Artificial Neural Network Inference (ANNI): A Study on Gene-Gene Interaction for Biomarkers in Childhood Sarcomas

Dong Ling Tong[1], David J. Boocock[1], Gopal Krishna R. Dhondalay[2], Christophe Lemetre[3], Graham R. Ball[1]*

1 The John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, United Kingdom, 2 Imperial Centre for Translational and Experimental Medicine, National Heart and Lung Institute (NHLI), Imperial College London, London, United Kingdom, 3 Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America

Abstract

Objective: To model the potential interaction between previously identified biomarkers in children sarcomas using artificial neural network inference (ANNI).

Method: To concisely demonstrate the biological interactions between correlated genes in an interaction network map, only 2 types of sarcomas in the children small round blue cell tumors (SRBCTs) dataset are discussed in this paper. A backpropagation neural network was used to model the potential interaction between genes. The prediction weights and signal directions were used to model the strengths of the interaction signals and the direction of the interaction link between genes. The ANN model was validated using Monte Carlo cross-validation to minimize the risk of over-fitting and to optimize generalization ability of the model.

Results: Strong connection links on certain genes (TNNT1 and FNDC5 in rhabdomyosarcoma (RMS); FCGRT and OLFM1 in Ewing's sarcoma (EWS)) suggested their potency as central hubs in the interconnection of genes with different functionalities. The results showed that the RMS patients in this dataset are likely to be congenital and at low risk of cardiomyopathy development. The EWS patients are likely to be complicated by EWS-FLI fusion and deficiency in various signaling pathways, including Wnt, Fas/Rho and intracellular oxygen.

Conclusions: The ANN network inference approach and the examination of identified genes in the published literature within the context of the disease highlights the substantial influence of certain genes in sarcomas.

Introduction

Although computer technologies have evolved over the past decades, debate regarding on the suitability of data mining techniques to identify "true" biomarkers continues due to the fact that these techniques are fundamentally based on mathematical paradigms. Furthermore, the connection between the statistical and the biological significance of the findings are not well described and validated. Questions regarding the advantage of these techniques and the relevance of the selected biomarkers to biological processes might explain why very few biomarkers that have been discovered using these approaches have seen clinical applications.

Additionally, many of the published studies assumed that biomarker discovery involves merely marker selection and classification. Markers with high statistical power and able to accurately predict the disease group are treated as "best" markers

for clinical use, even though their biological interactivities are not tested *in silico*. This might explained why clinical trials on these markers have failed. We believe that the identification and validation of biomarkers *in silico* are equally important in biomarker discovery and are vital for clinical trial development. An *in silico* simulation of possible biological interaction between the selected candidate markers provides information on the nature of the markers (i.e. proactive or inactive), state of the markers (i.e. on or off) and possible chemical changes on the markers. These information can subsequently improve the success rate in clinical trials and patient care. In short, the biology of phenotype is more than just a list of markers; it is the complex interaction of biological components that defines phenotype.

We previously identified a list of high potential marker candidates that are able to differentiate small round blue cell tumors (SRBCTs) in children [1]. This study builds on previous work which has modeled the interaction between these markers to

reveal their potential biological relevance in child sarcoma cancers using a bespoke artificial neural network based interactive algorithm.

The sarcoma groups in the SRBCTs dataset reported by Khan *et al.* [2] were used in this study. The selection of biomarker panels for the SRBCTs dataset was performed using a hybrid genetic algorithm-neural network (GANN) model, as has been reported in our previous work [1]. The aim of this GANN approach was to identify sets of features that possess significant statistics information and statistical comparison between classification methods based upon gene sets reported by Khan *et al.* and the GANN model has been elaborated. This study focused on modeling interactions between these features in order to infer their potential biological significance in sarcoma cancers.

It is difficult to objectively compare the biological relevance of individual genes identified between studies, as complex biology cannot be quantified using numerical analyses. Published literature [3,4] has argued that the biological relevance between features can be achieved using correlation analysis and a pre-defined baseline value of the parameter which is known to be biologically meaningful. However, this information alone is not sufficient for comparing which feature has more biological meaning than another, as the full biological function of the features and the biochemical reaction mechanisms underlying regulatory interactions between features cannot be fully known without conducting a thorough assessment in clinical materials relevant to the disease status in order to describe behaviors of these features *in vivo*. This clinical assessment and validation is not within the remit of this paper and instead of looking for more biological meaningful features, this paper reports a complementary set of genes to those reported by Khan *et al* [2].

For the sake of conciseness of the interpretation of the biological relevance on the selected genes in the dataset, the biological functionality of the genes associating with 2 types of sarcomas, and interactions that are of potential relevance on the basis of plausible biological explanations and the correlation analysis of the genes were studied in this paper.

Materials and Methods

In this section, we describe the SRBCTs dataset and the selected biomarkers by the GANN model. We then describe the framework of the interactome network analysis constructed using ANNs.

Small round blue cell tumors (SRBCTs) dataset

The SRBCTs cDNA microarray dataset was originally studied by Khan *et al.* [2], with the scope of identifying marker genes that are able to distinguish 4 different types of blue cell tumors which often masquerade as one another in childhood. This was performed with ANN classification models in conjunction with Principle Component Analysis (PCA). The original dataset contained 88 samples with 2,308 genes, distributed into 4 different tumor types, *i.e.* rhabdomyosarcoma (RMS), Ewing's sarcoma (EWS), Burkitt lymphoma (BL) and neuroblastoma (NB). Amongst the 88 samples, 25 were in the RMS group, 29 in EWS, 11 in BL, 18 in NB and the remaining 5 samples were unknown.

Feature selection using genetic algorithm-neural network (GANN)

The GANN is a hybrid model which we have previously applied to the screening of datasets for genes of statistical relevance [1,5–7]. In the GANN model, genetic algorithm (GA) and ANN co-evolve in the learning process. In brief, a population of chromosomes each representing a subset of microarray genes was first generated and the fitness for each chromosome was computed as a solution to the problem using multilayered ANNs. These fitness values were then iteratively evaluated using GA's operators and ANNs, and a rank order of the genes based on their fitness values was produced. The evaluation process was iterated 3,000 times and the whole process cycle was repeated 5,000 times. The complete parameter settings on the GANN model can be found in our work [1].

The previously reported panel of 96 genes [1] is summarized in Table 1. Among these genes, 44 were complementary to the top 96 genes reported by Khan *et al.* [2].

Interactome Network Map

The concept of the interactome network map in which the internal organization and functional regulation of cells can be presented using network/graph theory was initially set out by Barabási and Oltvai [8]. In the network map, a single gene is symbolized by a *node,* and the link between genes is known as an *edge,* which can be presented with an arrow to indicate the direction of the link from a *source* node to a *target* node.

Interactome network maps have been used to demonstrate interactions between biological components. These originally utilized off-the-shelf or publicly available modeling tools to analyze associations between biological molecules [9–13], but later used customized data mining tools to comprehensively model the interaction between biological molecules. These customized regulatory networks were initially Boolean-based [14,15], then evolved to Bayesian probability [16–18], dynamic ordinary differential equations (ODEs) [19] and in recent years, ANNs [20–25]. These and other approaches have been reviewed elsewhere [26–30]. An ANN network inference approach was chosen to model the interactions between sarcoma-related microarray genes for the SRBCTs dataset in this study. For reference, we refer to this approach as Artificial Neural Network Inference (ANNI).

Artificial neural network inference (ANNI)

ANNs have been extensively used for biomarker identification and classification [31–37] due to their ability to cope with complexity and nonlinearity within the biology datasets. These features enable ANNs to address a particular question by identifying and modeling patterns in the data [38,39]. The underlying structure of the multilayer perceptron (MLP) is a weighted, directed graph [24], interconnecting artificial neurons (i.e. nodes) organized in layers with artificial synapses (i.e. links) which carry a value, (i.e. weight) transmitting data (i.e. signals) from one node to the other nodes. All incoming signals from the input layer will be processed based upon a set of defined parameters (i.e. error computation function, acceleration measure, input weights) by the nodes in the intermediate layer (i.e. hidden layer) and an activation function is applied to the resulting sum. This sum is then used to determine the output result (i.e. predicted value) generated by the nodes in the output layer. Due to the connectionist computation in ANNs, the architecture of the ANN can be easily modified to address different questions and able to compose complex hypotheses that can explain a high degree of correlation between features without any prior information from the datasets. Hence, a backpropagation MLP was chosen as ANN to model the gene-gene interaction in this paper.

This study hypothesized that the expression (i.e. up- or down-regulation) of a biomarker can be explained using the remaining biomarkers in the gene pool, if these biomarkers are able to explain one particular categorical outcome (i.e. a disease status). Herein, we explored the influences of all biomarkers among

Table 1. Summary of the previously selected 96 genes by GANN.

GANN rank	Symbol	Image Id.	GeneID	Truncated p-value for each cancer			
				EWS	BL	NB	RMS
1	MLLT11	812105	10962	8.65×10^{-12}	4.12×10^{-16}	7.27×10^{-17}	1.33×10^{-10}
2	IGF2	207274	3481	1.03×10^{-07}	5.65×10^{-11}	7.74×10^{-06}	4.78×10^{-09}
3	FCGRT	770394	2217	1.03×10^{-13}	6.87×10^{-13}	5.69×10^{-10}	2.27×10^{-07}
4	CAV1	377461	857	5.21×10^{-11}	1.00×10^{-10}	1.02×10^{-08}	8.49×10^{-08}
5	FGFR4	784224	2264	5.20×10^{-08}	5.44×10^{-08}	2.90×10^{-08}	8.32×10^{-11}
6	CD99	1435862	4267	8.12×10^{-12}	6.96×10^{-13}	8.43×10^{-09}	1.04×10^{-05}
7	IGF2	296448	3481	1.33×10^{-07}	9.71×10^{-10}	5.37×10^{-05}	7.89×10^{-09}
8	MAP1B	629896	4131	1.30×10^{-05}	4.24×10^{-11}	1.90×10^{-07}	1.41×10^{-04}
9	KDSR	814260	2531	2.11×10^{-09}	2.49×10^{-08}	1.13×10^{-07}	8.03×10^{-08}
10	FNDC5	244618	252995	2.53×10^{-07}	4.58×10^{-09}	3.63×10^{-06}	2.13×10^{-08}
11	CDH2	325182	1000	9.87×10^{-07}	4.86×10^{-07}	2.95×10^{-08}	1.41×10^{-04}
12	OLFM1	52076	10439	3.29×10^{-08}	2.29×10^{-10}	1.26×10^{-03}	1.77×10^{-08}
13	RCVRN	383188	5957	1.00×10^{-07}	4.69×10^{-14}	8.68×10^{-08}	4.22×10^{-03}
14	PTPN13	866702	5783	2.74×10^{-08}	9.53×10^{-09}	1.26×10^{-06}	2.33×10^{-07}
15	HLA-DMA	183337	3108	3.76×10^{-04}	1.90×10^{-07}	1.15×10^{-06}	2.84×10^{-04}
16	MYL4	461425	4635	8.30×10^{-06}	1.03×10^{-06}	2.28×10^{-06}	1.25×10^{-06}
17	SGCA	796258	6442	1.83×10^{-07}	1.02×10^{-06}	5.97×10^{-07}	2.05×10^{-08}
18	GAP43	44563	2596	1.03×10^{-04}	1.71×10^{-05}	3.22×10^{-05}	9.14×10^{-05}
19	PSMB8*	624360	5696	2.76×10^{-03}	1.72×10^{-06}	2.00×10^{-05}	1.16×10^{-06}
20	PMS2L12*	878652	392713	2.25×10^{-03}	4.03×10^{-11}	3.34×10^{-07}	9.49×10^{-04}
21	EHD1*	745019	10938	1.13×10^{-05}	1.38×10^{-09}	2.90×10^{-04}	1.45×10^{-03}
22	TNNT2	298062	7139	3.20×10^{-05}	1.70×10^{-05}	1.97×10^{-05}	1.91×10^{-05}
23	CBX1*	786084	10951	3.86×10^{-08}	1.76×10^{-03}	1.86×10^{-09}	2.85×10^{-03}
24	RBM38	814526	55544	2.84×10^{-05}	2.50×10^{-07}	1.06×10^{-09}	2.41×10^{-03}
25	TNNT1	1409509	7138	4.90×10^{-05}	4.05×10^{-07}	5.91×10^{-06}	3.63×10^{-06}
26	CRMP1	878280	1400	1.12×10^{-04}	2.46×10^{-11}	4.80×10^{-07}	1.90×10^{-03}
27	HLA-DQA1	80109	3117	1.46×10^{-03}	1.80×10^{-05}	2.67×10^{-05}	2.64×10^{-04}
28	DPYSL4	395708	10570	7.62×10^{-04}	1.63×10^{-10}	3.05×10^{-06}	5.67×10^{-05}
29	PIM2	1469292	11040	2.74×10^{-03}	6.31×10^{-06}	9.72×10^{-06}	1.36×10^{-05}
30	CTNNA1	21652	1495	1.09×10^{-05}	5.03×10^{-11}	1.13×10^{-06}	8.77×10^{-04}
31	SELENBP1	80338	8991	2.12×10^{-07}	2.77×10^{-10}	1.46×10^{-08}	7.56×10^{-04}
32	ELF1	241412	1997	7.03×10^{-04}	4.75×10^{-05}	5.19×10^{-06}	9.91×10^{-04}
33	KIF3C	784257	3797	4.80×10^{-03}	8.36×10^{-11}	1.90×10^{-05}	3.69×10^{-05}
34	GYG2	43733	8909	9.70×10^{-07}	5.56×10^{-08}	8.77×10^{-06}	1.03×10^{-05}
35	LSP1*	143306	4046	1.22×10^{-05}	4.34×10^{-10}	4.77×10^{-04}	6.24×10^{-07}
36	MT1L	297392	4500	2.20×10^{-03}	2.50×10^{-06}	1.84×10^{-04}	7.29×10^{-04}
37	CHD3*	379708	1107	1.99×10^{-07}	4.37×10^{-09}	1.09×10^{-06}	4.37×10^{-04}
38	EST (CDK6)	295985	1021	3.08×10^{-10}	3.26×10^{-03}	2.66×10^{-06}	1.63×10^{-04}
39	TNFAIP6	357031	7130	7.93×10^{-07}	1.92×10^{-07}	3.07×10^{-06}	1.34×10^{-05}
40	WAS*	236282	7454	7.36×10^{-04}	7.88×10^{-08}	3.27×10^{-07}	1.10×10^{-03}
41	GAS1	365826	2619	1.72×10^{-04}	4.78×10^{-11}	4.69×10^{-09}	2.04×10^{-03}
42	HCLS1	767183	3059	1.42×10^{-03}	6.05×10^{-06}	3.06×10^{-06}	1.66×10^{-04}
43	MYO1B	377048	4430	2.41×10^{-05}	2.77×10^{-15}	5.34×10^{-07}	1.71×10^{-02}
44	ARPC1B*	626502	10095	1.37×10^{-03}	8.58×10^{-05}	1.49×10^{-04}	1.47×10^{-03}
45	HOXB7*	1434905	3217	4.75×10^{-07}	4.65×10^{-06}	1.80×10^{-11}	1.89×10^{-02}
46	PRKAR2B	609663	5577	3.77×10^{-04}	2.19×10^{-05}	4.23×10^{-03}	1.16×10^{-07}
47	G6PD*	768246	2539	1.67×10^{-05}	2.72×10^{-03}	7.13×10^{-06}	5.18×10^{-04}
48	GATA2	135688	2624	3.42×10^{-03}	5.37×10^{-08}	1.67×10^{-05}	1.01×10^{-05}
49	CSDA*	810057	8531	3.77×10^{-04}	2.10×10^{-02}	4.99×10^{-10}	7.21×10^{-04}

Table 1. Cont.

GANN rank	Symbol	Image Id.	GeneID	Truncated p-value for each cancer			
				EWS	BL	NB	RMS
43	MYO1B	308231	4430	3.59×10^{-05}	2.84×10^{-11}	5.94×10^{-06}	1.31×10^{-02}
51	EID1*	244637	23741	2.25×10^{-03}	4.54×10^{-05}	5.89×10^{-07}	7.23×10^{-04}
52	PSMB10*	68977	5699	1.24×10^{-03}	2.80×10^{-05}	3.74×10^{-04}	1.76×10^{-03}
53	FHL3*	796475	2275	2.56×10^{-05}	3.26×10^{-11}	4.42×10^{-07}	5.07×10^{-03}
54	ITPR3*	435953	3710	1.31×10^{-03}	8.08×10^{-05}	1.15×10^{-07}	1.22×10^{-06}
55	DPYSL2	841620	1808	2.99×10^{-05}	4.26×10^{-15}	2.83×10^{-04}	5.63×10^{-07}
56	BIN1	788107	274	2.93×10^{-05}	2.53×10^{-07}	3.05×10^{-03}	1.97×10^{-05}
57	PFN2	486110	5217	5.41×10^{-03}	1.02×10^{-15}	9.16×10^{-06}	2.33×10^{-02}
58	TLE2	1473131	7089	2.88×10^{-07}	5.70×10^{-09}	2.24×10^{-08}	1.06×10^{-03}
59	PGAM2*	283315	5224	3.25×10^{-03}	1.21×10^{-05}	2.51×10^{-02}	5.78×10^{-08}
60	ISG20*	740604	3669	9.18×10^{-04}	1.04×10^{-04}	1.26×10^{-04}	7.18×10^{-04}
61	RDX*	740554	5962	6.12×10^{-07}	4.86×10^{-11}	2.57×10^{-05}	1.62×10^{-03}
62	PPP1R18*	208699	170954	1.94×10^{-07}	2.46×10^{-06}	3.19×10^{-03}	3.19×10^{-03}
63	CCND1	841641	595	6.10×10^{-04}	1.67×10^{-09}	2.10×10^{-03}	3.21×10^{-06}
64	SMPD1*	729964	6609	3.03×10^{-05}	2.00×10^{-08}	1.07×10^{-04}	2.74×10^{-03}
65	MEIS3P1*	450152	4213	3.42×10^{-04}	2.82×10^{-10}	7.84×10^{-06}	3.17×10^{-03}
66	MYH10*	823886	4628	2.77×10^{-03}	7.30×10^{-13}	4.38×10^{-06}	1.72×10^{-02}
67	IFITM3	809910	10410	1.45×10^{-03}	2.50×10^{-11}	6.00×10^{-09}	5.59×10^{-04}
68	ARSB*	502055	411	2.91×10^{-04}	9.78×10^{-10}	2.01×10^{-02}	4.32×10^{-03}
69	BCKDHA*	740801	593	1.62×10^{-05}	4.15×10^{-11}	9.91×10^{-06}	1.08×10^{-02}
70	NF2*	769716	4771	3.40×10^{-04}	3.62×10^{-07}	6.50×10^{-09}	9.26×10^{-08}
71	CLEC3B*	345553	7123	1.24×10^{-03}	2.45×10^{-04}	1.07×10^{-03}	2.73×10^{-03}
72	HSPB2	324494	3316	1.77×10^{-02}	1.37×10^{-09}	2.10×10^{-09}	1.74×10^{-06}
73	NFKB1*	789357	4790	2.05×10^{-05}	4.94×10^{-05}	1.92×10^{-05}	1.31×10^{-02}
74	GNA11*	221826	2767	4.08×10^{-07}	1.12×10^{-14}	8.04×10^{-04}	8.58×10^{-03}
75	IGF2	245330	3481	7.93×10^{-04}	4.43×10^{-04}	8.16×10^{-04}	7.51×10^{-04}
76	APLP1	289645	333	1.39×10^{-03}	1.10×10^{-11}	1.18×10^{-05}	8.82×10^{-08}
77	NFIC*	265874	4782	1.18×10^{-04}	3.84×10^{-09}	2.50×10^{-05}	1.90×10^{-02}
78	TEAD4*	346696	7004	1.61×10^{-02}	1.37×10^{-09}	1.92×10^{-05}	9.80×10^{-06}
79	HLA-DPB1	840942	3115	3.38×10^{-02}	7.46×10^{-06}	4.75×10^{-05}	8.64×10^{-04}
80	HMGA1*	782811	3159	3.37×10^{-04}	7.05×10^{-05}	5.88×10^{-04}	8.40×10^{-07}
81	MEST*	898219	4232	1.06×10^{-05}	2.25×10^{-06}	5.69×10^{-05}	1.20×10^{-06}
82	PTPN12*	774502	5782	1.32×10^{-07}	4.28×10^{-07}	2.82×10^{-05}	9.16×10^{-05}
83	IGLL1*	344134	3543	2.55×10^{-03}	7.44×10^{-05}	7.19×10^{-06}	3.55×10^{-02}
84	PTTG1IP*	505491	754	5.37×10^{-06}	1.63×10^{-12}	4.99×10^{-03}	1.83×10^{-03}
85	AKAP7*	195751	9465	9.97×10^{-04}	1.06×10^{-02}	9.20×10^{-04}	1.05×10^{-03}
86	SERPINH1*	142788	871	1.73×10^{-02}	3.87×10^{-11}	3.42×10^{-04}	1.14×10^{-02}
87	SEPT4*	66714	5414	8.07×10^{-04}	3.62×10^{-07}	2.87×10^{-04}	5.90×10^{-05}
88	CITED2*	491565	10370	5.18×10^{-09}	9.60×10^{-10}	8.45×10^{-04}	1.00×10^{-05}
89	TXNRD1*	789376	7296	3.23×10^{-03}	3.25×10^{-02}	2.39×10^{-06}	6.32×10^{-07}
90	EST (RND3)	784593	390	1.92×10^{-06}	1.30×10^{-13}	1.51×10^{-04}	7.13×10^{-04}
91	TRIP6*	811108	7205	3.74×10^{-07}	4.99×10^{-10}	3.88×10^{-08}	3.27×10^{-02}
92	EST (YAP1)	308163	10413	3.09×10^{-05}	7.70×10^{-16}	4.73×10^{-09}	8.68×10^{-04}
93	TAF15*	1474955	8148	6.55×10^{-03}	1.38×10^{-05}	2.34×10^{-03}	3.45×10^{-02}
94	RXRG*	358433	6258	2.31×10^{-02}	1.70×10^{-08}	5.13×10^{-07}	2.87×10^{-04}
95	SERPING1	756556	710	2.82×10^{-03}	2.01×10^{-09}	2.45×10^{-05}	7.04×10^{-03}
96	MYL1*	628336	4632	1.74×10^{-03}	1.72×10^{-03}	2.51×10^{-02}	1.51×10^{-03}

*indicates complement genes. Truncated p-value is the product of p-value of the gene expression value by its rank order in the GANN model and subsequently adjusted using Benjamini and Hochberg false discovery rate.
doi:10.1371/journal.pone.0102483.t001

themselves and provide a complete view of all of the possibilities of network interactions for all biomarkers. Therefore, the principle of the algorithm is to show the relationship between genes from the same pool, to shed light on how these molecules interact with each other and to identify new relationships between these molecules by iteratively calculating the influence that multiple variables may have upon a single one. The main advantage of this algorithm lays in its multi-factorial consideration of each input which allows the magnitude of interaction (i.e. inhibitory, stimulatory, bi- or unidirectional) of a given pair of parameters to be determined on the basis of a matrix of full interaction, and by iteratively examining the weights and prediction performance of each single input expression from all the others within the set.

Figures 1 and 2 present a schematic and a diagrammatic representation of the interaction algorithm, respectively. A summary of the parameter settings for the algorithm is depicted in Table 2.

Multilayer perceptron (MLP). A 3-layered MLP with backpropagation learning and sigmoid activation function was applied to model gene-gene interactions for sarcoma cancers in SRBCTs data. To prevent any relationships being omitted in the interaction analysis, iterative calculation of the influence that multiple genes may have on a single gene was performed in the following way: The process begins with the first input gene in the gene pool, which was defined as an output node in the ANN and omitted from the pool. The remaining genes were then used to predict the omitted gene, and the weights of the trained ANN model were stored. This process was repeated by omitting the second input gene as an output node in ANN and the remaining genes as the input nodes in ANN and so on. The process was iterated until all the genes in the dataset were used as an output. The average values of these iterations were then computed as the interaction score values.
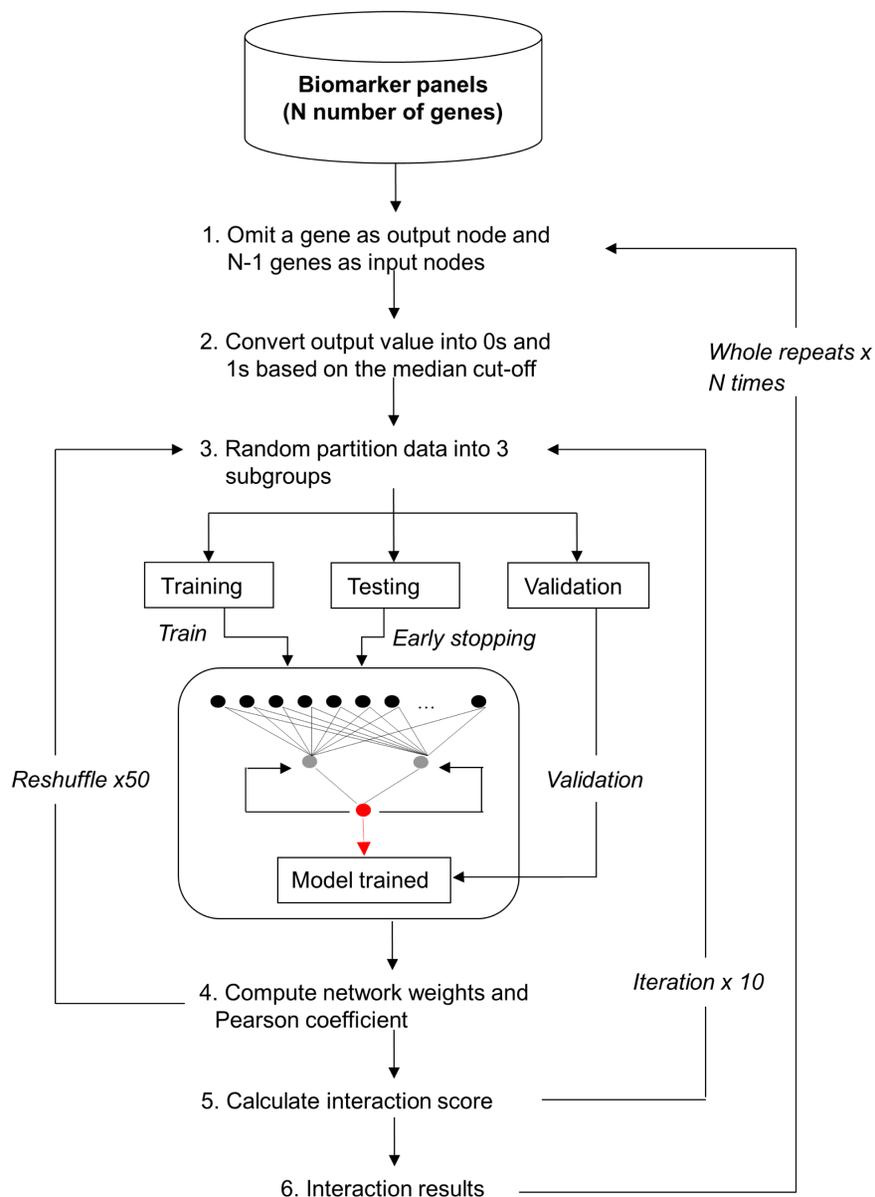
Figure 1. Overview of the interaction algorithm.
doi:10.1371/journal.pone.0102483.g001

The learning process. Given the initial connection weight ω is randomized in-between range -1 to 1, u_k is the sum linear output value from the input signal x of node k and b is the bias value, the input signal x and the output signal y of node k can be expressed as follows:

$$x_k = \sum_{i=1}^{n} \omega_{ki} x_i \qquad (1)$$

and

$$y_k = f(u_k + b_k), \qquad (2)$$

where $\omega \in \{1,2,3,...,n\}$ is the connection weights of the input signals $x \in \{1,2,3,...,n\}$ and $f(\cdot)$ is the sigmoid activation function which can be defined as:

$$f(v) = \frac{1}{1 + \exp(-av)}, \qquad (3)$$

where v is the local signal in the node and a is the slope parameter of the sigmoid function.

In the iterative learning cycle, the weights are adjusted based on the output error δ (see Eq. 4), the total sum-of-squares error based on the difference between the network output y and the target output d of the sample case. In Eq. 4, N is the total number of sample patterns and s is the sample pattern.

$$\delta = \frac{1}{2} \sum_{s=1}^{N} (d_s - y_s)^2 \qquad (4)$$

If both the network output and the target output are identical, no adjustment on the weight in the current learning cycle T is required for this sample pattern and a new sample pattern is fed into the network and the learning cycle continues. If the match is not perfect, the adjustment of the weight $\Delta\omega$ is the proportion of the input signal x_i of node k, the learning rate η and the size of the error δ:

$$\Delta\omega(T) = \eta \delta_k x_i \qquad (5)$$

The new weight for the next learning cycle $T+1$ can then be written as:

$$\omega_{ki}(T+1) = \omega_{ki}(T) + \Delta\omega(T) \qquad (6)$$

To reduce the chance of trapping in endless learning cycle, a momentum α is applied in the learning process and thus, the Eq. 6 can be rewritten as:

$$\omega_{ki}(T+1) = \omega_{ki}(T) + \eta \delta_k(T) x_i + \alpha \omega_{kj}(T-1) \qquad (7)$$

Interaction. To define an interaction map for the genes, the weights of the trained ANN models were used to illustrate and score the interaction between genes, such as the intensity of the relationship between a source gene and a target gene, and the nature of the relationship either stimulatory or inhibitory (i.e. positive or negative sign in correlation result, respectively). The Pearson correlation coefficient r with a cutoff value of 0.7 was implemented in the algorithm to remove the least significant interaction scores.

Monte-Carlo cross-validation (MCCV). To prevent the ANN model from being over-trained, the MCCV strategy was applied as follows. The sample set was randomly partitioned (without replacement in the sample set) into 3 smaller groups: training, test and validation, with the ratio of 0.6:0.2:0.2, respectively. The training and test sets were used to train the ANN model and the validation set was used to examine the trained model. After the model was trained, all samples were re-shuffled and re-partitioned into training, test, and validation sets. This re-shuffle and re-partition process was repeated 50 times, and each time, new different sets of training, test and validation samples were randomly generated.

The entire process cycle was repeated 10 times for single gene. The algorithm was coded in C and empirical work on simulated dataset was presented in the subsequent section.

Visualization of interactome network maps

The Cytoscape software platform (version 2.8) for molecular interaction display was used in this study. Cytoscape [40,41] is an open-source software for analyzing complex biological networks by visually interrogating the relationship of their components using a variety of plug-ins.

Assessment of ANNI using a simulated dataset

A simulated dataset has been generated to assess the prediction ability and the robustness of the optimum ANN parameters (see Table 2) for correctly identify correlated features and to compute their correlation values. This dataset was created using R and contains 100 samples with 25,000 features in which 32 were highly correlated, distributed into 2 major classes. Amongst these 25,000 features, 100 best predictive including 29 pre-defined correlating features were pre-selected using a data mining algorithm.

Despite the obvious advantages of using well-characterised simulated datasets for the testing of new analysis tools, it is important to note that human biological data are complex and the lack in the knowledge of actual biological correlation between sample replicates, molecular relationship between a biological state of a cell and transcript expression, biochemical reaction mechanisms underlying regulatory interactions between features and activity changes from one state to another. This makes artificial data valuable for algorithm development, but is not of value for comparing different methods.

To assess the predictive ability of the algorithm, criteria such as number of hidden nodes used in the network, correlation analysis comparing the predicted correlation scores for each pair of the features with their actual correlation values, interaction signs analysis comparing between the sign of the actual correlation value and the sign of the predicted interaction score and true positive rate (TPR) have been considered. Table 3 shows a summary of the results.

High accuracy on the TPR, correlation result and predicted interaction sign confirm the feasibility of this approach to accurately identify the simulated features having strong correlations. In terms of network architecture, there is no significant improvement on TPR when the number of hidden nodes increases, thereby suggesting that the number of hidden nodes does not affect the predictive ability of the algorithm. A model with 2 hidden nodes performs equally good, or better than those equipped with higher number of hidden nodes and lesser

Omit G1 as output for ANN modelling. The network weights for trained ANN model were stored and the scores were calculated.

Omit G2 as output for ANN modelling. The network weights for trained ANN model were stored and the scores were calculated.

Omit G3 as output for ANN modelling. The network weights for trained ANN model were stored and the scores were calculated.

Process continues with next variable as output for ANN…

Final interaction scores were computed by averaging network weights to represent the direction and signal of the edge.

Figure 2. Diagrammatic representation of the interaction algorithm.
doi:10.1371/journal.pone.0102483.g002

computational time is needed to process the query. Thus, 2 hidden nodes were implemented in the algorithm.

A full, comprehensive empirical validation on the algorithm can be found in Lemetre's PhD Thesis [42].

Results and Discussion

Analysis of all 96 genes from the interaction analysis produced a matrix of $(96 \times (96-1))$ 9,120 potential interactions (Figure 3). Due to the high dimensionality and complexity of the interactions between all genes, it clearly appeared that no relevant information could be elucidated in the map. Thus, only the 96 strongest associations (based upon the averaged values of the scores leading from a given input to the output), each associated to one of the 96 genes, were imported in Cytoscape and are presented in Figure 4. The consequence of this is that for each of these 96 genes, only the strongest interaction among all interactions with the other 95 genes was modeled. This greatly simplified the map, and facilitated the interpretation and understanding of the key features within the map. Table 1 presents the list of 96 genes with its truncated p-values based upon the product of the p-values and its rank order in

Table 2. Summary of the interaction algorithm parameters.

Parameter	Setting
Architecture	N-2-1, where N = total number of genes −1
Learning algorithm	Backpropagation
Activation function	Sigmoid
Epochs	300
Threshold of mean squared error (MSE)	0.01
Window of MSE	100
Momentum	0.5
Learning rate	0.1
Pearson *r* cutoff	0.7
Cross-validation	Monte Carlo cross-validation
Random reshuffle	50 times
Ratio (%) for training: testing: validation	60:20:20
Number of time the whole process repeat for single variable	10

doi:10.1371/journal.pone.0102483.t002

the GANN model. These truncated p-values are adjusted using Benjamini and Hochberg false discovery rate procedure [43].

Genes (i.e. target genes) with most associations include MT1L, HLA-DPB1, GATA2, OLFM1, FCGRT, TNNT1, and FNDC5. The repetitive genes in the map were IGF2 in ranks 2, 7, and 75; and MYO1B in ranks 43 and 50. These repetitive genes show consistent expression pattern in which all 3 selected IGF2 genes show high expression values and strong negative interactions in RMS, and the 2 selected MYO1B genes are in non-sarcoma group. Low truncated p-value (stringent p-value<0.005) for RMS- and EWS-regulated genes are observed in Table 1.

RMS and EWS are soft tissue sarcomas that can be found virtually anywhere in the body and share common clinical characteristics, more frequently occurring in males than females and normally found in children. Early-stage RMS patients are often confused with EWS, consequently inducing lymphatic-related cancer when the RMS tumor migrates to lymph node. Although these tumors have similar clinical conditions, EWS is commonly developed in bones, whereas RMS is more frequently found in skeletal muscle. In the map, genes which have exhibited an up-regulation in both RMS and EWS groups include CTNNA1, GAS1, IFITM3, YAP1, SERPING1, CSDA, FHL3, NFIC, TRIP6, TAF15, and RXRG.

CSDA is a repressor gene involved in various biological processes including skeletal muscle tissue development and organ growth. FHL3 is only expressed in skeletal muscle and could be involved in tumor suppression and repression of MyoD expression. IFITM3 is IFN-induced antiviral protein that plays a role in innate immune response to virus infections. NFIC is a cellular transcription factor involved in DNA binding transcription factor activity. Although the function of TRIP6 is not fully understood, it has been associated with ligand binding of the thyroid receptor in the presence of thyroid hormone. TAF15 plays specific roles during transcription initiation in RNA binding and it may be involved in protein-protein-interaction. RXRG is a retinoic acid (RA) receptor that regulates gene expression in various biological processes, including skeletal muscle tissue development, heart development, and response to hormone stimulus.

Rhabdomyosarcoma (RMS)

RMS is a connective tissue related cancer. The cause of this sarcoma is unknown, but its development has been associated with desmin, MyoD1 and myogenin (MYOG) [44]. MyoD1 is a regulator involved in muscle regeneration and muscle cell differentiation and MYOG is the muscle-specific transcription factor that plays a role in the development of functional skeletal muscle.

TNNT1 (mainly expressed in heart muscle) and FNDC5 (normally induced by the expression of PGC-1 alpha in muscle) are two highly associated genes in the RMS cancer, which act as target genes (i.e. interaction hubs) that interconnect genes which also show high expression values in the RMS cancer. These 2 genes are bridged by FGFR4, which plays a role in the regulation of cell proliferation, differentiation and migration.

TNNT1 interaction cluster. TNNT1 protein is inhibited by genes RXRG, MYL1, RND3, FHL3 and FGFR4. Amongst these genes, RXRG, MYL1 and FHL3 were complementary genes to the genes reported by Khan *et al.*

RXRG is a tumor suppressor gene that mediates the anti-proliferative effect of retinoic acid (RA), an essential metabolite of vitamin A for the growth, development and cell differentiation of vertebrate species. This protein suppresses tumor growth by increasing the anti-proliferative effects of RA in the tumor cells. MYL1 is the motor protein known for the role in muscle cell

activities including vesicle transportation inside the muscle cell. A negative interaction on this protein indicates that tumor cells are constrained in a particular location rather than move randomly. FHL3 expressed only in skeletal muscle has been known by its role in skeletal myogenesis [45–47], although its actual function is unknown. This gene has been related to cell spreading and actin stress fiber disassembly [46] and is involved in tumor suppression/ repression of MyoD expression. RND3 is a member of the Rho family GTPase protein superfamily that acts as a negative regulator in cytoskeletal organization. It is known to have a role in myoblast fusion [48] and to be responsible for down-regulation in focal adhesions and stress fibers. FGFR4 is a tyrosine kinase receptor responsible for signal transduction activities in the cell. Although the activity of this protein is undetectable in normal tissues, it becomes active when a tumor is formed. High expression of FGFR4 has been associated with advanced-stage in RMS cancer and a poor survival rate [49].

Although over-expression of FGFR4 in the map (see Figure 4), is inhibited by other low/moderately expressed genes, it may suggest the formation of tumor cell in skeletal muscle. Over-expression of FGFR4 may stimulate the expression of TNNT1; the regulator for striated muscle contraction, despite the fact that a negative interaction between these 2 genes was detected. Low expression of FHL3, RXRG, MYL1 and RND3 may also promote the mutation of TNNT1 genes due to a suppression of the anti-proliferative effects of RA in the RMS cells. The low activity of these genes might influence normal cell spreading, actin stress fiber disassembly and, consequently, tumor cells migration. TNNT1 is the diagnostic marker for nemaline myopathy [50,51]. The high expression value in mutated TNNT1 genes suggests that RMS tumors may be congenital.

FNDC5 interaction cluster. FNDC5 protein expressed in muscle is inhibited by RMS-expressed genes including TNNT2, BIN1, SEPT4, MYL4 and HSPB2. Up-regulation of this gene suggests that the increased level of irisin hormone [52,53] promotes the beneficial effects of exercise on metabolic pathways. TNNT2 and BIN1 are proteins that play important roles in cardiac muscle development. Over-expression of TNNT2 has been correlated to myocardial stunning in hemodialysis patients [54,55], and the disruption of this gene could lead to impaired cardiac development in the embryo and infant [56–58]. Deficiency of the BIN1 gene has been correlated with cardiomyopathy. SEPT4 is the nucleotide binding proteins highly expressed in heart and brain. It regulates cytoskeletal organization during embryonic and adult life [59]. MYL4 is the hexameric ATPase cellular motor proteins that commonly found in embryonic muscle and adult atria. Over-expression of this gene was normally found in patients with hypertrophic cardiomyopathy and congenital heart diseases [60]. HSPB2 is another protein that expressed in the heart and skeletal muscles. Over-expression of this gene indicates the efficient recovery of motor neurons following nerve injury [61].

In the map (Figure 4), observation on the over-expression pattern on TNNT2, BIN1 and MYL4, moderately expression on HSPB2 and SEPT4 is lowly expressed. This suggests that RMS patients have low chance to develop cardiomyopathy due to high expressions of TNNT2, BIN1, SEPT4 and HSPB2 in RMS pathway could potentially suppress expression level of FNDC5.

Ewing's sarcoma (EWS)

EWS is a bone malignancy which commonly affects areas including the pelvis, femur, humerus and the ribs [44]. The cell origin of this tumor is uncertain. However, this cancer has a shared cytogenetic abnormality with the primitive neuroectodermal tumor (PNET) which arises from the soft tissue or bone. The

Table 3. Summary of the assessment results.

No. of hidden nodes	No. of features	Person's coefficient	% of correctly assigned signs	Ave. true positive rate (%)
2	32	0.805	89.16	93.16
	100	0.865	91.26	70.33
5	32	0.653	80.31	89.00
	100	0.871	90.89	68.50
10	32	0.607	79.32	92.66
	100	0.866	88.55	81.16

doi:10.1371/journal.pone.0102483.t003

shared cytogenetic abnormality involves the EWS/FLI fusion in t(11;22)(q24;q12), a signature marker for EWS/PNET from other small round tumors [62].

Two of the highest associated genes observed were FCGRT and OLFM1 (Figure 4). FCGRT and OLFM1 acted as interaction hubs, in which FCGRT plays a role in interconnecting genes associated to the EWS/FLI fusion and OLFM1 integrates variety biological process performed by other genes.

FCGRT interaction cluster. FCGRT protein is known as the promoter marker to EWS-FLI fusion, one of the common

cytogenetic abnormalities on the t(11;22) translocation for Ewing tumors. In the map, highly expressed FCGRT was suppressed by TLE2, CITED2, CAV1, PTTG1IP and KDSR. Amongst these associated genes, CITED2 and PTTG1IP were in addition to those genes reported by Khan *et al.*

CITED2 is a cardiac transcription factor responsible for inhibiting transactivation activity of hypoxia-induced genes. Mutation of this gene decreases its ability to mediate the expressions of VEGF and PITX2C genes, suggesting that it may play a role in the development of congenital heart disease [63].
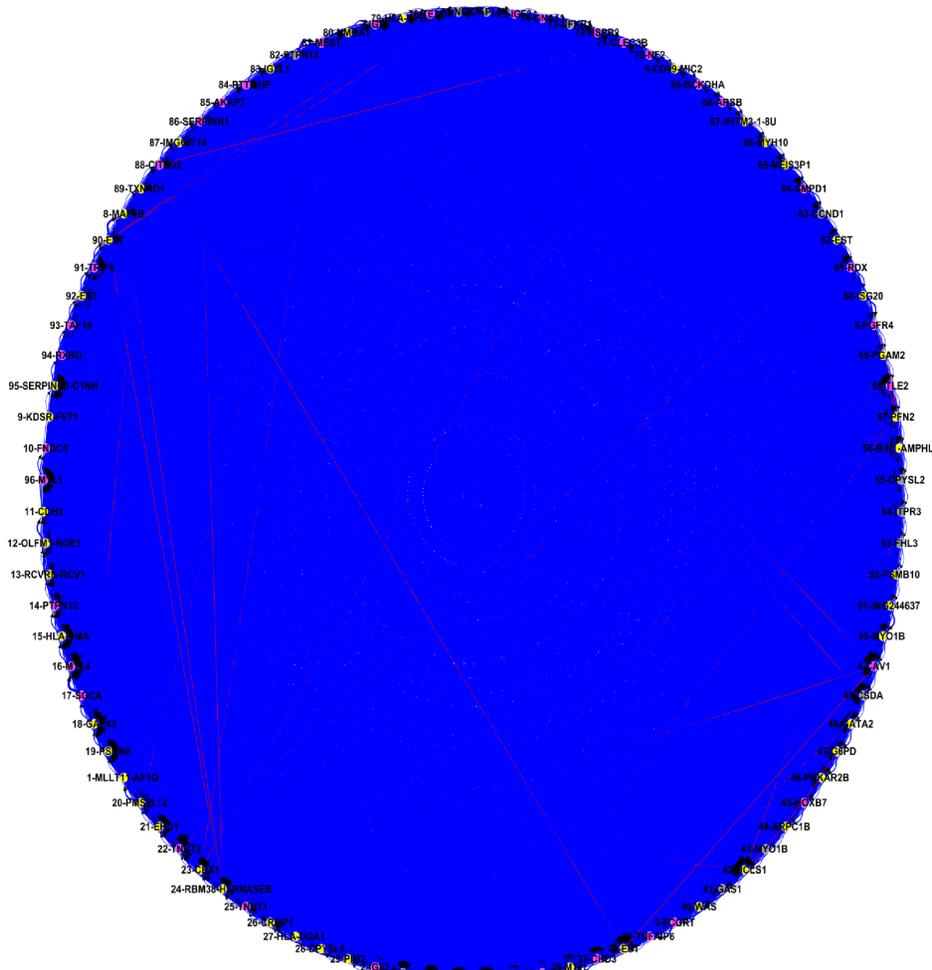
Figure 3. A complete interactome network map.
doi:10.1371/journal.pone.0102483.g003
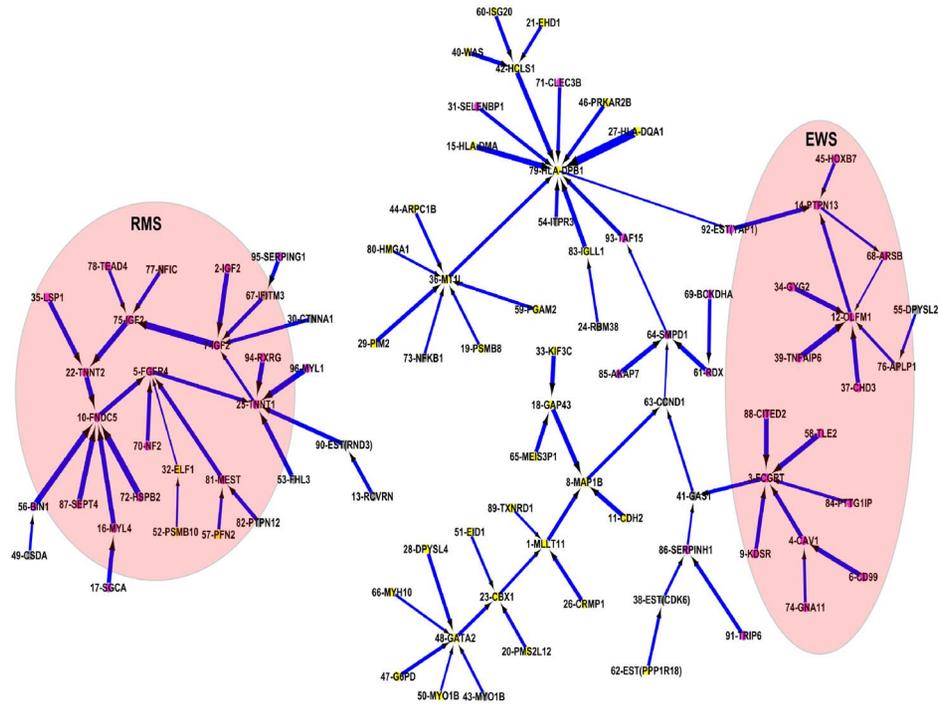
Figure 4. A simplified interactome network map for the 96 selected genes by ANN network inference algorithm. The red node is the genes with high expression values in either of the sarcoma cancers. The gray node is the gene with high expression values in more than one cancer groups in which one of these groups is sarcoma cancer. Yellow node is the genes with low expression values in both sarcoma cancers.
doi:10.1371/journal.pone.0102483.g004

Over-expressed CITED2 induces resistance to cisplatin [64], a commonly used chemotherapeutic drug for sarcomas and other solid malignancies. CAV1 is another promoter marker associates to EWS-FLI fusion. Over-expression of this gene promotes metastasis on Ewing tumor [65,66]. PTTG1IP is the promoter gene that has been correlated with the Runx2 gene expression in osteoblastic differentiation and skeletal morphogenesis [67]. Runx2 is the master regulator of osteoblast differentiation and study reviewed that blockade of this gene by EWS-FLI induce osteoblast specification of a mesenchymal progenitor cell and thus, disrupting interactions between Runx2 and EWS-FLI may promote differentiation of the tumor cell [68]. TLE2 proteinTLE2 is the member of TLE family and its actual function is not well known. It is believed that the expression of TLE family members is required for EWS oncogenic transformation [62] and acts as repressor protein in histone modification when recruited by NKX2–2 gene [62,69]. NKX2–2 is the known immunohisto-chemical marker for Ewing sarcoma [70]. KDSR is the murine gene and has been known by its role in the development of spinal muscular atrophy (SMA) disease in cattle [71]. Although its actual function in human SMA is unknown, a study has claimed that the mutation of this gene does not contribute significantly to the cause of human SMA [72], rather that it associated with chromosomal rearrangement in chronic lymphatic leukemia [73,74].

In Figure 4, over-expression of CAV1, KDSR and CITED2 inhibited up-regulated FCGRT suggests these EWS patients are EWS-FLI affected sarcoma and have deficiency of cisplatin controlled repression of EWS-FLI fusion. Low expression values of TLE2 and PTTC1IP further indicate the promotion of tumor cell differentiation and histone modification due to inhibition on Runx2 and NKX2–2 genes.

OLFM1 interaction cluster. OLFM1 protein which has abundant expression in brain has been associated to the biological process in nerve tissue, is inhibited by GYG2, APLP1, CHD3, TNFAIP6, and ARSB genes. This cluster shows the interaction between wide ranges of glycoproteins, histone deacetylase and Wnt signaling pathway. GYG2 and APLP1 are glycoproteins expressed in liver and membrane, respectively. GYG2 is a self-glucosylating protein found in chromosome X and involved in the initiation reactions of glycogen biosynthesis and blood glucose homeostasis. Over-expression of this gene resulted in excess glycogen storage levels in liver and may lead to glycogenosis due to the deficiency of liver phosphorylase. APLP1 is a membrane-associated glycoprotein that is cleaved by secretases and up-regulation of this protein reduced endocytosis of amyloid precursor protein (APP), and makes more APP available for alpha-secretase cleavage [75].

ARSB is a sulfatase protein responsible for protein fragmenta-tion and mediating intracellular oxygen signaling. Deficiency of this gene due to hypoxia leads to an accumulation of GAGs protein in lysosomes, resulting to mucopolysaccharidosis [76]. CHD3 is part of a histone deacetylase complex participating in chromatin remodeling. Histone deacetylase removes acetyl group on a histone so that the DNA can be tightly wrapped by histones. TNFAIP6 is a multifunctional protein that exhibited in many pathological and physiological contexts. It plays important roles in inflammation and tissue/bone remodeling. It acts as a protector by suppressing tumor necrosis factor (TNFSF11)-induced bone erosion caused by inflammatory processes in arthritis diseases, and exhibits a homeostatic function by interacting with bone morphogenetic protein 2 (BMP2) and TNFSF11 to balance mineralization by osteoblasts and bone resorption by osteoclasts [77]. The binding of TNFSF11 with GAGs and other proteins ligands enhances its activity in various biological processes, including leucocyte adhesion and anti-plasmin activity [78].

OLFM1 protein has been associated with Wnt signaling pathway in which the activation of Wnt signaling pathway in receptive endometrium suppresses the OLFM1 gene and leads to ectopic pregnancy in humans [79]. Wnt signaling pathway plays a critical role in the normal development of multiple neuroectodermal tissues, cell degradation and cell death. This pathway has been associated with neuroectodermal Ewing sarcoma [80], osteosarcoma [81] and possibly embryogenesis [82].

In the map (Figure 4), up-regulation of the OLFM1 gene suppresses the activities of extracellular inhibitors in Wnt signaling pathway. This may promote tumor cell proliferation, learning impairment, suffocation of normal cells (i.e. hypoxia) in ARSB and a deficiency of APP for alpha secretase cleavage in EWS patients. Alpha secretases have been implicated in the regulation of learning and memory formation. Furthermore, over-expression of CHD3 in tumor cells accelerates the interactions between DNA and nuclear proteins, and contributes to tumor cell differentiation, proliferation and migration. Although the moderately expressed genes GYG2 and TNFAIP6 suppress the activity of tumor proliferation, intervention from over-expressed CHD3 and down-regulated APLP1 and ARSB promote diversification of Ewing tumors.

A triangular affair between Wnt signaling pathways, Fas/Rho signaling pathways and intracellular oxygen signaling is observed in the map. Down-regulated ARSB (mediates oxygen in intracellular process) due to intervention of moderately expressed Fas receptor PTPN13 (a large intracellular protein mediates programmed cell death and regulates Rho signaling pathway) fails to activate inhibitors in Wnt signaling pathways (over-expression of OLFM1). Both the Fas receptor and Wnt signaling pathways mediate programmed cell death and the Rho signaling pathway responsible for cell proliferation, apoptosis and gene expression. The over-expression of PTPN13 was suppressed by OLFM1.

Conclusions

We have previously shown that assessing the statistical significance of genes, based on classification accuracy alone, may not be useful for cancer diagnosis, as classification results are subject to the proposed computational methods and a lack of biological association to those genes to provide a global view on the development of the tumor of interest. Thus, this paper discussed the feasibility of using ANNI to model the potential interactions between the correlated genes in childhood sarcoma cancers. This interaction information could be useful for other researchers who are assessing the practicality of using these genes as new target markers for sarcoma cancers in the clinical setting.

Due to the high dimensional and complexity of the interactions between genes, this paper showed only the strongest association for each of the genes, rather than the complete potential 9,120 interactions existing among them. Even so, this approach highlighted that certain genes were highly influential on sarcoma cancers.

Based on the interaction map, several observations on sarcoma cancers in SRBCTs dataset were noted. The RMS patients in this dataset are likely to be congenital (due to up-regulation of mutated TNNT1 stimulated by FGFR4 and suppression of anti-proliferative effects of RA in FHL3, RXRG, MYL1 and RND3) and these patients have a lower risk of developing cardiomyopathy (due to over-expression in TNNT2, BIN1, SEPT4 and HSPB2 potentially suppress the expression level of FNDC5). The EWS patients are likely to be affected by EWS-FLI fusion (due to deficiency of cisplatin controlled repression, up-regulation of CAV1, KDSR and CITED2, and suppression of TLE2 and PTTC1IP) and involves various signaling pathway complications including Wnt, Fas/Rho and intracellular oxygen pathways (due to over-expression in OLFM1 and CHD3, and low expression in APLP1 and ARSB).

Additional information: The proposed algorithm will be available by contacting the corresponding author. Individuals from all disciplines are invited to access the software on a collaborative basis. Information on the development and background to the algorithm are publicly available in Lemetre's PhD thesis, which is referenced in the manuscript.

Author Contributions

Conceived and designed the experiments: DLT GRB. Performed the experiments: DLT. Analyzed the data: DLT DJB GKRD. Contributed reagents/materials/analysis tools: DLT. Wrote the paper: DLT DJB. Development of algorithm: CL GRB. Reviewed the paper: GRB GKRD.

References

1. Tong DL, Schierz AC (2011) Hybrid genetic algorithm-neural network: feature extraction for unpreprocessed microarray data. Artif Intell Med 53: 47–56. Available: http://www.ncbi.nlm.nih.gov/pubmed/21775110. Accessed 10 March 2013.
2. Khan J, Wei JS, Ringnér M, Saal LH, Ladanyi M, et al. (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 7: 673–679. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1282521&tool=pmcentrez&rendertype=abstract.
3. Taylor R (1990) Interpretation of the Correlation Coefficient: A Basic Review. J Diagnostic Med Sonogr 6: 35–39. Available: http://jdm.sagepub.com/cgi/doi/10.1177/875647939000600106. Accessed 23 August 2013.
4. Martínez-Abraín A (2008) Statistical significance and biological relevance: A call for a more cautious interpretation of results in ecology. Acta Oecologica 34: 9–11. Available: http://linkinghub.elsevier.com/retrieve/pii/S1146609×08000337. Accessed 9 August 2013.
5. Tong DL (2009) Hybridising Genetic Algorithm-Neural Network (GANN) in marker genes detection. In: Wang X, Yeung DS, Lai LL, editors. Proceedings of 2009 International Conference on Machine Learning and Cybernetics (ICMLC) - Vol. 2. Baoding, Hebei, China: IEEE. pp. 1082–1087. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5212372. Accessed 3 April 2013.
6. Tong DL, Phalp K, Schierz A, Mintram R (2009) Innovative hybridisation of genetic algorithms and neural networks in detecting marker genes for leukaemia cancer. In: Kadirkamanathan V, Sanguinetti G, Girolami M, Niranjan M, Noirel J, editors. Proceedings of the 4th IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB) - Supplementary. Sheffield, UK: Springer-Verlag Berlin, Heidelberg. pp. 1–7.
7. Tong DL, Mintram R (2010) Genetic Algorithm-Neural Network (GANN): a study of neural network activation functions and depth of genetic algorithm search applied to feature selection. Int J Mach Learn Cybern 1: 75–87. Available: http://www.springerlink.com/index/10.1007/s13042-010-0004-x. Accessed 2 March 2013.
8. Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101–113. Available: http://www.ncbi.nlm.nih.gov/pubmed/14735121. Accessed 27 February 2013.
9. Schwikowski B, Uetz P, Fields S (2000) A network of protein-protein interactions in yeast. Nat Biotechnol 18: 1257–1261. Available: http://www.ncbi.nlm.nih.gov/pubmed/11101803.
10. Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci 100: 12123–12128. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=218723&tool=pmcentrez&rendertype=abstract.
11. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, et al. (2004) A map of the interactome network of the metazoan C. elegans. Science (80-) 303: 540–543. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1698949&tool=pmcentrez&rendertype=abstract. Accessed 4 March 2013.
12. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2: 2366–2382. Available: http://www.ncbi.nlm.nih.gov/pubmed/17947979. Accessed 2 April 2013.
13. Wu G, Zhu L, Dent JE, Nardini C (2010) A comprehensive molecular interaction map for rheumatoid arthritis. PLoS One 5: e10137. Available:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2855702&tool = pmcentrez&rendertype = abstract. Accessed 4 March 2013.

14. Albert R (2004) Boolean Modeling of Genetic Regulatory Networks. Complex Networks, Lect Notes Phys 650: 459–481. doi:10.1007/978-3-540-44485-5_21.

15. Giacomantonio CE, Goodhill GJ (2010) A Boolean model of the gene regulatory network underlying Mammalian cortical area development. PLoS Comput Biol 6: e1000936. Available: http://www.pubmedcentral.nih.gov/articlerender. fcgi?artid = 2940723&tool = pmcentrez&rendertype = abstract. Accessed 23 March 2013.

16. Hartemink AJ, Gifford DK, Jaakkola TS, Young RA (2002) Bayesian methods for elucidating genetic regulatory networks. IEEE Intell Syst 17: 37–43. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber = 999218. Accessed 16 April 2013.

17. Perrin B-E, Ralaivola L, Mazurie A, Bottani S, Mallet J, et al. (2003) Gene networks inference using dynamic Bayesian networks. Bioinformatics 19 Suppl 2: ii138–ii148. Available: http://bioinformatics.oxfordjournals.org/cgi/doi/10. 1093/bioinformatics/btg1071. Accessed 27 February 2013.

18. Ram R, Chetty M (2009) MCMC based Bayesian inference for modelling gene networks. In: Kadirkamanathan V, Sanguinetti G, Girolami M, Niranjan M, Noirel J, editors. Proceedings of the 4th IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB 2009). Sheffield, UK: Springer-Verlag Berlin, Heidelberg. pp. 293–306.

19. Christley S, Nie Q, Xie X (2009) Incorporating existing network information into gene network inference. PLoS One 4: e6799. Available: http://www.pubmedcentral. nih.gov/articlerender.fcgi?artid = 2729382&tool = pmcentrez&rendertype = abstract. Accessed 10 September 2013.

20. Xu R, Hu X, Wunsch II DC (2004) Inference of genetic regulatory networks with recurrent neural network models. In: Hudson DL, Liang Z-P, Dumont G, editors. Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS 2004). San Francisco, CA, USA: IEEE, Vol. 4. pp. 2905–2908. Available: http://www.ncbi.nlm.nih. gov/pubmed/17975278.

21. Xu R, Venayagamoorthy G, Wunsch II DC (2006) A study of particle swarm optimization in gene regulatory networks inference. In: Wang J, Yi Z, Zurada JM, Lu B-L, Yin H, editors. Proceedings of the 3rd international conference on Advances in Neural Networks (ISNN 2006) - Volume Part III. Springer-Verlag Berlin, Heidelberg. pp. 648–653. doi:10.1007/11760191_95.

22. Lemetre C, Lancashire LJ, Rees RC, Ball GR (2009) Artificial Neural Network Based Algorithm for Biomolecular Interactions Modeling. In: Cabestany J, Sandoval F, Prieto A, Corchado JM, editors. Proceedings of the 10th International Work-Conference on Artificial Neural Networks (IWANN 2009) - Part I: Bio-Inspired Systems: Computational and Ambient Intelligence. Salamanca, Spain: Springer-Verlag Berlin, Heidelberg. pp. 877–885. doi:10.1007/978-3-642-02478-8_110.

23. Maraziotis IA, Dragomir A, Thanos D (2010) Gene regulatory networks modelling using a dynamic evolutionary hybrid. BMC Bioinformatics 11: 140. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2848237&tool = pmcentrez&rendertype = abstract.

24. Günther F, Wawro N, Bammann K (2009) Neural networks for modeling gene-gene interactions in association studies. BMC Genet 10: 87. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2817696&tool = pmcentrez&rendertype = abstract. Accessed 9 August 2013.

25. Cheng L, Hou Z-G, Lin Y, Tan M, Zhang WC, et al. (2011) Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks. IEEE Trans Neural Netw 22: 714–726. Available: http://www.ncbi.nlm.nih.gov/pubmed/21427022.

26. Schlitt T, Brazma A (2007) Current approaches to gene regulatory network modelling. BMC Bioinformatics 8 Suppl 6: S9. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 1995542&tool = pmcentrez&rendertype = abstract. Accessed 7 March 2013.

27. Karlebach G, Shamir R (2008) Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol 9: 770–780. Available: http://www.ncbi.nlm. nih.gov/pubmed/18797474. Accessed 27 February 2013.

28. De Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9: 67–103. Available: http://www.ncbi.nlm. nih.gov/pubmed/11911796.

29. Chipman KC, Singh AK (2009) Predicting genetic interactions with random walks on biological networks. BMC Bioinformatics 10: 17. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2653491&tool = pmcentrez&rendertype = abstract. Accessed 19 August 2013.

30. Lee W-P, Tzou W-S (2009) Computational methods for discovering gene networks from expression data. Brief Bioinform 10: 408–423. Available: http:// www.ncbi.nlm.nih.gov/pubmed/19505889. Accessed 8 August 2013.

31. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, et al. (2002) Novel artificial neural network for early detection of prostate cancer. J Clin Oncol 20: 921–929. Available: http://www.ncbi.nlm.nih.gov/pubmed/11844812.

32. O'Neill MC, Song L (2003) Neural network analysis of lymphoma micro-array data: prognosis and diagnosis near-perfect. BMC Bioinformatics 4: 13. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 155539&tool = pmcentrez&rendertype = abstract.

33. Wei JS, Greer BT, Westermann F, Steinberg SM, Son C-G, et al. (2004) Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. Cancer Res 64: 6883–

6891. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 1298184&tool = pmcentrez&rendertype = abstract.

34. Lancashire LJ, Powe DG, Reis-Filho JS, Rakha E, Lemetre C, et al. (2010) A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks. Breast Cancer Res Treat 120: 83–93. Available: http://www.ncbi.nlm.nih.gov/pubmed/19347577. Accessed 10 September 2013.

35. Lancashire LJ, Rees RC, Ball GR (2008) Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach. Artif Intell Med 43: 99–111. Available: http://www.ncbi.nlm.nih.gov/pubmed/ 18420392. Accessed 10 September 2013.

36. Matharoo-Ball B, Hughes C, Lancashire L, Tooth D, Ball G, et al. (2007) Characterization of biomarkers in polycystic ovary syndrome (PCOS) using multiple distinct proteomic platforms. J Proteome Res 6: 3321–3328. Available: http://www.ncbi.nlm.nih.gov/pubmed/17602513.

37. Dhondalay GK, Tong DL, Ball GR (2011) Estrogen receptor status prediction for breast cancer using artificial neural network. Proceedings of the 2011 International Conference on Machine Learning and Cybernetics (ICMLC 2011) - Vol. 2. Guilin, China: IEEE. pp. 727–731. Available: http://ieeexplore.ieee. org/lpdocs/epic03/wrapper.htm?arnumber = 6016771. Accessed 10 March 2013.

38. Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, et al. (2009) MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. Breast cancer Res 11: R27. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 2716495&tool = pmcentrez&rendertype = abstract. Accessed 2 April 2013.

39. Lancashire LJ, Lemetre C, Ball GR (2009) An introduction to artificial neural networks in bioinformatics–application to complex microarray and mass spectrometry datasets in cancer studies. Brief Bioinform 10: 315–329. Available: http://www.ncbi.nlm.nih.gov/pubmed/19307287. Accessed 11 September 2013.

40. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–2504. Available: http://www.pubmedcentral.nih.gov/ articlerender.fcgi?artid = 403769&tool = pmcentrez&rendertype = abstract. Accessed 28 February 2013.

41. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27: 431–432. Available: http://www.pubmedcentral.nih.gov/articlerender. fcgi?artid = 3031041&tool = pmcentrez&rendertype = abstract. Accessed 1 March 2013.

42. Lemetre C (2010) Artificial neural network techniques to investigate potential interactions between biomarkers Nottingham Trent University.

43. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Pratical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B 57: 289–300.

44. Rajwanshi A, Srinivas R, Upasana G (2009) Malignant small round cell tumors. J Cytol 26: 1–10. Available: http://www.pubmedcentral.nih.gov/articlerender. fcgi?artid = 3167982&tool = pmcentrez&rendertype = abstract. Accessed 26 April 2013.

45. Morgan MJ, Madgwick AJ (1999) The LIM proteins FHL1 and FHL3 are expressed differently in skeletal muscle. Biochem Biophys Res Commun 255: 245–250. doi:10.1006/bbrc.1999.0179.

46. Coghill ID, Brown S, Cottle DL, McGrath MJ, Robinson PA, et al. (2003) FHL3 is an actin-binding protein that regulates alpha-actinin-mediated actin bundling: FHL3 localizes to actin stress fibers and enhances cell spreading and stress fiber disassembly. J Biol Chem 278: 24139–24152. doi:10.1074/jbc.M213259200.

47. Cottle DL, McGrath MJ, Cowling BS, Coghill ID, Brown S, et al. (2007) FHL3 binds MyoD and negatively regulates myotube formation. J Cell Sci 120: 1423–1435. doi:10.1242/jcs.004739.

48. Fortier M, Comunale F, Kucharczak J, Blangy A, Charrasse S, et al. (2008) RhoE controls myoblast alignment prior fusion through RhoA and ROCK. Cell Death Differ 15: 1221–1231. doi:10.1038/cdd.2008.34.

49. Taylor JG, Cheuk AT, Tsang PS, Chung J-Y, Song YK, et al. (2009) Identification of FGFR4-activating mutations in human rhabdomyosarcomas that promote metastasis in xenotransplanted models. J Clin Invest 119: 3395–3407. doi:10.1172/JCI39703.

50. Johnston JJ, Kelley RI, Crawford TO, Morton DH, Agarwala R, et al. (2000) A novel nemaline myopathy in the Amish caused by a mutation in troponin T1. Am J Hum Genet 67: 814–821. doi:10.1086/303089.

51. Clarkson E, Costa CF, Machesky LM (2004) Congenital myopathies: diseases of the actin cytoskeleton. J Pathol 204: 407–417. doi:10.1002/path.1648.

52. Huh JY, Panagiotou G, Mougios V, Brinkoetter M, Vamvini MT, et al. (2012) FNDC5 and irisin in humans: I. Predictors of circulating concentrations in serum and plasma and II. mRNA expression and circulating concentrations in response to weight loss and exercise. Metabolism 61: 1725–1738. doi:10.1016/ j.metabol.2012.09.002.

53. Lecker SH, Zavin A, Cao P, Arena R, Allsup K, et al. (2012) Expression of the irisin precursor FNDC5 in skeletal muscle correlates with aerobic exercise performance in patients with heart failure. Circ Heart Fail 5: 812–818. doi:10.1161/CIRCHEARTFAILURE.112.969543.

54. Breidthardt T, Burton JO, Odudu A, Eldehni MT, Jefferies HJ, et al. (2012) Troponin T for the detection of dialysis-induced myocardial stunning in

hemodialysis patients. Clin J Am Soc Nephrol 7: 1285–1292. doi:10.2215/CJN.00460112.

55. Pianta TJ, Horvath AR, Ellis VM, Leonetti R, Moffat C, et al. (2012) Cardiac high-sensitivity troponin T measurement: a layer of complexity in managing haemodialysis patients. Nephrology (Carlton) 17: 636–641. doi:10.1111/j.1440–1797.2012.01625.x.

56. Gomes A V, Barnes JA, Harada K, Potter JD (2004) Role of troponin T in disease. Mol Cell Biochem 263: 115–129.

57. Nishii K, Morimoto S, Minakami R, Miyano Y, Hashizume K, et al. (2008) Targeted disruption of the cardiac troponin T gene causes sarcomere disassembly and defects in heartbeat within the early mouse embryo. Dev Biol 322: 65–73. doi:10.1016/j.ydbio.2008.07.007.

58. Lopes DN, Ramos JMM, Moreira MEL, Cabral JA, de Carvalho M, et al. (2012) Cardiac troponin T and illness severity in the very-low-birth-weight infant. Int J Pediatr 2012: 479242. doi:10.1155/2012/479242.

59. Liu W (2012) SEPT4 is regulated by the Notch signaling pathway. Mol Biol Rep 39: 4401–4409. doi:10.1007/s11033-011-1228-x.

60. Abdelaziz AI, Pagel I, Schlegel W-P, Kott M, Monti J, et al. (2005) Human atrial myosin light chain 1 expression attenuates heart failure. Adv Exp Med Biol 565: 283–92; discussion 92, 405–15. doi:10.1007/0-387-24990-7_21.

61. Sharp P, Krishnan M, Pullar O, Navarrete R, Wells D, et al. (2006) Heat shock protein 27 rescues motor neurons following nerve injury and preserves muscle function. Exp Neurol 198: 511–518. doi:10.1016/j.expneurol.2005.12.031.

62. Owen LA, Kowalewski AA, Lessnick SL (2008) EWS/FLI mediates transcriptional repression via NKX2.2 during oncogenic transformation in Ewing's sarcoma. PLoS One 3: e1965. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2291578&tool=pmcentrez&rendertype=abstract. Accessed 26 April 2013.

63. Li Q, Pan H, Guan L, Su D, Ma X (2012) CITED2 mutation links congenital heart defects to dysregulation of the cardiac gene VEGF and PITX2C expression. Biochem Biophys Res Commun 423: 895–899. doi:10.1016/j.bbrc.2012.06.099.

64. Wu Z-Z, Sun N-K, Chao CC-K (2011) Knockdown of CITED2 using short-hairpin RNA sensitizes cancer cells to cisplatin through stabilization of p53 and enhancement of p53-dependent apoptosis. J Cell Physiol 226: 2415–2428. doi:10.1002/jcp.22589.

65. Sáinz-Jaspeado M, Lagares-Tena L, Lasheras J, Navid F, Rodriguez-Galindo C, et al. (2010) Caveolin-1 modulates the ability of Ewing's sarcoma to metastasize. Mol Cancer Res 8: 1489–1500. doi:10.1158/1541–7786.MCR-10–0060.

66. Sengupta A, Mateo-Lozano S, Tirado OM, Notario V (2011) Auto-stimulatory action of secreted caveolin-1 on the proliferation of Ewing's sarcoma cells. Int J Oncol 38: 1259–1265. doi:10.3892/ijo.2011.963.

67. Stock M, Schäfer H, Fliegauf M, Otto F (2004) Identification of novel genes of the bone-specific transcription factor Runx2. J Bone Miner Res 19: 959–972. doi:10.1359/jbmr.2004.19.6.959.

68. Li X, McGee-Lawrence ME, Decker M, Westendorf JJ (2010) The Ewing's sarcoma fusion protein, EWS-FLI, binds Runx2 and blocks osteoblast differentiation. J Cell Biochem 111: 933–943. doi:10.1002/jcb.22782.

69. Patel N, Black J, Chen X, Marcondes AM, Grady WM, et al. (2012) DNA methylation and gene expression profiling of ewing sarcoma primary tumors

reveal genes that are potential targets of epigenetic inactivation. Sarcoma 2012: 498472. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3447379&tool=pmcentrez&rendertype=abstract. Accessed 26 April 2013.

70. Yoshida A, Sekine S, Tsuta K, Fukayama M, Furuta K, et al. (2012) NKX2.2 is a useful immunohistochemical marker for Ewing sarcoma. Am J Surg Pathol 36: 993–999. doi:10.1097/PAS.0b013e31824ee43c.

71. Krebs S, Medugorac I, Röther S, Strässer K, Förster M (2007) A missense mutation in the 3-ketodihydrosphingosine reductase FVT1 as candidate causal mutation for bovine spinal muscular atrophy. Proc Natl Acad Sci U S A 104: 6746–6751. doi:10.1073/pnas.0607721104.

72. Parkinson NJ, Baumer D, Rose-Morris A, Talbot K (2008) Candidate screening of the bovine and feline spinal muscular atrophy genes reveals no evidence for involvement in human motor neuron disorders. Neuromuscul Disord 18: 394–397. doi:10.1016/j.nmd.2008.03.003.

73. Rimokh R, Gadoux M, Berthéas MF, Berger F, Garoscio M, et al. (1993) FVT-1, a novel human transcription unit affected by variant translocation t(2;18)(p11;q21) of follicular lymphoma. Blood 81: 136–142.

74. Czuchlewski DR, Csernus B, Bubman D, Hyjek E, Martin P, et al. (2008) Expression of the follicular lymphoma variant translocation 1 gene in diffuse large B-cell lymphoma correlates with subtype and clinical outcome. Am J Clin Pathol 130: 957–962. doi:10.1309/AJCP12HIRWSRQLAN.

75. Neumann S, Schöbel S, Jäger S, Trautwein A, Haass C, et al. (2006) Amyloid precursor-like protein 1 influences endocytosis and proteolytic processing of the amyloid precursor protein. J Biol Chem 281: 7583–7594. doi:10.1074/jbc.M508340200.

76. Bhattacharyya S, Tobacman JK (2012) Hypoxia reduces arylsulfatase B activity and silencing arylsulfatase B replicates and mediates the effects of hypoxia. PLoS One 7: e33250. doi:10.1371/journal.pone.0033250.

77. Mahoney DJ, Mikecz K, Ali T, Mabilleau G, Benayahu D, et al. (2008) TSG-6 regulates bone remodeling through inhibition of osteoblastogenesis and osteoclast activation. J Biol Chem 283: 25952–25962. doi:10.1074/jbc.M802138200.

78. Milner CM, Higman VA, Day AJ (2006) TSG-6: a pluripotent inflammatory mediator? Biochem Soc Trans 34: 446–450. doi:10.1042/BST0340446.

79. Kodithuwakku SP, Pang RTK, Ng EHY, Cheung ANY, Horne AW, et al. (2012) Wnt activation downregulates olfactomedin-1 in Fallopian tubal epithelial cells: a microenvironment predisposed to tubal ectopic pregnancy. Lab Invest 92: 256–264. doi:10.1038/labinvest.2011.148.

80. Uren A, Wolf V, Sun Y-F, Azari A, Rubin JS, et al. (2004) Wnt/Frizzled signaling in Ewing sarcoma. Pediatr Blood Cancer 43: 243–249. doi:10.1002/pbc.20124.

81. Geryk-Hall M, Hughes DPM (2009) Critical signaling pathways in bone sarcoma: candidates for therapeutic interventions. Curr Oncol Rep 11: 446–453.

82. Gupta A, Verma A, Mishra AK, Wadhwa G, Sharma SK, et al. (2013) The wnt pathway: emerging anticancer strategies. Recent Pat Endocr Metab Immune Drug Discov 7: 138–147.