EPIGENETICS AND GENETICS

A census of amplified and overexpressed human cancer genes

Thomas Santarius, Janet Shipley‡, Daniel Brewer‡, Michael R. Stratton* and Colin S. Cooper#*

Abstract | Integrated genome-wide screens of DNA copy number and gene expression in human cancers have accelerated the rate of discovery of amplified and overexpressed genes. However, the biological importance of most of the genes identified in such studies remains unclear. In this Analysis, we propose a weight-of-evidence based classification system for identifying individual genes in amplified regions that are selected for during tumour development. In a census of the published literature we have identified 77 genes for which there is good evidence of involvement in the development of human cancer.

Over the past 25 years considerable effort has been directed towards the identification of genes involved in cancer development. A census of cancer genes[1] listed 291 human genes for which there is sufficient evidence to support a causal role in sporadic or familial cancer development when mutated. This was recently updated to 384 genes, therefore almost 2% of genes in the human genome are thought to be causally implicated in cancer development when appropriately mutated. The genes listed in this census can be altered by several types of genetic alterations, including point mutations, deletions and re-arrangements. However, in the original list only six were altered by amplification and consequent overexpression (*AKT2*, *ERBB2*, *MYC*, *MYCL1*, *MYCN* and *REL*). The shortness of this list was not a result of gene amplification being an infrequent mechanism for converting a gene into a cancer gene. On the contrary, amplified regions are common in human cancer genomes. The shortness of this list reflects the difficulty encountered in identifying the true cancer gene on amplicons that often include several candidate genes.

For the purposes of this article the term gene amplification refers to the somatically acquired increase in copy number of a restricted region of the genome that is the underlying genomic mechanism that results in overexpression of a dominantly acting cancer gene. A more complete definition of the term 'amplification' and the criteria for inclusion of amplified genetic regions in this article are given in Supplementary information S1(text). The mechanism of amplification can be complex, involving breakage-fusion-bridges cycles, formation and reinsertion of double minute chromosomes or the formation of clusters of small genomic fragments[2–5]. Amplification events often include multiple genes, so consideration of the pattern of genetic alteration alone is usually insufficient to identify the cancer gene that is being selected for in the amplicon and is contributing to oncogenesis. More information is usually required, including physical mapping of the amplicon in multiple cancers, evidence that amplified genes are accompanied by overexpression in tumours that have the amplicon, correlation of amplification and/or overexpression with clinical outcome data, biological investigations of function and in some cases the efficacy of drugs targeted against the overexpressed proteins. For most putative amplified cancer genes, these datasets are incomplete and different combinations of data are available for different amplicons. The interpretation of these data might also be difficult if more than one gene can contribute to the biological effect of an amplicon or if the identity of the tumour promoting genes in a genetically defined amplicon are different in distinct cancer types. In this Analysis we use a classification scheme for amplified and overexpressed genes that takes into account the complex and sometimes distinct datasets that are available for different cancer amplicons.

Assessing amplified genes

The criteria used for assessing amplified genes are summarised in BOX1. Using searches in the PubMed database, we have reviewed all data relevant to gene amplification that we could identify. However, references were only included when they contained information relevant to our new classification system (see Supplementary information S1 (text)). The cut off date for including all references was April 2009, although a few recent and relevant publications have been included.

**The Wellcome Trust Sanger Centre, Wellcome Trust Genome Campus Hinxton, Cambridge CB10 1SA,UK.*
‡The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK.
Correspondence to C.S.C.
e-mail: colin.cooper@icr.ac.uk
doi:10.1038/nrc2771

At a glance

- Integrated screens of DNA copy number and gene expression in human cancers using microarray platforms have accelerated the rate of discovery of amplified and overexpressed genes. However, the biological importance of most of the genes identified in such studies remains unclear.
- Amplification events often include multiple genes, so consideration of the pattern of genetic alteration alone is usually insufficient to identify which gene in an amplicon is being selected for owing to its contribution to oncogenesis. Supplementary datasets are usually required, including physical mapping, determination of overexpression, correlation with clinical outcome, biological investigations of function and, in some cases, efficacy of drugs targeted against the encoded overexpressed proteins.
- In this Analysis we propose a weight-of-evidence based classification system for identifying individual genes in amplified regions that are selected for in an amplicon and so contribute to cancer development. The proposed classification scheme takes into account the complex and sometimes distinct datasets that are available for different amplicons in cancer.
- Using this classification scheme in a census of the published literature, we have identified 77 genes for which there is evidence of involvement in human cancer development.
- The 77 genes were divided into three classes based on the weight of supporting evidence. We consider that for class II (12 genes), and class I (3 genes) the evidence is sufficiently strong for their inclusion in the census of human cancer genes.
- Linking newly generated integrated datasets to supporting evidence using the criteria outlined in this Analysis will aid in the future identification of amplified and overexpressed genes that contribute to cancer development.

Minimum region of amplification. For point mutations and translocations the pattern and position of the observed genetic alterations makes cancer gene assignment unambiguous. In isolated cases, genomic amplifications can encompass a single gene; for example calcium channel voltage dependent alpha-1E subunit (*CACNA1E*) in Wilm's tumour[6] and *KIT* in testicular germ cell tumours (TGCT)[7]. More often an amplicon consists of a large chromosomal region encompassing many genes. To identify the tumour promoting gene in an amplicon it is necessary to map the amplicon in many cancers to identify a common genomic region, known as the minimal region of amplification (MRA), that is amplified in all the cancers examined. For example, at the 2p24 amplicon found in neuroblastomas this approach identified *MYCN* as the single protein encoding gene contained in a common 600kb MRA[8]. However, even in this straight forward case the biological contribution of adjacent genes is not excluded and involvement of the frequently co-amplified gene, *DDX1*, has been proposed[8–11]. Large common regions of genetic gain can be identified that cannot be further defined by analyses of additional cancers. This suggests the involvement of multiple tumour promoting genes in the amplicon. For example, the 12p amplicon in TGCT is 5Mb in length and contains at least 22 genes[12–14]. Complex regions of genetic gain also exist. One example is the 11q13 amplicon in breast cancer where several separate but closely linked peaks of amplification have been reported — cyclin D1 (*CCND1*), *EMSY* and p21/CDC42/RAC1-activated kinase 1 (*PAK1*) are the candidate tumour promoting genes[15,16] (TABLE 1; see Supplementary information S2 (table)).

Amplification causes overexpression. Expression profiling allows the identification of a subset of candidate oncogenes in a MRA where DNA copy number gain is accompanied by overexpression. Although a good statistical correlation between amplification and overexpression is expected, there is not always an exact match between the level of DNA gain and gene expression. Genes that contribute to tumour development (driver genes) can be overexpressed by different mechanisms in the absence of DNA amplification, as observed for the *MDM2* gene in the 12q13-15 amplicon in human sarcomas[17]. In addition, where there is more than one driver gene, absence of expression in a subset of tumours cannot be used to exclude involvement.

Knowledge of control pathways. Knowledge of other cancer genes and control pathways can help implicate amplified and overexpressed genes. This is exemplified by the *KIT* gene at 4q12 that can be either amplified or activated by point mutations in TGCT[7]. Crucially, mutation and amplification are mutually exclusive, consistent with the view that they represent alternative mechanisms of gene activation. The established amplification and overexpression of other family members in the same cancer type can also support involvement. This is true for Myc family members in the development of lung cancer: amplification and overexpression of *MYCN*, *MYCL1* or *MYC* have all been reported in different individual cancers (TABLE 1; see Supplementary information S2 (table)). Inherited mutations can predispose to the same cancer type as illustrated by cyclin dependent kinase 4 (*CDK4*) amplification and overexpression in malignant melanoma (TABLE 1). Rare melanoma prone families also have activating mutations in *CDK4* (REF. 18). Knowledge of the pathways crucial for the development of the cancer under investigation can also be helpful. For example, mutation of the *RB1* gene has been reported in bladder cancer and this information was used to implicate *E2F3*, which encodes a regulator of RB, as the driver gene in the MRA in the 6p22 amplicon[19,20].

Clinical correlations. Correlations between amplified genes and clinical parameters are often carried out. Such correlations can be used to implicate an amplification event in determining clinical outcome, but do not usually provide the detail required to determine the relative importance of individual genes. When array comparative genomic hybridization (aCGH) approaches are used in genome wide screens, correlations with clinical outcome can be useful in focusing attention on amplicons that are biologically most important. For example, a screen of Wilm's tumour samples identified an amplicon at 1q25.3 as potentially clinically important[6]. Correlation of expression with clinical outcome can also provide supporting evidence in implicating a specific amplified gene (BOX 1).

Biological activity. Short interfering RNA (siRNA) knockdown of amplified and overexpressed genes in cell lines, or ectopic expression of the gene in appropriate cell lines that lack the amplification, can support the involvement of individual genes in contributing to the

biological consequences of an amplicon. Such experiments need to be interpreted with care because they can detect genes important for the growth of cells in culture that have no special relevance to cancer development. The results from siRNA knockdown experiments can be especially valuable when the siRNA does not affect cancer cell lines of the same type that lack amplification and overexpression of the genes under investigation. We discounted studies where this important control had not been included. siRNA studies can provide particularly interesting insights into the structure and biological contribution of genes in an amplicon. For example, it is well established that *ERBB2* is the gene selected for in the amplification at 17q12 in breast cancer (TABLE 1; see Supplementary information S2 (table)). However, siRNA knockdown studies in cells lines with the 17q12 amplicon showed that *ERBB2* and the adjacent co-amplified genes — growth factor receptor bound protein 7 (*GRB7*) and STAR-related lipid transfer domain containing protein 3 (*STARD3*) — contribute to the biological effect of this amplicon. Indeed, siRNA knockdown of these genes had no effect on breast cancer lines that lacked the 17q12 amplicon[21]. Analogous *in vitro* knockdown experiments can also be undertaken using drugs that target a gene or its encoded protein, as demonstrated by drugs that target MET and only inhibit the growth of gastric cancer cell lines containing the 7q31 amplicon that includes *MET*[22].

Box 1 | **Criteria for assigning amplified and overexpressed genes**

To be considered in this analysis, genes must lie within a minimal region of amplification and overexpression of the normal gene must accompany the amplification. Genes that meet this criteria are classified as class IV genes. Genes were further classified according to their scores as determined by the evidence shown in the table.

- Class III genes require 1 or 2 points indicating significant evidence of their involvement in cancer development.
- Class II genes require 3 or more points and indicate substantial evidence for their involvement in cancer development.
- Class I genes require that a drug that targets the encoded protein is used to treat patients for which efficacy must have been shown in clinical trials.

Supporting evidence	Criteria	Assigned score
Clinical correlation	Expression correlated with clinical outcome	1 point
Knowledge of cancer genes and control pathways	Mutation and amplification mutually exclusive	1 point
	Sole identified gene in the amplicon	1 point
	Inherited mutation predisposed to the same cancer type	1 point
	Mutation or amplification and overexpression of other genes in the same pathway	1 point
Biological evidence	Overexpression causes biological effect	1 point*
	siRNA knockdown or targeted drug causes biological effect in the cell containing the amplified and overexpressed gene, but not in cells lacking the amplified and overexpressed gene	1 point*
Animal studies	Substantial evidence from animal experiments	1 point

* Modulation of gene expression affecting biological properties can only be counted once.

Evidence from animal experiments. Well designed animal experiments can provide supporting evidence for the involvement of a gene in human cancer development. This was shown by the work of Zender *et al.*[23] who infected hepatoblasts from *Trp53*$^{-/-}$ mice with retroviruses that express *Myc* and these were used to colonize the livers of host animals. An aCGH analysis of the liver cancers that arose from the seeded cells identified a recurrent amplicon at 9qA1. Importantly, an amplicon at the syntenic region, 11q22, was found in human hepatocellular carcinomas. *YAP1* and baculovirus IAP repeat containing protein 2 (*BIRC2*) were amplified and overexpressed at this locus both in human and mouse cancers. Experiments involving changes in the expression of *Yap1* and *Birc2* in this mouse model showed the biological importance of and cooperation between the two genes. These data support the involvement of these genes in the syntenic amplicon found in human liver cancer (TABLE 1; see Supplementary information S2 (table)).

Drug activity in patients. The ultimate test of whether an amplified gene makes a contribution to the growth of a cancer *in vivo* is whether a drug targeted against its encoded protein can be used to effectively treat primary cancers containing the amplicon. This has been shown for the ERBB2 antibody trastuzumab, which improves survival in patients with breast cancers that have amplified and/or overexpressed *ERBB2* (REFS 24,25).

Classification

To facilitate the identification of key cancer genes we have developed a weight-of-evidence based classification system that takes into account the different sets of published data that are available for each amplicon. Genes are categorised into four classes (I–IV) based on the type of supporting data. To qualify as class IV the gene must lie in the MRA and the amplified gene must be accompanied by overexpression. Each line of evidence discussed above is then assessed. All types of evidence, listed in BOX 1, are scored with equal weighting (one point). Different views may be taken on the relative weights that should be assigned to different types of evidence. However, we believe that our system of equal weighting, in which only one type of biological evidence can be scored (BOX 1), provides an appropriate balance between genetic and biological information. A gene must score at least one point to be considered as a class III gene (significant evidence of involvement) and three or more points to be considered a class II gene (substantial evidence of involvement). Reproducible demonstration (in multiple Phase II or Phase III trials) that therapeutic targeting of the gene or gene product improves clinical outcome in patients whose cancers contain the amplification is required for assignment as a class I gene.

In applying this system we considered each cancer type separately. This requirement arose because the same cytogenetically defined amplicon can have different driver genes in different cancer types, or even subtypes. For example, in breast, colon and lung cancers with an amplicon at 8q24 the proposed driver gene is *MYC*,

Table 1 | **Amplified and overexpressed genes in human cancer**

Cancer type	Class III	Class II	Class I
Acute myeloid leukaemia	*TRIB1*		
Bladder	*YWHAQ, E2F3, YWHAZ, ERRB2, AURKA*		
Breast	*SHC1, CKS1B, RUVBL1, C8orf4, LSM1, FGFR1, BAG4, MTDH, MYC, EMSY, PAK1, CDK4, MDM2, PLA2G10, STARD3, GRB7, RPS6KB1, PPM1D, CCNE1, YWHAB, ZNF217, AURKA, PTK6*	*CCND1, NCOA3*	*ERBB2*
Colorectal	*MYC*		*EGFR*
Diffuse large B cell lymphoma	*REL*		
Endometrial	*ERBB2*		
Gastric	*RAB23, MET, MYC, ERBB2, CDC6*		
Glioma	*MDM4, EGFR, CDK4, MDM2, AKT3, CCND2, CDK6, MET*		
Head and neck	*DCUN1D1*		
Hepatocellular carcinoma	*CHD1L*		
Hodgkin's lymphoma	*REL*		
Laryngeal squamous cell carcinoma	*FADD*		
Liver		*YAP1, BIRC2*	
Lung	*MYCN, EGFR, MET, WHSC1L1, YWHAZ, MYC, CCND1, MDM2, BCL2L2, PAX9, NKX2-1, KIAA0174, DCUN1D1, EEF1A2*	*MYCL1, SKP2, NKX2-8*	
Malignant melanoma	*MITF, CCND1, CDK4*		
Medulloblastoma	*MYC*		
Neuroblastoma	*MDM2*	*MYCN*	
Oesophageal	*PRKCI, ZNF639, SKP2, EGFR, SHH, DYRK2, ERBB2, CCNE1, AURKA*		
Oral squamous cell carcinoma	*CCND1*		
Osteosarcoma	*COPS3*		
Ovarian	*EIF5A2, EVI1, EMSY, ERBB2, RPS6KB1, AKT2*	*RAB25, PIK3CA*	
Pancreatic	*ARPC1A, SMURF1, MED29*		
Pancreatobillary	*GATA6*		
Prostate	*MYC*		*AR*
Retinoblastoma	*E2F3*	*MDM4*	
Rhabdomyosarcoma	*MYCN, FGFR1, GPC5*		
Sarcoma	*JUN, MAP3K5, YEATS4, CDK4, DYRK2*	*MDM2*	
Soft tissue sarcoma	*SKP2*		
Testicular germ cell tumour	*KIT, KRAS*		
Wilm's tumour	*CACNA1E*		

AR, androgen receptor; *ARPC1A*, actin-related protein complex 2/3 subunit A; *AURKA*, Aurora kinase A; *BAG4*, BCL-2 associated anthogene 4; BCL2L2, BCL-2 like 2; *BIRC2*, Baculovirus IAP repeat containing protein 2; *CACNA1E*, calcium channel voltage dependent alpha-1E subunit; *CCNE1*, cyclin E1; *CDK4*, cyclin dependent kinase 4; *CHD1L*, chromodomain helicase DNA binding domain 1-like; *CKS1B*, CDC28 protein kinase 1B; *COPS3*, COP9 subunit 3; *DCUN1D1*, DCN1 domain containing protein 1; *DYRK2*, dual specificity tyrosine phosphorylation regulated kinase 2; *EEF1A2*, eukaryotic elongation transcription factor 1 alpha 2; *EGFR*, epidermal growth factor receptor; *FADD*, Fas-associated via death domain; *FGFR1*, fibroblast growth factor receptor 1, *GATA6*, GATA binding protein 6; *GPC5*, glypican 5; *GRB7*, growth factor receptor bound protein 7; *MAP3K5*, mitogen activated protein kinase kinase kinase 5; *MED29*, mediator complex subunit 5; *MITF*, microphthalmia associated transcription factor; *MTDH*, metadherin; *NCOA3*, nuclear receptor coactivator 3; *NKX2-1*, NK2 homeobox 1; *PAK1*, p21/CDC42/RAC1-activated kinase 1; *PAX9*, paired box gene 9; *PIK3CA*, phosphatidylinositol-3 kinase catalytic α; *PLA2G10*, phopholipase A2, group X; *PPM1D*, protein phosphatase magnesium-dependent 1D; *PTK6*, protein tyrosine kinase 6; *PRKCI*, protein kinase C iota; *RPS6KB1*, ribosomal protein S6 kinase 70kDa; *SKP2*, S-phase kinase associated protein; *SMURF1*, SMAD specific E3 ubiquitin protein ligase 1; *SHH*, sonic hedgehog homologue; *STARD3*, STAR-related lipid transfer domain containing protein 3; *YWHAQ*, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta isoform; *ZNF217*, zinc finger protein 217.

but in acute myeloid leukaemias with an 8q24 amplicon, *TRIB1* is the driver gene implicated (TABLE 1). The fibroblast growth factor receptor 1 (*FGFR1*) gene had been excluded as a driver gene for the breast cancer 8p11–12 amplicon where *PPAPDC1B* and *WHSC1L1* are considered better candidates[26,27]. However, a recent study has provided evidence that *FGFR1* is the driver gene for this amplicon specifically in lobular breast cancer[28]. In this

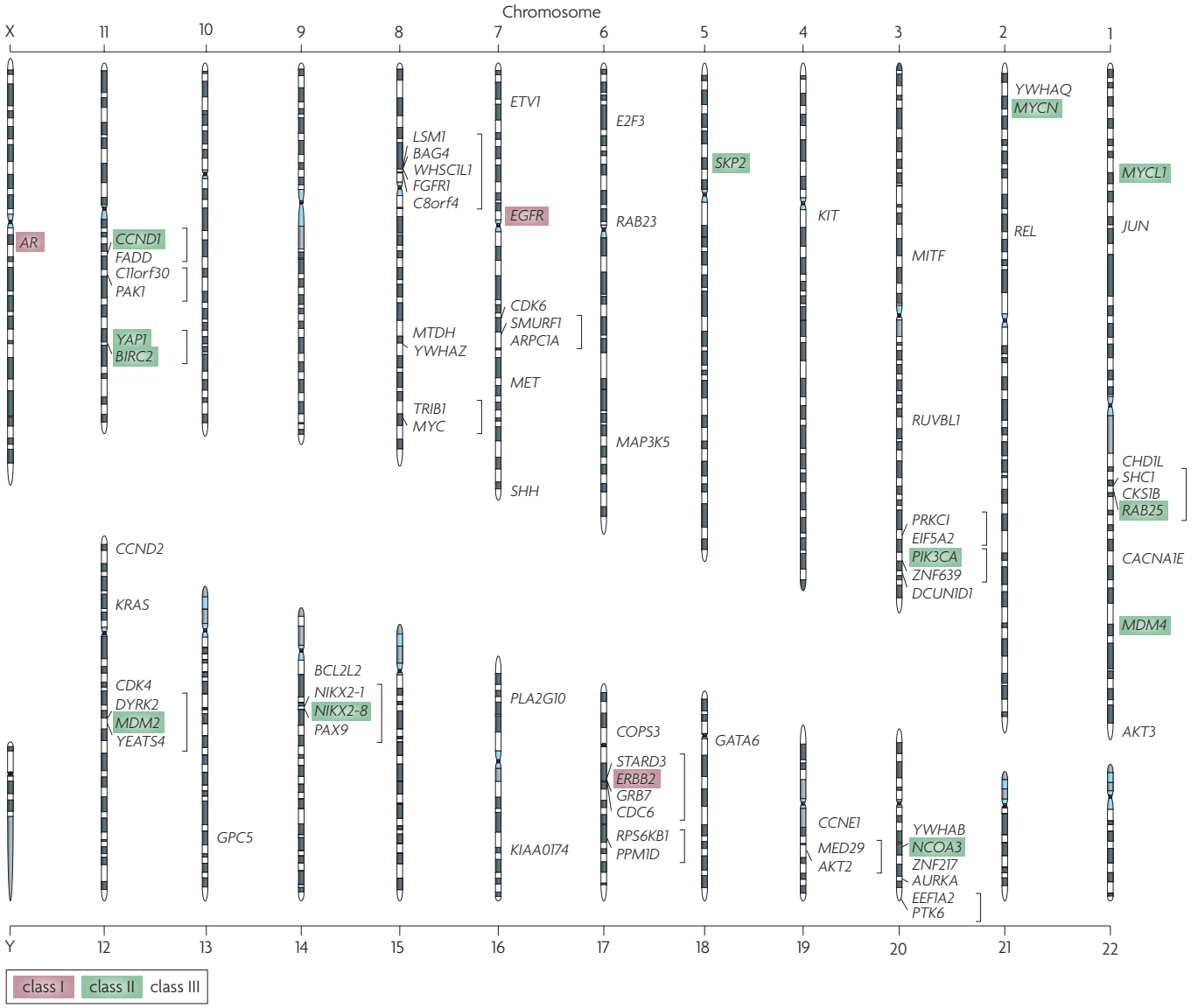

Figure 1 | **Idiogram.** This idiogram shows the genomic positions of amplified and overexpressed genes in cancer. Genes that reside at a similar genetic position are joined by a square bracket (]). The strength of evidence for involvement in human cancer, as defined by the criteria outlined in BOX 1, is as follows: class I genes have the strongest evidence, followed by class II and then class III.

census the classification system is only applied where normal cellular genes are thought to contribute to cancer development by amplification and overexpression: mutated genes that are also amplified and conventional drug resistance genes are not considered. This distinction is important because activation of cancer genes by mutation or translocation can also be associated with their amplification[29–31]. This analysis is also restricted to the examination of genes: evidence establishing the importance of microRNAs in amplified regions is starting to emerge but was considered too premature to include.

Many hundreds of class IV genes have been identified in integrated genome-wide studies. These are not comprehensively listed. Class I–class III genes are listed in TABLE 1. There are 62 class III genes, 12 class II genes and 3 class I genes (*ERBB2,* epidermal growth factor receptor (*EGFR*) and androgen receptor (*AR*)). Detailed justification for the classification of each gene is presented in Supplementary information S1 (text). A summary of the all of the genes and the cancers in which they have been implicated is shown in TABLE 1 and Supplementary information S2 (table). Their genomic positions are show in FIG. 1. We think that the evidence for class II and class I genes is sufficiently strong for their inclusion in the Cancer Gene Census.

Integrated genome analysis

Integrated genomic analyses involving parallel assessment of DNA copy number, expression and in some cases mutation have now been completed for many cancers, including breast and colon cancer and glioblastoma. Such analyses reveal complex patterns of genomic

changes. In a screen of 191 breast cancers, for example, Nikolsky *et al.*[32] identified 1,747 recurrently amplified genes organized into 30 amplicons, which equates to 5.6% of the genome. To assess each dataset we cross-referenced recurrent gene amplifications and overexpression events with established cancer genes that are altered by other mechanisms (point mutation, translocation, etc) in the same cancer type using the Cancer Gene Census database. These studies identified an additional 9 genes (tyrosine 3-monoxygenase/typtophan 5-monooxygenase activation protein β isoform (*YWHAB*), *YWHAQ*, *CDC6*, *SHC1* phopholipase A2, group X (*PLA2G10*), *CDK6*, *CCND2*, *AKT3* and *RUVBL1*) where there were genetic alterations in the same gene or in other genes in the same control pathway, providing class III evidence (see Supplementary information S3 (text)).

Gene function

Of the 77 class I, class II and class III amplified and overexpressed genes identified, only 14 (phosphatidylinositol-3 kinase catalytic α (*PIK3CA*), *KIT*, *EGFR*, *MET*, *WHSC1L1*, *FGFR1*, *MYC*, *CCND1*, *KRAS*, *ERBB2*, *CDK4*, *AR, CDK6, CCND2)* corresponded to cancer genes that are altered by other mechanisms[1], such as translocation or mutation. However, most of the 77 genes are involved in pathways already known to be deregulated in cancer development (Supplementary information S2 (table)), including MAPK signalling, ERBB signalling, apoptosis and the p53 pathway, hedgehog signalling and the transforming growth factor β signalling pathway. Notably, three of the new genes identified in our analyses of integrated datasets, *YWHAB*, *YWHAQ* and *CDC6,* together with the class III gene *YWHAZ,* are all thought to be components of the origin of replication complex that serves as a foundation for DNA replication.

Conclusions

This weight-of-evidence based analysis is complementary to the previously published census of oncogenes in human cancer[1]. In addition to the 6 amplified genes originally reported, we found significant or substantial evidence for the importance of 71 amplified genes in cancer development. Technologies are now readily available for carrying out integrated genomic analysis of DNA copy number and expression. Linking newly generated datasets to supporting evidence using the criteria outlined in this analysis will aid in identifying new genes that contribute to cancer development.

1. Futreal PA. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).
2. Bignell GR. *et al.* Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome Res.* **17**, 1296–1303 (2007).
3. Savelyeva L. & Schwab M. Amplification of oncogenes revisited: from expression profiling to clinical application. *Cancer Lett.* **167**, 115–123 (2001).
4. Myllykangas S. & Knuutila S. Manifestation, mechanisms and mysteries of gene amplifications. *Cancer Lett.* **232**, 79–89 (2006).
5. Schwab M. *Oncogene* amplification in solid tumors. *Semin. Cancer Biol.* **9**, 319–325 (1999).
6. Natrajan R. *et al.* Amplification and overexpression of CACNA1E correlates with relapse in favorable histology Wilms' tumors. *Clin. Cancer Res.* **12**, 7284–7293 (2006).
7. McIntyre A. *et al.* Amplification and overexpression of the KIT gene is associated with progression in the seminoma subtype of testicular germ cell tumors of adolescents and adults. *Cancer Res.* **65**, 8085–8089 (2005).
8. Fix A. *et al.* Characterization of amplicons in neuroblastoma: high-resolution mapping using DNA microarrays, relationship with outcome, and identification of overexpressed genes. *Genes Chromosom. Cancer* **47**, 819–834 (2008).
9. Weber A, Imisch P, Bergmann E & Christiansen H. Coamplification of DDX1 correlates with an improved survival probability in children with MYCN-amplified human neuroblastoma. *J. Clin. Oncol.* **22**, 2681–2690 (2004).
10. Kaneko S., Ohira M, Nakamura Y, Isogai E, Nakagawara A & Kaneko M. Relationship of DDX1 and NAG gene amplification/overexpression to the prognosis of patients with MYCN-amplified neuroblastoma. *J. Cancer Res. Clin. Oncol.* **133**, 185–192 (2007).
11. De Preter K. *et al.* No evidence for correlation of DDX1 gene amplification with improved survival probability in patients with MYCN-amplified neuroblastomas. *J. Clin. Oncol.* **23**, 3167–3168 (2005).
12. Rodriguez S. *et al.* Expression profile of genes from 12p in testicular germ cell tumors of adolescents and adults associated with i(12p) and amplification at 12p11.2-p12.1. *Oncogene* **22**, 1880–221891 (2003).
13. Zafarana G. *et al.* 12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH. *Oncogene* **22**, 7695–7701 (2003).
14. Bourdon V. *et al.* Genomic and expression analysis of the 12p11-p12 amplicon using EST arrays identifies two novel amplified and overexpressed genes. *Cancer Res.* **62**, 6218–6223 (2002).
15. Ormandy CJ, Musgrove EA, Hui R., Daly RJ & Sutherland RL. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. *Breast Cancer Res. Treat.* **78**, 323–335 (2003).
16. Albertson DG. Gene amplification in cancer. *Trends Genet.* **22,** 447–455 (2006).
17. Cordon-Cardo C. *et al.* Molecular abnormalities of mdm2 and p53 genes in adult soft tissue sarcomas. *Cancer Res.* **54**, 794–799 (1994).
18. Hayward NK. Genetics of melanoma predisposition. *Oncogene* **22**, 3053–3062 (2003).
19. Feber A. *et al.* Amplification and overexpression of E2F3 in human bladder cancer. *Oncogene* **23**, 1627–1630 (2004).
20. Oeggerli M. *et al.* E2F3 amplification and overexpression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer. *Oncogene* **23**, 5616–5623 (2004).
21. Kao J. & Pollack JR. RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. *Genes Chromosom. Cancer* **45**, 761–769 (2006).
22. Comoglio PM, Giordano S. & Trusolino L. Drug development of MET inhibitors: targeting oncogene addiction and expedience. *Nature Rev. Drug Discov.* **7**, 504–516 (2008).
23. Zender L. *et al.* Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125**, 1253–1267 (2006).
24. Vogel CL. *et al.* Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J. Clin. Oncol.* **20**, 719–726 (2002).
25. Mass RD. *et al.* Evaluation of clinical outcomes according to HER2 detection by fluorescence *in situ* hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin. Breast Cancer* **6**, 240–246 (2005).
26. Ray ME. *et al.* Genomic and expression analysis of the 8p11–12 amplicon in human breast cancer cell lines. *Cancer Res.* **64**, 40–47 (2004).
27. Garcia MJ. *et al.* A 1 Mb minimal amplicon at 8p11–12 in breast cancer identifies new candidate oncogenes. *Oncogene* **24**, 5235–5245 (2005).
28. Reis-Filho JS. *et al.* FGFR1 emerges as a potential therapeutic target for lobular breast carcinomas. *Clin. Cancer Res.* **12**, 6652–6662 (2006).
29. Campbell PJ. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
30. Sugawa N., Ekstrand AJ, James CD & Collins VP. Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas. *Proc. Natl Acad. Sci. USA* **87**, 8602–8606 (1990).
31. Weber-Hall S. *et al.* Novel formation and amplification of the PAX7-FKHR fusion gene in a case of alveolar rhabdomyosarcoma. *Genes Chromosom. Cancer* **17**, 7–13 (1996).
32. Nikolsky Y. *et al.* Genome-wide functional synergy between amplified and mutated genes in human breast cancer. *Cancer Res.* **68**, 9532–9540 (2008).

Acknowledgements

The authors thank the National Cancer Research Institute, the Grand Charity of Freemasons, the Prostate Cancer Charity, the Bob Champion Cancer Trust, the Prostate Cancer Research Foundation, and the Wellcome Trust for Support.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

AKT2 | *AKT3* | *AR* | *BIRC2* | *CACNA1E* | *CCND1* | *CCND2* | *CDC6* | *CDK4* | *CDK6* | *E2F3* | *EGFR* | *EMSY* | *ERBB2* | *FGFR1* | *GRB7* | *KIT* | *MDM2* | *MET* | *MYC* | *MYCL1* | *MYCN* | *PAK1* | *PIK3CA* | *PLA2G10* | *REL* | *RUVBL1* | *SHC1* | *STARD3* | *TRIB1* | *WHSC1L1* | *YAP1* | *YWHAB* | *YWHAQ*

FURTHER INFORMATION

Cancer Genetics Programme at the Sanger Centre: http://www.sanger.ac.uk/genetics/CGP/Census/

Institute for Cancer Research Amplified and Overexpressed Genes in Cancer (AOGIC) dataset: http://www.amplicon.icr.ac.uk

SUPPLEMENTARY INFORMATION

See online article: S1 (text) | S2 (table) | S3 (text)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF