Recurrent inactivation of *STAG2* in bladder cancer is not associated with aneuploidy

Cristina Balbás-Martínez[1], Ana Sagrera[1,19], Enrique Carrillo-de-Santa-Pau[1,19], Julie Earl[1,2], Mirari Márquez[3], Miguel Vazquez[4], Eleonora Lapi[1], Francesc Castro-Giner[5], Sergi Beltran[5], Mònica Bayés[5], Alfredo Carrato[2], Juan C Cigudosa[6], Orlando Domínguez[7], Marta Gut[5], Jesús Herranz[3], Núria Juanpere[8], Manolis Kogevinas[9–12], Xavier Langa[1], Elena López-Knowles[10], José A Lorente[13], Josep Lloreta[8,14], David G Pisano[15], Laia Richart[1], Daniel Rico[4], Rocío N Salgado[6], Adonina Tardón[16], Stephen Chanock[17], Simon Heath[5], Alfonso Valencia[4], Ana Losada[18], Ivo Gut[5], Núria Malats[3] & Francisco X Real[1,14]

Urothelial bladder cancer (UBC) is heterogeneous at the clinical, pathological and genetic levels. Tumor invasiveness (T) and grade (G) are the main factors associated with outcome and determine patient management[1]. A discovery exome sequencing screen (*n* = 17), followed by a prevalence screen (*n* = 60), identified new genes mutated in this tumor coding for proteins involved in chromatin modification (*MLL2*, *ASXL2* and *BPTF*), cell division (*STAG2*, *SMC1A* and *SMC1B*) and DNA repair (*ATM*, *ERCC2* and *FANCA*). *STAG2*, a subunit of cohesin, was significantly and commonly mutated or lost in UBC, mainly in tumors of low stage or grade, and its loss was associated with improved outcome. Loss of expression was often observed in chromosomally stable tumors, and *STAG2* knockdown in bladder cancer cells did not increase aneuploidy. *STAG2* reintroduction in non-expressing cells led to reduced colony formation. Our findings indicate that *STAG2* is a new UBC tumor suppressor acting through mechanisms that are different from its role in preventing aneuploidy.

The most commonly mutated oncogene in UBC is *FGFR3* (50–60% of cases): mutations are more frequent in non-muscle-invasive bladder cancers (NMIBCs) with a low risk of progression (stage Ta low-grade tumors), here designated 'non-aggressive' (Online Methods)[2,3]. *PIK3CA* mutations occur in 15–20% of tumors and tend to associate with *FGFR3* mutations[4]. p53 and RB pathway inactivation has been associated with NMIBCs with a high risk of progression (stage Ta or T1 high-grade tumors) and with muscle-invasive bladder cancer (MIBC) (here designated 'aggressive')[5,6]. *RAS* mutations are less common and are mutually exclusive with *FGFR3* mutations[7]. There is now extensive evidence indicating that NMIBCs of high grade are genomically similar to MIBCs[8,9]: non-aggressive UBCs are genomically stable, whereas aggressive UBCs are genomically unstable[2,8–10]. Recently, exome sequencing has identified chromatin remodeling as an important pathway involved in UBCs[11]; this study focused mainly on MIBCs.

To discover new genes mutated in UBC, we sequenced the exomes of 17 tumors of variable stage and grade and corresponding normal leukocyte DNA; all neoplastic samples used had a tumor cellularity of ≥70% (**Supplementary Table 1**). Because there are major initiatives for the sequencing of MIBC (for example, The Cancer Genome Atlas (TCGA) project), we have focused mainly on NMIBC. Metrics for enrichment and depth of coverage are shown in **Supplementary Table 2**: the mean coverage for tumors and leukocytes was 79 ± 16 × and 82 ± 18 ×, respectively. We identified 2,927 somatic mutations, of which 1,263 and 798 were predicted to be relevant (nonsynonymous) and damaging (have a functional effect) (**Supplementary Table 3**), respectively (Online Methods). The average number of somatic mutations per tumor was 169 ± 114, with wide interindividual variation (range of 4–360 mutations) (**Fig. 1a**), a figure that falls in the midrange for exome studies in solid tumors in adults. C>T transitions were the most common nucleotide substitution (mean of 44%), followed by C>G transversions (**Fig. 1b**). The ratio of

[1]Epithelial Carcinogenesis Group, Molecular Pathology Programme, CNIO (Spanish National Cancer Research Centre), Madrid, Spain. [2]Servicio de Oncología Médica, Hospital Ramón y Cajal, Madrid, Spain. [3]Genetic and Molecular Epidemiology Group, Human Cancer Genetics Programme, CNIO (Spanish National Cancer Research Centre), Madrid, Spain. [4]Structural Computational Biology Group, Structural Biology and Biocomputing Programme, CNIO (Spanish National Cancer Research Centre), Madrid, Spain. [5]Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain. [6]Molecular Cytogenetics Group, Human Cancer Genetics Programme, CNIO (Spanish National Cancer Research Centre), Madrid, Spain. [7]Biotechnology Programme, CNIO (Spanish National Cancer Research Centre), Madrid, Spain. [8]Department of Pathology, Hospital del Mar–Parc de Salut Mar, Barcelona, Spain. [9]Centre de Recerca d'Epidemiologia Ambiental, Barcelona, Spain. [10]Institut Municipal d'Investigació Médica (IMIM)–Institut de Recerca Hospital del Mar, Barcelona, Spain. [11]Centro de Investigación Biomédica en Red (CIBER) Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. [12]National School of Public Health, Athens, Greece. [13]Urology Service, Hospital del Mar–Parc de Salut Mar, Barcelona, Spain. [14]Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain. [15]Bioinformatics Unit, Structural Biology and Biocomputing Programme, CNIO (Spanish National Cancer Research Centre), Madrid, Spain. [16]Departamento de Medicina, Departamento de Medicina, Universidad de Oviedo, Oviedo, Spain. [17]Translational Genomics Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA. [18]Chromosome Dynamics Group, Molecular Oncology Programme, CNIO (Spanish National Cancer Research Centre), Madrid, Spain. [19]These authors contributed equally to this work. Correspondence should be addressed to F.X.R. (preal@cnio.es).

Received 17 February; accepted 16 September; published online 13 October 2013; doi:10.1038/ng.2799

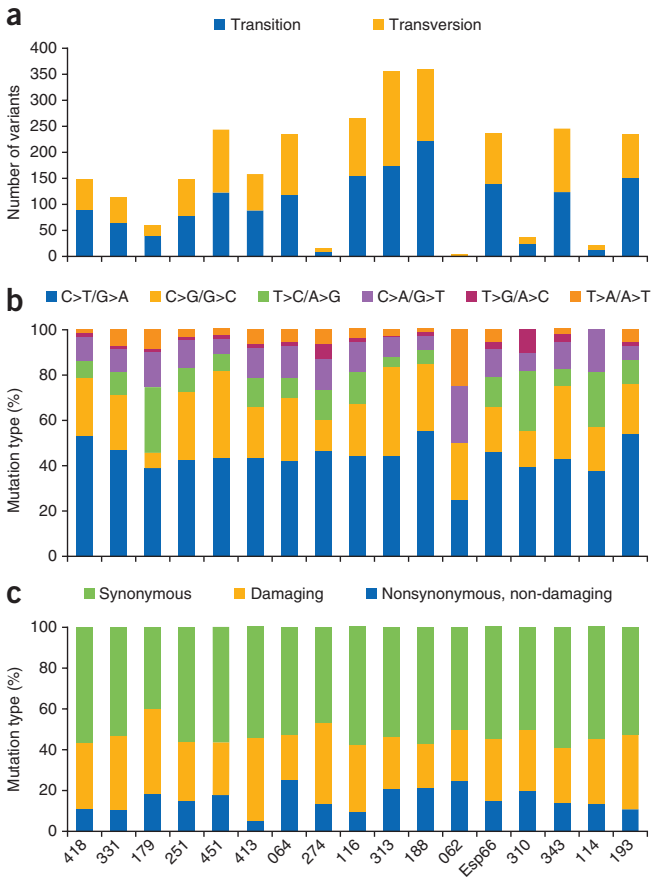

Figure 1 Distribution of SNVs identified in the discovery screen through exome sequencing. (**a**–**c**) Shown for each tumor are the total number of SNVs (**a**) categorized according to type of nucleotide substitution (**b**) and predicted effect (**c**).

nonsynonymous to synonymous (NS:S) changes was <1 in 15 of 17 samples (**Fig. 1c** and **Supplementary Fig. 1a**). We compared the total number of single-nucleotide variants (SNVs), indels, transitions, transversions, synonymous mutations, non-damaging nonsynonymous mutations and damaging nonsynonymous mutations in aggressive versus non-aggressive tumors; all variables were highly similar in both tumor groups. We performed the same analysis according to smoking status: the number of damaging mutations was higher in tumors from smokers than in those from non-smokers, but differences did not reach statistical significance ($P = 0.09$). The number of mutations in tumors from individuals who were >60 years old was also slightly but not significantly higher than for younger individuals (**Supplementary Fig. 2**). The NS:S ratio was similar regardless of tumor aggressiveness and smoking status, but this ratio was slightly lower in individuals diagnosed at >60 years of age (**Supplementary Fig. 1b**). It will be necessary to sequence more tumors to further investigate these relationships.

We assessed the reliability of the exome analysis and strategies for somatic variant calling using Sanger sequencing: we assayed 226 variants and verified 219, of which 214 (94.7%) were confirmed to be somatic (**Supplementary Table 3**). The list of genes with nonsynonymous mutations in ≥3 tumors that were expressed in >30% of UBCs, on the basis of Affymetrix expression analyses of an independent tumor sample series ($n = 43$) covering the full spectrum of the disease, is shown in **Table 1**. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses identified chromatin modification, DNA repair and DNA damage response, apoptosis and the cell cycle among the most significant processes to which mutated genes were ascribed (false discovery rate (FDR) < 0.1) (**Supplementary Table 4**).

To extend our findings, we performed a mutation prevalence screen including mainly NMIBCs ($n = 60$ tumors) (**Supplementary Table 5**) using the HaloPlex Target Enrichment System followed by sequencing. We included selected genes that were recurrently mutated in the discovery screen as well as additional genes from the pathways in which these genes participate (**Supplementary Table 6**). We identified 260 SNVs: 200 were predicted to be relevant, and 143 of these were predicted to be damaging. We analyzed 95 mutations identified by HaloPlex by Sanger sequencing; 73 were verified (76.13%), and 72 of these (98.6%) were confirmed to be somatic.

The joint distribution of mutations in the discovery and prevalence screens is shown in **Figure 2** and **Table 1**. Among the genes recurrently mutated, we identify new genes involved in chromatin remodeling different from those reported by Gui *et al.*[11] (*MLL2*, *ASXL2* and *BPTF*). The BPTF protein binds histone H3 that is trimethylated at lysine 4 (H3K4me3) and has been found to be mutated in hepatocarcinoma[12], but little is known about its function in cancer. We show that *BPTF* knockdown led to a marked reduction in colony formation in the three UBC lines tested (**Supplementary Fig. 3**), suggesting that it has a role in cancer cell proliferation. We confirmed recurrent mutations in *ARID1A*, *KDM6A* (*UTX*), *CREBBP*, *EP300*, *MLL* and *MLL3* (ref. 11), in agreement with recent reports implicating mutations in a wide range of chromatin remodelers in human cancer[12–15]. Notably, we identified recurrent, previously unreported somatic mutations in genes involved in DNA repair (*ATM*, *ERCC2* and *FANCA*, among others) and in the cohesin subunits *STAG2*, *STAG1*, *SMC1A* and *SMC1B*, indicating that these pathways have an important role in UBC. *FGFR3*, *TP53*, *PIK3CA* and *RB1* are among the recurrently mutated genes, providing evidence of the representativeness of the tumors analyzed.

We have focused on *STAG2* because it was significantly mutated in our exomes (**Table 1**); an additional mutation was found in 9 published UBC exomes[11], and we identified 2 mutations in 21 MIBC exomes from the TCGA project (overall damaging mutation rate of 13% (6/47)). Our prevalence screen identified nine additional somatic mutations predicted to be damaging. Altogether, we have identified damaging somatic *STAG2* mutations in 12 of 77 tumors (15.6%; 5 nonsense, 4 exon junction, 2 missense and 1 indel) (**Supplementary Fig. 4** and **Supplementary Table 7**), and 9 of 11 were verified by Sanger sequencing (**Supplementary Fig. 5**). Damaging mutations were found in both non-aggressive (6/29; 20.7%) and aggressive (5/47; 10.6%) tumors. *STAG2*-inactivating mutations leading to loss of protein expression have recently been reported in non-epithelial tumors[16]. STAG2 expression was low or undetectable in 6 of 7 (85%) UBCs with damaging mutations and in 3 of 34 (9%) tumors with wild-type *STAG2* ($P = 0.0001$) (**Fig. 3**, **Supplementary Fig. 5** and **Supplementary Table 7**). Together with the exome significance analysis, these data indicate that *STAG2* is a new gene commonly mutated in UBC.

We then analyzed STAG2 expression in tissue microarrays of incident tumors representative of the disease spectrum (**Supplementary Tables 8** and **9**)[3,17]. Loss of STAG2 expression in tumors, defined as a histoscore of ≤50 with detectable stromal expression, was observed in 197 of 671 tumors (29.3%) (**Fig. 3** and **Supplementary Fig. 5**). Loss of STAG2 expression was significantly associated with multicentricity ($P = 0.011$), tumor size ($P = 0.002$), low stage ($P = 5.7 \times 10^{-15}$) and low grade ($P = 1.96 \times 10^{-15}$) (**Supplementary Table 10**). Abnormal

Table 1 Genes frequently mutated in UBC assessed through exome sequencing or targeted HaloPlex resequencing ($n = 77$)

Gene	Number of mutations ($n = 17$)	P value[a]	Number of mutations ($n = 60$)	Number of mutations in all tumors ($n = 77$)	Number of non-aggressive mutant cases ($n = 29$)[b]	Number of aggressive mutant cases ($n = 47$)[b]	P value[c]
ARID1A	7	0.0001	3	10	3	7	0.732
STAG2	3	0.019	9	12	6	5	0.315
KDM6A	4	0.019	6	10	3	7	0.732
PDZD2	3	0.019	0	3	0	2	0.521
MYCBP2	3	0.061	2	5	2	2	0.999
LPHN3	3	0.096	0	3	0	2	0.521
CREBBP	2	0.098	9	11	4	7	1
EP300	2	0.098	5	7	3	4	1
ATM	3	0.138	6	9	4	4	0.702
TP53	3	0.2117	8	11	2	9	0.188
RREB1	3	0.237	0	3	1	1	1
PIK3CA	6	0.239	4	10	5	4	0.289
WHSC1L1	2	0.241	1	3	1	2	1
MYO5B	3	0.430	0	3	0	2	0.521
MLL2	2	0.636	13	15	6	5	0.315
FGFR3	2	0.659	12	14	10	4	0.011
TEX15	3	0.778	0	3	0	2	0.521
BRAF	1	NA	6	7	2	5	0.701
ERCC2	0	NA	8	8	5	1	0.040
MAPK8IP3	1	NA	4	5	3	2	0.363
MLL	1	NA	5	6	3	2	0.363
NUP93	1	NA	4	5	1	3	1
STAG1	0	NA	5	5	0	3	0.282
RB1	1	NA	3	4	1	3	1
FANCA	0	NA	4	4	0	3	0.282
MLL3	1	NA	4	5	2	3	1
NOTCH1	0	NA	4	4	0	3	0.282
ASXL2	3	NA	1	4	1	1	1

Data are shown for the discovery screen and the prevalence screen. Discrepancies between number of mutations and numbers of mutant tumors result from the occurrence of ≥2 mutations in the same gene in a given tumor sample. NA, not available.
[a]*P*-value calculations are based on the mutations identified in the discovery screen (Online Methods). [b]One sample with a mutation in *STAG2* is excluded from the non-aggressive versus aggressive tumor comparison owing to insufficient information for classification. [c]*P* value for the frequency of mutant tumors with non-aggressive versus aggressive features.

STAG2 expression patterns included focal losses within otherwise positive tumors and a predominant cytoplasmic distribution of the protein (**Supplementary Fig. 6**). Because non-aggressive tumors are more differentiated, lack of STAG2 might reflect urothelial cell maturation. Arguing against this possibility, STAG2 expression was detected in all cell layers of the normal urothelium (**Supplementary Fig. 7**). We also analyzed whether mechanisms other than mutation might account for loss of STAG2 expression. Using SNP arrays, we found *STAG2* losses in 1 of 18 TaG1 or TaG2 STAG2-negative tumors (5%) from male subjects. Similar findings have been reported in leukemia[16,18–20] (**Supplementary Fig. 8**).

STAG2 encodes a subunit of cohesin, a complex that mediates sister chromatid cohesion to ensure accurate chromosome segregation and DNA repair. Cohesin also regulates gene expression through mechanisms involving DNA looping and interactions with transcriptional regulators such as Mediator and CTCF[21,22]. Somatic human cells contain two versions of this complex consisting of SMC1A, SMC3, RAD21 and either STAG2 or STAG1 (ref. 21). Although there is still an incomplete understanding of the functional redundancy of these complexes, differential roles in centromeric versus telomeric cohesion have been proposed, and STAG1 has been preferentially implicated in transcriptional control[23,24]. Using immunohistochemistry, we showed that STAG1 is expressed in the normal urothelium and in the majority of UBCs, including most tumors lacking STAG2 expression (**Supplementary Figs. 7** and **9**). Interestingly, all six tumors that lost both STAG1 and STAG2 expression were of high grade and were wild type for *FGFR3*, suggesting partial functional compensation in the maintenance of integral chromosome segregation machinery. The more frequent loss of the expression of STAG2 in tumors (**Supplementary Fig. 9a**) may reflect the fact that only one hit is required for its inactivation, given its location on the X chromosome.

Recently, *STAG2* mutations in glioblastoma, melanoma and Ewing sarcoma have been proposed to participate in tumor development by promoting aneuploidy[16]. This hypothesis is at odds with our finding

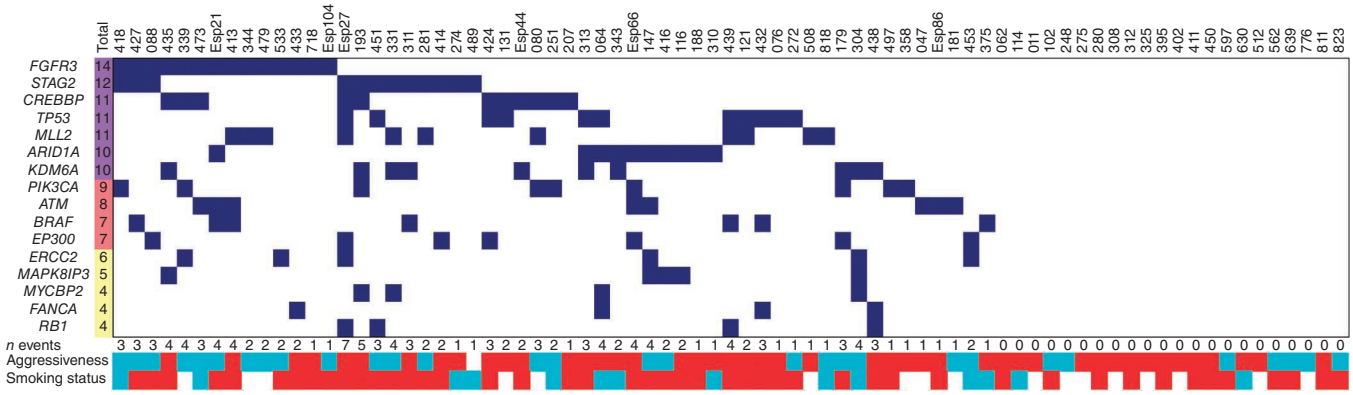

Figure 2 Distribution of mutations in genes recurrently mutated in UBC that are expressed in >30% of tumors: joint analysis of the discovery and prevalence screens. In 22 of 77 tumors (28.6%), none of the genes listed were found to be mutated. For aggressiveness, red indicates aggressive tumors, and blue indicates non-aggressive tumors. Aggressiveness was defined as described in the Online Methods. For smoking status, red indicates smokers, and blue indicates non-smokers. White squares indicate cases where information was not available.

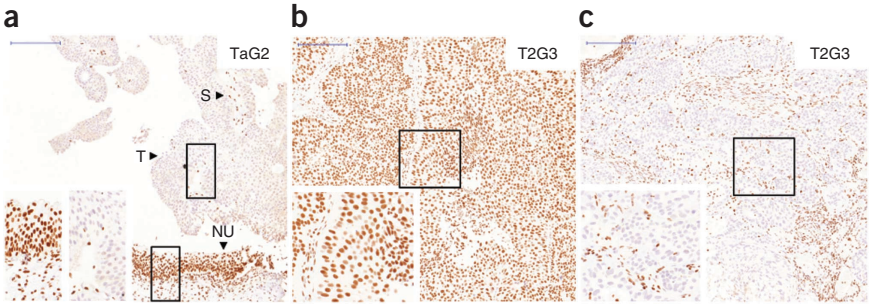

Figure 3 Immunohistochemical analysis of STAG2 expression in bladder tumors of different stage and grade. (**a**–**c**) A TaG2 tumor (**a**) and a T2G3 tumor (**c**) show lack of STAG2 expression, and one T2G3 tumor shows strong expression (**b**). Strong STAG2 protein expression was found in the normal urothelium of a subject with a STAG2-negative tumor (**a**) and in the stroma of all tumor samples. T, tumor; S, stroma; NU, normal urothelium. Scale bars, 200 µm. Boxes highlight areas shown at higher magnification (×2) at the lower left.

that loss of STAG2 expression occurs mainly in non-aggressive UBCs that are genomically stable. To address this issue, we analyzed chromosome number changes in a panel of 23 TaG1 and TaG2 tumors using high-resolution SNP or BAC arrays. Of 11 tumors without STAG2 expression, 9 lacked aneuploidy, and 2 showed loss of 1 copy of chromosome 9; similarly, 9 of 12 tumors that expressed STAG2 had normal chromosomal content (**Fig. 4a**, **Supplementary Fig. 10** and **Supplementary Table 11**). Consistent with these findings, mutations in *STAG2* and other cohesin genes have recently been reported not to be associated with aneuploidy in acute myeloid leukemia[18–20]. We next knocked down *STAG2* in three UBC lines displaying a broad range of phenotypes. Efficient knockdown was achieved in the three lines, but there were no consistent effects on chromosome number at metaphase (**Fig. 4b**,**c**, **Supplementary Fig. 11** and **Supplementary Table 12**), unlike what was previously reported in HCT116 cells[16]. This discrepancy may reflect the fact that different cell types have variable tolerance to aneuploidy. We also introduced *STAG2* cDNA in three cell lines lacking STAG2 protein expression: UM-UC-6 cells harbor a p.Arg305* stop-gain alteration (exon 11) and a p.Phe1228Leu alteration (exon 33), VM-CUB-3 cells harbor a 10-bp deletion in exon 6 and LGW0 1 G600 cells have wild-type sequence in exons 3–35. In the three cell lines, we observed a significant decrease in colony formation upon *STAG2* lentiviral expression (**Fig. 4d** and **Supplementary Fig. 12**). Intriguingly, *STAG2* knockdown was also associated with reduced colony formation in five different cell lines (**Supplementary Fig. 13**). Similar effects have been reported with knockdown of the tumor suppressor *ARID1A* in pancreatic and bladder cancer cells[25,26].

To place these findings in the context of the known pathways of UBC progression, we assessed the association of STAG2 alterations with *FGFR3* mutation or overexpression, p53 nuclear accumulation and Ki67 expression (**Supplementary Tables 13–16**)[2]. In NMIBC, loss of STAG2 was significantly more common in tumors with mutant *FGFR3* (42.7% versus 27.2%; $P = 0.001$), tumors lacking p53 overexpression ($P = 0.002$) and tumors with a low Ki67 index ($P = 0.002$). These results indicate that loss of STAG2 is associated with less aggressive tumors. Within the low-risk NMIBC subgroup, loss of STAG2 expression was associated with *FGFR3* mutant status ($P = 0.059$) and with low p53 expression ($P = 0.011$). In individuals with high-risk NMIBC, loss of STAG2 expression was associated with high FGFR3 expression ($P = 0.037$), *FGFR3* mutation ($P = 0.12$) and a low Ki67 index ($P = 0.049$). In both high-risk NMIBC and MIBC, there was no association with p53 expression as detected by immunohistochemistry (**Supplementary Tables 13–16**).

We then analyzed the association of loss of STAG2 expression with recurrence and progression among individuals with NMIBC and with progression and mortality in individuals with MIBC. We applied both Kaplan-Meier curves and multivariable Cox regression analyses. The large sample size of our study allowed us to perform a more informative stratified analysis. Loss of STAG2 expression was associated with lower risk of tumor recurrence and progression in individuals with NMIBC (**Fig. 5a**,**b**). However, in multivariable analyses, STAG2 expression was not an independent predictor of recurrence or progression after adjusting for stage, grade and *FGFR3* mutation status (**Supplementary Tables 17** and **18**), as these parameters were highly correlated. In individuals with MIBC, loss of STAG2 expression was associated with lower risk of progression

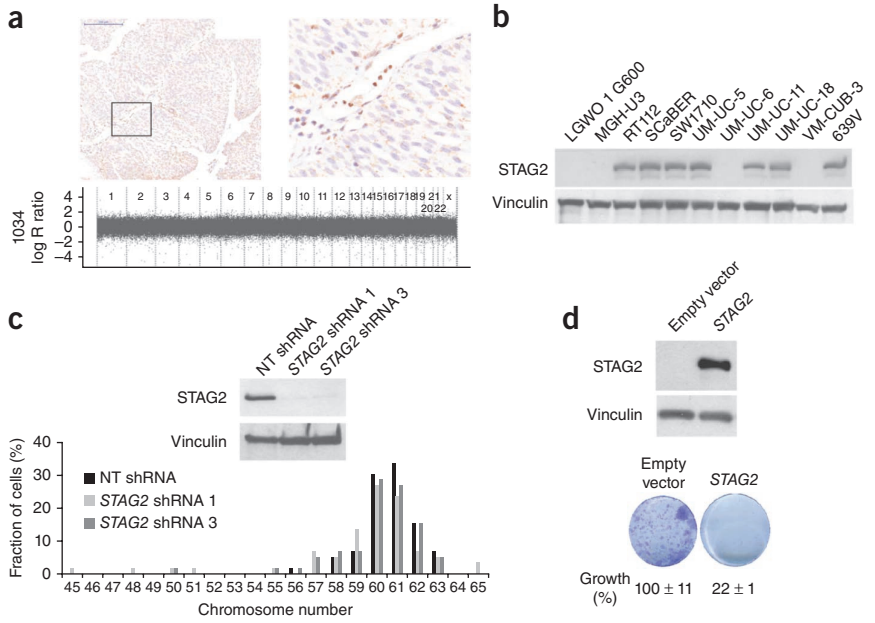

Figure 4 Loss of STAG2 expression is not associated with aneuploidy in primary tumors and in 639V bladder cancer cells. Effects of *STAG2* reconstitution on cell growth. (**a**) SNP array genomic plots show lack of chromosomal changes (aneuploidy) in a TaG1 tumor lacking STAG2 expression (note strong STAG2 labeling of normal stroma). Scale bar, 200 µm. The box highlights the region shown at higher magnification (×400) to the right. (**b**) Protein blotting analysis of STAG2 in UBC lines shows undetectable expression in 4 of the 11 lines used for functional studies. (**c**) Efficient *STAG2* knockdown in 639V cells demonstrated by protein blotting does not lead to consistent changes in chromosome number (quantification shown in **Supplementary Table 12**). NT shRNA, non-targeting short hairpin RNA. (**d**) *STAG2* overexpression in UM-UC-6 cells demonstrated by protein blotting leads to significantly reduced ($P < 0.005$) colony formation efficiency.

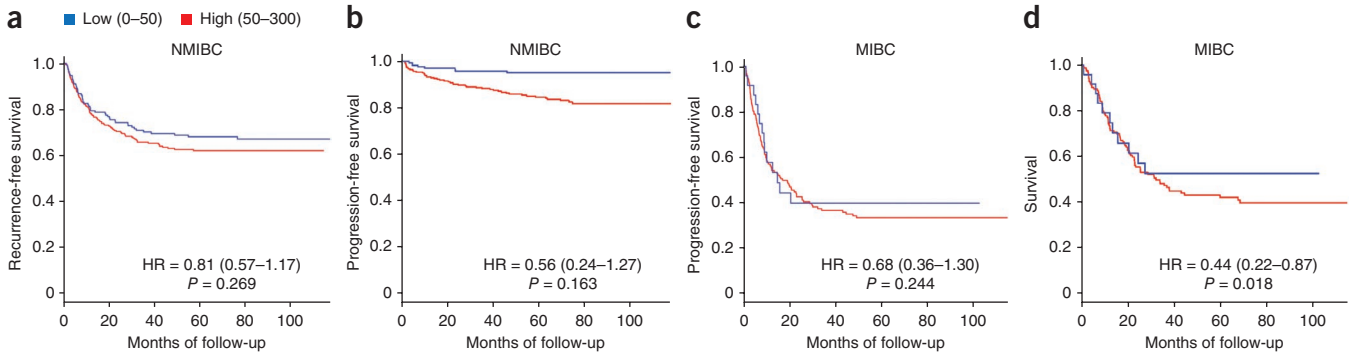

Figure 5 Kaplan-Meier plots of the association of STAG2 expression with outcome in individuals with UBC. (**a**) Recurrence in subjects with NMIBC with high STAG2 expression ($n = 309$) versus low STAG2 expression ($n = 171$). (**b**) Progression in individuals with NMIBC with high STAG2 expression ($n = 309$) versus low STAG2 expression ($n = 171$). (**c**) Progression in individuals with MIBC with high STAG2 expression ($n = 158$) versus low STAG2 expression ($n = 24$). (**d**) Cancer-specific survival of individuals with MIBC with high STAG2 expression ($n = 158$) versus low STAG2 expression ($n = 24$). Expression was defined by histoscore, calculated as described in the Online Methods. *P* values correspond to results from multivariable analysis. Details on results and variables used for adjustment are given in **Supplementary Tables 17–20**.

(hazards ratio (HR) = 0.68; $P = 0.244$), and it was an independent predictor of survival in the multivariable analysis (HR = 0.44; $P = 0.018$) (**Fig. 5c**,**d** and **Supplementary Tables 19** and **20**). Therefore, we conclude that loss of STAG2 expression is associated with better prognosis in individuals with both NMIBC and in MIBC; additional studies are required to determine its clinical value.

In summary, we find that both previously reported and newly identified genes coding for proteins involved in chromatin modification are recurrently mutated in UBC. In addition, we identify mutations in genes involved in the cell cycle, DNA repair and the regulation of apoptosis. The frequent alteration of genes in these pathways may provide opportunities for novel therapies, including those based on synthetic lethality. *STAG2* is significantly mutated in UBC; mutations and loss of expression are common, particularly in tumors of low stage and grade, and are associated with improved clinical outcome. In non-aggressive tumors, STAG2 alterations occur in the absence of chromosomal instability. Our findings strongly suggest that *STAG2* is a new tumor suppressor in UBC through mechanisms that are distinct from its role in cohesion to prevent aneuploidy.

URLs. FAR, http://sourceforge.net/projects/flexbar/; R software, http://www.r-project.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequencing data have been deposited in the Sequence Read Archive (SRA) database. Whole-exome sequencing data are accessible under SRP029936, and validation data are accessible under SRP029935. SNP array data have been deposited in the Gene Expression Omnibus (GEO).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank F. Algaba, Y. Allory, A. Cuadrado, C. González, E. López, P. Lapunzina, T. Lobato, M. Malumbres, S. Remeseiro, V.J. Sánchez-Arévalo, F. Waldman and the CNIO core facilities for valuable contributions. We also thank TCGA investigators for providing unpublished information for analysis. This work was supported, in part, by grants from Ministerio de Economia y Competitividad, Madrid (grants Consolíder ONCOBIO, Consolider INESGEN, SAF-2010-21517 and SAF2011-15934-E), Instituto de Salud Carlos III (grants G03/174, 00/0745, PI051436, PI061614, G03/174, PI080440, PI120425 and Red Temática de Investigación Cooperativa en Cáncer (RTICC)), Asociación Española Contra el Cáncer, EU-FP7-201663 and US National Institutes of Health grant RO1 CA089715. C.B.-M. is the recipient of a La Caixa International PhD Fellowship. E.L. is supported by a grant from the Fundación Banco Santander Postdoctoral Programme.

AUTHOR CONTRIBUTIONS

C.B.-M., E.L. and L.R. designed and performed *in vitro* functional studies. C.B.-M. performed immunohistochemical analysis of tumor samples. A.S. and C.B.-M. designed and performed mutation validation analyses. A.S. designed and prepared HaloPlex libraries for sequencing. E.C.-d.-S.-P., M.V., F.C.-G., S.B. and D.G.P. processed and analyzed exome sequencing and targeted resequencing data. J.E., E.L.-K., D.R. and S.C. performed gene copy number analyses of tumors. M.M. coordinated subject and sample data management. A.C., M.K., J.A.L. and A.T. contributed to subject recruitment and data collection. J.H. performed statistical analyses. X.L. provided technical support with subject samples. M.B. and M.G. coordinated library preparation and sequencing. O.D. contributed to library preparation and sequencing. J.C.C. and R.N.S. contributed to the *in vitro* analysis of the effects of *STAG2* knockdown on aneuploidy. N.J. and J.L. performed pathological review of samples. I.G. coordinated exome sequencing and targeted resequencing. I.G., S.H. and A.V. supervised bioinformatics analyses. A.L. provided scientific insight and contributed with reagents. N.M. coordinated subject recruitment and the collection of clinical and pathological data and supervised clinical-pathological-molecular association and outcome analyses. F.X.R. and N.M. conceived the study. F.X.R. supervised the overall way in which the study was conducted. F.X.R. wrote the manuscript with N.M., C.B.-M., A.S., E.C.-d.-S.-P., A.V., A.L. and I.G. contributed to manuscript writing.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Sylvester, R.J. Natural history, recurrence, and progression in superficial bladder cancer. *ScientificWorldJournal* **6**, 2617–2625 (2006).
2. Luis, N.M., Lopez-Knowles, E. & Real, F.X. Molecular biology of bladder cancer. *Clin. Transl. Oncol.* **9**, 5–12 (2007).
3. Hernández, S. *et al. FGFR3* mutations as a prognostic factor in non-muscle invasive urothelial bladder carcinomas: results of a prospective study. *J. Clin. Oncol.* **24**, 3664–3671 (2006).
4. López-Knowles, E. *et al. PIK3CA* mutations are an early genetic alteration associated with *FGFR3* mutations in superficial papillary bladder tumors. *Cancer Res.* **66**, 7401–7404 (2006).
5. Real, F.X. p53: it has it all, but will it make to the clinic as a marker in bladder cancer? *J. Clin. Oncol.* **25**, 5341–5344 (2007).

6. López-Knowles, E. *et al.* The p53 pathway and outcome among patients with T1G3 bladder tumors. *Clin. Cancer Res.* **12**, 6029–6036 (2006).
7. Jebar, A.H. *et al. FGFR3* and *Ras* gene mutations are mutually exclusive genetic events in urothelial cell carcinoma. *Oncogene* **24**, 5218–5225 (2005).
8. Lindgren, D. *et al.* Combined gene expression and genomic profiling define two intrinsic molecular subtypes of urothelial carcinoma and gene signatures for molecular grading and outcome. *Cancer Res.* **70**, 3463–3472 (2010).
9. Höglund, M. The bladder cancer genome; chromosomal changes as prognostic makers, opportunities, and obstacles. *Urol. Oncol.* **30**, 533–540 (2012).
10. Blaveri, E. *et al.* Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clin. Cancer Res.* **11**, 7012–7022 (2005).
11. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* **43**, 875–878 (2011).
12. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
13. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
14. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
15. Wilson, B.G. & Roberts, C.W. SWI/SNF nucleosome remodelers and cancer. *Nat. Rev. Cancer* **11**, 481–492 (2011).
16. Solomon, D.A. *et al.* Mutational inactivation of *STAG2* causes aneuploidy in human cancer. *Science* **333**, 1039–1043 (2011).
17. Amaral, A.F.S. *et al.* Plasma 25-hydroxyvitamin D3 levels and bladder cancer risk according to tumor stage and *FGFR3* status: a mechanism-based epidemiological study. *J. Natl. Cancer Inst.* **104**, 1897–1904 (2012).
18. Welch, J.S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
19. Walter, M.J. *et al.* Acquired copy number alterations in adult acute myeloid leukemia genomes. *Proc. Natl. Acad. Sci. USA* **106**, 12950–12955 (2009).
20. Walter, M.J. *et al.* Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1090–1098 (2012).
21. Nasmyth, K. & Haering, C.H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
22. Remeseiro, S. & Losada, A. Cohesin, a chromatin engagement ring. *Curr. Opin. Cell Biol.* **25**, 63–71 (2013).
23. Remeseiro, S. *et al.* Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. *EMBO J.* **31**, 2076–2089 (2012).
24. Remeseiro, S. *et al.* A unique role of cohesin-SA1 in gene regulation and development. *EMBO J.* **31**, 2090–2102 (2012).
25. Shain, A.H. *et al.* Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proc. Natl. Acad. Sci. USA* **109**, E252–E259 (2012).
26. Balbás-Martínez, C. *et al. ARID1A* alterations are associated with *FGFR3* wild type, poor-prognosis, urothelial bladder tumors. *PLoS ONE* **8**, e62483 (2013).

ONLINE METHODS

Subjects and samples. Subjects and samples came from the Epicuro/Spanish Bladder Cancer Study (SBCS)[3,27] and from the Integrated Study of Bladder Cancer (ISBLAC) (**Supplementary Tables 1** and **5**). STAG2 expression was analyzed using tissue microarrays containing tumors from the Epicuro/SBCS, including tumors from individuals with newly diagnosed UBC. Staging, grading and follow-up were performed as described[3,17]. Expert pathologists reviewed diagnostic slides from all tumor blocks. We categorized TaG1 and TaG2 tumors as low-risk NMIBC or non-aggressive; TaG3, T1G2 and T1G3 tumors were categorized as high-risk NMIBC; and ≥T2 tumors were categorized as MIBC. The latter two groups were pooled as aggressive tumors. Subject characteristics are summarized in **Supplementary Tables 8** and **9**. In subjects with NMIBC, recurrence was defined as the reappearance of NMIBC following a negative follow-up medical evaluation. Progression was defined as a transition from NMIBC to MIBC or the development of new local or metastatic tumors after primary treatment for individuals with MIBC. Median follow-up time was 62.6 months (range of 1–98 months). All deaths were recorded, but only UBC-related deaths were considered for survival analysis. Cases dying from other causes were censored at the time of death. Survival was computed as the period comprised between diagnosis and death or last control. All subjects provided written informed consent[3]. The ethics committees of all participating institutions approved both studies.

Exome sequencing, targeted resequencing, bioinformatic analyses and mutation verification. The Agilent SureSelect Human All Exon plus v3 50Mb (samples 114, 116, 193, 251, 310, 331, 413, 418 and Esp66) or v4 51Mb (samples 062, 064, 179, 188, 274, 313, 343 and 451) was used for library preparation and enrichment. Libraries were applied to an Illumina flow cell; sequencing was performed on HiSeq 2000 instruments using paired-end 75-bp reads.

Base calling and quality control were performed on the Illumina Real-Time Analysis pipeline. Sequence reads were trimmed to the first base with quality >10 that mapped to human genome build hg19 (GRCh37) using GEM, allowing ≤4 mismatches. Reads not mapped by Genome Multitool Mapper (GEM)[27] (~4%) were submitted to a last round of mapping with BFAST. Results were merged; only uniquely mapping, non-duplicate read pairs were used. SAMtools suite version 0.1.18 with default settings was used to call SNVs and short indels. Variants on regions with low mappability, read depth <10, tail distance bias $P < 0.05$ or strand bias $P < 0.001$ were filtered out. Somatic mutations were called by comparing tumor and blood exomes; Fisher's exact tests were performed using variant-supporting read counts. Only variants with Fisher's exact test P value <0.0001 were considered.

All SNVs in exon junctions or that led to a nonsynonymous change were considered to be 'relevant'. SNVs that led to an amino acid substitution were evaluated using MutationAssessor[28] and SIFT[29] to predict their damaging effect; both scores were normalized into a range from 0 to 1. P values from SIFT were subtracted from 1. MutationAssessor predictions were scored as follows: high risk of damage was assigned 1, medium risk of damage was assigned 0.7 and low risk of damage was assigned 0.5. When both predictions were available, scores were averaged; if one prediction was missing, the other score was used. Variants with a final score of >0.8 were considered to be damaging. Stop gains and frameshifts were considered to be damaging if they ablated >30% of the sequence or a protein domain annotated in InterPro. Variants close to an exon boundary were considered to be relevant and damaging if the distance from the exon junction was eight bases into the intron or two bases into the exon of donor junctions or if the distance from the exon junction was eight bases into the intron or three bases into the exon of acceptor junctions. Scores from both methods were used as input to calculate P value for the associated genes. We used the Oncodrive-fm approach[30] that combines data on recurrence and functional impact (**Table 1**).

Statistical analyses within R software (2.15.1) were performed on stage, smoking status and age groups using Mann-Whitney U tests (**Fig. 1**). Fisher's exact tests (non-aggressive versus aggressive) were performed for recurrently altered genes (**Table 1**); P value <0.05 was considered to be statistically significant.

Pathway analysis was performed as reported[31]. First, a gene list was selected. We processed different combinations of three lists: (i) all relevant genes (those with mutations leading to nonsynonymous substitutions or affecting exon junctions); (ii) damaged genes (those with mutations predicted to be damaging); and (iii) recurrently altered genes (those with relevant mutations in ≥2 samples). The resulting lists were examined for enrichment in terms from Gene Ontology (biological process) and KEGG pathways. For the latter, pathways associated with diseases were filtered out, as reported[32]. Enrichment analysis was based on a hypergeometric test. P values were adjusted using Benjamini-Hochberg's FDR; only FDR < 0.1 was considered. A correction for genes in overlapping clusters was applied.

For targeted resequencing, the HaloPlex Target Enrichment System (Agilent) was used in an independent tumor series ($n = 60$) following the manufacturer's instructions. Five cases from the discovery screen were included for targeted resequencing. A library of genomic DNA fragments was created by digestion in eight restriction reactions and was hybridized with probes against target regions incorporating Illumina paired-end sequencing motifs and index sequences. DNA was captured and PCR amplified with KAPA HiFi HotStart polymerase or Herculase II Fusion Enzyme, and products were purified using AMPure XP beads. Amplicons were sequenced in multiplex using Illumina protocols. Adaptors and primers were removed from both ends of the reads with FAR. Trimmed reads were mapped; SNPs and indels were called as described above, without excluding duplicates and filtering out variants with tail distance bias $P < 0.05$ or strand bias $P < 0.001$. Variants outside of the regions selected for enrichment, with low mappability, with read depth of <10, occurring in >1% of the reads in blood or annotated in the 1000 Genomes Project as SNPs (release 20110521) were filtered out. Variants with Fisher's exact test P value for somatic comparison <0.0001 were considered. Somatic mutations were called by comparing tumor and blood samples. Damaging and effect annotations were performed as described for exomes.

Mutations were verified by Sanger sequencing. Using this bioinformatics pipeline, we verified 94.7% of the SNVs called as somatic mutations by exome sequencing.

***STAG2* sequencing of UBC lines.** Cell lines were obtained from the American Type Culture Collection or from the investigators who established them and were authenticated by gene mutation analyses; all cultures were mycoplasma free. Exons 3–7, 11–31 and 33–35 of *STAG2* were sequenced from overlapping amplicons generated from cDNA; exons 8–10 were sequenced from genomic DNA. The predominant transcript lacked exon 32 (Ensembl variant STAG2-0001) and coded for a 1,231-residue protein.

Immunohistochemistry. STAG2 was detected using clone J-12 (0.5 µg/ml) (Santa Cruz Biotechnology, sc-81852) and affinity-purified rabbit polyclonal antibodies (0.5 µg/ml) raised against a synthetic peptide (DPASIMDESVLGVSMF)[23]. Both antibodies yielded concordant results in 92% of tumors. To detect STAG1, we used affinity-purified rabbit polyclonal antibodies (2 µg/ml) raised against a synthetic peptide (EDDSGFGMPMF)[23]. Antibodies D5/16B4 (1:2,000 dilution) and Ks20.8 (1:50 dilution) detecting KRT5/KRT6 and KRT20, respectively, were from Dako. Antigen retrieval and reactions were performed as described[6,17]. A histoscore was calculated as the product of the staining intensity (0–3) and the percentage of positive cells (0–100%). Unsupervised clustering analysis was performed using scores and the heatmap.2 function of the gplots package within the R 2.15.1 statistical environment.

Gene copy number analyses. Copy number changes were analyzed using manually microdissected fresh tissue samples containing ≥60% tumor cells ($n = 55$). DNA was hybridized to Illumina HumanHap 1M BeadChip SNP arrays; 20 tumors were TaG1 or TaG2. Copy number changes were called as described[33]. An additional 76 samples were analyzed using Human 2.0 BAC arrays (UCSF Cancer Center)[10,34].

STAG2 functional assays. To knock down *STAG2*, control or *STAG2*-targeting lentiviral particles were produced in HEK293T cells using Sigma Mission plasmids. Viral supernatants were used to infect RT112, UM-UC-5, 639V, SW1710 and UM-UC-11 cells; after three rounds of infection with shRNA-encoding lentiviruses, cells were selected for 48 h in medium containing puromycin (2 µg/ml). To overexpress *STAG2*, human cDNA (b isoform; 1,231 residues) was amplified by PCR (Addgene pEGFP-*STAG2* plasmid, ref. 31972) and subcloned into the pLVX-puro lentiviral vector. After three rounds of infection,

doi:10.1038/ng.2799

cells were selected for 48 h with puromycin (2 μg/ml). Protein blotting was performed as described[17].

For colony formation assays, $8x10^3$ puromycin-selected cells were seeded; 7 days later, cells were methanol-fixed and crystal violet-stained; after elution (10% acetic acid), 680 nm absorbance was measured.

For chromosome analyses, puromycin-selected cells with knockdown were arrested with colcemid (0.1 mg/ml) for 6 h, collected, swollen in 75 mM KCl for 15 min (RT112), 25 min (639V) or 30 min (UM-UC-11) at 37 °C and fixed. Images of metaphase were captured, and chromosomes per metaphase were counted (Axioplan II Imaging MetaSystem Microsoft and Ikaros software, Metasystems). Chromosome number was compared using the Wilcoxon rank-sum test.

Other statistical analyses. Categorical data were reported as numbers and percentages. Associations between STAG2 expression and subject characteristics were assessed using the χ^2 test. Associations between markers were evaluated using the χ^2 test and using the odds ratio (OR) and 95% confidence interval (95% CI) as a measure of association between categorical variables.

Outcomes considered were recurrence-free and progression-free survival (NMIBC) and progression-free and cancer-specific mortality (MIBC). Survival was represented using Kaplan-Meier curves; differences between curves were assessed with the log-rank test. Cox proportional hazards models were applied for multivariable analysis. The adjusting factors used are indicated in **Supplementary Tables 17–20**. Statistical significance was considered as 0.05. R software (version 2.14) was used for statistical analysis.

27. Marco-Sola, S. *et al.* The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat. Methods* **9**, 1185–1188 (2012).
28. Reva, B. *et al.* Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
29. Kumar, P. *et al.* Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
30. Gonzalez-Perez, A. *et al.* Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **40**, e169 (2012).
31. Vazquez, M. *et al.* Chapter 14: cancer genome analysis. *PLoS Comput. Biol.* **8**, e1002824 (2012).
32. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor *SF3B1* gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2012).
33. Rodríguez-Santiago, B. *et al.* Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am. J. Hum. Genet.* **87**, 129–138 (2010).
34. Snijders, A.M. *et al.* Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* **29**, 263–264 (2001).