



On the absence of mutations in nucleotide excision repair genes in sporadic solid tumors

J.C.M. Mombach^{1,2}, M.A.A. Castro^{3,4}, J.C.F. Moreira⁴ and R.M.C. de Almeida⁵

¹Centro de Ciências Rurais, UNIPAMPA, São Gabriel

²Universidade Federal de Santa Maria, Santa Maria, RS, Brasil

³Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

⁴Universidade Luterana do Brasil, Gravataí, RS, Brasil

⁵Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

Corresponding author: R.M.C. de Almeida

E-mail: rita@if.ufrgs.br

Genet. Mol. Res. 7 (1): 152-160 (2008)

Received November 12, 2007

Accepted January 15, 2008

Published February 19, 2008

ABSTRACT. In general, stochastic tumors show genomic instability associated with the proliferation of DNA point mutations, that is, a mutator phenotype. This feature cannot be explained by a dysfunctional mismatch repair alone, and indicates that nucleotide excision repair (NER) and/or base excision repair should be suppressed. However, mutations in NER genes are not causally implicated in the oncogenesis of sporadic solid tumors, according to the Cancer Gene Census at <http://www.sanger.ac.uk/genetics/CGP/Census/>. This brings up an apparent paradox: how to explain the recurrent non-existence in NER genes of somatic mutations causally related to cancer? In a recent study, we have shown that the origin of point mutations in cancer cell genomes can be explained by a structurally conserved NER with a functional disorder generated from its

entanglement with a disabled apoptosis gene network. In the present study, we further characterize NER gene network properties and show that it has a highly connected architecture. This feature suggests that the absence of mutations in NER genes in sporadic solid tumors is a result of their participation in many essential cellular functions.

Key words: Nucleotide excision repair; Cancer; Gene network; Sporadic solid tumors

INTRODUCTION

Genome instability in solid tumors originates from either somatic mutations (observed in the majority of sporadic cancers) or germline mutations (associated with rare hereditary cancer syndromes) (Hoeijmakers, 2001). Germline mutations are present in all cells of an individual while somatic mutations occur in one or few cells. On the other hand, somatic mutations are favored by genomic instability, which increases the probability that a cell turns cancerous due to some insult such as, for example, exposure to a carcinogenic agent (Breivik and Gaudernack, 2004). We can then expect the genome instability per cell to be greater in the few mutated cells that may originate sporadic tumors than in the cells of individuals showing germline syndromes prone to cancer development. In fact, germline syndromes make an individual vulnerable to some specific cancer, characteristic of each syndrome: depending on the locus of the germline mutation of a given syndrome, different DNA repair mechanisms are affected, generating genome instability as a result of different kinds of insults (Chao and Lipkin, 2006).

Genomic stability is guaranteed by i) DNA repair mechanisms, ii) chromosome stability mechanisms (such as chromosome segregation control), and iii) as an extreme resource, apoptosis, which in fact plays the role of a tissue repair mechanism, since it guarantees tissue homeostasis (Hoeijmakers, 2001; Castro et al., 2006). Consequently, since cancer cells display genome instability, we can conclude that in neoplasms one or more of these mechanisms are dysfunctional.

Here, we consider genomic instability at the nucleotide level. There are three repair systems responsible for correcting DNA point damage: mismatch repair (MMR), base excision repair (BER) and nucleotide excision repair (NER). Germline mutations, associated with syndromes prone to cancer development, are found at genes related to these three systems, for example, adenomatosis polyposis coli at genes related to MMR and BER and xeroderma pigmentosum at genes related to NER (Friedberg, 2003). As expected, these mutations yield a mutator phenotype. However, somatic mutations causally related to cancer are never found in NER or BER (Futreal et al., 2004), which could be thought as an alternative way to produce a mutator phenotype.

Nevertheless, stochastic tumors show, in general, genomic instability associated with the proliferation of DNA point mutations, that is, a mutator phenotype. As this feature cannot be explained by a dysfunctional MMR alone, it indicates that NER and/or BER should be suppressed. In fact, in a previous study, we have found that the activity and diversity profiles of the gene transcripts related to the NER pathway are altered in the majority of tumors listed in NIH/CGAP database (Wheeler et al., 2003), while BER is functionally intact. In Figure 1 we reproduce our previous finding, showing that NER and apoptosis are the most impaired pathways of all tested pathways involved in genome stability (Castro et al., 2007).

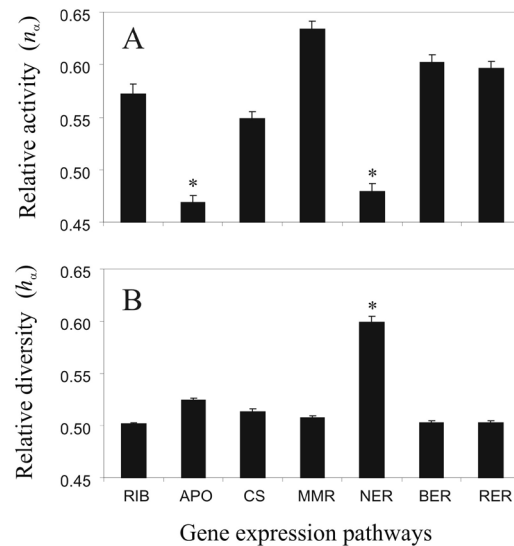


Figure 1. Analysis of gene expression of genome stability pathways. Pathways were evaluated according to two different expression features: gene activity level (i.e., number of transcripts in the pathway) and the distribution of transcripts among the set of genes that characterizes each pathway (i.e., diversity of transcripts). **A.** Relative activity n_a , as defined by Castro et al. (2007), expresses the gene activity in cancer cells in relation to normal cells. **B.** Relative diversity h_a , as defined by Castro et al. (2007), expresses the spread in the distribution of transcripts in cancer cells in relation to normal cells. The results were also compared with conserved pathways used as invariant expression controls (i.e., ribosome, ATP synthase, electron transport chain, and glycolysis). Here, we present the most conserved one: ribosome. RIB = ribosome; APO = apoptosis; CS = chromosome stability; MMR = mismatch repair; NER = nucleotide excision repair; BER = base excision repair; RER = recombinational repair. *Different from controls with $P < 0.001$.

As apoptosis and NER pathways are functionally entangled, it is possible that NER is dysfunctional due to apoptosis suppression. Another way would be through a somatic mutation. As somatic mutations in NER genes causally linked to cancer apparently do not exist, according to Cancer Gene Census (Futreal et al., 2004), how do we explain the recurrent non-existence in NER genes of somatic mutations causally related to cancer?

In this paper aimed at explaining the recurrent absence of somatic mutations causally linked to cancer, we compare the properties of the interaction network of NER genes with those of genes that may show these mutations.

Although cancer reduces an organism's life span, it enhances the fitness of the cancer cell. Consequently, stochastic mutations that reduce cell fitness are not likely to originate cancer. This implies that highly connected genes should not give rise to cancer when randomly mutated. Cell fitness should be taken here in a broad sense: not only should the mutated cell win the race for nutrients and show an enhanced proliferation rate to overcome a clonal selection, it must also evade internal tissue control mechanisms, such as apoptosis, that would remove transformed cells as compared to the original tissue.

In this study, we were interested in determining whether the recurrent absence in NER genes of somatic mutations causally linked to cancer is related to the fact that they are highly functionally connected. To verify this statement, we assessed the connectivity of NER genes and compared it to that of all genes whose somatic mutations have been identified as being causally implicated in cancer (Futreal et al., 2004). We emphasize that for all calculations, we considered

not only physical but also predicted functional interactions, as explained in the Material and Methods section. We found that the NER networks are much more connected than cancer genes. We propose that this feature may explain why stochastic mutations in NER genes, by affecting more neighbors in the network as compared to cancer genes, may reduce cell fitness.

MATERIAL AND METHODS

Data selection

The list of 79 human cancer genes used in this study was obtained from the Cancer Gene Census of the Sanger Institute. This census is an up-to-date catalog of genes causally implicated in cancer and is available at <http://www.sanger.ac.uk/genetics/CGP/Census/>. Here, we considered only somatically mutated genes described in epithelial tumors (i.e., we focused on sporadic solid tumors) and that are also annotated in STRING database (von Mering et al., 2007). The list of 28 genes representing the nucleotide-excision repair pathway follows the catalog of Human DNA repair genes (Wood et al., 2005) and is available at http://www.ejal.icnet.uk/DNA_Repair_Genes.html. The complete file showing Entry IDs of NER and cancer genes are listed in Table 1.

Analysis of protein/gene interaction networks

The protein-protein interaction networks and functional interactions associating genes are generated using the STRING database ('search tool for the retrieval of interacting genes/proteins') (von Mering et al., 2007) with input option "databases" (functional interactions), "experiments" (physical interactions), and 90% confidence level. STRING is a database of known and predicted protein-protein and functional interactions and it allows data mining for individual genes (e.g., a 'target gene'). In this way, all annotations in the database are searched concerning this target gene, which is the central node of the predicted network of interactions. In order to identify genes in the database, we used both HUGO ID (Wain et al., 2004) and Ensembl Peptide ID (Birney et al., 2006). Genes were also checked in NCBI Entrez (Wheeler et al., 2003) to link the approved gene name with respective STRING ID. Alternatively, the amino acid sequence of a given protein was supplied to identify the corresponding gene. The results from the search were saved in data files ('tab-delimited text fields') describing edge relationships. Graphs were built with the Medusa program (Hooper and Bork, 2005) and exported to postscript files to have figure quality improved and edited. Network parameters were obtained as proposed by Barabási and Oltvai (2004). Briefly, we calculated the connectivity for all genes specified in the Data selection section. The connectivity k_i of a given gene i is defined as the number of links that a given node has with other nodes. Genes that are not annotated in STRING database were not included in our analysis. We obtained a total of 107 protein-protein interaction networks (28 for NER and 79 for cancer genes). The complete file showing the network parameters is shown in Table 1.

Statistical analysis

For statistical comparison of the gene networks, we used the Mann-Whitney test. Tests were performed with the SPSS nonparametric statistical package (SPSS for Windows, release 14.0.0., SPSS Inc., Chicago, IL). Results are reported as means \pm SEM with $P < 0.05$ for significance.

Table 1. Network parameters of nucleotide excision repair and cancer genes.

Network ^{a,b}	Gene symbol	Entrez ID	Ensembl ID	k_j^c
CAN	AKAP9	10142	ENSP00000339128	10
CAN	AKT2	208	ENSP00000309428	19
CAN	APC	324	ENSP00000257430	11
CAN	ATF1	466	ENSP00000262053	3
CAN	BRAF	673	ENSP00000288602	10
CAN	BRCA1	672	ENSP00000337814	41
CAN	BRCA2	675	ENSP00000267071	5
CAN	BRD4	23476	ENSP00000263377	0
CAN	CCND1	595	ENSP00000227507	88
CAN	CDH1	999	ENSP00000261769	14
CAN	CDKN2A	1029	ENSP00000354638	12
CAN	COPEB	1316	ENSP00000173785	3
CAN	CTNNB1	1499	ENSP00000344456	69
CAN	CYLD	1540	ENSP00000308928	1
CAN	D10S170	8030	ENSP00000263102	0
CAN	EGFR	1956	ENSP00000275493	48
CAN	ELKS	23085	ENSP00000354158	2
CAN	EP300	2033	ENSP00000263253	65
CAN	ERBB2	2064	ENSP00000269571	14
CAN	ERG	2078	ENSP00000288319	0
CAN	ETV1	2115	ENSP00000242066	0
CAN	ETV6	2120	ENSP00000266427	0
CAN	FBXW7	55294	ENSP00000281708	4
CAN	FGFR2	2263	ENSP00000350166	8
CAN	FGFR3	2261	ENSP00000339824	10
CAN	GNAS	2778	ENSP00000302237	47
CAN	GOLGA5	9950	ENSP00000163416	1
CAN	HMGA1	3159	ENSP00000318322	1
CAN	HRAS	3265	ENSP00000309845	49
CAN	HRPT2	3279	ENSP00000256767	0
CAN	KRAS	3845	ENSP00000256078	16
CAN	KTN1	3895	ENSP00000348562	1
CAN	LIFR	3977	ENSP00000263409	6
CAN	MADH4	4089	ENSP00000341551	35
CAN	MAML2	84441	ENSP00000327563	1
CAN	MAP2K4	6416	ENSP00000262445	28
CAN	MECT1	94159	ENSP00000345001	0
CAN	MEN1	4221	ENSP00000337088	1
CAN	MET	4233	ENSP00000317272	31
CAN	MLH1	4292	ENSP00000231790	9
CAN	MSH2	4436	ENSP00000233146	8
CAN	MSH6	2956	ENSP00000234420	4
CAN	MYC	4609	ENSP00000259523	25
CAN	MYCL1	4610	ENSP00000335376	0
CAN	NCOA4	8031	ENSP00000351902	1
CAN	NONO	4841	ENSP00000298087	4
CAN	NRAS	4893	ENSP00000261444	15
CAN	NTRK1	4914	ENSP00000351486	11
CAN	NTRK3	4916	ENSP00000354207	1
CAN	NUT	256646	ENSP00000329448	0
CAN	PAX8	7849	ENSP00000263334	0
CAN	PCMI	5108	ENSP00000327077	0
CAN	PIK3CA	5290	ENSP00000263967	21
CAN	PLAG1	5324	ENSP00000325546	0
CAN	PPARG	5468	ENSP00000287820	9
CAN	PRCC	5546	ENSP00000271526	0

Continued on next page

Table 1. Continued.

Network ^{a,b}	Gene symbol	Entrez ID	Ensembl ID	k_i^c
CAN	PRKAR1A	5573	ENSP00000226090	21
CAN	PRO1073	29005	ENSP00000332769	0
CAN	PTCH	5727	ENSP00000332353	6
CAN	PTEN	5728	ENSP00000304973	39
CAN	RB1	5925	ENSP00000267163	80
CAN	RET	5979	ENSP00000344798	8
CAN	SFPQ	6421	ENSP00000263532	5
CAN	SMO	6608	ENSP00000249373	2
CAN	STK11	6794	ENSP00000324856	7
CAN	TCEA1	6917	ENSP00000353558	56
CAN	TCF1	6927	ENSP00000257555	2
CAN	TFE3	7030	ENSP00000338360	2
CAN	TFEB	7942	ENSP00000351685	2
CAN	TFG	10342	ENSP00000240851	0
CAN	TMPRSS2	7113	ENSP00000330330	0
CAN	TNFRSF6	355	ENSP00000347979	16
CAN	TP53	7157	ENSP00000269305	93
CAN	TPM3	7170	ENSP00000271850	2
CAN	TPR	7175	ENSP00000264142	25
CAN	TRIM33	51592	ENSP00000351250	0
CAN	TSHR	7253	ENSP00000298171	4
CAN	VHL	7428	ENSP00000256474	11
CAN	ZNF331	55422	ENSP00000345514	0
NER	CCNH	902	ENSP00000256897	98
NER	CDK7	1022	ENSP00000256443	98
NER	CETN2	1069	ENSP00000276365	2
NER	DDB1	1642	ENSP00000301764	20
NER	DDB2	1643	ENSP00000256996	19
NER	ERCC1	2067	ENSP00000013807	46
NER	ERCC2	2068	ENSP00000221481	85
NER	ERCC3	2071	ENSP00000285398	87
NER	ERCC4	2072	ENSP00000310520	46
NER	ERCC8	1161	ENSP00000265038	41
NER	GTF2H1	2965	ENSP00000265963	85
NER	GTF2H2	2966	ENSP00000274400	85
NER	GTF2H3	2967	ENSP00000228955	84
NER	GTF2H4	2968	ENSP00000259895	85
NER	GTF2H5	404672	ENSP00000333219	7
NER	LIG1	3978	ENSP00000263274	28
NER	MMS19L	64210	ENSP00000307263	0
NER	MNAT1	4331	ENSP00000261245	97
NER	RAD23A	5886	ENSP00000321365	1
NER	RAD23B	5887	ENSP00000350708	22
NER	RPA1	6117	ENSP00000254719	70
NER	RPA2	6118	ENSP00000263698	69
NER	RPA3	6119	ENSP00000223129	66
NER	XAB2	56949	ENSP00000351137	42
NER	XPA	7507	ENSP00000259463	19
NER	XPC	7508	ENSP00000285021	19

^aNucleotide excision repair (NER) genes according to Wood (2005). The gene list was checked in HGNC Database (Wain et al., 2004) and NCBI Entrez (Wheeler et al., 2003) to link the approved gene name with respective STRING ID.

^bList of somatically mutated genes (CAN) causally implicated in cancer, described for epithelial tissue in Cancer Gene Census (Futreal et al., 2004), and also annotated in STRING database (von Mering et al., 2007). The gene list was obtained from the Wellcome Trust Sanger Institute Cancer Genome Project web site, <http://www.sanger.ac.uk/genetics/CGP>. Genes were checked in HGNC Database (Wain et al., 2004) and NCBI Entrez (Wheeler et al., 2003) to link the approved gene name with respective STRING ID.

^cThe connectivity k_i of each gene i was estimated as described in Material and Methods.

RESULTS AND DISCUSSION

Four graphs of protein-protein interactions exemplifying gene network analysis are presented in Figure 2. The central node corresponds to the target gene for which the search is made. The predicted associations start from this central node and include all associations found in the STRING database (von Mering et al., 2007). Figure 2A and B represent the networks of NER target genes, while Figure 2C and D are two examples of networks generated from cancer genes as targets. It is clear that NER genes are more connected. Note that these networks show all genes connected with the target genes, including genes belonging to other functions or pathways. We generated 107 such networks (28 for NER and 79 for cancer genes), one for each target gene.

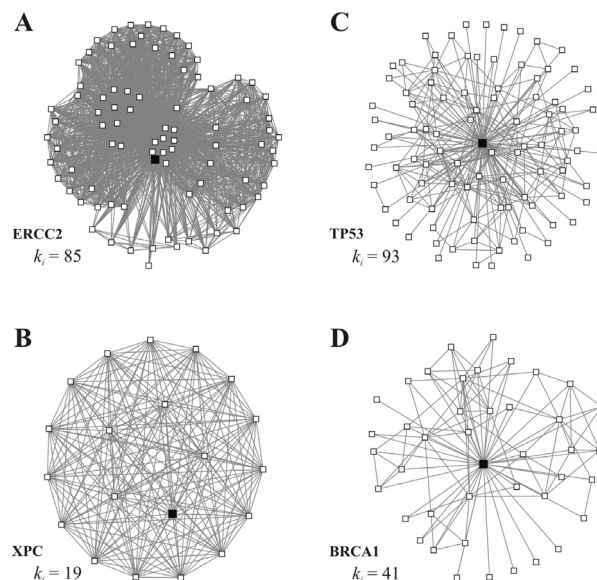


Figure 2. Graph of interactions illustrating nucleotide excision repair (NER) and cancer gene networks. The predicted associations are generated using database STRING (von Mering et al., 2007) with input options “Experimental/Biochemical Data”, “Association in Curated Databases” and 90% confidence level. **A** and **B**. ERCC2 and XPC genes, illustrating NER gene networks. **C** and **D**. TP53 and BRCA1 genes, illustrating somatically mutated gene networks. Black node represents the central gene for which the network is built (i.e., ‘target gene’). White nodes represent all genes connected to the central node. The connectivity k_i of the central node is indicated. Network parameters are listed in Table 1 for all NER genes and all cancer genes for which somatic mutations have been causally implicated in epithelial tumors (i.e., sporadic solid tumors).

The connectivity calculated for all NER and cancer genes is presented in Figure 3A and is provided in Table 1. As one can observe, the great majority of NER genes have connectivity higher than do cancer genes. These latter networks are significantly more sparse (Figure 3B; $P < 0.05$). In fact, this feature has been partially observed in a previous study for a much smaller set of cancer genes (17 genes), those restricted to genome stability functions (Castro et al., 2007). Here, we extended the comparison to all 79 cancer genes associated with sporadic solid tumors found in the Cancer Genes Census (Futreal et al., 2004), considering the 79 networks produced by STRING database for each cancer gene as a target gene. In Figure 4, we present the interaction network, considering only NER genes. As we can see, the network is highly connected, which confirms a cooperative functioning (Christensen et al., 2007).

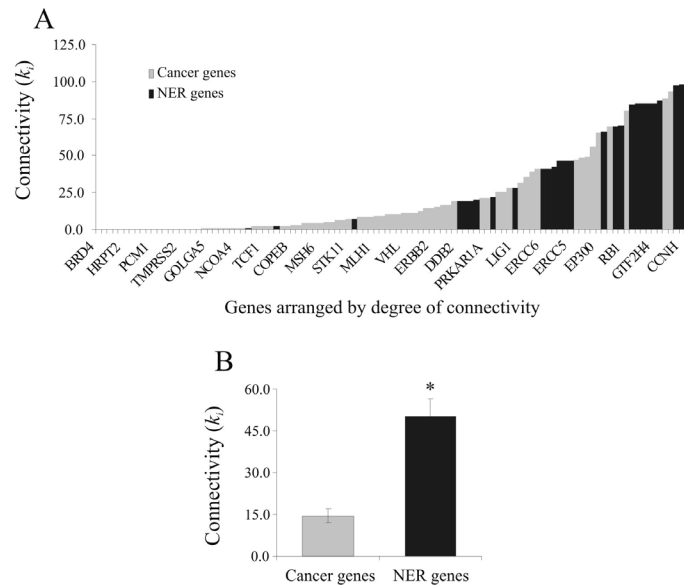


Figure 3. Contrasting nucleotide excision repair (NER) and cancer gene networks. Analysis of protein-protein interaction networks of NER and cancer genes for which somatic mutations have been causally implicated in epithelial tumors. **A.** Connectivity k_i of a given gene i , which shows the number of links that a given node has with other nodes as defined in Material and Methods. Genes are ranked by degree of connectivity. NER genes are indicated in black columns. **B.** Average connectivity of NER and cancer genes. Asterisk indicates significant difference between groups, with $P < 0.05$.

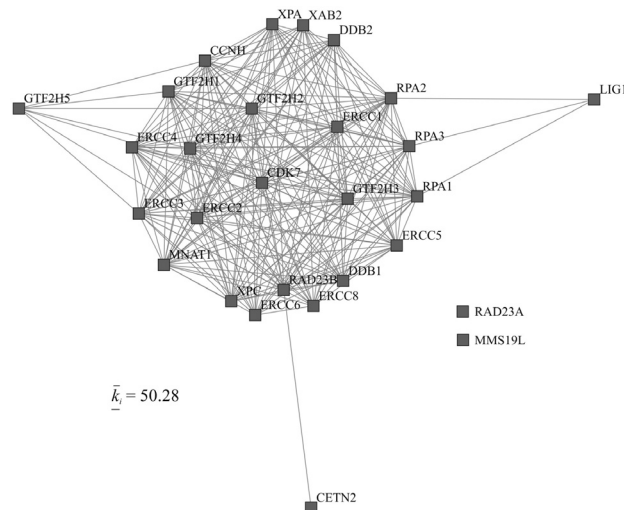


Figure 4. Graph of the nucleotide excision repair (NER) gene network built using software MEDUSA (Hooper and Bork, 2005). The network is highly connected. The network shows the known and predicted protein-protein and functional interactions from the database STRING (von Mering et al., 2007) with input options “databases”, “experiments” and 90% confidence level. Nodes without predicted associations at this level are indicated. Average connectivity k_i of NER genes is indicated (considering all interactions of NER genes - including external interactions). The list of NER gene identifiers (i.e., Entry IDs) is provided as a supplementary spreadsheet (Table 1).

We can explain these multiple interactions of NER genes by their participation in a variety of cellular processes. Some of the NER gene products are pre-associated tightly in complexes as heteromultimers. These include the 10 subunits of TFIIH, the ERCC1-XPF complex, the XPC-RAD23B complex, and the heterotrimeric RPA1-RPA2-RPA3 complex. Additionally, some known biochemical facts about NER gene products are unusual in comparison to other DNA repair pathways. Many NER genes have essential functions in the cell other than DNA repair. This is true for all 10 subunits of TFIIH (as an essential factor for RNA transcription), for ERCC1 and XPF (which also act in a recombination pathway), for XPG (which works with TFIIH as a transcription activator), for the RPA subunits (which form the major single-stranded binding protein, also necessary for semi-conservative DNA replication), and for LIG1, and XAB2.

An additional support to our proposal comes from experiments of gene silencing. A recent study has shown that silencing of some NER genes by means of RNA interference in cultures of HeLa and MCF-7 cells causes major growth disadvantages (Biard, 2007). Consequently, the contrast among NER and cancer genes in regard to their properties in the network suggests that, in general, random point mutations directly in NER genes may reduce cell fitness in sporadic solid tumors, explaining the absence of somatic mutations in these genes.

ACKNOWLEDGMENTS

Research partially supported by the Brazilian agencies FAPERGS, CNPq, and CAPES. We acknowledge STRING database and Wellcome Trust Sanger Institute Cancer Genome Project for providing public access to their data. We thank Marialva Sinigaglia for useful discussions.

REFERENCES

- Barabási AL and Oltvai ZN (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5: 101-113.
- Biard DS (2007). Untangling the relationships between DNA repair pathways by silencing more than 20 DNA repair genes in human stable clones. *Nucleic Acids Res.* 35: 3535-3550.
- Birney E, Andrews D, Caccamo M, Chen Y, et al. (2006). Ensembl 2006. *Nucleic Acids Res.* 34: D556-D561.
- Breivik J and Gaudernack G (2004). Resolving the evolutionary paradox of genetic instability: a cost-benefit analysis of DNA repair in changing environments. *FEBS Lett.* 563: 7-12.
- Castro MA, Onsten TG, Moreira JC and de Almeida RM (2006). Chromosome aberrations in solid tumors have a stochastic nature. *Mutat. Res.* 600: 150-164.
- Castro MA, Mombach JC, de Almeida RM and Moreira JC (2007). Impaired expression of NER gene network in sporadic solid tumors. *Nucleic Acids Res.* 35: 1859-1867.
- Chao EC and Lipkin SM (2006). Molecular models for the tissue specificity of DNA mismatch repair-deficient carcinogenesis. *Nucleic Acids Res.* 34: 840-852.
- Christensen C, Thakar J and Albert R (2007). Systems - level insights into cellular regulation: inferring, analysing, and modelling intracellular networks. *IET. Syst. Biol.* 1: 61-77.
- Friedberg EC (2003). DNA damage and repair. *Nature* 421: 436-440.
- Futreal PA, Coin L, Marshall M, Down T, et al. (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4: 177-183.
- Hoeijmakers JH (2001). Genome maintenance mechanisms for preventing cancer. *Nature* 411: 366-374.
- Hooper SD and Bork P (2005). Medusa: a simple tool for interaction graph analysis. *Bioinformatics* 21: 4432-4433.
- von Mering C, Jensen LJ, Kuhn M, Chaffron S, et al. (2007). STRING 7 - recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* 35: D358-D362.
- Wain HM, Lush MJ, Ducluzeau F, Khodiyar VK, et al. (2004). Genew: the Human Gene Nomenclature Database, 2004 updates. *Nucleic Acids Res.* 32: D255-D257.
- Wheeler DL, Church DM, Federhen S, Lash AE, et al. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Res.* 31: 28-33.
- Wood RD, Mitchell M and Lindahl T (2005). Human DNA repair genes, 2005. *Mutat. Res.* 577: 275-283.