January 2011

High-Resolution Characterization Of The Leukemic Cutaneous T-Cell Lymphoma Genome

William Lin

High-resolution characterization of the leukemic cutaneous T cell lymphoma genome

A Thesis Submitted to the
Yale University School of Medicine
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

by
William Michael Lin
2011

ACKNOWLEDGMENTS

This study was supported by the following grants and awards:

1) Argall and Anna Hull Cancer Research Award from the Yale Cancer Center

2) Skin Cancer Foundation Research Grant

3) American Skin Association Medical Student Award

4) Etta S. Chidsey Award in Cancer Research

5) Yale School of Medicine Student Fellowship.

6) This study was also supported in part by research funding from Affymetrix, Inc. to WML as part of their Collaborators in Cancer Research Program. Affymetrix had no scientific involvement in this study.

7) The authors thank Dr. Robert Tigelaar for critical review of this manuscript.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest to disclose pertaining to this study.

High-resolution characterization of the leukemic cutaneous T cell lymphoma genome

William M Lin,[1] Julie Lewis,[1] Kacie Carlson,[1] Badri Modi,[1] Renata Filler,[1] Adam Thornberg,[2] Gordon Saksena,[3] Swapna Reddy,[1] Sheila Umlauf,[2] Patrick A Oberholzer,[3,5] Maria Karpova,[4] Gad Getz,[3] Shrikant Mane,[2] Levi A Garraway[3,5] Reinhard Dummer,[4] Carole L Berger,[1] Richard L Edelson,[1] Michael Girardi.[1]

[1] Department of Dermatology, Yale University School of Medicine, New Haven, Connecticut, USA
[2] Microarray Facility, Keck Foundation, New Haven, Connecticut, USA
[3] Cancer Program, Broad Institute, Cambridge, Massachusetts, USA
[4] Department of Dermatology, University Hospital, Zurich, Switzerland
[5] Department of Oncology, Dana Farber Cancer Institute, Boston, Massachusetts, USA

ABSTRACT

Cutaneous T cell lymphoma (CTCL) is a heterogenous disease that can progress to leukemic involvement and is lethal in advanced stages. Toward a better understanding of the pathogenesis and identification of novel therapeutic targets in this malignancy of $CD4^+$ T cells, we conducted an integrative high-resolution genomic analysis combining DNA and mRNA data, along with clinical information, from 24 CTCL patients with blood involvement. We further performed a consensus analysis, totaling 108 samples, that confirms and narrows key loci to produce one of the most comprehensive views of the CTCL genome to date. The most significant regions of alteration are amplifications on 8q and 17q and deletions on 17p and chromosome 10. We also discover specific focal amplifications containing *KIT* and *EGFR*, and adjacent to *VEGFA*, raising the possibility of targeted therapeutic interventions. Correlating with clinical phenotypes, we determine 17q25.1 amplification's association with resistance to

skin disease improvement and nominate candidate targets in this locus. Finally, we identify a gene expression signature of immunosuppression among patients who developed infections and secondary malignancies. Overall, we determine patterns in the CTCL genome with implications for pathogenesis and immunosuppression that may focus future therapeutic developments and the genetic classification of CTCL.

TABLE OF CONTENTS

1) Introduction 6

2) Statement of Purpose and Specific Aims 12

3) Materials and Methods 13

4) Results 21

5) Discussion 27

6) References 34

7) Figures and Legends 43

8) Tables 45

INTRODUCTION

Cutaneous T cell lymphoma (CTCL) is a clinically heterogeneous malignancy of $CD4^+$ skin-homing T cells. Mycosis fungoides and Sézary syndrome represent the most common forms in the spectrum of CTCL (1, 2). First described nearly 200 years ago by French physician Jean Louis Alibert, mycosis fungoides can present as erythematous flat patches, thin plaques or tumors. These lesions tend to afflict areas of the skin infrequently exposed to sunlight, often have scale and can be numerous and of long-standing duration (3, 4). Sézary syndrome, recognized in 1938 by Albert Sézary, is a more aggressive variant with diffuse skin involvement in addition to evidence of leukemic disease. This malignant T cell disease, characterized by cerebriform nuclei known as Sézary cells, was classically identified by the triad of pruritic erythroderma affecting over 80% of body surface area, lymphadenopathy and other systemic manifestations and greater than 5% of Sézary cells (4). The term cutaneous T cell lymphoma was coined in 1975 by Edelson and Lutzner and initially referred to these two most common forms of disease. However, united by the common characteristic of a clonal expansion of T cells within at least the skin, many other entities are now also categorized under CTCL (2-4).

Epidemiologically, cutaneous T cell lymphoma is a relatively uncommon disease with an overall age-adjusted incidence of CTCL from 1973 to 2002 of 6.4 cases per million in the US. There has been an increasing incidence over the past few

decades, however, this is likely a result of improved diagnosis. Mycosis fungoides accounted for 72% of these cases whereas the leukemic form, Sézary syndrome accounted for 2.5%. CTCL more commonly affects men than women with roughly a 2:1 frequency and is more common among African Americans than Caucasians. Its incidence also increases greatly with age (3,4).

With significant heterogeneity in its presentation, the diagnosis of CTCL now relies upon integration of clinical, histopathological, immunopathological and molecular data. Persistent and/or progressive patches or thin plaques in non-sun exposed areas with variations in size or shape and poikiloderma should raise clinical suspicion of CTCL. Histopathlology, which may require serial biopsies to adequately identify disease, classically shows a superficial lymphoid infiltrate and atypical T cells with cerebriform nuclei displaying epidermotropism without spongiosis. Additionally, Pautrier microabscess, formations of malignant T cells and Langerhans cells aggregating in the epidermis, is a relatively specific, albeit not sensitive, finding for mycosis fungoides (3, 4). In this malignancy of $CD4^+$ T cells, immunologic markers can show an epidermal or dermal discordance of T cell antigens $CD2^+$, $CD3^+$, $CD5^+$, and particularly loss of $CD7^+$, and $CD26^+$(4). Molecular criteria include evidence of clonal TCR gene rearrangements by PCR (2, 3, 4).

Staging patients with CTCL at initial diagnosis can be helpful in guiding treatment and offers some prognostic information. Tumor stage (T) is based on the size of

patches and plaques, the presence of tumors >1 cm, or erythroderma, involving over 80% of body surface area. The presence or absence of clinically abnormal lymph nodes or evidence of metastases determine nodal (N) and metastases (M) stage. The degree of peripheral blood involvement (B) is based upon the percentage of circulating atypical cells present, and in more advanced stages additionally defined by an absolute malignant cell count >=1000 cells per microliter, a CD4/CD8 ratio >=10, aberrant expression of normal T cell markers, molecular evidence of clonality or chromosomal abnormality in a T cell clone (2, 4). T1 patients, with less than 10% body surface area involvement, often have a normal life expectancy whereas T4 erythrodermic patients have a median survival of 4-5 years and patients with visceral disease may only survive 1-2 years. Despite this TNMB staging, patients with mycosis fungoides do not uniformly progress through stages (4).

Currently, the pathogenesis of CTCL is still unknown and investigations of infectious agents and occupational exposures have not yielded definitive causative etiologic factors (3). Early stages of CTCL principally involve the skin, however the risk of peripheral blood involvement increases with advancing stage (5). The degree of clonal expansion of malignant cells within the peripheral blood is associated with a loss of T cell receptor diversity (6), T-regulatory activity (7), and diminished CD8 counts (8,9). CTCL patients are therefore at increased risk of complications of immune suppression including susceptibility to microbial infections and the development of secondary malignancies (10-13). While

treatments such as photopheresis (14) and interferon-α (15) have demonstrated CTCL's immunogenicity and responsiveness to immunotherapies, our understanding of the pathogenesis in this disease is still limited. Moreover, stage III and IV CTCL, especially in those with diminishing CD8 counts, is commonly lethal with survival not improved by conventional systemic therapies (8,9,14,15).

In recognition that cancer is fundamentally a genetic disease (16,17), knowledge of the critical mutations underlying tumor initiation and progression, along with a growing array of targeted therapeutics, offer tremendous potential in improving the diagnosis and treatment of cancer (18-20). Early cytogenetic studies in CTCL suggested that there might be commonalities in DNA copy number mutations (21,22) with 61% of samples showing chromosomal abnormalities (22). Importantly some of these alterations, such as gain of 8q and loss of 6q and 13q, were associated with shorter survival (23). These initial studies provided indicate that discernible patterns of mutations exist in CTCL and merit further investigation.

Additional characterizations have since bolstered this hypothesis with more potential candidate regions identified and previous regions narrowed (24-26). Vermeer *et al.* performed a copy number analysis that highlighted the significance of *MYC*, *TP53,* and genes in the IL-2 pathway as targets in the leukemic form of CTCL, Sézary Syndrome (24). Laharanne *et al.* confirmed 8q gain, including *MYC*, as well as loss of 10q and 17p as alterations more common

in Sézary Syndrome (26). *JUNB* has also been identified as a proto-oncogene and NAV3 as a potential tumor suppressor in this malignancy (4). Each of these studies has advanced our understanding of the CTCL genome, however, past studies of CTCL traditionally focused on a single type of mutation. In the context of multiple levels of cellular dysregulation, this singular approach with restricted sample size, low-resolution data and limited clinical information has encumbered the discernment of the primary biological pathways in this malignancy. More recently, in a matched copy number and expression analysis, Caprini *et al.* found genes in the *NF-κB* pathway to be perturbed and found better survival among patients with less than three copy number alterations(25). Since then, microarray resolution has further increased and now, advances in microarray technologies, coupled with improved computational algorithms, offer the unique opportunity to characterize the CTCL genome to the sub-gene level and study various modes of mutations.

The approach we now apply to CTCL has proven useful in identifying novel oncogenes and tumor suppressors in multiple other malignancies including some of the largest published studies in glioma, melanoma and lung adenocarcinoma (27-30). A testament to the successes of our methodology has been the subsequent functional validation of many genes initially identified with this genomic algorithm in melanoma (29). *SETDB1* was first identified in a region of DNA copy number gain on 1q21.3 by our approach and this methyltransferase has since been found to be oncogenic in a melanoma zebrafish model (31).

ETV1 amplification on 7p and its consequent over-expression was also identified by this approach and found to cause melanoma tumor formation in mice (32). Finally, growing *in vitro* and mouse model evidence supports that *KLF6*, first identified with our genomic algorithm (29), is a tumor suppressor in melanoma (33).

Building upon this methodological framework applied in other malignancies, our characterization of the genomic landscape in CTCL at both the DNA and RNA level in the context of detailed clinical data pinpoints new targets and advances our understanding of the mechanisms underlying pathogenesis, and immunosuppression in this malignancy.

STATEMENT OF PURPOSE AND SPECIFIC AIMS

The purpose of this study is to characterize the leukemic cutaneous T cell lymphoma (CTCL) genome and identify key oncogenes and tumor suppressors. By determining statistical outliers and patterns in DNA and RNA alterations in this disease and correlating them with clinical parameters, genes and biological pathways involved in pathogenesis can be identified and thereby focus the future development of biomarkers and targeted therapeutics in CTCL.

SPECIFIC AIMS

1. Generate high-resolution, patient-matched genomic data from primary leukemic cutaneous T cell lymphoma (CTCL) patients, including global DNA copy number gains/losses, and gene expression.
2. Determine statistically significant genetic alterations representing putative oncogenes and tumor suppressors through analysis of our data set and a meta-analysis of published data.
3. Integrate genomic data with gene expression data and correlate genetic alterations with clinical phenotypes to identify critical genes and biological pathways.

MATERIALS AND METHODS

Patient selection

CTCL patients were selected from the Yale Cutaneous Lymphoma Center clinic over a 3-month period in 2008. Patients chosen had leukemic disease meriting treatment with photopheresis based on clinical assessment and evidence of blood involvement from FACS data. Within this population, a reasonable diversity of clinical phenotypes and molecular markers on FACS analysis (such as CD4/CD8 ratios) ranging from predominantly early to some patients with late stage disease was also desired to characterize the spectrum of leukemic CTCL while identifying early genes involved in pathogenesis. Exclusion and inclusion criteria are described below. A final 24 patients were selected from which genomic data was generated. Written informed consent was obtained in accordance with the Declaration of Helsinki and approval was obtained from the Yale School of Medicine Human Investigation Committee
(HIC protocols: 25269 and 23636). All of these patients were treated with extracorporeal photopheresis on a monthly basis for CTCL with blood involvement. Some patients additionally received therapies such as bexarotene, methotrexate, interferon, or vorinostat.

Inclusion criteria:	***Exclusion criteria:***
1. CTCL patients age ≥18 and ≤90.	1. Previous use of immunosuppressive regimens including multi-drug chemotherapy within 3 months.
2. At least T2 skin involvement and evidence of blood involvement taking into account $CD3^+$, $CD4^+$, $CD8^+$, $CD4^+CD7^+$, $CD4^+CD7^-$, $CD4^+CD26^+$, $CD4^+CD26^-$, $CD25^+$, $CD4^+CD25^+$, CD45RO, CD4/CD8 ratios, presence of Vα or Vβ clones (34-36).	2. Known HIV or HTLV-1 positive patients.
	3. Pregnant or nursing mothers.

Clinical data collection

Clinical information was gathered at the time of sample collection and one year later (Table 1). A patient's current status, skin disease, PET/CT results and lymph node status on physical exam were categorized into "Improved," "Stable," or "Worsened." Current status was determined by physician global assessment. Skin disease was assessed taking into account changes over one year in body surface area and whether lesions were patches, plaques, tumors, or erythrodermic.

Isolation of malignant cell populations

In the final cycle of leukapheresis prior to 8-methoxypsoralen exposure, 10mL of lymphocyte-enriched blood was drawn from each patient. For normal controls, 50 mL of whole blood were drawn from four volunteers chosen as most similar to the patient population based on age and sex. The blood sample was immediately layered on 4 mL of Isolymph (per 10 mL of blood, CTL Scientific Supply Corp.; Deer Park, NY) and processed as described by the manufacturer. Wash steps

used 10 mL of RPMI (Gibco; Carlsbad, CA) with 1% penicillin-streptomycin followed by 10 min centrifugation at 1200 RPM, 20°C. An extra incubation at 4°C in lysis buffer (Sigma; St. Louis, MO) prior to the final centrifugation was performed on samples with a visible amount of red blood cells.

Based on FACS data, expanded clonal populations ≥20% of the peripheral blood, believed to be enriched for malignant cells, were identified for isolation, focusing on $CD4^+CD7^-$, $CD4^+CD26^-$, and $CD4^+CD7^-CD26^-$ populations. Otherwise, the $CD4^+$ population was isolated. Negative selection using CD7 and/or CD26 mAb conjugated to biotin (eBiosciences; San Diego, CA; Miltenyi Biotech; Auburn, CA) and anti-biotin microbeads (Miltenyi Biotech) was performed first followed by positive selection using anti-CD4 Microbeads (Miltenyi Biotech). Following manufacturer's instructions, populations were isolated using a Magnetic-bead Antibody Cell Sorting LS column and magnet (Miltenyi Biotech). Washes were done using filtered RPMI (Gibco) with 10% AB serum, 1% pencillin-streptomycin and 2mM EDTA. FACS was performed after isolation to confirm purity ≥70% (average 96%, median 99% $CD4^+$).

Genomic data generation

From the isolated cells, RNA was extracted using Qiagen RNeasy kit and stored in RNAlater (Qiagen; Valencia, CA) and DNA was extracted using Puregene (Gentra; Valencia, CA) and stored in TE buffer. After standard quality control measurements and quantitation, genomic data and total RNA were input into

their respective array assays. 500ng of genomic DNA was input in the PCR based Affymetrix Human SNP 6.0 assay so approximately 225ug of fragmented and labeled PCR product was hybridized to the human SNP 6.0 arrays (Affymetrix, Santa Clara, CA). Total RNA (300ng) was input in the Affymetrix Whole Transcript assay utilizing *in vitro* transcription and a second round of first strand cDNA synthesis to create the sense targeted single stranded DNA needed for hybridization. The sscDNA was fragmented, labeled and approximately 2.5ug was hybridized to the human Gene 1.0 ST arrays (Affymetrix) at the Keck Microarray Core facility according to manufacturer specifications.

Genomic Data Quality Controls

Affymetrix Expression Console software using the gene level RMA algorithm on all samples as a batch was used to assess quality metrics for hybridization and assay performance for the Affymetrix Human Gene 1.0 ST array. Perfect mean match were found to be in an acceptable range (298.2-402.7) compared to background values. Positive versus negative area under the ROC curve values to assess overall data quality were also acceptable (0.836-0.886). All samples run on the Affymetrix human Genome wide SNP 6.0 arrays were found to have acceptable contrast quality control values (1.56-3.30) indicating sufficient contrast resolution quality for genotyping.

Genomic Data Access

All genomic data will eventually be available in the GEO database (Accession number pending). Temporarily, the raw genomic data may be downloaded at: http://ymd.med.yale.edu/ymd_prod/cgi-bin/gz_login.cgi Login information: Username: WGEO Passcode: ctcl. Please click on "SIMPLE MENU," then "Download Affy, Illumina, Nimblegen Files from Keck." From there, click on individual CEL files to save them to a local directory.

Copy number analysis

Genomic data was processed using algorithms written in MATLAB and packages in GenePattern [http://genepattern.broadinstitute.org/gp/pages/index.jsf] (37). Affymetrix .CEL files were converted to .SNP files by an adaption of SNP File Creator for SNP 6.0. GISTIC preprocessing and normalization was done as described previously using 5 nearest normals (29). Determination of significantly altered regions for copy-number used the *g*enomic *i*dentification of *s*ignificant *t*argets in *c*ancer (GISTIC) algorithm [Figure 1, Supplementary Table 1] (27,29). GISTIC parameters include: genome build: hg18, amplifications threshold: 0.1, deletions threshold: 0.1, join.segment.size: 8, qv.thresh: 0.25, remove.X: 1. Copy number variants were removed as part of the GISTIC algorithm based on a database of genomic variants (file: variation.hg18.v8.txt, http://projects.tcag.ca/variation/tableview.asp?table=DGV_Content_Summary.txt)

Consensus analysis

Studies used in the consensus analysis focused on CTCL patients with blood involvement (Table 2). Lower resolution studies with sample number < 10 included all available data. Higher resolution studies with > 10 samples only included regions identified by each study's respective authors. Analyzing these 7 studies, including our own data set, common regions were defined by at least 2 studies and minimal common regions were defined by at least 1 study. Cancer related genes were annotated from genes identified in Beroukhim *et al*., 2010 (28).

Cluster analysis

Significant regions of amplifications and deletions defined by GISTIC, were analyzed using hierarchical clustering [Pearson correlation, complete linkage] (Figure 2A). The matrix is comprised of regions of significant alteration by individual samples. In a particular cell, no amplification or deletion was assigned a value of 0, a low-level amplification or deletion ($0.1 < \log_2(\text{copy number in region-1}) < 0.9$) was assigned a value of 1, and high level amplification or deletion ($0.9 < \log_2(\text{copy number in region-1})$) was assigned a value of 2. 17q25.1 was identified through visual analysis of clusters and confirmed with Fisher exact test (Figure 2B).

Gene expression integration with DNA copy number

Gene expression was integrated with copy number through Comparative Marker Selection. Samples with or without an amplification in the region were compared to identify differentially expressed genes (Figure 2C). CLS and GCT files were created and input into the GenePattern module (parameters in Supplementary Material). GSE17601 (25) was downloaded with GEOImporter and annotated using GeneCruiser. Comparative Marker Selection Parameters include: test.direction = 2, test.statistic = 0, number.of.permutations = 1000, complete = false, balanced = false, random.seed = 779948241, smooth.p.values = true.

Expression and pathway analysis

Differentially expressed genes between samples with the presence or absence of infections and secondary malignancies were identified with Class Neighbors analysis (default parameters) and Comparative Marker selection [parameters in Supplementary Material] (37). Overlapping genes were then input into Funcassociate 2.0 (38) (http://llama.med.harvard.edu/funcassociate/) to identify significant pathways (Figure 3).

Materials and Methods Contributions

WML performed 1)patient selection, 2)clinical data collection, 3)isolation of malignant cell populations, 4)DNA and RNA isolation, 6)genomic data deposition, 7)copy number analysis, 8)consensus analysis, 9)cluster analysis, 10)gene expression integration with DNA copy number, and 11)expression and pathway analysis. *KC* performed 2)clinical data collection. *SR* performed 4)DNA and RNA isolation. *AT* and *SU* performed 5)genomic data generation and 6)genomic data deposition. *GS* and *GG* performed .CEL to .SNP file conversion as part of 7) copy number analysis.

RESULTS

Copy number analysis identifies regions containing putative oncogenes and tumor suppressors

Given that cancer progression is an evolutionary process, mutations that confer a survival advantage should be selected for and altered at a higher rate. Therefore identifying statistical outliers should determine regions of the genome with genes of potential biological significance. Our DNA based analysis seeks to determine these “driver” copy number mutations using an algorithm, termed GISTIC (27), which has identified known and novel oncogenes and tumor suppressors in multiple cancers (27-30). GISTIC takes into account both the frequency of a DNA gain or loss and the magnitude of this gain or loss at each probe. Then, comparing this calculated score to the background null hypothesis score, a significance value at each probe throughout the genome is determined. Significantly altered regions of the genome can then be further investigated for potential oncogenes and tumor suppressors.

The CTCL landscape for amplifications based on GISTIC analysis of our samples is shown in Figure 1A with significance per probe plotted across the genome. We identified 17 regions of amplifications in the CTCL genome seen in red (Figure 1A, Supplementary Table 1A) and 40 regions of deletion in blue (Figure 1B, Supplementary Table 1B). The most significant amplifications involve large regions of gain on 8q and 17q, in addition to many focal amplifications. Overall, there were many more regions of deletion than amplification with the most

significant deletions on 17p and 10. These regions of chromosomal amplification and loss are shown in Supplementary Table 1.

We also identified focal amplifications of 4q12 including *KIT* and 7p11.2 including *EGFR.* While these amplifications were only found in a subset of samples, 4 of 23 (17%) and 3 of 23 (13%) respectively, half of those with *KIT* amplifications and all of those with *EGFR* amplifications had non-responsive or worsened skin disease after one year of treatment (Figure 4). The identification of these mutations in treatment non-responders is of potential clinical relevance given the availability of targeted therapeutics for these respective mutations. Moreover, *VEGFA* is another oncogene with an FDA-approved targeted therapy and is the closest well-validated gene to the focal amplification on 6p21.1 in 3 of 23 (13%) patients (Table 3). Thus, starting from the whole genome, the GISTIC algorithm begins to narrow and prioritize the list of candidate genes.

Consensus analysis of the leukemic CTCL genome

The clinical heterogeneity and relative rarity of CTCL combined with the difficulty in culturing and isolating malignant cells have hindered the genomic characterization of CTCL. To address some of these concerns, we performed a meta-analysis of published data, integrating regions of amplification and deletion from 108 samples comprising 7 studies, including our own, to create a consensus view of the CTCL genome (Table 2). In total, we determined 21 regions of amplification and 42 regions of deletion. Notably, all of our samples

had significant copy number alterations and many patterns in the consensus analysis mirrored our GISTIC analysis (Supplementary Table 1), corroborating our data set. A majority of minimal common regions were narrowed and defined by our study (38 of 63 regions) and there were no instances where a region was found in all of the studies except ours. Annotating these regions, we identified known oncogenes and tumor suppressors in addition to new candidates not previously reported. The most significant gain lies in a broad region on 8q, including *MYC.* This finding supports results from Vermeer *et al.* and Laharanne *et al.* suggesting *MYC*'s primary role in oncogenesis (24, 26). The second most significant region of gain encompasses a large region on 17q that has yet to be fully studied in CTCL. *STAT3* has previously been suggested as a candidate of gain at this locus (24), however in our analysis, *STAT3* is in a significant region but not at the peak, implying that there may be other gene targets of the amplification and drivers of oncogenesis on 17q. While our analysis identifies genes involved in many cancers, CTCL-specific oncogenes were also considered. A focal amplification of 10p13 including *GATA3*, a transcription factor believed to promote T-helper 2 (TH2) cytokine skewing (39, 40), was also identified. GATA3's importance in CTCL was previously proposed based on a supervised gene expression analysis (41). Its amplification in malignant cells may thus represent one mechanism of its activation and consequent CTCL mediated immune suppression via TH2 production. Deletions were most commonly found on chromosome 17p with *TP53* nearby, chromosome 10 including *PTEN*, 13q

including *RB1*, and 9p21.3 including *CDKN2A*, reaffirming previous findings (24-26).

17q amplification associated with stable/progressive skin disease

Ultimately, one of the goals of cancer genomics is tying patient genotype to clinical phenotype to better understand pathogenesis and identify opportunities to therapeutically intervene. After defining the most significant copy number alterations in CTCL, we used unsupervised pattern recognition algorithms to determine whether any clinical phenotypes were linked to copy number alterations (Figure 2A). Using hierarchical clustering, we found that 17q25.1 amplification was associated with patients who did not appear to improve and had stable/worsened skin disease after one year of treatment (Fisher exact test, p-value 0.026; Figure 2B). Seeking to identify candidate targets of amplification at this locus, we integrated gene expression data with our copy number analysis. Comparing samples with and without 17q25.1 gain, we identified genes that were both amplified and differentially over-expressed. This analysis was additionally done in an independent matched data set of CTCL samples with blood involvement by Caprini *et al.* (25). Two genes from our consensus analysis were identified: *RPS6KB1* from analysis of the Caprini *et al.* data set and *GRB2* from our data set. *RPS6KB1* is involved in the PI3K/mTOR pathways and its amplification and over-expression is believed to affect prognosis in breast cancer (42). *GRB2* is involved in MAPK signaling and therapeutic targets are already under development(43). Finally, overlapping genes in common include:

PRKAR1A and *USP32* (Figure 2C). In particular, *PRKAR1A* regulates cAMP levels and has been found to induce three-to-four fold increased expression of IL-2 when it is over-expressed (44). Moreover, we found that IL-2 receptor beta expression, the subunit which binds IL-2 and has been specifically linked to proto-oncogene induction (45), was concordantly highest in patients with *PRKAR1A* amplifications in comparison with those without amplifications. Thus, a theoretical oncogenic model would be that *PRKAR1A* amplification is one mechanism of IL-2 pathway activation. This then leads to further malignant T cell proliferation through up-regulation of JAK/STAT, PI3K/AKT, and MAP kinase pathways (44-46), ultimately resulting in a clinically observed treatment-resistant skin disease.

Correlating molecular alterations with clinical phenotypes of immunosuppression

Patients with CTCL often suffer from the immunosuppressive effects of the disease, however the mechanism by which this occurs is not well understood (7-10). A subset of the patients we studied developed new infections and secondary malignancies after their diagnosis and treatment of CTCL (Table 1). Comparing the gene expression of three patients with B-cell lymphoma, spinal abscess and herpes zoster to patients who did not develop any infections or secondary malignancies determined a gene expression signature shown in Figure 3. The genes in this expression signature range from chemokines to antibody receptors to glycoproteins to serine-threonine kinases to transcription

factors. Pathway analysis of these genes found to be relatively down-regulated in patients with infections and secondary malignancies revealed many key processes of the normal immune system including leukocyte chemotaxis, migration, chemokines and cytokine production, and cellular migration and movement. Overall, this gene expression signature suggests a degree of immunosuppression in patients who developed infections and secondary malignancies.

DISCUSSION

We are currently in the early stages of CTCL genomics still defining the genetic landscape, and CTCL continues to offer unique opportunities to study T cell biology and cancer. Advances in genomic technology along with improved modeling and an increasing number of genomic studies are steadily expanding our knowledge of the mutations underlying malignant T cell transformation of CTCL. Our high-resolution analysis and meta-analysis of patients with leukemic disease discerns some of the common DNA copy number alterations characteristic of CTCL, and we will continue to learn more as we study the mutations of patients with patch, plaque, tumor, erythrodermic and blood involvement and those ranging from the earliest to the most advanced stages of disease.

Despite their promise, genomic studies in CTCL must navigate many challenges. Foremost are limitations presented by sample size and data quality, which are affected by both the purity and yield of malignant cell isolation. These hurdles, largely due to the low incidence and heterogeneity of disease and technical options available, continue to restrain the potential contribution of genomics to our understanding of CTCL biology. Our desire to understand key genes in pathogenesis led us to incorporate a spectrum of CTCL patients, but primarily focusing on those with earlier stages of disease. Characterizing early leukemic disease makes these concerns about malignant cell isolation particularly critical to consider. Ultimately, we chose to use immunophenotypic markers CD7 and

CD27 to maximize the enrichment of the malignant cells in the population we isolated. The loss of CD7 expression has often been correlated with expansion of clonal T-cells in CTCL, and taken together with observations from Rappl and colleagues that a CD4+CD7- phenotype was found in all clonal TCR VBeta cells from seven CTCL patients suggests that the dominant clone can, in at least some cases, arise from this subset (36). While, this CD4+CD7- phenotype can also be found in reactive T-cell populations, it is nevertheless the most common major T-cell antigen lost, with 46% to 76% of leukemic CTCL patients showing decreased CD7 expression (47). Similarly, CD26 has also been used as another marker for leukemic CTCL. With its use first described by Bernengo *et al.* (48), CD26 loss has also been confirmed in clonal T-cell populations and has been estimated to occur in 59.3% of Sézary syndrome patients (47). Notably, it has been found to significantly correlate with the percentage of Sézary cells within the lymphoid population (48), and Bernengo *et al.* also observed significantly expanded CD4+CD26- populations in patients with CD4/CD8 ratios less than 10 (48). Overall, neither of these markers are ideal as there exist normal cells expressing these immunophenotypes; however, in the setting where there was a CD7- or CD26- population expanded outside of normal ranges, we utilized these markers in hopes to more specifically select and enrich our isolation for the dominant malignant clone.

The potential ramifications of reduced malignant cell purity affects all leukemic CTCL genomic studies and the resulting data must be viewed in light of isolation

methodology limitations and the patient population. Gene expression values can be biased as a direct result of cell population contaminants. DNA copy number changes are theoretically more robust, but there can be attenuation of absolute signal from normal cell admixture. Strikingly, all of our samples had copy number alterations, however had they not, due to isolation of the incorrect population of cells or selection of a patient with undetectable malignancy, one safeguard in our copy number algorithm is to correct for this theoretical case and to remove samples without any significant copy number changes from inclusion in further analysis. Additional concerns about the data quality are also partially alleviated through meta-analysis which enabled not only increased sample size but also allowed us to place our study in the proper context. Our consensus analysis included 6 other genomic studies which were selected focusing on Sézary syndrome patients, often with more advanced leukemic disease. The results of this analysis and the similarities of alterations on a genome-wide level between our data and other data sets corroborates our own analysis. Indeed, it appears our samples captured a large proportion of the genomic diversity in leukemic CTCL with some patterns characteristic of malignancy observed even in our patients with early stages of blood involvement. Also, including data and analysis from 108 patients enabled us to determine recurrent commonalities in DNA copy number alterations resulting in one of the more comprehensive descriptions of the leukemic CTCL genome. At the same time, we recognize that this view reflects an evolving picture which we hope will be built upon and refined by future studies to further our understanding of the genomic landscape in CTCL.

While the genomic landscapes are similar from a global view between our patients with early leukemic disease and other studies of patients with more advanced disease, we also make direct comparisons at individual loci. By generating our own high resolution data set and then using an algorithm that identifies driver genetic mutations, we narrow and verify significant regions of gain and loss in leukemic CTCL. Our analysis confirms findings from previous studies such as the significance of 8q amplification including *MYC*, chromosome 10 loss including *PTEN* and *FAS,* 9p21.3 deletion including *CDKN2A,* 13q deletion including *RB1,* and 17p loss including *TP53*. In addition, we begin to narrow known regions on 17q, and identify new loci and genes of interest, some with available targeted therapeutics.

Directly addressing the potential clinical importance of our work, we notably discover copy number amplifications within or adjacent to *KIT*, *EGFR*, and *VEGFA*. Cancer is increasingly classified and additionally defined by key mutations. In part driven by our expanding arsenal of targeted therapeutics and the knowledge that copy number amplifications can predict sensitivity to targeted therapeutics, leading to increased survival (49), the discovery of these amplifications obligates further study as to whether a subset of CTCL patients may benefit from targeted therapies such as imatinib, erlotinib, and bevacizumab, respectively. While *in vitro* validation and testing is certainly warranted, this question's importance is only further emphasized knowing that the skin

disease of a subset of the patients with apparent *KIT* and *EGFR* amplifications are not improving with currently available treatments.

In addition to the potential therapeutic ramifications of these findings, these focal amplifications also suggest that these genes and their products may play an intriguing role in mediating pathology. CTCL cells display a survival advantage within the skin, and the ligands for KIT (stem cell factor) and EGFR (epidermal growth factor) are both readily produced in the skin, raising the possibility that these factors stimulate CTCL within the cutaneous environment. In addition, VEGFA was shown to be spontaneously produced by CTCL cells, and is implicated in stage-dependent angiogenesis observed within cutaneous lesions (50). While further *in vitro* and *in* vivo studies are necessary, what role these genes play in the malignant milieu and whether patients who harbor these mutations may benefit from targeted therapeutics represent exciting prospects in the further understanding and treatment of CTCL.

Clinical data is of great benefit in modeling mutations in cancer and we demonstrate two instances where correlating genotype to clinical phenotype delineates important biology. Our cluster analysis of significant copy number alterations in correlation with skin disease severity led to identifying 17q25.1 amplification as associated with stable/progressive skin disease one year after treatment. While it could be argued that treatment was not optimized in these patients and it is difficult to account for any potential impact they may have, the

fact that 17q25.1 was also the second most significant region of amplification and that a total of ~44% samples from our consensus analysis included 17q25 gain suggests that there are genes at or near this locus driving oncogenesis. Utilizing orthogonal genomic expression data focusing on samples that are amplified and differentially over-expressed in 2 data sets, *PRKAR1A* was nominated from our list of candidates. Given the importance of the IL-2 pathway in CTCL and the availability of Denileukin difitox (45), *PRKARK1A* represents an interesting target meriting further study. It is important to note, however, that broad gain of 17q suggests multiple oncogenic targets and therefore *PRKAR1A* is likely not the sole focus of amplification. Indeed, *GRB2* and *RPS6KB1* represent other potential candidates. Thus, while *PRKAR1A's* potential role in activating the IL-2 pathway is conceivable, more importantly, our analysis draws attention to the 17q locus and provides a framework with which to prioritize future functional work.

Recognizing the immunosuppressive effects of CTCL and the complications that these patients consequently suffer, we were also interested in studying molecular markers of patients who developed infections and secondary malignancies versus those who did not. While sample number is severely limited, the statistically significant results of our pathway analysis essentially reflects what we see clinically. Further study of this signature may provide new insight into mechanisms of immunosuppression. Moreover, if biomarkers for this phenotype can be identified, this would enable identification of patients at risk for immunosuppression and facilitate stratification for increased screening or

prophylactic treatment.

The genomic characterization of CTCL presents both new opportunities and challenges. As our view of the genomic landscape focuses, there will be an increasing shift toward modeling the interactions between oncogenes, tumor suppressors and their mediators and ultimately understanding the malignant T cell in the context of its immunologic environment. While there is considerable complexity to these systems, insights gained from these models will likely facilitate the delineation of clinically relevant molecular subsets in CTCL. Ultimately, the foremost indicator of clinical relevance will be the translation of these discoveries to the development of biomarkers and targeted therapeutics. The identification of novel oncogenes in CTCL may spur new indications of approved therapies in addition to identifying rational drug targets. Looking toward pre-clinical testing, functional validation of genomic hypotheses will require robust, genetically characterized *in vitro* and *in vivo* models. Continued genomic discoveries correlating genotypes to clinical phenotypes will enable additional avenues and insights into the interface of cancer biology and immunology in this malignancy.

REFERENCES

1. Berger CL, Edelson RL. Current concepts of the immunobiology and immunotherapy of cutaneous T cell lymphoma: insights gained through cross-talk between the clinic and the bench. Leuk Lymphoma. 2003 Oct;44(10): 1697-703.
2. Willemze R, Meijer CJ. Classification of cutaneous T cell lymphoma: from Alibert to WHO-EORTC. J Cutan Pathol. 2006 Feb;33 Suppl 1:18-26.
3. Girardi M, Heald PW, Wilson LD. The pathogenesis of mycosis fungoides.N Engl J Med. 2004 May 6;350(19):1978-88.
4. Hwang ST, Janik JE, Jaffe ES, Wilson WH. Mycosis fungoides and Sézary syndrome. Lancet. 2008 Mar 15;371(9616):945-57.
5. Schechter GP, Sausville EA, Fischmann AB, Soehnlen F, Eddy J, Matthews M, Gazdar A, Guccion J, Munson D, Makuch R, et al. Evaluation of circulating malignant cells provides prognostic information in cutaneous T cell lymphoma. Blood. 1987 Mar;69(3):841-9.
6. Yawalkar N, Ferenczi K, Jones DA, Yamanaka K, Suh KY, Sadat S, Kupper TS. Profound loss of T cell receptor repertoire complexity in cutaneous T cell lymphoma. Blood. 2003 Dec 1;102(12):4059-66.
7. Berger CL, Tigelaar R, Cohen J, Mariwalla K, Trinh J, Wang N, Edelson RL. Cutaneous T cell lymphoma: malignant proliferation of T-regulatory cells. Blood. 2005 Feb 15;105(4):1640-7.

8. Vonderheid EC, Pena J, Nowell P. Sézary cell counts in erythrodermic cutaneous T cell lymphoma: implications for prognosis and staging. Leuk Lymphoma. 2006 Sep;47(9):1841-56.
9. Klemke CD, Mansmann U, Poenitz N, Dippel E, Goerdt S. Prognostic factors and prediction of prognosis by the CTCL Severity Index in mycosis fungoides and Sézary syndrome. Br J Dermatol. 2005 Jul;153(1):118-24.
10. Nguyen V, Huggins RH, Lertsburapa T, Bauer K, Rademaker A, Gerami P, Guitart J. Cutaneous T cell lymphoma and Staphylococcus aureus colonization. J Am Acad Dermatol. 2008 Dec;59(6):949-52.
11. Goldgeier MH, Cohen SR, Braverman IM, Stenn KS. An unusual and fatal case of disseminated cutaneous herpes simplex. Infection in a patient with cutaneous T cell lymphoma (mycosis fungoides). J Am Acad Dermatol. 1981 Feb;4(2):176-80.
12. Axelrod PI, Lorber B, Vonderheid EC. Infections complicating mycosis fungoides and Sézary syndrome. JAMA. 1992 Mar 11;267(10):1354-8.
13. Brownell I, Etzel CJ, Yang DJ, Taylor SH, Duvic M. Increased malignancy risk in the cutaneous T cell lymphoma patient population. Clin Lymphoma Myeloma. 2008 Apr;8(2):100-5.
14. Girardi M, Knobler R, Edelson R. Selective immmunotherapy through extracorporeal photochemotherapy: yesterday, today, and tomorrow. Hematol Oncol Clin North Am. 2003 Dec;17(6):1391-403. Review.
15. Olsen EA. Interferon in the treatment of cutaneous T cell lymphoma. Dermatol Ther. 2003;16(4):311-21.

16. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 2004 Aug;10(8):789-99.
17. Weber BL. Cancer genomics. Cancer Cell. 2002 Feb;1(1):37-47.
18. Kantarjian H, Sawyers C, Hochhaus A, Guilhot F, Schiffer C, Gambacorti-Passerini C, Niederwieser D, Resta D, Capdeville R, Zoellner U, Talpaz M, Druker B, Goldman J, O'Brien SG, Russell N, Fischer T, Ottmann O, Cony-Makhoul P, Facon T, Stone R, Miller C, Tallman M, Brown R, Schuster M, Loughran T, Gratwohl A, Mandelli F, Saglio G, Lazzarino M, Russo D, Baccarani M, Morra E; International STI571 CML Study Group. Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. N Engl J Med. 2002 Feb 28;346(9):645-52.
19. Demetri GD, von Mehren M, Blanke CD, Van den Abbeele AD, Eisenberg B, Roberts PJ, Heinrich MC, Tuveson DA, Singer S, Janicek M, Fletcher JA, Silverman SG, Silberman SL, Capdeville R, Kiese B, Peng B, Dimitrijevic S, Druker BJ, Corless C, Fletcher CD, Joensuu H. Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med. 2002 Aug 15;347(7):472-80.
20. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004 Jun 4;304(5676): 1497-500.

21. Mao X, Lillington D, Scarisbrick JJ, Mitchell T, Czepulkowski B, Russell-Jones R, Young B, Whittaker SJ. Molecular cytogenetic analysis of cutaneous T cell lymphomas: identification of common genetic alterations in Sézary syndrome and mycosis fungoides. Br J Dermatol. 2002 Sep;147(3):464-75.
22. Mao X, Lillington DM, Czepulkowski B, Russell-Jones R, Young BD, Whittaker S. Molecular cytogenetic characterization of Sézary syndrome. Genes Chromosomes Cancer. 2003 Mar;36(3):250-60.
23. Fischer TC, Gellrich S, Muche JM, Sherev T, Audring H, Neitzel H, Walden P, Sterry W, Tönnies H. Genomic aberrations and survival in cutaneous T cell lymphomas. J Invest Dermatol. 2004 Mar;122(3):579-86.
24. Vermeer MH, van Doorn R, Dijkman R, Mao X, Whittaker S, van Voorst Vader PC, Gerritsen MJ, Geerts ML, Gellrich S, Söderberg O, Leuchowius KJ, Landegren U, Out-Luiting JJ, Knijnenburg J, Ijszenga M, Szuhai K, Willemze R, Tensen CP. Novel and highly recurrent chromosomal alterations in Sézary syndrome. Cancer Res. 2008 Apr 15;68(8):2689-98.
25. Caprini E, Cristofoletti C, Arcelli D, Fadda P, Citterich MH, Sampogna F, Magrelli A, Censi F, Torreri P, Frontani M, Scala E, Picchio MC, Temperani P, Monopoli A, Lombardo GA, Taruscio D, Narducci MG, Russo G. Identification of key regions and genes important in the pathogenesis of Sézary syndrome by combining genomic and expression microarrays. Cancer Res. 2009 Nov 1;69(21):8438-46.
26. Laharanne E, Oumouhou N, Bonnet F, Carlotti M, Gentil C, Chevret E, Jouary T, Longy M, Vergier B, Beylot-Barry M, Merlio JP. Genome-wide analysis of

cutaneous T cell lymphomas identifies three clinically relevant classes. J Invest Dermatol. 2010 Jun;130(6):1707-18.

27. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, Du J, Kau T, Thomas RK, Shah K, Soto H, Perner S, Prensner J, Debiasi RM, Demichelis F, Hatton C, Rubin MA, Garraway LA, Nelson SF, Liau L, Mischel PS, Cloughesy TF, Meyerson M, Golub TA, Lander ES, Mellinghoff IK, Sellers WR. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A. 2007 Dec 11;104(50):20007-12. Epub 2007 Dec 6.
28. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, Tanaka KE, Chiang DY, Bass AJ, Loo A, Hoffman C, Prensner J, Liefeld T, Gao Q, Yecies D, Signoretti S, Maher E, Kaye FJ, Sasaki H, Tepper JE, Fletcher JA, Tabernero J, Baselga J, Tsao MS, Demichelis F, Rubin MA, Janne PA, Daly MJ, Nucera C, Levine RL, Ebert BL, Gabriel S, Rustgi AK, Antonescu CR, Ladanyi M, Letai A, Garraway LA, Loda M, Beer DG, True LD, Okamoto A, Pomeroy SL, Singer S, Golub TR, Lander ES, Getz G, Sellers WR, Meyerson M. The landscape of somatic copy-number alteration across human cancers. Nature. 2010 Feb 18;463(7283):899-905.
29. Lin WM, Baker AC, Beroukhim R, Winckler W, Feng W, Marmion JM, Laine E, Greulich H, Tseng H, Gates C, Hodi FS, Dranoff G, Sellers WR, Thomas RK,

Meyerson M, Golub TR, Dummer R, Herlyn M, Getz G, Garraway LA. Modeling genomic diversity and tumor dependency in malignant melanoma. Cancer Res. 2008 Feb 1;68(3):664-73.

30. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J, Fujii Y, Gazdar AF, Giordano T, Greulich H, Hanna M, Johnson BE, Kris MG, Lash A, Lin L, Lindeman N, Mardis ER, McPherson JD, Minna JD, Morgan MB, Nadel M, Orringer MB, Osborne JR, Ozenberger B, Ramos AH, Robinson J, Roth JA, Rusch V, Sasaki H, Shepherd F, Sougnez C, Spitz MR, Tsao MS, Twomey D, Verhaak RG, Weinstock GM, Wheeler DA, Winckler W, Yoshizawa A, Yu S, Zakowski MF, Zhang Q, Beer DG, Wistuba II, Watson MA, Garraway LA, Ladanyi M, Travis WD, Pao W, Rubin MA, Gabriel SB, Gibbs RA, Varmus HE, Wilson RK, Lander ES, Meyerson M. Characterizing the cancer genome in lung adenocarcinoma. Nature. 2007 Dec 6;450(7171):893-8.
31. Ceol C, Houvras Y, Jane-Valbuena J, Bilodeau S, Lin WM, Holman T, Ferre F, Bourque C, Burke C, Turner L, Uong A, Loda M, Garraway LA, Young RA, Zon LI. The SETDB1 histone methyltransferase is recurrently amplified in and accelerates formation of melanoma. Nature. 2010 in press.
32. Jané-Valbuena J, Widlund HR, Perner S, Johnson LA, Dibner AC, Lin WM, Baker AC, Nazarian RM, Vijayendran KG, Sellers WR, Hahn WC, Duncan LM, Rubin MA, Fisher DE, Garraway LA. An oncogenic role for ETV1 in melanoma. Cancer Research. 2010 Mar 1;70(5):2075-84.

33. Huh SJ, Chen YL, Friedman SL, Liao J, Huang HJ, Cavenee WK, Robertson GP. KLF6 Gene and early melanoma development in a collagen I-rich extracellular environment. J Natl Cancer Inst. 2010 Aug 4;102(15):1131-47.
34. Vonderheid EC, Boselli CM, Conroy M, Casaus L, Espinoza LC, Venkataramani P, Bigler RD, Hou JS. Evidence for restricted Vbeta usage in the leukemic phase of cutaneous T cell lymphoma. J Invest Dermatol. 2005 Mar;124(3):651-61.
35. Vonderheid EC. On the diagnosis of erythrodermic cutaneous T cell lymphoma. J Cutan Pathol. 2006 Feb;33 Suppl 1:27-42.
36. Rappl G, Muche JM, Abken H, Sterry W, Tilgen W, Ugurel S, Reinhold U. CD4 (+)CD7(-) T cells compose the dominant T cell clone in the peripheral blood of patients with Sézary syndrome. J Am Acad Dermatol. 2001 Mar;44(3):456-61.
37. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet. 2006 May;38(5):500-1.
38. Berriz GF, Beaver JE, Cenik C, Tasan M, Roth FP. Next generation software for functional trend analysis. Bioinformatics. 2009 Nov 15;25(22):3043-4.
39. Skapenko A, Leipe J, Niesner U, Devriendt K, Beetz R, Radbruch A, Kalden JR, Lipsky PE, Schulze-Koops H. GATA-3 in human T cell helper type 2 development. J Exp Med. 2004 Feb 2;199(3):423-8.
40. Ho IC, Tai TS, Pai SY. GATA3 and the T cell lineage: essential functions before and after T-helper-2-cell differentiation. Nat Rev Immunol. 2009 Feb;9 (2):125-35. Review.

41. Kari L, Loboda A, Nebozhyn M, Rook AH, Vonderheid EC, Nichols C, Virok D, Chang C, Horng WH, Johnston J, Wysocka M, Showe MK, Showe LC.Classification and prediction of survival in patients with the leukemic phase of cutaneous T cell lymphoma.J Exp Med. 2003 Jun 2;197(11): 1477-88.
42. Bärlund M, Forozan F, Kononen J, Bubendorf L, Chen Y, Bittner ML, Torhorst J, Haas P, Bucher C, Sauter G, Kallioniemi OP, Kallioniemi A. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. J Natl Cancer Inst. 2000 Aug 2;92(15):1252-9.
43. Dharmawardana PG, Peruzzi B, Giubellino A, Burke TR Jr, Bottaro DP. Molecular targeting of growth factor receptor-bound 2 (Grb2) as an anti-cancer strategy. Anticancer Drugs. 2006 Jan;17(1):13-20.
44. Elliott MR, Shanks RA, Khan IU, Brooks JW, Burkett PJ, Nelson BJ, Kyttaris V, Juang YT, Tsokos GC, Kammer GM. Down-regulation of IL-2 production in T lymphocytes by phosphorylated protein kinase A-RIIbeta. J Immunol. 2004 Jun 15;172(12):7804-12.
45. Manoukian G, Hagemeister F. Denileukin diftitox: a novel immunotoxin.Expert Opin Biol Ther. 2009 Nov;9(11):1445-51.
46. Marzec M, Halasa K, Kasprzycka M, Wysocka M, Liu X, Tobias JW, Baldwin D, Zhang Q, Odum N, Rook AH, Wasik MA.Differential effects of interleukin-2 and interleukin-15 versus interleukin-21 on CD4+ cutaneous T cell lymphoma cells.Cancer Res. 2008 Feb 15;68(4):1083-91.

47. Kelemen K, Guitart J, Kuzel TM, Goolsby CL, Peterson LC.The usefulness of CD26 in flow cytometric analysis of peripheral blood in Sézary syndrome. Am J Clin Pathol. 2008 Jan;129(1):146-56.
48. Bernengo MG, Novelli M, Quaglino P, Lisa F, De Matteis A, Savoia P, Cappello N, Fierro MT.The relevance of the CD4+ CD26- subset in the identification of circulating Sézary cells.Br J Dermatol. 2001 Jan;144(1): 125-35.
49. Cappuzzo F, Hirsch FR, Rossi E, Bartolini S, Ceresoli GL, Bemis L, Haney J, Witta S, Danenberg K, Domenichini I, Ludovini V, Magrini E, Gregorc V, Doglioni C, Sidoni A, Tonato M, Franklin WA, Crino L, Bunn PA Jr, Varella-Garcia M.Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. J Natl Cancer Inst. 2005 May 4;97(9): 643-55.
50. Krejsgaard T, Vetter-Kauczok CS, Woetmann A, Lovato P, Labuda T, Eriksen KW, Zhang Q, Becker JC, Ødum N. Jak3- and JNK-dependent vascular endothelial growth factor expression in cutaneous T cell lymphoma. Leukemia. 2006 Oct;20(10):1759-66.

FIGURE AND TABLE LEGENDS

Figure 1. Significant DNA copy number alterations in leukemic CTCL

Statistically significant (a) amplifications and (b) deletions pinpointed by GISTIC aggregate analysis of 23 CTCL patient samples. Chromosomal location is across the bottom with labeled cytobands corresponding to the center of the region and cancer related genes from Beroukhim *et al.* or known CTCL genes labeled above the region. * denotes a gene adjacent to the peak region. Significance is reported as false discovery rate corrected q-values.

Figure 2 17q25.1 amplification associated with stable/progressive skin disease (a) Unsupervised analysis of significant copy number mutations in CTCL as defined by GISTIC analysis was done with hierarchical clustering. Each patient's skin disease severity after one year of treatment is annotated next the the sample number. (b) Fisher exact test shows an association between 17q25.1 amplification and stable/worsened skin disease after one year of treatment. (c) Using Comparative Marker Selection, differential expression analysis of samples with 17q25.1 amplification versus no amplification reveals candidate targets that are amplified and differentially over-expressed in the Caprini *et al.* (shown in red), our (shown in blue) data sets and the overlap (shown in purple).

Figure 3 Gene signature in CTCL patients with infections and secondary malignancies Gene expression of patients with and without infections and secondary malignancies was compared using two differential gene expression

algorithms, ClassNeighbors analysis and Comparative Marker Selection. Pathway analysis with Funcassociate 2.0 was then performed with the genes in common from the two independent differential expression analyses.

Figure 4 Potential CTCL targets with available targeted therapies GISTIC analysis identifies focal amplifications of 4q12, 7p11.2, 6p21.1 including or adjacent to candidate genes *KIT*, *EGFR*, and *VEGFA*. (a) The relative frequency of these amplifications in our data set is shown in addition to skin disease response to treatment one year later. A panel of selected therapies commercially available targeting the specific candidate genes is also included. (b) Heatmap views of DNA copy number alterations for all the samples at the three loci are shown. Red indicates amplification, white indicates no alteration and blue indicates deletion with the degree of color intensity indicating the magnitude of the alteration. The samples that determined the peak from the GISTIC analysis are marked with an *. The darker * indicates stable/worse skin disease whereas the lighter * indicates improved skin disease. Genes are indicated on the bottom with their relative size.

Table 1 CTCL Patient Characteristics

Clinical data was collected from the 24 CTCL patients with blood involvement and various clinical parameters are shown to help describe our patient population.

Table 2 Consensus Analysis of CTCL amplifications and deletions

Analyzing 7 CTCL genomic studies, common regions of amplification (shown in red) and deletion (shown in blue) were defined by at least 2 studies and the minimal common region was determined by at least 1 study. The first column indicates the chromosome number. The second column indicates the number of studies including the minimal common region. Cancer related genes from Beroukhim *et al.,* 2010 are shown in the final column.

Supplementary Table 1 Significant amplifications and deletions in CTCL

Corresponding with Figure 1, a rank list of significant copy number alterations and further details of each peak region including genes within the region are shown from the GISTIC analysis.

Figure 1: Significant Amplifications and Deletions in CTCL

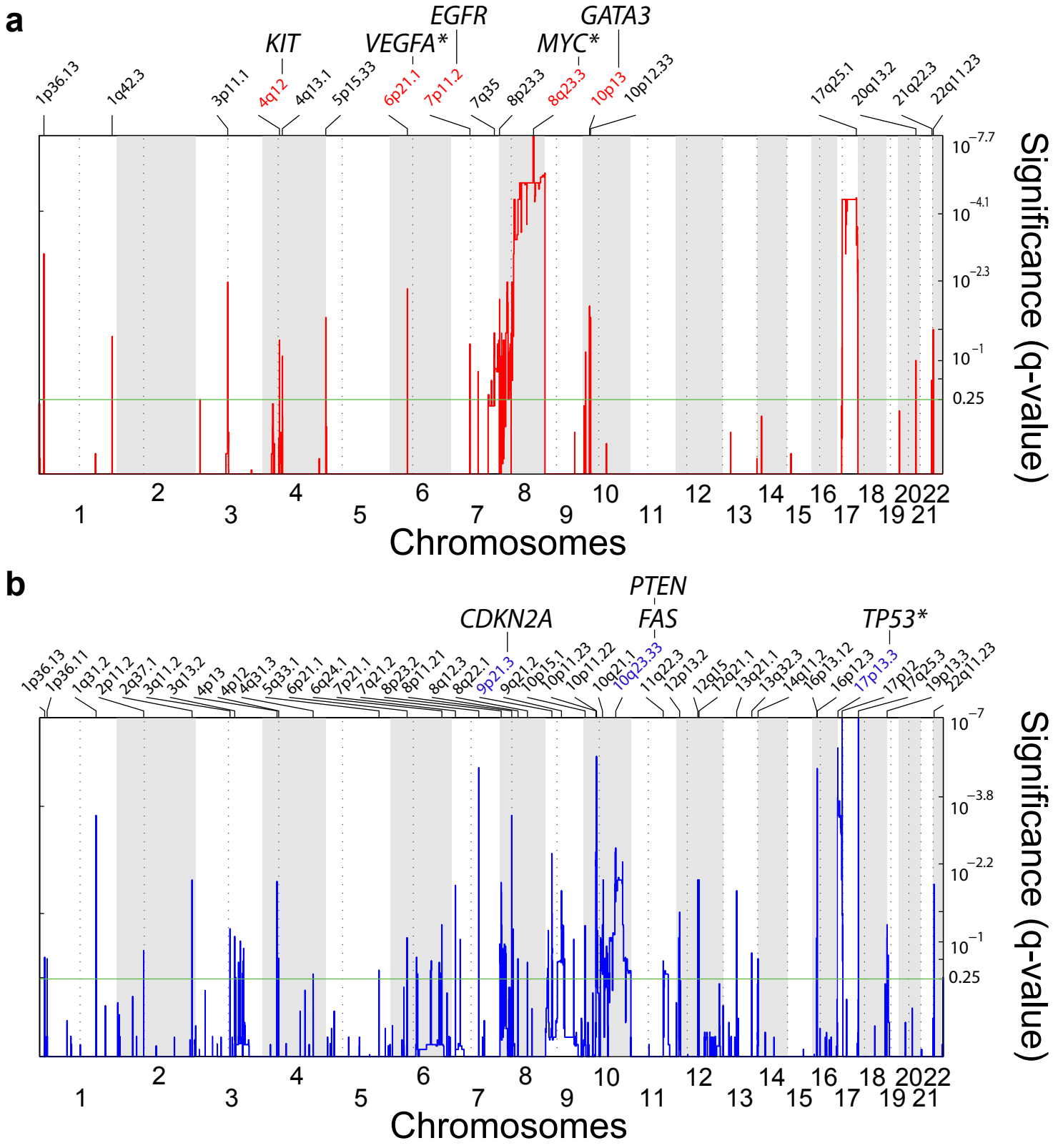
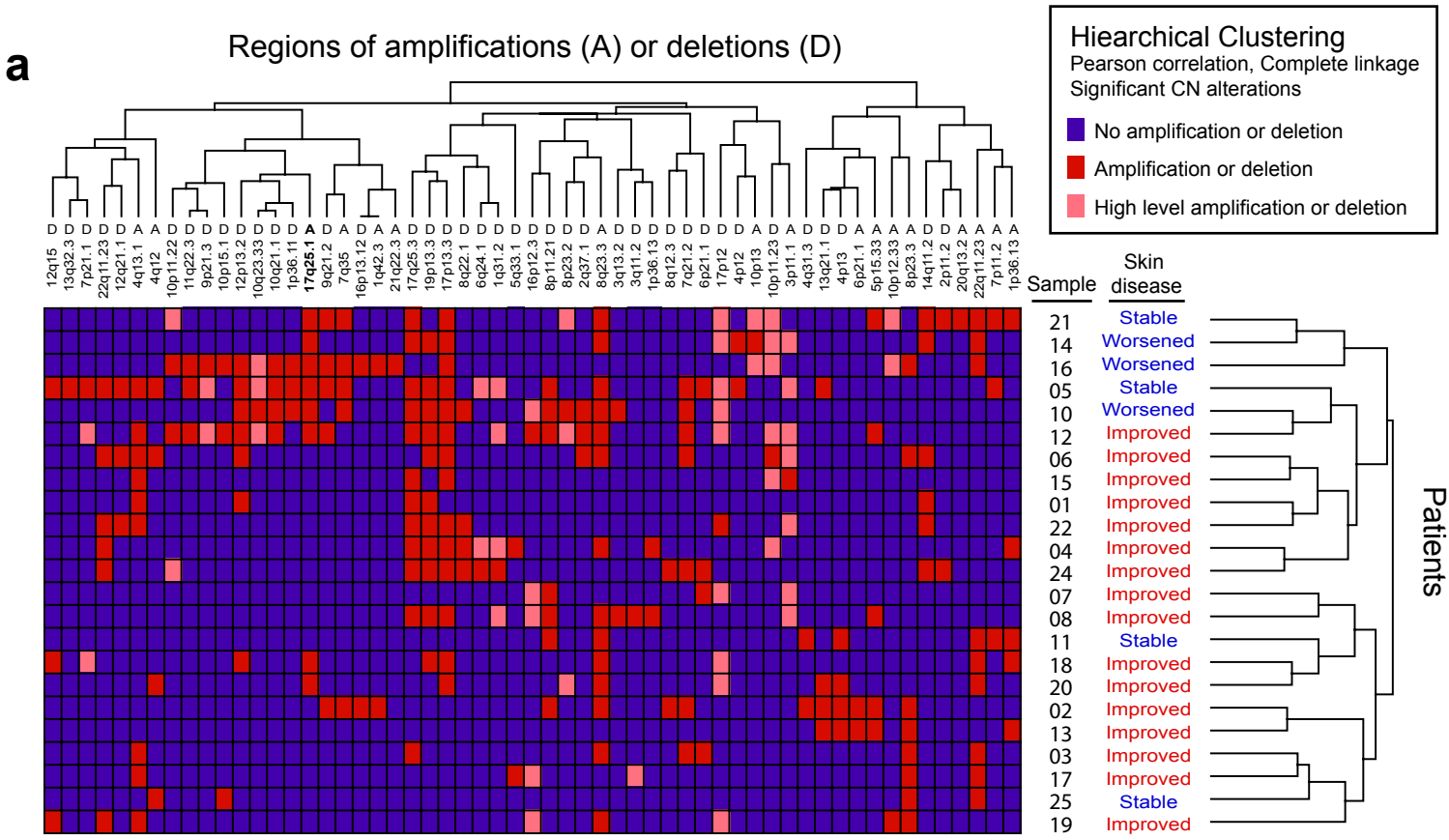
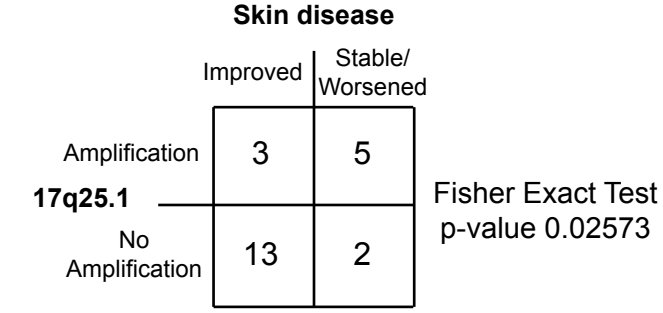
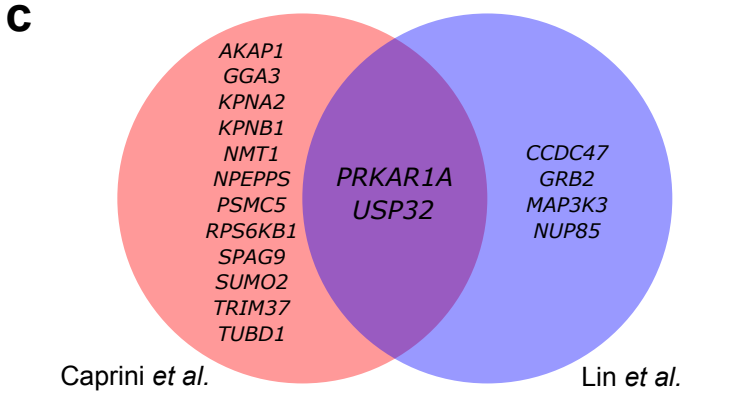


Figure 2: 17q25.1 amplification associated with stable/progressive skin disease







17q25.1 Amplified and differentially over-expressed

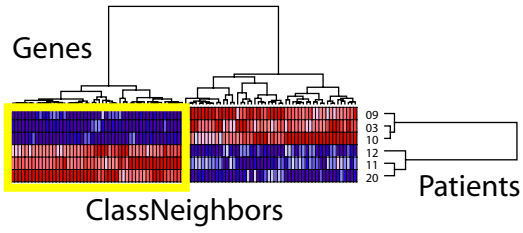
Figure 3: Gene signature in patients with infections and secondary malignancies

CLINICAL PHENOTYPE

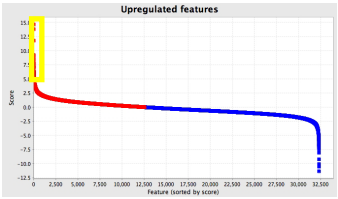
B-cell lymphoma
Spinal Abscess
VZV

vs.

Patients with no infections or secondary malignancies



ClassNeighbors



Comparative Marker Selection

GENES

CCL4
CD300E
CD83
CFP
CSDA
CXCL10
CXCL16
DUSP6
FCAR
HLA-DRA
HMOX1
ICAM1
IFI30
KLF4
LILRB2
MARCKS
MXD1
RASGEF1B
RIN2
RNASE2
SGK1
THBS1

Funcassociate 2.0

PATHWAYS

P-value	Description
4.51E-11	immune system process
2.03E-08	immune response
2.23E-06	regulation of leukocyte migration
4.37E-06	positive regulation of cell migration
6.14E-06	positive regulation of locomotion
1.51E-05	negative regulation of cytokine production
3.01E-05	chemokine activity
3.40E-05	chemokine receptor binding
3.82E-05	antigen processing and presentation
4.56E-05	chemotaxis

Genes and pathways upregulated in patients without infections or secondary malignancies and downregulated in patients who developed B-cell lymphoma, spinal abscess, VZV.

Figure 4: Potential CTCL targets with available targeted therapies

a

Significantly Amplified Region	Candidate Gene	Frequency	Treatment response stable/worse vs improved skin disease	Targeted therapy options
4q12	*KIT*	17% = 4/23	2 vs 2	imatinib, nilotinib
7p11.2	*EGFR*	13% = 3/23	3 vs 0	erlotinib, gefitinib, cetuximab
6p21.1	*VEGFA*	13% = 3/23	1 vs 2	bevacizumab

b

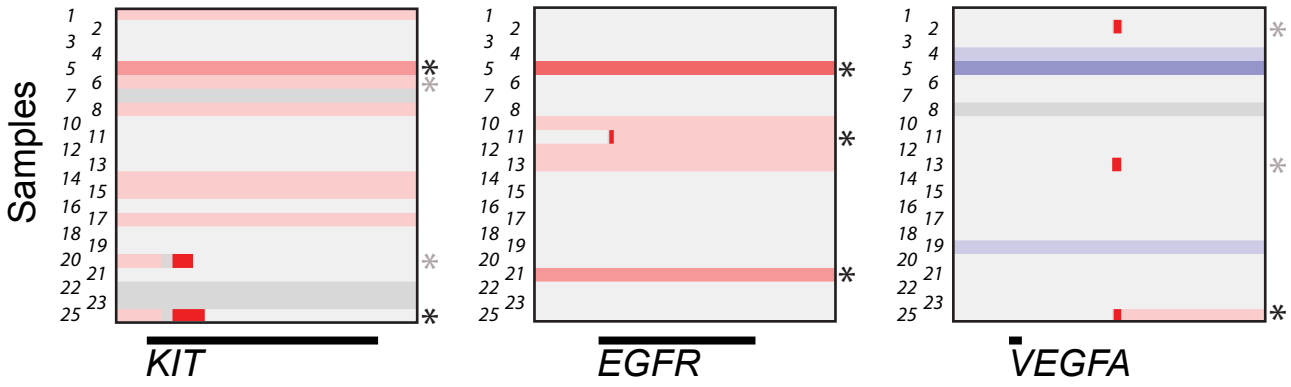


Table 1: CTCL Patient Characteristics

Patient Sample	Gender	Age	Current Status	Skin Score	PET/CT results	Lymph Node status on physical exam	Complete Remission	Infections and secondary malignancies	Valpha/ Vbeta	CD4/CD8 ratio	Type	Gene Expression Data	SNP data
1	Male	85	Stable	Improved	No Data	Stable	No	Past history of esophageal and prostate cancer	No	1.7	CD4+	-	Yes
2	Female	86	Stable	Improved	Stable	Worsened	No	Breast Cancer	No	1.6	CD4+CD7-CD26-	-	Yes
3	Male	67	Stable	Improved	Improved	Stable	No	B-cell lymphoma	Yes	2.9	CD4+	Yes	Yes
4	Female	50	Stable	Improved	No Lymphadenopathy	Stable	No	No Data	No	38	CD4+CD7-	Yes	Yes
5	Male	66	Deceased	Stable	No Data	Improved	No	No Data	No	99	CD4+CD7-CD26-	Yes	Yes
6	Male	52	Improved	Improved	Worsened	No Data	Yes	Past history of SCCs	No	2.3	CD4+	-	Yes
7	Female	76	Improved	Improved	No Lymphadenopathy	No Data	Yes	Past history of Uterine cancer. SCC	No	0.79	CD4+	Yes	Yes
8	Female	83	Improved	Improved	No Data	No Data	Yes	Osteomyelitis	Yes	0.28	CD4+	-	Yes
9	Male	54	Stable	Stable	No Data	Stable	No	Spinal abscess	No	1.2	CD4+	Yes	-
10	Female	87	Worsened	Worsened	Improved	Worsened	No	VZV	Yes	19.7	CD4+CD26-	Yes	Yes
11	Female	73	Worsened	Stable	No Data	Stable	No	No Post-CTCL disease	No	3.5	CD4+CD7-CD26-	Yes	Yes
12	Male	63	Improved	Improved	No Lymphadenopathy	Stable	Yes	No Post-CTCL disease	No	2.8	CD4+CD7-CD26-	Yes	Yes
13	Male	32	Improved	Improved	Worsened	Stable	No	No Post-CTCL disease	No	1.6	CD4+	-	Yes
14	Male	77	Stable	Worsened	Worsened	Stable	No	No Post-CTCL disease	No	18.5	CD4+CD26-	-	Yes
15	Female	57	Improved	Improved	No Data	Stable	Yes	No Post-CTCL disease	Yes	1.4	CD4+	-	Yes
16	Male	84	Worsened	Worsened	No Data	Worsened	No	No Post-CTCL disease	Yes	4.9	CD4+CD7-CD26-	-	Yes
17	Female	81	Improved	Improved	No Data	Improved	Yes	No Data	No	3.3	CD4+	Yes	Yes
18	Female	89	Deceased	Improved	No Data	Stable	No	No Data	Yes	3.6	CD4+CD26-	Yes	Yes
19	Male	54	Stable	Improved	Improved	Improved	No	Past history of SCC, possible VZV	No	3.1	CD4+	Yes	Yes
20	Male	66	Stable	Improved	No Data	Stable	No	No Post-CTCL disease	Yes	3.4	CD4+CD7-CD26-	Yes	Yes
21	Male	76	Deceased	Stable	Improved	Improved	No	Past history of VZV	No	7.5	CD4+CD7-CD26-	-	Yes
22	Male	64	Deceased	Improved	Stable	Improved	No	B-cell lymphoma	Yes	7.6	CD4+	-	Yes
24	Male	46	Improved	Improved	Improved	Improved	Yes	Vitiligo	No	3.1	CD4+	-	Yes
25	Female	38	No Data	Stable	Improved	Stable	No	No Data	No	1.2	CD4+	-	Yes

Table 2: Consensus analysis of the leukemic CTCL genome

	Mao *et al.*	Mao *et al.*	Fischer *et al.*	Vermeer *et al.*	Caprini *et al.*	Laharanne *et al.*	Lin *et al.*
Year	2002	2003	2004	2008	2009	2010	Unpublished
Samples	10	9	4	20	26	16	23
Method	CGH	CGH	CGH	CGH	SNP	CGH	SNP
Resolution	Low	Low	Low	3.5K	10K	105K	1800K
Region Defined	Table 1	Table 1	Table 1	By Authors	By Authors	By Authors	By Authors

Chr	Minimal Common Region	Number of studies	Cancer Genes
1	1p13-1p31(+)	2	
	1p22(-)	2	
	*1p36.11(-)**	3	***SFN***
	*1p36.13(+)**	2	
	*1q32-1q43(-)**	2	***RYR2, FH, ZNF678***
2	2p22-2p24(-)	2	
	*2q37.1(-)**	2	
3	3p14-3p21(+)	2	
	*3q13.2(-)**	2	***LSAMP***
4	4p16.1(+)	2	
	*4q12(+)**	2	***PDGFRA, KIT***
	*4q13.1(+)**	2	
5	5q14.3(-)	2	
	5q21-5q23(-)	2	***APC***
	*5q33.1(-)**	2	
6	*6p21.1(-)**	2	
7	*7p11.2(+)**	3	***EGFR***
	7p21.1(+)	3	
	*7q35(+)**	3	
8	*8p11.21(-)**	3	
	*8p23.2(-)**	3	***CSMD1***
	8p11.2-p11.1(+)	4	***PLAT***
	8p23.3(+)	4	
	*8q23.3(+)**	7	
	*8q24.1-q24.2(+)**	6	***NOV***
	*8q24.2-q24.3(+)**	6	***MYC***

Chr	Minimal Common Region	Number of studies	Cancer Genes
9	*9p21.3(-)**	3	***CDKN2A***
	*9q21.2(-)**	4	
	9q34(-)	2	
10	*10p11.22(-)**	5	
	*10p11.23(-)**	5	
	*10p15.1(-)**	5	
	*10q21.1(-)**	5	
	*10p13(+)**	5	
	10q11(-)	4	***PRKG1, DKK1***
	*10q23.33(-)**	6	***PTEN***
	10q24-q25(-)	5	
	10q26(-)	5	***MGMT, SYCE1***
11	*11q22.3(-)**	2	
12	*12p13.2(-)**	2	***ETV6***
	*12q15(-)**	2	
	*12q21.1(-)**	2	
	12q25(-)	2	
13	13q14(-)	4	***RB1***
	*13q21.1(-)**	4	
	13q22pter(-)	4	
	*13q32.3(-)**	4	
14	No consensus region		
15	No consensus region		
16	16q24(-)	2	***C16orf3***
	16p11.2(-)	2	
	*16p13.12(-)**	2	
	*16p12.3(-)**	2	

Chr	Minimal Common Region	Number of studies	Cancer Genes
17	*17p12(-)**	7	***MAP2K4, ZNF18***
	*17p13.3(-)**	7	
	17p11.2(+)	2	***MAPK7***
	17q21.31(+)	5	
	*17q23(+)**	6	***RPS6KB1***
	*17q25.1(+)**	5	***GRB2, ITGB4***
18	18(+)	2	
	18p11.1-18p11.3(-)	2	
19	*19p13.3(-)**	4	***GZMM, THEG, PPAP2C, C19orf20***
	19q13(-)	2	***BBC3, ZNF324***
20	*20q13.2(+)**	2	***ZNF217***
21	No consensus region		
22	22q12-22q13(-)	3	***TUBGCP6***

Minimal Common Regions
Gain: 21 Loss: 42

Red region indicates copy gain
Blue region indicates copy loss
(**Common region** defined from at least 2 studies)
Minimal common region defined by at least 1 study
38 Regions italicized with * were found in Lin *et al.* analysis
Number of studies indicates the number of studies that included the minimal common region
Cancer Genes represent significantly amplified or deleted genes in cancer from Beroukhim *et al.* (PMID: 20164920, Supplementary Table 2: GISTIC analysis of 3,131 cancer specimens, 26 tumor types).

Supplementary Table 1A: Significant regions of DNA copy number amplification in CTCL

Rank	Cytoband	Q-Value	Number of Samples	Region	Genes in region	Genes of interest related to region
1	8q23.3	3.44E-08	14	chr8:112360671-115742135	1 gene: CSMD3	***MYC****
2	17q25.1	1.84E-05	8	chr17:40531449-75546555	371 genes	**NGFR, PHB, RPS6KB1, BPTF, GRB2, ITGB4,** MIR21, *PRKAR1A, USP32*
3	1p36.13	0.000945	5	chr1:19256834-19274554	1 gene: UBR4	-
4	3p11.1	0.00429	8	chr3:90106010-96880237	5 genes: PROS1, NSUN3, ARL13B, DHFRL1, STX19	**POU1F1***
5	6p21.1	0.00596	2	chr6:43979350-43993064	1 gene: C6orf223	***VEGFA****
6	8p23.3	0.00950	8	chr8:577114-1646102	2 genes: DLGAP2, ERICH1	-
7	10p13	0.0128	3	chr10:4275516-17978213	75 genes: PFKFB3, PTPLA, ST8SIA6, STAM, OLAH, AKR1C3, CDC123, GATA3, SEC61A2, NUDT5, ACBD7, PRKCQ, VIM, SFMBT2, IL2RA, OPTN, FBXO18, USP6NL, RBM17, HSPA14, TAF3, AKR1CL1, tAKR, PHYH, TRDMT1, KIN, C10orf31, MCM10, ANKRD16, CALML5, ECHDC3, AKR1CL2, AKR1C2, CAMK1D, DHTKD1, C10orf111, IL15RA, DCLRE1C, MEIG1, ITGA8, LOC439951, FLJ45983, RPP38, PRPF18, ASB13, RSU1, TCEB1P3, FAM107B, CUGBP2, FRMD4A, AKR1C1, SEPHS1, ATP5C1, AKR1C4, GDI2, C10orf18, C10orf97, CUBN, SUV39H2, C10orf49, C10orf47, ARMETL1, UCN3, C1QL3, ITIH2, UPF2, TUBAL3, LOC100128356, NMT2, LOC389936, CALML3	*GATA3*
8	5p15.33	0.0204	5	chr5:2179815-3322657	2 genes: C5orf38, IRX2	**TERT***
9	10p12.33	0.0204	3	chr10:1-135374737	717 genes	-
10	22q11.23	0.0320	9	chr22:20471826-21907111	13 genes: hsa-mir-650, BCR, GNAZ, MAPK1, VPREB1, TOP3B, RAB36, PPM1F, PRAME, RTDR1, GGTLC2, ZNF280A, ZNF280B	**CRKL***
11	7q35	0.0362	5	chr7:142856201-158821424	103 genes: hsa-mir-595, hsa-mir-153-2, hsa-mir-671, ABP1, CDK5, DPP6, EN2, EZH2, GBX1, MNX1, HTR5A, INSIG1, KCNH2, NOS3, PTPRN2, RARRES2, RHEB, SHH, SLC4A2, SMARCD3, VIPR2, XRCC2, ARHGEF5, ZNF212, ZNF282, CUL1, ACCN3, PDIA4, UBE3C, FAM115A, DNAJB6, ABCF2, FASTK, ABCB8, PAXIP1, SSPO, CNTNAP2, GIMAP2, OR2F1, TPK1, ZNF777, TMEM176B, REPIN1, PRKAG2, NUB1, CSGlcA-T, NCAPG2, WDR60, GIMAP4, GIMAP5, TMEM176A, ACTR3B, FAM62B, ZNF398, MLL3, GALNT11, LMBR1, NOM1, LRRC61, ZNF767, TMUB1, KRBA1, C7orf29, CENTG3, NOBOX, OR2A14, OR6B1, OR2F2, ZNF786, ASB10, RNF32, GIMAP8, CRYGN, ZNF425, ZNF783, ZNF746, ATP6V0E2, CCT8L1, RBM33, GALNTL5, GIMAP7, ZNF467, GIMAP1, C7orf33, TAS2R41, CNPY1, FAM139A, ZNF775, ATG9B, CTAGE6, OR2A12, OR2A1, WDR86, OR2A25, OR2A5, OR2A7, OR2A42, LOC402715, LOC441294, OR2A2, FLJ43692, GIMAP6, LOC643641	
12	1q42.3	0.0410	2	chr1:232738038-232763374	1 gene: IRF2BP2	-
13	4q12	0.0467	4	chr4:55224058-55239617	1 gene: KIT	***KIT***
14	7p11.2	0.0531	3	chr7:54899474-65012069	20 genes: CCT6A, EGFR, GBAS, PHKG1, PSPH, ZNF138, ZNF273, SUMF2, CHCHD2, ZNF117, MRPS17, ZNF107, LANCL2, ECOP, ZNF479, VKORC1L1, ZNF92, ZNF680, 14-Sep, ZNF713	***EGFR***
15	4q13.1	0.0780	9	chr4:63139114-63362224	1 gene: LPHN3	-
16	20q13.2	0.0894	1	chr20:1-62435964	520 genes	**ZNF217**
17	21q22.3	0.156	1	chr21:39772858-46944323	88 genes: ADARB1, AIRE, PTTG1IP, C21orf2, CBS, COL6A1, COL6A2, CRYAA, CSTB, DSCAM, PRMT2, ITGB2, LSS, MX1, MX2, NDUFV3, PCNT, PCP4, PDE9A, PFKL, PKNOX1, PWP2, S100B, SH3BGR, SLC19A1, SUMO3, TFF1, TFF2, TFF3, TMEM1, TMPRSS2, TRPM2, U2AF1, UBE2G2, C21orf33, PDXK, RRP1, MCM3AP, ABCG1, B3GALT5, WDR4, FTCD, HSF2BP, RRP1B, DIP2A, POFUT2, ICOSLG, BACE2, C2CD2, DNMT3L, ZNF295, UBASH3A, SLC37A1, PCBP3, C21orf58, C21orf57, C21orf29, FAM3B, RIPK4, AGPAT3, PRDM15, TMPRSS3, COL18A1, LRRC3, C21orf56, C21orf70, RSPH1, UMODL1, IGSF5, SNF1LK, PLAC4, KRTAP12-2, KRTAP12-1, KRTAP10-10, C21orf123, KRTAP10-4, KRTAP10-6, KRTAP10-7, KRTAP10-9, KRTAP10-1, KRTAP10-11, KRTAP10-2, KRTAP10-5, KRTAP10-8, KRTAP10-3, KRTAP12-3, KRTAP12-4, KRTAP10-12	-

1. Rank according to Q-value. Q-value are false-discovery rate corrected.
2. Cytoband from the middle of the minimal common region.
3. Region represents the wider area that results when the samples defining each border of the minimal common region are left out.
4. Genes of interest related to region:
   Genes in bold represent significantly amplified genes in cancer from Beroukhim *et al.* (PMID: 20164920, Supplementary Table 2: GISTIC analysis of 3,131 cancer specimens of 26 tumor types).
   Genes in italics represent genes that are in the region that are mentioned in the text
   Genes with * represent genes adjacent to the region

Supplementary Table 1B: Significant regions of DNA copy number deletion in CTCL

Rank	Cytoband	Q-value	Number of samples	Region	Genes in region	Genes of interest related to region
1	17p12	1.73E-07	10	chr17:15524281-15794764	4 genes: ADORA2B, TRIM16, ZNF286A, TBC1D26	**MAP2K4*, ZNF18***
2	17q25.3	1.73E-07	12	chr17:78226063-78242141	1 gene: RAB40B	**ZNF750***
3	10p11.22	6.79E-06	4	chr10:30673943-31410094	4 genes: MAP3K8, PAPD1, LYZL2, ZNF438	-
4	7q21.2	1.74E-05	7	chr7:91195291-91198925	1 gene: MTERF	-
5	17p13.3	1.74E-05	14	chr17:2109320-2748547	10 genes: MNT, PAFAH1B1, SGSM2, GARNL4, KIAA0664, SMG6, TSR1, SRR, METT10D, LOC284009	***TP53*, ATP1B2***
6	16p12.3	1.86E-05	6	chr16:19711348-19743071	1 gene: IQCK	-
7	1q31.2	0.000481	5	chr1:191421896-191426345	2 genes: B3GALT2, CDC73	-
8	8p11.21	0.000481	7	chr8:42269520-42277830	1 gene: IKBKB	**C8orf4*, ZMAT4***
9	10q23.33	0.00288	4	chr10:60671322-97354704	211 genes	***PTEN***
10	10p11.23	0.00346	7	chr10:29478793-30143973	4 genes: hsa-mir-938, hsa-mir-604, SVIL, LYZL1	-
11	9p21.3	0.00376	3	chr9:21790514-22605798	4 genes: CDKN2A, CDKN2B, MTAP, DMRTA1	***CDKN2A***
12	2q37.1	0.0121	3	chr2:232785309-232823401	1 gene: DIS3L2	TMEM16G, ING5
13	12q15	0.0121	3	chr12:66948191-66954650	1 gene: IL22	-
14	12q21.1	0.0121	3	chr12:69853599-69876390	1 gene: TSPAN8	-
15	4p13	0.0128	4	chr4:43100100-43137404	1 gene: LOC389207	-
16	8p23.2	0.0133	4	chr8:3570501-12646620	57 genes: hsa-mir-598, hsa-mir-124a-1, hsa-mir-597, ANGPT2, BLK, CTSB, DEFA1, DEFA3, DEFA4, DEFA5, DEFA6, DEFB1, DEFB4, FDFT1, GATA4, MSRA, TNKS, MFHAS1, SPAG11B, PINX1, AGPAT5, DEFB103A, CSMD1, MTMR9, MCPH1, PPP1R3B, SOX7, C8orf13, AMAC1L2, FAM86B1, THEX1, LONRF1, RP1L1, CLDN23, DEFB104A, PRAGMIN, UNQ9391, C8orf74, DEFB105A, DEFB106A, DEFB107A, DEFB130, NEIL2, XKR6, DUB3, XKR5, DEFB103B, DEFB107B, DEFB104B, DEFB106B, DEFB105B, DEFB136, DEFB137, DEFB134, SPAG11A, LOC728358, LOC728957,	**CSMD1**
17	22q11.23	0.0144	6	chr22:22564665-23056324	10 genes: DDT, GGT5, GSTT1, GSTT2, MIF, SPECC1L, CABIN1, SUSD2, GSTT2B, DDTL	**VPREB1***
18	7p21.1	0.0150	3	chr7:16820392-17640380	2 genes: AHR, AGR3	-
19	9q21.2	0.0184	5	chr9:80289820-92881841	43 genes: hsa-mir-7-1, CKS2, CTSL1, DAPK1, EDG3, GAS1, HNRPK, NTRK2, SYK, TLE1, TLE4, SEMA4D, GADD45G, SPIN1, AGTPBP1, CCRK, UBQLN1, GOLM1, SHC3, DIRAS2, KIF27, MAK10, SLC28A3, SECISBP2, ZCCHC6, RMI1, GKAP1, ISCA1, C9orf64, NXNL2, RASEF, FRMD3, C9orf47, C9orf79, LOC286238, C9orf164, FLJ46321, C9orf153, CTSL3, FLJ45537, FAM75B, C9orf103, CHCHD9	-
20	13q21.1	0.0184	4	chr13:54591155-54593583	1 gene: FLJ40296	***RB1****
21	12p13.2	0.0399	7	chr12:10860835-11642691	18 genes: PRB1, PRB3, PRB4, PRH1, PRH2, PRR4, TAS2R13, TAS2R10, TAS2R14, TAS2R43, TAS2R44, TAS2R46, TAS2R47, TAS2R48, TAS2R49, TAS2R50, TAS2R42, PRB2	**ETV6***
22	6q24.1	0.0596	3	chr6:141733503-163657589	99 genes: ACAT2, ESR1, FUCA2, GRM1, HIVEP2, IGF2R, LPA, MAS1, MAP3K4, NMBR, OPRM1, PARK2, PCMT1, PLAGL1, PLG, SLC22A1, SLC22A3, SLC22A2, SOD2, TCP1, DYNLT1, UTRN, EZR, VIP, EPM2A, PEX3, STX11, SYNJ2, LATS1, WTAP, AKAP12, PHACTR2, UST, RAB32, KATNA1, RBM16, MAP3K7IP2, SASH1, SYNE1, MTHFD1L, PIP3-E, TIAM2, FBXO5, RGS17, MRPL18, CLDN20, NOX3, TFB1M, AIG1, SNX9, VTA1, MTRF1L, RMND1, AGPAT4, TULP4, GPR126, PLEKHG1, ARID1B, TMEM181, ZBTB2, ULBP3, C6orf211, ZDHHC14, C6orf103, C6orf97, MYCT1, ULBP2, ULBP1, LPAL2, PPP1R14C, SF3B5, RSPH3, FBXO30, FNDC1, LRP11, LTV1, SERAC1, PPIL4, SYTL3, C6orf72, TAGAP, ADAT2, STXBP5, PACRG, RAET1E, CNKSR3, RAET1L, PNLDC1, LOC202459, SHPRH, ZC3H12D, NUP43, RAET1G, C6orf98, SUMO4, SAMD5, IYD, LOC389435, GTF2H5,	-
23	10q21.1	0.0596	4	chr10:50229783-58152761	27 genes: hsa-mir-605, CHAT, ERCC6, MBL2, MSMB, PRKG1, SLC18A3, NCOA4, PARG, TIMM23, ZWINT, DKK1, CSTF2T, A1CF, FAM21B, OGDHL, ASAH2, PCDH15, SGMS1, PGBD3, C10orf53, FAM21A, CTGLF3, DRGX, CTGLF4, ASAH2B, CTGLF5	-
24	19p13.3	0.0596	11	chr19:3178103-5466792	54 genes: hsa-mir-7-3, hsa-mir-637, DAPK3, EEF2, MATK, NFIC, MAP2K2, PTPRS, SH3GL1, TBXA2R, APBA3, CHAF1A, EBI3, M6PRBP1, HMG20B, SEMA6B, JMJD2B, ZFR2, PIP5K1C, TJP3, ITGB1BP3, UHRF1, ZBTB7A, FZR1, SIRT6, PIAS4, FEM1A, STAP2, CCDC94, C19orf10, SHD, C19orf29, BRUNOL5, FSD1, UBXD1, DOHH, CREB3L3, HDGF2, RAX2, MPND, ATCAY, DPP9, MRPL54, LRG1, TMIGD2, TNFAIP8L1, C19orf28, GIPC3, TICAM1, ZNRF4, ANKRD24, LSDP5, ARRDC5, KIAA1881	**GZMM*, THEG*, PPAP2C*, C19orf20***
25	10p15.1	0.0610	3	chr10:3931317-3941491	1 gene: KLF6	-
26	3q11.2	0.0676	2	chr3:99583690-99616246	1 gene: OR5K3	-
27	3q13.2	0.0849	2	chr3:113854947-124236792	70 genes: hsa-mir-198, hsa-mir-568, ADPRH, ATP6V1A, CASR, CD80, CD86, CSTA, DRD3, GAP43, GOLGB1, GSK3B, GTF2E1, HCLS1, HGD, KPNA1, LSAMP, NDUFB4, SLC15A2, UPK1B, ZNF80, B4GALT4, NR1I2, STXBP5L, IQCB1, COX17, POLQ, FSTL1, C3orf17, ZBTB20, C3orf28, GTPBP8, C3orf1, PLA1A, FBXO40, SEMA5B, WDR5B, PARP14, GRAMD1C, SIDT1, TMEM39A, WDR52, EAF2, KTELC1, CDGAP, KIAA1407, POPDC2, HSPBAP1, QTRTD1, NAT13, PARP9, DIRC2, C3orf15, BOC, LRRC58, CCDC58, CD200R1, DTX3L, CCDC52, IGSF11, C3orf30, PARP15, GPR156, VSTM3, KIAA2018, ZDHHC23, RABL3, ILDR1, CD200R2, ARGFX	**LSAMP**

28	6p21.1	0.0877	4	chr6:41292490-41342376	1 gene: TREML4	-
29	16p13.12	0.0877	2	chr16:14139929-14166178	1 gene: MKL2	-
30	2p11.2	0.125	2	chr2:85892717-85927786	1 gene: ST3GAL5	-
31	13q32.3	0.133	1	chr13:1-114142980	302 genes	***RB1*, LATS2**
32	1p36.13	0.149	2	chr1:19552461-19610776	1 gene: CAPZB	-
33	4p12	0.153	2	chr4:47936218-47954681	1 gene: TEC	-
34	1p36.11	0.155	3	chr1:25408676-28385091	64 genes: CD52, EXTL1, EYA3, FGR, IFI6, SFN, GPR3, HMGN2, STMN1, PAFAH2, PPP1R8, PTAFR, RHCE, RHD, RPA2, RPS6KA1, SLC9A1, SLC30A2, ARID1A, NR0B2, FCN3, MAP3K6, C1orf38, WASF2, CNKSR1, NUDC, WDTC1, TMEM50A, STX12, SYF2, LDLRAP1, AHDC1, SMPDL3B, ZNF593, ATPBD1B, AIM1L, XKR8, TMEM57, PIGV, FAM54B, C1orf63, MAN1C1, SEPN1, GPATCH3, CCDC21, C1orf135, LIN28, GRRP1, DHDDS, SH3BGRL3, C1orf160, ZDHHC18, TRIM63, SYTL1, UBXD5, FAM46B, C1orf172, PDIK1L, PAQR7, FAM76A, ZNF683, CATSPER4, LOC388610, CD164L2	**SFN**
35	8q12.3	0.155	2	chr8:60048458-67666248	23 genes: hsa-mir-124a-2, ASPH, CA8, CRH, MYBL1, PDE7A, RAB2A, TTPA, GGH, CYP7B1, MTFR1, TOX, RRS1, BHLHB5, ARMC1, CHD7, TRIM55, DNAJC5B, ADHFE1, RLBP1L1, YTHDF3, C8orf46, NKAIN3	-
36	14q11.2	0.155	6	chr14:19800971-22103068	43 genes: ANG, APEX1, HNRNPC, NP, RNASE1, RNASE2, RNASE3, RNASE4, RNASE6, SALL2, TEP1, TOX4, PARP2, FAM12A, SUPT16H, OR10G3, OR10G2, OR4E2, SLC39A2, ZNF219, OSGEP, FLJ10357, METTL3, RPGRIP1, NDRG2, CHD8, CCNB1IP1, FAM12B, METT11D1, RNASE7, RAB2B, TMEM55B, TTC5, RNASE11, TPPP2, RNASE8, KLHL33, RNASE10, OR6S1, RNASE9, OR5AU1, RNASE13, RNASE12	**ZNF219, NDRG2**
37	11q22.3	0.163	3	chr11:103469638-116739377	89 genes: hsa-mir-34c, ACAT1, APOA1, APOA4, APOC3, ATM, CASP1, CASP4, CASP5, CRYAB, DDX10, DLAT, DRD2, FDX1, GRIA4, GUCY1A2, HSPB2, HTR3A, IL18, NCAM1, NNMT, NPAT, PAFAH1B2, POU2AF1, PPP2R1B, PTS, RDX, SDHD, SLN, TAGLN, ZBTB16, CUL5, ZNF259, PCSK7, HTR3B, ZW10, RBM7, CEP164, EXPH5, SNF1LK2, KIAA0999, BACE1, CADM1, REXO2, TIMM8B, SIDT2, C11orf71, SLC35F2, RAB39, BTG4, FAM55D, TTC12, C11orf57, ELMOD1, TEX12, ARHGAP20, USP28, ICEBERG, AASDHPPT, C11orf1, ALG9, PDGFD, TMPRSS5, BCDO2, KIAA1826, BUD13, DIXDC1, ZC3H12C, ALKBH8, LOC91893, C11orf52, COP1, APOA5, PIH1D2, FAM55A, KBTBD3, CWF19L2, KDELC2, LAYN, C11orf65, ANKK1, RNF214, C11orf53, LOC399947, FLJ45803, FLJ46266, INCA, LOC643923, LOC644672	**ATM, CADM1**
38	8q22.1	0.169	4	chr8:95625564-95627290	1 gene: KIAA1429	-
39	5q33.1	0.205	2	chr5:151480736-151502520	1 gene: GLRA1	-
40	4q31.3	0.223	2	chr4:153897161-153915365	1 gene: TIGD4	-

1. Rank according to Q-value. Q-value are false-discovery rate corrected.
2. Cytoband from the middle of the minimal common region.
3. Region represents the wider area that results when the samples defining each border of the minimal common region are left out.
4. Genes of interest related to region:
   Genes in bold represent the most significantly deleted genes in cancer from Beroukhim *et al.* (PMID: 20164920, Supplementary Table 2: GISTIC analysis of 3,131 cancer specimens of 26 tumor types).
   Genes in italics represent genes that are in the region that are mentioned in the text
   Genes with * represent genes adjacent to the region