unique properties (Abruzzo et al., 2018; Autore et al., 2019; Beekman et al., 2018; Bulian et al., 2017; Decker et al., 2012; Fiorcari et al., 2019; Vendramini et al., 2019; Zucchetto et al., 2013), but targeted therapies have not arisen. Tumor microenvironment (Cutrona et al., 2018), B cell receptor (BCR) signaling (Smith et al., 2020), and epigenetic regulation (Gaiti et al., 2019) are important considerations that affect CLL disease pathogenesis. These factors, in addition to treatment history and patient background, may contribute to the difficulty in target identification using CLL patient-derived cells.

Our approach using human tri12 PSCs and normal controls demonstrates the ability to capture important aspects of a complex mutation observed in human cancers. We then pragmatically applied these tools to uncover cell-autonomous features of cancer and provide targets for chemical inhibition as well and biomarkers of disease progression, EDNRB and IRAK4. Endothelin signaling is strongly associated with cancerous properties (Rosanò et al., 2013) and is reportedly regulated through epigenetic mechanisms (Martinelli et al., 2017). tri12 CLL was recently reported to have an epigenetic state different from the other cytogenetic subgroups of CLL (Beekman et al., 2018), thereby providing a plausible mechanism for its activation and uniqueness to tri12 CLL specifically. The compound identified in our study with significant *in vivo* efficacy, IRAK4-i, has been used in a phase II trial for arthritis (ClinicalTrials.gov: NCT02996500), which holds more immediate potential for phase I testing in tri12 CLL patients.

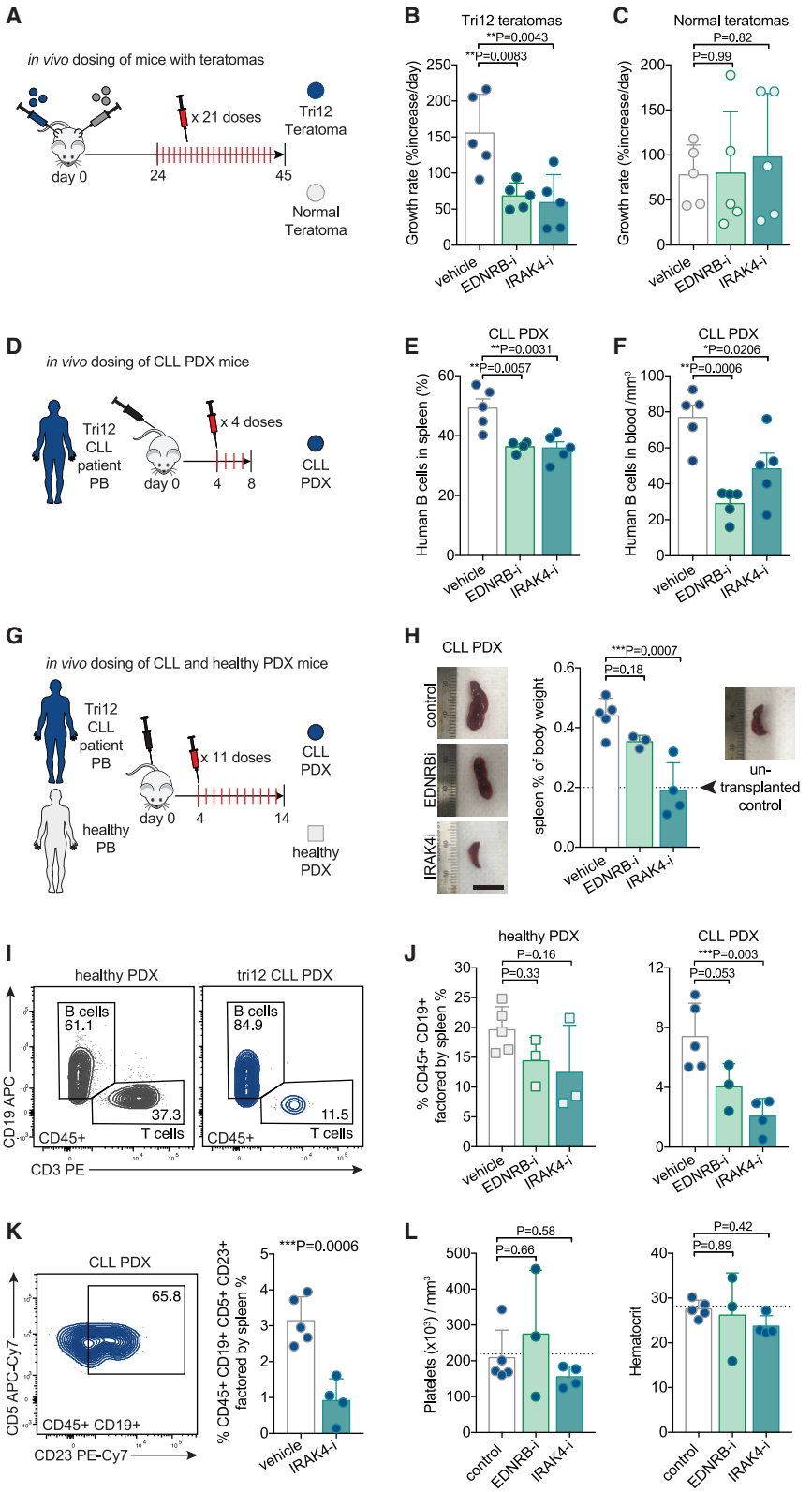


Figure 5. Chemically targeting tri12 tumor and PDX cells *in vivo*

(A) Schematic depicting s.c. transplantation of tri12 and normal PSCs into separate flanks of mice. Tumors were established for 24 days, then mice received daily treatment for 3 weeks.

(B) tri12 PSC-derived tumor growth rate, percent increase per day. Mean ± SD. Each dot is one tumor in a separate mouse, n = 5 mice per group. One-way ANOVA with Dunnett's multiple comparisons test.

(C) Normal PSC-derived tumor growth rate, percent increase per day. Mean ± SD. Each dot is one tumor in a separate mouse, n = 5 mice per group. One-way ANOVA with Dunnett's multiple comparisons test.

(D) Schematic depicting establishment of tri12 CLL PDXs for 2 weeks.

(E) Human CLL chimerism (%CD45$^+$CD19$^+$) of live cells harvested from CLL PDX mouse spleens. Mean ± SD. Each dot is one PDX mouse, n = 5 mice per group. One-way ANOVA with Dunnett's multiple comparisons test.

(F) Human CLL chimerism (PB MNC count was quantified using a blood analyzer, and % CD45$^+$CD19$^+$ was quantified using flow cytometry analysis) of live cells harvested from CLL PDX mouse PB. Mean ± SD. Each dot is one PDX mouse, n = 5 mice per group. One-way ANOVA with Dunnett's multiple comparisons test.

(G) Schematic depicting establishment of healthy MPB or tri12 CLL PDXs for 2 weeks.

(H) Left: images of spleens harvested from CLL PDX mice after 2 weeks. Ruler provides scale, and scale bar is 1 cm. Right: spleen weight divided by body weight of CLL PDX mouse. Mean ± SD. Each dot is a separate mouse. Dotted line indicates a mouse in this cohort (age-matched, sex-matched, and received preconditioning radiation) that was not transplanted. One-way ANOVA with Dunnett's multiple comparisons test.

(I) Flow cytometry plots of live cells harvested from healthy and CLL PDX mouse spleens.

(J) Human B cell chimerism of live cells harvested from healthy and CLL PDX mouse spleens multiplied by "spleen % body weight" shown in Figures 5H and S5H for each mouse. Mean ± SD. Each dot is one PDX mouse. One-way ANOVA with Dunnett's multiple comparisons test.

(K) Left: representative flow cytometry plot of CD45$^+$CD19$^+$ human B cells co-expressing CD5 and CD23. Right: leukemic B cell chimerism of CLL PDX mouse spleens. Mean ± SD. Each dot is one PDX mouse. Unpaired one-tailed t test.

(L) Freshly collected PB from CLL PDX mice was quantified with a blood analyzer for platelet counts and hematocrit. Mean ± SD. Each dot is one PDX mouse. One-way ANOVA with Dunnett's multiple comparisons test. Dotted line indicates a mouse in this cohort (age-matched, sex-matched, and received preconditioning radiation) that was not transplanted.

See also Figure S5.

At this time, the two targets identified here provide proof of principle for CLL patient application and thereby reveal the importance of using controlled genetic systems to find selective compounds and unique disease targets. We propose that tri12-selective targets will play an important role in at least three major areas of precision medicine: (1) better disease management using tri12 biomarkers, (2) therapeutically targeting tri12 gene targets in both CLL and MBL, and (3) to adjuvant and synergize with combination therapy in tri12 cases of disease progression and relapse. tri12 and normal hPSCs offer the potential of high-throughput automated screening platforms for pre-clinical testing and validation studies for new drugs and combinations moving forward. Our tri12 targets may be applied to other diseases where tri12 is the sole cytogenetic abnormality, including adenocarcinoma, leiomyoma, gonadal stromal cell tumors, and Wilms tumors (Mitelman et al., 2003). This further supports the rationale of effectively targeting cell-autonomous pathways and leads to a broader application beyond CLL. This approach also offers a plausible basis to screen for targets in the preleukemic state, allowing for early intervention of MBL patients at high clinical risk for leukemic progression.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study Design
  - Patient samples
  - Human PSC culture
  - Human pluripotent reprogramming
  - Mouse strain used for transplantation experiments
- METHOD DETAILS
  - Cytogenetics
  - Colony initiating cell assay
  - B cell differentiation
  - Flow cytometry
  - High-content drug testing
  - Xenotransplantation
  - Histology
  - STRING protein-protein analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gene expression analysis
  - Survival and hazard ratio analysis
  - One-class logistic regression machine learning
  - Principal component analysis and K-nearest neighbors machine learning
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2021.108845.

ACKNOWLEDGMENTS

We thank the patients and their families for their participation. We thank G. Markous, Y. Kim, and Dr. C. Campbell for valuable comments. J.C.R. was supported by the Canadian Institutes of Health Research Doctoral Award, an Ontario Graduate Fellowship, Michael G. DeGroote Doctoral Scholarship, and the H.G. Thode Postdoctoral Fellowship. D.G. was supported by the Jans Graduate Scholarship. L.A. and A.N. were supported by Associazione Italiana Ricerca sul Cancro (AIRC IG16722, IG10136). The Quebec Leukemia Cell Bank (Banque de cellules leucémiques du Québec, Montreal, Quebec, Canada) receives funds from the Cancer Research Network of the Fonds de Recherche du Québec Santé (FRQS). The Manitoba Tumour Bank (Winnipeg, Manitoba, Canada) receives funds from the CancerCare Manitoba Foundation and the Canadian Institutes of Health Research and is a member of the Canadian Tissue Repository Network. This work was supported by research grants to M.B. from the Leukemia & Lymphoma Society of Canada (353141 and 567994) and support to M.B. from the Canada Research Chair Program (Tier 1) in Human Stem Cell Biology and Michael G. DeGroote Chair in Stem Cell Biology. The work in this study related to human leukemia was funded and supported by the Canadian Cancer Society Impact Grant (706694), whereas results and studies using hPSCs were supported by a Canadian Institutes of Health Research (CIHR) foundation grant FRN 159925 to M.B. approved by the Stem Cell Oversight Committee in Canada.

AUTHOR CONTRIBUTIONS

J.C.R. designed and performed experiments, analyzed and interpreted data, and wrote the manuscript. D.G., A.L.B., C.G.H., C.H., and L.O. performed experiments and analyzed data. A.L. provided clinical annotation. J.H. and B.L. provided patient samples and clinical expertise. F.M., G.C., L.A., M.G., M.F., and A.N. provided patient annotation and clinical expertise. M.B. supervised the project, designed experiments, provided final interpretation, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure sex balance in the selection of non-human subjects. We worked to ensure diversity in experimental samples through the selection of the genomic datasets.

Received: June 11, 2020
Revised: January 13, 2021
Accepted: February 17, 2021
Published: March 16, 2021

REFERENCES

Abruzzo, L.V., Herling, C.D., Calin, G.A., Oakes, C., Barron, L.L., Banks, H.E., Katju, V., Keating, M.J., and Coombes, K.R. (2018). Trisomy 12 chronic lymphocytic leukemia expresses a unique set of activated and targetable pathways. Haematologica *103*, 2069–2078.

Arruga, F., Bracciamà, V., Vitale, N., Vaisitti, T., Gizzi, K., Yeomans, A., Coscia, M., D'Arena, G., Gaidano, G., Allan, J.N., et al. (2020). Bidirectional linkage between the B-cell receptor and NOTCH1 in chronic lymphocytic leukemia and in Richter's syndrome: therapeutic implications. Leukemia *34*, 462–477.

Autore, F., Strati, P., Innocenti, I., Corrente, F., Trentin, L., Cortelezzi, A., Visco, C., Coscia, M., Cuneo, A., Gozzetti, A., et al. (2019). Elevated Lactate Dehydrogenase Has Prognostic Relevance in Treatment-Naïve Patients Affected by Chronic Lymphocytic Leukemia with Trisomy 12. Cancers (Basel) *11*, 896.

Bahlo, J., Kutsch, N., Bauer, K., Bergmann, M.A., Byrd, J.C., Chaffee, K.G., Döhner, H., Eichhorst, B., Else, M., Geisler, C., et al.; International CLL-IPI

working group (2016). An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. Lancet Oncol. *17*, 779–790.

Banno, K., Omori, S., Hirata, K., Nawa, N., Nakagawa, N., Nishimura, K., Ohtaka, M., Nakanishi, M., Sakuma, T., Yamamoto, T., et al. (2016). Systematic cellular disease models reveal synergistic interaction of trisomy 21 and GATA1 mutations in hematopoietic abnormalities. Cell Rep. *15*, 1228–1241.

Beekman, R., Chapaprieta, V., Russiñol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, N., Martens, J.H.A., Duran-Ferrer, M., Kulis, M., Serra, F., Javierre, B.M., et al. (2018). The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia. Nat. Med. *24*, 868–880.

Ben-David, U., Arad, G., Weissbein, U., Mandefro, B., Maimon, A., Golan-Lev, T., Narwani, K., Clark, A.T., Andrews, P.W., Benvenisty, N., and Carlos Biancotti, J. (2014). Aneuploidy induces profound changes in gene expression, proliferation and tumorigenicity of human pluripotent stem cells. Nat. Commun. *5*, 4825.

Bendall, S.C., Stewart, M.H., Menendez, P., George, D., Vijayaragavan, K., Werbowetski-Ogilvie, T., Ramos-Mejia, V., Rouleau, A., Yang, J., Bossé, M., et al. (2007). IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. Nature *448*, 1015–1021.

Bertilaccio, M.T., Scielzo, C., Simonetti, G., Ten Hacken, E., Apollonio, B., Ghia, P., and Caligaris-Cappio, F. (2013). Xenograft models of chronic lymphocytic leukemia: problems, pitfalls and future directions. Leukemia *27*, 534–540.

Bien, S., Riad, A., Ritter, C.A., Gratz, M., Olshausen, F., Westermann, D., Grube, M., Krieg, T., Ciecholewski, S., Felix, S.B., et al. (2007). The endothelin receptor blocker bosentan inhibits doxorubicin-induced cardiomyopathy. Cancer Res. *67*, 10428–10435.

Binet, J.L., Auquier, A., Dighiero, G., Chastang, C., Piguet, H., Goasguen, J., Vaugier, G., Potron, G., Colona, P., Oberling, F., et al. (1981). A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer *48*, 198–206.

Boyd, A.L., Aslostovar, L., Reid, J., Ye, W., Tanasijevic, B., Porras, D.P., Shapovalova, Z., Almakadi, M., Foley, R., Leber, B., et al. (2018). Identification of chemotherapy-induced Leukemic-Regenerating Cells reveals a transient vulnerability of human AML recurrence. Cancer Cell *34*, 483–498.e5.

Brown, J.R. (2018). How I treat CLL patients with ibrutinib. Blood *131*, 379–386.

Bulian, P., Bomben, R., Bo, M.D., Zucchetto, A., Rossi, F.M., Degan, M., Pozzo, F., Bittolo, T., Bravin, V., D'Agaro, T., et al. (2017). Mutational status of *IGHV* is the most reliable prognostic marker in trisomy 12 chronic lymphocytic leukemia. Haematologica *102*, e443–e446.

Chao, M.P., Gentles, A.J., Chatterjee, S., Lan, F., Reinisch, A., Corces, M.R., Xavy, S., Shen, J., Haag, D., Chanda, S., et al. (2017). Human AML-iPSCs Reacquire Leukemic Properties after Differentiation and Model Clonal Variation of Disease. Cell Stem Cell *20*, 329–344.e7.

Chilloux, J., Brial, F., Everard, A., Smyth, D., Zhang, L., Plovier, H., Myridakis, A., Hoyles, L., Fuchs, J.E., Blancher, C., et al. (2018). Microbiome Inhibition of IRAK-4 by Trimethylamine Mediates Metabolic and Immune Benefits in High-Fat-Diet-induced Insulin Resistance. bioRxiv. https://doi.org/10.1101/277434.

Cutrona, G., Tripodo, C., Matis, S., Recchia, A.G., Massucco, C., Fabbi, M., Colombo, M., Emionite, L., Sangaletti, S., Gulino, A., et al. (2018). Microenvironmental regulation of the IL-23R/IL-23 axis overrides chronic lymphocytic leukemia indolence. Sci. Transl. Med. *10*, eaal1571.

Decker, S., Zirlik, K., Djebatchie, L., Hartmann, D., Ihorst, G., Schmitt-Graeff, A., Herchenbach, D., Jumaa, H., Warmuth, M., Veelken, H., and Dierks, C. (2012). Trisomy 12 and elevated GLI1 and PTCH1 transcript levels are biomarkers for Hedgehog-inhibitor responsiveness in CLL. Blood *119*, 997–1007.

Del Giudice, I., Rossi, D., Chiaretti, S., Marinelli, M., Tavolaro, S., Gabrielli, S., Laurenti, L., Marasca, R., Rasi, S., Fangazio, M., et al. (2012). NOTCH1 mutations in +12 chronic lymphocytic leukemia (CLL) confer an unfavorable prognosis, induce a distinctive transcriptional profiling and refine the intermediate prognosis of +12 CLL. Haematologica *97*, 437–441.

Döhner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Kröber, A., Bullinger, L., Döhner, K., Bentz, M., and Lichter, P. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. N. Engl. J. Med. *343*, 1910–1916.

Draper, J.S., Smith, K., Gokhale, P., Moore, H.D., Maltby, E., Johnson, J., Meisner, L., Zwaka, T.P., Thomson, J.A., and Andrews, P.W. (2004). Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. Nat. Biotechnol. *22*, 53–54.

Fiorcari, S., Benatti, S., Zucchetto, A., Zucchini, P., Gattei, V., Luppi, M., Marasca, R., and Maffei, R. (2019). Overexpression of CD49d in trisomy 12 chronic lymphocytic leukemia patients is mediated by IRF4 through induction of IKAROS. Leukemia *33*, 1278–1302.

Fischer, K., Al-Sawaf, O., Bahlo, J., Fink, A.M., Tandon, M., Dixon, M., Robrecht, S., Warburton, S., Humphrey, K., Samoylova, O., et al. (2019). Venetoclax and Obinutuzumab in Patients with CLL and Coexisting Conditions. N. Engl. J. Med. *380*, 2225–2236.

Fraser, G., Cramer, P., Demirkan, F., Silva, R.S., Grosicki, S., Pristupa, A., Janssens, A., Mayer, J., Bartlett, N.L., Dilhuydy, M.-S., et al. (2019). Updated results from the phase 3 HELIOS study of ibrutinib, bendamustine, and rituximab in relapsed chronic lymphocytic leukemia/small lymphocytic lymphoma. Leukemia *33*, 969–980.

French, A., Yang, C.T., Taylor, S., Watt, S.M., and Carpenter, L. (2015). Human induced pluripotent stem cell-derived B lymphocytes express sIgM and can be generated via a hemogenic endothelium intermediate. Stem Cells Dev. *24*, 1082–1095.

Gaiti, F., Chaligne, R., Gu, H., Brand, R.M., Kothen-Hill, S., Schulman, R.C., Grigorev, K., Risso, D., Kim, K.-T., Pastore, A., et al. (2019). Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. Nature *569*, 576–580.

Hallek, M. (2019). Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment. Am. J. Hematol. *94*, 1266–1287.

Hallek, M., Cheson, B.D., Catovsky, D., Caligaris-Cappio, F., Dighiero, G., Döhner, H., Hillmen, P., Keating, M.J., Montserrat, E., Rai, K.R., and Kipps, T.J.; International Workshop on Chronic Lymphocytic Leukemia (2008). Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood *111*, 5446–5456.

Hallek, M., Cheson, B.D., Catovsky, D., Caligaris-Cappio, F., Dighiero, G., Döhner, H., Hillmen, P., Keating, M., Montserrat, E., Chiorazzi, N., et al. (2018). iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood *131*, 2745–2760.

Hentze, H., Soong, P.L., Wang, S.T., Phillips, B.W., Putti, T.C., and Dunn, N.R. (2009). Teratoma formation by human embryonic stem cells: Evaluation of essential parameters for future safety studies. Stem Cell Res *2*, 198–210.

Herman, S.E., Gordon, A.L., Wagner, A.J., Heerema, N.A., Zhao, W., Flynn, J.M., Jones, J., Andritsos, L., Puri, K.D., Lannutti, B.J., et al. (2010). Phosphatidylinositol 3-kinase-δ inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. Blood *116*, 2078–2088.

Herszfeld, D., Wolvetang, E., Langton-Bunker, E., Chung, T.L., Filipczyk, A.A., Houssami, S., Jamshidi, P., Koh, K., Laslett, A.L., Michalska, A., et al. (2006). CD30 is a survival factor and a biomarker for transformed human pluripotent stem cells. Nat. Biotechnol. *24*, 351–357.

Hillmen, P., Rawstron, A.C., Brock, K., Muñoz-Vicente, S., Yates, F.J., Bishop, R., Boucher, R., MacDonald, D., Fegan, C., McCaig, A., et al. (2019). Ibrutinib Plus Venetoclax in Relapsed/Refractory Chronic Lymphocytic Leukemia: The CLARITY Study. J. Clin. Oncol. *37*, 2722–2729.

Honigberg, L.A., Smith, A.M., Sirisawad, M., Verner, E., Loury, D., Chang, B., Li, S., Pan, Z., Thamm, D.H., Miller, R.A., and Buggy, J.J. (2010). The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. Proc. Natl. Acad. Sci. USA *107*, 13075–13080.

Hu, B., Patel, K.P., Chen, H.C., Wang, X., Luthra, R., Routbort, M.J., Kanagal-Shamanna, R., Medeiros, L.J., Yin, C.C., Zuo, Z., et al. (2019). Association of gene mutations with time-to-first treatment in 384 treatment-naive chronic lymphocytic leukaemia patients. Br. J. Haematol. *187*, 307–318.

Imhof, A.K., Glück, L., Gajda, M., Bräuer, R., Schaible, H.G., and Schulz, S. (2011). Potent anti-inflammatory and antinociceptive activity of the endothelin receptor antagonist bosentan in monoarthritic mice. Arthritis Res. Ther. *13*, R97.

International Stem Cell Initiative; Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B., Beil, S., et al. (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nat. Biotechnol. *29*, 1132–1144.

Kater, A.P., and Brown, J.R. (2019). Ibrutinib: searching for a partner drug. Lancet Oncol. *20*, 3–5.

Kikushige, Y., Ishikawa, F., Miyamoto, T., Shima, T., Urata, S., Yoshimoto, G., Mori, Y., Iino, T., Yamauchi, T., Eto, T., et al. (2011). Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. Cancer Cell *20*, 246–259.

Kotini, A.G., Chang, C.-J., Chow, A., Yuan, H., Ho, T.-C., Wang, T., Vora, S., Solovyov, A., Husser, C., Olszewska, M., et al. (2017). Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia. Cell Stem Cell *20*, 315–328.e7.

Landau, D.A., Tausch, E., Taylor-Weiner, A.N., Stewart, C., Reiter, J.G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Böttcher, S., et al. (2015). Mutations driving CLL and their evolution in progression and relapse. Nature *526*, 525–530.

Lee, J.-H., Salci, K.R., Reid, J.C., Orlando, L., Tanasijevic, B., Shapovalova, Z., and Bhatia, M. (2017). Brief Report: Human Acute Myeloid Leukemia Reprogramming to Pluripotency Is a Rare Event and Selects for Patient Hematopoietic Cells Devoid of Leukemic Mutations. Stem Cells *35*, 2095–2102.

Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell *177*, 1022–1034.e6.

Ljungars, A., Mårtensson, L., Mattsson, J., Kovacek, M., Sundberg, A., Tornberg, U.C., Jansson, B., Persson, N., Emruli, V.K., Ek, S., et al. (2018). A platform for phenotypic discovery of therapeutic antibodies and targets applied on Chronic Lymphocytic Leukemia. NPJ Precis. Oncol. *2*, 18.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al.; Cancer Genome Atlas Research Network (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell *173*, 338–354.e15.

Martinelli, S., Maffei, R., Fiorcari, S., Quadrelli, C., Zucchini, P., Benatti, S., Potenza, L., Luppi, M., and Marasca, R. (2017). The expression of endothelin-1 in chronic lymphocytic leukemia is controlled by epigenetic mechanisms and extracellular stimuli. Leuk. Res. *54*, 17–24.

Matas-Céspedes, A., Vidal-Crespo, A., Rodriguez, V., Villamor, N., Delgado, J., Giné, E., Roca-Ho, H., Menéndez, P., Campo, E., López-Guillermo, A., et al. (2017). The Human CD38 Monoclonal Antibody Daratumumab Shows Antitumor Activity and Hampers Leukemia-Microenvironment Interactions in Chronic Lymphocytic Leukemia. Clin. Cancer Res. *23*, 1493–1505.

Mitelman, F., Johansson, B., and Mertens, F. (2003). Mitelman Database of Chromosome Aberrations in Cancer. https://mitelmandatabase.isb-cgc.org/.

Morabito, F., Mosca, L., Cutrona, G., Agnelli, L., Tuana, G., Ferracin, M., Zagatti, B., Lionetti, M., Fabris, S., Maura, F., et al. (2013). Clinical monoclonal B lymphocytosis versus Rai 0 chronic lymphocytic leukemia: a comparison of cellular, cytogenetic, molecular, and clinical features. Clin. Cancer Res. *19*, 5890–5900.

Muñoz-López, A., Romero-Moya, D., Prieto, C., Ramos-Mejía, V., Agraz-Doblas, A., Varela, I., Buschbeck, M., Palau, A., Carvajal-Vergara, X., Giorgetti, A., et al. (2016). Development Refractoriness of MLL-Rearranged Human B Cell Acute Leukemias to Reprogramming into Pluripotency. Stem Cell Reports *7*, 602–618.

Muñoz-Novas, C., Poza-Santaella, M., González-Gascón Y Marín, I., Hernández-Sánchez, M., Rodríguez-Vicente, A.E., Infante, M.S., Heras, C., Foncillas, M.A., Marín, K., Hernández-Rivas, J.M., and Hernández, J.Á. (2018). The International Prognostic Index for Patients with Chronic Lymphocytic Leukemia Has the Higher Value in Predicting Overall Outcome Compared with the Barcelona-Brno Biomarkers Only Prognostic Model and the MD Anderson Cancer Center Prognostic Index. BioMed Res. Int. *2018*, 9506979.

Nakanishi, M., Mitchell, R.R., Benoit, Y.D., Orlando, L., Reid, J.C., Shimada, K., Davidson, K.C., Shapovalova, Z., Collins, T.J., Nagy, A., and Bhatia, M. (2019). Human Pluripotency Is Initiated and Preserved by a Unique Subset of Founder Cells. Cell *177*, 910–924.e22.

Papaemmanuil, E., Gerstung, M., Bullinger, L., Gaidzik, V.I., Paschka, P., Roberts, N.D., Potter, N.E., Heuser, M., Thol, F., Bolli, N., et al. (2016). Genomic Classification and Prognosis in Acute Myeloid Leukemia. N. Engl. J. Med. *374*, 2209–2221.

Parikh, S.A. (2018). Chronic lymphocytic leukemia treatment algorithm 2018. Blood Cancer J. *8*, 93.

Parikh, S.A., Strati, P., Tsang, M., West, C.P., and Shanafelt, T.D. (2016). Should IGHV status and FISH testing be performed in all CLL patients at diagnosis? A systematic review and meta-analysis. Blood *127*, 1752–1760.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Rai, K.R., Sawitsky, A., Cronkite, E.P., Chanana, A.D., Levy, R.N., and Pasternack, B.S. (1975). Clinical staging of chronic lymphocytic leukemia. Blood *46*, 219–234.

Reid, J.C., Tanasijevic, B., Golubeva, D., Boyd, A.L., Porras, D.P., Collins, T.J., and Bhatia, M. (2018). CXCL12/CXCR4 Signaling Enhances Human PSC-Derived Hematopoietic Progenitor Function and Overcomes Early In Vivo Transplantation Failure. Stem Cell Reports *10*, 1625–1641.

Rhyasen, G.W., and Starczynowski, D.T. (2015). IRAK signalling in cancer. Br. J. Cancer *112*, 232–237.

Roberts, A.W., Davids, M.S., Pagel, J.M., Kahl, B.S., Puvvada, S.D., Gerecitano, J.F., Kipps, T.J., Anderson, M.A., Brown, J.R., Gressick, L., et al. (2016). Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. N. Engl. J. Med. *374*, 311–322.

Roos-Weil, D., Nguyen-Khac, F., Chevret, S., Touzeau, C., Roux, C., Lejeune, J., Cosson, A., Mathis, S., Feugier, P., Leprêtre, S., et al.; FILO working group (2018). Mutational and cytogenetic analyses of 188 CLL patients with trisomy 12: A retrospective study from the French Innovative Leukemia Organization (FILO) working group. Genes Chromosomes Cancer *57*, 533–540.

Rosanò, L., Spinella, F., and Bagnato, A. (2013). Endothelin 1 in cancer: biological implications and therapeutic opportunities. Nat. Rev. Cancer *13*, 637–651.

Séguin, C.A., Draper, J.S., Nagy, A., and Rossant, J. (2008). Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. Cell Stem Cell *3*, 182–195.

Serafim, K.G., Navarro, S.A., Zarpelon, A.C., Pinho-Ribeiro, F.A., Fattori, V., Cunha, T.M., Alves-Filho, J.C., Cunha, F.Q., Casagrande, R., and Verri, W.A., Jr. (2015). Bosentan, a mixed endothelin receptor antagonist, inhibits superoxide anion-induced pain and inflammation in mice. Naunyn Schmiedebergs Arch. Pharmacol. *388*, 1211–1221.

Smith, L.D., Minton, A.R., Blunt, M.D., Karydis, L.I., Dutton, D.A., Rogers-Broadway, K.R., Dobson, R., Liu, R., Norster, F., Hogg, E., et al. (2020). BCR signaling contributes to autophagy regulation in chronic lymphocytic leukemia. Leukemia *34*, 640–644.

Sokolov, A., Carlin, D.E., Paull, E.O., Baertsch, R., and Stuart, J.M. (2016). Pathway-Based Genomics Prediction using Generalized Elastic Net. PLoS Comput. Biol. *12*, e1004790.

Souers, A.J., Leverson, J.D., Boghaert, E.R., Ackler, S.L., Catron, N.D., Chen, J., Dayton, B.D., Ding, H., Enschede, S.H., Fairbrother, W.J., et al. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nat. Med. *19*, 202–208.

Stewart, B.W., and Wild, C.P. (2014). Haematopoietic and lymphoid malignancies. In World Cancer Report 2014, B.W. Stewart and C.P. Wild, eds. (World Health Organization), pp. 482–494.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Van Dyke, D.L., Werner, L., Rassenti, L.Z., Neuberg, D., Ghia, E., Heerema, N.A., Dal Cin, P., Dell Aquila, M., Sreekantaiah, C., Greaves, A.W., et al. (2016). The Dohner fluorescence in situ hybridization prognostic classification of chronic lymphocytic leukaemia (CLL): the CLL Research Consortium experience. Br. J. Haematol. *173*, 105–113.

Vendramini, E., Bomben, R., Pozzo, F., Benedetti, D., Bittolo, T., Rossi, F.M., Dal Bo, M., Rabe, K.G., Pozzato, G., Zaja, F., et al. (2019). KRAS, NRAS, and BRAF mutations are highly enriched in trisomy 12 chronic lymphocytic leukemia and are associated with shorter treatment-free survival. Leukemia *33*, 2111–2115.

Vodyanik, M.A., Bork, J.A., Thomson, J.A., and Slukvin, I.I. (2005). Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood *105*, 617–626.

Werbowetski-Ogilvie, T.E., Bossé, M., Stewart, M., Schnerch, A., Ramos-Mejia, V., Rouleau, A., Wynder, T., Smith, M.J., Dingwall, S., Carter, T., et al. (2009). Characterization of human embryonic stem cells with features of neoplastic progression. Nat. Biotechnol. *27*, 91–97.

Zucchetto, A., Caldana, C., Benedetti, D., Tissino, E., Rossi, F.M., Hutterer, E., Pozzo, F., Bomben, R., Dal Bo, M., D'Arena, G., et al. (2013). CD49d is overexpressed by trisomy 12 chronic lymphocytic leukemia cells: evidence for a methylation-dependent regulation mechanism. Blood *122*, 3317–3321.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Alexa Fluor 488 mouse anti-TRA-1-60	BD Biosciences	Cat#560173; RRID:AB_1645379
Alexa Fluor 647 mouse anti-TRA-1-60	BD Biosciences	Cat#560122; RRID:AB_1645448
APC mouse anti-CD34	BD Biosciences	Cat#555824; RRID:AB_398614
mouse anti-CD19 microbeads	Miltenyi	Cat#130-050-301; RRID:AB_2848166
mouse anti-CD34 microbeads	Miltenyi	Cat#130-046-702; RRID:AB_2848167
Donkey anti-mouse Alexa 647	BD Biosciences	Cat#A31571; RRID:AB_2536180
FITC mouse anti-CD19	BD Biosciences	Cat#555412; RRID:AB_395812
PE mouse anti-OCT4	BD Biosciences	Cat#560186; RRID:AB_1645331
PE mouse anti-IRAK4	BD Biosciences	Cat#560303; RRID:AB_1645518
PE mouse anti-Ki67	BD PharMingen	Cat#556027; RRID:AB_2266296
PE-Cy7 mouse anti-KIT	BD Biosciences	Cat#339195; RRID:AB_647418
mouse anti-EDNRB purified	R&D Systems	Cat#AF4496-SP; RRID:AB_2096682
v450 mouse anti-CD45	BD Biosciences	Cat#642275; RRID:AB_1645755
Biological samples		
Primary CLL and MBL patient samples, #1-159	This paper and Morabito et al. (2013)	N/A
Primary CLL patient samples, #160-161	Juravinksi Hospital and Cancer Centre	N/A
Primary CLL patient samples, #162-166, 173-177	Obtained under MTA from The Manitoba Tumor Bank	N/A
Primary CLL patient samples, #167-172, 178-203	Obtained under MTA from The Quebec Leukemia Cell Bank	N/A
Healthy human donor hematopoietic samples	Juravinksi Hospital and Cancer Centre	N/A
Xenotransplant samples	This paper	N/A
Teratoma sections	This paper	N/A
Chemicals, peptides, and recombinant proteins		
7AAD	Beckman Coulter	Cat#A07704
Live/Dead Fixable Violet Dead Cell Stain	Invitrogen	Cat#L34955
Annexin-FITC, 1:100	BD Biosciences	Cat#556419; RRID:AB_2665412
AIM-V Serum Free Medium	ThermoFisher Scientific	Cat#12055083
Hoechst 33342	ThermoFisher Scientific	Cat#561908
CellTrace Calcein Green, AM	ThermoFisher Scientific	Cat#C34852
LS Columns	Miltenyi	Cat#130-042-401
Stem Cell Factor (SCF)	R&D Systems	Cat#255-SC/CF
Fms-related tyrosine kinase 3 ligand	R&D Systems	Cat#308-Fk/CF
Interleukin 3 (IL-3)	Miltenyi	Cat#130-095-069
Interleukin 7 (IL-7)	PeproTech	Cat#200-07-50UG
mouse embryonic feeder (MEF)-conditioned media (MEF-CM)	This paper	N/A
basic fibroblast growth factor	Corning	Cat#TXC359282
Fetal bovine serum, Canada	HyClone	
Bosentan	R&D Systems	Cat#6232/50; CAS:147536-97-8
PF06650833	Ontario Chemicals Inc	CAS:1817626-54-2

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Prestwick Chemical Library	PerkinElmer, Guelph, Ontario	https://www.prestwickchemical.com/screening-libraries/prestwick-phytochemical-library/
OICR Kinase Inhibitor Library	Dr. Rima Al-war, Ontario Institute of Cancer Research	N/A
Critical commercial assays		
Fixation/Permeabilization solution kit	BD Cytofix/Cytoperm	Cat#554714
DNeasy blood and tissue kit	QIAGEN	Cat#69506
RNeasy Micro Kit	QIAGEN	Cat#74004
Affymetrix Human Gene 2.0 ST microarray	London Regional Genomics Centre, London, Ontario, Canada	N/A
Karyotyping	The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada	N/A
hPSC Genetic Analysis Kit	STEMCELL Technologies	Cat#07550
Cytoscan HD microarray	The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada	N/A
Paraffin processing and embedding, hematoxylin and eosin (H&E) slide staining	The MIRC Histology Core Facility, McMaster University, Ontario, Canada	N/A
blood analyzer	scil animal care company, Barrie, Ontario, Canada	scil Vet abc Plus+
Deposited data		
Microarray raw data: CA2 embryonic stem cells	This paper	GSE133697
Microarray raw data: CA2 embryonic stem cells, NoCre	Séguin et al., 2008	GSE10809
Microarray raw data: leukemic peripheral blood CD19+ cells from 159 patients	Morabito et al., 2013	GSE40570
Experimental models: cell lines		
normal karyotype CA2 embryonic stem cells	Obtained under MTA from Dr. A. Nagy, Lunenfeld-Tanenbaum Research Institute, Toronto, Ontario	hPSCreg Name: MSHRIe002-A
trisomy 12 CA2 embryonic stem cells	This paper	N/A
OP9 mouse bone marrow stromal cells	ATCC	Cat#CRL-2749
MS-5 mouse bone marrow stromal cells	DSMZ	Cat#ACC 441
Experimental models: organisms/strains		
NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG)	The Jackson laboratory	RRID:IMSR_ARC:NSG. JAX:05557
Software and algorithms		
Code Availability	All code used in this study	https://github.com/reidjco
FACSDiva	BD Biosciences	https://www.bdbiosciences.com/us/instruments/research/software/flow-cytometry-acquisition
FlowJo v10.6.2	FlowJo, LLC	https://www.flowjo.com
Prism v7.0a	Graphpad	https://www.graphpad.com
Drug Gene Interaction Database v2.22	Griffith lab	https://www.dgidb.org/
Volocity	PerkinElmer	https://www.perkinelmer.com/lab-products-and-services/resources/whats-new-volocity-6-3.html
Columbus Image Data Storage and Analysis System v2.9.1	PerkinElmer	https://www.perkinelmer.com/product/image-data-storage-and-analysis-system-columbus

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Genomics Suite v6.6 software	Partek Inc.	https://www.partek.com/partek-genomics-suite/
GSEA vs2.1.0	Broad Institute	https://www.gsea-msigdb.org/gsea/downloads.jsp
STRING v10.5	STRING Consortium	https://string-db.org/
R software v4.0.0	The R Foundation for Statistical Computing	https://www.r-project.org/
RStudio v1.2.5042	RStudio Inc.	https://rstudio.com/
R package survival	CRAN	https://cran.r-project.org/web/packages/survival/index.html
R package survminer	CRAN	https://cran.r-project.org/web/packages/survminer/index.html
Python software v3.8.2	Python Software Foundation	https://www.python.org/
JupyterLab Notebook	JupyterLab	https://jupyterlab.readthedocs.io/en/stable/
Python module itertools v2.3	Python Software Foundation	https://docs.python.org/3/library/itertools.html
Python library matplotlib v3.2	Matplotlib	https://matplotlib.org/stable/index.html
Python library numpy v1.18.1	NumPy	https://numpy.org/
Python library pandas v1.0.1	pandas	https://pandas.pydata.org/
Python library seaborn v0.10.0	seaborn	https://seaborn.pydata.org/
Python library Scikit-learn (sklearn) v0.23	Pedregosa et al., 2011	https://scikit-learn.org/stable/
Chromosome Analysis Suite v3.1	ThermoFisher Scientific	https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/chromosome-analysis-suite.html
One-class logistic regression (OCLR)	Malta et al., 2018	https://bioinformaticsfmrp.github.io/PanCanStem_Web/
FCS Express v5	De Novo Software	https://denovosoftware.com/installation-instructions/installing-previous-versions/
ImageJ	NIH	https://imagej.nih.gov/ij/
Power of Analysis, Stat UBC	Rollin Brant, University of British Columbia	https://www.stat.ubc.ca/~rollin/stats/ssize/n2.html
Power of Analysis, ClinCalc	Kane SP. Sample Size Calculator.	https://clincalc.com/stats/samplesize.aspx

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Mickie Bhatia, (mbhatia@mcmaster.ca).

Materials availability

PB MNC from CLL patients were obtained through Material Transfer Agreements (MTAs) from both the Quebec Leukemia Cell Bank and Manitoba Tumor Bank. CA2 human embryonic stem cells were obtained by MTA from Dr. A. Nagy, Lunenfeld-Tanenbaum Research Institute, Toronto, Ontario. There are restrictions to the availability of variant CA2 hESCs due to a requirement for an MTA.

Data and code availability

Analysis packages using code in Python and R software are detailed in the key resources table and the complete computational analysis workflows are publicly available on GitHub in the form of R scripts and Python code at https://github.com/reidjco. The accession numbers for datasets are as follows: original data presented in this study (GEO accession: GSE133697), CA2 from initial study

(GEO accession: GSE10809, CA2 NoCre), and our previous study with human patients with Rai 0 clinical MBL and Rai 0 CLL diagnosis PB samples (GEO accession: GSE40570).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study Design

Due to the lack of genetic models for trisomy 12 CLL with suitable biological controls (Table S1), the objective of this study was to investigate the effect of tri12 on cell autonomous signaling in human PSC and CLL and identify druggable targets. The two lead compounds discovered were EDNRB and IRAK4, where IRAK4 was more potent. Three models were used to address whether IRAK4 and EDNRB affect tri12 CLL: (i) using transcriptomics of purified B cells from CLL patients, (ii) *in vitro* chemical treatment, and (iii) *in vivo* chemical treatment using PDX mice. *Sample size*: Power of Analysis using two different calculations (Stat UBC and ClinCalc), determined 5 patient samples per group (tri12 and normal) would substantiate our drug testing claims (power = 0.95 and alpha = 0.05). This was determined using the GI50 and variance values of tri12 and normal CLL patient B cells (CLL#160 and CLL#161). In total, we screened N = 12 tri12 CLL patient B cell samples, and N = 18 $FISH^{NEG}$ CLL patient B cell samples, and N = 14 CLL patient B cell samples with 1 to 3 deletions in 11q, 13q, or 17p chromosomal regions (Table 1). *Rules for stopping data collection*: No experiments were stopped prematurely. *Data inclusion/exclusion criteria*: Patient data and samples were included based on diagnostic criteria set by the International Workshop on CLL (IWCLL) guidelines, where the diagnosis of CLL requires a circulating B-lymphocyte count greater than or equal to $5x10^9$ cells/L in the peripheral blood (PB) for a duration of at least 3 months (Hallek et al., 2008). The IWCLL guidelines have been updated since the original report (Hallek et al., 2018). No patient data were excluded. Cytogenetic classification was performed using the standard hierarchical classification with fluorescence *in situ* hybridization (FISH) probes for: tri12, del(13q), del(11q), del(17p) (Döhner et al., 2000). *Outliers*: No outliers were removed. *Selection of endpoints*: Drug screening was completed in 4 days based on our previous screening platform (Nakanishi et al., 2019). Mice were killed 7 weeks after injection of hPSCs due to large tumor size, and consistent with our previous study (Nakanishi et al., 2019). CLL PDX mice were killed 8 days or 15 days after transplant. *Replicates*: Independent experiments were completed at a different time with different reagents. 'n' biological replicates represent testing in separate wells completed concurrently but analyzed separately. *Research subjects or units of investigation*: 'n' number of patient samples tested indicates different people. 'nN' number of mice tested indicates different mice. *Experimental design*: controlled laboratory experiments. *Randomization*: Mice were assigned randomly to the various experimental groups, based on weight and sex, and distributed among cages. *Blinding*: Clinical colleagues were blinded to the study purpose and results when providing patient samples for analysis. Animal caretakers were blinded to the allocation of mice and for endpoint monitoring. Investigators who assessed, measured, and quantified the results were not blinded to the intervention, with the exception of drug screening analysis which was completed with coded ID's that were revealed after analysis.

Patient samples

Diagnostic criteria for CLL patients was based on the 2008 International Workshop on CLL (IWCLL) guidelines, where the diagnosis of CLL requires a circulating B-lymphocyte count greater than or equal to $5x10^9$ cells/L in the peripheral blood (PB) for a duration of at least 3 months 4 (Hallek et al., 2008). The IWCLL guidelines have been updated since the original report (Hallek et al., 2018). Tier 1 BRISQ reporting guidelines were used to describe CLL patient characteristics including age and sex (Table 1; Table S4). Heparinized PB was obtained from consenting CLL patients and healthy volunteers (Juravinksi Hospital and Cancer Centre, Hamilton, Ontario), and protocols were approved by the McMaster Research Ethics Board. Mononuclear cells (MNC) were collected and isolated as previously described (Reid et al., 2018), cryopreserved in 10% DMSO and stored in liquid nitrogen until analysis (5 years, −150°C). PB MNC from CLL patients were also provided by the Quebec Leukemia Cell Bank as well as the Manitoba Tumor Bank, which were cryopreserved in 10% DMSO and stored in liquid nitrogen (9 ± 6 years), transported overnight to McMaster University on dry ice (−109°C), and then returned to liquid nitrogen storage until analysis (< 1 month). CLL cells were thawed and enriched for CD19+ cells if below 90% CD19+ (Miltenyi, cat# 130-050-301), cultured at a density of $1.0x10^5$/mL in serum-free AIM-V medium (GIBCO) with β-mercaptoethanol in 96-well optical glass plates (Corning) at 37°C in 5% CO_2. Drug testing was conducted for 4 days, with drugs and doses indicated in the figures, conducted similarly to previously established *in vitro* treatment of primary CLL patient cells (Table S1).

Human PSC culture

Human embryonic stem cell (hESC) research received Canadian Stem Cell Oversight Committee (SCOC; Canadian Institutes of Health Research, CIHR) approval, and Research Ethics and Biohazard Utilization Protocols approval at McMaster University, following the principles of the 2016 ISSCR Guidelines for Stem Cell Research and Clinical Applications of Stem Cells. CA2 hESC were cultured in feeder-free conditions in mouse embryonic feeder (MEF)-conditioned media (MEF-CM) with 8 ng/mL basic fibroblast growth factor (bFGF) as previously described (Reid et al., 2018). The media we used for cell culture of hESCs creates an OCT4 negative population which allows for microenvironmental cues as well as differentiation priming, which we have previously reported (Bendall et al., 2007). Interestingly and inexplicably, our tri12 PSCs will not expand in mTeSR and equivalent media, despite many attempts to have them do so (data not shown). Therefore, we used one media which permitted the tri12 and normal PSCs to grow similarly to each other.

Human pluripotent reprogramming

CD19 positive cells were purified from CLL#160 sample, which tested positive for tri12. CD19 negative cells were retained. Cells were transduced with the CytoTune iPS 2.0 Sendai Reprogramming Kit (ThermoFisher Scientific) using manufacturer's protocol, while keeping the cells in AIM-V media for 48 hours at a density of $2.0x10^5$ per 6-well. $1x10^6$ and $2.0x10^5$ CD19 positive and negative cells were transduced, respectively (a five-fold difference). Cells were then washed and cultured on top of a irradiated MEF (iMEF) monolayer. Colonies were observed and individually transferred to new wells with iMEF monolayers, 3 weeks after transduction. At passage 2, 3-5 colonies were manually picked for hPSC genetic analysis, as described in the 'Cytogenetics' section below.

Mouse strain used for transplantation experiments

Immunodeficient NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice were bred in a barrier facility and all experimental protocols were approved by the Animal Research Ethics Board of McMaster University. NSG mice were originally purchased from Jackson Laboratories (Bar Harbor, Maine, USA). NSG mouse sex and age were controlled within each experiment, and mice were randomly assigned to experimental groups which included male and female mice and were over 8 weeks old.

METHOD DETAILS

Cytogenetics

CLL patient sample cytogenetics were tested using the standard hierarchical classification with fluorescence *in situ* hybridization (FISH) probes for: tri12, del(13q), del(11q), del(17p) (Döhner et al., 2000). Additional karyotyping and comparative genomic hybridization using Cytoscan HD Array (ThermoFisher Scientific) were performed by The Centre for Applied Genomics, at The Hospital for Sick Children, Toronto, Ontario, Canada. Cytoscan HD Array analysis was performed using Chromosome Analysis Suite (NetAffx 33.1, hg19) and default settings. Normal CA2 human ESC were received at passage 42, and then maintained between passage number 52 and 71 (less than 5 months of continuous culture). Tri12 CA2 human ESC were first identified at passage number 127, and then maintained between passage numbers 127 and 163 for all experiments, except for the following exception: we continuously cultured tri12 PSC up to passage 186 (a total of 59 weeks (> 1 year) of continuous culture) following its original discovery in order to analyze genetic stability, which is reported in Table S2. iPSC colonies derived from CLL#160 were analyzed for chromosomal copy number in house with the hPSC Genetic Analysis Kit (STEMCELL Technologies) following manufacturer's protocol, in 384-well plate format with a ViiA 7 Real-Time PCR System instrument (ThermoFisher Scientific).

Colony initiating cell assay

Colony initiating cell (CIC) assay was completed with a FACSAria II (BD); 1000 TRA-1-60+ and TRA-1-60– live cells were sorted into 6-well plates with irradiated MEF cells (iMEFs, seeded 24 h in advance), in MEF-CM with 8 ng/mL bFGF. For Oct4 staining, cells were incubated with Live/Dead Fixable Violet Dead Cell Stain 1:7000 (Invitrogen) for 30 min at 4°C, washed, fixed with Fix/Perm for 20 min at RT, washed with 1X Perm/Wash, incubated with Oct4-PE 1:100 (all from BD Biosciences) overnight, and washed three times prior to whole well images acquired via automated high-content confocal fluorescence microscopy (Operetta, PerkinElmer). For all experiments, tri12 and normal hPSCs were tested in parallel.

B cell differentiation

B cell differentiation was adapted from a previous report (Vodyanik et al., 2005). TC-treated 10 cm dishes were coated with 0.1% gelatin (Sigma) for 1 h at 37°C prior to seeding stromal cells. MS5 cells were irradiated (iMS5) to prevent growth and cryopreserved until use. OP9 cells were thawed and seeded at $4.0x10^5$ cells per 10 cm dish one day prior to hPSC harvest, in OP9 media: αMEM supplemented with 10% FBS (HyClone), 1 mM non-essential amino acids (GIBCO), 1 mM L-glutamine (GIBCO), and 100 μM monothioglycerol (Sigma). Human PSCs were seeded as clumps onto OP9 cells, half of one 6-well of hPSC per 10 cm dish. Half media changes (10 mL) were performed on day 2 and 4 with OP9 media. One day prior to day 8 harvest, iMS5 cells were thawed and seeded at a density of $8.0x10^5$ per 10 cm dish in iMS5 media; the same as OP9 media except FBS was previously heat inactivated (2 h at 56°C). Cells were harvested on day 8 by 20 min collagenase IV and 10 min 0.05% trypsin, keeping all non-adherent cells throughout. Day 8 single cells were stained with CD34 microbeads (Miltenyi) and enriched with LS columns following manufacturer's protocol. Purity of isolated CD34+ cells was > 80%. CD34-enriched cells were seeded on to iMS5 cells at a density of $3.0x10^5$ per 10 cm dish in iMS5 media, supplemented with 50 ng/mL stem cell factor (SCF, R&D Systems), 50 ng/mL FMS-like tyrosine kinase 3 ligand (FLT3L, R&D Systems), 20 ng/mL Interleukin-7 (IL-7, PeproTech), and 10 ng/mL IL-3 (R&D Systems). 10 mL half media changes occurred every 5 days for a total of 35 days, and SCF, FLT3L, and IL-7 were added fresh each time. Day 42 cells were collected as described for day 8 and were assessed by flow cytometry (see *Flow cytometry* section).

Flow cytometry

Cells were stained using reagents in the key resources table. For intracellular staining, cells were incubated with Live/Dead Fixable Violet Dead Cell Stain 1:7000 (Invitrogen, 30 min at 37°C), stained with cell surface antibodies (if applicable, 30 min at 4°C), washed,

fixed using Fix/Perm Buffer (20 min at RT), stained with antibodies overnight in Perm/Wash Buffer, and then washed three times prior to acquisition. For IRAK4 antibody staining: we performed antibody staining overnight and then 2 hours at room temperature, as well as performing a total of 4 wash steps to remove more background signal intensity. Live cells were stained for 30 min at 4°C in PBS with 3% FBS and 1 mM EDTA (Invitrogen), washed, and analyzed with an LSRII or Aria II flow cytometers (BD Biosciences). 7AAD 1:50 (Beckman Coulter) or Live/Dead Fixable Violet Dead Cell Stain was used to exclude nonviable cells. Fluorescence minus one (FMO; if applicable, secondary without primary antibody) were used to set gates in FlowJo (TreeStar, v10). The 'modal' setting was used to display histograms of flow cytometry results in FlowJo.

High-content drug testing

96-well plates of hPSCs were stained with Live/Dead Fixable Far Red Dead Cell Stain 1:7000 (Invitrogen, 30 min at 37°C), fixed using Fix/Perm Buffer (ThermoFisher Scientific, 20 min at RT), and nuclei were stained with Hoechst 33342 1:7000 (ThermoFisher Scientific) 10 min at RT. CLL cells were stained with 1 μg/mL calcein green in AIM-V media (both from ThermoFisher Scientific, 1 hour at RT). Whole 96-wells were evaluated for cytotoxicity by automated high-content confocal fluorescence microscopy (Operetta, PerkinElmer). Acquired images were analyzed using Acapella 2.0 (PerkinElmer) and FCS Express (*De Novo* Software v5) software. We thank Dr. Rima Al-war from the Ontario Institute of Cancer Research for supplying the OICR kinase library. Additional compounds were purchased as part of a chemical library from PerkinElmer (Guelph, ON).

Xenotransplantation

NSG mouse sex and age were controlled within each experiment, and mice were randomly assigned to experimental groups which included male and female mice and were over 8 weeks old. No statistical method was used to predetermine sample size. For teratoma experiments, representative 6-wells of confluent PSCs were counted, and determined 1/2 6-wells contained $2x10^6$ cells. PSCs were then collected as clumps and pelleted individually in Eppendorf tubes. 500 μL of Matrigel (Corning) was added on ice. NSG mice were injected subcutaneously (SQ) with separate injections of tri12 and normal PSCs into the right and left flank, respectively. Tumors were not observed to migrate and stayed stationary during the entire experiment. At endpoint, tumors were measured with calipers, excised, and tumor mass was weighed. EDNRB-i (Bosentan) and PF06650833 (IRAK4-i) were administered in 10% DMSO in corn oil, at 15 mg/kg daily for 21 consecutive days. For human CLL PDX modeling, mice were sublethally irradiated (single dose of 315 cGy, ^{137}Cs) 24 h before transplant. Cells were transplanted by intravenous injection with $100x10^6$ PB MNCs as previously described (Ljungars et al., 2018), from patient CLL#160. EDNRB-i (Bosentan) and PF06650833 (IRAK4-i) were administered in 10% DMSO in corn oil, at 100 mg/kg daily for 4 or 10 consecutive days, starting on the 4th day post-transplant. Mice were harvested on day 8 or 14 post-transplant and euthanized by cervical dislocation. PB was immediately collected using trunk bleeding into K2 EDTA tubes, mixed immediately by inverting three times, and then immediately analyzed on the blood analyzer (scil Vet abc Plus+). Spleens were dissected, weighed immediately, and cells were recovered by mechanical dissociation in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 3% FBS, and 1 mM EDTA. Red blood cells were lysed using ammonium chloride. Phenotyping was analyzed using FMO gating (refer to *Flow cytometry* section). Given the significant changes in spleen size, measured as 'spleen weight divided by body weight' and termed 'spleen % body weight' (Figure 5I), we factored the chimerism level by spleen % body weight to reflect the overall contribution of human cells in the spleens of PDX mice (Figure 5J). Data were combined from all independent experiments and outliers were not excluded.

Histology

Tumors were excised from mice, weighed after draining of cysts (if any), and cut longitudinally to increase surface area for fixation in 10% formalin for 72 h. Tissues were then washed with phosphate buffered saline (PBS, then 50% ethanol for 2 h, and submitted to the MIRC Histology Core Facility (McMaster University), for standard Paraffin processing and embedding (Leica ASP300 Enclosed Tissue Processor), and hematoxylin and eosin (H&E) slide staining (Leica RM2255 Automated Microtome, Leica Autostainer XL, and CV5030 Coverslipper).

STRING protein-protein analysis

STRING is an online database consisting of protein-protein interaction networks that have been built by a consortium to bring together different types of evidence: textmining, experiments, co-expression, neighborhood, and gene fusion data etc., to find genes that may regulate each other as proteins (https://string-db.org/). We used all genes in the Tri12 gene signature (Table S5), the highest confidence setting which removes many of the 'putative' and '*in silico*' based relationships, and the setting which removes any nodes that are not connected.

QUANTIFICATION AND STATISTICAL ANALYSIS

Gene expression analysis

Total RNA was extracted from human PSCs using the RNeasy Micro Kit (QIAGEN) following manufacturer's protocol. RNA was analyzed with the Affymetrix Human Gene 2.0 ST microarray (London Regional Genomics Centre, London, Ontario,

Canada). Differential gene expression analysis was conducted using Partek Gene Suite (Partek v6.6) and included original data (GEO accession: GSE133697), as well as the initial CA2 study (GEO accession: GSE10809, CA2 NoCre). In our previous study (Morabito et al., 2013), human patients with clinical MBL and Rai 0 CLL donated PB samples, and the transcriptome of purified leukemic B cells (co-expressing CD19, CD5, and CD23) was analyzed and is publicly available (GEO accession: GSE40570). Genes were considered differentially regulated with fold change > |2|, unless indicated to be > |1.25|, and false discovery rate (FDR)-corrected $p < 0.05$. Gene set enrichment analysis (GSEA, Broad Institute) was performed with default settings.

Survival and hazard ratio analysis

Leading edge genes were individually mapped from 0 and 1 using a linear transformation which subtracted the minimum and divided by the maximum for N = 22 patients. This was done to assist with interpretation and integration of the relative expression of each gene into a single 'leading edge genes' sum, and to provide each gene the same weight. These N = 22 CLL and MBL patients (N = 9 tri12, N = 13 normal cytogenetics, all *IGHV*-unmutated) were used for survival analysis as they received a minimum of 3 years of continuous disease monitoring. Patients were stratified based on LE gene expression into two groups (above and below-mean LE gene score), and survival differences were estimated with R package 'survival' and 'survminer' (Kaplan–Meier method), fit using survfit function and plotted using ggsurvplot function. To assess the statistical significance of the survival stratification observed between patients with higher and lower LE gene expression, we computed a log-rank *P value* with survdiff function, using the R package 'survival'. To assess the statistical significance of feature coefficients in multivariate Cox Proportional Hazards model, we computed each feature's hazard ratio (HR), 95% confidence intervals, and *P value* with coxph function, and plotted with ggforest, using the R package 'survival'. A positive coefficient indicates a worse prognosis. Analysis packages and software are detailed in the key resources table and the complete computational analysis workflows are publicly available on GitHub in the form of R scripts at https://github.com/reidjco.

One-class logistic regression machine learning

To calculate a tri12 scoring index based on mRNA expression, we built a predictive model using one-class logistic regression (OCLR) as described by Malta et al. (2018), and R script was adapted from deposited code (https://bioinformaticsfmrp.github.io/PanCanStem_Web/). Briefly, to ensure compatibility across datasets we selected only genes commonly shared across platforms, and the resulting training matrix consisted of 17,318 mRNA expression values common to all samples. Data was mean-centered and then OCLR was applied to only the tri12 hPSC samples. We validated our approach using leave-one-out cross-validation, by withholding one tri12 hPSC sample in each iteration. A separate signature was then trained on the other tri12 hPSC samples and used to score the withheld sample as well as the non-tri12 samples. Each iteration was measured using the area under the curve (AUC) metric, which can be interpreted as the probability that the model correctly ranks a positive sample above a negative sample. Every withheld tri12 hPSC sample was scored higher than all the non-tri12 samples, yielding an overall AUC of 1.0. The Tri12 Score was then applied to score new samples by computing Spearman correlations between the model's weight vector and the new sample's expression profile (which were accessed from GEO, NCBI: GSE40570) (Morabito et al., 2013). The tri12 indices were then mapped from 0 and 1 using a linear transformation which subtracted the minimum and divided by the maximum. Mapping was done to assist with interpretation and integration of datasets. Analysis packages and software are detailed in the key resources table and the complete computational analysis workflows are publicly available on GitHub in the form of R scripts https://github.com/reidjco.

Principal component analysis and K-nearest neighbors machine learning

KNN predicts the class (tri12 versus not-tri12) of the sample based on its nearest neighbors. As such, KNN accuracy will improve when tri12 patients cluster more closely together and away from not-tri12 patients. To compute principal component analysis (PCA) and K-nearest neighbor (KNN), we used Python (v3), and imported the following packages: itertools (v2.3), matplotlib.pyplot and matplotlib.ticker (v3.2), numpy (v1.18.1), pandas, and seaborn; as well as: KNeighborsClassifier from sklearn.neighbors, metrics from sklearn, NullFormatter from matplotlib.ticker, PCA from sklearn.decomposition, and train_test_split from sklearn.model_selection. Sklearn refers to Scikit-learn (Pedregosa et al., 2011). We defined features sets using numpy arrays for N = 63 *IGHV*-U CLL and MBL patients (N = 17 tri12 and N = 46 not-tri12; patient annotation provided in Table S4). 'Top chr. 12 genes' refers to genes in Table S5 (gray shading) and 'all genes' refers to all genes available in global transcriptome. Two datasets were created and independently analyzed using 'top chr. 12 genes' or 'all genes' for these 63 patients. PCA was computed with 2 components using PCA(n_components = 2), fit using pca.fit, and the data was transformed using pca.transform. The percentage of variance explained by the principal components was 14.24% and 9.95% for 'all genes' (total: 24.19%) and was 38.13% and 30.95% for 'top chr. 12 genes' (total: 69.08%). PCA coordinates were exported and plotted using Prism (GraphPad v7). The K-nearest neighbor (KNN) machine learning algorithm trains on part of the dataset (training set) in order to predict the class (tri12 or not-tri12) of part of the dataset which it has not been trained on (testing set). KNN classifies a test sample using the class held by the majority of its closest neighboring samples, where the number of neighbor is set by the value k. Training and testing sets were randomly created with the same random seed in both analyses. KNN model iterated through K = 1-9, trained the model using neighbors = KNeighborsClassifier(n_neighbors = n).fit(X_train,y_train), predicted class (tri12 versus not-tri12) using y_predict = neighbors.predict(X_test), and calculated model accuracy using metrics.accuracy_score(y_test,y_predict). The accuracy values were exported, plotted using Prism, and are shown in Figure 1I. Analysis packages and software are detailed in the key resources

table and the complete computational analysis workflows are publicly available on GitHub in the form of Python code at https://github.com/reidjco.

Statistical analysis

Data are represented as means ± standard deviation (SD) of all replicates, where all data was included, and no outliers were excluded. Prism (GraphPad v7) software was used for all statistical analyses unless otherwise indicated. The criterion for statistical significance was $p < 0.05$, unless stated as more stringent at $p < 0.01$ for chemical genomic drug testing (Figure 3; Table S8). All GI50 values were calculated using log(inhibitor) versus response – Variable slope (four parameters) statistics, constrained between the values 0 (10 μM venetoclax, positive control) and 100 (0.1% DMSO, negative control). If the GI50 curve was not converged, GI50 is represented as the highest dose tested (i.e., 10 μM is $\log_{10}$ GI50 = −5). Z-factor scores for CLL patient drug testing were 0.69 ± 0.08, and for hPSC drug testing were 0.56 ± 0.05. A Z-factor of 0.5-1.0 is an excellent assay (GraphPad, Article #1153). Power of Analysis using two different calculations (Stat UBC and ClinCalc), determined 5 patient samples per group (tri12 and normal) would substantiate our drug testing claims, at a power of 0.95 and alpha 0.05 error rate. This was determined using the GI50 and variance values determined using tri12 and normal CLL patient B cells (CLL#160 and CLL#161). In total, we screened n = 12 tri12 CLL patient B cell samples, and n = 18 $FISH^{NEG}$ CLL patient B cell samples (Table 1). Additional statistics are described in figure legends.