Data in Brief 34 (2021) 106646

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Genotyping data of routinely processed matched primary/metastatic tumor samples

Vassiliki Kotoula[a,b,*], Kyriakos Chatzopoulos[a,b,1,*], Kyriaki Papadopoulou[a], Eleni Giannoulatou[c,d], Georgia-Angeliki Koliou[e], Vasilios Karavasilis[f], Elissavet Pazarli[g], Stavroula Pervana[g], Georgia Kafiri[h], Georgios Tsoulfas[i], Sofia Chrisafi[a], Helen Sgouramali[a], Pavlos Papakostas[j], Dimitrios Pectasides[j], Prodromos Hytiroglou[b], George Pentheroudakis[k,l], George Fountzilas[a,m,n]

[a] *Laboratory of Molecular Oncology, Hellenic Foundation for Cancer Research/Aristotle University of Thessaloniki, Thessaloniki, Greece*
[b] *Department of Pathology, School of Health Sciences, Faculty of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[c] *Bioinformatics and Systems Medicine Laboratory, Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia*
[d] *The University of New South Wales, Kensington, NSW, Australia*
[e] *Section of Biostatistics, Hellenic Cooperative Oncology Group, Data Office, Athens, Greece*
[f] *Department of Medical Oncology, Papageorgiou Hospital, Aristotle University of Thessaloniki, School of Health Sciences, Faculty of Medicine, Thessaloniki, Greece*
[g] *Department of Pathology, Papageorgiou Hospital, Aristotle University of Thessaloniki, School of Health Sciences, Faculty of Medicine, Thessaloniki, Greece*
[h] *Department of Pathology, Hippokration Hospital, Athens, Greece*
[i] *Department of Surgery, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[j] *Oncology Section, Second Department of Internal Medicine, Hippokration Hospital, Athens, Greece*
[k] *Department of Medical Oncology, Medical School, University of Ioannina, Ioannina, Greece*
[l] *Society for Study of Clonal Heterogeneity of Neoplasia (EMEKEN), Ioannina, Greece*
[m] *Aristotle University of Thessaloniki, Thessaloniki, Greece*
[n] *German Oncology Center, Limassol, Cyprus*

DOI of original article: 10.1016/j.humpath.2020.10.009

* Corresponding authors at: Laboratory of Molecular Oncology, Hellenic Foundation for Cancer Research/Aristotle University of Thessaloniki, Thessaloniki, Greece.

E-mail addresses: vkotoula@auth.gr (V. Kotoula), Chatzopoulos.Kyriakos@mayo.edu (K. Chatzopoulos).

[1] Present address: Department of Laboratory Medicine and Pathology, Mayo Clinic, 55905, Rochester, MN, United States of America.

https://doi.org/10.1016/j.dib.2020.106646

ARTICLE INFO

Article history:
Received 4 November 2020
Revised 7 December 2020
Accepted 8 December 2020
Available online 11 December 2020

Keywords:
Next generation sequencing
Formalin fixed paraffin embedded tissue
Tumor cell content
Ampliseq panel
Targeted NGS
Matched sample genotypes
Colorectal cancer

ABSTRACT

Genotypic and phenotypic comparisons of tumors in multiple tissue samples from the same patient are important for understanding disease evolution and treatment possibilities. Panel NGS genotyping is currently widely used in this context, whereby NGS variant filtering and final evaluation constitute the basis for meaningful comparisons. Here, we present the genotype data used for genotype / phenotype comparisons between matched primary / metastatic colorectal tumors in the work by Chatzopoulos et al (doi: 10.1016/j.humpath.2020.10.009), as well as the process followed for obtaining these data. We describe key issues while processing routinely formalin-fixed paraffin-embedded (FFPE) tumors for genotyping, NGS application (Ion Torrent), a stringent variant filtering algorithm for genotype analyses in FFPE tissues and particularly in matched tumor samples, and provide the respective datasets. Apart from research, tumor NGS genotyping is currently applied for clinical diagnostic purposes in Oncology. The datasets and method description provided herein (a) are important for comprehending the peculiarities of FFPE tumor genotyping, which is still mostly based on principles of germline DNA genotyping; (b) can be used in pooled analyses, e.g., of primary / metastatic tumors for the investigation of tumor evolution.

Specifications Table

Subject	Health and medical sciences Pathology and Medical Technology
Specific subject area	Molecular Pathology; Pathology; cancer genotyping; tumor evolution
Type of data	Targeted next generation sequencing; custom panel; multiple matched samples; formalin-fixed paraffin-embedded (FFPE) tissue; germline DNA
How data were acquired	Ampliseq panel sequencing (Ion Torrent; Proton sequencer); tissue sectioning and staining; light microscope; macrodissection
Data format	Filtered and analysed NGS data; raw microscopy data; processed histology report data
Parameters for data collection	FFPE tissue blocks; histology reports; matched primary and metastatic tumors; FFPE and germline DNA
Description of data collection	Microscopic evaluation of hematoxylin and eosin stained FFPE tissue sections for determination of histopathologic parameters and tumor cell content (TCC%). Macrodissection, DNA extraction and measurement. FFPE and germline NGS genotyping (Ion Torrent), library construction, library evaluation, sequencing on an Ion Proton sequencer, variant calling, multiple steps for variant filtering in matched tumor samples, and variant annotation.
Data source location	Laboratory of Molecular Oncology (Hellenic Foundation for Cancer Research / Aristotle University of Thessaloniki), Thessaloniki, Greece; Hellenic Cooperative Oncology Group (HeCOG), Athens, Greece

(*continued on next page*)

Data accessibility	Matched genotype datasets suitable for pooled analyses are hosted with the article. Raw data can be publicly accessed through HeCOG server at the link below: https://files.hecog.gr/DATA_DIB-D-20-02216.7z
Related research article	K. Chatzopoulos, V. Kotoula, G.A. Koliou, E. Giannoulatou, K. Papadopoulou, V. Karavasilis et al. Genotype – phenotype associations in colorectal adenocarcinomas and their matched metastases. Hum Pathol. 2020 Nov 5:S0046-8177(20)30215-X. doi: 10.1016/j.humpath.2020.10.009. Online ahead of print. PMID: 33161028

Value of the Data

- FFPE tumor processing and NGS variant filtering are important for understanding the results of the currently widely applied tumor NGS genotyping on clinical material, i.e., FFPE tumors that are routinely processed in pathology laboratories for DNA studies.
- Matched genotype data are important for comprehending intra-patient tumor heterogeneity and evolution from primary site to metastasis.
- FFPE tissue and panel NGS variant processing are of interest for all scientific and medical disciplines involved with FFPE genotyping, particularly if matched samples are concerned.
- The provided matched genotype data can be reused for pooled analyses in colorectal cancer; in studies concerning tumor evolution; in pooled bioinformatics studies for deciphering additional characteristics of variant DNA in tumor tissues that does not only concern “clinically actionable” or “driver” variants.
- The description of FFPE tissue processing and the matched genotype data can be used for educational purposes in Molecular Pathology.

1. Data Description

“Genotype – phenotype associations in colorectal adenocarcinomas and their matched metastases” is a study on geno/phenotyping in patient samples, matched primary and metastatic colorectal carcinomas [1]. FFPE tumor sample characteristics with respect to sample origin (site of primary and metastatic tumor), classification of metastasis with respect to elapsed time (synchronous or metachronous metastasis) based on histology report dates, as well as tumor cell content (TCC%), DNA content, and NGS data eligibility are included in the Supplementary File S1 dataset. NGS was performed with a custom panel, which is provided as Supplementary File S2. The genotyping dataset with all eligible NGS variants per case and relevant annotations, detailed data per primary tumor, matched metastatic tumor, and germline DNA, where available, is provided as Supplementary File S3. This dataset includes technical variant and genotype characteristics for matched samples; Ion Reporter v.5.0 variants were matched in samples from the same patient based on best coverage for identical chromosomal positions. Details on FFPE tumor section processing for molecular studies are also provided.

2. Experimental Design, Materials and Methods

The data presented here refer to a study on geno/phenotyping in matched primary and metastatic colorectal carcinomas [1] that primarily describes differences in synchronous vs. metachronous metastases. FFPE tumor blocks were retrieved from the biologic material repository of the Hellenic Cooperative Oncology Group. FFPE tissue processing, histology review, DNA

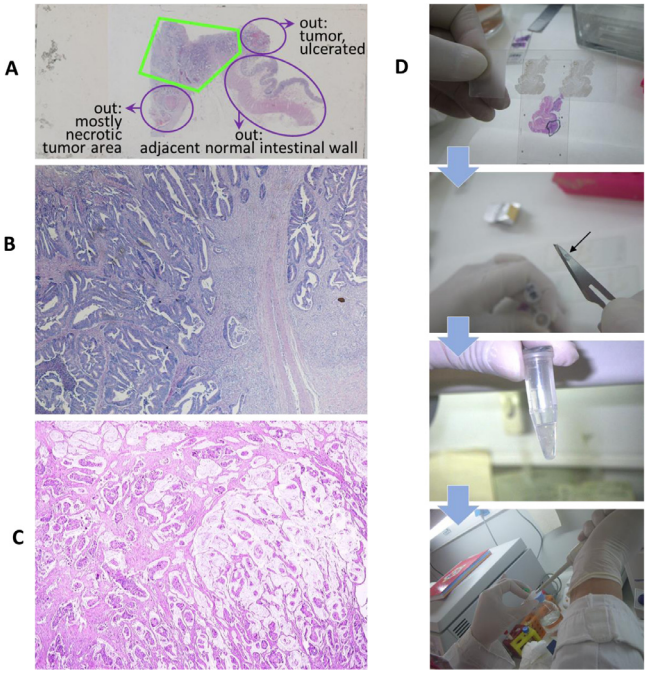

Fig. 1. FFPE tumor processing for molecular studies. A – C: Histologic evaluation of hematoxylin & eosin (H&E) sections with examples of common pitfalls in TCC% evaluation. In **A**, the tumor area of interest is marked on the H&E section avoiding necrotic, hemorrhagic and normal tissue elements. In **B**, higher magnification of the section in A is shown; malignant cells occupy 70% of this tumor area. However, the surrounding dense lymphocytic infiltrates by far outnumber malignant cells. Hence, the tumor DNA percentage in the extracted sample will be only 30%. In **C**, malignant cells appear to occupy 30% of this tumor area (the tumor produces a lot of mucin but also has desmoplastic stroma). However, the malignant cells outnumber non-malignant cells in this case and the tumor DNA percentage in this sample will be 80%. **D:** The process of microdissection for enrichment of molecular templates in tumor DNA. Unstained sections (5 – 10 micron thick) are oriented on simple sterilized glass slides to the respective H&E section that has been marked for the tumor area of interest → By using a thin scalpel, the tissue corresponding to the marked area is removed from the unstained section (arrow) directly after deparaffinization and partial rehydration → All macrodissected tissue fragments are placed in an Eppendorf tube with lysis buffer → Proteinase K is added to the lysis buffer and the tissue fragments are allowed to dissolve completely.

extraction, NGS application and evaluation were performed in the Laboratory of Molecular Oncology (MOL; Hellenic Foundation for Cancer Research / Aristotle University of Thessaloniki). Variant quality control, filtering and matching were performed at the Bioinformatics and Systems Medicine Laboratory, Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia.

2.1. FFPE tumor processing for genotyping

Hematoxylin & eosin (H&E) sections from an initial number of 235 FFPE blocks were examined for tumor presence and eligibility for the assessment of histological parameters [1]. Out of these, 170 FFPE blocks corresponding to 85 matched primary / metastatic tumor pairs (and to an equal number of patients) were considered eligible for both phenotypic evaluation and genotyping. These tumor blocks were processed for DNA extraction, starting with histological evaluation of the interrogated tissues. H&E sections from these blocks were marked for the most solid tumor areas by avoiding to the extent possible hemorrhagic infiltrates, large necrotic areas, and normal tissue elements (Fig. 1A). These marked areas were evaluated for tumor DNA content, by

providing the ratio of malignant cell nuclei vs. all nuclei in the area. Tumor DNA content is otherwise known as tumor cell content or tumor cell percentage (TCC%). Reasonably, TCC% should reflect the percentage of the molecular template that derives from malignant cells in a given sample. TCC% should for no reason represent the percentage of the area occupied by the tumor in a tissue section. The latter is a misconception of what the marker stands for, unfortunately not uncommon among pathologists who assess TCC%. The TCC% values obtained either way may greatly differ, as exemplified in Fig. 1B and 1C. Because of such differences in TCC% assessment, the marker has been considered as unreliable over the years, difficult to communicate to biologists and clinicians [2] and is not obligatorily included in the recommendations for genotyping reports (e.g., by the Spanish pathologists and geneticists [3]). However, TCC% is important not only for determining the eligibility of tumor samples for molecular studies relevant to method sensitivity, but also for interpreting profiling and genotyping results [2,4,5]. For DNA studies in particular, where malignant cell ploidy is not measured at baseline, TCC% may serve as a rough marker of malignant cell DNA content in a given sample and has been introduced in bioinformatics algorithms for the assessment of gene amplification (e.g., Oncomine assays by Thermo Fisher Scientific, Paisley, UK). The marked areas on H&E sections were subsequently manually macrodissected for DNA extraction (Fig. 1D), in order to increase representation of malignant cell DNA content in the molecular template and genotyping accuracy [6]. Macrodissection was considered necessary in the majority of sections from surgical specimens and was also applied in biopsy sections, where possible. Macrodissection was performed on unstained deparaffinized sections, whereby tissue fragments were directly processed for DNA extraction. DNA was extracted with the QIAamp® DNA mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Next to tumor samples, peripheral blood samples were available at study initiation for 65 patients. Germline DNA was extracted with a classic desalting method [7]. Histology data of the tumor samples and technical data of all examined samples are provided in the Supplementary File S1 dataset.

2.2. NGS panel

For colorectal tumor genotyping [1] we used the IAD47763_31 Ampliseq panel (Applied Biosystems / Thermo Fisher Scientific). The panel was designed in 2013 to interrogate frequently mutated genes in colorectal cancer, as reported in the literature [8,9,10]. The panel comprised 424 amplicons covering a total genomic sequence of 45,689 nucleotides in 57 genes (52 coding, 5 microRNAS); coordinates for the regions of interest were spotted and exported from the UCSC Genome Browser with GRCh37 (hg19) as the genomic reference (Supplementary File S2). The returned Ampliseq primers and amplicons were checked for specificity with BLAST; the finally accepted panel design was 100% specific. The panel primers were provided in two pools by the manufacturer. Among the 424 amplicons, 24 (5.6%) constantly failed in all runs, i.e., never produced end-to-end reads, mostly due to very high GC content.

2.3. NGS application

DNA concentration for all samples was measured in a Qubit fluorometer by using the dsDNA HS Assay (Thermo Fisher Scientific, Paisley, UK). In addition, amplifiability of FFPE DNA was checked with the BIOMED2 multiplex PCR control assay [6]. FFPE DNA samples were processed for NGS if DNA concentration was >2ng/ul and if DNA quality allowed for PCR products of at least 200bp to be amplified, given that the Ampliseq amplicons were maximally 191bp. Twenty ng per DNA sample (10ng per primer pool) were used for library construction that was carried out with the Ampliseq Library Kit and Ion Xpress barcodes, and submitted for multiplex PCR, according to the manufacturer's instructions (Life Technologies / Ion Torrent, Carlsbad, CA). Library concentration was again normalized to 15ng/ml upon a second measurement with

Qubit and processed on a One-Touch-2 instrument followed by enrichment on a One-Touch-ES station. Initial sample and library DNA concentration data are provided in the Supplementary File S1 dataset. Templating was performed using the Ion PI™ template OT2-200 Kit, and sequencing was performed on a Proton™ sequencer using PI chips (Ion PI™ Sequencing 200 Kit v2) with multiplexing up to 36 samples. FFPE and germline DNA libraries were sequenced in separate runs.

2.4. NGS variant processing

Base calling was performed on Torrent Server using Torrent Suite v.5.0.2 under default settings. Variant calling was carried out using the embedded Variant Caller high stringency plug-in for somatic variant detection to evaluate sample stability. Reads were then uploaded by the coverage analysis plug-in to the Ion Reporter v.5.0 pipeline for variant annotation under high stringency conditions (for SNVs, min coverage for each strand: 5%; min variant score: 20; min coverage: 100; strand bias: 0.09), since this pipeline has been considered as the most reliable one for Ion Torrent NGS data [11,12]. Upon removing overlapping reads, 9844 variants were returned automatically by the system. Taking into account the intrinsically damaged status of FFPE DNA templates, additional quality filters were applied for avoiding unreliably read variants. Thus, removed were: variants covered by amplicons with read depth <100 (remaining: 6006 variants); SNVs and indels with position coverage <100 and <200, respectively (remaining: 5588 variants); variants with variant coverage <40 (remaining: 3036 variants); variants with p-value (a system quality metric) >0.0001 (remaining: 2741 variants); stretches with Gs and Cs (occurrences of more than 2 Gs or Cs in either variant or reference; remaining: 2709 variants); off target & unannotated variants (no gene / chromosomal band assigned; remaining: 2653 variants). In total, the rate of variants eligible for analysis out of all variants automatically returned by the system was 26.9%. These variants corresponded to 85 patient cases (multiple samples per case).

2.5. Eligible samples and matched pairs

Based on the above described panel efficiency, informative samples should have at least 75% of the amplicons read end-to-end; mean depth of at least 300; uniformity of reads at least 50%; and, at least 10 variants. Samples eligible for analysis based on these criteria were: 158 out of 170 (92.9%) processed FFPE and 59 out of 65 (90.8%) processed germline DNA. However, 10 patients were subsequently removed from the geno/phenotyping study [1] due to incomplete clinical data or sample mismatches. Hence, the final number of patients was 75; for these, 141 FFPE tumor, 49 germline DNA genotypes and 2380 variants were analyzed. The technical characteristics of these samples are shown in Table 1 and detailed per sample in the Supplementary File S1 dataset.

2.6. Variant annotation

Variants with minor allele frequency (MAF) <0.1% (based on dbSNP, ExAC and 5000Exomes) that were amino acid (coding non-synonymous) or splice site changing, were called mutations (n=438). The incidence of the 45 mutated genes in all examined cases is shown in Fig. 2A. Mutations covered the entire spectrum of functional annotations by Ion Reporter and ANNOVAR [13]; scores and predictions suitable for somatic variants, i.e., fathmm/fathmm-mkl scores, ClinVar and COSMIC, were used, while those commonly used in population genetics (e.g., sift, grantham, polyphen, etc) were omitted. As shown in Supplementary File S3, mutations were called pathogenic if registered as pathogenic/likely pathogenic in ClinVar and/or COSMIC, and/or if predicted to be deleterious by both fathmm scores [14]. With this approach 224 pathogenic

Table 1
FFPE tumor DNA content (TCC%) and NGS quality metrics for informative samples

		Tumor DNA %		Mean Depth			Uniformity %			Number of Variants		
N Samples		Primary 69*	Meta 68*	Primary 71	Meta 70	Blood 49	Primary 71	Meta 70	Blood 49	Primary 71	Meta 70	Blood 49
Quantiles												
100.00%	maximum	90	90	14391.0	10669.0	5273.0	91.6	91.8	88.4	298.0	340.0	40.0
99.50%		90	90	14391.0	10669.0	5273.0	91.6	91.8	88.4	298.0	340.0	40.0
97.50%		90	90	13235.8	9194.2	5085.0	91.5	91.3	88.4	211.6	311.3	38.8
90.00%		85	80	4218.2	4169.0	1901.0	88.9	89.5	87.7	73.6	165.8	33.0
75.00%	quartile	75	70	3130.0	3432.3	1081.0	87.4	87.9	86.6	40.0	39.0	28.5
50.00%	median	60	50	1931.0	1570.0	857.9	85.3	85.1	84.5	35.0	32.5	25.0
25.00%	quartile	40	35.25	513.4	694.5	731.7	81.6	78.9	80.2	31.0	30.0	19.0
10.00%		35	24	392.7	502.5	565.7	75.8	73.0	75.1	28.0	27.0	16.0
2.50%		17.5	5	362.0	399.2	509.0	45.8	56.9	70.6	15.8	18.6	12.3
0.50%		10	2.5	358.7	366.8	509.0	41.5	55.2	70.5	15.0	17.0	12.0
0.00%	minimum	10	2.5	358.7	366.8	509.0	41.5	55.2	70.5	15.0	17.0	12.0
Mean		58.2	51.6	2381.3	2281.0	1136.1	82.7	82.2	83.0	48.7	52.7	24.1
Std Dev		19.0	22.1	2623.5	2032.8	925.7	8.8	8.1	4.8	46.8	62.2	6.2

Abbreviations: NGS = next generation sequencing; meta = metastasis; Std Dev = standard deviation

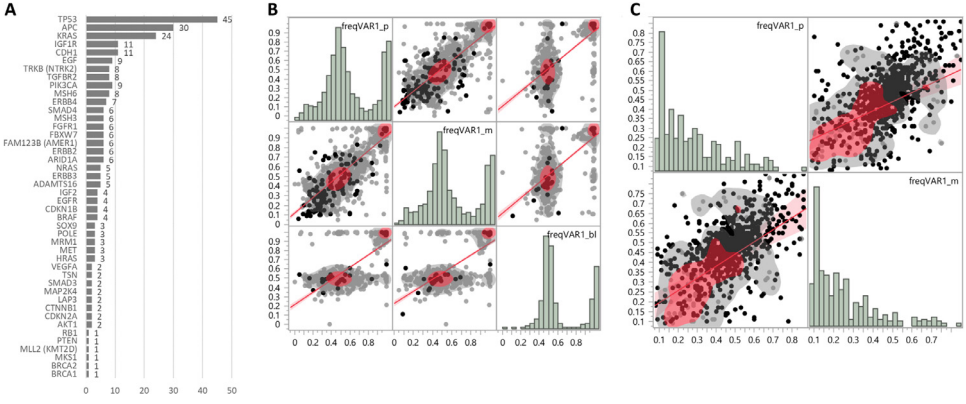

Fig. 2. Common / rare SNPs and mutations as assessed in this dataset. A. Number of cases (X axis) with tumors mutated in the respective genes. **B.** Paired comparisons of variant allele frequencies (VAFs or freqVAR1) for common and rare SNPs between primary / metastatic tumors (1081 comparisons; correlation 0.8204; 95%CI 0.80-0.84), primary tumor / blood (695 comparisons; correlation 0.7406; 95%CI 0.70-0.77), metastatic tumor / blood (686 comparisons; correlation 0.7219; 95%CI 0.68-0.76). **C.** Paired comparisons of shared mutation VAFs between primary / metastatic tumors (107 comparisons; correlation 0.5465; 95%CI 0.39-0.67).

out of 438 (51.1%) mutations were identified in tumors from 63 out of 64 (98.4%) patients with at least one mutated sample. However, (a) because pathogenicity of somatic mutations is conditional and depends on the molecular context and the underlying mechanisms driving the mutational process [15] and (b) because the limited number of matched primary / metastatic genotypes [1] did not allow for profiled pathogenic mutation evaluation, we did not address pathogenic mutations as a separate category in this study. We did however assess variant load in paired tumors, after adjusting for TCC% in each sample. Clonal mutations were considered for variant allele frequency (VAF) >25% [16] taking into account TCC%. As also shown in Supplementary File S3, common SNPs, i.e., variants with MAF >0.1%, had always VAFs 40 – 55% or close to 100% in blood samples; these were used for the validation of matched pairs. Rare SNPs (MAF 0.1% - 0.01%) were also annotated; 56 rare SNPs were identified among cases, 15 of

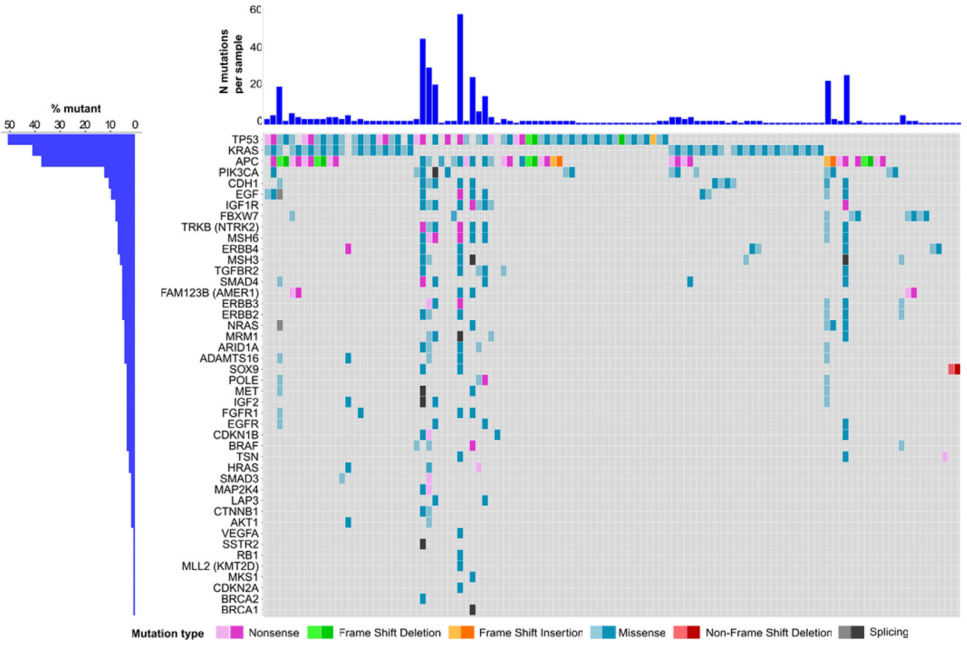

Fig. 3. Distribution of mutations in primary and matched metastatic colorectal carcinomas. The map shows shared and private mutations in matched primary / metastatic tumor samples, along with mutation types (color key), number of mutations per sample (top) and prevalence of mutated genes among tumors. Tumor pairs are shown in columns; the lighter color shade stands for the primary, the darker shade for the metastatic tumor. The map was created with the free R software.

which (26.8%) were pathogenic in ClinVar. The data in Supplementary File S3 are of interest for further investigation in pooled cohorts.

2.7. Genotypes in matched primary / metastatic tumor samples

For variant comparison between matched tumor pairs, positions inadequately read in either side were excluded. Nevertheless, few inadequately read positions that did exhibit the matched variant were considered for matched comparisons, in order to cope with the high diversity of FFPE DNA quality and PCR efficiency. Matched pairs were considered informative if samples shared at least 50% of the identified common/rare SNPs per case. With this criterion, 66 primary / metastatic pairs were eligible for analysis [1]. As a test for variant performance in matched pairs, common / rare SNP VAFs were compared in matched primary / metastatic tumors, primary tumor / blood, metastatic tumor / blood showed the anticipated associations when studying FFPE DNA (Fig. 2B). In comparison, paired comparisons of mutations in primary / metastatic tumors showed diffuse patterns (Fig. 2C). The distribution of private and shared mutations in matched primary / metastatic tumors is shown in Fig. 3. These data form the basis of the geno/phenotyping comparisons reported [1].

2.8. Germline variants in tissues

Five germline variants that affected coding regions but could not be characterized as common SNPs were identified (Table 2). These were: 2 VUS's, 1 rare SNP, 1 unknown in ClinVar, and 1

Table 2
Germline variants in the examined samples

Chromosome:Position	Gene	Protein	Coding	ClinVar (Year)	Sample	TCC%	Genotype	Position Coverage	Variant Coverage	VAF
16:68849520	CDH1	p.Val475Met	c.1423G>A	VUS (2018)						
					blood		G/A	215	116	0.54
					primary	50	G/A	1443	798	0.55
					metastatic	10	G/A	1976	1,056	0.53
7:116403245	MET	p.Val854Ile	c.2560G>A	unknown (2020)*						
					blood		G/A	842	355	0.42
					primary	65	G/A	1287	420	0.33
					metastatic	75	G/A	589	71	0.12
2:212812268	ERBB4	p.Arg103His	c.308G>A	unknown (2020)						
					blood		C/T	1153	582	0.5
					primary	65	C/T	1614	635	0.39
					metastatic	0	C/T	1994	829	0.42
16:68862107	CDH1	p.Arg732Leu	c.2195G>T	VUS (2019)						
					blood		G/T	623	299	0.48
					primary	not evaluable	G/T	1986	993	0.5
					metastatic	not evaluable	G/T	797	354	0.44
2:48030612	MSH6	p.Arg1076Cys	c.3226C>T	likely patho (2018)						
					blood		C/T	876	431	0.49
					primary	65	C/T	131	68	0.52
					metastatic	75	NA	failed	failed	failed

VUS: variant of unknown significance;
* : rare SNP (5000 Exomes); patho: pathogenic; TCC%: tumor cell content; VAF: variant allele frequence (variant load)

likely pathogenic in the *MSH6* gene. The latter was heterozygous in the matched primary tumor but the position was not properly read in the metastatic tumor.

Ethics Statement

The study was approved by the Bioethics Committee of the Aristotle University of Thessaloniki (AUTH, #356/15/7/2016). Patients had signed informed consent for the use of their biological material for research purposes.

Declaration of Competing Interest

The authors declare no conflicts of interest.

Acknowledgements

The authors wish to thank Mrs. Emily Daskalaki for excellent technical assistance and Mr. Dinos Lilakos (Antisel S.A., Greece) for technical support with Ion Torrent NGS.

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2020.106646.

References

[1] K. Chatzopoulos, V. Kotoula, G.A. Koliou, E. Giannoulatou, K. Papadopoulou, V. Karavasilis, et al., Genotype – phenotype associations in colorectal adenocarcinomas and their matched metastases, Hum. Pathol. (2020) Nov 5S0046-8177(20)30215-X, doi:10.1016/j.humpath.2020.10.009.
[2] A.J. Smits, J.A. Kummer, P.C. de Bruin, M. Bol, J.G. van den Tweel, K.A. Seldenrijk, et al., The estimation of tumor cell percentage for molecular testing by pathologists is not accurate, Mod. Pathol. 27 (2014) 168–174.
[3] J. Palacios, M. de la Hoya, B. Bellosillo, I. de Juan, X. Matias-Guiu, C. Lazaro, et al., Mutational Screening of BRCA1/2 genes as a predictive factor for therapeutic response in epithelial ovarian cancer: a consensus guide from the spanish society of pathology (SEAP-IAP) and the Spanish Society of Human Genetics (AEGH), Virchows Arch 476 (2020) 195–207.
[4] B. Lhermitte, C. Egele, N. Weingertner, D. Ambrosetti, B. Dadone, V. Kubiniek, et al., Adequately defining tumor cell proportion in tissue samples for molecular testing improves interobserver reproducibility of its assessment, Virchows Arch. 470 (2017) 21–27.
[5] K. Dufraing, J.H. van Krieken, G. De Hertogh, G. Hoefler, A. Oniscu, T.P. Kuhlmann, et al., Neoplastic cell percentage estimation in tissue samples for molecular oncology: recommendations from a modified Delphi study, Histopathology 75 (2019) 312–319.
[6] V. Kotoula, E. Charalambous, B. Biesmans, A. Malousi, E. Vrettou, G. Fountzilas, et al., Targeted KRAS mutation assessment on patient tumor histologic material in real time diagnostics, PloS One 4 (2009) e7746.
[7] A.V. Stavropoulou, F. Fostira, M. Pertesi, M. Tsitlaidou, G.E. Voutsinas, O. Triantafyllidou, et al., Prevalence of BRCA1 mutations in familial and sporadic greek ovarian cancer cases, PloS One 8 (2013) e58182.
[8] Cancer Genome Atlas Network, Comprehensive molecular characterization of human colon and rectal cancer, Nature 487 (2012) 330–337.
[9] T.R. Geiger, J.Y. Song, A. Rosado, D.S. Peeper, Functional characterization of human cancer-derived TRKB mutations, PloS One 6 (2011) e16871.
[10] S. Savas, A. Hyde, S.N. Stuckless, P. Parfrey, H.B. Younghusband, R. Green, Serotonin transporter gene (SLC6A4) variations are associated with poor survival in colorectal cancer patients, PloS One 7 (2012) e38953.
[11] A. Deshpande, W. Lang, T. McDowell, S. Sivakumar, J. Zhang, J. Wang, et al., Strategies for identification of somatic variants using the Ion Torrent deep targeted sequencing platform, BMC Bioinform. 19 (2018) 5.
[12] Q. Wang, V. Kotoula, P.C. Hsu, K. Papadopoulou, J.W.K. Ho, G. Fountzilas, et al., Comparison of somatic variant detection algorithms using Ion Torrent targeted deep sequencing data, BMC Med. Genom. 12 (2019) 181.
[13] K. Wang, M. Li, H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, Nucleic Acids Res. 38 (2010) e164.

[14] V. Kotoula, S. Lakis, I. Tikas, E. Giannoulatou, G. Lazaridis, K. Papadopoulou, et al., Pathogenic BRCA1 mutations may be necessary but not sufficient for tissue genomic heterogeneity: deep sequencing data from ovarian cancer patients, Gynecol. Oncol. 152 (2019) 375–386.
[15] M. Mina, F. Raynaud, D. Tavernari, E. Battistello, S. Sungalee, S. Saghafinia, et al., Conditional selection of genomic alterations dictates cancer evolution and oncogenic dependencies, Cancer Cell 32 (2017) 155-168.e6.
[16] N. McGranahan, F. Favero, E.C. de Bruin, N.J. Birkbak, Z. Szallasi, C. Swanton, Clonal status of actionable driver events and the timing of mutational processes in cancer evolution, Sci. Transl. Med. 283 (2015) 283ra54.