Genome **Biology**

Integrated analysis of recurrent properties of cancer genes to identify novel drivers

Matteo D'Antonio and Francesca D Ciccarelli[*]

Abstract

The heterogeneity of cancer genomes in terms of acquired mutations complicates the identification of genes whose modification may exert a driver role in tumorigenesis. In this study, we present a novel method that integrates expression profiles, mutation effects, and systemic properties of mutated genes to identify novel cancer drivers. We applied our method to ovarian cancer samples and were able to identify putative drivers in the majority of carcinomas without mutations in known cancer genes, thus suggesting that it can be used as a complementary approach to find rare driver mutations that cannot be detected using frequency-based approaches.

Keywords: Driver mutations, cancer genetic heterogeneity, interaction network, gene duplication, gene origin, gene expression

Background

In recent years, the completion of dozens of high-throughput sequencing screenings of cancer genomes led to the identification of >10,000 genes that bear at least one non-synonymous mutation. The discovery of such a wealth of mutations that progressively accumulate in the cancer genome was to some extent surprising and substantiated the idea of tumours as evolutionary systems where most acquired variations are 'passenger' because they do not have any direct role in promoting cancer. These mutations are fixed in the cancer cell population owing to the presence in the same cells of 'driver' mutations that instead confer growth advantages [1]. The identification of the (few) driver mutations among the (many) passenger variants is therefore key to pinpoint genes and pathways that play an active role in cancer development and may be used as therapeutic targets. Unfortunately, the distinction between driver and passenger mutations is not straightforward, because of the high heterogeneity of the mutational landscape among and within cancer types [2]. One of the most widely used approaches to identify novel cancer genes (that is, genes that harbour driver mutations) measures the gene mutation frequency, relying on the assumption

that genes that are important for the development of a certain cancer type are recurrently mutated in several tumours [2-17]. Frequency-based methods led to the detection of unexpectedly high mutation frequency of isocitrate dehydrogenases 1 and 2, eventually linking these enzymes to the onset of leukaemia and glioma [12,18]. They also contributed to better understand the genetic heterogeneity of cancer, leading to the observation that only few genes are mutated in the vast majority of tumour types, while most cancer genes are mutated at high frequency in one or few cancer types [19]. Also the analysis of pathways instead of genes contributed to reduce the heterogeneity of cancer mutational landscape, because often the de-regulation of cancer-associated pathways can occur through the mutations of different components [20]. Pathway analysis for example identified significant enrichment of mutations in *BRCA1* and *ATM* pathways in breast cancer, and *WNT* and *TGFβ* signalling pathways in colorectal cancer [21]. Although these processes were already known to be involved in tumorigenesis [20], only a systematic approach led to assign a likely driver role to new pathway components. Following conceptually similar approaches, several groups have analysed the proteins encoded by cancer genes in the context of the human protein-protein interaction network and identified network modules that are significantly associated

* Correspondence: francesca.ciccarelli@ieo.eu
Department of Experimental Oncology, European Institute of Oncology, IFOM-IEO Campus, Via Adamello 16, 20139 Milan, Italy

with mutations [22-24]. Network analysis showed that cancer genes encode proteins that are highly connected and central inside the network [25,26]. This has been interpreted as a sign of fragility of cancer genes towards perturbations, because modifications of proteins at the crossroad of multiple biological processes are likely to have harmful consequences [27]. In addition to encoding highly connected and central proteins, cancer genes share also other systems-level properties (that is, global properties that do not strictly depend on the individual gene function) that distinguish them from the rest of human genes. For example, they tend to maintain only one single copy in the genome, which suggests an intrinsic sensitivity of cancer genes towards gene dosage imbalance [26]. Moreover, cancer genes mostly appeared at two time points in evolution: caretakers and tumour suppressors are ancient genes that have orthologs also in prokaryotes, while gatekeepers and oncogenes were acquired with metazoans [28]. This suggests that tumorigenesis may arise from the impairment of either very basic or regulatory processes [29]. The existence of properties that distinguish cancer genes from the rest of human genes may be used to discriminate between driver and passenger mutations because mutated genes that have properties similar to known cancer genes are, in principle, more likely to harbour driver mutations, particularly when the mutations alter the protein function. In the last years, several methods to predict damaging mutations have been developed taking into account the site conservation throughout evolution and the possible effects on protein structure, as well as on splice-sites and UTRs [30-34]. In this study we developed an integrative method that uses tumour, gene and mutation properties to eventually predict novel drivers. As a proof of principle we applied our selection procedure to a panel of >300 ovarian carcinoma patients and identified genes with a putative driver role in >70% of tumours with previously unknown genetic determinants.

Results
The mutational landscape is cancer-specific and recurrently mutated genes are long
We collected 10,681 human genes with at least one non-synonymous mutation from 39 high-throughput mutational screenings conducted in 3,052 cancer samples and 20 cancer types [2,4-18,35-57]. We divided these mutated genes into three groups: (1) 444 known cancer genes that are part of the Cancer Gene Census, a literature-based collection of genes that play experimentally-proven driver roles in cancer development [58,59]; (2) 608 candidate cancer genes that are likely to play a driver role (see Additional file 2, Table S1 for the definition of candidates in each study); and (3) 9,629 genes with no evidence of active involvement in cancer (Table 1 and Additional

file 2, Table S1). As already reported [2,27], we confirmed the heterogeneity of cancer mutational landscape and the overall tendency of genes to be mutated only in few cancer types (Figure 1A) and samples (Figure 1B). In particular, 40% of genes with no evidence of involvement in cancer are mutated in only one cancer type or sample, and <10% recur in more than four cancer types or samples (Figure 1A, B). This indicates the likely enrichment of these genes in passenger mutations. Similarly, the observed tendency of candidates to mutate in several samples (Figure 1B) likely reflects the frequency-based methods that were used to identify them (Additional file 2, Table S1).

Next, we checked whether gene length might influence the recurrent mutations of the same gene in multiple tumours, since longer genes are likely to host a higher number of mutations. Indeed we found positive correlation between the tendency of a gene to be recurrently mutated and its length, particularly in the case of mutated genes with no evidence of cancer involvement, but surprisingly also for candidates (Figure 1C). In both groups the vast majority of genes that are mutated in >10 cancer types have a coding portion longer than 4,450 bp (top 5% of the longest human genes). As a comparison, only five known cancer genes that are mutated in >10 cancer types (*NF1*, *EP300*, *BRCA2*, *MLL*, *ARID1A*) are longer than 4,450 bp. Although a positive correlation between gene length and the number of mutations was expected for genes harbouring passenger mutations, the fact that it was observed also for candidates, but not for known cancer genes, show that current methods do not completely correct for this effect.

Our survey of cancer somatic mutations confirmed that most of them are cancer- and sample-specific. Furthermore, gene length influences the recurrence of mutations and it should be taken into account when selecting candidates only on the basis of gene mutation frequency.

The majority of mutated genes are tissue-selective and lowly expressed
Indirect pieces of evidence have recently shown that gene expression may be useful for discriminating between driver and passenger mutations. For example, mutations of expressed genes in lung carcinomas are overall negatively selected, while the mutation rate of non-expressed genes is similar to the genome-wide average [43]. Based on this observation we reasoned that mutations affecting coding sequences are more likely to exert their function if the gene is expressed. To check whether this is true we investigated the breadth of expression (that is, the number of tissues where a gene is expressed) of mutated genes in a panel of 109 healthy human tissues. Overall we found that known cancer

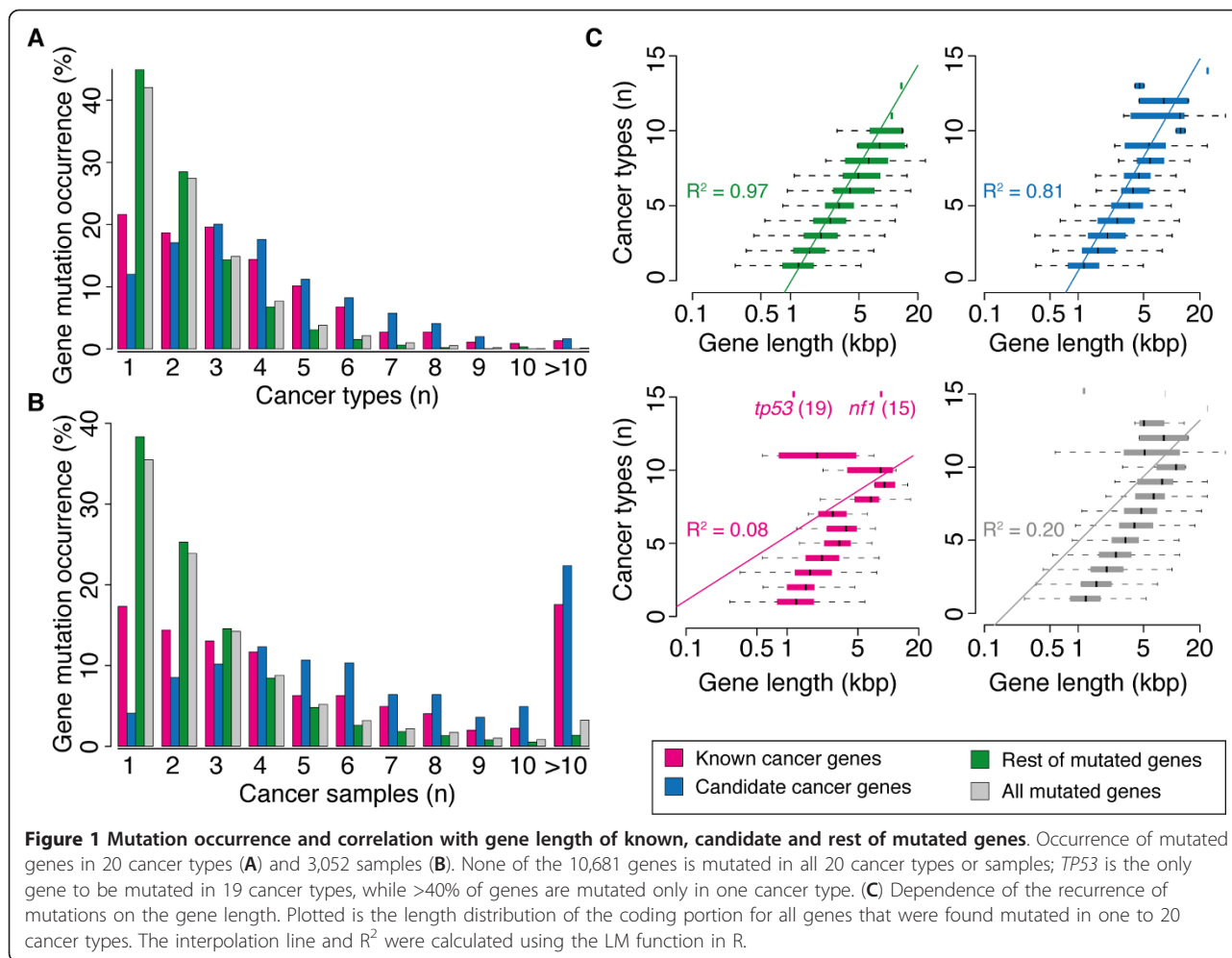**Table 1 Dataset of known, candidates and mutated cancer genes**

Cancer type	Known cancer genes	Candidate cancer genes	Rest of mutated genes	Tumour samples	References
Bladder	24	9	380	97	[8]
Breast	165	280	1,520	196	[2,11,40,52]
Colorectal	42	124	654	11	[2]
Gastric	203	14	5,160	109	[16]
Glioblastoma	127	87	1,679	197	[5,12,39]
HNSCC	221	67	5,901	194	[15,35]
Kidney	71	15	655	517	[7,9,55]
Leukaemia	58	169	877	393	[18,41,44,46,50,51]
Liver	26	66	370	140	[45,53]
Lung	218	143	1,012	324	[4,11,43,49]
Lymphoma	46	47	574	133	[14,47]
Medulloblastoma	16	1	194	88	[13]
Melanoma	96	159	2,845	16	[17,48,54]
Myelodysplasia	22	4	206	29	[57]
Myeloma	69	6	1,260	37	[6]
Oligodendroglioma	12	0	149	34	[38]
Ovarian	60	80	109	58	[11]
Pancreas	77	194	1,070	205	[10,11,42,56]
Prostate	54	161	81	67	[11,37]
Sarcoma	11	10	0	207	[36]
Total	**444**	**608**	**9,629**	**3,052**	NA

Starting from 10,681 human genes with at least one cancer-specific non-synonymous mutation, three groups were identified: known cancer genes that are part of the cancer gene census [58]; candidate genes that are likely to play a role in cancer because they are recurrently mutated; and the rest of mutated genes that carry at least one non-synonymous mutation. The full description of mutated genes in each of the 39 sequencing screenings is reported in Additional file 2, Table S1. HNSCC; head and neck squamous cell carcinoma.

genes are expressed in a significantly higher number of tissues than non-mutated human genes, while candidates and other mutated genes show narrow expression breadth (Figure 2A, Additional file 1, Figure S1, Wilcoxon test). Moreover, known cancer genes are significantly depleted in tissue selective genes (that is, genes expressed in <25% of the total, Fisher's exact test), while candidates and other mutated genes are significantly depleted in housekeeping genes (that is, genes expressed in at least 98% of the total, Figure 2B, Fisher's exact test). These results confirm that known cancer genes are housekeeping and broadly expressed.

We further investigated whether mutated genes are expressed in the same tissues where they are mutated. Unfortunately, such a direct comparison was possible only for three studies that had both mutation and expression data on the same samples, including the whole genomes of acute myeloid leukaemia [44] and primary lung tumour [43], and the mutational screenings of 722 protein-coding genes in 207 sarcoma samples [36]. In all three studies we found a clear distinction between known cancer genes, which are expressed in higher fraction than the rest of human genes, and other mutated genes, which instead are expressed in lower fraction (Figure 2C, Additional file 2,

Table S2, chi squared test). To confirm that this is a general trend in all 20 cancer types with available mutation data, we checked for the expression of mutated genes in the corresponding healthy counterparts (Additional file 2, Table S3). We found that in the normal tissues corresponding to 18 of the 20 cancer types, the fraction of expressed known cancer genes is higher than the rest of expressed human genes, and in 14 cases this difference is statistically significant (Figure 2D and Additional file 2, Table S4, chi squared test). The majority of both candidates and other mutated genes are instead not expressed in the tissues where they were found mutated (Figure 2D, Additional file 1, Figure S2A, chi-squared test). The only significant exception were candidate cancer genes in myeloma, which were expressed in higher fraction than the rest of human genes, probably also because of an overall low expression of human genes in blood and bone marrow (Additional file 2, Table S4). Interestingly, even when mutated genes are expressed, their expression levels are lower than the median expression of non-mutated genes in the same tissues, while known or candidate cancer genes are expressed at levels comparable with the overall tissue median (Figure 2E, Additional file 1, Figure S2B, and Additional file 2, Table S5, chi-squared test).

Figure 1 Mutation occurrence and correlation with gene length of known, candidate and rest of mutated genes. Occurrence of mutated genes in 20 cancer types (**A**) and 3,052 samples (**B**). None of the 10,681 genes is mutated in all 20 cancer types or samples; *TP53* is the only gene to be mutated in 19 cancer types, while >40% of genes are mutated only in one cancer type. (**C**) Dependence of the recurrence of mutations on the gene length. Plotted is the length distribution of the coding portion for all genes that were found mutated in one to 20 cancer types. The interpolation line and R^2 were calculated using the LM function in R.
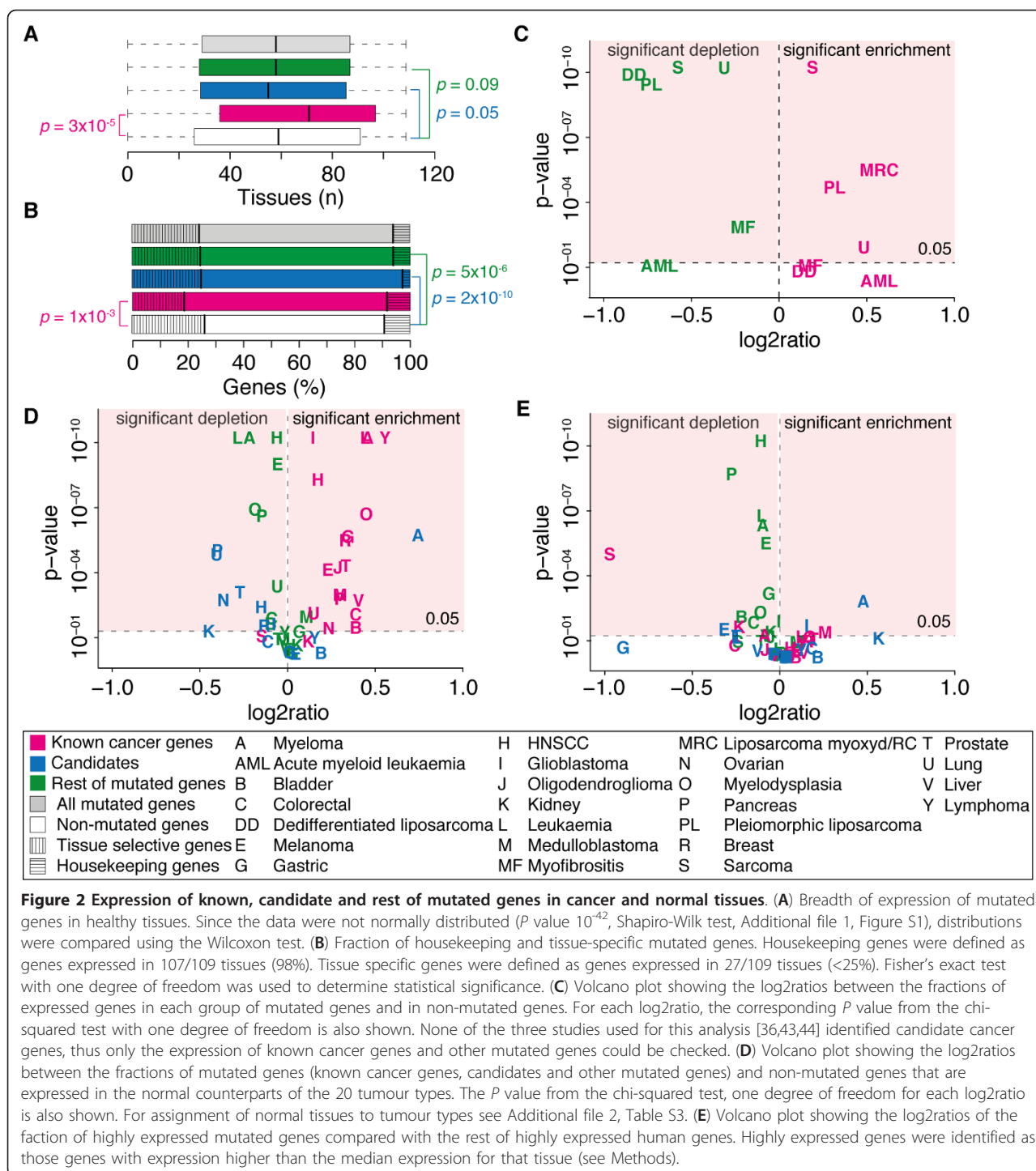
Altogether, these data showed that cancer genes with driver mutations tend to be expressed in the tissue where they are mutated, while genes likely harbouring passenger mutations are generally not expressed. Expression can be therefore used as a further filter to distinguish passenger from driver mutations. Although this might be expected, so far gene expression has not been thoroughly exploited for identifying driver mutations and only a small fraction of published re-sequencing screenings of cancer genomes takes it into account to directly discriminate between driver and passenger mutations [43,60] or to assess the background mutation rate [61].

Identification of novel drivers in ovarian carcinomas
To identify novel cancer genes from mutation data, we developed an integrated pipeline that identifies putative drivers on the basis of the similarity between their properties and those of known cancer genes (Figure 3A). The starting point were cancer samples that underwent both sequencing and expression profiling, since we found that
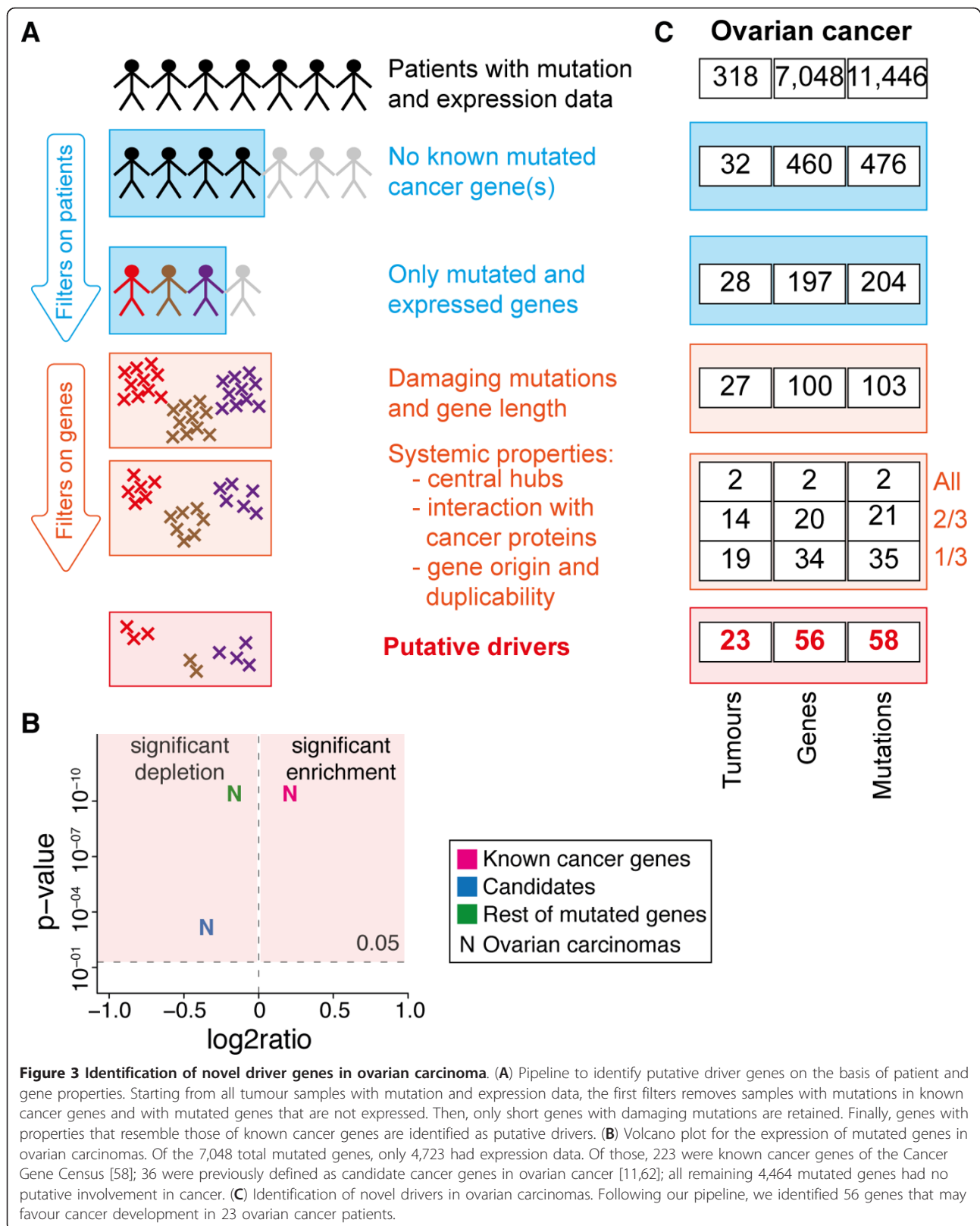
driver mutations occur in genes that are also expressed in the cancer tissue. As a first filter, we removed tumours with at least one known mutated and expressed cancer gene, because these genes are the most likely, albeit not the only, cancer drivers in these tumours. Since our main purpose was to prioritize the selection, we reasoned that it was more likely to find novel drivers in tumours with no mutations in known cancer genes. Further filters were then applied at the gene level. First, mutations were analysed for their putative effects on the encoded proteins, in order to eliminate passenger mutations with no functional consequences. Second, since a positive correlation between gene length and gene mutation frequency exists (Figure 1C), all genes in the top 5% of gene length (>4,450 bp) and mutated in more than five different cancer types (Figure 1A) were removed. Finally, four systems-level properties were evaluated to prioritize genes that resemble known cancer genes. We considered in particular high connectivity and centrality of the protein products in the human protein-protein interaction network [25,26]; direct interaction with a known cancer

Figure 2 Expression of known, candidate and rest of mutated genes in cancer and normal tissues. (**A**) Breadth of expression of mutated genes in healthy tissues. Since the data were not normally distributed (*P* value 10^{-42}, Shapiro-Wilk test, Additional file 1, Figure S1), distributions were compared using the Wilcoxon test. (**B**) Fraction of housekeeping and tissue-specific mutated genes. Housekeeping genes were defined as genes expressed in 107/109 tissues (98%). Tissue specific genes were defined as genes expressed in 27/109 tissues (<25%). Fisher's exact test with one degree of freedom was used to determine statistical significance. (**C**) Volcano plot showing the log2ratios between the fractions of expressed genes in each group of mutated genes and in non-mutated genes. For each log2ratio, the corresponding *P* value from the chi-squared test with one degree of freedom is also shown. None of the three studies used for this analysis [36,43,44] identified candidate cancer genes, thus only the expression of known cancer genes and other mutated genes could be checked. (**D**) Volcano plot showing the log2ratios between the fractions of mutated genes (known cancer genes, candidates and other mutated genes) and non-mutated genes that are expressed in the normal counterparts of the 20 tumour types. The *P* value from the chi-squared test, one degree of freedom for each log2ratio is also shown. For assignment of normal tissues to tumour types see Additional file 2, Table S3. (**E**) Volcano plot showing the log2ratios of the faction of highly expressed mutated genes compared with the rest of highly expressed human genes. Highly expressed genes were identified as those genes with expression higher than the median expression for that tissue (see Methods).

protein [20]; gene evolutionary appearance and duplicability. In the latter case, we prioritized genes that originated either early in evolution or with metazoans and vertebrates [29].

We applied our pipeline to 318 ovarian carcinomas with available sequencing and expression data that could be obtained from the Cancer Genome Atlas and used with no restrictions. Furthermore, all ovarian carcinomas underwent whole exome sequencing and matched expression profiling [62], therefore they constituted the ideal cases for our analysis. Before applying the pipeline for the selection of new drivers, we confirmed that also for this

Figure 3 Identification of novel driver genes in ovarian carcinoma. (**A**) Pipeline to identify putative driver genes on the basis of patient and gene properties. Starting from all tumour samples with mutation and expression data, the first filters removes samples with mutations in known cancer genes and with mutated genes that are not expressed. Then, only short genes with damaging mutations are retained. Finally, genes with properties that resemble those of known cancer genes are identified as putative drivers. (**B**) Volcano plot for the expression of mutated genes in ovarian carcinomas. Of the 7,048 total mutated genes, only 4,723 had expression data. Of those, 223 were known cancer genes of the Cancer Gene Census [58]; 36 were previously defined as candidate cancer genes in ovarian cancer [11,62]; all remaining 4,464 mutated genes had no putative involvement in cancer. (**C**) Identification of novel drivers in ovarian carcinomas. Following our pipeline, we identified 56 genes that may favour cancer development in 23 ovarian cancer patients.

set of patients, similarly to other cancer types (Figure 2C), known cancer genes tend to be expressed in the tumour where they are mutated, while the rest of mutated genes are poorly expressed (Figure 3B and Additional file 1, Figure S3). The vast majority of the analysed ovarian carcinomas (286/318, 90% of the total) had at least one known cancer gene (mostly *TP53*) that was mutated and expressed and were therefore discarded from further analysis. After applying all other filters, we identified 58 putative driver mutations in 56 genes that were mutated and expressed in 23 of the 32 ovarian carcinomas with previously unknown genetics determinants (72%, Figure 3C).

To test the performance of our method in detecting known cancer drivers, we applied it to 130 of the 318 ovarian carcinomas that had mutations in 31 known tumour suppressor genes (Additional file 1, Figure S4). We correctly identified the mutated tumour suppressor genes as the cancer drivers in almost all tumours (123 out of 130 Additional file 2, Table S6). Furthermore, in the same samples we also identified additional putative drivers that are known to co-operate in tumour development. For example, in tumours where we found *TP53* mutations, we also identified genes such as *CDH1* and *CDKN2C* that often co-mutate with *TP53* and are known to have synergic tumour-suppressor activity [63-66]. Therefore, in addition to pinpoint novel drivers, our method could also be applied to search for second hits or co-operating genes that help tumour development. In this respect one interesting putative co-driver is *NUMB*, a gene that encodes a negative regulator of NOTCH [67] and prevents TP53 ubiquitination and degradation [68]. The functional impairment of this gene upon damaging mutation might thus enhance tumour development because of the activation of the *NOTCH* oncogene and the degradation of *TP53* tumour suppressor.

Novel drivers of ovarian cancer resemble tumour suppressors and affect gene transcription, cell proliferation and survival

We had several indications that the mutated genes that we identified as putative drivers might indeed play an active role in ovarian carcinogenesis.

First, in addition to being all predicted as damaging by at least two out of three predictors (see Methods), 60% of the 58 mutations either modified protein functional domains or removed >50% of the protein sequence (Additional file 2, Table S7). Furthermore, the vast majority (77%) of the genomic sites where the mutation occurred are highly conserved among vertebrates (MultiZ score >0.95) [69] (Additional file 2, Table S7). Both these observations suggest a likely functional role of the mutations.
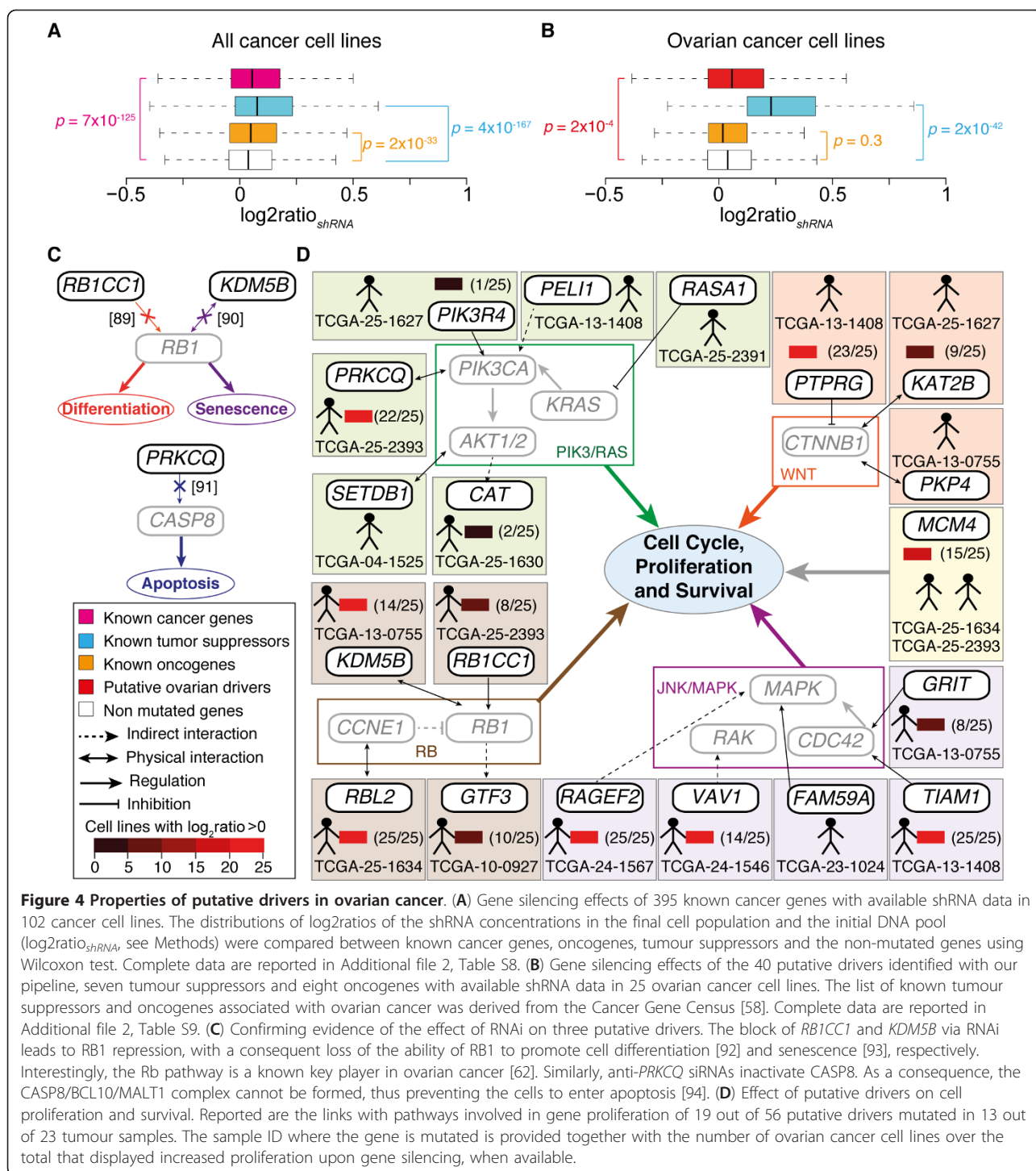
Second, we measured the effect of silencing the putative driver genes via RNA interference (RNAi), which mimics the effect of loss-of-function mutations and can therefore be used to infer the effect of gene impairment in cancer [70]. We derived large-scale gene silencing data from short hairpin RNA (shRNA) screens of approximately 11,000 genes in 102 cancer cell lines [71]. To check whether our assumption of an overall increased cell proliferation upon impairment of genes harbouring driver mutations was correct, we compared the gene silencing effect of known cancer genes with that of the rest of non-mutated human genes in all cell lines (see Methods). As expected, we observed that the silencing of known cancer genes, and in particular of tumour suppressors, favoured cell growth significantly more than non-mutated genes (Figure 4A, Additional file 1, Figure S5 and Additional file 2, Table S8, Wilcoxon test). We then analysed the silencing effect of the putative driver genes identified with our pipeline in the 25 ovarian cancer cell lines used in the screen [71]. Out of the 56 predicted driver genes, 40 were screened via RNAi and 35 of them led to increased cell proliferation in at least one ovarian cancer cell line (Table 2). Furthermore, their silencing effect overall resembled that of known tumour suppressors on the same ovarian cell lines (Figure 4B and Additional file 2, Table S9, Wilcoxon test). Thus, as expected, our selection procedure mainly identified tumour suppressor genes, since we retained putative damaging mutations that disrupt the protein function (Additional file 2, Table S7). For at least three of these genes (*RBICCI*, *KDM5B*, *PRKCQ*) we also found direct literature support that confirmed the effect of their impairment (Figure 4C). Interestingly, all three genes are strong candidate drivers of ovarian cancer (see below).

Finally, we investigated the association of the 56 putative driver genes with pathways known to be involved in ovarian cancer onset. We found that 13 of the 23 tumours (57% of the total) harboured mutations in 19 genes belonging to pathways that control cell proliferation and survival, including the RB and PI3K/RAS signalling pathways, which are altered in 67% and 45% of ovarian cancers, respectively [62] (Figure 4D). RNAi data were available for 14 out of these 19 genes and in all cases gene silencing led to increased proliferation in at least one ovarian cancer cell line and the block of eight genes (*KDM5B*, *TIAM1*, *RAGEF2*, *PRKCQ*, *VAV1*, *PTPRG*, *RBL2* and *MCM4*) favoured cell growth in the majority of cell lines (Figure 4D). Although for the remaining 10 tumours no such a direct link with ovarian cancer could be drawn, six of them had alterations in gene transcription and in other two cases a general association with cancer could be made (Table 2). Therefore, overall >90% of tumours harboured genomic alterations in pathways associated with cancer.

Discussion

The central tenet of our study was that cancer driver mutations occur in genes with peculiar properties and, therefore, such properties can be used to identify novel

Figure 4 Properties of putative drivers in ovarian cancer. (**A**) Gene silencing effects of 395 known cancer genes with available shRNA data in 102 cancer cell lines. The distributions of log2ratios of the shRNA concentrations in the final cell population and the initial DNA pool (log2ratio$_{shRNA}$, see Methods) were compared between known cancer genes, oncogenes, tumour suppressors and the non-mutated genes using Wilcoxon test. Complete data are reported in Additional file 2, Table S8. (**B**) Gene silencing effects of the 40 putative drivers identified with our pipeline, seven tumour suppressors and eight oncogenes with available shRNA data in 25 ovarian cancer cell lines. The list of known tumour suppressors and oncogenes associated with ovarian cancer was derived from the Cancer Gene Census [58]. Complete data are reported in Additional file 2, Table S9. (**C**) Confirming evidence of the effect of RNAi on three putative drivers. The block of *RB1CC1* and *KDM5B* via RNAi leads to RB1 repression, with a consequent loss of the ability of RB1 to promote cell differentiation [92] and senescence [93], respectively. Interestingly, the Rb pathway is a known key player in ovarian cancer [62]. Similarly, anti-*PRKCQ* siRNAs inactivate CASP8. As a consequence, the CASP8/BCL10/MALT1 complex cannot be formed, thus preventing the cells to enter apoptosis [94]. (**D**) Effect of putative drivers on cell proliferation and survival. Reported are the links with pathways involved in gene proliferation of 19 out of 56 putative drivers mutated in 13 out of 23 tumour samples. The sample ID where the gene is mutated is provided together with the number of ovarian cancer cell lines over the total that displayed increased proliferation upon gene silencing, when available.

cancer genes. For example we showed that cancer genes with an established driver role are usually expressed in the tissue where they are mutated, thus suggesting that mutations in genes that are not expressed are neutral or passenger. In support to our results, the vast majority of cancer somatic mutations have been shown to occur in genomic regions associated with repressive chromatin marks [72].

This indicates that indeed most cancer mutations are neutral and occur in transcription-silent regions of the genome.

In addition to expression profiles, we analysed the evolutionary, genomics and network properties of genes mutated in 32 ovarian cancer carcinomas with previously unknown genetics determinants. These tumours constitute only a small fraction of ovarian carcinomas

Table 2 Putative novel drivers in ovarian cancer

Ovarian carcinoma sample	Gene	Ovarian cell lines with increased proliferation	Pathway/Biological process	Reference
TCGA-04-1525	SETDB1	NA	PI3K/RAS	[95]
	SIPA1L1	NA	Cytoskeleton organisation	[96]
TCGA-09-2051	SPTLC1	1	MYC	[97]
TCGA-10-0927	GTF3	10	RB	[98]
	SQRDL	1	Amino acid metabolism	[99]
TCGA-13-0717	GRIT	NA	JNK/MAPK	[100]
TCGA-13-0755	ANXA2	2	Vesicle transport	[101]
	FBLN1	1	Cell adhesion	[97]
	FEN1	4	DNA repair	[102]
	GSS	3	Glutathione metabolism	[103]
	KDM5B	14	RB	[104]
	MBNL1	3	Splicing regulation	[105]
	PKP4	NA	WNT	[106]
	U2SURP	NA	WNT	[97]
	WDR61	NA	Chromatin regulation	[107]
TCGA-13-1408	F11R	0	Cell adhesion	[108]
	PELI1	NA	PI3K/RAS	[109]
	PTPRG	23	WNT	[110]
	TAF12	18	Transcriptional activation	[111]
	TIAM1	25	JNK/MAPK	[112]
TCGA-13-1506	BTBD2	NA	NF-kB (transcription)	[113]
TCGA-13-1510	RIOK3	7	Unknown	NA
TCGA-23-1024	FAM59A	NA	JNK/MAPK	[114]
	PBX2	22	Transcriptional activation	[115]
TCGA-24-1431	ANKZF1	NA	Unknown	NA
	ATP1B1	0	Ion transport	[116]
TCGA-24-1544	CCDC93	10	Unknown	NA
	HDAC6	1	Chromatin regulation	[117]
TCGA-24-1546	SUGP1	7	Splicing regulation	[99]
	VAV1	14	JNK/MAPK	[118]
TCGA-24-1567	HTATSF1	5	Splicing regulation	[119]
	INADL	NA	Cell-cell interactions	[120]
	KAL1	0	Cell adhesion	[121]
	RAGEF2	25	JNK/MAPK	[122]
TCGA-25-1623	PSD3	NA	Transcriptional activation	[123]
TCGA-25-1627	KAT2B	9	WNT	[124]
	PIK3R4	1	PI3K/RAS	[125]
	TTF2	25	Transcriptional repression	[126]
TCGA-25-1630	APEX1	9	DNA repair	[127]
	ATG3	NA	Autophagy	[128]
	CAT	2	PI3K/RAS	[129]
	DCAF6	NA	Transcriptional activation	[130]
TCGA-25-1633	SASH3	0	Antigen receptor signalling	[131]
TCGA-25-1634	GSN	5	TP53	[132]
	MCM4	15	G1/S transition	[133]
	RBL2	25	RB	[134]
TCGA-25-2391	RASA1	5	PI3K/RAS	[135]

Table 2 Putative novel drivers in ovarian cancer (Continued)

TCGA-25-2393	ANO1	0	Ion transport	[131]
	FUT1	1	Antigen synthesis	[136]
	MCM4	15	G1/S transition	[137]
	PRKCQ	22	PI3K/RAS	[138]
	RB1CC1	8	RB	[121]
	ZC3H14	NA	Translation	[139]
TCGA-36-1570	PSD3	NA	Transcriptional activation	[140]
TCGA-36-1574	NRCAM	NA	Cell adhesion	[141]
	NSMAF	NA	NF-kB (transcription)	[142]
	PASK	24	Translation	[143]
TCGA-36-1577	PRKD1	1	Transcriptional activation	[144]

Shown are 23 ovarian carcinomas hosting 58 putative driver mutations in 56 genes. Two genes (*MCM4* and *PSD3*) were mutated in two tumours. For 40 genes with available RNAi data [71], the number of ovarian cell lines where an increased in cell growth was observed upon gene silencing is shown. Thirteen carcinomas (57% of the total) bear mutations in pathways previously associated with ovarian cancer, and six of them (26% of the total) had alterations in transcription-related pathways. A detailed description of the effects of each mutation is shown in Additional file 2, Table S7.

(approximately 10% of the initial set) since the large majority of affected individuals bear mutations in known cancer genes, in particular in *TP53*. Although cancer is usually the outcome of the alteration of several genes and multiple drivers are required for cancer progression [73], we reasoned that focusing on tumours with no mutation in known cancer genes could increase the chances to find novel drivers. Furthermore, this would also help identifying a possible cause of cancer onset and development also in tumours that harbour rare mutations. With our approach we were indeed able to find 56 putative cancer genes in >70% of previously uncharacterized tumours, thus significantly reducing the fraction of patients with unknown cancer determinants. In the vast majority of cases, at least one of the putative drivers exerts a function in pathways that are altered in ovarian cancer. This confirms that the high heterogeneity of the cancer mutational landscape is reduced when considering biological processes rather than single genes [19].

As a comparison with our method, we investigated whether the 56 putative cancer genes had also been detected in the original study on the same set of ovarian carcinomas [62], which also identified possible cancer genes using a variety of approaches, from gene mutation frequency to pathway and network analysis [62]. Our list of putative drivers showed very poor overlap with the genes identified in the original study, mainly because the latter were for the vast majority already known cancer genes or had no expression data, and were therefore discarded from our analysis. Interestingly, some overlap existed between our list of 56 drivers and the network modules that were significantly mutated in ovarian cancer [24]. In particular, we identified five genes in common between the two lists. The silencing via RNA interference of three of these five genes (*VAV1*, *TAF12* and *GTF3*) resulted in increased proliferation in at least 10 ovarian

cancer cell lines. This strongly suggests a role of tumour suppression of these genes, and this is worth further experimental investigation.

Conclusions

Our analysis showed that the integration of several sources of information allows the identification of rare cancer genes. This may be of particular utility in tumours with no known driver mutations or where frequency-based methods cannot be applied. However, we also showed that an integrated analysis may be useful for the identification of mutated genes that may cooperate in promoting tumour development. The poor overlap with previous findings in the same set of tumour samples demonstrates that our approach is complementary to frequency-based methods. The integration of several methods based upon different theoretical assumptions may therefore result in a better and more complete characterization of the mutational landscape of cancer.

Methods
Gene sets used in the analysis

To derive a dataset of unique human genes (that is, genes with a unique locus in the genome), 33,398 protein sequences were retrieved from RefSeq v.51 [74] and aligned to the human reference genome (hg19) using BLAT [75]. In case of multiple isoforms aligning to the same locus, only the longest was retained [26]. Only genes located on autosomal chromosomes and chromosome × were considered for further analysis, for a total of 19,009 unique human genes. Gene length was calculated as the coding portion of the longest isoform for each locus.

The dataset of 10,681 genes with at least one somatic non-synonymous mutation in cancer was collected from 39 mutational screenings of cancer tissues [2,4-18,35-57] (Table 1, Additional file 2, Table S1). Genes were grouped

into three classes: (1) known cancer genes included all genes whose mutations or amplifications are known to be involved in tumorigenesis (Cancer Gene Census, frozen on 15 November 2011, and Census of Amplified Genes in Cancer) [58,59]; (2) candidate cancer genes that were found recurrently mutated in different tumour samples and, therefore, likely to harbour driver mutations (candidates were extracted directly from the corresponding experiments, Additional file 2, Table S1); (3) genes with low frequency non-synonymous mutations. The rest of human genes used for comparison were defined as all human genes with either no mutations or only synonymous mutations (Table 1).

Expression of mutated genes in normal and cancer tissues

Expression data for 12,397 genes in 109 healthy tissues were derived from two microarray experiments on 36 [76] and 73 [77] normal human tissues, respectively, for a total of 109 unique tissues. The raw CEL files were downloaded from the corresponding series (GSE2361 and GSE1133) stored in the Gene Expression Omnibus (GEO) [78], normalised and analysed using the MAS5 algorithm included in the R affy package [79,80] (Additional file 2, Table S10). Given that more than one probe could be associated with a single gene, a gene was labelled as 'expressed' if at least half of the corresponding probes had detection P values <0.05. Housekeeping genes were defined as genes expressed in at least 98% of the tissues (107/109), while tissue-specific genes were expressed in <25% of the tissues (27/109).

To test whether the fraction of housekeeping mutated genes (known, candidates and rest of genes with non-synonymous mutations) was different from the fraction of housekeeping genes among the rest of human genes, Fisher's exact test with one degree of freedom was used. Fisher's test was used because of the small number of genes that were compared (only 10 candidate genes were housekeeping). The same test was applied to assess the differences in the fraction of tissue-specific genes between mutated and non-mutated genes.

To check whether mutated genes tend to be expressed in the corresponding healthy tissue, one or more of the 109 normal tissues with expression data were associated with the 20 tumour types with mutation data (Additional file 2, Table S3). For each of the three groups of mutated genes (known, candidates and rest of genes with non-synonymous mutations), the fraction of expressed genes over the total ($f_{exp\_mutated}$) was calculated in the tissues corresponding to each of the 20 tumour types. Similarly, the fraction of expressed non-mutated human genes in the same tissue (f_{exp_rest}) was also measured and the two proportions were compared using chi-squared test with one degree of freedom to determine whether they were

statistically different. Results were visualised as volcano plots that reported the log2ratios between the two fractions of expressed genes and the corresponding P value as measured with chi-squared test:

$$\log 2 ratio = \log_2 \frac{f_{exp\_mutated}}{f_{exp\_rest}}$$

To verify whether mutated genes were expressed at higher or lower levels than the rest of human genes, the median expression level was calculated in each of the 109 tissues. All genes with expression higher than the median were considered as highly expressed, while all genes with expression lower than the median were defined as lowly expressed. In each tissue, the fraction of highly expressed genes over the total in each of the three groups of mutated genes ($h_{exp\_mutated}$) and the fraction of highly expressed non-mutated genes (h_{exp_rest}) were compared using the chi-squared test with one degree of freedom. Results were displayed as volcano plots that reported the log2ratio between the fractions of highly expressed mutated and non-mutated genes and the corresponding P value assessed with chi-squared test:

$$\log 2 ratio = \log_2 \frac{h_{exp\_mutated}}{h_{exp\_rest}}$$

For three of the 39 mutational screenings [36,43,44], both expression and mutation data were available for each analysed tumour sample. The raw CEL files were downloaded from GEO and the data were processed as described for the normal tissues (Additional file 2, Table S10). Since the study by Barretina *et al.* [36] reported the mutational screen of 722 genes and only a small number of mutations were detected in each sample, tumours were clustered into four groups, on the basis of the tumour subtype (Additional file 2, Table S4). A pipeline similar to that described for the analysis of normal tissues was applied to determine whether higher fraction of cancer genes were expressed in the cancer tissues where they were also mutated. Briefly, the fractions of expressed mutated and non-mutated genes in each tumour sample were compared using chi-squared test with one degree of freedom, in each sample. As for the other analyses, the results were displayed as volcano plots where each log2-ratios of the fractions of expressed genes between mutated genes and non-mutated genes were displayed in association with the corresponding P values of the chi-squared test.

Analysis of ovarian carcinoma samples

Genes mutated in ovarian carcinomas were derived from the Cancer Genome Atlas [81]. In addition to all validated somatic mutations (data level 3), the raw CEL files of the expression data corresponding to the same tumour sample

were also retrieved (platform HG_U133A, data level 1, Additional file 2, Table S10). Of the 323 tumours, five were removed because they did not undergo whole exome sequencing. The fraction of expressed and mutated genes was calculated for each carcinoma as described above, and compared with the corresponding fraction of expressed and non-mutated human genes using the chi-squared test (one degree of freedom). Starting from the list of all mutated genes, several filters were applied to identify putative driver mutations (Figure 3A). First, carcinomas with mutations in at least one known cancer gene from the Cancer Gene Census [58] and those with no expression data for any mutated gene were discarded. Second, three different predictors (SIFT [30], Polyphen [31] and MutationTaster [32]) were applied to infer the effect of mutations. Only frameshift, nonsense and splice-site mutations, as well as missense mutations predicted as damaging by two out the three predictors (SIFT score >0.95, Polyphen score >0.9, or labelled as 'disease causing' by MutationTaster [82]) were retained. Third, the gene length of the coding portion was taken into account and all genes in the bottom 95% of gene length were retained (coding length <4,450 bp). Genes longer than 4,450 bp were retained only if mutated in less than five different cancer types. This filter discarded genes that mutate at high frequency because of their length. Finally four systemic properties were investigated: protein connectivity and centrality in the protein-protein interaction network; interaction(s) with known cancer proteins; evolutionary origin; and duplicability. To measure protein connectivity and to determine the occurrence of direct interactions with known cancer proteins, data on 98,492 experimentally proven protein-protein interactions between 13,531 human proteins were integrated from five databases (HPRD [83], BioGRID [84], IntAct [85], MINT [86] and DIP [87]), as previously described [88]. The *IGRAPH* module for R [89,90] was used to measure degree, betweenness and direct interactions with known cancer proteins. Central hubs were defined as the 25% most connected (degree >14) and most central (betweenness >9,198) proteins of the network. Evolutionary origin and gene duplicability were defined as previously described [29]. Briefly, gene origin was traced as the most ancient node of the tree of life where orthologs for a given human gene could be found. A gene was defined as duplicated if at least one human paralog was present in the corresponding cluster of orthologs, otherwise it was considered as singleton. All scripts used to run this pipeline are available as Additional file 3.

Effect of gene silencing on cell proliferation using RNA interference

Short hairpin RNA (shRNA) data were derived from the high throughput analysis on 10,941 genes (corresponding to 52,209 probes) in 102 cancer cell lines (including 25 ovarian cancer cell lines) and analysed as described in the

original study [71], with slight modifications. Briefly, the raw GCT file with the measurements of the shRNA abundance in all cell lines (20110303_achilles2.gct) was downloaded and normalized to obtain the corresponding shRNA score for each gene probe. The effect of the individual gene silencing on cell proliferation was calculated in comparison with the initial DNA pool, using an in-house modified version of the R *shRNAscores* package from the Integrative Genomics Portal at the BROAD Institute [91]. In order to determine the silencing effect of each gene, the concentration of its corresponding shRNA in the final cell population and the initial DNA pool was compared. To have a single comparison for each gene probe *i*, the log2ratio was calculated between the means of all replicates in each cell line and the means of replicates in the initial DNA pool:

$$\log 2ratio_{shRNA.h.i} = \log_2 \left[\frac{\frac{1}{m}\sum_{j=1}^{m} shRNA\_score_{h,i,j}}{\frac{1}{n}\sum_{k=1}^{n} shRNA\_score_{DNA,i,k}} \right]$$

Where *m* and *n* are the number of replicates in the considered cell line *h* and in the reference DNA pool, respectively. Having a median of five probes associated with a single gene, only the top-scoring shRNA value among all probes was considered as the representative effect of that gene on cell proliferation in order to minimise the false positives [71]. The ratio was preferred to the difference between cell lines and DNA pool (as in the original paper [71]) in order to better appreciate the modifications in the cell proliferation caused by gene silencing. To measure the overall effect on gene proliferation of the silencing of known cancer genes, the log2ratio$_{shRNA}$ distributions between 395 genes (95 tumour suppressors and 300 oncogenes) from the Cancer Gene Census with at least one shRNA probe and the rest of 10,546 non-mutated genes in all 102 cancer cell lines were compared (Figure 4A). Shapiro-Wilk test was applied to control for the shape of the distributions. Since the distribution could not be considered as normal (*P* value <10^{-50}, Additional file 1, Figure S5), Wilcoxon test was used to assess the differences between them. For the analysis on ovarian cancer, only the 25 ovarian cancer cell lines and 15 known cancer genes (seven tumour suppressors and eight oncogenes) that were associated with ovarian cancer in the original annotation of the Cancer Gene Census were considered (Figure 4B).

Additional material

Additional file 1: Supplemental figures. This file contains Figures S1-S5.

Additional file 2: Supplemental tables. This file contains Tables S1-S10.

Additional file 3: Scripts to identify putative drivers. This file contains a collection of scripts to run the pipeline for the identification of cancer drivers.

References
1. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, et al: **Patterns of somatic mutation in human cancer genomes.** *Nature* 2007, **446**:153-158.
2. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, et al: **The genomic landscapes of human breast and colorectal cancers.** *Science* 2007, **318**:1108-1113.
3. Attolini CS, Michor F: **Evolutionary theory of cancer.** *Ann N Y Acad Sci* 2009, **1168**:23-51.
4. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, Wendl MC, Lawrence MS, Larson DE, Chen K, Dooling DJ, Sabo A, Hawes AC, Shen H, Jhangiani SN, Lewis LR, Hall O, Zhu Y, Mathew T, Ren Y, Yao J, Scherer SE, Clerc K, et al: **Somatic mutations affect key pathways in lung adenocarcinoma.** *Nature* 2008, **455**:1069-1075.
5. The Cancer Genome Atlas Research Network: **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**:1061-1068.
6. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet JP, Ahmann GJ, Adli M, Anderson KC, Ardlie KG, Auclair D, Baker A, Bergsagel PL, Bernstein BE, Drier Y, Fonseca R, Gabriel SB, Hofmeister CC, Jagannath S, Jakubowiak AJ, Krishnan A, Levy J, Liefeld T, Lonial S, Mahan S, Mfuko B, Monti S, Perkins LM, et al: **Initial genome sequencing and analysis of multiple myeloma.** *Nature* 2011, **471**:467-472.
7. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, et al: **Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes.** *Nature* 2010, **463**:360-363.
8. Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, Wu R, Chen C, Li X, Zhou L, Zhou L, He M, Li Z, Sun X, Jia W, Chen J, Yang S, Zhou F, Zhao X, Wan S, Ye R, Liang C, Liu Z, Huang P, Liu C, Jiang H, Wang Y, Zheng H, Sun L, Liu X, Jiang Z, et al: **Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder.** *Nat Genet* 2011, **43**:875-878.
9. Guo G, Gui Y, Gao S, Tang A, Hu X, Huang Y, Jia W, Li Z, He M, Sun L, Song P, Sun X, Zhao X, Yang S, Liang C, Wan S, Zhou F, Chen C, Zhu J, Li X, Jian M, Zhou L, Ye R, Huang P, Chen J, Jiang T, Liu X, Wang Y, Zou J, Jiang Z, et al: **Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma.** *Nat Genet* 2011, **44**:17-19.
10. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, et al: **Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.** *Science* 2008, **321**:1801-1806.
11. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, Yue P, Haverty PM, Bourgon R, Zheng J, Moorhead M, Chaudhuri S, Tomsho LP, Peters BA, Pujara K, Cordes S, Davis DP, Carlton VE, Yuan W, Li L, Wang W,

Eigenbrot C, Kaminker JS, Eberhard DA, Waring P, Schuster SC, Modrusan Z, Zhang Z, Stokoe D, de Sauvage FJ, et al: **Diverse somatic mutation patterns and pathway alterations in human cancers.** *Nature* 2010, **466**:869-873.
12. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, et al: **An integrated genomic analysis of human glioblastoma multiforme.** *Science* 2008, **321**:1807-1812.
13. Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, Boca SM, Carter H, Samayoa J, Bettegowda C, Gallia GL, Jallo GI, Binder ZA, Nikolsky Y, Hartigan J, Smith DR, Gerhard DS, Fults DW, VandenBerg S, Berger MS, Marie SK, Shinjo SM, Clara C, Phillips PC, Minturn JE, Biegel JA, Judkins AR, Resnick AC, Storm PB, Curran T, et al: **The genetic landscape of the childhood cancer medulloblastoma.** *Science* 2010, **331**:435-439.
14. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, Wells VA, Grunn A, Messina M, Elliot O, Chan J, Bhagat G, Chadburn A, Gaidano G, Mullighan CG, Rabadan R, Dalla-Favera R: **Analysis of the coding genome of diffuse large B-cell lymphoma.** *Nat Genet* 2011, **43**:830-837.
15. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence M, Sougnez C, McKenna A, Shefler E, Ramos AH, Stojanov P, Carter SL, Voet D, Cortés ML, Auclair D, Berger MF, Saksena G, Guiducci C, Onofrio RC, Parkin M, Romkes M, Weissfeld JL, Seethala RR, Wang L, Rangel-Escareño C, Fernandez-Lopez JC, Hidalgo-Miranda A, Melendez-Zajgla J, et al: **The mutational landscape of head and neck squamous cell carcinoma.** *Science* 2011, **333**:1157-1160.
16. Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, Chan TL, Kan Z, Chan AS, Tsui WY, Lee SP, Ho SL, Chan AK, Cheng GH, Roberts PC, Rejto PA, Gibson NW, Pocalyko DJ, Mao M, Xu J, Leung SY: **Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer.** *Nat Genet* 2011, **43**:1219-1223.
17. Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S, Stemke-Hale K, Davies MA, Gershenwald JE, Robinson W, Robinson S, Rosenberg SA, Samuels Y: **Exome sequencing identifies GRIN2A as frequently mutated in melanoma.** *Nat Genet* 2011, **43**:442-446.
18. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, McGrath SD, McGrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, McMichael JF, et al: **Recurring mutations found by sequencing an acute myeloid leukemia genome.** *N Engl J Med* 2009, **361**:1058-1066.
19. Ding L, Wendl MC, Koboldt DC, Mardis ER: **Analysis of next-generation genomic data in cancer: accomplishments and challenges.** *Hum Mol Genet* 2010, **19**:R188-196.
20. Vogelstein B, Kinzler KW: **Cancer genes and the pathways they control.** *Nat Med* 2004, **10**:789-799.
21. Lin J, Gan CM, Zhang X, Jones S, Sjoblom T, Wood LD, Parsons DW, Papadopoulos N, Kinzler KW, Vogelstein B, Parmigiani G, Velculescu VE: **A multidimensional analysis of genes mutated in breast and colorectal cancers.** *Genome Res* 2007, **17**:1304-1318.
22. Cerami E, Demir E, Schultz N, Taylor BS, Sander C: **Automated network analysis identifies core pathways in glioblastoma.** *PLoS One* 2010, **5**:e8918.
23. Girvan M, Newman ME: **Community structure in social and biological networks.** *Proc Natl Acad Sci USA* 2002, **99**:7821-7826.
24. Vandin F, Upfal E, Raphael BJ: **Algorithms for detecting significantly mutated pathways in cancer.** *J Comput Biol* 2011, **18**:507-522.
25. Jonsson PF, Bates PA: **Global topological features of cancer proteins in the human interactome.** *Bioinformatics* 2006, **22**:2291-2297.
26. Rambaldi D, Giorgi FM, Capuani F, Ciliberto A, Ciccarelli FD: **Low duplicability and network fragility of cancer genes.** *Trends Genet* 2008, **24**:427-430.
27. Ciccarelli FD: **The (r)evolution of cancer genetics.** *BMC Biol* 2010, **8**:74.
28. Domazet-Loso T, Tautz D: **Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa.** *BMC Biol* 2010, **8**:66.
29. D'Antonio M, Ciccarelli FD: **Modification of gene duplicability during the evolution of protein interaction network.** *PLoS Comput Biol* 2011, **7**:e1002029.
30. Kumar P, Henikoff S, Ng PC: **Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.** *Nat Protoc* 2009, **4**:1073-1081.

31. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: **A method and server for predicting damaging missense mutations.** *Nat Methods* 2010, **7**:248-249.

32. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D: **MutationTaster evaluates disease-causing potential of sequence alterations.** *Nat Methods* 2010, **7**:575-576.

33. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R: **Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations.** *Cancer Res* 2009, **69**:6660-6667.

34. Carter H, Samayoa J, Hruban RH, Karchin R: **Prioritization of driver mutations in pancreatic cancer using cancer-specific high-throughput annotation of somatic mutations (CHASM).** *Cancer Biol Ther* 2010, **10**:582-587.

35. Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J, Wang J, Zhang N, El-Naggar AK, Jasser SA, Weinstein JN, Treviño L, Drummond JA, Muzny DM, Wu Y, Wood LD, Hruban RH, Westra WH, Koch WM, Califano JA, Gibbs RA, Sidransky D, Vogelstein B, Velculescu VE, Papadopoulos N, Wheeler DA, Kinzler KW, *et al*: **Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1.** *Science* 2011, **333**:1154-1157.

36. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, Shah K, Socci ND, Weir BA, Ho A, Chiang DY, Reva B, Mermel CH, Getz G, Antipin Y, Beroukhim R, Major JE, Hatton C, Nicoletti R, Hanna M, Sharpe T, Fennell TJ, Cibulskis K, Onofrio RC, Saito T, Shukla N, Lau C, Nelander S, Silver SJ, Sougnez C, *et al*: **Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy.** *Nat Genet* 2010, **42**:715-721.

37. Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K, Sivachenko AY, Sboner A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, Carter SL, Park K, Habegger L, Ambrogio L, Fennell T, Parkin M, Saksena G, Voet D, Ramos AH, Pugh TJ, Wilkinson J, Fisher S, Winckler W, Mahan S, Ardlie K, Baldwin J, Simons JW, Kitabayashi N, MacDonald TY, *et al*: **The genomic complexity of primary human prostate cancer.** *Nature* 2011, **470**:214-220.

38. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, Hruban RH, Rodriguez FJ, Cahill DP, McLendon R, Riggins G, Velculescu VE, Oba-Shinjo SM, Marie SK, Vogelstein B, Bigner D, Yan H, Papadopoulos N, Kinzler KW: **Mutations in CIC and FUBP1 contribute to human oligodendroglioma.** *Science* 2011, **333**:1453-1455.

39. Clark MJ, Homer N, O'Connor BD, Chen Z, Eskin A, Lee H, Merriman B, Nelson SF: **U87MG decoded: the genomic sequence of a cytogenetically aberrant human cancer cell line.** *PLoS Genet* 2010, **6**:e1000832.

40. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, Harris CC, McLellan MD, Fulton RS, Fulton LL, Abbott RM, Hoog J, Dooling DJ, Koboldt DC, Schmidt H, Kalicki J, Zhang Q, Chen L, Lin L, Wendl MC, McMichael JF, Magrini VJ, Cook L, McGrath SD, Vickery TL, Appelbaum E, Deschryver K, Davies S, Guintoli T, Lin L, *et al*: **Genome remodelling in a basal-like breast cancer metastasis and xenograft.** *Nature* 2010, **464**:999-1005.

41. Greif PA, Eck SH, Konstandin NP, Benet-Pages A, Ksienzyk B, Dufour A, Vetter AT, Popp HD, Lorenz-Depiereux B, Meitinger T, Bohlander SK, Strom TM: **Identification of recurring tumor-specific somatic mutations in acute myeloid leukemia by transcriptome sequencing.** *Leukemia* 2011, **25**:821-827.

42. Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, Schulick RD, Tang LH, Wolfgang CL, Choti MA, Velculescu VE, Diaz LA Jr, Vogelstein B, Kinzler KW, Hruban RH, Papadopoulos N: **DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors.** *Science* 2011, **331**:1199-1203.

43. Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, Ha C, Johnson S, Kennemer MI, Mohan S, Nazarenko I, Watanabe C, Sparks AB, Shames DS, Gentleman R, de Sauvage FJ, Stern H, Pandita A, Ballinger DG, Drmanac R, Modrusan Z, Seshagiri S, Zhang Z: **The mutation spectrum revealed by paired genome sequences from a lung cancer patient.** *Nature* 2010, **465**:473-477.

44. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, *et al*: **DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome.** *Nature* 2008, **456**:66-72.

45. Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA, Pawlik TM, Daniel HD, Kannangai R, Offerhaus GJ, Velculescu VE, Wang L, Zhou S, Vogelstein B, Hruban RH, Papadopoulos N, Cai J, Torbenson MS, Kinzler KW: **Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma.** *Nat Genet* 2011, **43**:828-829.

46. Lilljebjorn H, Rissler M, Lassen C, Heldrup J, Behrendtz M, Mitelman F, Johansson B, Fioretos T: **Whole-exome sequencing of pediatric acute lymphoblastic leukemia.** *Leukemia* 2011, **26**:1602-1607.

47. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, *et al*: **Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma.** *Nature* 2011, **476**:298-303.

48. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordonez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, *et al*: **A comprehensive catalogue of somatic mutations from a human cancer genome.** *Nature* 2010, **463**:191-196.

49. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordoñez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, *et al*: **A small-cell lung cancer genome with complex signatures of tobacco exposure.** *Nature* 2010, **463**:184-190.

50. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, Escaramis G, Jares P, Bea S, Gonzalez-Diaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JM, López C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernández JM, Puente DA, Freije JM, Velasco G, Gutiérrez-Fernández A, Costa D, Carrió A, Guijarro S, Enjuanes A, *et al*: **Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia.** *Nature* 2011, **475**:101-105.

51. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente DA, Velasco G, Freije JM, Tubío JM, Royo R, Gelpí JL, Orozco M, Pisano DG, Zamora J, Vázquez M, *et al*: **Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia.** *Nat Genet* 2012, **44**:47-52.

52. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S: **Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution.** *Nature* 2009, **461**:809-813.

53. Totoki Y, Tatsuno K, Yamamoto M, Arai Y, Hosoda F, Ishikawa S, Tsutsumi S, Sonoda K, Totsuka H, Shirakihara T, Sakamoto H, Wang L, Ojima H, Shimada K, Kosuge T, Okusaka T, Kato K, Kusuda J, Yoshida T, Aburatani H, Shibata T: **High-resolution characterization of a hepatocellular carcinoma genome.** *Nat Genet* 2011, **43**:464-469.

54. Turajlic S, Furney SJ, Lambros MB, Mitsopoulos C, Kozarewa I, Geyer FC, Mackay A, Hakas J, Zvelebil M, Lord CJ, Ashworth A, Thomas M, Stamp G, Larkin J, Reis-Filho JS, Marais R: **Whole genome sequencing of matched primary and metastatic acral melanomas.** *Genome Res* 2011, **22**:196-207.

55. Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LF, Richard S, Kahnoski RJ, Anema J, *et al*: **Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma.** *Nature* 2011, **469**:539-542.

56. Wang L, Tsutsumi S, Kawaguchi T, Nagasaki K, Tatsuno K, Yamamoto S, Sang F, Sonoda K, Sugawara M, Saiura A, Hirono S, Yamaue H, Miki Y, Isomura M, Totoki Y, Nagae G, Isagawa T, Ueda H, Murayama-Hosokawa S, Shibata T, Sakamoto H, Kanai Y, Kaneda A, Noda T, Aburatani H: **Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency.** *Genome Res* 2011, **22**:208-219.

57. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M,

Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, *et al*: Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 2011, **478**:64-69.

58. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: A census of human cancer genes. *Nat Rev Cancer* 2004, **4**:177-183.

59. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS: A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 2010, **10**:59-64.

60. The Cancer Genome Atlas Research Network: Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012, **487**:330-337.

61. The Cancer Genome Atlas Research Network: Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012, **489**:519-525.

62. The Cancer Genome Atlas Research Network: Integrated genomic analyses of ovarian carcinoma. *Nature* 2011, **474**:609-615.

63. Shimada S, Mimata A, Sekine M, Mogushi K, Akiyama Y, Fukamachi H, Jonkers J, Tanaka H, Eishi Y, Yuasa Y: Synergistic tumour suppressor activity of E-cadherin and p53 in a conditional mouse model for metastatic diffuse-type gastric cancer. *Gut* 2012, **61**:344-353.

64. Boyault S, Drouet Y, Navarro C, Bachelot T, Lasset C, Treilleux I, Tabone E, Puisieux A, Wang Q: Mutational characterization of individual breast tumors: TP53 and PI3K pathway genes are frequently and distinctively mutated in different subtypes. *Breast Cancer Res Treat* 2012, **132**:29-39.

65. Craig DW, O'Shaughnessy JA, Kiefer JA, Aldrich J, Sinari S, Moses TM, Wong S, Dinh J, Christoforides A, Blum JL, Aitelli CL, Osborne CR, Izatt T, Kurdoglu A, Baker A, Koeman J, Barbacioru C, Sakarya O, De La Vega FM, Siddiqui A, Hoang L, Billings PR, Salhia B, Tolcher AW, Trent JM, Mousses S, Von Hoff D, Carpten JD: Genome and transcriptome sequencing in prospective metastatic triple-negative breast cancer uncovers therapeutic vulnerabilities. *Mol Cancer Ther* 2013, **12**:104-116.

66. Lafarga V, Cuadrado A, Nebreda AR: p18(Hamlet) mediates different p53-dependent responses to DNA-damage inducing agents. *Cell Cycle* 2007, **6**:2319-2322.

67. Roegiers F, Jan YN: Asymmetric cell division. *Curr Opin Cell Biol* 2004, **16**:195-205.

68. Colaluca IN, Tosoni D, Nuciforo P, Senic-Matuglia F, Galimberti V, Viale G, Pece S, Di Fiore PP: NUMB controls p53 tumour suppressor activity. *Nature* 2008, **451**:76-80.

69. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W: Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* 2004, **14**:708-715.

70. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, Arai Y, Takahashi H, Shirakihara T, Nagasaki M, Shibuya T, Nakano K, Watanabe-Makino K, Tanaka H, Nakamura H, Kusuda J, Ojima H, Shimada K, Okusaka T, Ueno M, Shigekawa Y, Kawakami Y, Arihiro K, Ohdan H, Gotoh K, Ishikawa O, *et al*: Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 2012, **44**:760-764.

71. Cheung HW, Cowley GS, Weir BA, Boehm JS, Rusin S, Scott JA, East A, Ali LD, Lizotte PH, Wong TC, Jiang G, Hsiao J, Mermel CH, Getz G, Barretina J, Gopal S, Tamayo P, Gould J, Tsherniak A, Stransky N, Luo B, Ren Y, Drapkin R, Bhatia SN, Mesirov JP, Garraway LA, Meyerson M, Lander ES, Root DE, Hahn WC: Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc Natl Acad Sci USA* 2011, **108**:12372-12377.

72. Schuster-Bockler B, Lehner B: Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* 2012, **488**:504-507.

73. Stratton MR, Campbell PJ, Futreal PA: The cancer genome. *Nature* 2009, **458**:719-724.

74. Pruitt KD, Tatusova T, Klimke W, Maglott DR: NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* 2009, **37**: D32-36.

75. Kent WJ: BLAT–the BLAST-like alignment tool. *Genome Res* 2002, **12**:656-664.

76. Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, Aburatani H: Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* 2005, **86**:127-141.

77. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB: A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004, **101**:6062-6067.

78. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muertter RN, Holko M, Ayanbule O, Yefanov A, Soboleva A: NCBI GEO: archive for functional genomics data sets–10 years on. *Nucleic Acids Res* 2011, **39**: D1005-1010.

79. Gautier L, Cope L, Bolstad BM, Irizarry RA: affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004, **20**:307-315.

80. Hubbell E, Liu WM, Mei R: Robust estimators for expression analysis. *Bioinformatics* 2002, **18**:1585-1592.

81. The Cancer Genome Atlas.. [http://cancergenome.nih.gov/].

82. Liu X, Jian X, Boerwinkle E: dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 2011, **32**:894-899.

83. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Harrys Kishore CJ, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A: Human Protein Reference Database–2009 update. *Nucleic Acids Res* 2009, **37**:D767-772.

84. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M: The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res* 2010, **39**:D698-704.

85. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K, Hermjakob H: The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 2009, **38**:D525-531.

86. Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res* 2009, **38**:D532-539.

87. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* 2004, **32**: D449-451.

88. D'Antonio M, Pendino V, Sinha S, Ciccarelli FD: Network of Cancer Genes (NCG 3.0): integration and analysis of genetic and network properties of cancer genes. *Nucleic Acids Res* 2012, **40**:D978-983.

89. [http://igraph.sourceforge.net/].

90. Csardi G, Nepusz T: The igraph software package for complex network research. *InterJournal* 2006, Complex Systems 1695.

91. The Integrative Genomics Portal at the BROAD Institute.. [http://www.broadinstitute.org/IGP/].

92. Watanabe R, Chano T, Inoue H, Isono T, Koiwai O, Okabe H: Rb1cc1 is critical for myoblast differentiation through Rb1 regulation. *Virchows Arch* 2005, **447**:643-648.

93. Nijwening JH, Geutjes EJ, Bernards R, Beijersbergen RL: The histone demethylase Jarid1b (Kdm5b) is a novel component of the Rb pathway and associates with E2f-target genes in MEFs during senescence. *PLoS One* 2011, **6**:e25235.

94. Bidere N, Snow AL, Sakai K, Zheng L, Lenardo MJ: Caspase-8 regulation by direct interaction with TRAF6 in T cell receptor-induced NF-kappaB activation. *Curr Biol* 2006, **16**:1666-1671.

95. Gao H, Yu Z, Bi D, Jiang L, Cui Y, Sun J, Ma R: Akt/PKB interacts with the histone H3 methyltransferase SETDB1 and coordinates to silence gene expression. *Mol Cell Biochem* 2007, **305**:35-44.

96. Chih DY, Park DJ, Gross M, Idos G, Vuong PT, Hirama T, Chumakov AM, Said J, Koeffler HP: Protein partners of C/EBPepsilon. *Exp Hematol* 2004, **32**:1173-1181.

97. He Q, Johnson VJ, Osuchowski MF, Sharma RP: Inhibition of serine palmitoyltransferase by myriocin, a natural mycotoxin, causes induction of c-myc in mouse liver. *Mycopathologia* 2004, **157**:339-347.

98. Tussie-Luna MI, Bayarsaihan D, Ruddle FH, Roy AL: Repression of TFII-I-dependent transcription by nuclear exclusion. *Proc Natl Acad Sci USA* 2001, **98**:7789-7794.

99. Haw R, Stein L: Using the reactome database. *Curr Protoc Bioinformatics* 2012, Chapter 8:Unit8 7.

100. Taniguchi S, Liu H, Nakazawa T, Yokoyama K, Tezuka T, Yamamoto T: **p250GAP, a neural RhoGAP protein, is associated with and phosphorylated by Fyn.** *Biochem Biophys Res Commun* 2003, **306**:151-155.

101. He KL, Deora AB, Xiong H, Ling Q, Weksler BB, Niesvizky R, Hajjar KA: **Endothelial cell annexin A2 regulates polyubiquitination and degradation of its binding partner S100A10/p11.** *J Biol Chem* 2008, **283**:19192-19200.

102. Twal WO, Czirok A, Hegedus B, Knaak C, Chintalapudi MR, Okagawa H, Sugi Y, Argraves WS: **Fibulin-1 suppression of fibronectin-regulated cell adhesion and motility.** *J Cell Sci* 2001, **114**:4587-4598.

103. Querol-Audi J, Yan C, Xu X, Tsutakawa SE, Tsai MS, Tainer JA, Cooper PK, Nogales E, Ivanov I: **Repair complexes of FEN1 endonuclease, DNA, and Rad9-Hus1-Rad1 are distinguished from their PCNA counterparts by functionally important stability.** *Proc Natl Acad Sci USA* 2012, **109**:8528-8533.

104. Chicas A, Kapoor A, Wang X, Aksoy O, Evertts AG, Zhang MQ, Garcia BA, Bernstein E, Lowe SW: **H3K4 demethylation by Jarid1a and Jarid1b contributes to retinoblastoma-mediated gene silencing during cellular senescence.** *Proc Natl Acad Sci USA* 2012, **109**:8971-8976.

105. Ho TH, Charlet BN, Poulos MG, Singh G, Swanson MS, Cooper TA: **Muscleblind proteins regulate alternative splicing.** *EMBO J* 2004, **23**:3103-3112.

106. Hatzfeld M: **The p120 family of cell adhesion molecules.** *Eur J Cell Biol* 2005, **84**:205-214.

107. Will CL, Urlaub H, Achsel T, Gentzel M, Wilm M, Luhrmann R: **Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein.** *EMBO J* 2002, **21**:4978-4988.

108. Zhu B, Mandal SS, Pham AD, Zheng Y, Erdjument-Bromage H, Batra SK, Tempst P, Reinberg D: **The human PAF complex coordinates transcription with events downstream of RNA synthesis.** *Genes Dev* 2005, **19**:1668-1673.

109. Kim JH, Sung KS, Jung SM, Lee YS, Kwon JY, Choi CY, Park SH: **Pellino-1, an adaptor protein of interleukin-1 receptor/toll-like receptor signaling, is sumoylated by Ubc9.** *Mol Cells* 2010, **31**:85-89.

110. Lin G, Aranda V, Muthuswamy SK, Tonks NK: **Identification of PTPN23 as a novel regulator of cell invasion in mammary epithelial cells from a loss-of-function screen of the 'PTP-ome'.** *Genes Dev* 2011, **25**:1412-1425.

111. Gazit K, Moshonov S, Elfakess R, Sharon M, Mengus G, Davidson I, Dikstein R: **TAF4/4b × TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes.** *J Biol Chem* 2009, **284**:26286-26296.

112. Adithi M, Venkatesan N, Kandalam M, Biswas J, Krishnakumar S: **Expressions of Rac1, Tiam1 and Cdc42 in retinoblastoma.** *Exp Eye Res* 2006, **83**:1446-1452.

113. Xu L, Yang L, Moitra PK, Hashimoto K, Rallabhandi P, Kaul S, Meroni G, Jensen JP, Weissman AM, D'Arpa P: **BTBD1 and BTBD2 colocalize to cytoplasmic bodies with the RBCC/tripartite motif protein, TRIM5delta.** *Exp Cell Res* 2003, **288**:84-93.

114. Tashiro K, Tsunematsu T, Okubo H, Ohta T, Sano E, Yamauchi E, Taniguchi H, Konishi H: **GAREM, a novel adaptor protein for growth factor receptor-bound protein 2, contributes to cellular transformation through the activation of extracellular signal-regulated kinase signaling.** *J Biol Chem* 2009, **284**:20206-20214.

115. Qiu Y, Wang ZL, Jin SQ, Pu YF, Toyosawa S, Aozasa K, Morii E: **Expression level of pre-B-cell leukemia transcription factor 2 (PBX2) as a prognostic marker for gingival squamous cell carcinoma.** *J Zhejiang Univ Sci B* 2012, **13**:168-175.

116. Wollscheid B, Bausch-Fluck D, Henderson C, O'Brien R, Bibel M, Schiess R, Aebersold R, Watts JD: **Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins.** *Nat Biotechnol* 2009, **27**:378-386.

117. Scott GK, Marx C, Berger CE, Saunders LR, Verdin E, Schafer S, Jung M, Benz CC: **Destabilization of ERBB2 transcripts by targeting 3' untranslated region messenger RNA associated HuR and histone deacetylase-6.** *Mol Cancer Res* 2008, **6**:1250-1258.

118. Shigematsu H, Iwasaki H, Otsuka T, Ohno Y, Arima F, Niho Y: **Role of the vav proto-oncogene product (Vav) in erythropoietin-mediated cell proliferation and phosphatidylinositol 3-kinase activity.** *J Biol Chem* 1997, **272**:14334-14340.

119. Miller HB, Robinson TJ, Gordan R, Hartemink AJ, Garcia-Blanco MA: **Identification of Tat-SF1 cellular targets by exon array analysis reveals dual roles in transcription and splicing.** *RNA* 2011, **17**:665-674.

120. Michel D, Arsanto JP, Massey-Harroche D, Beclin C, Wijnholds J, Le Bivic A: **PATJ connects and stabilizes apical and lateral components of tight junctions in human intestinal cells.** *J Cell Sci* 2005, **118**:4049-4057.

121. Clemente D, Ortega MC, Arenzana FJ, de Castro F: **FGF-2 and Anosmin-1 are selectively expressed in different types of multiple sclerosis lesions.** *J Neurosci* 2012, **31**:14899-14909.

122. Severson EA, Lee WY, Capaldo CT, Nusrat A, Parkos CA: **Junctional adhesion molecule A interacts with Afadin and PDZ-GEF2 to activate Rap1A, regulate beta1 integrin levels, and enhance cell migration.** *Mol Biol Cell* 2009, **20**:1916-1925.

123. Cannavo E, Gerrits B, Marra G, Schlapbach R, Jiricny J: **Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2.** *J Biol Chem* 2007, **282**:2976-2986.

124. Yang M, Waterman ML, Brachmann RK: **hADA2a and hADA3 are required for acetylation, transcriptional activity and proliferative effects of beta-catenin.** *Cancer Biol Ther* 2008, **7**:120-128.

125. Panaretou C, Domin J, Cockcroft S, Waterfield MD: **Characterization of p150, an adaptor protein for the human phosphatidylinositol (PtdIns) 3-kinase. Substrate presentation by phosphatidylinositol transfer protein to the p150.PtdIns 3-kinase complex.** *J Biol Chem* 1997, **272**:2477-2485.

126. Leonard D, Ajuh P, Lamond AI, Legerski RJ: **hLodestar/HuF2 interacts with CDC5L and is involved in pre-mRNA splicing.** *Biochem Biophys Res Commun* 2003, **308**:793-801.

127. Krwawicz J, Arczewska KD, Speina E, Maciejewska A, Grzesiuk E: **Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and DNA-end processors and their implication in mutagenesis and human disease.** *Acta Biochim Pol* 2007, **54**:413-434.

128. Tanida I, Tanida-Miyake E, Komatsu M, Ueno T, Kominami E: **Human Apg3p/Aut1p homologue is an authentic E2 enzyme for multiple substrates, GATE-16, GABARAP, and MAP-LC3, and facilitates the conjugation of hApg12p to hApg5p.** *J Biol Chem* 2002, **277**:13739-13744.

129. Takeuchi A, Miyamoto T, Yamaji K, Masuho Y, Hayashi M, Hayashi H, Onozaki K: **A human erythrocyte-derived growth-promoting factor with a wide target cell spectrum: identification as catalase.** *Cancer Res* 1995, **55**:1586-1589.

130. Jin J, Arias EE, Chen J, Harper JW, Walter JC: **A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1.** *Mol Cell* 2006, **23**:709-721.

131. Astoul E, Laurence AD, Totty N, Beer S, Alexander DR, Cantrell DA: **Approaches to define antigen receptor-induced serine kinase signal transduction pathways.** *J Biol Chem* 2003, **278**:9267-9275.

132. An JH, Kim JW, Jang SM, Kim CH, Kang EJ, Choi KH: **Gelsolin negatively regulates the activity of tumor suppressor p53 through their physical interaction in hepatocarcinoma HepG2 cells.** *Biochem Biophys Res Commun* 2011, **412**:44-49.

133. Ishimi Y: **A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex.** *J Biol Chem* 1997, **272**:24508-24513.

134. Wimuttisuk W, Singer JD: **The Cullin3 ubiquitin ligase functions as a Nedd8-bound heterodimer.** *Mol Biol Cell* 2007, **18**:899-909.

135. Giglione C, Gonfloni S, Parmeggiani A: **Differential actions of p60c-Src and Lck kinases on the Ras regulators p120-GAP and GDP/GTP exchange factor CDC25Mm.** *Eur J Biochem* 2001, **268**:3275-3283.

136. Perez-Cornejo P, Gokhale A, Duran C, Cui Y, Xiao Q, Hartzell HC, Faundez V: **Anoctamin 1 (Tmem16A) Ca2+-activated chloride channel stoichiometrically interacts with an ezrin-radixin-moesin network.** *Proc Natl Acad Sci USA* 2012, **109**:10376-10381.

137. Koda Y, Soejima M, Johnson PH, Smart E, Kimura H: **Missense mutation of FUT1 and deletion of FUT2 are responsible for Indian Bombay phenotype of ABO blood group system.** *Biochem Biophys Res Commun* 1997, **238**:21-25.

138. Yeh PY, Kuo SH, Yeh KH, Chuang SE, Hsu CH, Chang WC, Lin HI, Gao M, Cheng AL: **A pathway for tumor necrosis factor-alpha-induced Bcl10 nuclear translocation. Bcl10 is up-regulated by NF-kappaB and phosphorylated by Akt1 and then complexes with Bcl3 to enter the nucleus.** *J Biol Chem* 2006, **281**:167-175.

139. Chano T, Ikebuchi K, Ochi Y, Tameno H, Tomita Y, Jin Y, Inaji H, Ishitobi M, Teramoto K, Nishimura I, Minami K, Inoue S, Isono T, Saitoh M, Shimada T, Hisa Y, Okabe H: **RB1CC1 activates RB1 pathway and inhibits proliferation and cologenic survival in human cancer.** *PLoS One* 2010, **5**:e11404.

140. Kelly SM, Pabit SA, Kitchen CM, Guo P, Marfatia KA, Murphy TJ, Corbett AH, Berland KM: **Recognition of polyadenosine RNA by zinc finger proteins.** *Proc Natl Acad Sci USA* 2007, **104**:12306-12311.

141. Dry K, Kenwrick S, Rosenthal A, Platzer M: **The complete sequence of the human locus for NgCAM-related cell adhesion molecule reveals a novel alternative exon in chick and man and conserved genomic organization for the L1 subfamily.** *Gene* 2001, **273**:115-122.

142. Clarke CJ, Guthrie JM, Hannun YA: **Regulation of neutral sphingomyelinase-2 (nSMase2) by tumor necrosis factor-alpha involves protein kinase C-delta in lung epithelial cells.** *Mol Pharmacol* 2008, **74**:1022-1032.

143. da Silva Xavier G, Rutter J, Rutter GA: **Involvement of Per-Arnt-Sim (PAS) kinase in the stimulation of preproinsulin and pancreatic duodenum homeobox 1 gene expression by glucose.** *Proc Natl Acad Sci USA* 2004, **101**:8319-8324.

144. Eiseler T, Doppler H, Yan IK, Kitatani K, Mizuno K, Storz P: **Protein kinase D1 regulates cofilin-mediated F-actin reorganization and cell motility through slingshot.** *Nat Cell Biol* 2009, **11**:545-556.