

RESEARCH NOTE

Open Access



# Genetic variants of the *EGFR* ligand-binding domain and their association with structural alterations in Arab cancer patients

Maryam Marzouq<sup>1</sup>, Ali Nairouz<sup>1</sup>, Nouredine Ben Khalaf<sup>1</sup>, Sonia Bourguiba-Hachemi<sup>1</sup>, Raed Quaddorah<sup>1</sup>, Dana Ashoor<sup>1</sup> and M. Dahmani Fathallah<sup>1,2\*</sup>

## Abstract

**Objective:** This study aimed to identify novel genetic variants in the CR2 extracellular domain of the epidermal growth factor receptor (EGFR) in healthy individuals and patients with six different types of adenocarcinoma, in Arabian peninsula populations. It also aimed to investigate the effects of these variants on the EGFR structure and their eventual relevance to tumorigenesis.

**Results:** We detected seven new EGFR genetic variants in 168 cancer patients and 114 controls. A SNP rs374670788 was more frequent in bladder cancer but not significantly associated to. However, a missense mutation (V550M) was significantly associated to colon, ovary, lung, bladder and thyroid cancer samples ( $p < 0.05$ ). Three mutations (H590R, E602K and T605T) were found in the heterozygous form only in colon cancer patients. Genomic analysis of the synonymous mutation (G632G) showed that the T/A genotype could be associated to thyroid cancer in Arab patients ( $p < 0.05$ ). An additional novel SNP rs571064657 was observed in control individuals. Computational analysis of the genetic variants revealed a reduction in the stabilization of the EGFR tethered form for both V550M and the common R521K variant with low energetic state ( $-\Delta\Delta G$ ). Molecular interactions analysis suggested that these mutations might affect the receptor's function and promote tumorigenesis.

**Keywords:** EGFR protein, Missense mutation, Arabs, Adenocarcinoma, Single nucleotide polymorphism, CR1/CR2 EGFR domains, EGFR R521K polymorphism, Arabian Peninsula population

## Introduction

It has been reported that cancer prevalence is increasing in the Arabian peninsula populations [1]. This trend was linked to a number of contributing factors such as increased life expectancy, diet, young smoking, obesity, and pollution [2, 3]. These observations spurred better cancer patients characterization particularly at the level of cancer markers such as the epidermal growth

factor receptor (EGFR). Indeed, upon binding to ligands like EGF or TGF- $\alpha$ , the EGFR undergoes autophosphorylation that leads to the activation of several signal-transduction cascades, and cell cycle-progression [4]. Furthermore, molecular alterations of the EGFR activate pro-oncogenic signaling pathways, including the RAS-RAF-MEK-ERK MAPK and AKT-PI3K-mTOR [5]. In these cases, the activation of EGFR turns on tumorigenesis promoting event such as cellular proliferation, resistance to apoptosis, and angiogenesis [6, 7]. Alongside, the *EGFR* gene amplification leads to the receptor overexpression, accelerating its uncontrolled activation which is associated with several malignancies including lung, breast, stomach, colorectal, head and neck, pancreatic

\*Correspondence: d.fathallah@agu.edu.bh

<sup>1</sup> Department of Life Sciences, Health Biotechnology Program, College of Graduate Studies, Arabian Gulf University, PO Box 26671 Manama, Kingdom of Bahrain

Full list of author information is available at the end of the article



carcinomas and glioblastoma [8–11]. The *EGFR* gene encodes a protein that have different functional domains with exons 1 to 16 encoding the extracellular ligand-binding domain [8]. The later contains 4 sub-domains: I=L1, II=CR1, III=L2, and IV=CR2. L1 and L2 are leucine-rich ligand-binding domains. Whereas, CR1 and CR2 are cysteine-rich and involves the formation of structurally important disulfide bonds [7–9]. EGFR has a tethered monomeric form and an extended dimeric complex. X-ray crystals and molecular dynamic simulations studies described these structural changes and linked them to EGFR regulation in cancer [12–14]. In the tethered conformation, EGFR ectodomain II and IV are folded into each other's forming the hairpin loop of domain II spanning residues 240–260 and interacts with C11Vc and C11Vd modules of domain IV (spanning residues 561–569 and 572–585, respectively). EGF receptor is concomitant to domains rearrangement in a way that domain I and III are accessible for EGF and when domain II dimerizes with another EGFR unfolded tethered form that rotated 90° on its vertical axis. Alterations of the extracellular domains' protein sequences have been related to cancer prevalence and the effectiveness of targeted immunotherapy [14, 15]. Indeed, variant R521K is a widespread functional variant that plays an important role in EGFR expressing tumors and impacts the effectiveness of anticancer agents [16, 17].

## Main text

### Methods

#### Study population

168 Cancer patients from the Arabian peninsula populations (N=74 colon, 20 breast, 20 ovary, 20 bladder, 14 lung, 20 thyroid) alongside 114 healthy controls matched for gender, age, and ethnicity.

#### Genomic DNA isolation

DNA was extracted from FFPE tissues (patients) and peripheral blood (controls) using QIAamp DNA FFPE Tissue Kit (Cat#0.56404) and QIAamp DNA Mini Kit (Cat#0.51306) respectively, according to the manufacturer's instructions. DNA samples were stored at – 20 °C.

#### Mutational analysis

PCR was used to amplify the exons coding for EGFR-CR2 domain using primers [18] and conditions as shown in Additional file 1. PCR was conducted in 50 µl containing 33 ng of DNA, 5 × GoTaq buffer, 0.2 mM dNTP mixture, 4 mM MgCl<sub>2</sub>, 0.025 µM each primer, 0.96 U/µl GoTaq Enzyme, and nuclease-free water. PCR products were checked on 2% agarose gel and sequenced using the dye termination method [19]. All detected gene alterations were confirmed by two independent sense and

antisense PCR. Some samples were dropped because one of the two PCR reactions failed to confirm the nucleotide change. Indeed, the extraction of high-quality DNA (suitable for PCR reactions and sequencing) from some cancer samples was certainly affected by the storage of FFPE biopsy specimens (old blocks).

#### Bioinformatics study

**Model preparation** The untethered structure of EGFR (id:3NJP) and the tethered EGFR/EGF complex structure (id:1NQL) were downloaded from the RCSB PDB database (<https://www.rcsb.org/>). The Pymol program (Schrödinger) was used for structure visualization and mutants models generation.

**Interaction analysis for energy minimization** Wild type and mutants (R521K,V550M) forms of tethered EGFR/EGF complexes were solvated in a periodic boundary cube of water (2.4 Å × 2.4 Å × 2.4 Å) using VMD solvation plugin [20]. NaCl ions were added to neutralize the molecule. The steepest descent energy minimization was applied to relax the structures at 310 K (50,000 steps) in NVT mode, using NAMD (Nanoscale Molecular Dynamics program; v 2.9) [21]. Average structures obtained with UCSF Chimera (UCSF, USA) were analyzed for inter-chain residue H-bond interactions with LigPlot+ [22].

**Protein stability prediction** Site directed mutator (SDM) server (<http://marid.bioc.cam.ac.uk/sdm2>) was used to predict the effect of the mutations on protein stability. The variation of amino acid replacements that occurs a specific structural environment is analyzed by SDM according to substitution probability tables, generated from tolerated substitution in homologous proteins with known 3D-structures [23].

#### Statistical analysis

Standard contingency table and Chi-square test were used to assess the association of the genotype and allele frequencies of the EGFR-CR2 variants with cancer. A *p*-value < 0.05 was considered statistically significant.

## Results

### Gene alterations analysis

In *exon 13*, we found a novel SNP, a transition 1536C > T yielding synonymous substitution (P512P) (Table 1; Fig. 1a). This new SNP was assigned the ID rs374670788 by the dbSNP (ss825678873) data bank. Although the frequency of the heterozygous variant CT (Table 1; Fig. 1b) was high in bladder sample (6% vs. 1% in control), no statistically significant association was observed (*p* = 0.231). Genotypes distributions of tested polymorphisms were consistent with

**Table 1** Novel genetic alterations in CR2 domain: allele and genotype frequencies

Exon	SNP ID	Type of Samples (n)	Allele	AF (%)	Genotype	GF (%)	Samples (n)	HWE $p$ values	Fisher's $p$ value
13	SNP 1536 C>T <b>rs374670788</b>	Control (114)	C	99	CC	99	113	0.962	-
			T	1	CT	1	1		
					TT	0	0		
		Colon (54)	C	99	CC	98	53	0.945	0.540
			T	1	CT	2	1		
					TT	0	0		
Bladder (16)	C	97	CC	94	15	0.897	0.231		
	T	3	CT	6	1				
			TT	0	0				
14	Mutation 1648 G>A	Control (113)	G	100	GG	100	113	-	-
			A	0	GA	0	0		
					AA	0	0		
		Colon (64)	G	95.3	GG	91	58	0.693	0.0019
			A	4.7	GA	9	6		
					AA	0	0		
		Lung (10)	G	75	GG	50	5	0.291	0.000001
			A	25	GA	50	5		
					AA	0	0		
		Ovary (11)	G	86.33	GG	73	8	0.600	0.0005
			A	13.64	GA	27	3		
					AA	0	0		
Bladder (16)	G	81.25	GG	63	10	0.355	0.000001		
	A	18.75	GA	38	6				
			AA	0	0				
Thyroid (9)	G	88.89	GG	78	7	0.707	0.004		
	A	11.11	GA	22	2				
			AA	0	0				
15	Mutation 1769 A>G	Control (97)	A	100	AA	100	97	-	-
			G	0	AG	0	0		
					GG	0	0		
		Colon (37)	A	99	AA	97	36	0.933	0.246
			G	1	AG	3	1		
					GG	0	0		
	Mutation 1804 G>A	Control (97)	A	100	AA	100	97	-	-
			G	0	AG	0	0		
					GG	0	0		
		Colon (37)	G	99	GG	97	36	0.933	0.246
			A	1	GA	3	1		
					AA	0	0		
Mutation 1815 C>T	Control (97)	C	100	CC	100	97	-	-	
		T	0	CT	0	0			
				TT	0	0			
	Colon (37)	C	99	CC	97	36	0.933	0.246	
		T	1	CT	3	1			
				TT	0	0			
16	Mutation 1896T>A	Control (114)	T	100	TT	100	114	-	-
			A	0	TA	0	0		
					AA	0	0		
		Colon (64)	T	98.4	TT	90	62	0.898	0.127
			A	1.6	TA	10	2		
					AA	0	0		
		Ovary (10)	T	95	TT	90	9	0.867	0.08
			A	5	TA	10	1		
					AA	0	0		
		Thyroid (9)	T	77.8	TT	55.6	5	0.391	0.00001
			A	22.2	TA	44.4	4		
					AA	0	0		
Control (114)	C	99.12	CC	98.25	112	0.924	-		
	T	0.88	CT	1.75	2				
			TT	0	0				
SNP 1913 C>T <b>rs571064657</b>	Control (114)	C	99.12	CC	98.25	112	0.924	-	
		T	0.88	CT	1.75	2			
				TT	0	0			

**Table 1** (continued)

NCBI assigned new SNPs are highlighted in bold

AF allele frequency, GF genotype frequency, HWE Hardy–Weinberg equilibrium

(See figure on next page.)

**Fig. 1** The new variants in CR-2 domain of *EGFR* gene in cancer patients and healthy subjects. **a** Alignment of nucleic and amino acid sequences of exon 13 of the *EGFR* gene showing the new SNP 1536C>T (P512P) marked with an asterisk. **b** Chromatographic patterns of direct sequencing showing new mutation in exon 13 of *EGFR*, 1536 C>T (P512P) (indicated by the arrows). **c** Alignment of nucleic and amino acid sequences of exon 14 of the *EGFR* gene showing the novel mutation 1468G>A (V550M) marked with an asterisk. **d** Chromatographic patterns of direct sequencing of *EGFR* exon 14 showing the novel mutation 1648 G>A (V550M) (pointed by an arrows). **e** Alignment of nucleic and amino acid sequences of exon 15 of the *EGFR* gene showing the new mutations 1769A>G (H590R), 1804G>A (E602K), and 1815C>T (T605T) respectively, all marked with an asterisk. **f–h** Chromatographic patterns of direct sequencing of *EGFR* exon 15 showing the novel mutations 1769**A>G** (H590R), 1804**G>A** (E602K), and 1815**C>T** (T605T) respectively, the sequence change detected pointed by an arrows. **i** Alignment of nucleic and amino acid sequences of exon 16 of the *EGFR* gene showing two new variants 1896T>A (G632G) and 1913C>T (**T638M**) respectively marked with an asterisk. **j, k** Chromatographic patterns of direct sequencing of *EGFR* exon 16 showing the new mutation 1896 T>A (G632G) and the novel SNP 1913 C>T (**T638M**), respectively

the Hardy–Weinberg equilibrium ( $p > 0.05$ ). Also, the presence of 5 reported variants were confirmed (Additional file 2), including the common variant rs2227983, **R521K** which has been reported in NCBI-ClinVar database as related to cancer [24, 25]. There was no association of SNP rs2227983 with any cancer type, 47.9% in patients versus 47.4% in controls (Additional file 2). In *exon 14*, we observed a novel missense mutation (1648**G>A**) resulting in the substitution (V550M) (Table 2; Fig. 1c, d). The rate of heterozygous **GA** variant was statistically significant in colon, lung, ovary, bladder, and thyroid tumor samples ( $p < 0.05$ ). Also, one reported SNP was observed (Additional file 2). In *exon 15*, three new mutations were detected (1769**A>G**, 1804**G>A**, and 1815**C>T**) in cancer patients but not in healthy subjects (Table 1; Fig. 1e) and confirmed by reverse sequencing. These alterations yielded respectively the missense mutations H590R, E602K and T605T synonymous substitution (Fig. 1f–h, respectively). Also, three reported SNPs were identified (Additional file 2). The frequency of the heterozygous variant rs17290169, was statistically highly significant in controls ( $p < 0.05$ ). In *exon 16*, two novel alterations were found: a 1896**T>A**, transversion (Table 1; Fig. 1i) that was detected in 3.1% of colon tumor samples, 10% of ovary tumor samples and 44.4% of thyroid tumor samples (Table 1; Fig. 1j). The frequency of heterozygous (**T/A**) variant was statistically highly significant in thyroid tumor samples ( $p < 0.05$ ). Likewise, we observed a new SNP 1913**C>T**, a transition yielding synonymous substitution **T638M** (Fig. 1i). We observed this rare allele only in 2 controls (Fig. 1k). The newly identified SNP was assigned the ID rs571064657, by the dbSNP ss825678874 data base.

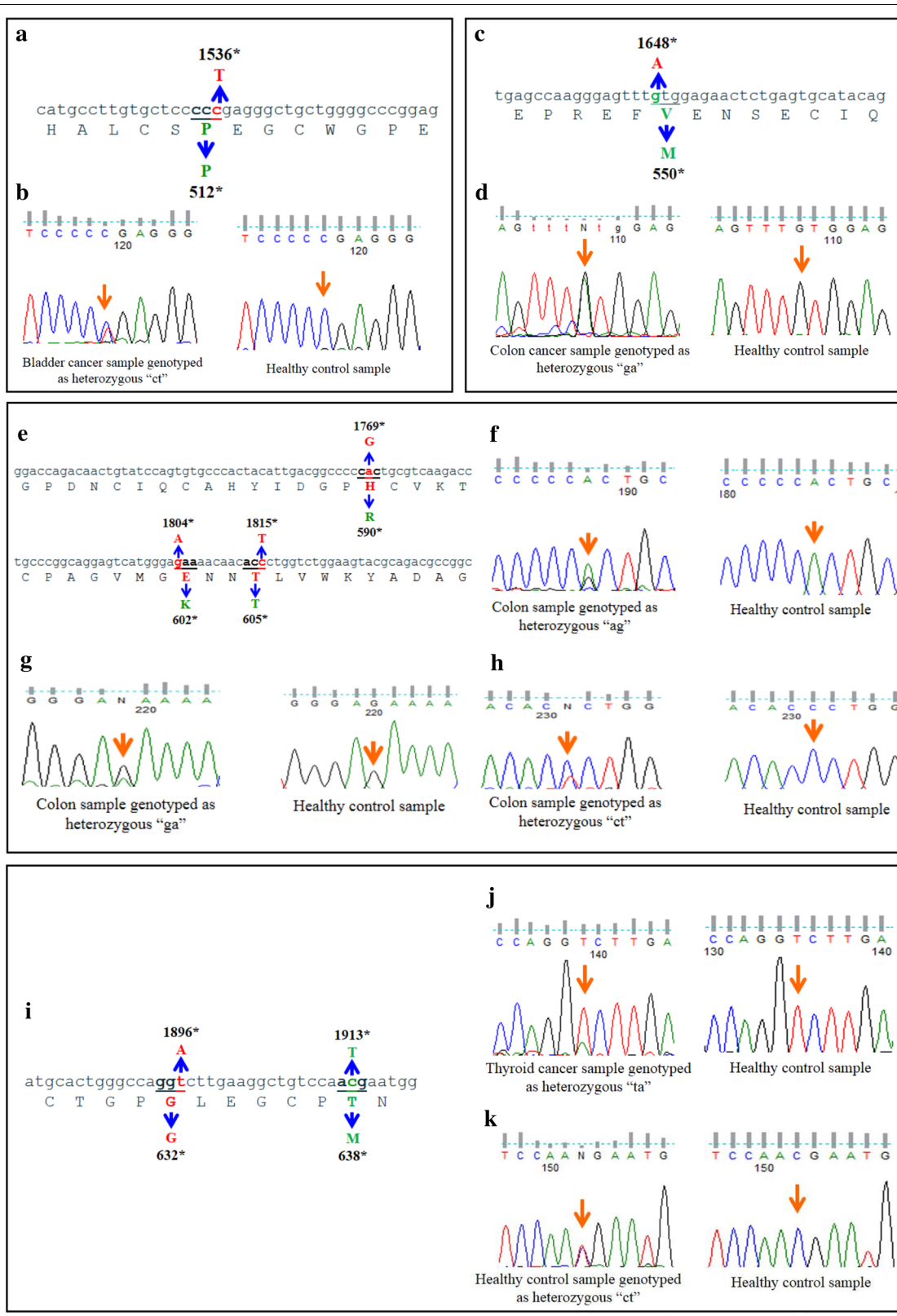
In addition, one reported SNP was detected (Additional file 2).

#### In silico analysis

This study focused on the interaction of *EGFR* variants V550M and R521K with EGF. As shown in Table 2, fifteen residues of wild type *EGFR* were found to interact with EGF through H-bonds. No differences in H-bonds were observed in the interface with mutant R521K. However, one H-bond was missing and one extra H-bond was identified in mutant V550M interaction with EGF (Additional file 3). Moreover, 5 residues of the *EGFR* were involved in internal interactions between CR1 and CR2 domains (Table 3). For mutants R521K and V550M, one H-bond was found to be missing in the interaction between CR1 and CR2 (Additional file 4). No differences in H-bonds were observed in the other alterations. SDM analysis on the human *EGFR* revealed that both mutants V550M and R521K were predicted to be destabilizing mutations in term of complex stability with  $\Delta\Delta G$  values of  $-1.16$  and  $-0.29$ , respectively.

#### Discussion

While a number of studies showed that mutations of *EGFR* intracellular signaling region can lead to tumorigenesis, less information is available about the extracellular region of *EGFR*. This region is structurally instrumental to the molecule function as it contains the binding site of the antibody cetuximab successfully used in cancer therapy [26]. This study focused on the CR2 domain that was shown to interact with the CR1 domain and play a major role in the receptor's dimerization and functions like ligand binding, growth stimulation and tyrosine kinase activation [27]. We identified 7 new gene variants in the *EGFR*-CR2 domain coding



**Table 2 Analysis of polar interactions between EGF/EGFR in the untethered form of EGFR.** The polar interactions between EGFR wild type (3NJP) and its mutated forms R521K and V550M revealed no changes in H-bonds with mutant R521K besides 1 missing and 1 extra H-bond (highlighted in bold) with the mutant V550M

EGF (ligand)-EGFR (wild type)	EGFR with (mutant R521K)	EGFR with (mutant V550M)
Lys <sup>28</sup> -Glu <sup>90</sup>	Lys <sup>28</sup> -Glu <sup>90</sup>	Lys <sup>28</sup> -Glu <sup>90</sup>
Ala <sup>25</sup> -Ser <sup>99</sup>	Ala <sup>25</sup> -Ser <sup>99</sup>	Ala <sup>25</sup> -Ser <sup>99</sup>
Asp <sup>11</sup> -Ser <sup>356</sup>	Asp <sup>11</sup> -Ser <sup>356</sup>	Asp <sup>11</sup> -Ser <sup>356</sup>
Glu <sup>51</sup> -Ser <sup>468</sup>	Glu <sup>51</sup> -Ser <sup>468</sup>	Glu <sup>51</sup> -Ser <sup>468</sup>
Trp <sup>50</sup> -Lys <sup>465</sup>	Trp <sup>50</sup> -Lys <sup>465</sup>	Trp <sup>50</sup> -Lys <sup>465</sup>
Lys <sup>48</sup> -Gln <sup>411</sup>	Lys <sup>48</sup> -Gln <sup>411</sup>	Lys <sup>48</sup> -Gln <sup>411</sup>
<b>Asp<sup>46</sup>-His<sup>409</sup></b>	<b>Asp<sup>46</sup>-His<sup>409</sup></b>	<b>Asp<sup>46</sup>-His<sup>409</sup></b>
<b>Lys<sup>48</sup>-His<sup>409</sup></b>	<b>Lys<sup>48</sup>-His<sup>409</sup></b>	-
Arg <sup>41</sup> Arg <sup>41</sup> -Asp <sup>355</sup>	Arg <sup>41</sup> Arg <sup>41</sup> -Asp <sup>355</sup>	Arg <sup>41</sup> Arg <sup>41</sup> -Asp <sup>355</sup>
<b>Gln<sup>43</sup>-Gln<sup>384</sup></b>	<b>Gln<sup>43</sup>-Gln<sup>384</sup></b>	<b>Gln<sup>43</sup>-Gln<sup>384</sup></b>
<b>Arg<sup>45</sup>-Gln<sup>384</sup></b>	<b>Arg<sup>45</sup>-Gln<sup>384</sup></b>	<b>Arg<sup>45</sup>-Gln<sup>384</sup></b>
		<b>His<sup>16</sup>-Gln<sup>384</sup></b>
Tyr <sup>44</sup> -His <sup>346</sup>	Tyr <sup>44</sup> -His <sup>346</sup>	Tyr <sup>44</sup> -His <sup>346</sup>
Cys <sup>33</sup> Asn <sup>32</sup> -Gly <sup>18</sup>	Cys <sup>33</sup> Asn <sup>32</sup> -Gly <sup>18</sup>	Cys <sup>33</sup> Asn <sup>32</sup> -Gly <sup>18</sup>
Cys <sup>33</sup> Cys <sup>31</sup> -Gln <sup>16</sup>	Cys <sup>33</sup> Cys <sup>31</sup> -Gln <sup>16</sup>	Cys <sup>33</sup> Cys <sup>31</sup> -Gln <sup>16</sup>
Cys <sup>31</sup> Glu <sup>40</sup> -Thr <sup>15</sup>	Cys <sup>31</sup> Glu <sup>40</sup> -Thr <sup>15</sup>	Cys <sup>31</sup> Glu <sup>40</sup> -Thr <sup>15</sup>
Gly <sup>39</sup> -Asn <sup>12</sup>	Gly <sup>39</sup> -Asn <sup>12</sup>	Gly <sup>39</sup> -Asn <sup>12</sup>
Glu <sup>40</sup> -Ser <sup>11</sup>	Glu <sup>40</sup> -Ser <sup>11</sup>	Glu <sup>40</sup> -Ser <sup>11</sup>

**Table 3 Analysis of polar interactions between CR1/CR2 domains in the tethered form of EGFR.** The polar interaction between EGFR wild type (1NQL) and its mutated forms R521K and V550M showed 1 missing H-bond (highlighted in bold)

CR1/CR2 domains (wild type)	CR1/CR2 domains (mutant R521K)	CR1/CR2 domains (mutant V550M)
Asn <sup>247</sup> -Glu <sup>578</sup>	Asn <sup>247</sup> -Glu <sup>578</sup>	Asn <sup>247</sup> -Glu <sup>578</sup>
<b>Tyr<sup>246</sup>-Glu<sup>578</sup></b>	-	-
Tyr <sup>246</sup> -Met <sup>576</sup>	Tyr <sup>246</sup> -Met <sup>576</sup>	Tyr <sup>246</sup> -Met <sup>576</sup>
Tyr <sup>246</sup> -Lys <sup>585</sup>	Tyr <sup>246</sup> -Lys <sup>585</sup>	Tyr <sup>246</sup> -Lys <sup>585</sup>
Tyr <sup>246</sup> -Asp <sup>563</sup>	Tyr <sup>246</sup> -Asp <sup>563</sup>	Tyr <sup>246</sup> -Asp <sup>563</sup>

exons. The V550M mutation was found to be significantly associated to colon, ovary, lung, bladder and thyroid cancer samples ( $p < 0.05$ ), which suggest that this mutation could be associated to tumorigenesis. The 3 novel mutations observed in exon 15 were found in the heterozygous form and only in colon cancer patients. The new alteration 1896T > A leading to a synonymous mutation (G632G) was mostly encountered in thyroid tumors but also less frequently in ovary and colon tumors. The data suggest that further study on larger sample size is mandatory to get accurate significance of the association between T/A genotype and thyroid cancer risk in Arab patients. Meanwhile, we did not observe a significant variant of the SNP rs2227983

genotypes frequencies between cancers and controls samples with over 50% being G/G in both groups. The impact of mutations on the function of protein associated to cancer can be predicted through in silico studies of its structure. Such information is crucial in understanding genotype-phenotype correlations and disease biology. The analysis of V550M substitution showed no striking differences in the interactions interface between EGF and EGFR wild type and R521K mutated form. However, SDM analysis revealed a reduction in the stabilization states of the EGFR/EGF complex for both V550M and R521K mutations. The impact of the V550M mutation on the complex stability can be described as a result of size difference, where the large methionine residue cannot fit within the available space which might disrupt the original core structure of CR2 domain and have an impact on its function. In turn this alters the interaction with its protein partner thereby affecting the signaling pathways. In the tethered (inactive) configuration of EGFR, the CR1 loop interacts closely with the CR2 domain and several side-chain to the backbone, and backbone to backbone hydrogen bonds are formed between CR1/CR2 domains. The side chain of tyrosine (Tyr<sup>246</sup>) is critical for both CR1/CR2 interactions, and it's found close to the tip of the CR1  $\beta$ -hairpin/loop interact via hydrogen bonds with the Glu<sup>578</sup>, Met<sup>576</sup>, Lys<sup>585</sup> and Asp<sup>563</sup> in the side chains of CR2 domain. Only one hydrogen bond out of five (Tyr<sup>246</sup>-Glu<sup>578</sup>) was missing in both R521K and V550M.

These mutations might disrupt the intramolecular domain CR1/CR2 interactions of EGFR leading up to exposure of the EGFR dimerization arm and enhance the reaction to EGF ligand by increasing its binding affinity. Therefore, the H-bond between (Tyr<sup>246</sup>-Glu<sup>578</sup>) is crucial for the regulation of receptor activation because it's involved in the stabilization of the receptor dimer interface, auto-inhibition, and impairment of receptor function. These data have guided our decision to raise monoclonal antibodies that distinguish the two variants R and K to ultimately develop these antibodies as an anti-cancer drug to be administered to a patient according to their rs2227983 genotype. Indeed, R521K variant is situated at the boundary of EGFR domains III and IV, at the location of the anti-cancer mAb cetuximab specific epitope [28]. This variant is known to impact the outcome of antibody-based therapy. It is also associated with the weakening of the EGFR functions as compared to the wild type [16]. The R521K variant has also been described as being associated with cancer severity in EGFR-expressing tumors, like gliomas, lung cancer and breast cancer [29–31].

## Conclusion

This study revealed new genetic variants in the EGFR-CR2 domain in cancer patient from the Arabian peninsula. The in-silico study highlighted the effect of two variants on the receptor structure–function and their eventual implication in tumorigenesis.

## Limitations

The genetic association of mutations in the EGFR extracellular domains requires validation on a larger number of patients and their family members. This will also allow haplotypes analysis. Moreover, the patients' clinicopathological data was not accessible.

## Abbreviations

EGFR: Epidermal growth factor receptor; FFPE: Formalin-fixed paraffin-embedded; PCR: Polymerase chain reaction; CR1: Cysteine-rich domain 1; CR2: Cysteine-rich domain 2.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13104-021-05559-y>.

**Additional file 1:** Primers used to amplify exons encoding CR2 sub-domain of human EGFR.

**Additional file 2:** A list of *EGFR* gene variants. The previously reported alterations found in the CR2 domain of *EGFR* gene in patients and healthy control samples from the Arabian peninsula region.

**Additional file 3:** Polar interactions between wild type and mutated EGFR with EGF (untethered monomer, 3NJP). A) Wild EGF/EGFR complex shows

15 polar interactions. B) EGF/EGFR-V550M showing 1 missing and 1 extra polar interaction. (Blue residues represent EGFR wild type; green residues represent EGF wild type).

**Additional file 4:** Polar interactions between wild type and mutated EGFR with EGF (untethered monomer, 3NJP). A) Wild CR1/CR2 shows 5 polar interactions. B) CR1/CR2-R521K showing 4 polar interactions. C) CR1/CR2-V550M showing 4 polar interactions. (Blue residues represent CR1 domain, green residues represent CR2 domain).

## Acknowledgements

The authors are grateful to all the patients and individuals for their participation. We also thank the clinicians and other hospital staff in respective centers [Kuwait National Cancer Control Center (KCCC) and from multiple medical centers in the Eastern Province of Saudi Arabia] who contributed to data collection for this study.

## Authors' contributions

All authors contributed to the study. MDF: conceptualization; data analysis MM, AN, RQ and DA: data collection and samples preparation; MM, AN and SBH: investigation and sequence analysis; SBH, NBK, DA and MM: modelling and in silico analysis; MM, SBH, NBK and MDF: writing; MM and MDF: funding acquisition, all authors commented on the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by Grants from Arabian Gulf University Research Fund -Kingdom of Bahrain and Civil Service Commission—State of Kuwait represents by the Ministry of Interior. The funding body had no role in the design of the study, collection, analysis, interpretation of data and in writing the manuscript.

## Availability of data and materials

The data sets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate

This study was considered and approved by an Arabian Gulf university research and ethics committee. Consent forms were signed by patient before collecting blood samples. Tissue samples were collected under the approval of Kuwait National Cancer Control Center.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that there is no conflict of interest.

### Author details

<sup>1</sup> Department of Life Sciences, Health Biotechnology Program, College of Graduate Studies, Arabian Gulf University, PO Box 26671 Manama, Kingdom of Bahrain. <sup>2</sup> King Fahad Chair for Health Biotechnology, College of Graduate Studies, Arabian Gulf University, PO Box 26671 Manama, Kingdom of Bahrain.

Received: 25 November 2020 Accepted: 8 April 2021

Published online: 19 April 2021

## References

1. Al Hamdan N, Ravichandran K, Al Sayyad J, Al Lawati J, Khazal Z, Al Khaateeb F, Abdulwahab A, Al Asfour A. Incidence of cancer in Gulf Cooperation Council countries, 1998–2001. *EMHJ Eastern Mediterranean Health Journal*. 2009;15(3):600–11.
2. Salim EI, Moore MA, Al-Lawati JA, Al-Sayyad J, Bazawir A, Bener A, Corbex M, El-Saghir N, Habib OS, Maziak W, et al. *Cancer epidemiology and*

- control in the arab world—past, present and future. *Asian Pac J Cancer Prev.* 2009;10(1):3–16.
3. Al-Madouj A, Eldali A, Al-Zahrani AJGcfc, prevention: ten-year cancer incidence among nationals of the GCC states 1998–2007. *Gulf Center for Cancer Control and Prevention* 2011.
  4. Harari PM. Epidermal growth factor receptor inhibition strategies in oncology. *Endocr Relat Cancer.* 2004;11(4):689.
  5. Wee P, Wang Z. Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers.* 2017;9(5):52.
  6. Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW, Burgess AW. Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp Cell Res.* 2003;284(1):31–53.
  7. Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *FEBS J.* 2010;277(2):301–8.
  8. Reiter JL, Threadgill DW, Eley GD, Strunk KE, Danielsen AJ, Sinclair CS, Pearsall RS, Green PJ, Yee D, Lampland AL, et al. Comparative genomic sequence analysis and isolation of human and mouse alternative EGFR transcripts encoding truncated receptor isoforms. *Genomics.* 2001;71(1):1–20.
  9. Oliveira S, van Bergen en Henegouwen PM, Storm G, Schifferers RM. Molecular biology of epidermal growth factor receptor inhibition for cancer therapy. *Expert Opin Biol Ther.* 2006;6(6):605–17.
  10. Holbro T, Civenni G, Hynes NE. The ErbB receptors and their role in cancer progression. *Exp Cell Res.* 2003;284(1):99–110.
  11. Roskoski R Jr. The ErbB/HER family of protein-tyrosine kinases and cancer. *Pharmacol Res.* 2014;79:34–74.
  12. Ferguson KM, Berger MB, Mendrola JM, Cho HS, Leahy DJ, Lemmon MA. EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell.* 2003;11(2):507–17.
  13. Burgess AW, Cho HS, Eigenbrot C, Ferguson KM, Garrett TP, Leahy DJ, Lemmon MA, Sliwkowski MX, Ward CW, Yokoyama S. An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell.* 2003;12(3):541–52.
  14. Li S, Schmitz KR, Jeffrey PD, Wiltzius JJ, Kussie P, Ferguson KM. Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell.* 2005;7(4):301–11.
  15. Martin-Fernandez ML, Clarke DT, Roberts SK, Zanetti-Domingues LC, Gervasio FL. Structure and dynamics of the EGF receptor as revealed by experiments and simulations and its relevance to non-small cell lung cancer. *Cells.* 2019;8(4):316.
  16. Wang WS, Chen PM, Chiou TJ, Liu JH, Lin JK, Lin TC, Wang HS, Su Y. Epidermal growth factor receptor R497K polymorphism is a favorable prognostic factor for patients with colorectal carcinoma. *Clin Cancer Res Off J Am Assoc Cancer Res.* 2007;13(12):3597–604.
  17. Martinelli M, Ugolini G, Scapoli L, Rivetti S, Lauriola M, Mattei G, Rosati G, Montroni I, Manaresi A, Zattoni D, et al. The EGFR R521K polymorphism influences the risk to develop colorectal cancer. *Cancer Biomarkers Sect A Disease Markers.* 2010;8(2):61–5.
  18. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med.* 2004;350(21):2129–39.
  19. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA.* 1977;74(12):5463–7.
  20. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):33–8 (27–38).
  21. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem.* 2005;26(16):1781–802.
  22. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model.* 2011;51(10):2778–86.
  23. Pandurangan AP, Ochoa-Montaño B, Ascher DB. Blundell TLJNar: SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res.* 2017;45(W1):W229–35.
  24. Zheng P, Ren L, Feng Q, Zhu D, Chang W, He G, Ji M, Jian M, Lin Q, Yi T, et al. Differences in clinical characteristics and mutational pattern between synchronous and metachronous colorectal liver metastases. *Cancer Manage Res.* 2018;10:2871.
  25. Han C, Liao X, Qin W, Yu L, Liu X, Chen G, Liu Z, Lu S, Chen Z, Su H, et al. EGFR and SYNE2 are associated with p21 expression and SYNE2 variants predict post-operative clinical outcomes in HBV-related hepatocellular carcinoma. *Sci Rep.* 2016;6(1):31237.
  26. Goncalves A, Esteyries S, Taylor-Smedra B, Lagarde A, Ayadi M, Monges G, Bertucci F, Esterni B, Delpero JR, Turrini O, et al. A polymorphism of EGFR extracellular domain is associated with progression free-survival in metastatic colorectal cancer patients receiving cetuximab-based treatment. *BMC Cancer.* 2008;8:169.
  27. Walker F, Orchard SG, Jorissen RN, Hall NE, Zhang HH, Hoyne PA, Adams TE, Johns TG, Ward C, Garrett TP, et al. CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor. *J Biol Chem.* 2004;279(21):22387–98.
  28. Moriai T, Kobrin MS, Hope C, Speck L, Korc M. A variant epidermal growth factor receptor exhibits altered type alpha transforming growth factor binding and transmembrane signaling. *Proc Natl Acad Sci USA.* 1994;91(21):10217–21.
  29. Hsieh YY, Tzeng CH, Chen MH, Chen PM, Wang WS. Epidermal growth factor receptor R521K polymorphism shows favorable outcomes in KRAS wild-type colorectal cancer patients treated with cetuximab-based chemotherapy. *Cancer Sci.* 2012;103(4):791–6.
  30. Sasaki H, Okuda K, Shimizu S, Takada M, Kawahara M, Kitahara N, Okumura M, Matsumura A, Iuchi K, Kawaguchi T. EGFR R497K polymorphism is a favorable prognostic factor for advanced lung cancer. *J Cancer Res Clin Oncol.* 2009;135(2):313.
  31. Leite MS, Giacomini LC, Piranda DN, Festa-Vasconcellos JS, Indio-do-Brasil V, Koifman S, de Moura-Neto RS, de Carvalho MA, Vianna-Jorge R. Epidermal growth factor receptor gene polymorphisms are associated with prognostic features of breast cancer. *BMC Cancer.* 2014;14:190.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

