Integrated Somatic and Germline Whole-Exome Sequencing Analysis in Women with Lung Cancer after a Previous Breast Cancer

Simona Coco [1,*,†], Silvia Bonfiglio [2,†], Davide Cittaro [2,†], Irene Vanni [1,3], Marco Mora [4], Carlo Genova [1,3], Maria Giovanna Dal Bello [1], Simona Boccardo [1], Angela Alama [1], Erika Rijavec [1], Claudio Sini [1], Valeria Rossella [5], Giulia Barletta [1], Federica Biello [1], Anna Truini [1], Cristina Bruzzo [1], Maurizio Gallo [3], Dejan Lazarevic [2], Alberto Ballestrero [3] and Francesco Grossi [6]

[1] Lung Cancer Unit, IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy; irenevanni85@yahoo.it (I.V.); carlo.genova1985@gmail.com (C.G.); mariagiovanna.dalbello@hsanmartino.it (M.G.D.B.); simona.boccardo@hsanmartino.it (S.B.); angela.alama@hsanmartino.it (A.A.); ery80x@yahoo.it (E.R.); audiosini@tiscali.it (C.S.); giulia.barletta@yahoo.it (G.B.); febiello@gmail.com (F.B.); anna.truini@gmail.com (A.T.); cristina.bruzzo@hsanmartino.it (C.B.)
[2] Centre for Translational Genomics and Bioinformatics, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy; bonfiglio.silvia@hsr.it (S.B.); cittaro.davide@hsr.it (D.C.); lazarevic.dejan@hsr.it (D.L.)
[3] Department of Internal Medicine and Medical Specialties (DIMI), University of Genoa, IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy; maurizio.gallo@unige.it (M.G.); aballestrero@unige.it (A.B.)
[4] Department of Pathology, IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy; marco.mora@hsanmartino.it
[5] Swiss Stem Cell Biotech, 6830 Vacallo, Switzerland; valeria.rossella@stembiotech.ch
[6] UOC Oncologia Medica Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, 20122 Milan, Italy; francesco.grossi@policlinico.mi.it
* Correspondence: simona.coco@hsanmartino.it; Tel.: +39-010-555-8316; Fax: +39-010-555-8335
† These authors contributed equally to the work.

check for updates

Abstract: Women treated for breast cancer (BC) are at risk of developing secondary tumors, such as lung cancer (LC). Since rare germline variants have been linked to multiple cancer development, we hypothesized that BC survivors might be prone to develop LC as a result of harboring rare variants. Sixty patients with LC with previous BC (the study population; SP) and 53 women with either BC or LC and no secondary cancer (control population; CP) were enrolled. Whole exome sequencing was performed in both tumors and unaffected tissues from 28/60 SP patients, and in germline DNA from 32/53 CP. Candidate genes were validated in the remaining individuals from both populations. We found two main mutational signature profiles: S1 (C>T) in all BCs and 16/28 LCs, and S2 (C>A) which is strongly associated with smoking, in 12/28 LCs. The burden test over rare germline variants in S1-LC vs CP identified 248 genes. Validation confirmed *GSN* as significantly associated with LC in never-smokers. In conclusion, our data suggest two signatures involved in LC onset in women with previous BC. One of these signatures is linked to smoking. Conversely, regardless of smoking habit, in a subgroup of BC survivors genetic susceptibility may contribute to LC risk.

Keywords: breast cancer; lung cancer; multiple cancer susceptibility; genetic predisposition; exome sequencing

1. Introduction

Breast cancer (BC) is one of the most common cancers, together with lung and colon, and it represents the most frequent malignancy in females worldwide [1]. Currently, the optimal management of early BC patients by integrated treatments, including radio-, chemo- and hormone therapies [1], have led to significant mortality reduction, with almost 90% of patients surviving at five years [2]. Despite the improvement in survival, approximately 10% of women with BC develop a second primary tumor, and lung cancer (LC) is one of the most frequently diagnosed tumors [3,4]. Several factors have been linked to LC co-occurrence in BC survivors, such as primary cancer treatment-related effects as well as exposure to carcinogens in tobacco products [5–7]. For decades, clinicians have hypothesized that radiotherapy could increase the risk of developing a second unrelated malignancy. Indeed, many evidences have shown a significant increase in the relative risk of developing LC after a BC in women receiving radiation therapy after a latency period of 5–20 years [8–10]. In particular, the risk increased in the ipsilateral lung and was more clearly related to post-mastectomy radiation [9]. A synergic effect of radiotherapy with smoking habit has also been reported [6,8]. In particular, smoking women who received post-mastectomy radiotherapy showed a 40-fold increase risk of an ipsilateral LC compared with never smokers who did not receive any adjuvant radio-treatment [11]. Although the exposure to carcinogens has been considered a risk factor of LC in BC survivors, substantial evidences argue that the genetic predisposition may also play a role in the development of multiple cancers [12]. Although several genetic variants have been associated with LC risk [13,14], a cancer susceptibility signature in the development of LC in BC survivors has not been elucidated yet. The advent of Next Generation Sequencing (NGS) has deeply revolutionized our knowledge of cancer molecular landscape [15–17]. The screening of the entire DNA coding region by whole exome sequencing (WES) represents the best option, compared to the expensive whole genome sequencing, in the detection of rare genetic variants involved in unknown genetic syndromes and cancer susceptibility [18,19].

In the present study we performed a WES analysis of both tumor (BC and LC) and germline DNA in a cohort of 28 patients that developed LC after BC (study population; SP) to identify both somatic and germline signatures clarifying the genetic susceptibility associated to LC development. To unravel any genetic predisposition, we applied a gene-based burden test over rare germline variants in the SP compared to a control population (CP), i.e., patients with a BC without any history of secondary cancers. Then, the candidate genes were validated in an independent cohort of 28 women with LC after a previous BC and compared to controls (10 women with BC and 11 women with LC but without any history of secondary cancers).

2. Results

2.1. Patients and Clinical Characteristics

Sixty women who developed LC after a previous BC (SP) and 42 patients with a primary BC without any secondary unrelated tumor (CP) were enrolled in the study. In addition, 11 women surgically resected for an adenocarcinoma of the lung, with no additional cancer history, were also recruited as control (CP-LC) for the Validation Set (VS). The clinical features and treatment information of SP and CP are summarized in Tables 1 and 2, respectively.

Table 1. Clinical data of study and control populations.

Study Population								Control Population							
Breast Cancer n. 60			Lung Cancer n. 60					Breast Cancer n. 42			Lung Cancer n. 11				
Age at diagnosis Breast Cancer	median 60	range 37–76	Age at diagnosis Lung Cancer		median 68	range 48–88		Age at diagnosis Breast Cancer	median 60	range 41–81	Age at diagnosis Lung Cancer		median 59	range 37–81	
Smoking habit	N.	%	Smoking habit		N.	%		Smoking habit	N.	%	Smoking habit		N.	%	
Smoker	28	46.7	Smoker		25	41.7		Smoker	9	21.4	Smoker		3	27.3	
Former smoker	8	13.3	Former smoker		11	18.3		Former smoker	2	4.8	Former smoker		1	9.1	
Never smoker	23	38.3	Never smoker		23	38.3		Never smoker	27	64.3	Never smoker		6	54.5	
n.a.	1	1.7	n.a.		1	1.7		n.a.	4	9.5	n.a.		1	9.1	
BC histology	N.	%	LC histology		N.*	% *		BC histology	N.	%	LC histology		N.	%	
Ductal Infiltrating	50	83.3	Adenocarcinoma		51	83.6		Ductal Infiltrating	33	78.6	Adenocarcinoma		11	100.0	
Lobular	5	8.3	Squamous cell carcinoma		5	8.2		Lobular	7	16.7	Squamous cell carcinoma		0	-	
Ductal/Lobular	1	1.7	Large cell		2	3.3		Ductal/Lobular	2	4.8	Large cell		0	-	
In situ ductal	2	3.3	Small cell lung cancer		3	4.9		In situ ductal	0	-	Small cell lung cancer		0	-	
n.a.	2	3.3	n.a.		0	-		n.a.	0	-	n.a.		0	-	
Surgery type	N.	%	Surgery type		N.	%		Surgery type	N.	%	Surgery Type		N.	%	
Quadrantectomy	35	58.3	Lobectomy		48	80.0		Quadrantectomy	12	28.6	Lobectomy		10	90.9	
Lumpectomy	8	13.3	Segmentectomy		3	5.0		Lumpectomy	19	45.2	Segmentectomy		1	9.1	
Mastectomy	17	28.3	Resection		2	3.3		Mastectomy	11	26.2	Resection		0	-	
No surgery [1]	0	-	No surgery [1]		6	10.0		No surgery [1]	0	-	No surgery [1]		0	-	
n.a.	0	-	n.a.		1	1.7		n.a.	0	-	n.a.		0	-	
Disease stage	N.	%	Disease stage		N.	%		Disease stage	N.	%	Disease stage		N.	%	
Non-metastatic	53	88.3	Non-metastatic		50	83.3		Non-metastatic	42	100.0	Non-metastatic		11	100.0	
Metastatic	0	-	Metastatic		7	11.7		Metastatic	0	-	Metastatic		0	-	
n.a.	7	11.7	n.a./n.a. [2]		3	5.0		n.a.	0	-	n.a./n.a.[2]		0	-	

* The sum and frequency of the lung cancer histotypes were calculated on 61 samples, since one patient developed two synchronous tumors of the lung (adenocarcinoma and squamous cell carcinoma). [1] Patient underwent a tumor biopsy. [2] Not applicable for small cell lung cancer histotype. Abbreviations: BC: Breast Cancer; LC: Lung Cancer; n.a.: not available.

Table 2. Breast cancer adjuvant therapies in study and control populations.

Study Population			Control Population		
Breast Cancer n. 60			**Breast Cancer n. 42**		
Adjuvant radiotherapy	N.	%	Adjuvant radiotherapy	N.	%
Yes	42	70.0	Yes	34	81.0
No	17	28.3	No	8	19.0
n.a.	1	1.7	n.a.	0	-
Adjuvant chemotherapy	N.	%	Adjuvant chemotherapy	N.	%
Yes	24	40.0	Yes	28	66.7
No	32	53.3	No	14	33.3
n.a.	4	6.7	n.a.	0	-
Hormone therapy	N.	%	Hormone therapy	N.	%
Yes	48	80.0	Yes	37	88.1
No	9	15.0	No	5	11.9
n.a.	3	5.0	n.a.	0	-

Abbreviations: n.a.: not available.

The median age for SP at BC diagnosis was 60 years (range: 37–76 years) and LC occurred nearly 8 years after BC. At BC diagnosis, 46.7% (28/60) and 13.3% (8/60) of the patients were current or former smokers, respectively, whereas three women quit smoking at the onset of LC. The most frequent BC histotype was invasive ductal carcinoma (86.6%), two of which were in situ. The majority of women (71.7%, 43/60) underwent a breast-conserving surgical approach, while 28.3% (17/60) had mastectomy. Every BC patient received at least one local and/or adjuvant systemic treatment (Table 2). Among the radio-exposed women, more than half (52.4%; 22/42) developed an ipsilateral LC. Approximately one third of women (23/60) of our cohort were high-risk patients, i.e. smoker at the time of radiotherapy, although only three underwent mastectomy before radiotherapy [11]. With regard to LC, the most common histotype was adenocarcinoma (83.6%) followed by squamous carcinoma (8.2%), small cell lung cancer (4.9%) and large cell carcinoma (3.3%); notably, one patient (#SP-60) developed two synchronous tumors of the lung: adenocarcinoma and squamous LCs, both contralateral to the primary BC.

2.2. Mutational Profile and Copy Number Alteration (CNA) of SP Breast Cancers

Breast cancer genomic DNA (gDNA) samples were isolated from formalin-fixed and paraffin-embedded (FFPE) tumor specimens from the 28 patients of the study population discovery set (SP-DS), who were enrolled in a wide period of time (1990–2014). Although the samples exhibited a heterogeneous degradation status (Supplementary Table S2), WES library preparation and sequencing were fully successful. However, somatic analysis revealed an unusual high mutational rate in 7/28 samples, with several hyper-mutated genes, suggesting false positive mutational calls probably linked to the formalin fixation. Therefore, we excluded these samples from the downstream analyses. Within the remaining 21 BC samples, we found a total of 420 mutated genes already described in cancer (see Appendix A). Each tumor showed at least one mutated gene (2–128), except sample #BC13 (Supplementary Table S3). Among the most frequently mutated genes in our cohort (Figure 1), some of them (*CDH1, KMT2C, GATA3, PIK3CA* and *TP53*) are known driver genes in breast cancer [20]. Other frequently mutated genes were also found in cBioPortal BC studies, although at lower frequencies than in our cohort (e.g., *BUB1B, MAGI2, MGA, PCNT, PDCD11, SPEN,* and *TRIOBP*). Notably, two triple negative BCs (ER−/PR−/HER2−) reported activating mutations in genes already described in this molecular subtype, such as *ATP8B2, CDC27, DHX37, FIN* and *SHROOM4* [21,22]. We then evaluated whether the 420 cancer genes carrying somatic mutations in our samples were also altered by the occurrence of Copy Number Alteration (CNA). Notably, some genes (*ATR, CDH1, KMT2C, GATA3, MGA, PIK3CA* and *TP53*) were also deregulated as a consequence of CNA (Supplementary Table S4; Figure 1).
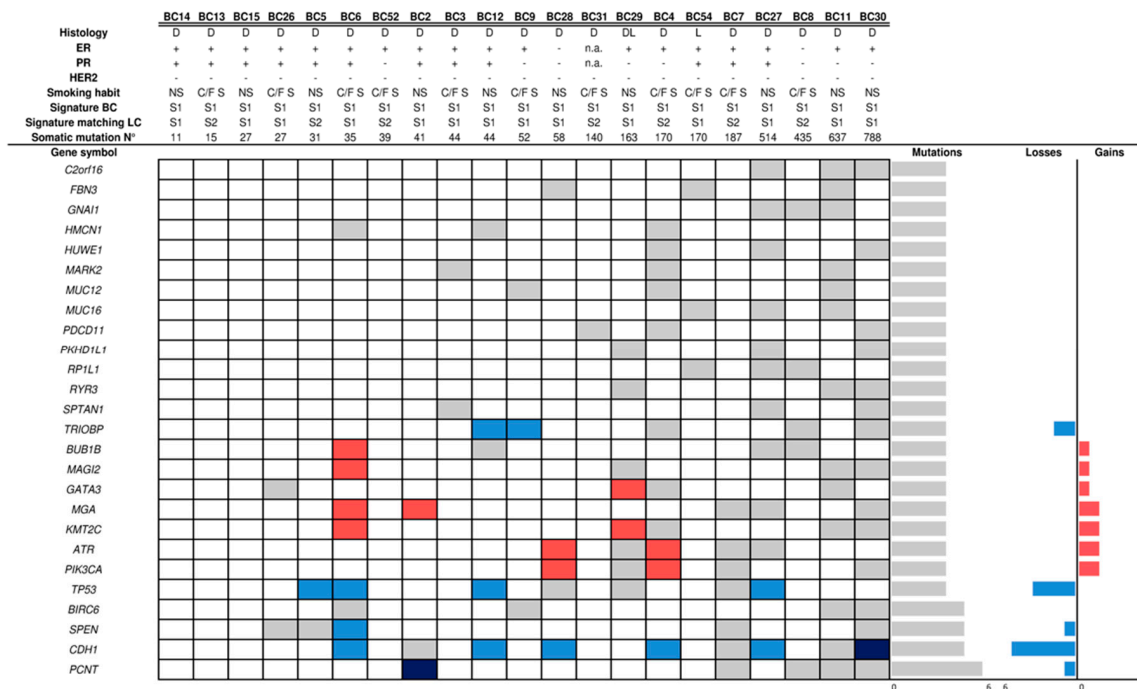
Figure 1. Mutated genes in breast cancer samples. The figure shows the mutational and copy number status of the 25 most frequently altered genes found in the cohort of BC samples obtained from SP patients. Grey, light red and light blue cells indicate the presence of a gene mutation, copy number gain and copy number loss, respectively. Dark red and dark blue cells indicate the concomitant presence of a gene mutation plus either gene copy number gain or loss, respectively. On the right of the picture, the number of mutations, and gene copy number losses and gains are shown. Abbreviations: BC: breast cancer; Histology: DL: ductal-lobular breast carcinoma, D: ductal breast carcinoma, L: lobular breast carcinoma; ER+: positive estrogen receptor ($\geq$ 1%), PR+: positive progesterone receptor ($\geq$ 1%); HER-2+: positive status (3+); Smoking Habit: NS: no smoker, C/F S: current/former smoker; Mutational Signature: S1: signature 1; S2: signature 2; N°: number; n.a.: not available.

2.3. Mutational Profile and CNAs of SP Lung Cancers

All 28 LC samples were collected at the Pathology Unit of the IRCCS Ospedale Policlinico San Martino (Genoa, Italy). Similarly to the BC group, despite FFPE LC gDNA exhibited a heterogeneous fragmentation, WES library preparation and sequencing were successful for all samples (Supplementary Table S2). The pair comparison between LC and matching BC samples (*n* = 21) revealed no common alteration. We then looked for pathways enriched for genes carrying somatic mutations. We found a total of five pathways significantly enriched (false discovery rate FDR < 0.01) in two tumors (#BC29 and #LC7), but no shared with paired samples. All pathways (R-HSA-5083636, R-HSA-5083625, R-HSA-5083632, R-HSA-977068, R-HSA-3906995) are related to the O-linked protein glycosylation; their enrichment is mainly caused by mutations in the mucin gene family (*MUC12*, *MUC17*, *MUC3A*, *MUC4*, and *MUC5B*).

Among the 28 LC samples, a total of 666 genes described in cancer was found mutated (see Appendix A), which were hit by at least one mutation (1–140), except sample #LC02 (Supplementary Table S5). The most frequently mutated gene in our cohort was *TP53* (42.8%; 12/28) (Figure 2), in agreement with data from the literature [20]. Other known LC driver genes were found mutated in our cohort. For example, 10/23 (43.5%) adenocarcinoma LCs harbored pathogenic mutations in *EGFR* (8/10 in exon 19 and 2/10 in exon 21). Across the variants of exon 19, beyond the canonical InDels (COSM6223, COSM6225 and COSM12678) we also found a deletion (p.Glu746_Arg748del) and rare nucleotide substitutions (COSM24267 and COSM51497) that were previously described [23]. In addition, two *KRAS* somatic mutations (Gly12Asp and Gly12Cys) were found in 2/23 adenocarcinoma LCs.

Other cancer genes frequently altered in LC according to cBioPortal (*e.g., ZFHX4, RYR2, CSMD3, FAT3,* and *RP1L1*) were also found mutated at a high frequency in our cohort (Figure 2 and Supplementary Table S6). The integration of mutational and CNA analysis, as a further mechanism of gene deregulation, showed both gains (*EGFR*) and losses (*TP53*) in some driver genes (Supplementary Table S6).

Since an increase in length of deletions across the genome has been described in some radiation-associated second malignancies [24], we investigated the potential radiotherapy related effect in our cohort of patients by regressing the extent of InDels (as found in our WES data) and chromosome aberrations (by Fraction of Genome Altered FGA) of LC samples, comparing radiation-exposed patients with unirradiated subjects. We did not find any significant association for either InDels (best *p*-value for logistic regression: 0.08 for deletions in the range [−5,0]) or FGA (logistic regression *p*-value = 0.844).
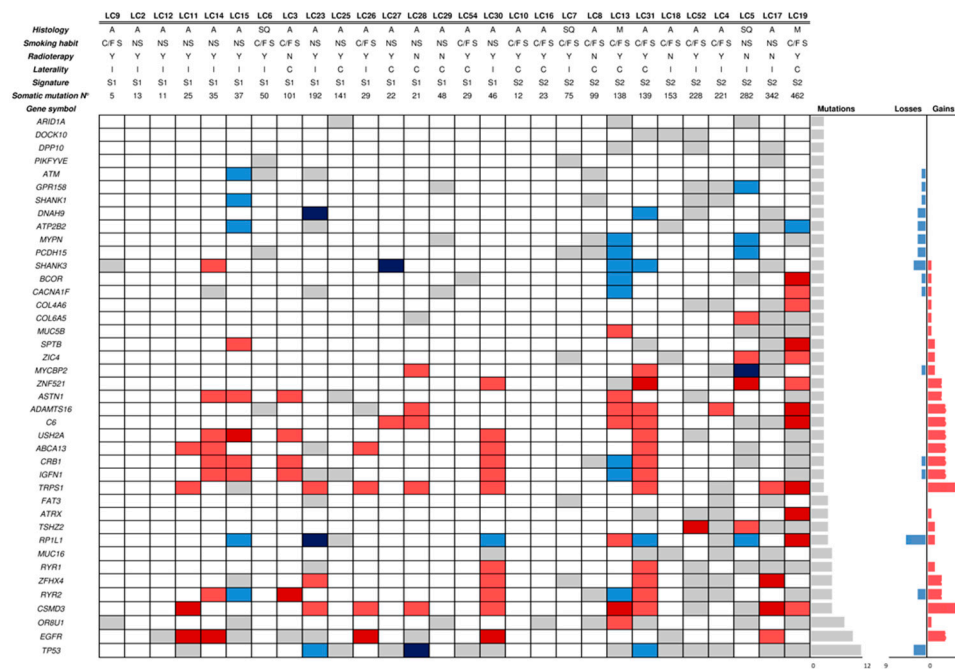

Figure 2. Mutated genes in lung cancer samples. The figure shows the mutational and copy number status of the 41 genes found most frequently altered in LC samples obtained from SP patients. Grey, light red and light blue cells represent the presence of a gene mutation, copy number gain and copy number loss, respectively. The dark red and dark blue cells represent the concomitant presence of a gene mutation plus either gene copy number gain or loss, respectively. On the right of the picture, the number of mutations, and gene copy number losses and gains are shown. Abbreviations: LC: lung cancer; A: adenocarcinoma; SC: squamous carcinoma; M: Microcitoma; NS: Never Smoker; C/F S: Current/Former Smoker; Y: radiotherapy; N: no-radiotherapy; I: LC ipsilateral versus BC; C: LC contralateral versus BC; Mutational Signature: S1: signature 1; S2: signature 2; N°: number; n.a.: not available.

2.4. Microsatellite Instability (MSI) Analysis

Alteration of the DNA mismatch repair system has been associated to hereditary cancer syndromes [25]. To this end, we evaluated the MSI status by assessing five nearly monomorphic markers on both BC and LC samples from 27/28 SP-DS patients (in patient #SP-23 the amplification of 4/5 markers of the normal gDNA failed, preventing MSI assessment in both tumors). In addition, all markers of five BC samples, previously excluded from WES, failed to amplify as a probable consequence of highly fragmented gDNA (Supplementary Table S2). MSI was assessed by comparing BC and LC profiles with the matching normal sample (Supplementary Figure S2a). No MSI was found in BC or LC; these data were also confirmed on WES data as measured by MANTIS score

(Supplementary Table S7). In one out of four patients, who showed a heterozygosity status within the monomorphic markers, an allelic imbalance at the NR21 was identified (Supplementary Figure S2b). In addition, we also found imbalances at the polymorphic loci in three BCs and in three LCs, although no loss of heterozygosity was shared between the tumor pairs from the same patient.

2.5. Mutational Signatures

In order to identify potential mutagenic processes underlying the development of LC in women with previous BC, we performed a mutational signature analysis. Two main mutational signatures were extracted from our cohort of 49 tumor samples (21 BC and 28 LC) [15]. All mutational signatures were compared to COSMIC signatures previously described in human cancers (Figure 3a,b; Supplementary Table S8) [15,26–28]. Signature 1 (S1), characterized predominantly by 'C>T substitutions', was found weakly similar to COSMIC S30 (cosine distance = 0.52). This signature was associated to all BC samples (21/21) as well as to a subgroup of LC (16/28). Conversely, Signature 2 (S2), dominated by 'C>A transversions', was exclusively found in LC (more specifically, 8/12 adenocarcinoma, 2/12 squamous and 2/12 small cell carcinoma) and showed a positive correlation with smoking (logistic regression *p*-value = 0.004; Table 3); specifically, 10 out of 12 patients were current/former smokers. Indeed, the S2 profile matched with COSMIC S4 (cosine distance = 0.35) which has been associated with smoking behavior. Since LC samples revealed two distinct mutational signatures, we next examined the clinical data of the two LC subgroups, as well as the potential effect linked to previous therapy for BC (Table 3). The S1-LC subgroup was mainly enriched in adenocarcinoma and the occurrence interval from primary BC was generally shorter (S1-LC 3.7 years, range: 0–11 versus S2-LC 7.7 years; range: 2–21). Overall, the therapies were not associated with the specific signature. Taken together, our data suggest that two different mutagenic processes drove the development of LC in our cohort. The presence of a tobacco-related signature hints that exposure to tobacco products played a predominant role in LC development in S2-LC subgroup of BC survivors. Conversely, the shared signature between BC and S1-LC group may underpin a genetic predisposition contributing to the development of both cancers.
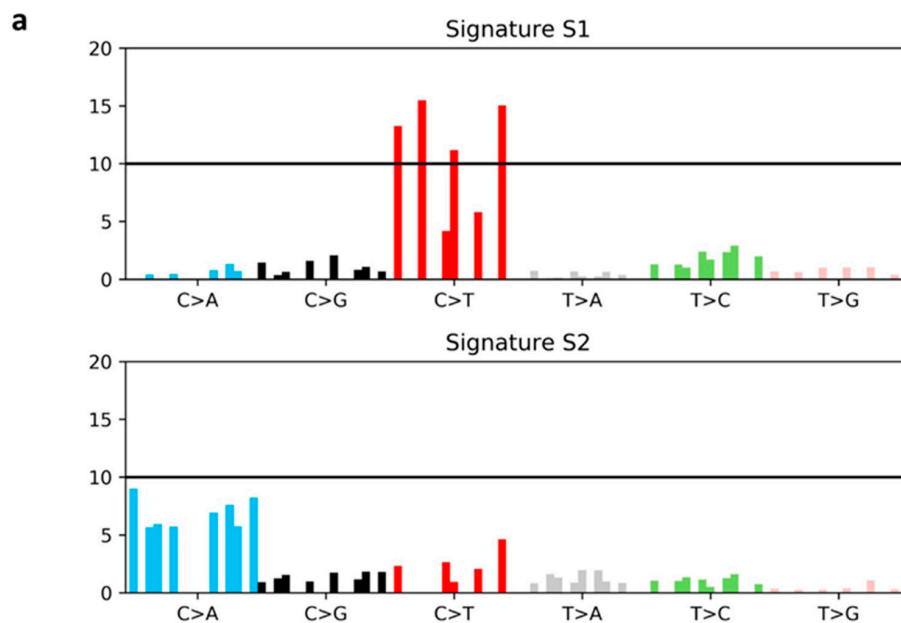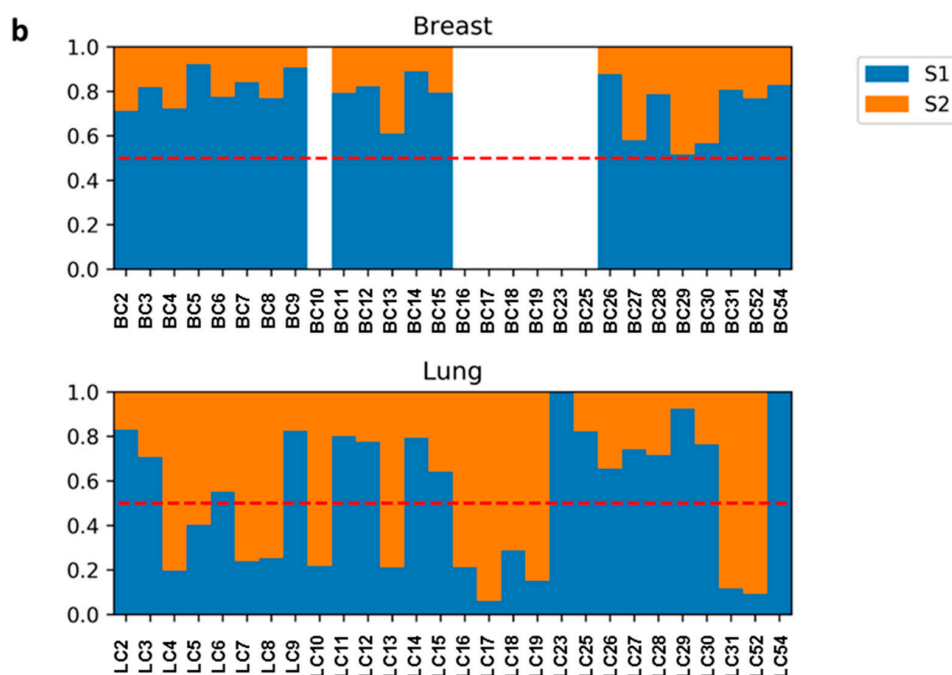

Figure 3. *Cont.*

Figure 3. Mutational Signatures (S) of breast and lung cancer samples. (**a**) Mutational signatures extracted from WES data. Each signature is displayed according to the 96 pyrimidine substitutions in their trinucleotide context, grouped and colored by the 6 classes of base substitutions. X-axis reports the mutation types; Y-Axis shows the frequency of mutations attributed to a specific mutation type. (**b**) Contribution of each tumor type to the extracted mutational signatures. The BC samples are reported at the top whereas the LC samples were depicted at the bottom. BC samples not used in the analysis are masked in white. The dotted red line represents the threshold used to perform signature attribution. Abbreviations: BC: breast cancer; LC: lung cancer; A: adenine; C: cytosine; G: guanine; T: thymine.

Table 3. Clinical data of study population patients according to mutational signature of lung cancer.

Clinical Data of Lung Cancer	S1		S2	
	N.	%	N.	%
Lung cancer	16	57.1%	12	42.9%
Age at diagnosis	67.5	56–75	68	53–78
LC occurrence after BC (years)	3.7	0–11	7.7	2–21
LC histotype	**N.**	%	**N.**	%
Adenocarcinoma	15	93.8%	8	66.7%
No adenocarcinoma	1	6.3%	4	33.3%
Smoking habit	**N.**	%	**N.**	%
Former/current smoker	5	31.3%	10	62.5%
Never smoker	11	68.8%	2	12.5%
Adjuvant radiotherapy for BC	**N.**	%	**N.**	%
Yes	13	81.3%	8	66.7%
No	3	18.8%	4	33.3%
Adjuvant chemotherapy for BC	**N.**	%	**N.**	%
Yes	6	37.5%	3	25.0%
No	10	62.5%	9	75.0%
Adjuvant hormone therapy for BC	**N.**	%	**N.**	%
Yes	15	93.8%	10	83.3%
No	1	6.3%	2	16.7%

Abbreviations: BC: breast cancer; LC: lung cancer; S: mutational signature.

2.6. WES Germline Analysis and Validation

In order to unravel a genetic contribution to the development of LC after BC in SP patients characterized by S1-LC signature (*n* = 16), we analyzed the differential distribution of rare germline variants (Allele Frequency in ExAC < 1%) in this group versus 32 out of 42 CP patients affected by BC only. Our gene-based burden test was performed on a subset of 4618 genes carrying more than one rare variant, identifying 248 candidate genes enriched with rare variants in SP versus control individuals (combined FDR < 0.05) (Supplementary Table S9). Interestingly, our panel included genes already described to be cancer predisposition genes (*FAS, ITK, KIF1B, PGR* and *MET*) or previously reported as playing important roles in cancer (*ABI1, ARID5B, DNMT3A, HEY1, KDM5C, NCOA1, NUP98,* and *SLC34A2*), or in DNA repair process (*FANCF*) [29–34]. We validated the genes found in the discovery cohort by a custom panel targeting all the exons of the 248 candidate genes. The library preparation was performed in 49 patients from VS (28 SP, 10 BC-CP, and 11 LC-CP) (Supplementary Figure S1). Library preparation and sequencing were successful for all samples (Supplementary Table S10). The distribution of rare germline variants was initially tested in a population with the highest similarity to our initial cohort. Initially, we excluded 19 women with a history of smoking (former/current smokers) from the cases, assuming that the carcinogenic effect of tobacco may be a major determinant of LC development. Burden test, performed on 63/248 genes carrying more than one rare variant in this cohort, revealed that only *GSN* gene resulted significantly associated to LC in non-smoker patients (*p*-value = 1.06e−5, FDR = 6.67e−4, Supplementary Table S9). We identified three rare single nucleotide variants in four SP patients, two in DS and two in VS. Specifically, the missense variants rs766602997 (ExAC-MAF = 0.0033%) and rs116185403 (ExAC-MAF = 0.27%) with a predicted moderate impact, were carried by one (#SP-43) and two patients (#SP-27 and #SP-59) respectively. The third variant, (c.1927G>T; p.Glu643*) identified in #SP-03 patient, is a novel one and has a predicted loss-of-function. No gene was found significantly enriched in rare variants when smokers were included in the analysis, although 25 genes harbored at least two more rare variants in cases as compared to controls (Supplementary Table S9). Finally, to investigate additional biological processes that may contribute to the development of LC in patients with a previous BC, we aggregated the genes carrying at least one rare germline variant (allele frequency in ExAC < 1%) at the level of reactome pathways [35], and we then performed the burden test in SP patients with S1-LC signature versus 32/42 CP patients affected by BC only. We retrieved two pathways significantly enriched in cases versus controls (combined FDR < 0.05) (Supplementary Table S11).

3. Discussion

This study includes a population of 60 women surviving a BC, who developed a LC in an interval of approximately 7 years, in agreement with previous studies [8–10]. WES was performed in 28 patients including both tumors and normal tissue. No common mutation was shared between the two primary cancers, indicating their independent origin. Similarly, the pathway analysis did not disclose any common enrichment across paired samples but processes linked to mucin genes were highlighted in two unrelated tumor samples (a BC and a LC from two patients); their role in tumorigenesis remains questionable [36] as they are highly polymorphic [37] and many of them have been ruled out from the list of known cancer genes [38].

Analysis of somatic mutations revealed that the distribution of mutated genes in both BC and LC samples from SP-DS cohort is in line with reported studies for these tumors. Specifically, BCs harbored mutations in *PI3KCA, CDH1, TP53* and some other driver cancer genes already found in BC [20], although with weak differences in their frequencies. A possible explanation for this discrepancy might be due to the small number of samples analyzed as well as the unbalanced early stage BCs in our cohort; e.g., *TP53* is found more frequently mutated in metastatic patients than in early stage tumor [39]. The most frequent driver alterations in LC samples involved *TP53* and *EGFR* genes. We observed a higher frequency of *EGFR* mutations (36%) compared to published data (15–20%) in Caucasian patients [40]. However, our finding should not be surprising, since our cohort includes exclusively

women and was enriched in never smokers with adenocarcinoma histology [41]. Deficiency of mismatch repair has also been described in the development of multiple tumors as well as in sporadic cancers, but at very low incidence (0.5–2%) [25,42]. Since data on this peculiar BC-LC population are still lacking, we investigated the MSI status. No significant result has been disclosed and data were also confirmed across all-coding regions by MANTIS score. Emerging evidences have demonstrated that each oncogenic process would leave a mutational signature in tumor DNA deriving from a combination of DNA repair deregulation and DNA-damaging agents e.g., smoking and radiations [28]. In this regard, Behjati and colleagues [24], investigating the genome of radio-induced tumors, identified radiation-related signatures irrespective of the tumor type. However, we did not find any significant pattern of alterations across the exome between radio-exposed *vs.* non-exposed women, hence we excluded any direct effect of radiotherapy on LC development in our cohort. Beyond radiation-related signatures, 30 distinct mutational profiles have been described from all cancers and some of them were linked to specific causes [15,26–28]. Analysis of mutational signatures in our cohort identified two main profiles characterized by the enrichment of distinct base substitutions: "C>T" (S1) and "C>A" (S2). Transition of C>T is one of the most common DNA modifications across the COSMIC signatures, and some of them have been linked to specific etiology, e.g., patient age [15]. However, our S1 was peculiar showing a slight similar pattern with COSMIC S30, described in a subgroup of BCs and of unknown etiology [43]. Conversely, the S2 profile clearly showed that tobacco smoking is linked to the emergence of LC in at least 12 samples; in this subset, the somatic signature resembles COSMIC S4, already linked to tobacco and confirmed in this study by regression with smoking habits [15,27,28]. According to these findings, we performed a gene-based burden test over rare germline variants restricted to a S1-LC of SP identifying 248 candidate genes enriched of rare variants compared to the CP. Among the genes significantly enriched in rare variants in SP-DS cohort, 14 were already known to be associated to cancer [32–34], suggesting that our cohort may at least cover known mechanisms of genetic predisposition to cancer. When we validated the panel on an independent cohort, a single gene (*GSN*) was found significantly enriched in rare variants in non-smoker cases vs. controls. The *GSN* gene (located on 9q33.2) encodes for a calcium-regulated protein mainly involved in the regulation of actin dynamics. The GSN protein has multifunctional roles inside the cell by coordinating several signal transduction pathways or acting as transcriptional regulator in both physiological and pathological conditions. In particular, GSN has been demonstrated to regulate apoptosis, proliferation as well as metastatic process in several cancers [44,45]. Evidences suggest that *GSN* may act as a tumor suppressor gene, indeed, several studies showed a GSN down-regulation in human cancers including breast carcinoma [44,46]. Notably, Wang and colleagues in 2014 reported that GSN may regulate the sensitivity to chemotherapy in patients with head-and-neck cancer [47], whereas more recently Zhao and colleagues have reported the GSN involvement in radio-resistant LC patients [48]. In support of this, Li and colleagues demonstrated a protective role of GSN against radiation-induced secondary injury in mice, by promoting the tissue regeneration [49]. Beyond the protein deregulation, genetic changes within its exon regions have been also linked to cancer predisposition such as breast and oral squamous cell carcinoma [46,50]. Taken together, these data suggest that patients harboring *GSN* gene defects may increase their susceptibility for developing a cancer as well as they would be more prone to the cancerous effects of the anticancer therapies. However, the validation on a larger cohort including smokers failed to score significant genes enriched in rare variants, possibly suggesting that the confounding effect of smoke might cover any genetic signal.

Although our study shows that genetic background may influence the LC onset in BC patients providing new insights in LC etiology, our results should be read in light of some considerations. We are aware that the size of SP is relatively small, but it should be taken into account that the occurrence of secondary LC remains a rare event in patients with a previous BC. Indeed, when we checked public data (TCGA: Lung adenocarcinoma [LUAD] and Lung squamous cell carcinoma [LUSC] studies) we found only seven out of 411 female individuals with LC and previous BC and all of them were reported to be current/former smokers [51]. Contrarily to genome-wide association studies,

WES-based studies are generally underpowered to evaluate the impact of single variants [52]; therefore, we chose to perform a burden test, aiming at the identification of genes which accumulate rare variants in our SP. We are aware that such tests do not weigh variants for their position in gene sequence nor the mutation type [53]. Therefore, we limited our analysis on variants predicted altering protein sequence. To enforce our results, we validated our candidate genes in an independent cohort of SP patients and CP. In particular, our CP included, over patients with BC without secondary malignancies, an additional group of women with lung adenocarcinoma and no history of multiple cancers, aiming to exclude any gene involvement in the LC carcinogenesis alone. This allowed us to select a list of genes, some of them already linked to cancer risk, among which *GSN* was confirmed to have a statistical significance. Although we analyzed only a small subgroup of BC survivors we found out rare variants in *GSN* in 16% (four out of 25) of the patients that are supposed to exhibit a genetic susceptibility to LC (S1-LC). In addition, the burden analysis at pathway level revealed that rare variants in the nicotinic acetylcholine receptor genes are enriched in S1-LC as compared with controls; interestingly, genetic variations in these receptors have been associated to lung cancer risk [54,55]. Finally, despite we excluded any direct effect of the radiation therapy on DNA, previous therapeutic treatments may also have contributed to LC onset, e.g., by lipid peroxidation and protein oxidation [56]. Therefore, we can assume that the development of LC in women with a previous BC might be the result of a combination of different factors, of which the genetic susceptibility would represent only a piece of the "omics" puzzle.

4. Materials and Methods

4.1. Patient Enrolment and Sample Collection

Between January 2006 and December 2017, 60 women (SP) with a LC diagnosis were admitted to the Lung Cancer Unit of the IRCCS Ospedale Policlinico San Martino (Genoa, Italy), reporting a medical history of previous BC in an interval ranging from few months to 31 years before LC (Supplementary Table S1A). In 28/60 patients (DS), FFPE surgical specimens from both BC and LC were collected, together with normal tissue specimens, for the WES and MSI analysis (Supplementary Figure S1). All tumor samples were checked by experienced pathologists for the presence of adequate tumor cell content (>50%). For the remaining 32/60 patients, tumor tissue from both tumors was unavailable for WES, therefore they were included in the VS and a normal tissue sample was provided for the germline analysis. In addition, two CPs including 42 women with primary BC (CP-BC) and 11 women with primary LC (CP-LC), without any history of multiple cancers, were enrolled and normal tissue samples were collected for each subject (Supplementary Figure S1). Both CPs had a follow-up of at least 10 years (range: 10–16 years) from the cancer diagnosis (Supplementary Table S1B,C). The present study was approved by the Local Ethics Committee (TrPo12.006) and conducted according to the provisions of the Declaration of Helsinki. For each patient included in the study a written informed consent was obtained.

4.2. Genomic DNA Extraction

Genomic DNA (gDNA) from FFPE tumor and normal tissues was isolated using GeneRead DNA FFPE Kit or QIAamp®DNA Mini Kit (Qiagen, Hilden, Germany). The gDNA concentration was assessed by Qubit®2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA), whereas the FFPE gDNA degradation status was assessed by DNA Integrity Number (DIN) value on TapeStation 2200 (Agilent Technologies, Santa Clara, CA, USA).

4.3. WES and Custom NGS Panel Library Preparation and Sequencing

WES was performed using the Agilent SureSelect Human All Exon kit v6 (Agilent Technologies) as previously described [57]. Libraries were sequenced on an Illumina HiSeq 2500 platform, 2×100 bp (Illumina Inc., San Diego, CA, USA). A custom design targeting all exons of 248 genes

(VS) was performed by Agilent SureDesign software (https://earray.chem.agilent.com/suredesign/). Libraries were prepared using Agilent SureSelect XT HS kit (Agilent Technologies). Briefly, 200 ng of gDNA were fragmented on a Covaris E220 focused ultrasonicator (Covaris, Woburn, MA, USA) and libraries were prepared following the manufacturer's protocol and sequenced on the NextSeq 500 platform, 2×100 bp (Illumina). Detailed information on the bioinformatics analysis is provided in the Appendix A.

4.4. Burden Test

Burden tests were performed using the AssotesteR R package (https://cran.rproject.org/web/packages/AssotesteR/index.html). In particular, the WSS test and C-alpha test were employed. The *p*-values of both tests were combined using Stouffer's method and corrected for multiple tests [53,58].

4.5. MSI Analysis

MSI was evaluated by assessment of the status of 5 nearly monomorphic markers on both BC and LC of the 28 SP patients enrolled in the DS. MSI analysis was carried out by MSI Analysis System v. 1.2, (Promega, Madison, WI, USA) covering five quasi-monomorphic mononucleotide markers (NR21, BAT26, BAT25, NR24, and MONO 27) for MSI detection and two highly-polymorphic penta-nucleotide repeat markers (Penta C and Penta D) for specimen identification. A multiplex PCR assay was carried out according to the manufacturer's protocol. Briefly, 8 ng of input gDNA was amplified according to manufacturer's instruction. All PCR products were denatured in Hi-Di™ Formamide (ThermoFisher Scientific, Carlsbad, CA, USA) with Internal Lane Standard 600 (Promega) and analyzed on a 3130xl Genetic Analyzer using the GeneMarker analysis software (Softgenetics LLC, State College, PA, USA). Loss of heterozigosity was defined when the ratio of tumor alleles and normal alleles resulted <0.5 or >2.0 [59].

5. Conclusions

To the best of our knowledge, this study is the first attempt to probe the genetics underlying the development of consecutive primary breast and lung tumors in the same patient. Although the sample size of this study is relatively small, our data provide a comprehensive picture of this peculiar population. In particular, we confirm their independent origin; in addition, we also show that smoke, a well-established risk factor in the general LC population, also represents the main cause of LC in a subgroup of women with BC. However, we also report that in a subgroup of patients, supposedly no-smoking women, the genetic background may contribute to increase their susceptibility to LC development. In particular, we identified a list of candidate genes among which *GSN* was significantly associated to LC risk. More extensive well-powered and in-depth sequencing studies are required to confirm the link of *GSN* gene to the predisposition of LC in BC survivors.

Author Contributions: Conceptualization, S.C. and F.G.; methodology, S.C., S.B. (Silvia Bonfiglio) and D.C.; software, D.C.; validation, S.B. (Silvia Bonfiglio); formal analysis, S.B. (Silvia Bonfiglio) and D.C.; investigation, S.C., S.B. (Silvia Bonfiglio), I.V., M.M., M.G.D.B., S.B. (Simona Boccardo), A.A., V.R., A.T., C.B. and D.L.; resources, C.G., E.R., C.S., G.B., F.B., M.G., A.B. and F.G.; data curation, D.C.; writing—original draft preparation, S.C.; writing—review and editing, S.B. (Silvia Bonfiglio) and D.C.; visualization, S.C. and S.B. (Silvia Bonfiglio); supervision, S.C. and D.C.; project administration, S.C.; funding acquisition, S.C.

Conflicts of Interest: A. Truini is currently an employee of MSD, however, these studies were conducted when she was working at IRCCS Ospedale Policlinico San Martino, Genoa, Italy. The other authors declare no conflicts of interest.

Appendix A

WES Bioinformatics Analysis

WES analysis was performed as previously described [60]. The sequencing raw data were converted to fastq data by bcl2fastq software (Illumina) and then aligned to the human reference genome (hg19) using BWA-MEM aligner [61]. Duplicated reads were assessed by Picard MarkDuplicates and WES metrics were determined by Picard HsMetrics and Samtools.

Somatic Analysis

Variant calling was performed with FreeBayes 1.0.2 across the Agilent v6 target region using the following parameters: F 0.2; C 2; q 20—pooled-continuous—min-repeat-entropy1. In order to avoid artifacts, the variant list was further filtered to select variants having a variant quality (QUAL) higher than 1; at least one alternate observation on the forward strand (Number of alternate observations on the forward strand; SAF > 0) and one on the reverse strand (Number of alternate observations on the reverse strand; SAR > 0); more than one read supporting the alternate balanced to both the left (5′) and the right (3′) of the alternate allele; mean mapping quality of observed alternate allele and mean mapping quality of observed reference alleles >50. In each trio (BC, LC, and normal samples), BC and LC somatic variants were identified by applying a somatic filter described in the SpeedSeq pipeline [62]. Only the variants with a variant allele frequency ≤5% in the normal sample, and covered at least 20X in tumor sample and at least 10X in normal sample, were considered for further analysis. Variant effect prediction was assessed by SnpEff and the following databases were used for annotation [63]: dbSNP 146, COSMIC v. 80 [64], ExAC [65], and dbNSFP v. 2.9 [66]. Genes bearing somatic mutations with "high" or "moderate" impact were retained and checked against comprehensive lists of known cancer genes. The gene lists mutated in the BC and the LC samples were checked in three published catalogues: IntoGen, Cancer Gene Census, and DriverDBv2 (http://driverdb.tms.cmu.edu.tw/driverdbv2) [67–69]. The mutational signature analysis was performed as described by Alexandrov and colleagues using custom scripts [15]. Counts of all 96 possible pyrimidine to purine substitution in their tri-nucleotide context were used to build a frequency matrix. Non-negative Matrix Factorization was then applied; the choice of the number of factors was based on the profile of the reconstruction error (β-divergence) curve. The MSI analysis was performed using MANTIS (v. 1.0.3) as previously described [70]. MANTIS score value higher than 0.4 identifies a tumor MSI, with a range from 0.0 (entirely stable) to 2.0 (entirely unstable). Pathway analysis was performed on Reactome Pathways using Enrichr [71].

Copy Number Alteration Analysis

CNA were called using CNV kit pipeline [72], using default parameters; all the analyses were performed on the aforementioned aligned reads. Individuals were analyzed separately and samples from normal tissues were used to build the reference for the corresponding tumors, either BC or LC. Segmentation data were used to calculate FGA as the span of segments with $|\log_2(CN)| > 0.2$ over the total span of segments analyzed.

Germline Analysis

Germline variant calling was performed in normal samples of SP and CP DS patients with FreeBayes 1.0.2 across the Agilent v6 target region, using the following parameters: F 0.2; C 2; q 20—min-repeat-entropy1. The same filters described above were applied in order to avoid artifacts. Variants were further filtered so that at least 10 samples reach at least 10X coverage for each genomic position in the vcf file. Germline variants, with an Allele Frequency < 10% in our samples and with a predicted moderate or high impact on protein function, were further filtered to select rare variants, according to their frequency in ExAC database (Minor Allele Frequency: MAF < 1%).

Custom Panel Bioinformatics Analysis

Raw sequences were preprocessed using the Agilent Genomics NextGen Toolkit (Agilent Technologies) v 4.0.1. SureCallTrimmer (Agilent Technologies) was used to trim low quality basecalls with the parameters '-xt -qualityTrimming 10'. After alignment with bwa, molecular barcodes were processed with LocatIt with the parameters '-i-IB-OB-d 1'. Variant calling across the custom target region and variant annotation were performed as described above for WES analysis.

References

1. Harbeck, N.; Gnant, M. Breast cancer. *Lancet* **2017**, *389*, 1134–1150. [CrossRef]
2. Surveillance, Epidemiology, and End Results Program (SEER). SEER Stat Fact Sheets: Female Breast Cancer. Available online: https://seer.cancer.gov/statfacts/html/breast.html (accessed on 4 October 2017).
3. Raymond, J.S.; Hogue, C.J. Multiple primary tumours in women following breast cancer, 1973–2000. *Br. J. Cancer* **2006**, *94*, 1745–1750. [CrossRef] [PubMed]
4. Levi, F.; Randimbison, L.; Te, V.C.; La Vecchia, C. Cancer risk after radiotherapy for breast cancer. *Br. J. Cancer* **2006**, *95*, 390–392. [CrossRef]
5. Neugut, A.I.; Robinson, E.; Lee, W.C.; Murray, T.; Karwoski, K.; Kutcher, G.J. Lung cancer after radiation therapy for breast cancer. *Cancer* **1993**, *71*, 3054–3057. [CrossRef]
6. Neugut, A.I.; Murray, T.; Santos, J.; Murray, T.; Karwoski, K.; Kutcher, G.J. Increased risk of lung cancer after breast cancer radiation therapy in cigarette smokers. *Cancer* **1994**, *73*, 1615–1620. [CrossRef]
7. Kirova, Y.M.; Gambotti, L.; De Rycke, Y.; Vilcoq, J.R.; Asselain, B.; Fourquet, A. Risk of second malignancies after adjuvant radiotherapy for breast cancer: A large-scale, single-institution review. *Int. J. Radiat. Oncol. Biol. Phys.* **2007**, *68*, 359–363. [CrossRef]
8. Prochazka, M.; Granath, F.; Ekbom, A.; Shields, P.G.; Hall, P. Lung cancer risks in women with previous breast cancer. *Eur. J. Cancer* **2002**, *38*, 1520–1525. [CrossRef]
9. Zablotska, L.B.; Neugut, A.I. Lung carcinoma after radiation therapy in women treated with lumpectomy or mastectomy for primary breast carcinoma. *Cancer* **2003**, *97*, 1404–1411. [CrossRef]
10. Huang, Y.J.; Huang, T.W.; Lin, F.H.; Chung, C.H.; Tsao, C.H.; Chien, W.C. Radiation Therapy for Invasive Breast Cancer Increases the Risk of Second Primary Lung Cancer: A Nationwide Population-Based Cohort Analysis. *J. Thorac. Oncol.* **2017**, *12*, 782–790. [CrossRef]
11. Kaufman, E.L.; Jacobson, J.S.; Hershman, D.L.; Desai, M.; Neugut, A.I. Effect of breast cancer radiotherapy and cigarette smoking on risk of second primary lung cancer. *J. Clin. Oncol.* **2008**, *26*, 392–398. [CrossRef]
12. Cybulski, C.; Nazarali, S.; Narod, S.A. Multiple primary cancers as a guide to heritability. *Int. J. Cancer* **2014**, *135*, 1756–1763. [CrossRef] [PubMed]
13. McKay, J.D.; Hung, R.J.; Gaborieau, V.; Boffetta, P.; Chabrier, A.; Byrnes, G.; Zaridze, D.; Mukeria, A.; Szeszenia-Dabrowska, N.; Lissowska, J. Lung cancer susceptibility locus at 5p15.33. *Nat. Genet.* **2008**, *40*, 1404–1406. [CrossRef] [PubMed]
14. Timofeeva, M.N.; Hung, R.J.; Rafnar, T.; Christiani, D.C.; Field, J.K.; Bickeböller, H.; Risch, A.; McKay, J.D.; Wang, Y.; Dai, J.; et al. Influence of common genetic variation on lung cancer risk: Meta-analysis of 14 900 cases and 29 485 controls. *Hum. Mol. Genet.* **2012**, *21*, 4980–4995. [CrossRef] [PubMed]
15. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Aparicio, S.A.; Behjati, S.; Biankin, A.V.; Bignell, G.R.; Bolli, N.; Borg, A.; Børresen-Dale, A.L.; et al. Signatures of mutational processes in human cancer. *Nature* **2013**, *500*, 415–421. [CrossRef] [PubMed]

16. Lawrence, M.S.; Stojanov, P.; Mermel, C.H.; Robinson, J.T.; Garraway, L.A.; Golub, T.R.; Meyerson, M.; Gabriel, S.B.; Lander, E.S.; Getz, G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **2014**, *505*, 495–501. [CrossRef]

17. Coco, S.; Truini, A.; Vanni, I.; Dal Bello, M.G.; Alama, A.; Rijavec, E.; Genova, C.; Barletta, G.; Sini, C.; Burrafato, G.; et al. Next generation sequencing in non-small cell lung cancer: New avenues toward the personalized medicine. *Curr. Drug Targets* **2015**, *16*, 47–59. [CrossRef]

18. Mathur, P.; Medicherla, K.M.; Chaudhary, S.; Patel, M.; Bagali, P.; Suravajhala, P. Whole exome sequencing reveals rare variants linked to congenital pouch colon. *Sci. Rep.* **2018**, *8*, 6646. [CrossRef] [PubMed]

19. Artomov, M.; Stratigos, A.J.; Kim, I.; Kumar, R.; Lauss, M.; Reddy, B.Y.; Miao, B.; Daniela Robles-Espinoza, C.; Sankar, A.; Njauw, C.N.; et al. Rare Variant, Gene-Based Association Study of Hereditary Melanoma Using Whole-Exome Sequencing. *J. Natl. Cancer Inst.* **2017**, *109*. [CrossRef] [PubMed]

20. Bailey, M.H.; Tokheim, C.; Porta-Pardo, E.; Sengupta, S.; Bertrand, D.; Weerasinghe, A.; Colaprico, A.; Wendl, M.C.; Kim, J.; Reardon, B.; et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **2018**, *174*, 1034–1035. [CrossRef] [PubMed]

21. Vaca-Paniagua, F.; Alvarez-Gomez, R.M.; Maldonado-Martínez, H.A.; Pérez-Plasencia, C.; Fragoso-Ontiveros, V.; Lasa-Gonsebatt, F.; Herrera, L.A.; Cantú, D.; Bargallo-Rocha, E.; Mohar, A.; et al. Revealing the Molecular Portrait of Triple Negative Breast Tumors in an Understudied Population through Omics Analysis of Formalin-Fixed and Paraffin-Embedded Tissues. *PLoS ONE* **2015**, *10*, e0126762. [CrossRef]

22. Shah, S.P.; Roth, A.; Goya, R.; Oloumi, A.; Ha, G.; Zhao, Y.; Turashvili, G.; Ding, J.; Tse, K.; Haffari, G.; et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **2012**, *486*, 395–399. [CrossRef] [PubMed]

23. Coco, S.; Truini, A.; Vanni, I.; Genova, C.; Rosano, C.; Dal Bello, M.G.; Alama, A.; Venè, R.; Rijavec, E.; Barletta, G.; et al. Uncommon EGFR Exon 19 Mutations Confer Gefitinib Resistance in Advanced Lung Adenocarcinoma. *J. Thorac. Oncol.* **2015**, *10*, e50–e52. [CrossRef] [PubMed]

24. Behjati, S.; Gundem, G.; Wedge, D.C.; Roberts, N.D.; Tarpey, P.S.; Cooke, S.L.; Van Loo, P.; Alexandrov, L.B.; Ramakrishna, M.; Davies, H.; et al. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **2016**, *7*, 12605. [CrossRef] [PubMed]

25. Calistri, D.; Presciuttini, S.; Buonsanti, G.; Radice, P.; Gazzoli, I.; Pensotti, V.; Sala, P.; Eboli, M.; Andreola, S.; Russo, A.; et al. Microsatellite instability in colorectal-cancer patients with suspected genetic predisposition. *Int. J. Cancer* **2000**, *89*, 87–91. [CrossRef]

26. Nik-Zainal, S.; Alexandrov, L.B.; Wedge, D.C.; Van Loo, P.; Greenman, C.D.; Raine, K.; Jones, D.; Hinton, J.; Marshall, J.; Stebbings, L.A.; et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **2012**, *149*, 979–993. [CrossRef]

27. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Campbell, P.J.; Stratton, M.R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **2013**, *3*, 246–259. [CrossRef] [PubMed]

28. Helleday, T.; Eshtad, S.; Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **2014**, *15*, 585–598. [CrossRef]

29. Zhang, X.; Miao, X.; Sun, T.; Tan, W.; Qu, S.; Xiong, P.; Zhou, Y.; Lin, D. Functional polymorphisms in cell death pathway genes FAS and FASL contribute to risk of lung cancer. *J. Med. Genet.* **2005**, *42*, 479–484. [CrossRef]

30. Zhang, B.; Sun, T.; Xue, L.; Han, X.; Zhang, B.; Lu, N.; Shi, Y.; Tan, W.; Zhou, Y.; Zhao, D.; et al. Functional polymorphisms in FAS and FASL contribute to increased apoptosis of tumor infiltration lymphocytes and risk of breast cancer. *Carcinogenesis* **2007**, *28*, 1067–1073. [CrossRef]

31. Yeh, I.T.; Lenci, R.E.; Qin, Y.; Buddavarapu, K.; Ligon, A.H.; Leteurtre, E.; Do Cao, C.; Cardot-Bauters, C.; Pigny, P.; Dahia, P.L. A germline mutation of the KIF1B beta gene on 1p36 in a family with neural and nonneural tumors. *Hum. Genet.* **2008**, *124*, 279–285. [CrossRef]

32. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **2014**, *505*, 302–308. [CrossRef] [PubMed]

33. Chen, Z.; Wang, S.H.; Zhou, J.; Wu, T.; Li, J. Contribution of PGR genetic polymorphisms to the pathogenesis of endometrial cancer: A meta-analysis. *J. Cancer Res. Ther.* **2015**, *11*, 810–817. [PubMed]

34. Pritchard, A.L.; Johansson, P.A.; Nathan, V.; Howlie, M.; Symmons, J.; Palmer, J.M.; Hayward, N.K. Germline mutations in candidate predisposition genes in individuals with cutaneous melanoma and at least two independent additional primary cancers. *PLoS ONE* **2018**, *13*, e0194098. [CrossRef] [PubMed]

35. Croft, D.; O'Kelly, G.; Wu, G.; Haw, R.; Gillespie, M.; Matthews, L.; Caudy, M.; Garapati, P.; Gopinath, G.; Jassal, B.; et al. Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.* **2011**, *39*, D691–D697. [CrossRef] [PubMed]

36. King, R.J.; Yu, F.; Singh, P.K. Genomic alterations in mucins across cancers. *Oncotarget* **2017**, *8*, 67152–67168. [CrossRef] [PubMed]

37. Fuentes Fajardo, K.V.; Adams, D.; NISC Comparative Sequencing Program; Mason, C.E.; Sincan, M.; Tifft, C.; Toro, C.; Boerkoel, C.F.; Gahl, W.; Markello, T. Detecting false-positive signals in exome sequencing. *Hum. Mutat.* **2012**, *33*, 609–613. [CrossRef] [PubMed]

38. Lawrence, M.S.; Stojanov, P.; Polak, P.; Kryukov, G.V.; Cibulskis, K.; Sivachenko, A.; Carter, S.L.; Stewart, C.; Mermel, C.H.; Roberts, S.A.; et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **2013**, *499*, 214–218. [CrossRef]

39. Lefebvre, C.; Bachelot, T.; Filleron, T.; Pedrero, M.; Campone, M.; Soria, J.C.; Massard, C.; Lévy, C.; Arnedos, M.; Lacroix-Triki, M.; et al. Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med.* **2016**, *13*, e1002201. [CrossRef]

40. Rosell, R.; Moran, T.; Queralt, C.; Porta, R.; Cardenal, F.; Camps, C.; Majem, M.; Lopez-Vivanco, G.; Isla, D.; Provencio, M.; et al. Screening for epidermal growth factor receptor mutations in lung cancer. *N. Engl. J. Med.* **2009**, *361*, 958–967. [CrossRef]

41. Pao, W.; Miller, V.; Zakowski, M.; Doherty, J.; Politi, K.; Sarkaria, I.; Singh, B.; Heelan, R.; Rusch, V.; Fulton, L.; et al. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13306–13311. [CrossRef]

42. Bonneville, R.; Krook, M.A.; Kautto, E.A.; Miya, J.; Wing, M.R.; Chen, H.Z.; Reeser, J.W.; Yu, L.; Roychowdhury, S. Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis. Oncol.* **2017**, *1*, 1–15. [CrossRef] [PubMed]

43. Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **2019**, *534*, 47–54; Erratum in **2019**, *566*, E1. [CrossRef] [PubMed]

44. Li, G.H.; Arora, P.D.; Chen, Y.; McCulloch, C.A.; Liu, P. Multifunctional roles of gelsolin in health and diseases. *Med. Res. Rev.* **2012**, *32*, 999–1025. [CrossRef] [PubMed]

45. Tanaka, H.; Shirkoohi, R.; Nakagawa, K.; Qiao, H.; Fujita, H.; Okada, F.; Hamada, J.; Kuzumaki, S.; Takimoto, M.; Kuzumaki, N. siRNA gelsolin knockdown induces epithelial-mesenchymal transition with a cadherin switch in human mammary epithelial cells. *Int. J. Cancer* **2006**, *118*, 1680–1691. [CrossRef] [PubMed]

46. Baig, R.M.; Mahjabeen, I.; Sabir, M.; Masood, N.; Ali, K.; Malik, F.A.; Kayani, M.A. Mutational spectrum of Gelsolin and its down regulation is associated with breast cancer. *Dis. Markers* **2013**, *34*, 71–80. [CrossRef]

47. Wang, P.W.; Abedini, M.R.; Yang, L.X.; Ding, A.A.; Figeys, D.; Chang, J.Y.; Tsang, B.K.; Shieh, D.B. Gelsolin regulates cisplatin sensitivity in human head-and-neck cancer. *Int. J. Cancer* **2014**, *135*, 2760–2769. [CrossRef]

48. Zhao, R.S.; Wang, W.; Li, J.P.; Liu, C.M.; He, L. Gelsolin Promotes Radioresistance in Non-Small Cell Lung Cancer Cells Through Activation of Phosphoinositide 3-Kinase/Akt Signaling. *Technol Cancer Res. Treat.* **2017**, *16*, 512–518. [CrossRef] [PubMed]

49. Li, M.; Cui, F.; Cheng, Y.; Han, L.; Wang, J.; Sun, D.; Liu, Y.L.; Zhou, P.K.; Min, R. Gelsolin: Role of a functional protein in mitigating radiation injury. *Cell Biochem. Biophys.* **2015**, *71*, 389–396. [CrossRef]

50. Yang, X.D.; Zhao, S.F.; Zhang, Q.; Li, W.; Wang, Y.X.; Hong, X.W.; Hu, Q.G. Gelsolin rs1078305 and rs10818524 polymorphisms were associated with risk of oral squamous cell carcinoma in a Chinese Han population. *Biomarkers* **2016**, *21*, 267–271. [CrossRef]

51. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A.; Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **2016**, *375*, 1109–1112. [CrossRef] [PubMed]

52. Stitziel, N.O.; Kiezun, A.; Sunyaev, S. Computational and statistical approaches to analyzing variants identified by exome sequencing. *Genome Biol.* **2011**, *12*, 227. [CrossRef]

53. Neale, B.M.; Rivas, M.A.; Voight, B.F.; Altshuler, D.; Devlin, B.; Orho-Melander, M.; Kathiresan, S.; Purcell, S.M.; Roeder, K.; Daly, M.J. Testing for an unusual distribution of rare variants. *PLoS Genet.* **2011**, *7*, e1001322. [CrossRef]

54. Dang, N.; Meng, X.; Song, H. Nicotinic acetylcholine receptors and cancer. *Biomed. Rep.* **2016**, *4*, 515–518. [CrossRef]

55. Zhao, Y. The Oncogenic Functions of Nicotinic Acetylcholine Receptors. *J. Oncol.* **2016**, *2016*, 9650481. [CrossRef]

56. Reisz, J.A.; Bansal, N.; Qian, J.; Zhao, W.; Furdui, C.M. Effects of ionizing radiation on biological molecules mechanisms of damage and emerging methods of detection. *Antioxid. Redox Signal.* **2014**, *21*, 260–292. [CrossRef]

57. Bonfiglio, S.; Vanni, I.; Rossella, V.; Truini, A.; Lazarevic, D.; Dal Bello, M.G.; Alama, A.; Mora, M.; Rijavec, E.; Genova, C.; et al. Performance comparison of two commercial human whole-exome capture systems on formalin-fixed paraffin-embedded lung adenocarcinoma samples. *BMC Cancer* **2016**, *16*, 692. [CrossRef]

58. Madsen, B.E.; Browning, S.R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **2009**, *5*, e1000384. [CrossRef]

59. Zauber, P.; Marotta, S.; Sabbath-Solitare, M. Copy number of the Adenomatous Polyposis Coli gene is not always neutral in sporadic colorectal cancers with loss of heterozygosity for the gene. *BMC Cancer* **2016**, *16*, 213. [CrossRef]

60. Vanni, I.; Coco, S.; Bonfiglio, S.; Cittaro, D.; Genova, C.; Biello, F.; Mora, M.; Rossella, V.; Dal Bello, M.G.; Truini, A.; et al. Whole exome sequencing of independent lung adenocarcinoma, lung squamous cell carcinoma, and malignant peritoneal mesothelioma: A case report. *Medicine (Baltim.)* **2016**, *95*, e5447. [CrossRef]

61. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *1*, 589–595. [CrossRef]

62. Chiang, C.; Layer, R.M.; Faust, G.G.; Lindberg, M.R.; Rose, D.B.; Garrison, E.P.; Marth, G.T.; Quinlan, A.R.; Hall, I.M. SpeedSeq: Ultra-fast personal genome analysis and interpretation. *Nat. Methods* **2015**, *12*, 966–968. [CrossRef]

63. Cingolani, P.; Platts, A.; Wang, L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **2012**, *6*, 80–92. [CrossRef]

64. Forbes, S.A.; Beare, D.; Boutselakis, H.; Bamford, S.; Bindal, N.; Tate, J.; Cole, C.G.; Ward, S.; Dawson, E.; Ponting, L.; et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **2017**, *4*, D777–D783. [CrossRef]

65. Lek, M.; Karczewski, K.J.; Minikel, E.V.; Samocha, K.E.; Banks, E.; Fennell, T.; O'Donnell-Luria, A.H.; Ware, J.S.; Hill, A.J.; Cummings, B.B.; et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **2016**, *18*, 285–291. [CrossRef]

66. Liu, X.; Jian, X.; Boerwinkle, E. dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* **2011**, *32*, 894–899. [CrossRef]

67. Rubio-Perez, C.; Tamborero, D.; Schroeder, M.P.; Antolín, A.A.; Deu-Pons, J.; Perez-Llamas, C.; Mestres, J.; Gonzalez-Perez, A.; Lopez-Bigas, N. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals novel targeting opportunities. *Cancer Cell* **2015**, *9*, 382–396. [CrossRef]

68. Futreal, P.A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M.R. A census of human cancer genes. *Nat. Rev. Cancer* **2004**, *4*, 177–183. [CrossRef]

69. Cheng, W.C.; Chung, I.F.; Chen, C.Y.; Sun, H.J.; Fen, J.J.; Tang, W.C.; Chang, T.Y.; Wong, T.T.; Wang, H.W. DriverDB: An exome sequencing database for cancer driver gene identification. *Nucleic Acids Res.* **2014**, *42*, D1048–D1054. [CrossRef]

70. Kautto, E.A.; Bonneville, R.; Miya, J.; Yu, L.; Krook, M.A.; Reeser, J.W.; Roychowdhury, S. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget* **2017**, *31*, 7452–7463. [CrossRef]

71. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **2013**, *14*, 128. [CrossRef]

72. Talevich, E.; Shain, A.H.; Botton, T.; Bastian, B.C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput. Biol.* **2016**, *12*, e1004873. [CrossRef]