Table I. Differences in the immunohistochemical staining of 8-nitroguanine and HIF-1α between deceased and living patients with malignant fibrous histiocytoma.

Staining rate (%)	8-Nitroguanine		HIF-1α	
	Living patients (n=20)	Deceased patients (n=16)	Living patients (n=20)	Deceased patients (n=16)
0.0-7.5	9 (45.00)[a]	1 (6.25)	11 (55.00)	3 (18.75)
7.5-15.0	9 (45.00)	2 (12.50)	5 (25.00)	2 (12.50)
15.0-22.5	2 (10.00)	9 (56.25)	4 (20.00)	4 (25.00)
22.5-30.0	0 (0.00)	2 (12.50)	0 (0.00)	5 (31.25)
>30.0	0 (0.00)	2 (12.50)	0 (0.00)	2 (12.50)
P-value	0.00075		0.01359	

[a]No. of patients (%).

life-table method of Kaplan-Meier, and then statistically analyzed by the generalized Wilcoxon test. P-values of <0.05 were considered to be statistically significant.

Results

Histopathological findings, 8-nitroguanine formation and HIF-1α expression in tumor tissues of MFH patients. Fig. 1A shows the histopathological observations of 8-nitroguanine formation and HIF-1α expression in the specimens from MFH patients. 8-Nitroguanine was predominately observed in the nuclei of the tumor cells and inflammatory cells within the MFH tissue specimens. HIF-1α expression was detected in the cytoplasm and nuclei of tumor cells. Little or no immunoreactivity of 8-nitroguanine and HIF-1α was observed in adjacent non-tumor tissues (Fig. 1A, right). Significantly higher levels of both 8-nitroguanine and HIF-1α were observed in the tissue specimens of deceased patients than in those of living subjects (Fig. 1A and Table I).

Fig. 1B shows the colocalization of 8-nitroguanine (red) with HIF-1α (green) visualized using a double immunofluorescence technique. The expression of HIF-1α was strongly detected in the nucleus of tumor cells and was co-localized with 8-nitroguanine (Fig. 1B). In addition, immunoreactivities of 8-nitroguanine and HIF-1α were observed in giant cells and inflammatory cells from MFH tissues (data not shown).

Comparison of the immunoreactivity of 8-nitroguanine and HIF-1α, in relation to prognosis. Table I demonstrates that the immunoreactivity of 8-nitroguanine was significantly stronger in the tumor tissues of deceased patients than in those of living patients. The generalized Wilcoxon test, using the Kaplan-Meier method, was used to evaluate the association of 8-nitroguanine and HIF-1α with the prognosis of MFH patients (Fig. 2). MFH patients with high-grade staining (≥15%) of 8-nitroguanine (p=0.00003) had a significantly shorter survival than those with low-grade staining (<15%)

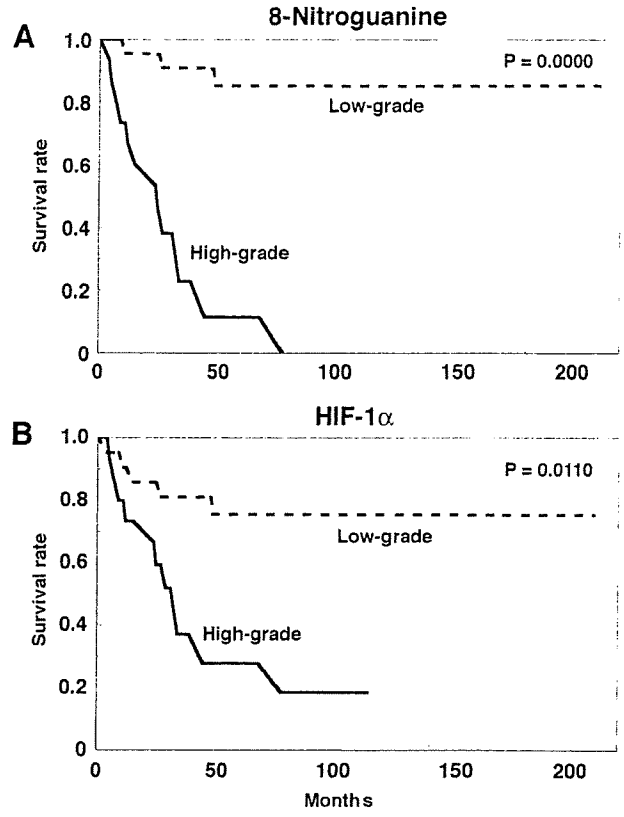


Figure 2. The survival curves of MFH patients as determined by the Kaplan-Meier method. Solid lines, patients with high-grade staining (≥15% detected by Lumina Vision) of 8-nitroguanine (A) or HIF-1α (B). Broken lines, patients with low-grade staining (<15%).

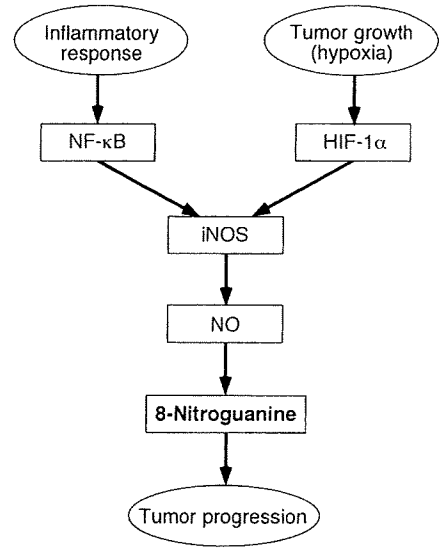


Figure 3. The proposed mechanism of inflammation and hypoxia-induced 8-nitroguanine formation leading to tumor progression.

(Fig. 2A). Meanwhile, patients with high-grade staining of HIF-1α (p=0.01104) also exhibited a significantly poor prognosis in comparison to those with low-grade staining (Fig. 2B).

Discussion

The accumulation of 8-nitroguanine formation and HIF-1α expression was examined in surgical specimens of MFH patients and was found to correlate with the prognosis of inflammation-related cancer. 8-Nitroguanine was clearly co-localized with HIF-1α in the nuclei of tumor cells, giant cells and inflammatory cells in MFH tissues. There was a significant difference in the immunoreactivity of 8-nitroguanine (p=0.00075) between the deceased and living patients, with HIF-1α (p=0.01359) exhibiting similar differences. The survival curves of MFH patients, according to the Kaplan-Meier method, differed more clearly between the groups distinguished by 8-nitroguanine (p=0.00003) rather than by HIF-1α (p=0.01104).

Intratumoral hypoxia induces a rapid increase of HIF-1 protein (22,23), a heterodimer consisting of α and ß subunits, in tumor cells. HIF-1α expression is related to cellular oxygen status (24). Under hypoxic conditions, the degradation of HIF-1α is suppressed and, subsequently, HIF-1α dimerizes with HIF-1ß in the nucleus, thus promoting the expression of numerous target genes (18). iNOS is one of the known HIF-1-inducible proteins. Previous studies have shown that HIF-1α is expressed in the nuclei of tumor cells, suggesting that HIF-1α is activated in MFH tumor tissues under hypoxic conditions. HIF-1α expression in tumor cells correlated with iNOS expression in MFH tissues (R=0.58923, p=0.00194), and double immunofluorescence revealed that 8-nitroguanine and HIF-1α were colocalized in cancerous tissues. In addition, our previous findings demonstrated that 8-nitroguanine was formed via HIF-1α-dependent iNOS expression in relation to the progression of cholangiocarcinoma (25). It therefore appears likely that tumor hypoxia induces HIF-1α expression, which mediates iNOS expression and results in 8-nitroguanine formation.

NF-κB is considered to be a key player in inflammation since it regulates the expression of various genes involved in controlling inflammatory response, including iNOS expression (26). NF-κB functions as a tumor promoter in inflammation-associated cancers (27). Previous studies have reported that 8-nitroguanine is colocalized with iNOS and NF-κB in tumor cells of MFH patients (10). Therefore, 8-nitroguanine can be formed through iNOS expression, mediated by NF-κB activation.

8-Nitroguanine is formed under inflammatory conditions and plays a substantial role in inflammation-related carcinogenesis, including MFH (10). Chemically unstable 8-nitroguanine formed in DNA can be spontaneously released, thereby causing the formation of an apurinic site (5). This can form a base-pair with adenine during DNA synthesis, resulting in G→T transversions (28). Translesion DNA synthesis past an apurinic site mediated by DNA polymerase ζ may contribute to point mutations (29). As a result, inflammatory responses may participate in tumor progression through the formation of mutagenic DNA lesions, such as 8-nitroguanine.

Based on these findings and previous research, a mechanism of tumor progression by hypoxia and inflammation can be proposed, as illustrated in Fig. 3. One pathway is initiated by 'hypoxia in tumor growth', where HIF-1α triggers the iNOS expression, and then 8-nitroguanine formation facilitates tumor progression. The other pathway starts with the 'inflammatory response', where NF-κB is activated by a variety of stimuli including inflammatory cytokines, and iNOS is expressed. These pathways converge into a common pathway, namely iNOS-dependent 8-nitroguanine formation. Consequently 8-nitroguanine, together with HIF-1α, is considered to be an excellent candidate prognostic and predictive biomarker in inflammation-related cancer, including MFH.

Acknowledgements

This study was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology and the Ministry of Health, Labour and Welfare of Japan.

References

1. Coussens LM and Werb Z: Inflammation and cancer. Nature 420: 860-867, 2002.
2. Balkwill F and Mantovani A: Inflammation and cancer: back to Virchow? Lancet 357: 539-545, 2001.
3. Hussain SP, Hofseth LJ and Harris CC: Radical causes of cancer. Nat Rev Cancer 3: 276-285, 2003.
4. Ohshima H, Tatemichi M and Sawa T: Chemical basis of inflammation-induced carcinogenesis. Arch Biochem Biophys 417: 3-11, 2003.
5. Yermilov V, Rubio J, Becchi M, Friesen MD, Pignatelli B and Ohshima H: Formation of 8-nitroguanine by the reaction of guanine with peroxynitrite *in vitro*. Carcinogenesis 16: 2045-2050, 1995.
6. Kawanishi S, Hiraku Y, Pinlaor S and Ma N: Oxidative and nitrative DNA damage in animals and patients with inflammatory diseases in relation to inflammation-related carcinogenesis. Biol Chem 387: 365-372, 2006.
7. Kawanishi S and Hiraku Y: Oxidative and nitrative DNA damage as biomarker for carcinogenesis with special reference to inflammation. Antioxid Redox Signal 8: 1047-1058, 2006.
8. Ma N, Tagawa T, Hiraku Y, Murata M, Ding X and Kawanishi S: 8-Nitroguanine formation in oral leukoplakia, a premalignant lesion. Nitric Oxide 14: 137-143, 2006.
9. Pinlaor S, Ma N, Hiraku Y, Yongvanit P, Semba R, Oikawa S, Murata M, Sripa B, Sithithaworn P and Kawanishi S: Repeated infection with *Opisthorchis viverrini* induces accumulation of 8-nitroguanine and 8-oxo-7,8-dihydro-2'-deoxyguanine in the bile duct of hamsters via inducible nitric oxide synthase. Carcinogenesis 25: 1535-1542, 2004.
10. Hoki Y, Hiraku Y, Ma N, Murata M, Matsumine A, Nagahama M, Shintani K, Uchida A and Kawanishi S: iNOS-dependent DNA damage in patients with malignant fibrous histiocytoma in relation to prognosis. Cancer Sci 98: 163-168, 2007.
11. Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, Feuer EJ and Thun MJ: Cancer statistics, 2004. CA Cancer J Clin 54: 8-29, 2004.
12. Weiss SW and Enzinger FM: Malignant fibrous histiocytoma: an analysis of 200 cases. Cancer 41: 2250-2266, 1978.
13. Belal A, Kandil A, Allam A, Khafaga Y, El-Husseiny G, El-Enbaby A, Memon M, Younge D, Moreau P, Gray A and Schultz H: Malignant fibrous histiocytoma: a retrospective study of 109 cases. Am J Clin Oncol 25: 16-22, 2002.
14. Randall RL, Albritton KH, Ferney BJ and Layfield L: Malignant fibrous histiocytoma of soft tissue: an abandoned diagnosis. Am J Orthop 33: 602-608, 2004.
15. Hockel M, Schlenger K, Aral B, Mitze M, Schaffer U and Vaupel P: Association between tumor hypoxia and malignant progression in advanced cancer of the uterine cervix. Cancer Res 56: 4509-4515, 1996.
16. Schindl M, Schoppmann SF, Samonigg H, Hausmaninger H, Kwasny W, Gnant M, Jakesz R, Kubista E, Birner P and Oberhuber G: Overexpression of hypoxia-inducible factor 1α is associated with an unfavorable prognosis in lymph node-positive breast cancer. Clin Cancer Res 8: 1831-1837, 2002.

17. Birner P, Gatterbauer B, Oberhuber G, Schindl M, Rossler K, Prodinger A, Budka H and Hainfellner JA: Expression of hypoxia-inducible factor-1α in oligodendrogliomas: its impact on prognosis and on neoangiogenesis. Cancer 92: 165-171, 2001.
18. Shintani K, Matsumine A, Kusuzaki K, Matsubara T, Satonaka H, Wakabayashi T, Hoki Y and Uchida A: Expression of hypoxia-inducible factor (HIF)-1α as a biomarker of outcome in soft-tissue sarcomas. Virchows Arch 449: 673-681, 2006.
19. Weiss SW and Goldbum JR: Enzinger and Weiss's Soft Tissue Tumors. 4th edition. Mosby Co., St. Louis, 2002.
20. Greene FL, Page DL, Fleming ID, Fritz A, Balch CM, Haller DG and Morrow M (eds): AJCC Cancer Staging Manual. Springer, Berlin, 2002.
21. Pinlaor S, Hiraku Y, Ma N, Yongvanit P, Semba R, Oikawa S, Murata M, Sripa B, Sithithaworn P and Kawanishi S: Mechanism of NO-mediated oxidative and nitrative DNA damage in hamsters infected with *Opisthorchis viverrini*: a model of inflammation-mediated carcinogenesis. Nitric Oxide 11: 175-183, 2004.
22. Harris AL: Hypoxia - a key regulatory factor in tumour growth. Nat Rev Cancer 2: 38-47, 2002.
23. Semenza GL: Hypoxia, clonal selection, and the role of HIF-1 in tumor progression. Crit Rev Biochem Mol Biol 35: 71-103, 2000.
24. Wang GL, Jiang BH, Rue EA and Semenza GL: Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O_2 tension. Proc Natl Acad Sci USA 92: 5510-5514, 1995.
25. Pinlaor S, Sripa B, Ma N, Hiraku Y, Yongvanit P, Wongkham S, Pairojkul C, Bhudhisawasdi V, Oikawa S, Murata M, Semba R and Kawanishi S: Nitrative and oxidative DNA damage in intrahepatic cholangiocarcinoma patients in relation to tumor invasion. World J Gastroenterol 11: 4644-4649, 2005.
26. Surh YJ, Chun KS, Cha HH, Han SS, Keum YS, Park KK and Lee SS: Molecular mechanisms underlying chemopreventive activities of anti-inflammatory phytochemicals: down-regulation of COX-2 and iNOS through suppression of NF-κB activation. Mutat Res 480-481: 243-268, 2001.
27. Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, Gutkovich-Pyest E, Urieli-Shoval S, Galun E and Ben-Neriah Y: NF-κB functions as a tumour promoter in inflammation-associated cancer. Nature 431: 461-466, 2004.
28. Loeb LA and Preston BD: Mutagenesis by apurinic/apyrimidinic sites. Annu Rev Genet 20: 201-230, 1986.
29. Wu X, Takenaka K, Sonoda E, Hochegger H, Kawanishi S, Kawamoto T, Takeda S and Yamazoe M: Critical roles for polymerase ζ in cellular tolerance to nitric oxide-induced DNA damage. Cancer Res 66: 748-754, 2006.

Analysis of Multiple Sarcoma Expression Datasets: Implications for Classification, Oncogenic Pathway Activation and Chemotherapy Resistance

Panagiotis A. Konstantinopoulos[1☯], Elena Fountzilas[1☯], Jeffrey D. Goldsmith[2], Manoj Bhasin[3], Kamana Pillay[1], Nancy Francoeur[1], Towia A. Libermann[3], Mark C. Gebhardt[4], Dimitrios Spentzos[1,3]*

1 Division of Hematology/Oncology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts, United States of America, **2** Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts, United States of America, **3** Genomics Center and Division of Interdisciplinary Medicine and Biotechnology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts, United States of America, **4** Department of Orthopedic Surgery, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Background: Diagnosis of soft tissue sarcomas (STS) is challenging. Many remain unclassified (not-otherwise-specified, NOS) or grouped in controversial categories such as malignant fibrous histiocytoma (MFH), with unclear therapeutic value. We analyzed several independent microarray datasets, to identify a predictor, use it to classify unclassifiable sarcomas, and assess oncogenic pathway activation and chemotherapy response.

Methodology/Principal Findings: We analyzed 5 independent datasets (325 tumor arrays). We developed and validated a predictor, which was used to reclassify MFH and NOS sarcomas. The molecular "match" between MFH and their predicted subtypes was assessed using genome-wide hierarchical clustering and Subclass-Mapping. Findings were validated in 15 paraffin samples profiled on the DASL platform. Bayesian models of oncogenic pathway activation and chemotherapy response were applied to individual STS samples. A 170-gene predictor was developed and independently validated (80-85% accuracy in all datasets). Most MFH and NOS tumors were reclassified as leiomyosarcomas, liposarcomas and fibrosarcomas. "Molecular match" between MFH and their predicted STS subtypes was confirmed both within and across datasets. This classification revealed previously unrecognized tissue differentiation lines (adipocyte, fibroblastic, smooth-muscle) and was reproduced in paraffin specimens. Different sarcoma subtypes demonstrated distinct oncogenic pathway activation patterns, and reclassified MFH tumors shared oncogenic pathway activation patterns with their predicted subtypes. These patterns were associated with predicted resistance to chemotherapeutic agents commonly used in sarcomas.

Conclusions/Significance: STS profiling can aid in diagnosis through a predictor tracking distinct tissue differentiation in unclassified tumors, and in therapeutic management via oncogenic pathway activation and chemotherapy response assessment.

Citation: Konstantinopoulos PA, Fountzilas E, Goldsmith JD, Bhasin M, Pillay K, et al. (2010) Analysis of Multiple Sarcoma Expression Datasets: Implications for Classification, Oncogenic Pathway Activation and Chemotherapy Resistance. PLoS ONE 5(4): e9747. doi:10.1371/journal.pone.0009747

Editor: Chad Creighton, Baylor College of Medicine, United States of America

Received September 28, 2009; **Accepted** January 21, 2010; **Published** April 1, 2010

Funding: This work was supported by institutional funds (Dr. Spentzos) and the Clinical Investigator Training Program (Dr. Konstantinopoulos) from Beth Israel Deaconess Medical Center and Harvard-MIT Division of Health Sciences and Technology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dspentzo@bidmc.harvard.edu

☯ These authors contributed equally to this work.

Introduction

Soft tissue sarcomas (STS) are a heterogeneous group of mesenchymal tumors traditionally classified according to their morphological resemblance to presumptive cells of origin such as fibroblasts, muscle cells, adipocytes or peripheral nerve-sheath cells [1,2,3]. Given their heterogeneity, sarcomas are ideal candidates for molecularly targeted therapies [4,5]. However, the therapeutic value of current histology-based classification remains unclear. In addition, precise classification is only partially possible, because current histopathologic classification criteria are often inconclusive reflecting the overlapping boundaries between conventional diagnostic groups [6]. This is best exemplified in the case of malignant fibrous histiocytoma (MFH), the second largest subtype by conventional criteria (approximately 20% of cases [7]), a controversial diagnosis which has lately been called in doubt [2,8]. Furthermore, a significant fraction of STS tumors are unclassifiable, presently called "not otherwise specified" (NOS) [8].

Gene expression profiling has been used in the study of STS [9,10,11]. However, these studies were limited by sample size or sample selection, thus clinically applicable diagnostic classifica-

tion models either have not been reported or have not been independently validated, particularly for the MFH, NOS, and pleomorphic subtypes [12,13]. Furthermore, the therapeutic utility of their findings was limited by the inability to predict activation status of relevant oncogenic pathways in a given individual tumor specimen, as opposed to presenting average expression patterns in predefined tumor subgroups. In order to address these challenges, we integrated five publicly available microarray datasets originating from different laboratories around the world, to develop and validate a class predictor, which we then used to molecularly characterize and reclassify MFH and uncategorized sarcomas. Further, we validated these findings in a new group of paraffin derived tumors. Finally, in order to assess the therapeutic relevance of genomic classification, we used computational models to determine the activation status of oncogenic pathways, for which specific targeted inhibitors are in clinical development in sarcoma, and studied the association of pathway activation with histologic subtype and chemoresponse.

Results

Development of a multi-gene predictor in the training dataset

We used study cohort 1 (NCI, the largest dataset including all 6 aforementioned subtypes) in order to train a nearest centroid classifier of the 6 subtypes (LEIO, LIPO, RHAB, MPNST, SYN and FIBRO) frequently presenting differential diagnosis problems, (Figure 1). A single step 6-class model could not be developed because of the significant gene expression overlap between MPNST and SYN classes. Therefore a two-step classifier was defined (Figure 2). In the first step, a 138-gene model distinguished LEIO, LIPO, FIBRO, RHAB and a composite class including MPNST and SYN. In the second step, the composite class is separated into MPNST and SYN tumors using a 35-gene model. Due to a partial gene overlap between the two-step predictors the combined model included 170 genes.

This optimal predictor was 85% accurate for the 6 classes (Figure 2 and Table S1). The genelists of the 1st and 2nd step predictors are shown in Tables S3 and S4. The distinct expression patterns of the first step and second step classifiers in the NCI dataset are displayed in Figures 3 and 4 respectively. Detailed training accuracies for all classes for the nearest centroid predictor are shown in Tables S1 and S2.

Importantly, the predictor included several genes associated with distinct differentiation states (i.e. fibroblastic, smooth muscle, adipocytic and peripheral nerve differentiation) (Figure 3). Appropriately, these genes were overexpressed in the corresponding subtypes.

The 170-gene predictor accurately classifies STS subtypes in four independent datasets

We mapped the 170-gene set across the different platforms of the 4 datasets in study cohort 2, and despite the many technical differences among them we were able to reproduce its performance. Specifically, its accuracy was 86%, 78%, 79%, and 84% in the MSKCC, Stanford, Japan and UK datasets, respectively by leave one out cross validation (permutation $p<0.001$ in all cases) (Figure 2, Table S1). Detailed accuracy, sensitivity and specificity for each class in the training and validation datasets are shown in Tables S1 and S2. Due to platform differences, a more direct validation of the predictor accuracy was only possible among the 3 U 133 datasets, after allowing for gene content mismatch compared with the NCI cDNA original predictor. Thus, training the (modified) predictor on each of the U 133 datasets and applying it to the other U 133 datasets, we obtained 70–75% accuracy.

Reclassification of MFH and NOS samples using the 170-gene STS predictor

We used the 170-gene predictor in study cohort 3, to reclassify 76 MFH and 10 NOS samples. As noted, the MFH and NOS samples were not used in the development or validation of the predictor.

The majority (68 out of 76) MFH tumors were predicted as liposarcomas (46%-35 samples), fibrosarcomas (29%-22 samples) and leiomyosarcomas (14%-11 samples), while 7 out of 10 NOS tumors (in the NCI dataset) were also predicted as liposarcomas (3), leiomyosarcomas (3) and fibrosarcomas (1). The remaining MFH and NOS tumors (11 samples) were predicted as malignant peripheral nerve sheath tumors (4 samples), synovial sarcomas (6 samples) and rhabdomyosarcoma (1 sample).

'Molecular match' between reclassified MFH and NOS samples and their corresponding STS subtypes

We reasoned that if our proposed MFH reclassification using the 170-gene predictor is valid, it should reflect the overall molecular similarity between reclassified MFH and their corresponding subtypes, above and beyond the prediction by the 170-gene predictor. We addressed this question both within each as well as across datasets.

a. Molecular match by clustering within each dataset

We performed unsupervised hierarchical clustering and assessed whether the MFH samples preferentially grouped with samples from their predicted subtype using the top-33% variant genes, i.e. 4200, 7428, 922 probe-sets in the NCI, Japan and MSKCC, and Stanford datasets, respectively. Indeed, 58 out of the 76 reclassified MFH samples (76%) clustered together with samples from their predicted STS subtypes suggesting that the 170-gene predictor reflects an overall 'molecular match' between them. Figure 5 and Figure S2 show hierarchical clustering of the four datasets, where MFH samples reclassified as LIPO, LEIO, FIBRO and SYN, clustered with conventional LIPO, LEIO, FIBRO and SYN samples respectively. The specific clustering results for the classification of MFH samples are presented in Table S7.

Furthermore, MFH-samples predicted as LIPO (MFH-LIPO) did not cluster exclusively with myxoid or with non-myxoid liposarcomas; rather certain MFH-LIPO clustered with myxoid while other MFH-LIPO clustered with non-myxoid liposarcomas suggesting that our predictor is capturing information associated with adipocyte differentiation irrespective of the myxoid or non-myxoid subclassification (Figure S1).

The same analysis was performed for the NOS sample predictions in the NCI dataset, and 6 out of the 10 NOS samples clustered with their predicted subtypes (Figure 5).

b. Molecular match by Subclass Mapping across different datasets

To strengthen the molecular relevance of the MFH reclassification we investigated whether MFH samples were molecularly similar with samples from their predicted STS subtype across different datasets. To achieve this, we used the Subclass Mapping (SubMap) methodology, specifically developed to assess the commonality of subtypes/subclasses in independent and disparate datasets (a candidate subclass is included in the analysis only if it

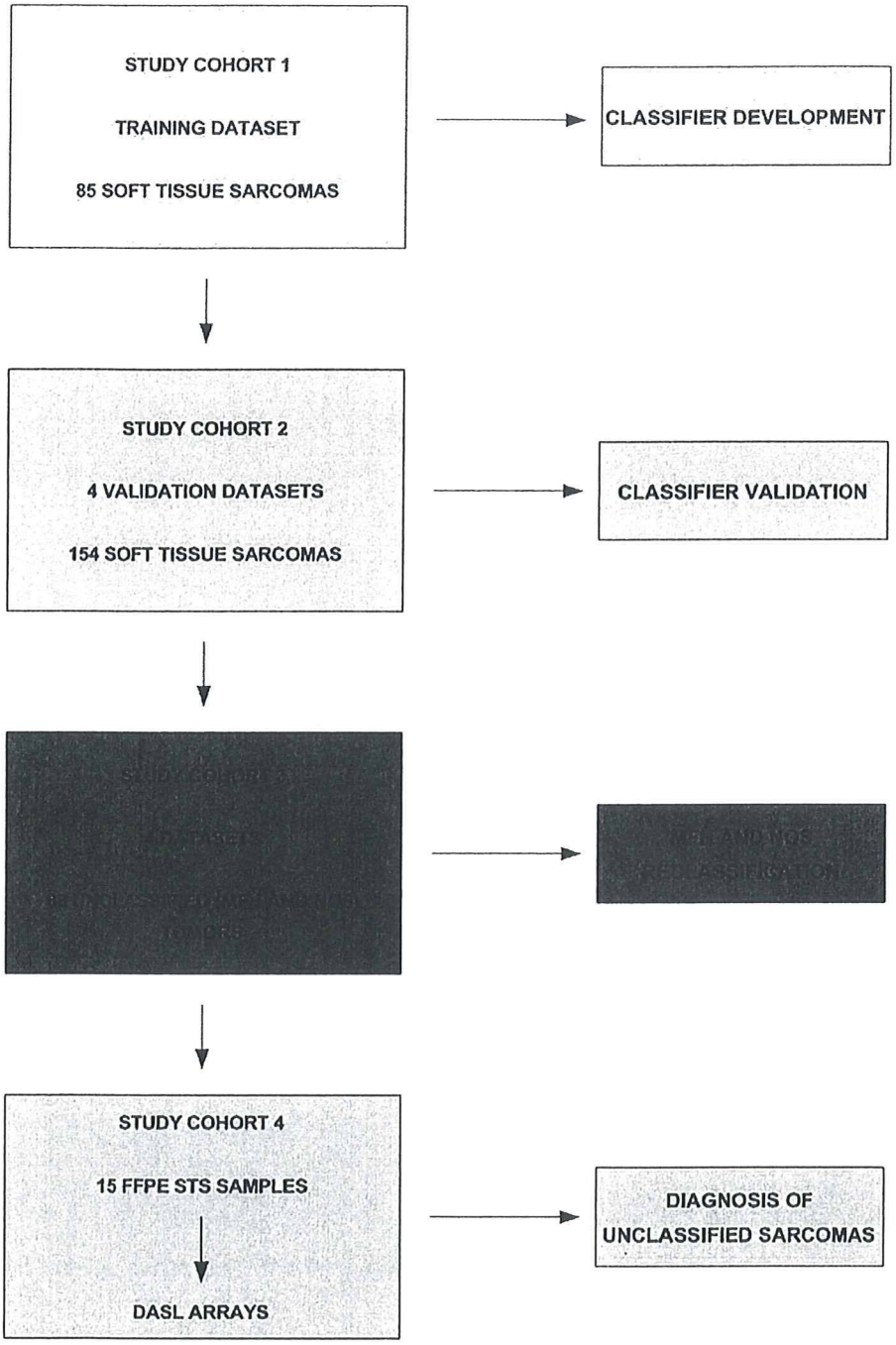


Figure 1. Consort Diagram (Study Design). A multi-class gene expression predictor for 6 major histologic subtypes* was developed in the training dataset (study cohort 1) and validated in 4 independent datasets (study cohort 2). The predictor was used to reclassify MFH (Malignant Fibrous Histiocytoma) and NOS (Not Otherwise Specified) tumors (study cohort 3) into known subtypes. The predictor's performance and capacity to classify unknown type sarcomas were also validated in paraffin STS samples (study cohort 4). * Liposarcoma (LIPO), Leiomyosarcoma (LEIO), Fibrosarcoma (FBR), Malignant Peripheral Nerve Sheath Tumor (MPNST), Synovial Sarcoma (SYN), Rhabdomyosarcoma (RHAB).
doi:10.1371/journal.pone.0009747.g001

contains at least 10% of all the samples of a dataset [14]). As shown in Figure 6 (upper panel), MFH samples from the NCI and Stanford datasets matched their predicted subtypes from the MSKCC and Japan datasets despite their many technical differences. Because of small sample size, this analysis could not be performed for NOS tumors.

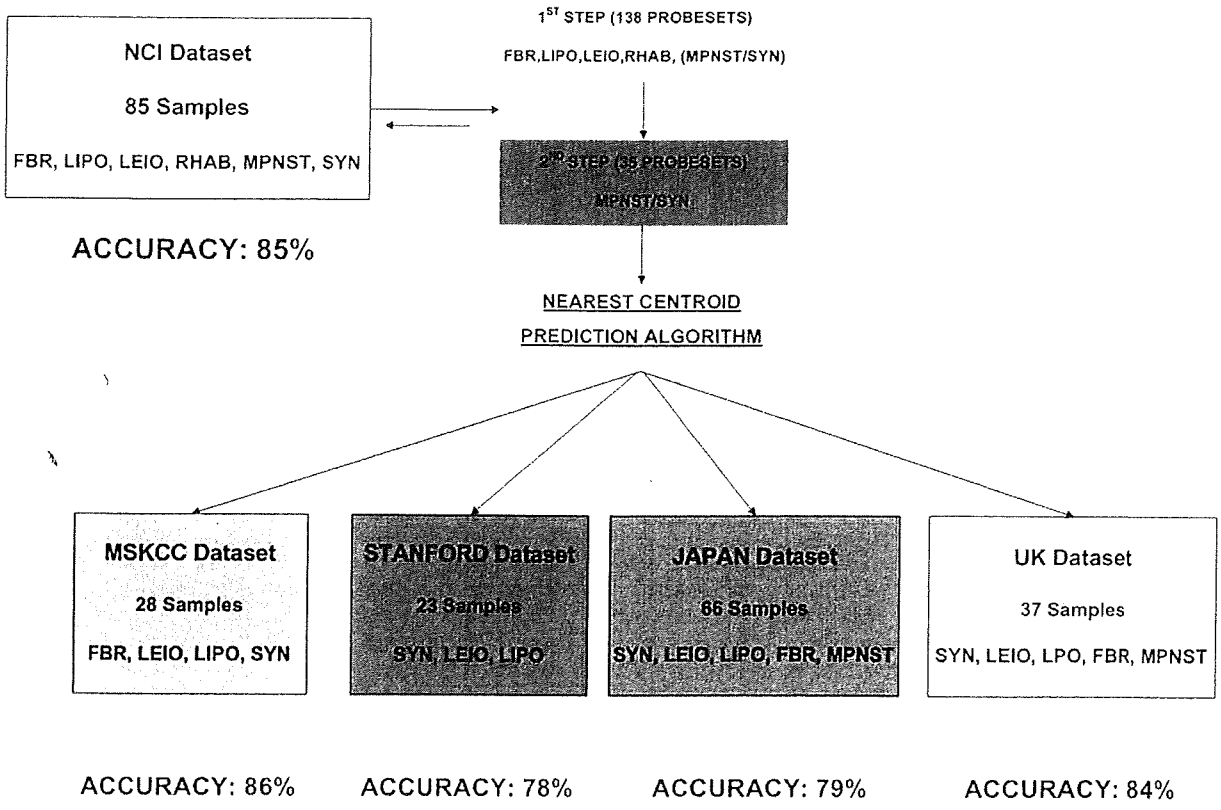


Figure 2. Predictor development and validation. A two-step 6-class predictor was identified in the NCI dataset and validated in the remaining four datasets. First step: A 138-gene model classifies LEIO, LIPO, FBR, RHAB and a composite class including MPNST and SS. Second step: The composite class is separated into MPNST and SYN tumors using a 35-gene model.
doi:10.1371/journal.pone.0009747.g002

Reclassified MFH tumors appropriately overexpress genes associated with distinct differentiation lines

To further demonstrate the molecular basis of reclassifying MFH, we examined whether MFH samples overexpressed genes associated with their predicted differentiation lines. Indeed, MFH tumors predicted as liposarcomas (MFH-LIPO) overexpressed genes associated with adipocyte differentiation compared to the rest of the MFH tumors. Similarly, MFH tumors predicted as leiomyosarcomas (MFH-LEIO) overexpressed genes associated with smooth muscle differentiation and MFH sarcomas predicted as fibrosarcomas (MFH-FIBRO) overexpressed genes associated with fibroblast differentiation (Figure 6). We could not reliably assess specific marker expression for MFH-MPNST, and MFH-SYN given the small number of tumors predicted as these categories. Figure S3 displays the fold upregulation of selected genes associated with smooth muscle, adipocyte and fibroblast differentiation in MFH tumors predicted as leio-, lipo- and fibrosarcomas.

Utility of the STS predictor in unclassifiable paraffin sarcoma specimens

We next evaluated the ability of our STS predictor to reclassify formalin fixed paraffin embedded NOS samples in order to assess its broader applicability for clinical practice and future large scale clinical research. These NOS samples had been previously evaluated by a sarcoma pathology expert (J.G) using state of the art current histopathologic methodology and could not possibly be classified into any of the known STS types.

Before applying our predictor to the unclassified samples, we verified its accuracy in 10 STS samples with known diagnosis. We trained the predictor (modified due to partial gene content mismatch) on the combined U 133 datasets and directly applied it on the independent DASL paraffin gene expression dataset. The STS classifier accurately predicted 8 of the 10 samples, demonstrating accuracy identical to that previously estimated in the 5 public frozen-tissue based datasets, thus validating its performance in samples with known diagnosis and in paraffin tissue. We then directly applied our predictor to the 5 unclassifiable (NOS samples) within the paraffin cohort, and 4 of them were classified as liposarcomas and 1 as leiomyosarcoma. We then examined expression of tissue specific genes in the 4 NOS samples classified as LIPO, and found that they appropriately overexpressed genes associated with adipocyte differentiation including adiponectin, insulin-like growth factor 1, and adipocyte fatty acid binding protein 4 (3.1 fold, 2.4 fold, and 1.5 fold upregulated (t test $p = 0.06$, 0.15 and 0.07 respectively), respectively, as compared to the known non-LIPO samples, (3 LEIO, 2 SYN and 2 MPNST). These findings confirm the utility of the classifier in real time and routinely collected paraffin sarcoma samples and its capacity to detect previously unrecognized tissue differentiation lineage in truly unclassified sarcoma tumors.

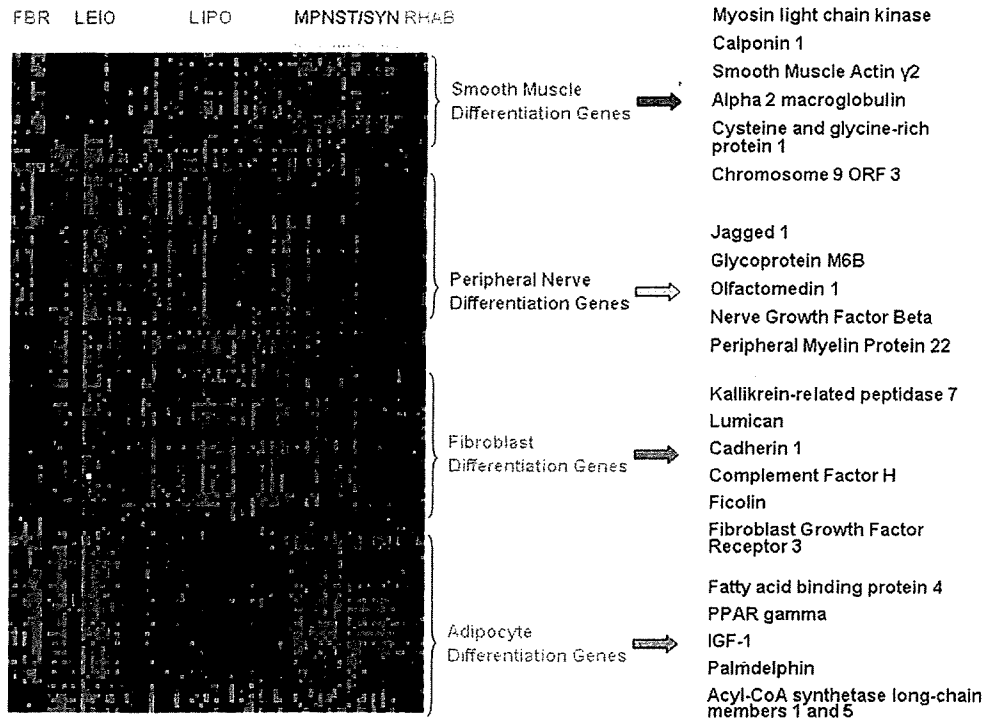


Figure 3. Distinct expression patterns of the first step 138-gene predictor in the NCI dataset. Selected predictor genes associated with distinct differentiation states (smooth muscle, peripheral nerve, fibroblast and adipocyte differentiation) based on Gene Ontology or literature evidence.
doi:10.1371/journal.pone.0009747.g003

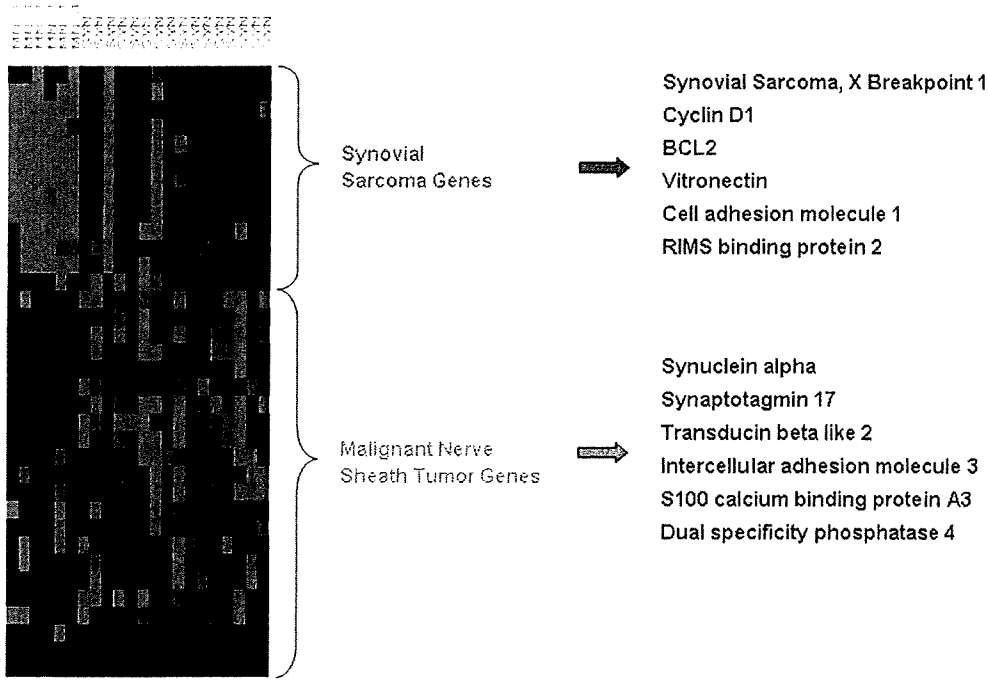


Figure 4. Distinct expression patterns of the second step 35-gene classifier in the NCI dataset. Selected genes overexpressed in synovial and malignant peripheral nerve sheath tumors, are shown on the right, representing potential novel tissue differentiation markers.
doi:10.1371/journal.pone.0009747.g004

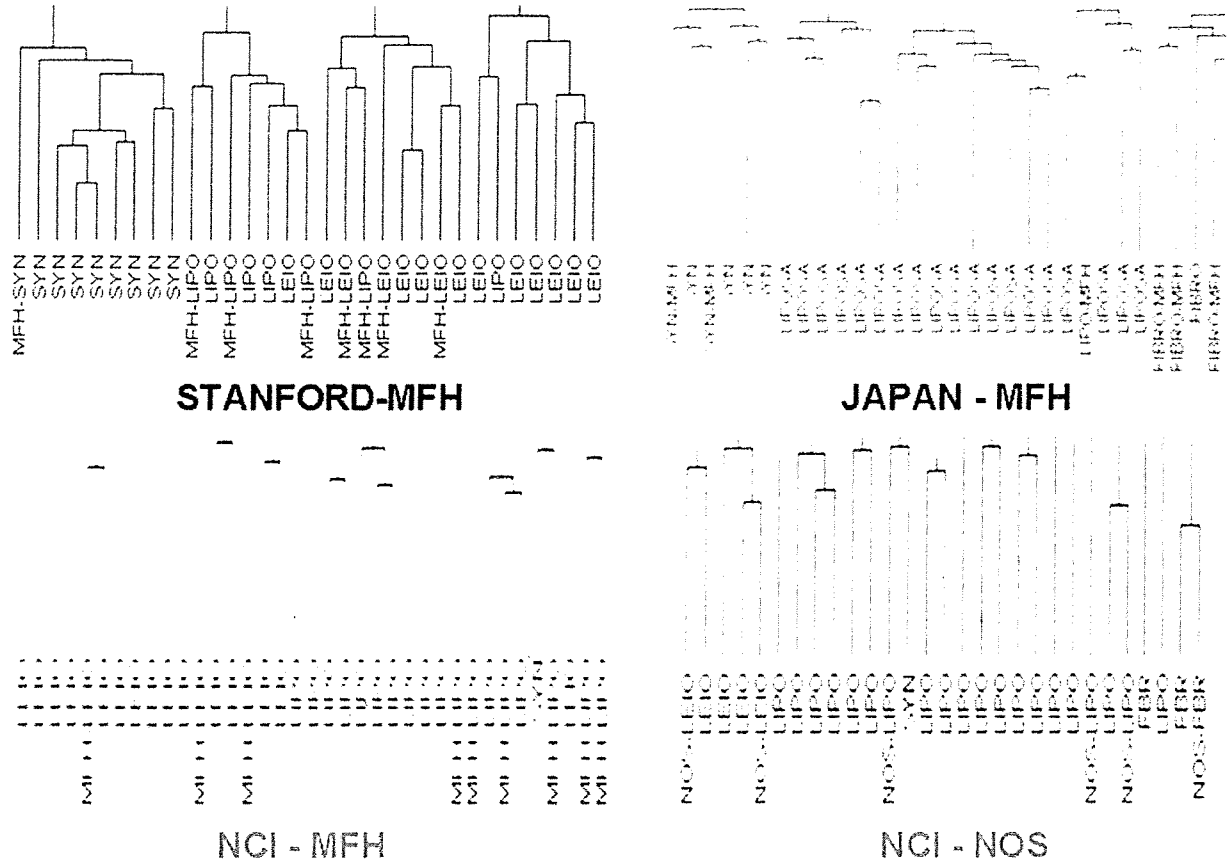


Figure 5. Assessment of MFH/NOS reclassification within each dataset. Genome-wide hierarchical clustering reveals that MFH samples predicted as LIPO, LEIO, and SYN, clustered together with conventional LIPO, LEIO and SYN samples respectively. Similarly, NOS samples reclassified as liposarcomas leiomyosarcomas and fibrosarcomas clustered together with conventional LIPO, LEIO and FBR samples respectively. For Stanford dataset, the complete dendrogram is presented here. For the remaining datasets, due to size limitations, representative portions of the dendrograms are shown. Full dendrograms are shown in Figure S2, which also demonstrate that 24% of MFH samples (in total) did not cluster with the predicted subtypes.
doi:10.1371/journal.pone.0009747.g005

Unique patterns of oncogenic pathway activation in STS subtypes

In order to evaluate whether STS classification bears potential biologic or therapeutic implications, we estimated the probability of activation of known oncogenic pathways in individual samples, using validated gene expression "read outs" previously generated in vitro as a result of controlled experimental activation of these pathways. We focused on Src, Ras and PI3K pathways, for which pharmacologic inhibitors are currently in clinical development in sarcoma, and we assessed tumors from the 3 Affymetrix oligonucleotide U133A datasets (Japan, MSKCC and UK datasets) in our study. Since these gene expression models of pathway activation were generated using oligonucleotide Affymetrix arrays, and most Affymetrix probesets included in these predictors were not present in the cDNA datasets, we did not assess pathway activation in tumors from the 2 cDNA datasets as these predictions would have been less reliable.

The probability of activation of the Src, Ras and PI3K pathways was statistically significantly different between different subtypes (Kruskal-Wallis $p<0.001$, $p=0.002$, $p=0.021$ respectively). More specifically, FIBRO demonstrated higher probability of Ras and PI3K pathway activation (Mann-Whitney $p=0.044$ and $p=0.013$ respectively) and LIPO demonstrated higher probability of Src pathway activation ($p<0.001$) and lower probability of PI3K pathway activation ($p=0.06$) compared to the rest of the samples (Table S8). Conversely, synovial sarcomas were associated with statistically significantly lower probability of Src and Ras pathway activation compared to the rest of the samples ($p=0.005$ and $p<0.001$ respectively). Finally, LEIO samples did not show any particular pattern of activation of any of the Src, Ras or PI3K pathways (Table S8).

Reclassified MFH share similar patterns of oncogenic pathway activation with their corresponding subtypes

In order to assess whether MFH reclassification using our 170-gene predictor is tracking specific oncogenic pathway activation patterns, we evaluated the activation status of Src, Ras and PI3K pathways in the 30 MFH samples of the U133 datasets using the aforementioned gene expression "readouts". Similar to their predicted STS subtypes, MFH sarcomas predicted as fibrosarcomas (MFH-FIBRO) had similarly high average probability of PI3K pathway activation (0.99 vs 0.99 in MFH-FIBRO and FIBRO respectively) and similarly low average probability of Src pathway activation (0.01 vs 0.13 respectively). Furthermore, MFH

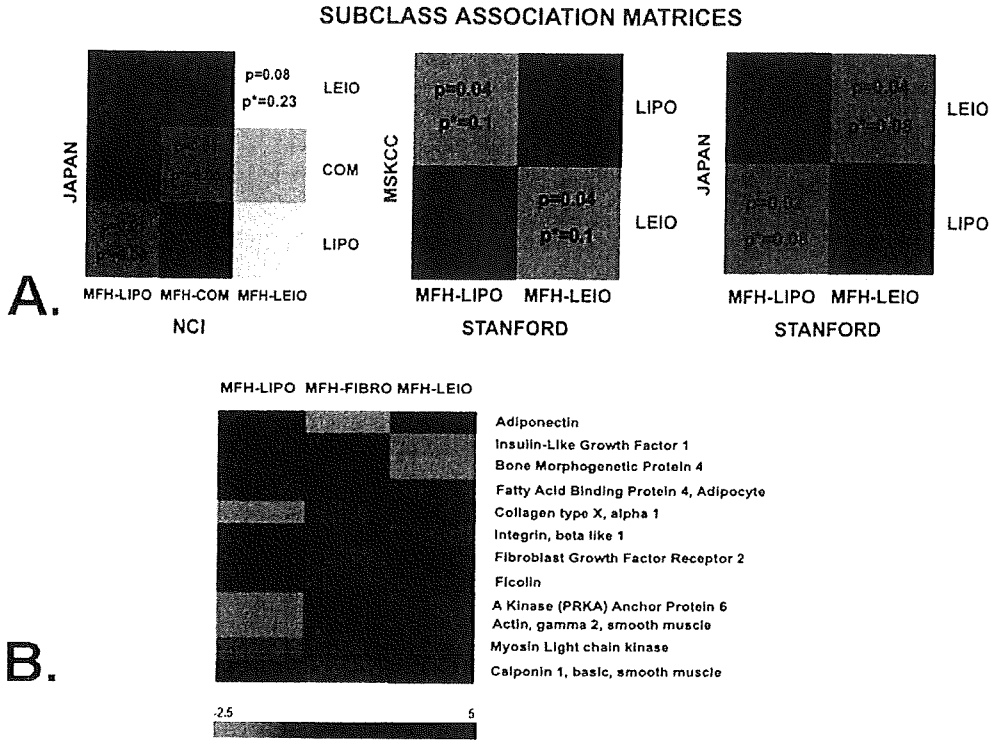


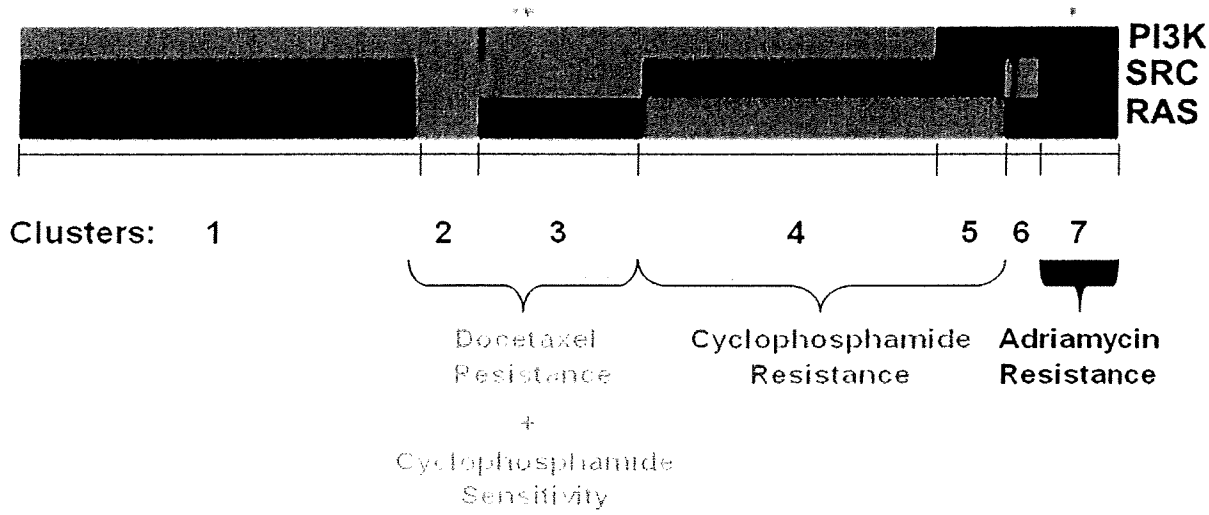
Figure 6. Reclassified MFH samples correspond to their respective STS subtypes across different datasets and express the tissue specific markers of their respective STS subtypes. A) Subclass Association Matrices assessing molecular correspondence of reclassified MFH samples across different datasets: *Left*: NCI versus Japan dataset (COM: composite MPNST-SYN group). *Center*: Stanford versus MSKCC dataset. *Right*: Stanford versus Japan dataset. In all cases p* is Bonferroni-corrected. (**Color scale**: Red and Yellow colors indicate $p<0.05$ and $0.05\leq p<0.1$, respectively suggesting strong molecular correspondence between subtypes in different datasets. Other colors indicate $p>0.1$ suggesting lack of molecular correspondence). Details are provided in Table S5). **B) Expression of tissue specific markers in MFH samples**: Heat map showing upregulation of selected genes associated with smooth muscle, fibroblast and adipocyte differentiation (based on Gene Ontology or literature) in MFH tumors predicted as leio-, lipo-, or fibrosarcoma, respectively, compared to the rest of MFH tumors (t-test $p<0.05$). **Color scale** (saturated at 5-fold upregulation) indicates average fold change for each predicted MFH subclass compared to the rest of MFH tumors. Detailed fold changes are provided in Figure S3.
doi:10.1371/journal.pone.0009747.g006

sarcomas predicted as liposarcomas (MFH-LIPO) had similar average probability of PI3K pathway activation (0.75 vs 0.73 in MFH-LIPO and LIPO respectively). Similar to LEIO, MFH-LEIO did not demonstrate any specific pattern of pathway activation compared to the rest of the samples. There were only two MFH samples predicted as synovial, so no firm conclusions could be reached for this subset.

Distinct patterns of oncogenic pathway activation are associated with chemotherapy resistance

Given the modest predictive value of histology for chemotherapy response, we investigated whether patterns of oncogenic pathway activation are associated with resistance to commonly used chemotherapy agents in sarcoma. For this purpose, we used gene expression readouts (previously developed using Affymetrix arrays from the NCI 60 cancer cell line panel) predicting the probability of resistance to adriamycin, docetaxel and cyclophosphamide, (which was used as a surrogate for ifosfamide, for which NCI 60 resistance data were not available). The probability of resistance to each chemotherapy drug was estimated for each individual sample included in the U133 datasets of our study. As with pathway activation predictions, we did not assess the possibility of chemotherapy resistance in tumors from the two cDNA datasets as these predictions would have been less reliable (most Affymetrix probesets from these predictors were not present in the cDNA datasets).

We then performed hierarchical clustering of the samples based on each sample's probability of activation of Src, Ras and PI3K pathways, and observed that samples were classified into seven clusters with distinct pathway activation patterns. As shown in Figure 7, different patterns of pathway activation were associated with resistance to different chemotherapy agents. Cluster 7 (Src, Ras and PI3K activation) was associated with higher probability of adriamycin resistance (Mann-Whitney $p = 0.002$), clusters 4 and 5 (Src with or without PI3K activation) were associated with higher probability of cyclophosphamide resistance ($p = 0.01$) and clusters 2 and 3 (Ras or no pathway activation) were associated with higher probability of docetaxel resistance ($p = 0.01$) compared to the rest of the samples. Finally, we examined the distribution of the different histologies within the pathway clusters and found that it was random with the exception of liposarcomas being overrepresented (16 out of 58) in cluster 3 and synovial sarcoma being overrepresented in cluster 1 (19 out of 30). However, these clusters did not recapitulate any previously reported associations with chemotherapy response patterns.



Histologies/Clusters	1	2	3	4	5	6	7
LIPO	13	6	16	7	6	5	5
LEIO	8	1	3	7	1	0	0
FIBRO	7	1	1	7	0	0	0
SYN	19	0	2	2	0	0	7
MPNST	1	1	0	4	1	0	0

Figure 7. Association between patterns of oncogenic pathway activation and resistance to chemotherapy drugs. Unsupervised hierarchical clustering of 161 tumor samples based on individual sample probability of Src, Ras and PI3K pathway activation reveals 7 clusters with distinct patterns of pathway activation and association with chemotherapy resistance: Cluster 7 is associated with higher probability of adriamycin resistance ($p = 0.002$), clusters 4 and 5 are associated with higher probability of cyclophosphamide resistance ($p = 0.01$), and clusters 2 and 3 are associated with higher probability of docetaxel resistance ($p = 0.01$), compared to the rest of the samples. The composition of STS histologies in each cluster is also presented.
doi:10.1371/journal.pone.0009747.g007

Gene expression patterns of tumors with activated Ras pathway are enriched for targets of the let-7 miRNA family

It has been postulated that expression profiles may partly be surrogates for microRNA alterations in sarcoma. Previous in vitro data indicate that Ras is regulated by the let-7 microRNA family [15]. In this regard, we tested the hypothesis that the gene expression pattern of clinical samples with predicted activation of Ras pathway, were enriched for gene targets of the let-7 microRNA family. We performed microRNA target gene set enrichment analysis in samples with or without predicted activation of Ras pathway using the functional scoring method. Indeed, we found that the gene expression pattern of the samples with predicted Ras pathway activation was enriched for targets of all microRNAs of the let-7 family ($p<0.05$ for all miRNAs of the let-7 family in all 3 datasets – Table S6), suggesting that let-7 may play a role in Ras activation in human sarcoma tumors.

Discussion

Soft tissue sarcomas are heterogeneous neoplasias, thus conceptually well-suited for application of targeted therapies [4,5,16]. Development of such therapies has been limited by the fact that traditional histopathologic classification has never been shown to carry substantial therapeutic value. This may be partly related to the additional challenge that current histomorphologic classification criteria are frequently inconclusive and do not fully capture the underlying molecular complexity of these tumors, leaving a sizable fraction of them unclassified or grouped in controversial entities, such as MFH [2,6,8,17]. Previous micro-

array studies have analyzed gene expression patterns [12,18,19] yielding evidence of substantial differences among sarcoma subtypes. However, most of the findings were not replicated, and the single previously reported diagnostic predictor was not independently validated [13], thus being subject to limitations related to overfitting [20,21] and single-study bias. Furthermore, despite its theoretical promise, the therapeutic relevance of proposed genomic classification of sarcomas has been difficult to assess, (particularly for those types not associated with a targetable "necessary and sufficient" molecular abnormality), partly because of the inability to predict activation status of multiple relevant oncogenic pathways in a given individual tumor specimen.

In this study, we integrated 5 publicly available sarcoma microarray datasets [12,13,18,19,22], with the intent to address current diagnostic as well as therapeutic challenges in sarcoma management. We, first, identified a 170-gene classifier of six major subtypes (Figure 2) that frequently pose differential diagnosis problems, as they can all present with pleomorphic poorly differentiated variants. This predictor was validated in four independent datasets derived from different laboratories in different parts of the world. Despite the significant inter-laboratory and platform differences (oligonucleotide versus cDNA arrays), the classifier demonstrated 78-86% accuracy across all 4 independent datasets (Figure 2). The classifier included several genes associated with distinct differentiation states (i.e. fibroblastic, myogenic, adipocytic and neural differentiation) (Figure 3) suggesting that assignment of a sample to a specific class reflects its molecular resemblance to differentiated mesenchymal cells such as fibroblasts, muscle cells, adipocytes or peripheral nerve-sheath cells, a resemblance that is otherwise not appreciated by standard histopathologic criteria. Several classifier genes are potential drug development targets, for instance the IGF-1 [23], PPAR gamma [24], NGF beta [25] and FGF receptor 3 genes [26]. While the biologic program responsible for the behavior of sarcomas is likely to encompass a much larger number of gene networks, these observations suggest that the relatively limited classifier gene set may still capture therapeutically relevant mechanisms.

A particular challenge for sarcoma classification remains the elusive nature of "MFH" sarcomas. Microarray studies indicated that MFH tumors comprise a complex group not forming a distinct cluster, raising the possibility that MFH does not represent a unique molecular category [12]. Also, previous morphologic studies have suggested that MFH tumors share similarity with pleomorphic variants of other known subtypes [2,17]. While it has been suggested by the WHO classification, that the terminology "MFH" will be abandoned when criteria for reclassification of pleomorphic sarcoma can be reproducibly defined [2,8], currently no such criteria exist, and the similarity between MFH and other subtypes in these studies was determined by a semi-quantitative histopathologic stains without regard to genome-wide molecular resemblance [2,17]. Of note, MFH tumors with myogenic differentiation behave more aggressively indicating that MFH reclassification may also be prognostically useful [27].

We showed that our validated multi-gene predictor can reclassify MFH and other uncharacterized tumors into one of the major subtypes, the majority predicted as leiomyosarcomas, liposarcomas and fibrosarcomas. The validity of this reclassification was confirmed by sophisticated bioinformatics approaches (Figures 5, 6) including a recently developed method (SubMap), uniquely suitable to assess the resemblance of subtypes identified in multiple, independent, and technically disparate datasets [14]. Furthermore, by examining the expression profiles of the MFH samples, we identified a number of tissue differentiation genes above and beyond the 170-gene predictor that were also appropriately overexpressed according to their predicted histology. In the absence of another "gold standard" metric, the finding of tissue specific genes over expressed in previously unclassified tumors in accordance to their predicted subtype, serves as confirmation of the classification potential of our predictor. Of note, for liposarcomas, we were able to show that our predictor is tracking adipocyte differentiation of MFH irrespective of the myxoid or non-myxoid sub-classification. Finally, but no less importantly, we again validated the performance on the predictor in real life paraffin specimens, including its capacity to reveal tissue lineage in specimens that are currently impossible to classify with state-of-the-art histopathologic examination. These findings, taken together, suggest that our proposed reclassification is not merely a mathematical model function; rather it is tracking real underlying molecular sarcoma "phenotypes" based on tissue differentiation lines. Our analysis supports the concept, proposed originally by Fletcher [17], that MFH represents the end stage of dedifferentiation of many sarcoma subtypes rather than a distinct entity.

It should be noted, however, that it is possible that certain MFH/NOS tumors may be so undifferentiated that they may not harbor any distinct lineage. In this case, our predictor would result in "overclassification" of these samples, forcing them into one of the conventional classes. The extent of this error is not possible to know with certainty. The fraction of MFH samples that did not co-cluster with their predicted subtype in Figure 5 (24%) may be an estimate suggesting that the subset of "overclassified" MFH tumors is small.

The second aim of our study was to explore whether genomic reclassification of sarcomas using our predictor bears biologic and therapeutic implications. Several oncogenic pathway inhibitors are currently undergoing clinical trial evaluation in sarcoma including drugs targeting the PI3K/mTOR (deferolimus and everolimus), Src (dasatinib and AZD0530), and the Ras/Raf pathway (sorafenib). However, the activation status of these pathways in individual sarcoma samples and across different histologic subtypes has not been possible to determine, making it difficult to prioritize patients for targeted therapies. For this purpose, we applied previously validated gene expression "read outs" of oncogenic pathway activation to individual samples, and discovered that different subtypes demonstrate distinct patterns of activation of these oncogenic pathways. Interestingly, reclassified MFH demonstrated similar patterns of oncogenic pathway activation as their corresponding predicted subtypes, providing further evidence that the 170-gene predictor reflects the overall molecular program in sarcomas, with therapeutic implications.

Although we showed an association between different histologies and patterns of oncogenic pathway activation, it is also well-known that the association between histology and chemotherapy resistance is only modest, and in many cases unproven, in soft tissue sarcomas. Our findings suggest that oncogenic pathway activation patterns, transcending histologic classes, and assessed by gene expression "read outs", may serve as useful predictors of resistance to chemotherapy drugs commonly used in sarcoma (Figure 7), perhaps overriding previously considered modest associations between histology and chemoresistance. Although proof of an etiologic association between specific patterns of oncogenic pathway activation and chemotherapy resistance, or *in vitro* demonstration of novel chemoresistance mechanisms were beyond the scope of this study, our findings reveal interesting therapeutic research strategies that can be studied in properly designed prospective studies. For example, prior knowledge of oncogenic pathway activation patterns in individual sarcoma samples may aid in prioritizing patients for novel molecularly targeted agents, conventional chemotherapeutic agents, or combinations thereof. However, our study was limited by the lack of clinical data (i.e. chemotherapy response or outcome

data) linked with the public microarray datasets we used, a fact that prevented us from being able to further test the chemoresistance patterns identified in our analysis.

Finally, the finding that the gene expression patterns of Ras-activated tumor samples are enriched for gene targets of the let-7 microRNA family, consistent with compelling experimental evidence of Ras regulation by let- 7 miRNAs [15], raises the possibility that microRNA alterations may contribute to oncogenic pathway activation, and is consistent with the concept that gene expression patterns may be partly surrogates for microRNA alterations [28,29,30] in primary sarcoma tumors.

Our study demonstrates the power of integrated analysis of multiple and diverse microarray datasets, in order to generate and validate clinically useful models and concepts, in a cost effective manner. We identified a multi-gene STS predictor, reproduced for the first time in multiple independent gene expression datasets, and in routinely collected paraffin tissue, which could serve as an aid to standard histopathologic methods, especially in the diagnosis of pleomorphic tumors that are impossible to classify based on state of the art histopathology. Our findings support the concept that MFH and unclassified (NOS) sarcomas can be reclassified into existing sarcoma subtypes, and proposes a tool for clinical application that reflects previously unrecognized lines of sarcoma differentiation. Finally, our results support the notion that genomic classification may carry potential therapeutic implications, and provide novel therapeutic research hypotheses regarding individualization of targeted therapies and overcoming chemotherapy resistance in soft tissue sarcomas.

Materials and Methods

Assembly and processing of Gene Expression Datasets

Our study included five public microarray datasets [12,13,18,19,22] (Table 1, Figures 1 and 2) with 325 tumors of the six STS subtypes [liposarcomas (LIPO), leiomyosarcomas (LEIO), rhabdomyosarcomas (RHAB), malignant peripheral nerve sheath tumors (MPNST), synovial sarcomas SYN and fibrosarcomas (FIBRO)] that frequently present differential diagnosis problems. Red/green channel expression data were retrieved from 2 cDNA datasets (NCI, Stanford) and raw data were retrieved from 3 oligonucleotide Affymetrix U133A datasets, (UK, Japan, MSKCC). cDNA data were normalized using the median normalization method and Affymetrix .CEL file data were processed using the Robust Multi-Array Average (RMA) algorithm [31]. Analyses described below were performed using the BRB Array Tools package (Dr Richard Simon, NCI), unless noted otherwise.

RNA Isolation from paraffin specimens and Illumina Whole Genome DASL array hybridization

Paraffin specimens from 15 soft tissue sarcomas including 5 NOS were cut into 1-3 mm cores at the BIDMC Histology Core facility. These included 10 paraffin STS samples with known diagnosis (3 LIPO, 3 LEIO, 2 SYN and 2 MPNST) and 5 paraffin NOS samples that had been previously evaluated by a sarcoma pathology expert (J.G) using state of the art current histopathologic methodology and could not possibly be classified into any of the known STS types. These samples, all archived between 2003 and 2006 at the Beth Israel Deaconess Medical Center Pathology Department, were chosen on the basis of tissue availability and adequate RNA yield for microarray studies. IRB approval for tissue utilization was obtained as per standard institutional protocols. Total RNA was isolated using the Qiagen RNeasy formalin-fixed, paraffin-embedded (FFPE) protocol according to the manufacturer's instructions. Whole genome DASL (cDNA-mediated, Annealing, Selection, and Ligation) arrays (Illumina, CA), containing 24,000 gene transcripts were used to profile the paraffin specimens on an Illumina BeadStation. The DASL array experiments were carried out at the Children's Hospital (Boston) Microarray Core facility as per manufacturer's instructions and as previously described [32,33].

Classification analysis design

Figure 1 shows our study workflow. We defined 4 "study cohorts". Study cohort 1 (NCI) was used to optimize a gene expression predictor. Study cohort 2 (4 datasets – Stanford, UK, Japan, MSKCC – not used in step 1) was used to independently validate the predictor. Study cohort 3 (MFH and NOS samples from all datasets, not used in prior steps) was used to reclassify previously uncategorized sarcomas based on the predictor. Finally, this predictor was applied to study cohort 4, which consisted of 15 paraffin STS samples that were profiled using whole-genome DASL (cDNA-mediated, Annealing, Selection, and Ligation) arrays. In this step, we used the predictor to classify NOS samples that were impossible to classify using current, state of the art histopathologic evaluation and staining.

Development and validation of a multi-gene predictor

Using the Nearest Centroid algorithm [34,35] we developed a predictor of 6 subtypes on the NCI dataset –the largest and only one that included all six subtypes. We trained the classifier selecting genes differentially expressed between classes by F-test. Classifier accuracy and statistical significance were assessed using

Table 1. Content of the 5 microarray expression datasets.

DATASETS	PLATFORM	SAMPLES	HISTOLOGIES
NCI	cDNA	133	SYN (16), LEIO (17), LIPO (33), MPNST (6), FIBRO (7), RHAB (6), **MFH (38)**, **NOS (10)**
STANFORD	cDNA	31	SYN (8), LEIO (11), LIPO (4), **MFH (8)**
UK	U133A	37	SYN (10), LEIO (8), LIPO (10), MPNST (4), FIBRO (5)
JAPAN	U133A	87	SYN (16), LEIO (6), LIPO (37), MPNST (3), FIBRO (4), **MFH (21)**
MSKCC	U133A	37	SYN (4), LEIO (6), LIPO (11), FIBRO (7), **MFH (9)**
TOTAL PATIENTS		325	SYN (54), LEIO (48), LIPO (95), MPNST (13), FIBRO (23), RHAB (6), **MFH (76)**, **NOS (10)**

doi:10.1371/journal.pone.0009747.t001

leave-one-out cross-validation and a random permutation test to control for over-fitting. The best-performing training classifier included 138 genes (F test cut off, $p<4\times10^{-7}$) and demonstrated an accuracy of 85%. Sensitivity analysis using an F-test threshold from $p<10^{-7}$ (100 genes) to $p<5\times10^{-7}$ (160 genes) demonstrated only minimal loss of performance with accuracy of 79–82%. This classifier was mapped across different platforms using Affymetrix annotation files and applied to the four independent public datasets as well as the paraffin based dataset. The overlap of predictor genes across different platforms was as follows: i) Among the NCI cDNA platform and the 3 U133 datasets there was complete overlap of the 1st and 2nd step predictor genes (i.e. 138 genes and 35 genes respectively), ii) Between the NCI cDNA platform and the DASL Illumina (paraffin) dataset the overlap was 136 genes for the 1st and 34 for the 2nd step predictor) and iii) Between the NCI cDNA platform and the cDNA (Stanford) dataset the overlap was 62 genes for the 1st and 19 genes for the 2nd step predictor).

Allowing for partial loss of classifier genes and the technical disparity between the different platforms, classifier accuracy was further assessed using leave one out cross validation and a random permutation test in each of the 4 independent public datasets. Further, for a more direct validation of accuracy, we also collated the 3 U133 datasets into one combined dataset after adjusting for platform effect using empirical Bayes method [36]. Then, we trained a modified predictor (using the same genes but creating a different model due to technical platform differences) in the largest U133 dataset (JAPAN), and directly applied it to the remaining datasets (UK and MSKCC) and vice versa. Finally, for prediction of the paraffin samples we collated all 4 datasets (3 U133 and the DASL paraffin dataset) after adjusting for platform effect using empirical Bayes method [36]. The final predictor (modified due to partial gene content mismatch with the U133 platform) was trained on the 3 combined U 133 datasets and directly applied to the DASL dataset.

MFH and NOS reclassification using the multi-gene predictor

We used the predictor to reclassify the 76 MFH and the 10 NOS samples into the 6 subtypes after mapping the predictor genes across different platforms (from cDNA to Affymetrix U133) as above. No mapping was necessary for the NOS samples that were all contained in the NCI dataset.

Validation of MFH and NOS reclassification within each dataset using genome-wide hierarchical clustering

In order to assess whether our MFH reclassification reflected true molecular similarity of the reclassified MFH samples with their corresponding STS subtypes, we performed unsupervised hierarchical clustering of all tumors (including MFH samples) within each dataset using the complete linkage method and the one minus centered correlation as a distance metric [37]. A large number of genes (top 33% variant) were included in this analysis in order to overcome the overfitting bias of the optimized 170-gene predictor. Then, we assessed whether each reclassified MFH sample clustered together with the samples of its corresponding STS class (the class it had been reclassified into) and repeated the same process for the NOS reclassification.

Validation of MFH reclassification across different datasets using Subclass Mapping (SubMap)

Hierarchical clustering cannot assess molecular correspondence between phenotypes across different datasets. Thus, we examined the molecular similarity between reclassified MFH samples and their predicted corresponding STS subtypes from different datasets using the Subclass Mapping (SubMap) methodology as previously described [14] (Gene Pattern Software, Version 3.0, Broad Institute, details in Data S1). This method relies on the principle of statistically assessed "enrichment" of the transcription program of one dataset, for marker gene lists derived from another dataset, as a function of relative differential ranking rather than absolute expression values, since the latter are platform and study-specific. "Mutual enrichment information" p values are generated to assess the "molecular match" between the different phenotypes and summarized in a Subclass Association Matrix. High mutual enrichment (low p value) indicates a strong molecular correspondence between subclasses in different datasets.

Prediction of probability of oncogenic pathway activation or chemotherapy resistance in individual samples

We used publicly available and validated gene expression "read outs" of oncogenic pathway activation previously generated by experimentally controlled activation of these pathways *in vitro* [38]. Furthermore, we retrieved publicly available and validated gene expression models predicting the probability of resistance to individual chemotherapeutic agents that were generated using U133 Affymetrix array and drug response data from the NCI 60 cancer cell line panel [39,40,41]. Bayesian probit regression models estimating the probability of activation of each pathway or resistance to specific chemotherapeutics were trained in the experimental systems used to develop these signatures and applied on individual samples included in the Affymetrix oligonucleotide U133 datasets of our study. Gene expression models ("read outs") of oncogenic pathway activation and chemoresistance are available at http://dig.genome.duke.edu/. Since these models were generated using oligonucleotide Affymetrix arrays and very few Affymetrix probesets were present in the cDNA datasets, our analysis was limited to the 3 Affymetrix oligonucleotide datasets of our study. Non-biological experimental variation between the *in vitro* arrays and the sarcoma datasets was corrected using a previously described batch effect adjustment algorithm [36]. Each individual sample was assigned a probability value (from 0 to 1) of pathway activation or resistance to a specific chemotherapeutic agent. A probability value higher than 0.5 was used as cut-off for predicted pathway activation.

Hierarchical clustering of samples based on their predicted probability values of oncogenic pathway activation was performed using the complete linkage algorithm with the Euclidean distance metric [37]. Non-parametric one-way Kruskal-Wallis and Mann-Whitney tests were applied to test whether the probability of oncogenic pathway activation and chemoresistance was different between different subtypes.

MicroRNA gene-target enrichment analysis

To assess whether the gene expression patterns of STS samples were enriched for targets of microRNAs of interest, we used the functional class scoring method which tests the null hypothesis that the list of differentially expressed genes from each microRNA target set is a random selection from the entire project differentially expressed gene list, implemented in the NCI BRB Array Tools software, as previously described [42].

Additional details of the statistical methodology and bioinformatics algorithms described above are found in Data S1.

Supporting Information

Data S1 Revised Supplementary Data File.
Found at: doi:10.1371/journal.pone.0009747.s001 (0.05 MB DOC)

Figure S1 MFH-samples predicted as LIPO (MFH-LIPOs) cluster together with both myxoid and non-myxoid liposarcomas.
Found at: doi:10.1371/journal.pone.0009747.s002 (0.27 MB TIF)

Figure S2 Complete Clustering Results.
Found at: doi:10.1371/journal.pone.0009747.s003 (0.43 MB TIF)

Figure S3 Fold upregulation of selected genes associated with smooth muscle, fibroblast and adipocyte differentiation in MFH tumors predicted as leio-, lipo or fibrosarcoma respectively (compared to the rest of MFH tumors).
Found at: doi:10.1371/journal.pone.0009747.s004 (0.19 MB TIF)

Table S1 Sensitivity and specificity of the 170-gene Nearest Centroid predictor for each STS class in the NCI, UK and JAPAN datasets (actual numbers of individual tumor subtypes are shown in parentheses).
Found at: doi:10.1371/journal.pone.0009747.s005 (0.02 MB XLS)

Table S2 Sensitivity and specificity of the 170-gene Nearest Centroid predictor for each STS class in the STANFORD and MSKCC validation datasets.
Found at: doi:10.1371/journal.pone.0009747.s006 (0.02 MB XLS)

Table S3 Genes included in 1st step predictor.
Found at: doi:10.1371/journal.pone.0009747.s007 (0.04 MB XLS)

Table S4 Genes included in 2nd step predictor.
Found at: doi:10.1371/journal.pone.0009747.s008 (0.02 MB XLS)

Table S5 Assessment of MFH reclassification across different datasets with the Sub Class Mapping methodology.
Found at: doi:10.1371/journal.pone.0009747.s009 (0.02 MB XLS)

Table S6 Gene expression pattern of the Ras activated samples was enriched for targets of microRNAs of the let-7 family.
Found at: doi:10.1371/journal.pone.0009747.s010 (0.02 MB XLS)

Table S7 Detailed clustering results of the classified MFH samples.
Found at: doi:10.1371/journal.pone.0009747.s011 (0.02 MB XLS)

Table S8 Patterns of oncogenic pathway activation in different STS histologies.
Found at: doi:10.1371/journal.pone.0009747.s012 (0.02 MB XLS)

Acknowledgments

We wish to thank Hal Schneider, PhD, Laboratory Supervisor, and the personnel of the Children's Hospital Molecular Genetics Core facility, for service and support provided in running the DASL microarray experiments.

Author Contributions

Conceived and designed the experiments: PAK JG TAL MCG DS. Performed the experiments: EF KP. Analyzed the data: PAK EF JG MB KP NF TAL MCG DS. Contributed reagents/materials/analysis tools: PAK EF MB NF TAL MCG DS. Wrote the paper: PAK DS.

References

1. Fletcher CD (1997) Soft tissue tumours: the impact of cytogenetics and molecular genetics. Verh Dtsch Ges Pathol 81: 318–326.
2. Nascimento AF, Raut CP (2008) Diagnosis and management of pleomorphic sarcomas (so-called "MFH") in adults. J Surg Oncol 97: 330–339.
3. Goldberg BR (2007) Soft tissue sarcoma: An overview. Orthop Nurs 26: 4–11; quiz 12–13.
4. Blanke CD, Rankin C, Demetri GD, Ryan CW, von Mehren M, et al. (2008) Phase III randomized, intergroup trial assessing imatinib mesylate at two dose levels in patients with unresectable or metastatic gastrointestinal stromal tumors expressing the kit receptor tyrosine kinase: S0033. J Clin Oncol 26: 626–632.
5. Demetri GD, von Mehren M, Blanke CD, Van den Abbeele AD, Eisenberg B, et al. (2002) Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 347: 472–480.
6. de Alava E (2007) Molecular pathology in sarcomas. Clin Transl Oncol 9: 130–144.
7. Lawrence W, Jr., Donegan WL, Natarajan N, Mettlin C, Beart R, et al. (1987) Adult soft tissue sarcomas. A pattern of care survey of the American College of Surgeons. Ann Surg 205: 349–359.
8. Daugaard S (2004) Current soft-tissue sarcoma classifications. Eur J Cancer 40: 543–548.
9. Brenton JD, Aparicio SA, Caldas C (2001) Molecular profiling of breast cancer: portraits but not physiognomy. Breast Cancer Res 3: 77–80.
10. Hegde U, Wilson WH (2001) Gene expression profiling of lymphomas. Curr Oncol Rep 3: 243–249.
11. Liotta L, Petricoin E (2000) Molecular profiling of human cancer. Nat Rev Genet 1: 48–56.
12. Baird K, Davis S, Antonescu CR, Harper UL, Walker RL, et al. (2005) Gene expression profiling of human sarcomas: insights into sarcoma biology. Cancer Res 65: 9226–9235.
13. Henderson SR, Guiliano D, Presneau N, McLean S, Frow R, et al. (2005) A molecular map of mesenchymal tumors. Genome Biol 6:-R76.
14. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP (2007) Subclass mapping: identifying common subtypes in independent disease data sets. PLoS ONE 2: e1195.
15. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, et al. (2005) RAS is regulated by the let-7 microRNA family. Cell 120: 635–647.
16. Dagher R, Cohen M, Williams G, Rothmann M, Gobburu J, et al. (2002) Approval summary: imatinib mesylate in the treatment of metastatic and/or unresectable malignant gastrointestinal stromal tumors. Clin Cancer Res 8: 3034–3038.
17. Fletcher CD (1992) Pleomorphic malignant fibrous histiocytoma: fact or fiction? A critical reappraisal based on 159 tumors diagnosed as pleomorphic sarcoma. Am J Surg Pathol 16: 213–228.
18. Nakayama R, Nemoto T, Takahashi H, Ohta T, Kawai A, et al. (2007) Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol 20: 749–759.
19. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, et al. (2002) Molecular characterisation of soft tissue tumours: a gene expression study. Lancet 359: 1301–1307.
20. Simon R (2005) Roadmap for developing and validating therapeutically relevant genomic classifiers. J Clin Oncol 23: 7332–7341.
21. Quackenbush J (2006) Microarray analysis and tumor classification. N Engl J Med 354: 2463–2472.
22. Detwiller KY, Fernando NT, Segal NH, Ryeom SW, D'Amore PA, et al. (2005) Analysis of hypoxia-related gene expression in sarcomas and effect of hypoxia on RNA interference of vascular endothelial cell growth factor A. Cancer Res 65: 5881–5889.
23. Yuen JS, Macaulay VM (2008) Targeting the type 1 insulin-like growth factor receptor as a treatment for cancer. Expert Opin Ther Targets 12: 589–603.
24. Blay JY, Ray-Coquard I, Alberti L, Ranchere D (2004) Targeting other abnormal signaling pathways in sarcoma: EGFR in synovial sarcomas, PPAR-gamma in liposarcomas. Cancer Treat Res 120: 151–167.
25. Adriaenssens E, Vanhecke E, Saule P, Mougel A, Page A, et al. (2008) Nerve growth factor is a potential therapeutic target in breast cancer. Cancer Res 68: 346–351.
26. Martinez-Torrecuadrada J, Cifuentes G, Lopez-Serra P, Saenz P, Martinez A, et al. (2005) Targeting the extracellular domain of fibroblast growth factor receptor 3 with human single-chain Fv antibodies inhibits bladder carcinoma cell line proliferation. Clin Cancer Res 11: 6280–6290.
27. Fletcher CD, Gustafson P, Rydholm A, Willen H, Akerman M, (2001) Clinicopathologic re-evaluation of 100 malignant fibrous histiocytomas: prognostic relevance of subclassification. J Clin Oncol 19: 3045–3050.

28. Ambros V (2004) The functions of animal microRNAs. Nature 431: 350–355.
29. Blower PE, Chung JH, Verducci JS, Lin S, Park JK, et al. (2008) MicroRNAs modulate the chemosensitivity of tumor cells. Mol Cancer Ther 7: 1–9.
30. Calin GA, Croce CM (2006) MicroRNA-cancer connection: the beginning of a new tale. Cancer Res 66: 7390–7394.
31. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, et al. (2003) Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31: e15.
32. Bibikova M, Talantov D, Chudin E, Yeakley JM, Chen J, et al. (2004) Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. Am J Pathol 165: 1799–1807.
33. Bibikova M, Yeakley JM, Wang-Rodriguez J, Fan JB (2008) Quantitative expression profiling of RNA from formalin-fixed, paraffin-embedded tissues using randomly assembled bead arrays. Methods Mol Biol 439: 159–177.
34. Dabney AR (2005) Classification of microarrays to nearest centroids. Bioinformatics 21: 4148–4154.
35. Dabney AR, Storey JD (2007) Optimality driven nearest centroid classification from genomic data. PLoS ONE 2: e1002.
36. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8: 118–127.
37. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–14868.
38. Bild AH, Yao G, Chang JT, Wang Q, Potti A, et al. (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 439: 353–357.
39. Potti A, Dressman HK, Bild A, Riedel RF, Chan G, et al. (2006) Genomic signatures to guide the use of chemotherapeutics. Nat Med 12: 1294–1300.
40. Acharya CR, Hsu DS, Anders CK, Anguiano A, Salter KH, et al. (2008) Gene expression signatures, clinicopathological features, and individualized therapy in breast cancer. JAMA 299: 1574–1587.
41. Dressman HK, Berchuck A, Chan G, Zhai J, Bild A, et al. (2007) An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. J Clin Oncol 25: 517–525.
42. Pavlidis P, Qin J, Arango V, Mann JJ, Sibille E (2004) Using the gene ontology for microarray data mining: a comparison of methods and application to age effects in human prefrontal cortex. Neurochem Res 29: 1213–1222.

Cross Species Genomic Analysis Identifies a Mouse Model as Undifferentiated Pleomorphic Sarcoma/ Malignant Fibrous Histiocytoma

Jeffrey K. Mito[1], Richard F. Riedel[2], Leslie Dodd[3], Guy Lahat[4], Alexander J. Lazar[5], Rebecca D. Dodd[6], Lars Stangenberg[7], William C. Eward[8], Francis J. Hornicek[7], Sam S. Yoon[7], Brian E. Brigman[8], Tyler Jacks[9,10], Dina Lev[3], Sayan Mukherjee[11,12], David G. Kirsch[1,6]*

1 Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina, United States of America, 2 Division of Medical Oncology, Department of Medicine, Duke University Medical Center, Durham, North Carolina, United States of America, 3 Department of Pathology, Duke University Medical Center, Durham, North Carolina, United States of America, 4 Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, 5 Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, 6 Department of Radiation Oncology, Duke University Medical Center, Durham, North Carolina, United States of America, 7 Department of Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States of America, 8 Division of Orthopaedic Surgery, Department of Surgery, Duke University Medical Center, Durham, North Carolina, United States of America, 9 Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 10 Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America, 11 Department of Statistical Sciences, Duke University, Durham, North Carolina, United States of America, 12 Institute for Genome Sciences and Policy, Duke University Medical Center, Durham, North Carolina, United States of America

Abstract

Undifferentiated pleomorphic sarcoma/Malignant Fibrous Histiocytoma (MFH) is one of the most common subtypes of human soft tissue sarcoma. Using cross species genomic analysis, we define a geneset from the *LSL-Kras^{G12D}; Trp53$^{Flox/Flox}$* mouse model of soft tissue sarcoma that is highly enriched in human MFH. With this mouse geneset as a filter, we identify expression of the RAS target FOXM1 in human MFH. Expression of *Foxm1* is elevated in mouse sarcomas that metastasize to the lung and tissue microarray analysis of human MFH correlates overexpression of FOXM1 with metastasis. These results suggest that genomic alterations present in human MFH are conserved in the *LSL-Kras^{G12D}; p53$^{Flox/Flox}$* mouse model of soft tissue sarcoma and demonstrate the utility of this pre-clinical model.

Citation: Mito JK, Riedel RF, Dodd L, Lahat G, Lazar AJ, et al. (2009) Cross Species Genomic Analysis Identifies a Mouse Model as Undifferentiated Pleomorphic Sarcoma/Malignant Fibrous Histiocytoma. PLoS ONE 4(11): e8075. doi:10.1371/journal.pone.0008075

Editor: Syed A. Aziz, Health Canada, Canada

Received October 9, 2009; **Accepted** November 3, 2009; **Published** November 30, 2009

Funding: This study was supported by the Howard Hughes Medical Institute (TJ) and by KO8 CA 114176 and RO1 CA 138265 (DGK), T32 GM-07171 (JM), and the Maria Garcia-Estrada Foundation (RR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: david.kirsch@duke.edu

Introduction

Malignant Fibrous Histiocytoma was first described in the 1960s and quickly became the most commonly diagnosed adult soft tissue sarcoma [1]. Because these tumors do not appear to arise from histiocytes, the term Malignant Fibrous Histiocytoma has recently fallen out of favor and many pathologists now classify these tumors as undifferentiated pleomorphic sarcomas. Despite this change in nomenclature, undifferentiated pleomorphic sarcoma (referred to here as MFH) remains one of the most common adult soft tissue sarcomas encountered in the clinic. However, the cell(s) of origin of MFH is unknown. Indeed, some have suggested that MFH is a collection of undifferentiated mesenchymal tumors sharing a common morphology rather than a single clinical entity [2,3,4]. This debate could be clarified by identifying the cell(s) of origin of a mouse model for MFH. Whether MFH describes a cancer that is a single pathogenic entity or an undifferentiated state shared by several sarcoma subtypes, the survival of patients with MFH has not improved for decades. Therefore, identifying a mouse model of MFH may also lead to better treatments for patients with this diagnosis.

Human MFH is characterized by a propensity to metastasize to the lungs and by a range of histologic appearances including spindle and pleomorphic cells. Although these features are recapitulated in a mouse model of soft tissue sarcoma initiated by conditional mutations in *Kras* and *Trp53* [5], it is not clear whether this model is most similar to human MFH or another soft tissue sarcoma. Therefore, we sought to more accurately classify the sarcoma subtype for this mouse model using gene expression profiling.

Methods

Mouse Genotyping, Tumor Generation, and Determination of Metastatic Potential

Both mouse genotyping and generation of tumors was carried out as described previously [5] in accordance with Duke University and MIT Institutional Animal Care and Use Committee approved protocols. Sarcomas were induced in the lower left limb and allowed to grow until ~200 mm^3 in volume. Tumors were then surgically excised via amputation of the limb and animals followed for a minimum of 4 months to determine the metastatic potential of the primary tumor.

RNA Isolation

RNA was extracted from *LSL-Kras^{G12D}; Trp53$^{Flox/Flox}$* tumors or normal muscle using TRIzol reagent (Invitrogen) and was purified using RNeasy mini kit (Qiagen).

Microarray Processing and Analysis

Full details can be found in supplementary methods S1. Briefly, gene expression was determined using Affymetrix 430A 2.0 arrays (Affymetrix) as described in detail online (http://www.genome.duke.edu/cores/microarray/). CEL files were processed using the RMA algorithm [6,7] to normalize the data. Genesets were identified using a signal-to-noise metric.

Human and Mouse datasets [8,9,10] were downloaded from GEO (GSE6461, GSE6481, GSE2553, and GDS1209), normalized with RMA (when appropriate). Genesets and array data were used in GSEA as described previously [11]. Classes were defined as one soft tissue sarcoma subtype versus controls (other sarcoma or normal muscle) present in their respective datasets.

Database Accession Numbers

Microarray data was generated in conformity to MIAME guidelines and has been deposited in the GEO database under accession number GSE16779.

Oncogenic Pathway Predictors

Human soft tissue sarcoma datasets [9,10] were combined using ComBat [12] and normal tissue samples removed from the combined dataset. An oncogenic pathway classifier for Ras pathway activity was developed as described previously [13]. This classifier was used to compare undifferentiated pleiomorphic sarcoma/MFH samples (n = 29) against all other soft tissue sarcomas. Significance was determined using a non-parametric Mann-Whitney test.

Histology and Immunohistochemistry and Image Analysis

All human samples were obtained from tissue repositories at Duke and MD Anderson. These samples were used in accordance with Duke and MD Anderson Cancer Center Institutional Review Board (IRB) approved protocols under a waiver of consent. Five micron thick sections were cut from formalin fixed paraffin embedded samples. Samples were subjected to standard hematoyxlin and eosin staining or immunohistochemistry. Immunohistochemistry was performed with the following antibodies: phospho-ERK (Invitrogen 29-2389) and FOXM1 (Abcam ab47808), using the Vectastain ABC Rabbit IgG kit with Vectastain Elite ABC Reagent (Vector Labs).

Brightfield images of slides taken at 40x were used for analysis using Image Pro AMS v6.1. The counting module was trained using both positive and negative nuclear staining for phospho-ERK. A minimum of 3000 nuclei were counted per sample and a ratio between total nuclei with positive nuclei to total nuclei was determined using a minimal and maximal area of 100 and 1000 pixels respectively.

Tissue Microarrays (TMAs)

TMAs were generated at MD Anderson Cancer Center and contained a clinically annotated set of 214 soft tissue sarcoma samples including: 166 MFH/Unclassified sarcomas, 19 synovial sarcomas, 6 leiomyosarcomas, 8 pleomorphic liposarcomas, 8 myxoid liposarcomas, 6 atypical lipomatous tumors, and 1 dedifferentiated liposarcoma. TMAs were stained as above and scored semiquantitatively on a scale from 0-3+ by a musculoskeletal pathologist (L.D.) blinded to patient outcome. Scores were correlated with both diagnosis and clinical outcome.

Statistical Analysis of TMAs

Scoring of TMAs was correlated with diagnosis, and metastasis-free survival.

Correlation of immunohistochemical staining with diagnosis was tested for normality using a chi-square test. This did not reach statistical significance, therefore comparison between MFH and other soft tissue sarcomas was performed using the non-parametric Mann-Whitney test.

Metastasis free survival analysis was performed on MFH samples comparing 3+ staining to 0-2+ staining. Survival was determined by Kaplan-Meier analysis.

Results

Cross-Species Genomic Analysis of *LSL-KRas^{G12D}; Trp53$^{Flox/Flox}$* Mouse Model of Soft Tissue Sarcoma

We hypothesized that genes most differentially expressed between mouse sarcoma and normal muscle would provide a useful molecular signature to interrogate human sarcoma datasets (Fig. 1A). To test this approach, we first analyzed published gene expression data from a previously validated mouse model of synovial sarcoma [14]. After identifying a geneset of 100 genes highly overexpressed in synovial sarcoma compared to normal muscle (Table S1) we used Gene Set Enrichment Analysis (GSEA) [11] to probe gene expression data from three studies of human soft tissue sarcomas [8,9,10]. We demonstrated strong statistical enrichment ($p<0.001$, FDR<0.02) of this geneset in human synovial sarcomas, but not in other subtypes of soft tissue sarcoma (Fig. S1, Table S2). This result is in agreement with GSEA for the mouse model of synovial sarcoma that was previously reported [14].

Having validated this approach, we identified a geneset of 100 genes highly overexpressed in the *LSL-Kras^{G12D}; Trp53$^{Flox/Flox}$* mouse model of soft tissue sarcoma (n = 17) compared to normal muscle (n = 4) (Table S3). When this geneset was analyzed in the three human datasets of soft tissue sarcoma [8,9,10], only MFH samples showed statistical enrichment ($p = 0.001$; FDR = 0.012) (Table 1, Fig. 1B). Enrichment was seen in all three datasets [8,9,10], which represent 325 sarcoma samples and two types of array platforms. Moreover, genesets derived from human MFH (Table S4) also enriched in the mouse sarcoma data ($p<0.001$; FDR<0.001) (Fig. 1C). These data indicate that this mouse sarcoma model and human MFH share common genomic features.

Ras Pathway Activity Is Enriched in Human MFH

The initiating events of human MFH are not well understood. Previous studies have shown mutations in p53 occur in 36% of human MFH [15] while the rate of canonical *RAS* mutations in human MFH varies from 0–50% [16,17]. We hypothesized that the RAS pathway may be activated in human MFH even in the absence of canonical RAS mutations. To explore a link between MFH and Ras, we utilized previously described oncogenic pathway predictors that correlate with *RAS* activity [13]. The Ras oncogenic signature is enriched in human MFH samples compared to a panel of other soft tissue sarcomas ($p = 0.002$) (Fig. 2A). Moreover, in human MFH (n = 8) lacking canonical RAS mutations, we observed nuclear staining of phospho-ERK by immunohistochemistry in greater than 30% of tumor cells in 7 of 8 samples (Fig. S2).

FoxM1 Is a Novel Marker of Metastasis in MFH

Because human MFH and the *LSL-Kras^{G12D}; Trp53$^{Flox/Flox}$* mouse model of soft tissue sarcoma share common genomic features, we wanted to determine if these shared features could be used to identify diagnostic or prognostic factors for human MFH. As MFH is considered a diagnosis of exclusion, we initially attempted to identify a marker that is specific to MFH. We identified a panel of 10 candidate

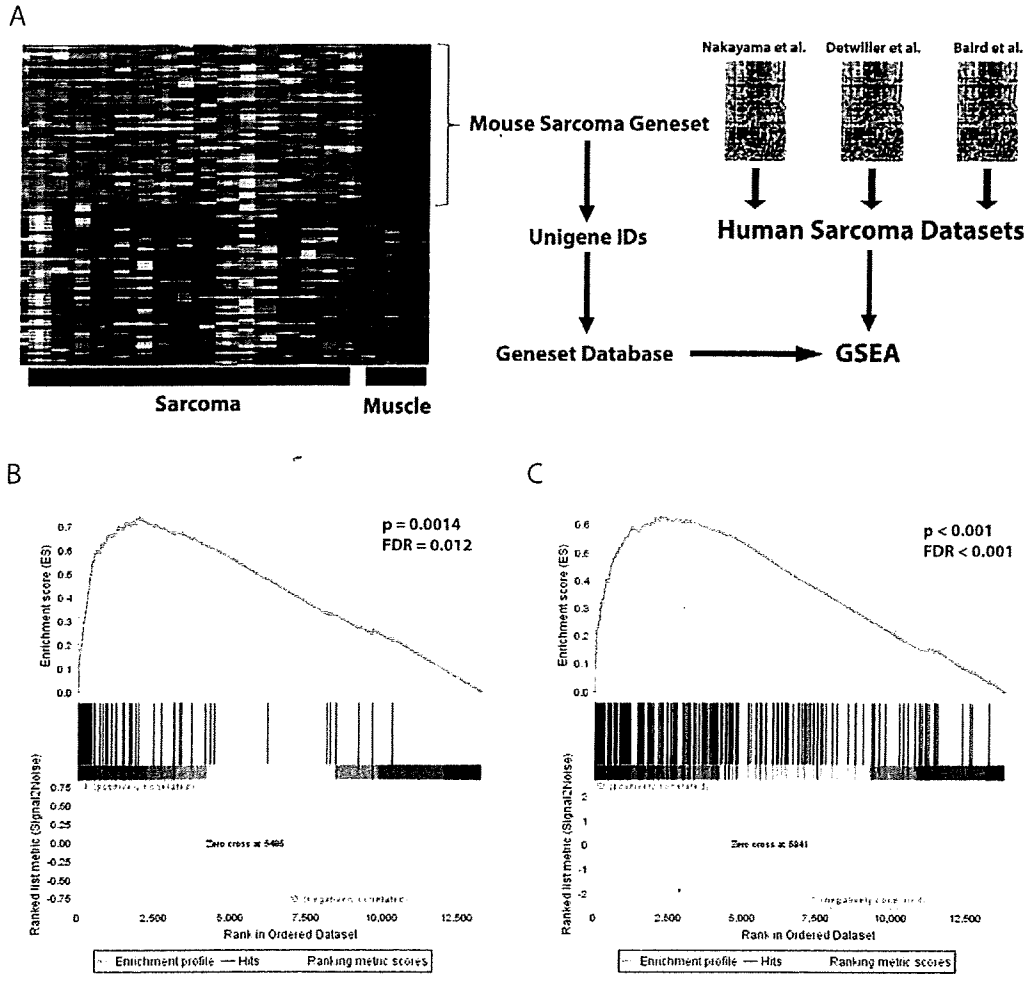


Figure 1. Cross-species genomic comparison. *A*, Schematic for cross species genomic comparison using GSEA. *B*, The *LSL-Kras*^{G12D}*; Trp53*^{FloxFlox} sarcoma geneset (Table S3) is highly enriched in human MFH in the Nakayama et al. soft tissue sarcoma dataset [10] [p = 0.0014; False Discovery Rate (FDR) = 0.012; Enrichment Score (ES) = 0.74; Normalized Enrichment Score (NES) = 2.09]. *C*, Conversely, a human MFH geneset (Table S4) is strongly enriched in the mouse model of soft tissue sarcoma (p<0.001, FDR<0.001, ES = 0.637 NES = 2.78).
doi:10.1371/journal.pone.0008075.g001

Table 1. GSEA results for the *LSL-Kras*^{G12D}*; Trp53*$^{Flox/Flox}$ geneset derived from the mouse model of soft tissue sarcoma (Table S3).

	Nakayama [10]	Detwiller [9]	Baird [8]
Malignant Fibrous Histiocytoma	**0.001 (0.012)**	**0.014 (0.149)**	**0.024 (0.190)**
Myxofibrosarcoma	0.162 (0.559)	-	-
Fibrosarcoma	0.177 (0.326)	0.159 (0.784)	DNE
Leiomyosarcoma	0.262 (0.673)	0.854 (0.981)	0.401 (0.715)
Synovial Sarcoma	DNE	DNE	DNE
Myxoid Liposarcoma	DNE	-	-
Dedifferentiated Liposarcoma	DNE	-	-
Rhabdomyosarcoma	-	-	0.586 (0.781)
Ewing's Sarcoma	-	-	DNE

The mouse sarcoma geneset was used to examine three human soft tissue sarcoma datasets [8,9,10]. Only undifferentiated pleomorphic sarcoma/MFH demonstrated statistically significant enrichment. Table denotes p-values with FDR in parentheses. Bolded results note significance with p<0.05; FDR<0.25. DNE = Did Not Enrich, dash marks represent insufficient data points to do comparison.
doi:10.1371/journal.pone.0008075.t001

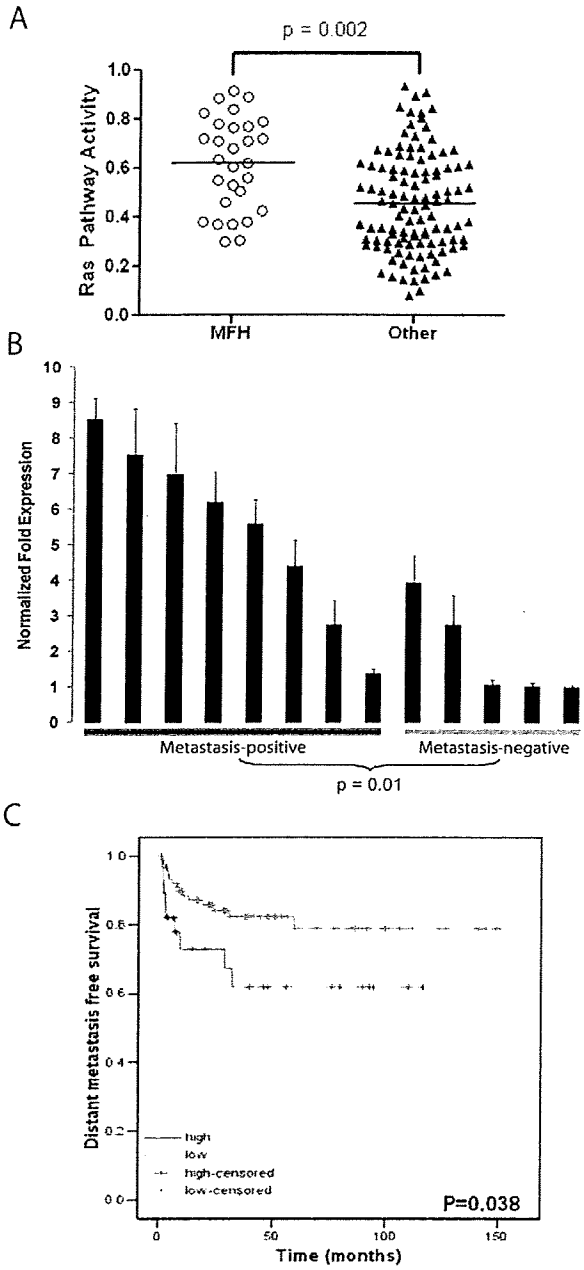


Figure 2. The *LSL-Kras*^{G12D}; *Trp53*$^{Flox/Flox}$ mouse model of soft tissue sarcoma provides insight into human MFH. *A*, An oncogenic Ras signature is enriched in human MFH samples compared to other types of soft tissue sarcoma ($p = 0.002$, non-parametric Mann-Whitney test). *B*, Q-RT-PCR for *Foxm1* in murine soft tissue sarcomas correlates with metastatic potential of primary tumors ($p = 0.01$, two-tailed student's T-test, scale bars represent one standard deviation). *C*, FOXM1 expression in a tissue microarray correlates with metastasis free survival in human MFH ($p = 0.038$).
doi:10.1371/journal.pone.0008075.g002

biomarkers based on their common upregulation in both human MFH and the mouse sarcoma model (Table S5). Expression of 9 of these candidates was validated in an independent cohort of mouse sarcomas by Q-RT-PCR (Fig. S3).

FOXM1, which is a member of the forkhead transcription factor family that enhances tumorigenesis in other solid tumors [18], was selected for further analysis because it is downstream of *Ras* [19]. Immunohistochemistry for FOXM1 in a panel of 8 human MFH samples demonstrated nuclear staining for FOXM1 in all 8 tumors (Fig. S2). We next analyzed the expression of FOXM1 in a clinically annotated tissue microarray (TMA) containing 166 MFH samples and 48 other soft tissue sarcomas. Although 84% of the MFH samples stained positive for FOXM1 ($p = 0.02$, Fig. S4), this marker was also expressed, but to a lesser degree, in other sarcoma subtypes. Therefore, FOXM1 may not be a useful diagnostic marker for human MFH.

Because FoxM1 has previously been shown to regulate the expression of matrix metalloproteases MMP-2 and -9, which are key mediators of cell invasion [20], we hypothesized that high FOXM1 expression may correlate with metastasis. We measured *Foxm1* gene expression in murine soft tissue sarcomas and found a correlation with the development of lung metastases (Fig. 2B, $p = 0.01$). Likewise, human MFH with high FOXM1 expression correlated with decreased metastasis-free survival compared to sarcomas with low to no FOXM1 expression (Fig. 2C). In contrast, overexpression of another candidate marker MELK (Table S5) did not correlate with metastasis-free survival (Fig. S5).

Discussion

We used cross species genomic analysis to determine which human sarcoma subtype is represented by the *LSL-Kras*^{G12D}; *Trp53*$^{Flox/Flox}$ mouse model of soft tissue sarcoma. We identified a geneset in the mouse sarcomas that is highly enriched in human MFH. This geneset was not enriched in other human sarcomas, such as fibrosarcoma or leiomyosarcoma, which can be difficult to distinguish from MFH. Furthermore, we have identified enrichment of Ras pathway activity in human MFH compared to other types of soft-tissue sarcoma. Additionally, the *LSL-Kras*^{G12D}; *Trp53*$^{Flox/Flox}$ mouse model of soft tissue sarcoma has a propensity to metastasize to the lungs much like human MFH [5]. Based on this genomic analysis, the pattern of lung metastasis, and the similarity of the mouse sarcomas to human MFH at the histological level (Fig. S6), we conclude that this model closely resembles MFH.

We recognize that the diagnosis of undifferentiated pleomorphic sarcoma (MFH) has recently been questioned as a distinct clinical entity [2,3,4]. Our results do not exclude the possibility that MFH is a collection of mesenchymal tumors derived from different cell types that share an undifferentiated state. However, our finding of a sarcoma geneset conserved between mouse sarcomas and human MFH suggests that this subtype of human sarcoma shares an underlying biology beyond a common histologic appearance. Moreover, the use of cross-species analysis to identify FOXM1 as a marker of metastasis-free survival in human MFH supports the use of this mouse model to understand mechanisms of metastasis, to investigate the cell(s) of origin, and to develop novel therapies for human MFH.

Supporting Information

Supplementary Methods S1 Supplementary Methods
Found at: doi:10.1371/journal.pone.0008075.s001 (0.07 MB DOC)

Table S1 Geneset derived from mouse model of synovial sarcoma versus control (normal muscle). Geneset was derived using signal-to-noise metric.
Found at: doi:10.1371/journal.pone.0008075.s002 (0.04 MB DOC)