Bridging the gap

implementation of whole genome sequencing in routine clinical care

Kris Samsom

Bridging the gap:

Implementation of whole genome sequencing in routine clinical care

Kris Samsom

Colofon

Bridging the gap: implementation of whole genome sequencing in routine clinical care

PhD thesis, Utrecht University, The Netherlands

ISBN/EAN: 978-94-93315-53-2
DOI:

The research described in this thesis was financially supported by ZonMw and Hartwig Medical Foundation.

Printing of this thesis was financially supported by the Netherlands Cancer Institute.

Cover & lay-out: Proefschrift all in one || www.proefschrift-aio.nl
Printed by: Proefschrift all in one || www.proefschrift-aio.nl

Bridging the gap: Implementation of whole genome sequencing in routine clinical care

Bruggen bouwen: implementatie van whole genome sequencing in de routine klinische zorg

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 3 mei 2023 des middags te 12.15 uur

door

Kristian Giulia Samsom

geboren op 15 juni 1993
te Utrecht

Promotoren:
Prof. dr. G.A. Meijer
Prof. dr. E.E. Voest

Copromotoren:
Dr. K. Monkhorst
Dr. L.J.W. Bosch

Beoordelingscommissie:
Prof. dr. R. Bernards
Prof. dr. I.H.M. Borel Rinkes
Prof. dr. P.J. van Diest
Prof. dr. W.T.A. van der Graaf
Dr. E.G. Klompenhouwer

Table of contents

CHAPTER 1

General introduction and thesis outline

Cancer development

The most common cause of cancer is acquired damage to the deoxyribonucleic acid (DNA) of a cell leading to certain processes that result in unregulated and abnormal cell growth. A minority of cancers are associated with pathogenic germline mutations (hereditary mutations). A germline mutation is defined as a mutation which occurs in reproductive cells and therefore is incorporated in every cell of the offspring. The functional capabilities acquired by cells as they transform from normalcy to malignancy have previously been described by Hanahan and Weinberg as the hallmarks of cancer (1). These hallmarks consist of (1) sustaining proliferative signaling, (2) evading growth suppressors, (3) avoiding immune destruction, (4) enabling replicative immortality, (5) tumor-promoting inflammation, (6) activating invasion and metastasis, (7) inducing or accessing vasculature, (8) genome instability and mutation, (9) resisting cell death and (10) deregulating cellular metabolism. These ten biological characteristics and their underlying genomic aberrations provide a rationale for cancer treatment.

DNA sequencing: a historical perspective

The first human genome, containing approximately 20,000 protein coding-genes, was sequenced in the late 20th century as part of the Human Genome Project (2). The initial sequencing technology was cumbersome and costly (3 billion US dollars), and the first complete sequence took ten years to complete. However, it enabled a detailed map of the location of genes on the human chromosomes to be built up, which provided insight into previously unstudied genes and became a fundamental resource for further research. The introduction of high-throughput next generation sequencing (NGS) in 2005 heralded a pivotal and transformative step in the history of DNA sequencing (3). NGS allows for the simultaneous sequencing of selected regions of the genome (targeted sequencing of 20 up to >500 genes), up to the whole exome (Whole Exome Sequencing (WES)) or the whole genome (Whole Genome Sequencing (WGS)). High-throughput sequencing enabled DNA to be sequenced within a relatively short period of time and at a fraction of the cost.

Molecular profiling for precision oncology

NGS revolutionized genomic studies of cancer and targeted NGS panels were readily adopted into clinical care thereby becoming the stepping stone of precision oncology (4). The premise of precision oncology is the use of therapeutics that are

expected to confer benefit to a subset of patients whose cancer displays specific molecular features and thereby offering the right treatment to the right patient at the right time (5). This premise is made possible by the use of molecular diagnostics which allow for the identification of molecular features which can be matched with targeted therapies. To illustrate, the discovery of epidermal growth factor receptor (*EGFR*) mutations and anaplastic lymphoma kinase (*ALK*) rearrangements in 2004 and 2007 respectively, as predictors of response to tyrosine kinase inhibitors, marked the beginning of the era of targeted therapies in lung cancer (6, 7).

Current standard of care molecular diagnostics

The ongoing development of a wide range of targeted therapies and their associated biomarkers has led to multiple diagnostic tests being performed as part of current standard of care (SOC) molecular diagnostics. These include targeted NGS panels, RNA-based NGS fusion analysis, Sanger sequencing, reverse transcription polymerase chain reaction (RT-PCR), fluorescence in situ hybridization (FISH) and immunohistochemistry. Each of these tests cover only a single or limited part of the spectrum of relevant genomic changes and are performed in a sequential tumor-type specific manner in clinical practice. The most prevalent biomarkers in certain tumor types are tested first, followed by less prevalent biomarkers if necessary. This occurs at the expense of time, available tumor tissue (which is scarce in metastatic lesions as they typically yield small biopsies) and efficient clinical decision making (8). In addition, this diagnostic approach is logistically complex and easily prone to errors.

Over the past years, the druggable genome has expanded rapidly which has led to an increase in the number and complexity of biomarkers that need to be assessed in each single patient. The current increase in number and complexity of biomarkers puts the logistics and sustainability of molecular diagnostics under constant pressure. Consequently, uptake of newly discovered biomarkers is often delayed, resulting in less than optimal access to rational treatment options and ultimately inequality of clinical care (9, 10). In addition, targeted drugs are increasingly used for tumor-agnostic indications (11-19). As biomarkers can now be used for selecting targeted therapies regardless of tumor type, a tumor-type specific molecular diagnostic approach is a less suitable strategy as it will limit patients in their possible treatment opportunities. Moreover, there are increasingly smaller populations of patients with rare biomarkers which are currently not covered by SOC molecular diagnostics.

Whole genome sequencing

WGS is a comprehensive method of analyzing the entire tumor DNA, which requires fresh or fresh frozen tumor material and a patient-matched blood sample. WGS is able to detect mutations, copy number variants (amplifications/losses), structural variants, oncogenic fusions, mutational signatures (i.e. tumor mutational burden (TMB), homologous recombination deficiency (HRD), microsatellite instability (MSI)) and viral insertions in tumor DNA. In addition, (pathogenic) germline variants can be identified in the accompanying blood reference sample. The applications of WGS therefore range from complete genomic characterization of the tumor and identification of all treatment relevant biomarkers to virus detection and pharmacogenomics in one single test.

Whole genome sequencing in rare cancers

Rare cancers are defined according to the Surveillance of Rare Cancers in Europe (RARECARE) project as cancers with fewer than 6 cases per 100,000 people each year (20). In the Netherlands, rare cancers account for 18% of all cancer diagnoses and overall 5-year survival for rare cancers is worse than for common cancers (52.0% vs. 68.7%) (21). This survival gap can partly be explained by an incomplete understanding of the molecular pathogenesis of these cancers and the lack of effective therapeutic options. Recently, it has been demonstrated that comprehensive genomic analysis in this patient population provides diagnostic and therapeutic benefits in a majority of patients (22).

Cancer of unknown primary

Cancer of Unknown Primary (CUP) is a heterogeneous group of cancers defined by the presence of metastatic disease without an identified primary tumor site despite modern imaging and extensive pathology work-up (23). CUP accounts for approximately 3-5% of all metastatic cancers (24). Due to the long diagnostic pathway and subsequent clinical deterioration, more than 50% of patients with CUP are currently not eligible for cancer treatment (25). Since most cancer treatments are based on primary tumor site, patients with CUP have limited number of therapy options, generally only consisting of non-selective chemotherapy regimens. Consequently, patients with CUP have a poor prognosis with a median overall survival of 3 months after diagnosis (24). The use of comprehensive genomic analysis in CUP is twofold; identification of primary tumor type with subsequent SOC therapy options and detection of actionable genomic events for targeted therapy.

Small intestinal neuroendocrine tumors

Neuroendocrine tumors of the small intestine (SI-NETs) are rare neoplasms arising from neuroendocrine cells in the bowel. The reported incidence of SI-NETs has been increasing in the last decades with an annual incidence of 1.05 per 100,000 (26). Up to 73% of patients have metastases at diagnosis, predominantly of the liver (27-30). In patients with metastatic disease, carcinoid syndrome which is characterized by diarrhea, episodic flushing, bronchospasm and valvular heart disease, is common (31). Although advancements have been made to alleviate symptoms of carcinoid syndrome and prolong the survival of patients with SI-NETs (5-year survival rate of 75% (32-34)), therapeutic options for patients with metastatic SI-NETs remain limited to i.e. surgery, liver-directed therapies, somatostatin receptor analogues and peptide receptor radionuclide therapy. At present, the molecular pathogenesis of SI-NETs is poorly elucidated and biomarkers for targeted therapy have not yet been identified. In order to move towards precision oncology, both the unravelling of the molecular landscape and identification of actionable biomarkers can be achieved by comprehensive genomic analysis

Outline of this thesis

Whilst the potential of WGS as a comprehensive diagnostic tool has been demonstrated in prior studies (35-38), its implementation into routine clinical care has been hampered by complexity of laboratory and analytical workflows to deliver results within clinically relevant timeframes and concerns about clinical validity and utility (38). This thesis therefore investigated the implementation and application of WGS in routine clinical care.

The first part of this thesis focusses on the implementation of WGS in routine clinical care. **Chapter 2** describes the study protocol of the *Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient* (WIDE) study, in which 1200 consecutive patients with (a suspicion) of stage IV disease of solid tumors from a single comprehensive cancer center were included without pre-selection on tumor type. WGS was performed prospectively in parallel with and independently of SOC diagnostics. Primary endpoints included feasibility and clinical validity and secondary endpoints clinical value of WGS in routine clinical care. **Chapter 3** describes a protocol based on the lessons learned from the WIDE study with essential steps for implementing WGS in routine pathology practice. Until now one of the biggest hurdle for clinical implementation of WGS has been the need for fresh frozen samples which requires a transition of laboratory workflow from mostly formalin

fixed paraffin embedded (FFPE) to a fresh frozen tissue orientated. This chapter highlights the optimizations needed in i.e. patient information, sample collection, laboratory and analytical logistics and integration into clinical decision-making, to guide uptake of WGS in hospitals worldwide. **Chapter 4** presents the results from the WIDE study on the feasibility, clinical validity and value of WGS in routine clinical care for patients with metastatic cancer. **Chapter 5** investigates the clinical value of a WGS-based 'cancer of unknown primary prediction algorithm' (CUPPA) in the routine diagnostic work-up of patients with CUP.

The second part of this thesis focusses on the use of (whole) genome sequencing in rare tumor types or to detect rare genetic events. **Chapter 6** provides an overview of the current literature regarding prognostic and predictive molecular factors in patients with SI-NETs. **Chapter 7** investigates the clinicopathological significance of driver mutations detected in metastatic well differentiated SI-NETs. In this chapter, the correlation between mutational status of SI-NETs and Ki-67 index, SSTR2A expression and disease specific survival is explored. **Chapter 8** investigates the performance of fluorescence in situ hybridization (FISH) as a diagnostic test for detecting functional *RET* fusions. Lung cancer tissue was analyzed in parallel for *RET* gene fusions by FISH and targeted RNA NGS in routine diagnostics. WGS data was used to explore whether disruptive events in the *RET* locus resulted in a functional *RET* fusion.

Last, **Chapter 9** presents a summary of this thesis and the future perspectives.

References

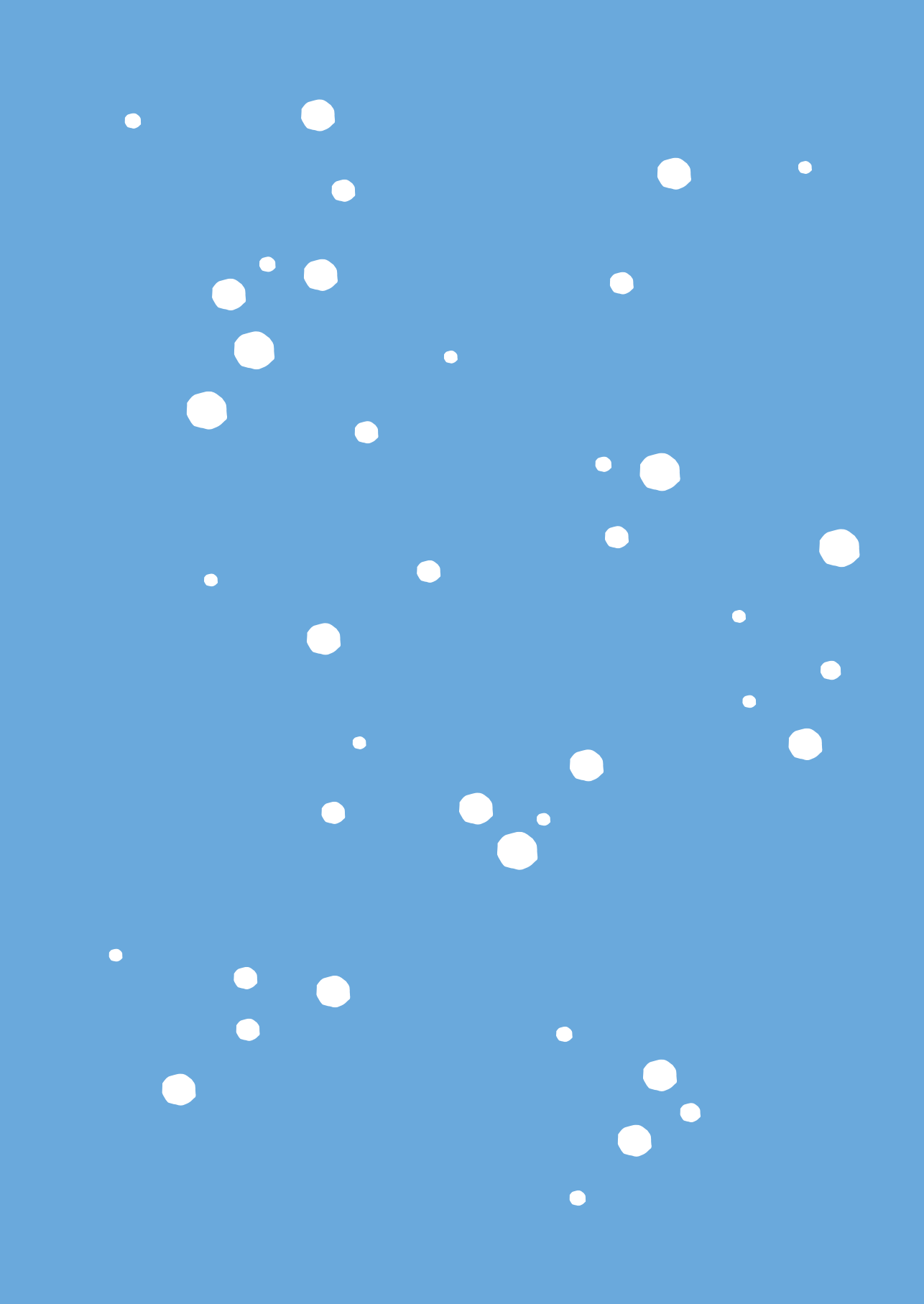
1. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31-46.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304-51.
3. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376-80.
4. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci Transl Med. 2011;3(111):111ra21.
5. Yates LR, Seoane J, Le Tourneau C, Siu LL, Marais R, Michiels S, et al. The European Society for Medical Oncology (ESMO) Precision Medicine Glossary. Ann Oncol. 2018;29(1):30-5.
6. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med. 2004;350(21):2129-39.
7. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561-6.
8. Roepman P, de Bruijn E, van Lieshout S, Schoenmaker L, Boelens MC, Dubbink HJ, et al. Clinical Validation of Whole Genome Sequencing for Cancer Diagnostics. J Mol Diagn. 2021;23(7):816-33.
9. Steeghs EMP, Groen HJM, Schuuring E, Aarts MJ, Damhuis RAM, Voorham QJM, et al. Mutation-tailored treatment selection in non-small cell lung cancer patients in daily clinical practice. Lung Cancer. 2022;167:87-97.
10. Lynch JA, Berse B, Rabb M, Mosquin P, Chew R, West SL, et al. Underutilization and disparities in access to EGFR testing among Medicare patients with lung cancer from 2010 - 2013. BMC Cancer. 2018;18(1):306.
11. Riaz N, Blecua P, Lim RS, Shen R, Higginson DS, Weinhold N, et al. Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. Nat Commun. 2017;8(1):857.
12. Okamura R, Boichard A, Kato S, Sicklick JK, Bazhenova L, Kurzrock R. Analysis of NTRK Alterations in Pan-Cancer Adult and Pediatric Malignancies: Implications for NTRK-Targeted Therapeutics. JCO Precis Oncol. 2018;2018.
13. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen HZ, et al. Landscape of Microsatellite Instability Across 39 Cancer Types. JCO Precis Oncol. 2017;2017.
14. van der Velden DL, Hoes LR, van der Wijngaart H, van Berge Henegouwen JM, van Werkhoven E, Roepman P, et al. The Drug Rediscovery protocol facilitates the expanded use of existing anticancer drugs. Nature. 2019;574(7776):127-31.
15. Hyman DM, Piha-Paul SA, Won H, Rodon J, Saura C, Shapiro GI, et al. HER kinase inhibition in patients with HER2- and HER3-mutant cancers. Nature. 2018;554(7691):189-94.
16. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay JY, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. N Engl J Med. 2015;373(8):726-36.
17. Subbiah V, Puzanov I, Blay JY, Chau I, Lockhart AC, Raje NS, et al. Pan-Cancer Efficacy of Vemurafenib in BRAF (V600)-Mutant Non-Melanoma Cancers. Cancer Discov. 2020;10(5):657-63.

18. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 2015;372(26):2509-20.
19. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, et al. Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. N Engl J Med. 2018;378(8):731-9.
20. Gatta G, van der Zwan JM, Casali PG, Siesling S, Dei Tos AP, Kunkler I, et al. Rare cancers are not so rare: the rare cancer burden in Europe. Eur J Cancer. 2011;47(17):2493-511.
21. de Heus E, Duijts SFA, van der Zwan JM, Kapiteijn E, Nieveen van Dijkum EJM, van Herpen CML, et al. The gap between rare and common cancers still exists: Results from a population-based study in the Netherlands. Eur J Cancer. 2022;167:103-11.
22. Horak P, Heining C, Kreutzfeldt S, Hutter B, Mock A, Hullein J, et al. Comprehensive Genomic and Transcriptomic Analysis for Guiding Therapeutic Decisions in Patients with Rare Cancers. Cancer Discov. 2021;11(11):2780-95.
23. Varadhachary GR. Carcinoma of unknown primary origin. Gastrointest Cancer Res. 2007;1(6):229-35.
24. Hemminki K, Bevier M, Hemminki A, Sundquist J. Survival in cancer of unknown primary site: population-based analysis by site and histology. Ann Oncol. 2012;23(7):1854-63.
25. Meijer L, Verhoeven RHA, de Hingh I, van de Wouw AJ, van Laarhoven HWM, Lemmens V, et al. Extensive diagnostic work-up for patients with carcinoma of unknown primary. Clin Exp Metastasis. 2021;38(2):231-8.
26. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, et al. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. JAMA Oncol. 2017;3(10):1335-42.
27. Hallet J, Law CH, Cukier M, Saskin R, Liu N, Singh S. Exploring the rising incidence of neuroendocrine tumors: a population-based analysis of epidemiology, metastatic presentation, and outcomes. Cancer. 2015;121(4):589-97.
28. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, et al. One hundred years after "carcinoid": epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. J Clin Oncol. 2008;26(18):3063-72.
29. Korse CM, Taal BG, van Velthuysen ML, Visser O. Incidence and survival of neuroendocrine tumours in the Netherlands according to histological grade: experience of two decades of cancer registry. Eur J Cancer. 2013;49(8):1975-83.
30. Fisher AT, Titan AL, Foster DS, Worth PJ, Poultsides GA, Visser BC, et al. Management of Ileal Neuroendocrine Tumors with Liver Metastases. J Gastrointest Surg. 2020;24(7):1530-9.
31. Hassan SA, Palaskas NL, Agha AM, Iliescu C, Lopez-Mattei J, Chen C, et al. Carcinoid Heart Disease: a Comprehensive Review. Curr Cardiol Rep. 2019;21(11):140.
32. Landerholm K, Zar N, Andersson RE, Falkmer SE, Jarhult J. Survival and prognostic factors in patients with small bowel carcinoid tumour. Br J Surg. 2011;98(11):1617-24.
33. Mocellin S, Nitti D. Gastrointestinal carcinoid: epidemiological and survival evidence from a large population-based study (n = 25 531). Ann Oncol. 2013;24(12):3040-4.

34. Kelly S, Aalberg J, Agathis A, Phillips K, Haile S, Haines K, et al. Predicting Survival of Small Intestine Neuroendocrine Tumors: Experience From a Major Referral Center. Pancreas. 2019;48(4):514-8.
35. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575(7781):210-6.
36. Campbell PJ GG, Korbel JO, Stuart JM, Jennings JL, Stein LD, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82-93.
37. Wong M, Mayoh C, Lau LMS, Khuong-Quang DA, Pinese M, Kumar A, et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. Nat Med. 2020;26(11):1742-53.
38. Shukla N, Levine MF, Gundem G, Domenico D, Spitzer B, Bouvier N, et al. Feasibility of whole genome and transcriptome profiling in pediatric and young adult cancers. Nat Commun. 2022;13(1):2485.

PART 1

Implementation of Whole Genome Sequencing in routine clinical care



CHAPTER 2

Study protocol: Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient (WIDE)

Kris G. Samsom*, Linda J.W. Bosch*, Luuk J. Schipper*, Paul Roepman, Ewart de Bruijn, Louisa R. Hoes, Immy Riethorst, Lieke Schoenmaker, Lizet E. van der Kolk, Valesca P. Retèl, Geert W.J. Frederix, Tineke E. Buffart, Jacobus J.M. van der Hoeven, Emile E. Voest, Edwin Cuppen, Kim Monkhorst & Gerrit A. Meijer

* *Equally contributing first authors*

BMC Medical Genomics, 2020

Abstract

Background

'Precision oncology' can ensure the best suitable treatment at the right time by tailoring treatment towards individual patient and comprehensive tumour characteristics. In current molecular pathology, diagnostic tests which are part of the standard of care (SOC) only cover a limited part of the spectrum of genomic changes, and often are performed in an iterative way. This occurs at the expense of valuable patient time, available tissue sample, and interferes with 'first time right' treatment decisions. Whole Genome Sequencing (WGS) captures a near complete view of genomic characteristics of a tumour in a single test. Moreover, WGS facilitates faster implementation of new treatment relevant biomarkers. At present, WGS mainly has been applied in study settings, but its performance in a routine diagnostic setting remains to be evaluated. The WIDE study aims to investigate the feasibility and validity of WGS-based diagnostics in clinical practice.

Methods

1200 consecutive patients in a single comprehensive cancer centre with (suspicion of) a metastasized solid tumour will be enrolled with the intention to analyse tumour tissue with WGS, in parallel to SOC diagnostics. Primary endpoints are (1) feasibility of implementation of WGS-based diagnostics into routine clinical care and (2) clinical validation of WGS by comparing identification of treatment-relevant variants between WGS and SOC molecular diagnostics. Secondary endpoints entail (1) added clinical value in terms of additional treatment options and (2) cost-effectiveness of WGS compared to SOC diagnostics through a Health Technology Assessment (HTA) analysis. Furthermore, the (3) perceived impact of WGS-based diagnostics on clinical decision making will be evaluated through questionnaires. The number of patients included in (experimental) therapies initiated based on SOC or WGS diagnostics will be reported with at least 3 months follow-up. The clinical efficacy is beyond the scope of WIDE. Key performance indicators will be evaluated after every 200 patients enrolled, and procedures optimized accordingly, to continuously improve the diagnostic performance of WGS in a routine clinical setting.

Discussion

WIDE will yield the optimal conditions under which WGS can be implemented in a routine molecular diagnostics setting and establish the position of WGS compared to SOC diagnostics in routine clinical care.

Background

Matching the best possible treatment with specific characteristics of a patient's tumour is the aim and challenge of 'precision oncology'. Development of a wide range of targeted drugs and their associated biomarkers has led to a large variety of diagnostic platforms being used in standard of care (SOC) molecular diagnostics. These include targeted next generation sequencing (NGS) panels, RNA-based NGS fusion analysis, Sanger sequencing, reverse transcription polymerase chain reaction (RT-PCR), fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC). Each of these tests covers only a single or limited part of the spectrum of relevant genomic changes. As a consequence, multiple tests are often performed iteratively in clinical practice, starting with profiling of the most prevalent biomarkers, followed by less prevalent biomarkers if necessary. However, this occurs at the expense of valuable patient time, tissue sample, and this strategy may interfere with making 'first time right' treatment decisions.

Hence, there is a need for an affordable comprehensive diagnostic approach, which optimally uses the available tumour tissue, can report within an acceptable time frame and is able to keep up with the pace of the rapidly changing current oncological landscape with respect to new treatment options and associated biomarkers. This rapidly changing oncological landscape poses a major challenge for standard molecular diagnostics since there is an increasing need for broader molecular testing of a growing patient population. Whole Genome Sequencing (WGS) captures a near complete overview of genomic characteristics of a tumour in one test, using a relatively low amount of tumour material. WGS thereby allows the streamlining of laboratory logistics – and thus patient care – with one comprehensive test. Furthermore, due to the steadily decreasing DNA sequencing costs over the past decades, WGS is becoming an attractive alternative to standard molecular diagnostics. Additional potential benefits of WGS include the option to test biomarkers for experimental clinical trials and the development of new biomarkers including complex biomarkers such as signatures. Most importantly, the potential therapeutic benefit of WGS for patients with metastatic cancer has clearly been documented. In recent years, the Hartwig Medical Foundation (HMF), in collaboration with the Center for Personalized Cancer Treatment (CPCT), has performed an in-depth retrospective pan-cancer WGS analysis on metastatic tumour and normal genome analysis of the first 2,500 patients. Based on an analysis of a subset of these patients (n=1,480), at least one 'clinically actionable' target could be identified for up to 62% of patients (1). In 31% of the subset, a match was found for an actionable target and a registered and approved therapy. In 13% of cases, this

match was an off-label indication (approved for another tumour type) for a target, which most likely would not have been detected using common panel-based NGS. These genetic variants were distributed over all possible mutation classes and across tumour types underpinning the importance of comprehensive genomic tumour profiling for precision medicine. In another study from the same consortium, it was shown that more than 30% of patients whom received such off-label treatment showed clinical benefit across a diversity of targeted treatments (1).

Whilst the potential of WGS as a comprehensive diagnostic tool has been demonstrated in a number of studies (2-6), its feasibility in a routine diagnostic setting has not yet been demonstrated. The WIDE (WGS Implementation in standard Diagnostics for Every cancer patient) study therefore aims to investigate feasibility of WGS-based diagnostics in routine practice and aspects such as clinical validity, cost-effectiveness, added value and the contribution of WGS and clinical data to a centralized database for facilitating cancer research and improving care for future patients (7). Importantly, WIDE adopts the approach of periodically (i.e. after every 200 patients) evaluating key performance indicators and optimizing procedures accordingly in order to achieve continuous improvement of the performance of WGS in a routine clinical setting.

Here, we describe the conceptual design and experimental conditions for this study.

Objectives

Primary objectives

Primary study endpoint	Primary outcome measure
Feasibility	Percentage of patients for whom processing from biopsy to WGS report is successful and the turnaround time (TAT) of biopsy until WGS report in working days.
Clinical validation	Percentage (%) of concordant variants (%) between WGS and SOC molecular based diagnostics.

Secondary objectives

Secondary study endpoint	Secondary outcome measure
Additional treatment options	Percentage of patients for whom potential treatment options (in clinical trials in the Netherlands) are identified by WGS, which have not been identified with SOC diagnostics.
Health Technology Assessment	Costs and benefits associated with WGS and SOC diagnostics.
Better informed decision making and experience of the treating clinician	Opinion of treating clinicians on the added value of WGS for clinical decision making compared to SOC diagnostics evaluated through questionnaires.
Expand HMF database	The number of patients for whom clinical and WGS data are added to the HMF database for biomarker discovery in cancer research.

Methods

Study design

WIDE is a prospective observational diagnostic study in the Netherlands Cancer Institute (NKI) which aims to include 1200 patients with metastasized cancer in a time frame of 18-24 months. The study has been approved by the Medical Ethical Committee of the Netherlands Cancer Institute. WIDE is a collaboration between NCI (a comprehensive cancer cancer), the Hartwig Medical Foundation (HMF, a non-profit organisation), and the Utrecht UMC (UMCU, an academic medical hospital).

Study population

1200 consecutive patients from the Netherlands Cancer Institute will be recruited from patients with (a suspicion) of stage IV disease of solid tumours who are treated at the NCI without pre-selection on tumour type. Patients are eligible when a biopsy or resection material can be safely obtained during a routine diagnostic procedure. Once a SOC biopsy procedure (4 biopsies if possible) has been performed, the

tumour material is used for both SOC diagnostics and WGS analysis. Patients of whom a fresh frozen tumour sample has been obtained earlier (either at the NCI or elsewhere), i.e. archival fresh frozen material, are also eligible when they will be treated at the NCI and routine molecular diagnostic analysis on the archival tumour material is requested. Use of archival fresh frozen material is excluded in patients whom have received tyrosine kinase inhibitors after the retrieval of tumour material as such treatments may shift the genomic profile by clonal selection. Patients who underwent allogeneic stem cell transplantation, or transplantation of the organ from which the tumour originated or is located, are excluded as well, because matching tumour and blood DNA is required for identifying a full tumour-specific mutational profile. For each WIDE patient no more than one successful WGS analysis will be obtained, except when molecular analysis is indicated in the context of resistance mechanisms which can be relevant for subsequent treatment options e.g. tyrosine kinase inhibitors in lung cancer. If so, patients can have multiple tumour samples analysed by WGS. Patients must be 18 years or older and willing and able to comply with the protocol as judged by the investigator, as well as sign a written consent.

Study workflow

WGS will be performed prospectively in parallel with and independently of SOC diagnostics, in which SOC diagnostics may or may not include molecular diagnostics (Figure 1).

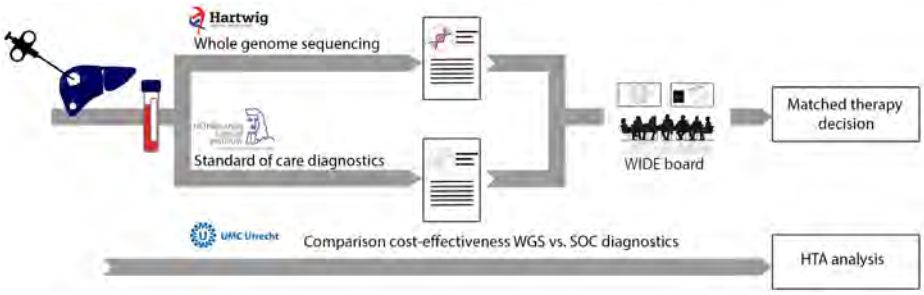


Figure 1. Workflow of the WIDE study. Patients with (a suspicion) of stage IV cancer undergoing a tumour biopsy as part of the routine standard of care at NCI are eligible for inclusion in the WIDE study. In addition, a blood sample is drawn. Subsequently, both a fresh frozen tumour and a blood sample are shipped to HMF for WGS analysis, and a tumour sample will be assessed according to standard of care (SOC). Both the results from WGS and SOC are discussed in a dedicated molecular tumour board and reported for clinical decision making. Alongside, a cost-effectiveness comparison of WGS versus SOC diagnostics will be performed.

Sample collection

A fresh sample of the primary or metastatic tumour will be obtained as part of routine standard of care (SOC) diagnostic procedures. This can be achieved by means of a needle biopsy, resection or collection of body fluids containing tumour cells, e.g. pleural effusion or ascites, all performed in the context of SOC. If possible, routine SOC diagnostic procedures usually involve multiple tumour biopsies (1-4 biopsies), depending on safety and risk of complications. The majority of biopsy procedures are image guided (e.g. CT or ultrasound). In addition, a 10ml whole blood sample will be obtained to isolate normal DNA in order to be able to discriminate somatic mutations from the patient's germline DNA background variations.

Sample processing

Depending on the amount of tumour material and the clinical question, a pathologist determines whether frozen sections are cut from the biopsies or resection specimens.. In case of insufficient tumour material, no frozen sections are cut and all material is used for standard diagnostics. If frozen sections are cut, a pathologist subsequently assesses the tumour cell percentage (TCP). In case of multiple biopsies, a pathologist decides based on the TCP which biopsy is most suitable for standard diagnostics (depending on the clinical question) and which biopsy for WGS analysis. Subsequently, the tumour area is marked for manual macrodissection when needed. For body fluids (e.g. pleural effusion or ascites) a TCP of 30% or higher is considered sufficient for WGS. For tissue samples (i.e. biopsies or resections), a TCP of 20% or higher is considered sufficient. When these conditions are met, the frozen tumour sample and a tube of blood are shipped to HMF within 24 hours by courier. The biopsy designated for standard diagnostics and leftover biopsy material from macrodissection is processed and embedded in paraffin (formalin fixed paraffin embedded (FFPE)) according to standard procedures. The NKI department of pathology operates under ISO15189 accreditation.

WGS and Bioinformatics

Whole genome DNA sequencing is performed according to standard procedures described previously (1). All procedures are automated as much as possible and the Illumina® NovaSeqX and NovaSeq6000 platforms are used. First, shallow whole-genome sequencing is used to determine an accurate tumour purity of the samples before continuing full sequencing at a sequencing depth of >90x. DNA isolated from blood is sequenced at depth of >30x as a germline control. Sequencing data is analysed with an optimized in-house bioinformatic pipeline designed to detect all types of somatic alterations, including single and multiple nucleotide substitutions (SNV and MNV), insertions and deletions (indels), copy number

alterations (amplifications and gene copy losses) and genomic rearrangements and structural variants (e.g. gene fusions). (8). Results from the tumour and germline sample are compared to filter out germline polymorphisms, which enables the reporting of somatic variants and therapeutically actionable germline variants (as described below) only. All other germline variations are automatically subtracted from the somatic mutations in the bioinformatics analysis and are not disclosed to the investigators, nor are they reported. All code and scripts used for analysis of the WGS data are publicly available via Github (9). HMF has established procedures for WGS under ISO17025 accreditation.

SOC molecular diagnostics

The standard of care molecular diagnostics portfolio at NCI comprises of targeted next generation sequencing (NGS) panel (Ampliseq, Cancer hotspot panel V2, Illumina Inc, San Deigo, United States of America), RNA-based NGS fusion analysis (Archer Fusionplex, Lung and Sarcoma panels, Archer DX Inc, Boulders, United States of America), Sanger sequencing, reverse transcriptase polymerase chain reaction (RT-PCR), Multiplex fragment analysis polymerase chain reaction, High Resolution Melting (HRM), fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC), respectively.

Reporting

A detailed molecular patient report (OncoAct, Hartwig Medical Foundation, Amsterdam, the Netherlands) is produced with all variants that are relevant for diagnostic purposes and cancer treatment decision making (actionable variants), including both somatic and germline variants. Upon patient preference declared in the informed consent, selected germline variants are being reported back as inherited variants, along with an offer for routine clinical genetics counselling. Moreover, actionable mutations are linked to existing biomarker-based clinical studies. In case a reported variant is identified by WGS but not in SOC or vice versa, additional verification tests will be performed in order to resolve the cause of the discordance. WGS and SOC results from all patients are discussed in a weekly WIDE-dedicated molecular tumour board, consisting of clinical molecular biologists, pathologists, clinical geneticists and medical oncologists. Ultimately, according to standard reporting procedures of the NKI, WGS results are included in the pathology report and included in the electronic patient record (Hix, Chipsoft, Amsterdam, The Netherlands) as well as the Dutch national pathology digital archive (PALGA) (9). Results from both SOC and WGS diagnostics are used for clinical decision making and can be discussed if requested by the treating clinician in a molecular tumour board (MTB). Treatment decisions will be made based on the expertise of these multiple

disciplines. In case of any persistent discordant results between WGS and SOC diagnostics, SOC findings will be leading in treatment decision making. Actionable variants identified by WGS can therefore potentially result in adjusting the (initial) treatment plan. The number of patients included in (experimental) therapies initiated based on SOC or WGS diagnostics will be reported with at least 3 months follow up. The clinical efficacy of (experimental) therapies is beyond the scope of WIDE.

Primary endpoints

Feasibility of WGS in routine clinical practice

In order to measure the feasibility of WGS in routine clinical practice, the percentage of patients for whom processing from biopsy to WGS report is successful in a turnaround time (TAT) of 12 working days, will be determined.

Clinical validation

For clinical validation, the percentage of variants for which WGS detects (at minimum) the same treatment-relevant variants as DNA-based SOC tests, will be reported. Variants considered include Single Nucleotide Variants (SNVs), Multi Nucleotide Variants (MNVs), Insertions/Deletions (INDELs), Copy Number Alterations (CNAs), Structural Variants (SVs) and other tumour characteristics (e.g. MSI status). For this endpoint, only those variants which are detectable with SOC will be evaluated.

Secondary endpoints

Health Technology Assessment

For the HTA, micro-costing and budget impact analysis of WGS compared to SOC diagnostics will be performed, based on the Activity Based Costing Method. The HTA will include the costs of personnel, turnaround time (TAT), equipment/material, tests and platforms used, potentially iterative sequences of SOC tests, consequences of treatment decisions made etc. compared to WGS. In order to extrapolate findings nation-wide, the global International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines will be used.

Additional treatment options

For the endpoint additional treatment options, the percentage of patients for whom treatment-relevant variants identified by WGS and not by SOC diagnostics, for which on-label or off-label drugs exist, will be reported and details will be described.

Better informed clinical decision making

The opinion of treating clinicians on the added value of WGS in terms of clinical decision making compared to SOC diagnostics will be collected through digital questionnaires (qualitative analysis) which are developed for this study and provided as Supplementary file. In addition, feedback of clinicians is collected by performing in depth interviews in order to ensure that the WGS report format meets their needs. If treating clinicians experience barriers in acting upon the patient WGS report format, additional educational activities will be arranged and/or the report format will be adapted accordingly.

Enriching the HMF database

In a learning healthcare system, it is pivotal to generate as much as possible potentially informative real world data, in this case clinically well annotated WGS results of metastatic tumours which are made available for future research. Therefore the data generated within the WIDE study will be added to the HMF database (https://www.hartwigmedicalfoundation.nl/data-aanvragen/). This database (currently storing >4500 patients) contains pseudonomized genetic and clinical data, including treatment and treatment outcomes, of individuals whose tumours have been sequenced by HMF in the context of multiple studies. Variables documented include date of biopsy, biopsy site, sample type, standard of care and WGS characteristics, biomarkers identified, diagnosis, pre-treatment, treatment and radiological or clinical treatment response (predominantly according to Response Evaluation Criteria In Solid Tumours (RECIST) 1.1). Only patients who provide additional consent for re-use of their genomic and (limited) clinical data for cancer research purposes will be added to the HMF database. The HMF database is accessible for international researchers through an access-controlled mechanism. The data access request procedure involves evaluation of scientific, legal and ethical aspects of the intended data usage and applications are assessed by an independent scientific and data access board.

Statistical analysis

Sample size calculation

The sample size of 1200 patients has been calculated based on the primary endpoint 'clinical validation'. The aim is to detect in at least 95% of the cases the same variants as SOC molecular tests. For the lower limit of the confidence interval to be at least 95% and under the assumption that the real concordance rate will be 97.5%, we will have a power of 96% to find the lower limit of the confidence interval to be at least 95% when we include 624 individual detected variants. Assuming that 30% of biopsies do not have detectable genomic aberrations in SOC, a total of 1200 patients need to be included in this study.

Continuous evaluation and improvement

In the WIDE study, the primary and secondary endpoints are evaluated periodically after every 200 patients. Based on these interim analyses, according to the PDCA approach (plan-do-check-act) and in line with ISO 151890 principles, procedures are being optimized and effects monitored, aiming to achieve continuous improvement of WGS implementation in a routine clinical setting. After completing the procedure for every 200 patients, feasibility, clinical validation and added value are systematically evaluated by descriptive statistics. In order to ensure that the sample size for the endpoint clinical validation is reached and the added value is safeguarded, there are multiple strategies in place. If the number of patients for whom SOC molecular diagnostics is performed is <70%, the inclusion strategy of the treating physicians will be discussed and adjusted to select more patients for whom routine molecular diagnostics is indicated. If for any given tumour type the added value is less than 10%, the inclusion criteria will be adjusted (e.g. excluding tumour types for which no additional targets are being found). At the start, halfway (upon inclusion of 600 patients) and at the end of the WIDE study, the (expectations of the) clinical value of WGS according to treating clinicians will be evaluated by means of a digital survey. In case less than 50% of the treating clinicians indicate WGS has added value compared to SOC diagnostics, inclusion criteria will be reconsidered. Similarly halfway, the patient report format will be evaluated based on comprehensibility and usefulness of the provided information. We aim to hereby identify important barriers and facilitators for integrating WGS-based diagnostics in clinical practice.

Discussion

In the WIDE study, for the first time the feasibility and impact of WGS-based diagnostics in routine clinical practice will be systematically investigated in a prospective setting. The WIDE study embeds WGS, with its unprecedented opportunities to match the specific biology of a patient's tumour with the right drug, in a routine pathology diagnostic workflow, thereby allowing streamlining of patient care and laboratory logistics. In patient care, treating clinicians play a crucial role in the implementation of WGS and therefore they will receive dedicated training on interpretation of the patient reports. Moreover, all patient reports based on WGS data as well as the results of SOC diagnostics will be discussed in a multidisciplinary MTB. Regarding laboratory logistics at HMF, experimental, bioinformatics and reporting procedures will be optimized for use in routine diagnostics (including reduction of TAT to 12 working days). Similarly, laboratory logistics at the NCI will be adjusted to accommodate WGS diagnostics. Importantly, logistics that can work with fresh or fresh-frozen samples (beyond standard frozen section diagnostics) rather than FFPE (formalin fixed paraffin embedded) need to be incorporated in the routine diagnostic workflow. Formalin fixation fragments DNA and causes artefacts, which interfere with genome-wide mutation and structural variant calling. Moreover, fixation renders a substantial amount of DNA unusable by which precious tumour material is lost. As a consequence, frozen sections need to be cut from every tumour sample for sample input evaluation. Since a tumour cell percentage of >20% is a prerequisite for successful and cost-effective WGS analysis (at 90/100x sequencing depth), the tumour cell percentage needs to be assessed first by a pathologist before shipping to HMF. As a result of the continuous evaluation and improvement approach, diagnostic and treatment logistics will frequently be adapted. As such this study will not only investigate feasibility, clinical value, added value and cost-effectiveness, but also identify a sound generic implementation strategy for WGS in routine clinical practice which can be adopted by other interested hospitals. Consequently, this study will generate new insights in the barriers and facilitators that can be encountered when WGS is integrated in routine clinical care and form the basis of a generic implementation strategy. Moreover, WGS data will be generated of 1200 metastasized solid tumours, which will shed more light on tumour molecular biology and options for targeted therapy, particularly in tumour types for which molecular diagnostics is not yet part of standard of care. The WGS and clinical data will be added to the HMF database which provides extensive opportunities for future research.

References

1. Van der Velden D, Hoes L, van der Wijngaart H, van Berge Henegouwen J, van Werkhoven E, Roepman P et al.The Drug Rediscovery protocol facilitates the expanded use of existing anticancer drugs. Nature. 2019;574(7776):127-131. doi: 10.1038/s41586-019-1600-x.
2. Priestley P, Baber J, Lolkema M, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumors. Nature. 2019;575 (7781):210-216. doi: 10.1038/s41586-019-1689-y.
3. Rusch M, Nakitandwe J, Shurtleff S, Newman S, Zhang Z, Edmonson M et al. Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. Nature Communications. 2018;9(1). doi: 10.1038/s41467-018-06485-7.
4. Horak P, Fröhling, S, Glimm, H. Integrating next-generation sequencing into clinical oncology: strategies, promises and pitfalls. ESMO Open. 2016;1(5)e000094.
5. Klintman J, Barmpouti K, Knight SL, Robbe P, Dreau H, Clifford R et al. Clinical-grade validation of whole genome sequencing reveals robust detection of low-frequency variants and copy number alterations in CLL. British Journal of Haematology 2018;182(3), 412-417. doi: 10.1111/bjh.15406.
6. Sosinsky A, Ambrose J, Zarowiecki M, Mitchell J, Henderson S, Murugaescu N et al. 100,000 genomes project: Integrating whole genome sequencing (WGS) data into clinical practice. Annals of Oncology. 2019;30 (suppl_7): vii1-vii35.10.1093/annonc/mdz238.
7. Database with genetic and clinical information of cancer patients [Internet]. Stichting Hartwig Medical Foundation. 2020 [cited 25 June 2020]. Available from: https://www.hartwigmedicalfoundation.nl/en/
8. Cameron D, Di Stefano L, Papenfuss A. Comprehensive evaluation and characterization of short read general-purpose structural variant calling software. Nature Communications. 2019;10(1). doi: 10.1038/s41467-019-11146-4.
9. Github [Internet]. Hartwig Medical Foundation. 2020 [cited 25 June 2020]. Available from: https://github.com/hartwigmedical/
10. Casparie M, Tiebosch A, Burger G, Blauwgeers H, van de Pol A, van Krieken J et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncology. 2-2007;29(1):19-24. doi: 10.1155/2007/971816.

CHAPTER 3

Optimization of the workflow for Whole Genome Sequencing in routine pathology practice

Kris G. Samsom, Linda J.W. Bosch, Luuk J. Schipper, Daoin Schout, Paul Roepman, Mirjam C. Boelens, Ferry Lalezari, Elisabeth G. Klompenhouwer, Adrianus J. de Langen, Tineke E. Buffart, Immy Riethorst, Lieke Schoenmaker, José G. van den Berg, Ewart de Bruijn, Jacobus J.M. van der Hoeven, Hans van Snellenberg, Lizet E. van der Kolk, Edwin Cuppen, Emile E. Voest, Gerrit A. Meijer & Kim Monkhorst

Nature Protocols for Clinical Use, 2023

Introduction

The recent and ongoing development of a wide range of targeted drugs and their associated biomarkers has led to multiple diagnostic platforms being utilized in routine molecular diagnostics. Each of these platforms covers limited parts of the spectrum of relevant genomic changes and is often performed in a sequential manner which occurs at the expense of time, available tumor tissue, and efficient clinical decision making. Moreover, pathology laboratories struggle with the rapidly changing oncological drug landscape, which substantially delays clinical implementation of newly approved biomarkers and patient access to these drugs[1]. Hence, there is a need for clinical implementation of a single comprehensive and future-proof test, such as whole-genome sequencing (WGS), which detects all possible biomarkers within an acceptable timeframe.

Development of the protocol

This protocol is the result of the clinical implementation process of WGS during the *whole-genome sequencing Implementation in standard Diagnostics for Every cancer patient* (WIDE) study. During the WIDE study, 1200 consecutive patients with metastatic cancer received WGS independent of and in parallel with standard of care (SOC) diagnostics on routinely obtained tumor samples. The study demonstrated that WGS is a feasible and clinically valid technique in routine clinical practice with a turnaround time of 11 workdays[2]. Given the success of the WIDE study, WGS was implemented at the Netherlands Cancer Institute (NKI), a comprehensive cancer center, as part of routine clinical care since January 2021. The introduction of WGS in a routine diagnostic setting has yielded valuable lessons learned and identified associated challenges, i.e., adopting a fresh frozen tissue workflow. In this protocol, we present a sound generic implementation strategy for WGS in clinical cancer care, which hospitals worldwide can adopt (for an overview of the protocol, see **Figure 1**).

Applications

WGS can detect mutations, copy number variants (amplifications/losses), structural variants, oncogenic fusions, mutational signatures (e.g., tumor mutational burden (TMB), homologous recombination deficiency (HRD), and microsatellite instability (MSI)), and viral insertions in tumor DNA. As WGS is typically performed on paired tumor and normal samples, it can also identify (pathogenic) germline variants in the accompanying blood reference sample. In addition, WGS can type HLA alleles of tumor and germline DNA (4-digit). The applications of WGS in cancer diagnostics, therefore, range from characterizing the complete genome of the tumor and identifying all the treatment-relevant biomarkers to detecting viruses and implementing

pharmacogenomics, for example dihydropyrimidine dehydrogenase (DPD) and UDP Glucuronosyltransferase Family 1 Member A1 (UGT1A1) for the use of 5-FU and irinotecan respectively. With these applications, WGS improves tumor diagnosis (e.g., genome-based classification, subclassification, and prediction of tumor origin in patients with cancer with an unknown primary tumor (CUP) and therapeutic decision making by, e.g., integration of germline information. Clinical applications of WGS beyond oncology include its use in Mendelian and complex diseases.

Alternative methods for WGS-based diagnostics

Current clinically accepted alternatives for WGS-based molecular profiling include conventional (small) next-generation sequencing (NGS) DNA panels combined with targeted RNA based panels for detecting rearrangements and multiple individual large gene panels, such as Illumina TruSight Oncology 500 (TSO500), MSKIMPACT™, Oncomine™ Plus, and whole-exome sequencing. While NGS DNA panels can potentially detect clinically actionable alterations in more than 500 cancer-related genes, provide a TMB estimate and determine MSI, and are compatible with FFPE material, their sustainability is limited due to the extensive wet lab validation of any update required in case of newly discovered biomarkers. Moreover, NGS DNA panels cannot effectively detect gene fusions, complex mutational signatures (i.e., HRD), or accurately detect copy number variants (CNVs). Because of these limitations, NGS DNA panels require additional tests for comprehensive analyses. In contrast, WGS can detect all genomic features in one test and has evident added value in the differential diagnosis of complex cases, especially CUP, for which WGS includes an integrated prediction algorithm called CUPPA[3]. Furthermore, because of its comprehensive nature, the WGS data can be used seamlessly for future research and analyses. Of note, WGS requires fresh frozen tumor material as input and currently requires a tumor cell percentage (TCP) of at least 20%. With steadily decreasing sequencing costs, sequencing with a deeper coverage will become possible and will potentially decrease the required TCP.

Experimental design

The presented protocol has been developed from the perspective of a routine diagnostic pathology laboratory in a comprehensive cancer center, operating under ISO15189:2012 accreditation and according to professional guidelines. The workflow of WGS in routine diagnostics is illustrated in **Figure 1**.

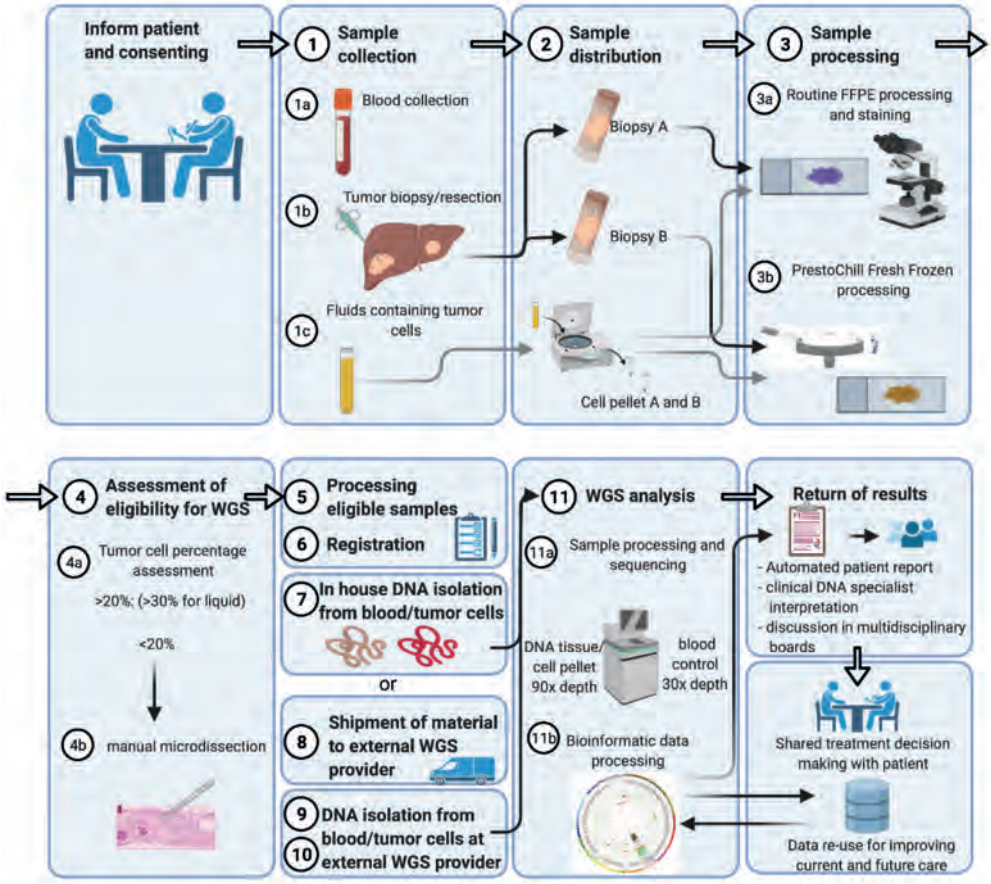


Figure 1. Overview of procedures involved in WGS implementation in routine diagnostics which are described in this protocol. This figure was created with BioRender.com.

Abbreviations: FFPE = Formalin Fixed Paraffin Embedded, WGS = Whole-genome sequencing

Streck tubes are needed for blood collection, which is required for the test to discriminate somatic mutations from germline background variations in bioinformatics analyses. Streck tubes are used because of their DNA stabilizing proprieties for up to 7 days at 15-30°C, allowing for high quality DNA isolation. Tumor samples consist of freshly obtained surgical resection specimens, diagnostic biopsies, or cytology specimens. Archival fresh frozen samples from prior surgical resection specimens may also qualify. Cytological specimens are processed with a cytospin to make cell pellets. Efficient tissue freezing of fresh biopsies is facilitated by a PrestoCHILL device (Milestone Medical, Michigan, USA), which limits freezing artifacts. These devices are already widely used in pathology for automated embedding and staining during Mohs micrographic surgery procedures [4]. The

specimens are cut into frozen sections of 5 µm on coated glass and hematoxylin- and eosin-stained. Next, a pathologist microscopically assesses a TCP on frozen sections and demarcates a tumor area for manual microdissection if needed. Performing WGS requires a TCP of ≥20% for tissue specimens, ≥30% for cytology specimens and at least 50 ng of tumor DNA. The sequencing process also needs a high-throughput DNA sequencing platform, e.g., NovaSeq6000 (Illumina, San Diego, USA), combined with a large-scale infrastructure for analyzing and storing human DNA data, either on the premise or via an external WGS provider.

Successful implementation of WGS in routine clinical practice requires well-trained, qualified staff operating under effective quality management during all process steps, from biopsy to interpreting sequencing data in a clinical context. However, these competences do not exceed the standard competences in a certified medical (laboratory) environment.

Overview of procedures

This protocol describes all laboratory logistics associated with WGS-based diagnostics in clinical care. Procedure 1 describes the collection of blood and tissue or fluid containing tumor cells. The subsequent procedures (procedures 2 and 3) highlight how to distribute samples for WGS-based and SOC diagnostics and how to process them according to the fresh frozen workflow. Procedure 4 explains how to assess TCPs on frozen sections and, if needed, how to manually microdissect samples to enhance the TCP. Procedures 5 and 6 outline the processing of tumor material and cytology specimens and the registration of samples before shipment to the external WGS provider. Instructions for DNA isolation according to the column-based extraction method are provided in procedure 7 for clinical laboratories that choose to perform DNA isolation from blood or tumor cells in-house. Procedure 8 describes the shipment process for clinical laboratories sending DNA for isolation at an external WGS provider site, and procedures 9 and 10 the DNA isolation from blood and tumor cells according to the magnetic-bead DNA extraction method. Lastly, procedure 11 outlines how the external WGS provider analyzes the DNA.

Limitations

Formalin-fixed paraffin-embedded (FFPE) samples are not suitable for generating high-quality, accurate genome-wide variant calls, which are needed for WGS analysis. Implementing a fresh frozen workflow can be slightly more complex than the commonly used FFPE procedures. However, modern pathology laboratories can easily accommodate such an FF workflow, provided appropriate staffing and equipment are available.

Another potential limitation is the required TCP of ≥20% to allow WGS analysis at the current sequencing depth of >90x. However, manual microdissection techniques can enhance the TCP and maximize the use of tumor material from small biopsy specimens. In addition, the steadily decreasing sequencing costs anticipate that WGS will eventually be able to reach a coverage of approximately 250x, which could enable the analysis of samples with TCPs as low as 8%[5].

Materials

Biological material

- Whole blood
- Fresh tumor samples from surgical resection specimens, diagnostic biopsies, or cytology specimens

Reagents

Procedures 3, 5, and 6

- Milestone Cryo Compound (MCC) (Milestone, medical@milestonesrl.com)
- Leica Cryofect (Leica Biosystems Nussloch GmbH, product no. 14038742801)
- (Dulbecco's) phosphate-buffered saline (PBS) (Life Technologies, product no. 21600069)
- Minimal Essential Medium (MEM) containing Earle's salts and L-glutamine (Thermo Fisher, cat. no. 11095080)
- FineFix Concentrate (Milestone, medical@milestonesrl.com)
- Nuclease-free Milli-Q water (MQ)

Procedure 7

- AllPrep DNA/RNA/miRNA Universal kit (Buffer EB) (Qiagen GmbH, cpc@qiagen.com, cat. no. 80224)
- RNAse-free ethanol absolute (Ethanol 100% (v/v), technical grade) (Boom BV, www.boomlab.nl, product no. 84050065)
- RNAse-free isopropanol (2-PROPANOL C+ Shell Pharma) (Chempoprack BV, product no. 3100915)
- DNAse/RNAse-free water
- DTT (Thermo Fisher Scientific Baltics UAB, product code R0861)
- Reagent DX (Qiagen GmbH, cpc@qiagen.com, cat no. 19088)
- Nuclease-free Milli-Q water (MQ)

Procedure 9

- QIAsymphony® DSP DNA kit (Qiagen GmbH, cpc@qiagen.com, cat. No. 937255)
  - Reagent cartridge
  - Enzyme tubes

Procedure 10

- QIAsymphony® DSP DNA kit (Qiagen GmbH, cpc@qiagen.com, cat. no. 937236)
  - Reagent cartridge
  - Enzyme tubes

- Buffer ATL
- Proteinase K
- DNAse/RNAse-free water

Procedure 11

- Illumina® TruSeq® Nano DNA LT Library Prep kit (Illumina®, https://www.illumina.com/)
- IDT 384 UDI UMI indexes (Illumina®, https://www.illumina.com/)
- Qubit™ dsDNA HS Assay Kit (high sensitivity; e.g. Thermo Fisher)
- KAPA™ Library Quantification Kit Illumina Platforms (Roche, sequencing.roche.com)

Tools

Procedure 1

- Streck Cell-Free DNA BCT® RUO & CE tube, 10 ml (STRECK, https://www.streck.com/products/stabilization/cell-free-dna-bct)

Procedures 3, 5, and 6

- Frozen section tables (Thermo Fisher or Leica)
- Disposable microtome blades
- Dissecting needles
- Transparent plastic spatulas (Milestone, medical@milestonesrl.com)
- Anatomical tweezers
- Lens paper disks (ø22 mm)
- Eppendorf tubes (1.5 ml and 2 ml)
- Scalpels
- Automatic pipettes (200 µl and 1000 µl)
- Nuclease-free filter pipette tips (200 µl and 1000 µl)
- Liquid nitrogen container (30x10 cm; Thermos) and nitrogen
- Cooling block for Eppendorf tubes (-20^{0}C)
- Screw skirted tube (2 ml; Sarstedt, https://www.sarstedt.com/en/products/, Stainless steel bead with a barcode; provided by external WGS provider)
- FluidX ™ black and orange tubes

Procedure 7

- Stainless steel beads (5 mm)
- Stainless steel tweezers (15-18 cm with or without ridges)
- FluidX™ tubes (Azenta, https://www.azenta.com/products) from external WGS provider

Procedure 9

Consumables for blood processing

- Streck cell-free DNA BCT preservatives tubes (10 ml; STRECK, https://www.streck.com/products/stabilization/cell-free-dna-bct/)

Consumables for QIAsymphony® Robotic system (Qiagen GmbH, cpc@qiagen.com)

- Piercing lid
- Reuse seal set
- Filter pipette tips (200 µl and 1500 µl)
- Cartridges (8 wells)
- Covers (8 rods)
- Cooling adaptor, SBS Universal (cat. no. 9243384)
- Screw skirted tubes (2 ml; Sarstedt, https://www.sarstedt.com/en/products/)
- Barcode tubes (1 ml) and rack

Procedure 10

Consumables for tissue processing

- DNA LoBind® Tubes (2 ml; Eppendorf)
- Dry ice
- Consumables for the QIAsymphony® robotic system cpg@qiagen.com Filter pipette tips (200 µl and 1500 µl)
- Cartridges (8 wells)
- Covers (8 rods)
- Accessory troughs
- Piercing lid
- Reuse seal set
- Cooling adaptor (SBS Universal, cat. no. 9243384)
- Screw skirted tubes (2 ml; Sarstedt, https://www.sarstedt.com/en/products/)
- Barcoded tubes and rack

Equipment

Procedures 3, 5, and 6

- Cryostat (-22°C to -40°C; Cryostar NX50 HOVPD – ref 957240, cat. no. S20071362)
- PrestoCHILL (-40°C; Milestone Medical, medical@milestonesrl.com, type 100850, cat. no. 182003139) and accessories
- Automated coverslipper (e.g., Tissue-Tek Film E2, cat. no. 47421411-0219)
- Centrifuge (Eppendorf, centrifuge 5810, cat. no. 5810CJ462997)
- Cytospin™ 4 (e.g., Epredia™, cat. no. A78300003)
- Laboratory refrigerator (2-8°C; e.g., Fisher Scientific, cat. no. 11317941)

Procedure 7

- Mini centrifuge
- Vortex
- Tissue Lyser LT (Qiagen GmbH, cpc@qiagen.com, cat. no. 85600)
- Heraeus pico™ 21 centrifuge (DNA; e.g., Thermo Fisher, cat. no. 41451811)
- Quantus™ Fluorometer (type E650, cat no. 517809534/840003636)

Procedure 9

- QIAsymphony® SP (Qiagen GmbH, cpc@qiagen.com)
- Plate centrifuge (Eppendorf, centrifuge 5810)
- Refrigerated centrifuge (Eppendorf, centrifuge 5810R)
- Vortex
- Whatman® paper
- Qubit™ dsDNA BR assay kit (broad range; e.g., Thermo Fisher)

Procedure 10

- TissueLyser II system (Qiagen GmbH, cpc@qiagen.com
- Stainless steel beads (5 mm)
- Centrifuge
- Vortex
- Nalgene® benchtop Cooler (-20°C; e.g. Thermo Scientific)
- Mini cooler (-20°C)
- Mini cooler (4°C)
- Mini cooler (0°C)

Procedure 11

- Illumina® NovaSeq6000 platform (Illumina®, https://www.illumina.com/)
- Covaris® shearing device (Covaris ®, https://www.covaris.com/)

Main processes involved in wgs implementation in routine diagnostics

All clinical diagnostic tests, including WGS, encompass the entire process, from clinical indication to diagnostic report and clinical decision-making. Successful implementation of WGS in routine diagnostics requires that the involved health care professionals follow a detailed protocol with all essential steps.

Clinical indication for requesting wgs

At the NKI, patients with metastasized cancer are eligible for WGS when there is a clinical indication for broad molecular testing. Clinical indications for broad molecular testing include, detection of genomic alterations to identify targeted treatment options, diagnostically challenging tumors or CUP. On the diagnostic request form in the electronic health record (EHR), treating physicians have to provide the reason for requesting WGS over other diagnostic options (e.g. progression after multiple lines of therapy, identification of targets for targeted therapy?).

Patient information

This guide highlights key points to consider when discussing the option of WGS-based diagnostics with patients with cancer. Informational materials, such as animated videos, can further support the discussion between the treating physician and patient.

1. Introduction and context of the test
WGS is a test to identify genomic alterations relevant for diagnostic purposes and clinical decision-making.
2. Sample collection
Samples: Blood draw and tumor material obtained by either fresh frozen surgical resection, needle biopsy, or cytological puncture.
3. Results
Main findings: All variants relevant for diagnostic purposes and cancer treatment (actionable gene variants for standard of care targeted therapy options or targeted therapy in clinical trial setting), including somatic and germline variants.
Germline findings: WGS uses the DNA from blood to discriminate somatic mutations in the tumor from the patient's background germline DNA variations. With respect to the national laws on germline diagnostics and upon patient request, the sequencing pipeline can report pathogenic germline variants (PGVs) in genes with diagnostic value or targeted tumor therapy implications as inherited variants. In case of an inherited variant, the patient is offered routine counseling by a clinical geneticist. A patient can also request not to be informed about inherited variants. In such cases, the sequencing pipeline reports variants present in the tumor sample without reporting the germline status. Germline variants without cancer-related actionability are not investigated or reported.
If a patient opted in, the absence of reported PGVs does not completely exclude possible genetic predisposition to cancer or other health conditions.
Interpretation and knowledge about results may change over time.
The treating physician confirms a timeline for the results (approximately 11 working days) and communicates the results to the patient.

4. Implications for the patient
Identification of a PGV, which may help eliminating diagnostic uncertainties and influence clinical decision-making, e.g., change in disease management or additional treatment options.
Possible referral to a clinical geneticist for counselling, which may result in screening and early monitoring.
Potential psychosocial impact of receiving the results.
Implications of family planning and reproductive choices.
5. Implications for family members
Possible preventive screening and management, predictive testing, and implications for reproductive choices.
6. Reuse of data
Upon patient consent, genomic and (limited) clinical data may be de-identified and used for cancer research. In such cases, the data are added to an external database belonging to the WGS provider (Hartwig Medical Foundation, a nonprofit organization, hereafter referred to as Hartwig).
De-identified data in the Hartwig database is accessible to other researchers through an access-controlled system.
Genomic data may be reanalyzed in the future as new evidence may change results and conclusions.

PROCEDURE 1: SAMPLE COLLECTION

Stage I: Blood collection

Timing 5-10 minutes

1. Use a 10-ml Streck blood collection tube to obtain a whole blood sample.
2. Mark the tube with the patient ID.
3. Complete the registration form via e-mail (an example can be found in **Supplementary 1**)
4. Place the tube in a blister with absorption material and a medical envelope.
5. Ship the blood sample to the external WGS provider via postal mail within 24 hours for a centralized analysis (see Procedure 9) after collection.

Lessons learned from the implementation process

- Streck blood collection tubes protect the blood cells from high temperatures (especially during the summer months), which may affect sequencing data quality.
- WGS logistics are streamlined by the combined ordering of the blood and tissue collection in the EHR and shipping them together to the external WGS provider.

Stage II: Tumor sample collection (tissue or body fluids containing tumor cells)

Tissue collection
Timing 15-45 minutes

6. Obtain a fresh sample of the primary or metastatic tumor through surgical resection or core needle biopsy.
7. For biopsy retrieval, follow standard of care (SOC) procedures, which usually involve multiple (preferentially at least 4) tumor biopsies, depending on the site and size of metastases and safety considerations.
8. After biopsy retrieval, place the samples in a container on a gauze soaked in PBS. **Note:** Surgical resection samples can be transferred dry (without PBS) immediately after retrieval.
9. Transfer biopsy samples to the pathology laboratory within 10-20 minutes.
10. Place samples in a laboratory refrigerator at 4^0C in the pathology laboratory.

Lessons learned from the implementation process

Immediate feedback on DNA and tumor yield and WGS success rates between pathologists, molecular biologists and radiologists facilitated the detection of factors reducing success rates. Factors included small biopsy sizes, unfavorable histological characteristics (e.g., calcification, fibrosis, necrosis, and mucinous tissue), and technically difficult and high-risk biopsy procedures (e.g., transthoracic and bone biopsies). As a result, radiologists started using ≤18 gauge needles and taking multiple biopsies (preferentially 4) whenever possible of the most viable tumor part, based on radiologic characteristics.

Body fluids containing tumor cells: collection and fine-needle aspirations
Timing 10-60 minutes

- Pleural effusion, pericardial effusion, and ascites are suitable body fluids for obtaining tumor cells for WGS. Collect approximately 10 ml of body fluids according to SOC procedures.
- Follow routine SOC procedures in case of endobronchial ultrasound (EBUS) guided punctures, endoscopic ultrasound (EUS) guided punctures, or fine-needle aspiration (FNA) punctures.
- Transfer the samples to the pathology laboratory immediately after collection.
- Place the samples in the laboratory refrigerator at 4^0C in the pathology laboratory.

PROCEDURE 2: SAMPLE DISTRIBUTION

Timing 5 minutes

11. Collect the samples from the laboratory refrigerator.
12. Determine the clinical question submitted with the samples (e.g., primarily WGS or morphological diagnosis, including immunohistochemistry) to prioritize samples accordingly.

PROCEDURE 3: SAMPLE FRESH FROZEN PROCESSING

Tissue fresh frozen processing
Timing 15 minutes

13. In the case of surgical resection specimens, take multiple 3-mm core punch biopsies from the vital tumor tissue (alternatively, use other methods to obtain similar-sized tissue pieces). Multiple punches (up to 5) can be combined for one frozen section.
14. Collect the biopsies and use one biopsy for each frozen section.
15. Label each biopsy with a unique tissue identification number, preferably barcoded when using a lab management system.
16. Print labels for the corresponding frozen sections and attach these to the coated slides (e.g., Superfrost).
17. Place a thin layer of MCC on the spatula and put a 22-mm lens paper disc into the MCC. To facilitate a flat embedding of the biopsy, it is essential that the paper disc completely absorbs the MCC without overflowing.
18. Place the biopsy on the paper using a clean anatomical tweezer. **Note:** Lightly bend one of the tips of the needle biopsy specimen to readily correlate the shape of the biopsy with the shape of the hematoxylin and eosin (HE) stained section produced in **step 23**. Immediately transfer the paper onto the mold of the PrestoCHILL. **Note:** Asymmetric tissue shapes prevent errors during microdissection.
19. Add MCC to the mold and place 1 corner of the label in the MCC with the written side facing down. Immediately place a frozen section table onto the MCC and close the lid. Freeze at -40^{0}C for 60 seconds.
20. Prepare the cryostat for cutting. In addition, clean the knife block with Leica Cryofect to prevent contamination.
21. Transfer the frozen section table from the PrestoCHILL to the cryostat and orientate the head to obtain a complete section.

22. Trim off the lens paper disc and use this margin to fine-tune the alignment of the head. Loss of tissue should be minimized.
23. Cut a 5-µm frozen section, mount it on coated glass, and stain the section with HE according to standard protocol. Be careful not to wash off thin needle biopsies during the staining procedure. Keep the remaining material in the cryostat or -20°C freezer for further processing (see procedure 5).

Body fluid (ascites, pleural- and pericardial effusion) fresh frozen processing
Timing 20 minutes

- Label glass slides with a patient ID.
- Transfer the respective body fluid into an appropriate centrifuge tube (15 or 50 ml) and label the tube with the patient ID.
- In case of clear fluid: spin down at 3000 rpm for 10 minutes. Make sure to balance the centrifuge by positioning a tube containing the same volume opposite the patient sample in the tube holder. A non-transparent fluid indicates the sample is rich in cells; spinning down is, therefore, not necessary, and the cytospin can be used immediately.
- Discard supernatant to leave a minimum volume of 5 ml and shake the tube until the remaining sediment is dissolved.
- Place the slides in the cytospin together with a cytofunnel.
- Using a sterile Pasteur pipette, add 2-4 drops of sediment to each funnel.
- Spin down the specimens in the cytospin using Program 1 (1250 rpm) for 5 minutes.
- After centrifuging, remove specimens from the holder and let them air dry.
- Stain specimens according to the May-Grünwald Giemsa (MGG) method.
- Add the remaining body fluid into a 50ml tube and store it in a refrigerator at 4^0C.

EBUS guided punctures, EUS guided punctures, and FNA fresh frozen processing
Timing 20 minutes

- Obtain a fine needle aspiration according to routine SOC procedures.
- Fill a small Eppendorf tube (1.5 ml) with a suitable medium (e.g., Minimum Essential Medium (MEM) containing Earle's salts and L-glutamine). Insert the punctured material in the medium and smear part of it onto a glass slide. Stain the slide and assess the types of material and cell richness.
- Ask the treating physician to perform another puncture of the tumor and process the material in the same manner as described above. Number each Eppendorf tube to correspond with the respective glass slides.

- Repeat the previous step multiple times until sufficient material is collected to perform the requested diagnostics. Of note, for WGS, at least 50 ng of tumor DNA is needed.

PROCEDURE 4: ASSESSMENT OF ELIGIBILITY FOR WGS

Stage I: TCP assessment on frozen sections
Timing 5-10 minutes

24. Let a pathologist or cytopathologist assess the tumor cell percentage on the frozen sections or cytology slides to determine the most suitable samples for standard diagnostics (i.e., FFPE) and WGS. TCPs <20% for tissues or <30% for cytology are not suitable for WGS.
25. If necessary (e.g., TCP <20%), let the pathologist mark a tumor area with the highest TCP for manual microdissection.

Stage II: Manual microdissection in case of TCP <20%
Timing 25 minutes

- Assess (if necessary with a pathologist) the number of 20 µm slides that need to be cut.

CRITICAL
Try to cut as much of the biopsy as possible but never more than 50% since SOC diagnostics may be required on the remaining material. Guidelines: 3-5x (cell-rich) to 10x (cell-poor) 20 µm blank slides contain enough DNA for WGS.

- Print labels for each 20-µm slide.
- Cut the 20-µm sections on glass slides, placing as many sections as possible per slide.
- Wash and fixate the sections in pre-cooled FineFix at -20^{0}C during 15-20 minutes.
- Cut a second 5-µm HE slide to check if tumor material is still present in the blank slides. Stain according to standard protocol. Be careful during staining as thin needle biopsies easily wash off. Ask the pathologist to assess the HE slides.
- Prepare all the materials at the workbench (nuclease-free MQ, cooling block for Eppendorf tubes, and scalpel).
- Place a 2-ml screw skirted tube in the cooling block.
- Open the 2-ml screw skirted tube and pipette 210 µl nuclease-free MQ into the lid.
- Place the container with cold FineFix in the flow cabinet.

- Take the glass with 20-µm sections out of the FineFix and let it drain.
- Work as quickly as possible to minimize exposure to room temperature.
- Scratch the relevant area off the glass with the scalpel tip based on the demarcated area on the freeze HE and freeze control HE. Submerge the tip of the scalpel into MQ (4^{0}C) in the lid of the screw skirted tube and wash the fragments off by gently rocking. Be careful to wash all fragments from the scalpel.
- Repeat the previous six steps for the other slides.
- Close the lid of the screw skirted tube. Shake the tube once to force the liquid to the bottom and immediately freeze the samples in liquid nitrogen.
- Store the screw skirted tube at -80^{0}C until shipping to the external WGS provider (within 24 hours; see shipping instructions under Procedure 8).
- After cutting off the blanks, the remaining material can be FFPE processed or stored in a biobank.

Lesson learned from the implementation process

The fresh-frozen fine fixed MCC embedded (4FME) method was developed during the implementation of WGS to maximize the yield of tumor material from biopsy procedures.

PROCEDURE 5: PROCESSING AFTER TCP ASSESSMENT

Timing 2-10 minutes

Processing of tumor material >0.5 cm for DNA isolation at external WGS provider

Timing 10 minutes

26. Prepare all the materials at the workbench (4^{0}C PBS, cooling block, scalpel, anatomical tweezer, and FluidX black and orange tubes).
27. Place a pre-cooled Eppendorf tube containing 1 ml PBS in the cooling block. Remove the frozen section table with the biopsy from the cryostat and put it in the cooling block. Cut a wedge around the biopsy in the MCC with a scalpel, trying to cut as close as possible to the edges of the biopsy. In case a partial biopsy is used, determine the relevant area based on the demarcated HE frozen section.
28. Transfer the cut or punched biopsy fragment or entire biopsy into the cold PBS. Tilt the Eppendorf tube in both directions for 10 seconds to wash off the MCC. Note: Keep the lid closed while tilting the tube.
29. Once the biopsy is clean, transfer it with the sharp end of a tweezer into a FluidX black and orange tube. Place the tube in the WGS container and immediately freeze in nitrogen.

30. The WGS liquid nitrogen container is stored at -80^{0}C until it is shipped to the external WGS provider (within 24 hours; see shipping instructions under Procedure 8).
31. The remaining material can be FFPE processed or stored in a biobank.

Processing of the tumor material <0.5 cm for in-house DNA isolation
Timing 10 minutes

- Prepare all the materials at the workbench (4^{0}C PBS, cooling block, scalpel, anatomical tweezer, and FluidX black and orange tube).
- Place a pre-cooled Eppendorf tube containing PBS in the cooling block. Remove the frozen section table containing the biopsy from the cryostat and place it in the cooling block. Cut a wedge around the biopsy in the MCC with a scalpel, trying to cut as close as possible to the edges of the biopsy. In case a partial biopsy is used, determine the relevant area of selection based on the demarcated HE frozen section.
- Transfer the cut or punched biopsy fragment or entire biopsy into the cold PBS. Tilt the Eppendorf tube in both directions for 10 seconds to wash off the MCC. **Note:** Keep the lid closed while tilting the tube.
- Transfer the biopsy to a 2-ml FluidX black and orange tube using the sharp end of a tweezer.
- Freeze the FluidX tube in a nitrogen container and transport the tube to the -80^{0}C freezer in a small nitrogen container or a container with dry ice.
- The remaining material can be FFPE processed or stored in a biobank.

Biopsies for FFPE and biobank
Timing 5 minutes

- If multiple biopsies from one diagnostic procedure need to be fixated, they should be placed in separate containers. The FFPE HE needs to correspond to the frozen section HE.
- Cut away as much MCC as possible and place the material on formalin; the remaining MCC will be washed off during the process. Perform the FFPE process according to standard procedures.
- The biopsies intended for the biobank can be cut out from the MCC and transferred to frozen Eppendorf tubes. **Note:** Try to exclude as much MCC as possible when cutting out the biopsies.

Serous fluids and cytological punctures for DNA isolation by the external WGS provider
Timing 5 minutes

- Transfer the fluid to a centrifuge tube (15 or 50 ml). Mark the patient ID on the lid of the tube. Make sure to balance the centrifuge by positioning a tube containing the same volume opposite the patient sample in the tube holder. Spin down the tube in a tabletop centrifuge for 10 minutes at 3000 rpm and discard as much supernatant as possible with a pipette.
- Transfer the cell pellet to a 2-ml Eppendorf tube (or Sarstedt tube).
- Place the tube in the WGS liquid nitrogen container.

Lesson learned from the implementation process

Immunohistochemistry (IHC) was validated on tissue samples that had been FFPE processed after initially being frozen. The validation was performed one time for all relevant antibodies. Slight membranous leakage was observed with mild freezing artifacts for the following IHC staining: MIB1, p63, p53, PMS2, MLH1, ER, GATA3, and PDL1. There is insufficient evidence of membranous leaking impacting routine diagnostics negatively. For more sensitive staining, e.g., HER2, it is advised to consider *in situ* hybridization techniques as external controls.

Communication to external WGS provider: WGS analysis is not possible due to insufficient TCP
Timing 2 minutes

- If TCP is insufficient (<20% for tissue and <30% for cytology specimens), the samples are not eligible for WGS analysis at the site of the WGS provider. If a blood sample has already been shipped separately (see shipping instructions for blood under **Procedure 1**), the external WGS provider needs to be contacted via e-mail using only the patient ID. At the site of the WGS provider, a blood sample is temporarily stored for 3 months in case new tissue or cytology is retrieved within that timeframe. After 3 months, the blood sample is destroyed.

PROCEDURE 6: REGISTRATION

Timing 3 minutes

32. Fill out the registration form (an example is provided in **Supplementary 1**).

PROCEDURE 7: IN-HOUSE DNA ISOLATION FROM BLOOD OR TUMOR CELLS

A clinical laboratory can either choose to perform DNA isolation from blood or tumor cells in-house or at the site of the external WGS provider. Further details on DNA isolation from blood are described under "Procedure 9". DNA isolation is performed in-house if complete or partial needle biopsies result in <0,5 cm biopsies since column-based DNA isolation from small tissue samples yields higher DNA concentrations than magnetic-bead DNA extraction.

In-house DNA isolation from tumor cells
Timing 110 minutes

Stage I: Preparation of buffers and aliquots
Timing 40 minutes

- Let all buffers reach room temperature.
- Prepare 0.5 ml 2M DTT. Make aliquots of 15 µl and store them at -20°C. Remnants of the aliquots are not reused.
- Before using FRN buffer for the first time, check whether a precipitate ha frmed. If necessary, dissolve by warming with gentle agitation. Add 42 ml RNase-free isopropanol (96-100%) to the buffer. Note: Use the fume hood while adding the isopropanol.
- In a fume hood, add RNase-free ethanol (96-100%) to the buffers: RPE (44 ml), AW1 (25 ml), and AW2 (30 ml).
- Inject 550 µl RNase-free water into the DNase I container using an RNase-free syringe and needle. Mix by inverting the container a few times. Do not vortex! Make aliquots of 25 µl and store them at -20°C.
- Clean the Tissuelyser LT with RNAaseAway and ethanol (70%).
- Clean the 5-mm beads with RNaseAway and ethanol (70%).
- Clean the tweezers with RNaseAway and ethanol (70%).
- Pre-cool the required beads on dry ice.

Stage II: DNA isolation on frozen material
Timing 70 minutes

- Collect the necessary samples from the -80°C freezer and put them on dry ice.
- Make a master mix of Buffer RTL Plus, Reagent DX (room temperature), and DTT (-20°C) using **Table 1**:

Table 1. Master mix using Buffer RTL Plus, Reagent DX and DTT for DNA isolation

	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10	n=11	n=12
Buffer RLT Plus	645	1290	1935	2580	3225	3870	4515	5160	5805	6450	7095	7740
Reagent DX	3.23	6.46	9.69	12.92	16.15	19.38	22.61	25.84	29.07	32.3	35.53	38.76
DTT	12.9	25.8	38.7	51.6	64.5	77.4	90.3	103.2	116.1	129	141.9	154.8

- Shortly centrifuge the prepared master mix and put it on ice.
- Add one cleaned and cooled bead per sample using clean tweezers.
- For each sample, pipette 600 µl of the Buffer RLT Plus master mix onto the frozen material inside the tube and put it on ice.

CRITICAL
For frozen biopsies (whole or partial), check that all frozen material is loosened from the tube wall. If the material is stuck to the wall, loosen it by gently swinging the tube.

- Homogenize the sample by placing the tube in the Tissuelyser LT and setting it at 30 Hz for 2 minutes. If the material has not dissolved, consider homogenizing for another 2 minutes at 30 Hz. Put the sample on ice immediately afterward.
- Spin down the sample and put it on ice again.
- Pipette the homogenized sample to an AllPrep DNA Mini spin column in an RNAse-free, marked 2-ml tube.
- Centrifuge at 20,000 x g for 30 seconds.
- Place the AllPrep DNA Mini spin column in a new RNAse-free, marked 2-ml tube for DNA isolation.
- Pipette 350 µl Buffer AW1 into the AllPrep DNA Mini spin column. Centrifuge at 20,000 x g for 15 seconds. Discard the filtrate and transfer the spin column to a new 2-ml collection tube.
- Make a master mix of Proteinase K (room temperature) and Buffer AW1 (room temperature) using **Table 2**. Gently mix the master mix by pipetting up and down.

Table 2. Master mix using Buffer AW1 and Proteinase K for DNA isolation

	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10	n=11	n=12
Buffer AW1	65	130	195	260	325	390	455	520	585	650	715	780
Proteinase K	21.7	43.3	65.0	86.7	108.3	130	151.7	173.3	195	216.7	238.3	260

- Centrifuge the master mix shortly.
- Add 80 µl Proteinase K master mix for each sample into the AllPrep DNA Mini column filter.
- Incubate at room temperature for 5 minutes.
- Pipette 350 µl Buffer AW1 (room temperature) onto the AllPrep DNA Mini spin column filter. Centrifuge at 20,000 x g for 15 seconds and discard the filtrate.
- Place the AllPrep DNA Mini spin column in a new 2-ml collection tube.
- Pipette 500 µl Buffer AW2 (room temperature) onto the AllPrep DNA Mini spin column filter. Centrifuge at 20,000 x g for 2 minutes and discard the filtrate.
- Place the AllPrep DNA Mini spin column in a new 1.5-ml collection tube. Mark the side of the collection tube in case the lid is lost during centrifugation.
- Add 30 µl Buffer EB onto the filter of the AllPrep DNA Mini spin column. Close the lid and incubate at room temperature for 5 minutes.
- Centrifuge at 8000 x g for 1 minute. Make sure no liquid remains on the column filter. Centrifuge again if the filter contains liquid.
- Transfer the eluted DNA from the collection tube to a labeled 500-µl tube.
- Measure the DNA concentration, e.g., with the Quantus fluorometer.
- Dilute the DNA to a concentration of 2 ng/µl using nuclease-free MQ.
- Transfer 30 µl of the dilution to a tube (e.g., FluidX).

PROCEDURE 8: SHIPMENT OF THE MATERIAL TO THE EXTERNAL WGS PROVIDER

Timing 20 minutes

33. Ship the material to the external WGS provider (a courier transports the samples to the external site every 24 hours). The frequency of shipment can differ depending on the logistics of a clinical laboratory provided that the tissue is stored in a in a liquid nitrogen container at -80^{o}C. Already isolated DNA can be transported at room temperature and frozen material on dry ice.
34. If minimal material is available, consider shipping sections in an Eppendorf tube instead of a piece of tumor material. Eppendorf tubes tagged with a code (BIXXXX) should be used for the shipment of sections.

PROCEDURE 9: DNA ISOLATION FROM BLOOD AT THE SITE OF THE EXTERNAL WGS PROVIDER

Timing 150 minutes

Stage I: Receiving blood samples for DNA isolation

Timing 5 minutes

35. Check the mail delivery (e.g., a biological bag or red envelope) every morning.
36. Place a Whatman paper on the workbench before working with blood.
37. Confirm that the required information (general information and tube barcodes) is complete on the digital registration form (**Supplementary 1**) and convert it to a CSV file for LIMS import. Personal information and reducible personal information must be destroyed or anonymized.

Stage II: Blood sample registration

Timing 5 minutes

38. Log in to LIMS and import the digital registration form.
39. Scan the barcode on the received blood tube to retrieve the corresponding patient information from LIMS.
40. Validate that the blood tube barcode corresponds to the barcode of the respective patient in the validation field. The barcode turns green when correct and red when incorrect.
41. Once the barcodes have been validated, click 'Received' followed by 'Add to LIMS'. LIMS automatically adds a date to the digital report corresponding to the date the samples were received.

Stage III: QIAsymphony SP: Create an experiment

Timing 5 minutes

42. Click on the 'Blood isolation' button to start a new experiment.
43. Click on 'Compose new' to find all samples waiting for isolation.
44. Click 'Add' to assign the new experiment number (IYY-XXXX; can be retrieved from LIMS) to the samples listed on the screen.
45. Click 'Export'. An export file containing all the necessary information of the composed experiment will appear on your desktop. Use the information to fill out the experiment form and save the document in a folder with the experiment number.
46. Collect the blood samples stored at 4°C or -20°C and let the latter thaw.

47. Prepare new 1-ml barcoded tubes and register the tube codes directly on the digital registration form.

Stage IV: QIAsymphony SP: DNA isolation
Timing 120 minutes

48. Switch on the QIAsymphony SP and log on to the instrument if necessary.
49. Start the UV light (duration 15 minutes) and press 'OK'.
50. Press 'OK' to continue once the UV-light procedure is completed.
51. Click on 'Tools' followed by 'Sample preparation' and wait until the initialization procedure is complete.
52. Click on the fork spanner/hammer icon.
53. Click 'Done' to complete the maintenance and press 'OK'.
54. Open the 'eluate' drawer of the QIAsymphony SP and select slot 1 on the screen:
    - Place a rack with scanned barcoded tubes on the cooling adaptor and load the adaptor into the 'eluate' drawer of the QIAsymphony SP. Ensure the caps from the barcoded tubes are removed before loading the rack into the QIAsymphony SP. Verify that the cooling slot is on:
      - Press 'Deep well' and select FL_TubeRack #68-1002-10*T1.0.
      - Press 'Rack ID' and create a rack ID (IYY-XXXX) and press 'OK'.
      - Close the 'eluate' drawer and press 'OK'.
55. Open the 'waste' drawer and make sure half of the waste drawer is empty. Replace the tip disposal bin if full, and press 'Scan'.
56. Open the 'reagent' and 'consumables' drawers.
57. Prepare one or several DSP Midi kit reagent cartridges:
    - Vigorously vortex the trough containing magnetic beads until fully resuspended.
    - Check that the buffers QSL1 and QSB1 from the QIAsymphony DSP DNA kit are clear and do not contain salt flakes (precipitates). If salt flakes are present, remove the buffers from the reagent cartridge, incubate at 37°C, and shake to dissolve the precipitates.
    - Place the cartridge in the appropriate holder. Ensure the piercing lid is placed on the reagent cartridge and the lid of the magnetic-bead trough has been removed. If using a previously used reagent cartridge, remove the reuse seal strips before placing the cartridges in the holder. Open the enzyme tubes.
58. Load to the Qiasymphony SP with:
    - The prepared reagent cartridge
    - Tips
    - 8-rod covers
    - 8-well cartridges

59. Close the 'reagent'' and 'consumable' drawers and select 'Scan'.
60. Mix by vortexing the sample tubes shortly.
61. Validate the tube barcodes on the DNA isolation registration form.
62. Carefully remove the tube caps and place the blood-sample tubes into the appropriate tube carrier in the 'sample' drawer, orientating the tubes so the barcodes face the barcode reader on the left side of the QIAsymphony SP.

CRITICAL

- Transfer 1 ml blood into a 2-ml Sarstedt tube if the sample volume in the Streck tube contains less than ~2 ml. Place a Sarstedt holder in the tube carrier.
- Once blood samples are defrosted, transfer 1 ml blood into a 2-ml Sarstedt tube. Place a Sarstedt holder in the tube carrier.

63. Open the 'sample' drawer.
64. Load the tube carrier into the 'sample' drawer (position 1, 2, 3, or 4).
    - Change the positions of the tube carrier in the drawer between each use to minimize wear and tear.
65. Select 'SP batch' on the QIAsymphony SP:
    - Check the sample ID. If needed, change the incorrect sample IDs: select the incorrect sample ID, press 'Sample ID', add the correct ID, and click 'OK'. A representation of a hand appears next to the sample ID.
66. Select 'NEXT'.
67. Select all samples and couple them to the isolation protocol 'Blood_1000_ custom' under 'Custom protocols' (when the sample and isolation protocol are coupled, a hand appears on the screen).
68. Select 'NEXT'.
69. Select slot 1. Elution volume is by default 200 µl.
70. Click 'QUEUED'.
71. If proceeding with more than one tube carrier (i.e., more than 24 samples), repeat the procedure from step 60 for the following tube carriers.
72. Press the 'Run' button to start the isolation procedure.
73. After isolation, remove the elution rack with purified DNA from the 'eluate' drawer.
    - Click 'Remove' followed by 'Yes' when the 'Do you want to remove the rack' pop-up appears. Close the 'eluate' drawer and press 'OK'.

74. If the reagent cartridge is partially used, seal it with the provided reuse seal strips and close the enzyme tubes with screw caps immediately after the end of the protocol run to avoid evaporation.
75. Remove full waste containers and the tube carrier.

Stage V: Quality control of DNA isolation
Timing 10 minutes

76. Measure the DNA concentration of the samples from **step 73** using the Qubit dsDNA BR assay kit (broad range).
77. Fill in the DNA concentration (ng/µl) in the corresponding digital registration form. After isolation, the final DNA concentration should be more than 2 ng/µl. Repeat the isolation step (starting at **step 72**) an additional time if the quality values are out of range of the kit or DNA concentration is less than 2 ng/µl. Contact the hospital and request new blood if isolation fails twice.

Stage VI: QIAsymphony SP: Finishing the DNA isolation experiment
Timing 5 minutes

78. Place the tube carriers with blood samples inside a storage box. Note the tube position, corresponding blood barcode, and sample name. If the storage box is full, perform a rack scan to confirm the sample position. Store the barcoded blood tubes at -20°C.
79. To finish an experiment, log in to LIMS.
80. Click 'Blood isolation'.
81. Select the corresponding isolation number in the column to the left of the screen (it turns black when selected).
82. Click on 'Import file' and upload sample barcodes and DNA concentrations into the registration form. Confirm that all barcodes and DNA concentrations are assigned to the respective sample IDs.
83. LIMS completes the experiment upon confirmation.

PROCEDURE 10: DNA ISOLATION FROM TUMOR CELLS AT EXTERNAL WGS PROVIDER

Timing 295 minutes

Stage I: QIAsymphony SP: Create an experiment

Time 5 minutes

84. Log in to LIMS and click on 'Tissue isolation'.
85. Click 'Compose new' to find all samples waiting for isolation.
86. Click 'Add' to assign the new experiment number (IYY-XXXX; can be retrieved from LIMS) to the samples listed on the screen.
87. Click 'Export'. An export file containing all the necessary information of the composed experiment appears on your desktop. Use the information to fill out the experiment form and save the document in a folder with the experiment number.
88. Assign barcodes to the respective samples and register the barcodes in the digital registration form.

Stage II: Receiving tumor samples for DNA isolation

Timing 5 minutes

89. When receiving or handling tumor material, always keep the samples on dry ice or at -80°C to prevent samples from thawing and degrading.
90. Upon arrival, confirm that the box containing the samples is intact and the corresponding information is correct.
91. Confirm that the digital registration form includes all the required information (general information and tube barcodes) and convert it into a CSV file for import. Personal information and reducible personal information must be destroyed or anonymized.

Stage III: Tissue sample registration

Timing 5 minutes

92. Log in to LIMS and import the digital registration form.
93. Scan the barcode of the received tissue tube. LIMS highlights the corresponding patient ID.
94. Validate the tissue tube in the validation field. The barcode turns green when correct and red when incorrect.
95. Click 'Received' followed by 'Add to LIMS'. LIMS automatically adds a date to the digital report corresponding to the date the samples were received.

Stage IV: Preparation of DNA isolation buffer
Timing 10 minutes

96. Print stickers with the corresponding BI numbers (see **Figure 2**) and barcodes and attach them to 2-ml Sarstedt tubes.

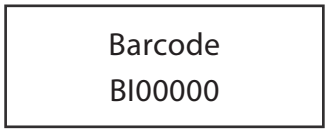


Figure 2: Sticker with barcode and BI number

97. Print -80°C stickers with the corresponding BI numbers and barcodes and attach them to separate 2-ml Sarstedt tubes, and add a 5-mm Stainless steel bead in the tube.
98. Keep the prepared -80°C Sarstedt tube in a -20°C cooler or on dry ice.
99. Prepare the DNA-isolation buffer in the fume hood following **table 3**:

Table 3. DNA isolation buffer (170 µl/sample). The buffer for 8 samples is calculated with an excess of 1.

# samples	1	8
Buffer ATL µl	154.5	1390.5
Proteinase K µl	15.5	139.5
Total µl:	170	1530

100. Invert the buffers several times to make sure they are thoroughly mixed. **Do not vortex the tubes!**
- Check the ATL buffer for precipitates. If necessary, warm the buffer while swirling under a lukewarm tap to dissolve precipitates.

101. Add the prepared DNA isolation buffer (170 µl; see **step 99**) into the barcoded 2-ml Sarstedt tubes.

Stage V: Sample preparation for DNA isolation
Timing 180 minutes

If there are multiple tumor samples per patient, only one is processed at a time. The second sample is stored for later processing. If the first sample fails, the second can be used.

CRITICAL
It is essential to work quickly to keep the samples as cool as possible.

102. Collect the samples from the shipment box or -80°C (see step 89).
103. The samples are prepared in batches (a maximum of 12 samples per batch) to prevent them from becoming too warm.
104. Keep the samples in a -20°C cooler or on dry ice before starting the process.

Processing whole biopsies

105. Verify the barcodes of the tissue tubes on the digital registration form (see **step 91**).
106. Transfer the complete biopsy into the barcoded 2-ml Sarstedt tube with stainless steel beads.
107. Add 210 µl DNAse/RNAse-free water (4°C) to the sample tubes.
108. Keep the samples in a 4°C cooler.

CRITICAL
If biopsies are tiny, use a broad tip for pipetting them up in the FluidX tissue tubes (**step 29**) with 210 µl DNAse/RNAse-free water. Transfer the water with the respective biopsy into the barcoded 2-ml Sarstedt tube. If biopsies are too large for pipetting with broad tips, use a scalpel, tweezer, and petri dish on dry ice to cut out a piece (2-3 mm^2) of the most likely tumor part (assessed macroscopically). When in doubt, isolate 2 parts from a single biopsy.

Processing sections

- Sections are delivered in barcoded 2-ml Sarstedt tubes containing a stainless steel bead.
- Verify that the 2-ml Sarstedt tube barcodes correspond to the patient ID on the digital registration form.
- Add 110 µl DNAse/RNAse-free water (4°C) to the coupes or sections.
- Keep the samples in a 4°C cooler.

Processing cell pellets

- Cell pellets are delivered in Eppendorf tubes.
- Verify that the tube barcodes correspond to patient IDs on the digital registration form.
- Add 110 µl DNAse/RNAse-free water(4°C) to the cell pellets, mix thoroughly by pipetting, and, using a pipette, transfer the total volumes to the barcoded Sarstedt tubes containing a stainless steel bead.
- Keep the samples in a 4°C cooler until further processing (see **step 108**).

109. Load the samples into the tube adaptors of the TissueLyser.
110. Ensure the TissueLyser's adaptor sets are balanced by evenly distributing the reaction tubes across the tube adaptors.
111. Place the tube adaptors in the TissueLyser and start Program 1 (2min at 25Hz).
112. Quick spin the tubes immediately after the TissueLyser has completed the program and place the tubes in a 0°C cooler.
113. Transfer the 0°C cooler containing the tubes with lysed products to the fume hood.
114. Transfer about 50 µl of the lysed product to the DNA isolation buffer tubes (**step 101**) and mix by pipetting up and down.
115. Vortex the samples for DNA isolation and incubate them at 56°C for 2 hours to lyse the cells completely.
- If further processing is not desired, store lysates in ATL buffer for 2 hours at 4°C.
116. After completing the lysis step, spin down the samples for DNA isolation.

Stage VI: QIAsymphony SP: DNA isolation
Timing 75 minutes

117. Switch on the QIAsymphony SP, and log on to the instrument if necessary.
118. Start the UV light (duration 15 minutes) and press 'OK'.
119. Press 'OK' to continue once the UV-light procedure is completed.
120. Click on 'Tools' followed by 'Sample preparation' and wait until the initialization procedure is complete.
121. Click on the fork spanner/hammer icon.
122. Click 'Done' to complete the maintenance and press 'OK'.
123. Open the 'eluate' drawer of the QIAsymphony SP and select slot 1.
- Place a scanned rack with barcoded tubes on the cooling adaptor and load the adaptor into the 'eluate' drawer of the QIAsymphony SP. Ensure the caps from the barcoded tubes are removed before loading the rack into the QIAsymphony SP. Verify that the cooling slot is on:
  - Press 'Deep well' and select FL_TubeRack #68-1002-10*T1.0.
- Press 'Rack ID', make a rack ID (IYY-XXXX), and press 'OK'.
- Close the 'eluate' drawer and press 'OK'.
124. Open the 'waste' drawer and make sure half of the waste drawer is empty. Replace the tip disposal bin if full and press 'Scan'.
125. Open the 'reagent' and 'consumables' drawers.
126. Prepare one or several DSP Mini kit reagent cartridges.
- Vigorously vortex the trough containing magnetic beads until fully resuspended.
- Place the cartridge in the appropriate holder. Ensure the piercing lid is placed on the reagent cartridge and the lid of the magnetic beads trough has been

removed. If using a previously used reagent cartridge, remove the reuse seal strips before placing the cartridges in the holder. Open the enzyme tubes.

127. Load the QIAsymphony SP with:
   - The prepared reagent cartridge
   - Tips
   - 8-rod covers
   - 8-well cartridges
128. Close the 'reagent' and 'consumable' drawers and select scan.
129. Open the sample drawer.
130. Validate the barcodes on the DNA isolation registration form.
131. Load the tube carrier into the 'sample' drawer (position 1, 2, 3, or 4).
   - Change the positions of the tube carrier in the drawer between each use to minimize wear and tear.
132. Select 'SP batch' on the QIAsymphony SP.
   - Check the box 'SAR-FIX_#72.694 T2.0 ScrewSkirt'.
133. Check the sample ID. If needed, change the incorrect sample IDs: select the incorrect sample ID, press 'Sample ID', add the correct ID, and press 'OK'). A representation of a hand appears next to the sample ID.
134. Select 'NEXT'.
135. Select all samples and couple them onto the isolation protocol 'Tissue_LC_200' under 'DNA tissue' (when sample and isolation protocol are coupled, a hand appears on the screen).
136. Select 'NEXT'.
137. Select slot 1. Elution volume is by default 50 µl.
138. Press 'QUEUED'.
139. If proceeding more than one tube carrier (i.e., more than 24 samples), repeat the procedure from step 129 for the following tube carrier(s).
140. Click 'Run' to start the purification procedure.
141. Pop-up appears to confirm the start of the procedure using FluidX tubes and 2 ml Sarstedt tubes, select 'Yes' if correct.
142. After isolation, remove the elution rack containing the purified DNA from the 'eluate' drawer.
143. Click 'Remove' followed by 'Yes' when the 'Do you want to remove the rack' pop-up appears. Close the 'eluate' drawer and press 'OK'. If a reagent cartridge is only partially used, seal it with the provided reuse seal strips and close the enzyme tubes with screw caps immediately after the end of the protocol run to avoid evaporation.
144. Discard the used sample tubes in the disposal tip bin.

Stage VII: Quality control of DNA isolation
Timing 10 minutes

145. Measure the DNA concentration of the samples from step 142 using the Qubit dsDNA BR assay kit (broad range).
146. Fill in the DNA concentration (ng/µl) in the digital registration form. After isolation, the final DNA concentration should be more than 1 ng/µl. Repeat the isolation step (see step 139) an additional time if the DNA concentration is less than 1 ng/µl. Ensure to report a Quality Control isolation fail on the WGS report if the concentration remains less than 1 ng/µl after repeating the isolation step and a spare biopsy is unavailable.
147. A technician may destroy the DNA if less than 1 ng/µl DNA isolate is available.

Stage VIII: QIAsymphony SP: Finishing the DNA isolation experiment
Timing 5 minutes

148. Place the tube carriers with tissue samples in a storage box. Note the position, corresponding tissue barcode, and sample name. If the storage box is full, perform a rackscan to confirm the sample positions.
149. Store the TissueLysed rest material (+/- 100 µl from the barcoded Sarstedt tubes) at -80°C in an appropriate storage box.
150. To finish the experiment, log in to LIMS.
151. Click 'Tissue isolation'.
152. Select the corresponding isolation number in the column to the left of the screen (it turns black when selected).
153. Click 'Import DNA + finish' in LIMS and upload to the digital registration form. Confirm that all the DNA barcodes and concentrations are assigned to sample IDs.
154. LIMS completes the experiment upon confirmation.

PROCEDURE 11: WGS ANALYSIS AT THE EXTERNAL WGS PROVIDER

Timing 4 working days

The sample prep is done using standard Illumina® TruSeq® Nano DNA LT Library Prep kit and IDT 384 UDI UMI indexes. The DNA (50-200ng) is fragmented using a Covaris shearing device (median 550bp) and quantification is performed with a Qubit dsDNA HS Assay Kit and KAPA Library Quantification Kit. The sample prep is diluted to 3 nM before pooling and sequencing. Sequencing is performed on Illumina NovaSeq6000 platforms with a sequencing depth of more than 90x for tumor DNA,

according to standard procedures described previously[5-7]. DNA isolated from blood is sequenced at an average depth of more than 30x. Sequencing data is analyzed with an optimized, fully open source in-house bioinformatic pipeline (code available through github.com/hartwigmedical).

TIMING

PROCEDURE 1: 15-70 minutes
- Stage I: Blood collection: 5-10 minutes
- Stage II: Tissue of fluid containing tumor cells collection: 10-60 minutes

PROCEDURE 2: 5 minutes

PROCEDURE 3: 15-20 minutes
- Tissue fresh frozen processing: 15 minutes
- Serous fluids fresh frozen processing: 20 minutes
- EBUS guided punctures, EUS guided punctures, and FNA fresh frozen processing: 20 minutes

PROCEDURE 4: 5-35 minutes
- Stage I: TCP assessment on frozen sections: 5-10 minutes
- Stage II: Manual microdissection in case of TCP <20%: 25 minutes

PROCEDURE 5: 2-10 minutes
- Processing of tumor material >0.5 cm for DNA isolation at external WGS provider: 10 minutes
- Processing tumor material <0.5 cm for in-house DNA isolation: 10 minutes
- Biopsies for FFPE and biobank: 5 minutes
- Serous fluids and cytological punctures for DNA isolation by the external WGS provider: 5 minutes
- Communication to external WGS provider: WGS analysis not possible due to insufficient TCP: 2 minutes

PROCEDURE 6: 3 minutes

PROCEDURE 7: 110 minutes
- Stage I: Preparation of buffers and aliquots: 40 minutes
- Stage II: DNA isolation on frozen material: 70 minutes

PROCEDURE 8: 20 minutes

PROCEDURE 9: 150 minutes
- Stage I: Receiving blood samples for DNA isolation: 5 minutes
- Stage II: Blood sample registration: 5 minutes
- Stage III: QIAsymphony SP: Create an experiment: 5 minutes
- Stage IV: QIAsymphony SP: DNA isolation: 120 minutes
- Stage V: Quality control of DNA isolation: 10 minutes
- Stage VI: QIAsymphony SP: Finishing the DNA isolation experiment: 5 minutes

PROCEDURE 10: 295 minutes
- Stage I: QIAsymphony SP: Create an experiment: 5 minutes
- Stage II: Receiving tumor samples for DNA isolation: 5 minutes
- Stage III: Tissue sample registration: 5 minutes
- Stage IV: Preparation of DNA isolation buffer: 10 minutes
- Stage V: Sample preparation for DNA isolation: 180 minutes
- Stage VI: QIAsymphony SP: DNA isolation: 75 minutes
- Stage VII: Quality control of DNA isolation: 10 minutes
- Stage VIII: QIAsymphony SP: Finishing the DNA isolation experiment: 5 minutes

PROCEDURE 11: 4 working days

Anticipated results

Reporting wgs results

The external WGS provider produces a detailed report (OncoAct, Hartwig, Amsterdam, the Netherlands) containing all variants with a high driver likelihood[6] relevant for cancer diagnostics and treatment decisions. Reports may include mutations, gene amplifications, or gene losses in 460 genes, 63 promiscuous fusion partners and 402 known oncogenic fusions, mutational signatures (TMB, HRD, and MSI), viral insertions (Human papillomavirus, Epstein-Barr virus, and Merkel cell polyomavirus) and DPD and UGT1A1 status. Actionable biomarkers are linked to existing biomarker-based clinical trials within Dutch hospitals using the iClusion database, in which ongoing clinical trials (phase I-III) in the Netherlands and their eligibility criteria are stored. An example of a WGS reports can be found in **Supplementary 2**.

Sequencing interpretation by clinical dna specialists

The external WGS provider uploads WGS reports for download on NextCloud (Google Cloud Platform) every 3 days. In addition, raw and analyzed data (i.e. variant calls) are made available. A clinical DNA specialist interprets the newly available WGS reports daily and subsequently composes a pathology report. Once the WGS reports are downloaded, clinical DNA specialists validate patient and tumor information (e.g., patient ID, sample ID, germline findings decision, and tumor type). Germline findings are excluded from the WGS report if the patient opted out of PGV reporting. If the information on the WGS report is incorrect, clinical DNA specialists contact the external WGS provider requesting a revised WGS report. Curation of the patient and tumor type information by clinical DNA specialists is essential for DNA annotation of the sequencing algorithm and future research. Curated data can be used optimally to promote a self-learning healthcare system.

Germline findings

If a PGV is reported on the WGS report, a clinical DNA specialist contacts the laboratory specialist in clinical genetics to confirm that the PGV is correctly classified and check whether the patient is a known PGV carrier. The clinical genetics laboratory specialist notifies the clinical geneticist of the variant, classification, and patient ID. Moreover, the clinical DNA specialist includes a standard phrase in the pathology report: '*The xxx-mutation has also been identified in the germline DNA. Consider referring the patient to a clinical geneticist*'. The clinical geneticist participates in the WGS dedicated board when discussing the implications of the PGV for the patients. The clinical geneticist aids in reporting the PGV and consults the treating physician

if necessary. All communication between clinical DNA specialists, the laboratory specialist in clinical genetics, and the clinical geneticist must be documented.

Molecular tumor cell percentages below 20%

A clinical DNA specialist checks the WGS report for molecular TCP (tumor purity as determined by WGS data) and the pathology report for the clinical request of the treating physician. If the molecular TCP is below 20% and does not meet the requirements for high-quality WGS, and there are suspicions of potentially missed mutations, the clinical DNA specialist requests the remaining DNA from the external WGS provider to perform SOC NGS. Considering that NGS requires a TCP above 10%, high-quality NGS is often possible using the remaining DNA.

A biopsy taken in the context of possible tyrosine kinase inhibitor (TKI) resistance with an mTCP below 20% may result in no new resistance mechanisms reported by WGS. Such cases may indicate that variants with low variant allele frequencies (VAFs) have not been reported. However, the raw data may contain these variants, e.g., T790M or C797S, in case of a resistant lung cancer biopsy. To retrieve this information, a clinical DNA specialist must contact the external WGS provider, which confirms possible variants present in the raw data within 1.5 hours.

Occasionally, WGS detects an unknown fusion with low driver likelihood, but potential clinical relevance or RNA expression of the fusion needs to be confirmed. In such cases, the clinical DNA specialist requests RNA corresponding to the tissue or blood from the external WGS provider and performs an Archer, an NGS fusion analysis that covers the unknown fusion partner.

Multidisciplinary clinical decision making

WGS results from all patients are discussed in a weekly WGS dedicated board consisting of clinical DNA specialists, pathologists, clinical geneticists, medical oncologists, and phase I clinicians. During multidisciplinary board meetings, the WGS results, including quality measures, germline, and CUPPA findings, are interpreted and translated into treatment recommendations and registered in the electronic patient record. Upon request by the treating physician (or delegated person), the WGS results and treatment recommendations can be discussed in the molecular tumor board (MTB) to ensure patient-centered decision-making. The MTB consists of clinicians from different medical divisions, pathologists, clinical DNA specialists, and clinical geneticists (if a PGV is identified).

Communication of results

According to standard reporting procedures, WGS results are included in the pathology report and electronic patient record (Hix, Chipsoft, Amsterdam, The Netherlands) and the Dutch national pathology digital archive (PALGA)[8]. The treating physician is responsible for communicating the WGS results to the patient.

Data usage

Upon patient consent, WGS data is stored in a centralized database of Hartwig. This database (currently storing more than 5000 patients) contains pseudonymized genetic and clinical data, including treatment and treatment outcomes according to Response Evaluation Criteria In Solid Tumors (RECIST) 1.1, of individuals whose tumors have been sequenced in the context of multiple studies or clinical care (https://catalog.hartwigmedicalfoundation.nl/). This database is accessible to international researchers through an access-controlled mechanism (https://www.hartwigmedicalfoundation.nl/data-aanvragen/). At the NKI, upon patient consent, pseudonymized genetic and clinical data is stored at a local CbioPortal. This database is accessible to internal and external researchers upon request and is reviewed by an independent reviewing board based on scientific merit. It is crucial to combine comprehensive genomic data with detailed clinical phenotyping and store the genomic data centrally to provide a solid basis for a learning healthcare system.

References

1 Lynch, J. A. *et al.* Underutilization and disparities in access to EGFR testing among Medicare patients with lung cancer from 2010 - 2013. *BMC Cancer* **18**, 306, doi:10.1186/s12885-018-4190-3 (2018).

2 Samsom, K. *et al.* Feasibility of whole-genome sequencing in routine clinical practice. *Journal of Clinical Oncology* **39**, 3013-3013, doi:10.1200/JCO.2021.39.15_suppl.3013 (2021).

3 Schipper LJ, S. P., Samsom KG, Bosch LJW, Lalezari F, Priestley P, Shale C, Jacobs N, van den Broek AJ, Roepman P, van der Hoeven JJM, Steeghs N, Cuppen E, Meijer G, Voest EE and Monkhorst K. Whole genome sequencing can classify diagnostically challenging tumors. *Annals of Oncology* **32** S921-S930., doi:10.1016/annonc/annonc707 (2021).

4 Orchard, G. E. *et al.* New embedding and staining systems PrestoCHILL and Presto stainer for application in the advancement of Mohs micrographic surgery. *Br J Biomed Sci* **74**, 203-208, doi:10.1080/09674845.2017.1348566 (2017).

5 Roepman, P. *et al.* Clinical Validation of Whole Genome Sequencing for Cancer Diagnostics. *J Mol Diagn* **23**, 816-833, doi:10.1016/j.jmoldx.2021.04.011 (2021).

6 Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210-216, doi:10.1038/s41586-019-1689-y (2019).

7 Samsom, K. G. *et al.* Study protocol: Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient (WIDE). *BMC Med Genomics* **13**, 169, doi:10.1186/s12920-020-00814-w (2020).

8 Casparie, M. *et al.* Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* **29**, 19-24, doi:10.1155/2007/971816 (2007).

CHAPTER 4

Feasibility of whole genome sequencing based tumor diagnostics in routine pathology practice

Kris G. Samsom*, Luuk J. Schipper*, Paul Roepman, Linda J.W. Bosch, Ferry Lalezari, Elisabeth G. Klompenhouwer, Adrianus J. de Langen, Tineke E. Buffart, Immy Riethorst, Lieke Schoenmaker, Daoin Schout, Vincent van der Noort, José G. van den Berg, Ewart de Bruijn, Jacobus J.M. van der Hoeven, Hans van Snellenberg, Lizet E. van der Kolk, Edwin Cuppen, Emile E. Voest, Gerrit A. Meijer** & Kim Monkhorst**

* *Equally contributing first authors, **Equally contributing last authors*

Journal of pathology, 2022

Abstract

The current increase in number and diversity of targeted anti-cancer agents poses challenges to the logistics and timeliness of molecular diagnostics (MolDx), resulting in underdiagnosis and treatment. Whole genome sequencing (WGS) may provide a sustainable solution for addressing current as well as future diagnostic challenges. The present study therefore aimed to prospectively assess feasibility, validity, and value of WGS in routine clinical practice. WGS was conducted independently of, and in parallel with standard of care (SOC) diagnostics on routinely obtained tumor samples from 1,200 consecutive patients with metastatic cancer. Results from both tests were compared and discussed in a dedicated tumor board. From 1,200 patients, 1,302 samples were obtained, of which 1,216 contained tumor cells. WGS was successful in 70% (854/1216) of samples with a median turnaround time of 11 days. Low tumor purity (<20%) was the main reason for not completing WGS. WGS identified 99.2% and SOC MolDx 99.7% of the total of 896 biomarkers found in genomic regions covered by both tests. Actionable biomarkers were found in 603/848 patients (71%). Of the 936 associated therapy options identified by WGS, 343 were identified with SOC MolDx (36.6%). Biomarker-based therapy was started in 147 patients. WGS revealed 49 not previously identified pathogenic germline variants. Fresh-frozen, instead of formalin-fixed and paraffin embedded, sample logistics were easily adopted as experienced by the professionals involved. WGS for patients with metastatic cancer is well feasible in routine clinical practice, successfully yielding comprehensive genomic profiling for the vast majority of patients.

Introduction

With the rapidly expanding tableau of (increasingly tumor-agnostic) targeted therapies (1-5), genome-driven cancer care has become the cornerstone of modern precision oncology (6). However, pathology laboratories are facing increasing challenges in keeping up with the speed at which targeted drugs and their associated biomarkers are entering clinical oncology practice. These challenges exist at multiple levels. First, the rapidly expanding druggable genome requires pathology laboratories to continuously update and validate their molecular diagnostic (MolDx) arsenal to cover the latest actionable genomic alterations. In practice, this inevitably causes substantial delays in clinical implementation of newly approved biomarkers (7), and contributes to inequality of clinical care (8). Second, indications for MolDx are still largely tumor type dependent (9). This leads to a multitude of complex and often sequential tumor type specific diagnostic routings that are error prone and easily outdated (10). This impedes effective MolDx for identifying rare targets in common cancers, as well as identifying therapeutic targets in less common cancers (11). Third, the interplay between somatic mutations and (possible) germline DNA alterations is becoming increasingly important in targeted therapies, e.g. carriers of germline BReast CAncer (*BRCA*) mutations without biallelic loss of function do not respond to Poly(ADP-Ribose) Polymerase (PARP) inhibitors (12, 13) and Whole Genome Sequencing (WGS) potentially offers insight in both tumor and germline. Fourth, there is an increasing demand for more complex biomarkers like signatures for homologous repair deficiency (HRD), microsatellite instability (MSI) and tumor mutational load, which increasingly guide therapeutic decisions and can act as valid proxies of epigenetic inactivation of druggable pathways. (1, 14-16). Fifth, combined large data sets on comprehensive genomic characterization, therapeutic interventions and patient outcome improves decision support in precision oncology.

Hence, there is a growing need for a future proof, tumor-type independent, comprehensive MolDx approach for all (metastatic) cancer patients, and the current fragmented and reactive MolDx approach does not meet these standards. Although most of these challenges could partly be covered by implementing large NGS panels, WGS has considerable advantages over panel sequencing. First, large NGS panels still have to be updated every couple of years. Second, since WGS is a stable test, the generated data will always be comparable in time and place and allows algorithms like Cancer of Unknown Primary Prediction Algorithm (CUPPA) and Cancer of Unknown Primary Location Resolver (CUPLR) to be implemented and improved over time (17, 18). Third, because of its completeness, it allows for better retrospective analysis in self learning healthcare systems and research endeavors. Furthermore, no biomarker

will be forgotten to be tested, a challenge especially relevant when the diagnosis is uncertain. WGS can successfully address the above mentioned challenges and would therefore offer an attractive solution (19-21), but its feasibility in routine pathology practice remains to be proven (22). WGS has proven its feasibility in multiple pediatric centers, however, the setting of pediatric oncology centers differs substantially in volume and scale compared to adult oncology centers (21, 23, 24). Historically, tissue handling in diagnostic pathology is based on formalin fixation and paraffin embedding (FFPE). While targeted gene panel-based diagnostics well with FFPE material, WGS ideally requires fresh frozen material to avoid FFPE induced sequencing artefacts and obtain genome-wide accurate variant calls, including comprehensive copy number and structural variant calling. Implementation of WGS in routine clinical practice therefore requires collecting and working with fresh frozen samples in routine pathology workflows. The WIDE study (WGS Implementation in standard Diagnostics for Each cancer patient) therefore aimed to prospectively generate evidence on the feasibility and clinical validity (primary endpoints), as well as clinical value of WGS (secondary endpoint) in routine clinical practice for patients with metastatic cancer (25).

Methods

Study design, setting and population

WIDE is a single center prospective, observational, diagnostic study in patients with (suspected) stage IV solid tumors of all occurring tumor types, approved by the Medical Ethical Committee of the Netherlands Cancer Institute (NKI) (NL68609.031.18) and conducted in concordance with the Declaration of Helsinki, Dutch law, and Good Clinical Practice. All patients provided written informed consent. Patients were eligible when biopsy, resection or suitable cytology (e.g. pleural effusion or ascites) samples could be obtained safely as part of routine diagnostic procedures. Patients, from whom archival fresh frozen tumor samples were available, were also eligible if not treated in-between with tyrosine kinase inhibitors, since this could have shifted the genomic profile by clonal selection (26). WGS was performed at Hartwig Medical Foundation (Amsterdam, The Netherlands; hereafter referred to as Hartwig) in parallel with and independently of standard of care (SOC) diagnostics. Depending on patient preference declared in the informed consent, pathogenic (class 4 and 5 (27)) germline variants in genes with targeted tumor therapy implications were either reported as inherited variants, along with an offer for routine clinical genetics counselling, or as variants present in the tumor sample without reporting germline status (supplementary material, Table S1). Germline variants without cancer related actionability were not investigated nor reported.

Sample collection and processing

SOC procedures aimed to collect 2–4 biopsies. A 10 ml whole blood sample was drawn for sequencing germline DNA as a reference, allowing us to discriminate somatic mutations from germline DNA background variations in bioinformatic analyses. The macroscopically best biopsy was prioritized for SOC diagnostics, which, depending on the clinical question at hand, either did or did not include MolDx. This biopsy then was FFPE processed. For the SOC MolDx portfolio at the NKI see supplementary materials and methods, Tables S2-S4. Next, WGS was performed in parallel with and independently of SOC diagnostics. To this end, specimens were cryoembedded using a PrestoCHILL (Milestone Medical, Kalamazoo, MI, USA) device for 60s at -40 °C. Subsequently, the specimens were cut into frozen sections of 5 µm thickness on coated glass then stained with hematoxylin and eosin.

Next, a pathologist microscopically assessed on frozen sections of every sample a tumor cell percentage (pTCP), and when needed demarcated a tumor area for manual microdissection. For tissue specimens, a pTCP of ≥20% and for cytology specimens ≥30% was required. Preferably within 24 h both tumor and blood samples were shipped by courier to Hartwig for WGS. Any remaining (frozen) tissue was processed to FFPE blocks. The NKI department of pathology operates under ISO15189:2012 accreditation.

WGS and bioinformatics

WGS was performed at Hartwig on Illumina NovaSeq6000 platforms (Illumina, San Diego, CA, USA) with a sequencing depth of >90*x* for tumor DNA according to standard procedures as described previously (20, 25). DNA isolated from blood was sequenced at an average depth of >30*x*. Sequencing data were analyzed with an optimized fully open source in-house bioinformatic pipeline (28) (code available through github.com/hartwigmedical, see further details in Supplementary materials and methods. Hartwig operates under ISO17025:2015 and ISO/NEN27001 accreditation.

Reporting

The WGS report (OncoAct, Amsterdam, the Netherlands) contained all variants with a high driver likelihood (20) relevant for diagnostic purposes and cancer treatment decision making, further referred to as biomarkers. These encompassed mutations and amplifications/losses in 460 genes, 63 promiscuous fusion partners and 402 known oncogenic fusions, mutational signatures (tumor mutational load, HRD and MSI) and viral insertions (Human Papilloma virus, Epstein-Barr virus, and Merkel cell polyomavirus). Although each tumor sample was analyzed for its whole genome characteristics, including all genes (exons and introns) and intergenic regions, the

WGS report is limited to variants with high driver likelihood in order to provide clinically manageable reports for the treating physicians. WGS, as well as SOC MolDx results when applicable, were discussed in a weekly dedicated molecular tumor board. In case of any discrepancies, additional verification tests were performed on the original input samples used for WGS and SOC according to a predefined workflow (supplementary material, Figure S1). Results of both WGS and SOC MolDx were communicated via routine pathology reporting to the treating physician.

Continuous evaluation and improvement

The design of the study allowed for continuous evaluation and improvement of procedures, in line with ISO15189:2012. Study progress was evaluated bi-weekly in a multidisciplinary team involving study coordinators, pathologists, radiologists, medical oncologists, clinical geneticists and support staff. As a result, multiple stages of the process underwent optimizations such as biopsy procedures, sample logistics (tissue and DNA handling and processing) and bioinformatics (supplementary material, Figure S2).

Sample size calculation

The objective sample size of 1,200 patients was based on the primary endpoint ''clinical validation.'' The aim was to detect the same variants by WGS as SOC MolDx in at least 95% of the cases with one sided 95% confidence. Under an assumed concordance rate of 97.5%, 624 individual genomic biomarkers were needed to achieve a lower limit of the confidence interval to be at least 95% with a power of 96%. Based on retrospective WGS data analysis of ~3,000 patients (20), the required 624 SOC biomarkers were expected to be identifiable in 1,200 patients.

Statistical analysis

Patient and tumor characteristics, feasibility, clinical validity, and clinical value were analyzed with descriptive statistics. Categorical variables are shown as percentages or frequencies and continuous variables as medians with ranges. Analyses were performed using the Matplotlib and NumPy packages in Python, 3.7.5 (29, 30).

Results

Feasibility of WGS in routine clinical care

One thousand two hundred patients with 32 different tumor types were included over a 22 month period (Table 1), and 95 patients underwent >1 sampling procedure, resulting in 1,302 samples in total. Of these, 86 (7%) did not contain tumor cells. The remaining 1,216 samples, consisting of biopsies (n=931), resections (n=247) and cytological specimens (n=38), entered the WGS procedure. Overall, WGS was successfully completed in 70% (854/1216) of tumor samples (Figure 1A). In 9% (113/1216) of samples pTCP was <20% and consequently WGS was not started, and in 15% (181/1,216) the molecular TCP (mTCP, as determined by WGS data) detected by WGS was <20% despite a pTCP ^{3}20%. Other reasons for drop out included poor DNA quality (technical failure) in 3% (31/1,216), low DNA yield in 2% (29/1,216) and patient-specific circumstances (e.g. allogenic stem cell transplantation) in 0.7% (8/1,216). Over time, feasibility of WGS improved from 50% in 111 WGS attempts during the first three months to 72% in 177 attempts during the last three months of the study, due to continuous improvements in sample retrieval, handling, and processing (Figure 1B). Direct feedback between pathologists and radiologists performing image guided biopsies contributed to this improvement. WGS was successfully performed with a median turnaround time (TAT) of 12 working days (range 4–52) (Figure 1C). TAT improved throughout the study down to 11 working days in the final three months, and 95% of the WGS results were available within 17 working days. Importantly, when WGS could not be completed, yet a clinical indication for MolDx existed, both panel sequencing and Archer fusion analysis was successful in 87% (186/214) of these cases, indicating that targeted sequencing approaches can still be performed in the majority of cases if WGS is not feasible.

Determinants of WGS feasibility

Success rates of WGS on tissues from biopsies and resections (including 187 fresh frozen archival tissues) were 70% and 77%, respectively, and could be performed on both freshly obtained samples and fresh frozen archival tissue (Figure 1D). Cytology specimens proved a less suitable source for WGS with a success rate of 21%, mainly due to a low pTCP (68% of cases). WGS success rates differed by biopsy sites (supplementary material, Figure S3), with the highest success rates from liver (78%, n=298), soft tissue (70%, n=138) and lymph nodes (62%, n=177), and lowest success rates from lung (transthoracic biopsies 49%, n=168), peritoneum (52%, n=58) and bone (54%, n=47). Details on all WGS attempts are listed in supplementary material, Table S5.

Table 1. Baseline table for all 1,200 included patients with metastatic cancer

Patient characteristics	N=1200
Age at WGS, years	
Mean	59.3
Range	18 – 98
Sex, male: female	43: 57
Primary tumour location, n (%)	344 (29%)
Lung cancer	
Colorectal cancer	210 (18%)
Breast cancer	143 (12%)
Sarcoma	80 (6.7%)
Other	67 (5.6%)
Prostate cancer	47 (3.9%)
CUP	34 (2.8%)
Ovarian cancer	32 (2.7%)
Melanoma	29 (2.4%)
Bladder cancer	25 (2.1%)
Lung NETs	23 (1.9%)
Esophageal cancer	21 (1.8%)
Renal cell cancer	19 (1.6%)
Head and Neck cancer	14 (1.2%)
Stomach cancer	13 (1.1%)
GEP-NETs	13 (1.1%)
Cervical cancer	11 (0.92%)
GIST	10 (0.83%)
Malignant mesothelioma	9 (0.75%)
Urothelial cell cancer of the bladder and renal pelvis	8 (0.67%)
Anal cancer	7 (0.58%)
Thymoma and Thymic cancer	7 (0.58%)
Vulva cancer	6 (0.50%)
Penile cancer	6 (0.50%)
Pancreatic cancer	5 (0.42%)
Endometrial cancer	4 (0.33%)
Lymphoma	4 (0.33%)
Cholangiocarcinoma	3 (0.25%)
Thyroid cancer	3 (0.25%)
Basal cell carcinoma	2 (0.17%)
Hepatocellular carcinoma	1 (0.083%)

CUP, Cancer of Unknown Primary; GEP-NETs = gastroenteropancreatic neuroendocrine tumor; GIST, gastrointestinal stromal tumor; NET, neuroendocrine tumor; WGS, whole genome sequencing

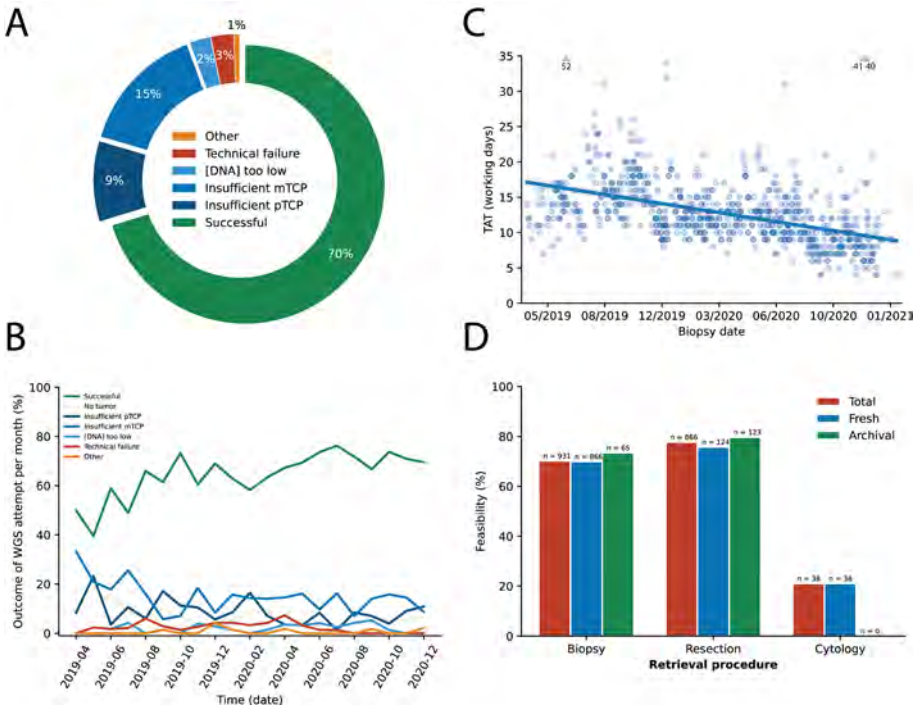


Figure 1. Feasibility of WGS in routine pathology practice. WGS was successfully completed in 854/1,216 (70%) samples containing tumor cells. The main reason for ineligibility for WGS was a low percentage of tumor cells in 24% of samples (9% insufficient pTCP and 15% insufficient mTCP) (A). Due to continuous optimizations during the course of the study, feasibility of WGS improved as samples with no tumor cells or insufficient TCP declined (B). WGS could be performed in a clinically acceptable time frame of a median of 12 working days. During the course of the study, the TAT decreased from 16 workings days in the first three months to 11 workings days in the last three months (C). Success rates of WGS procedures were high when using biopsy or resection specimens (70 and 77%, respectively), and could be performed both freshly obtained and fresh frozen archival tissue. Cytology specimens were less suitable for WGS with a success rate of 21% (D). mTCP, molecular tumor cell percentage; pTCP, tumor cell percentage as assessed by a pathologist.

Biomarker detection

Concordance between SOC MolDx and WGS was examined by looking at all biomarkers that could in principle be detected by both methods. Any discordant result was analyzed and classified using a pre-defined scheme (supplementary material, Figure S1). In total, 932 biomarkers, including 766 driver mutations (SNVs/MNVs/indels), 100 copy number alterations (amplifications/deletions), 46 fusion events, 13 viral insertions, and 7 genome-wide signatures were identified for comparison between WGS and SOC MolDx (Table 2). On top of these 932 biomarkers potentially detectable by SOC MolDx, WGS detected an additional 3860 biomarkers, including 249 genome-wide signatures (high mutational load [n=195], MSI [n=8],

and HRD [n=46] (supplementary material, Table S6). This included 2,018 biomarkers in patients who, by that time, did not have an indication for SOC MolDx and 1,842 biomarkers not covered by targeted sequencing panels in patients who did receive SOC MolDx.

Table 2. Concordance of WGS and SOC MolDx diagnostics

	Total (n=914)	SNVs/MNs/ indels (n=760)	Copy number variants		Fusions (n = 46)	Viral insertions (n=13)
			Amplifications (n=66)	Deletions (n=31)		
True positives	**903**	**760**	**64**	**25**	**41**	**13**
- WGS	887	749	64	25	36	13
- SOC MolDx	889	755	64	15	41	13
False negatives						
- WGS	16	11	0	0	5	0
- SOC MolDx	14	4	0	10	0	0
False positives						
- WGS	3	0	0	0	3	0
- SOC MolDx	8	0	2	6	0	0
Post-hoc analysis						
	Total (n=880)	**SNVs/MNs/ indels (n=760)**	**Copy number variants**		**Fusions (n = 46)**	**Viral insertions (n=13)**
			Amplifications (n=66)	**Deletions (NA)**		
True positives	**878**	**760**	**64**		**41**	**13**
- WGS	871	754	64	NA	40	13
- SOC MolDx	876	758	64		41	13
False negatives						
- WGS	7	6	0	NA	1	0
- SOC MolDx	2	2	0		0	0
False positives						
- WGS	0	0	0	NA	0	0
- SOC MolDx	2	0	2		0	0

Bold is used for total numbers, e.g. in the *post hoc analysis* first row there are 878 true positives with 871 biomarkers detected by WGS + 7 (false negatives) – 0 (false positives).

Indels, insertion/deletions; MNVs, multi-nucleotide variants; MolD, molecular diagnostics; NA, not available; SNVs, single nucleotide variants; SOC, standard of care; WGS, whole genome sequencing

Of the 924 true-positive biomarkers that were diagnostically reported during the course of the study, WGS detected 904 (97.8%, two sided 95% CI 96.7–98.7%), along with 3 false-positive calls (Figure 2A), while SOC MolDx detected 910 (98.5%, two sided 95% CI 97.5–99.2%) along with 5 false-positive calls. As part of continuous optimization, changes in the WGS pipeline included optimized calling of splice variants and detection of fusion events. In the SOC MolDx workflow, NGS-based

calling of gene deletions were abandoned because of large numbers of false-positive and false–negative results observed. In a *post hoc* analysis with the latest versions of these pipelines, biomarker detection rates were 889/896 (99.2%, two sided 95% CI 98.4–99.6%) for WGS and 894/896 (99.7%, two sided 95% CI 99.2–99.9%) for SOC MolDx, along with 0 and 2 false positives, respectively (Table 2, Figure 2B). These remaining false-negative and –positive results were attributable to various factors (supplementary material, Table S7).

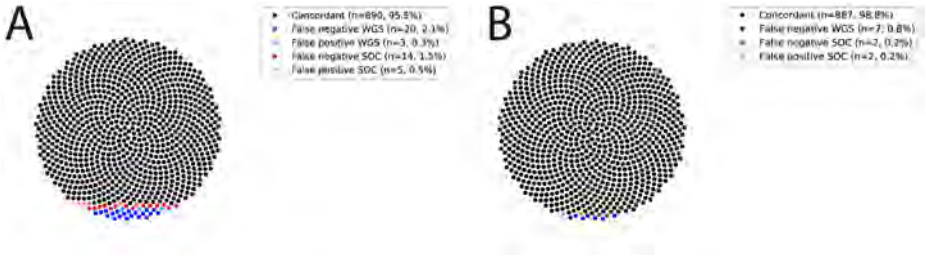


Figure 2. Concordance of WGS and SOC MolDx was determined in two ways. (A) by comparing WGS and SOC MolDx results as they were reported real time during the course of the study, while in accordance with ISO 15189:2012 a continuous optimization process of bioinformatic procedures took place and (B) in a *post hoc* analysis of all samples using the latest optimized pipelines as these emerged by the end of the study.

SOC, standard of care; WGS, whole-genome sequencing

Clinical value of WGS

Of 848 patients, 603 (71%) had ≥1 actionable event(s), i.e. biomarker-based eligibility for either regular therapy or a clinical trial in the Netherlands (supplementary material, Figure S4). In 250 patients, multiple biomarker-based therapy options were detected (supplementary material, Figure S5), resulting in a total of 936 different regular (n=145) or experimental (n=791) therapy options (supplementary material, Figure S6). Of these, 343 were identified with SOC MolDx (36.6%). Conversely, 593 therapy options in 431 patients remained undetected without WGS, either because SOC MolDx was not (yet) performed as part of the regular diagnostic work-up (345 options in 241 patients), or genomic biomarker regions were not covered by SOC MolDx (248 options in 190 patients).

At a median follow-up of 14 months, 147 out of 603 patients with actionable events (24%) had started with a biomarker-based therapy in a regular (n=54, 11% based on WGS-only findings) or clinical trial setting (n=93; 80% based on WGS-only findings). These numbers are likely to increase as not all patients have exhausted their regular treatment options (Figure 3, Table 3).

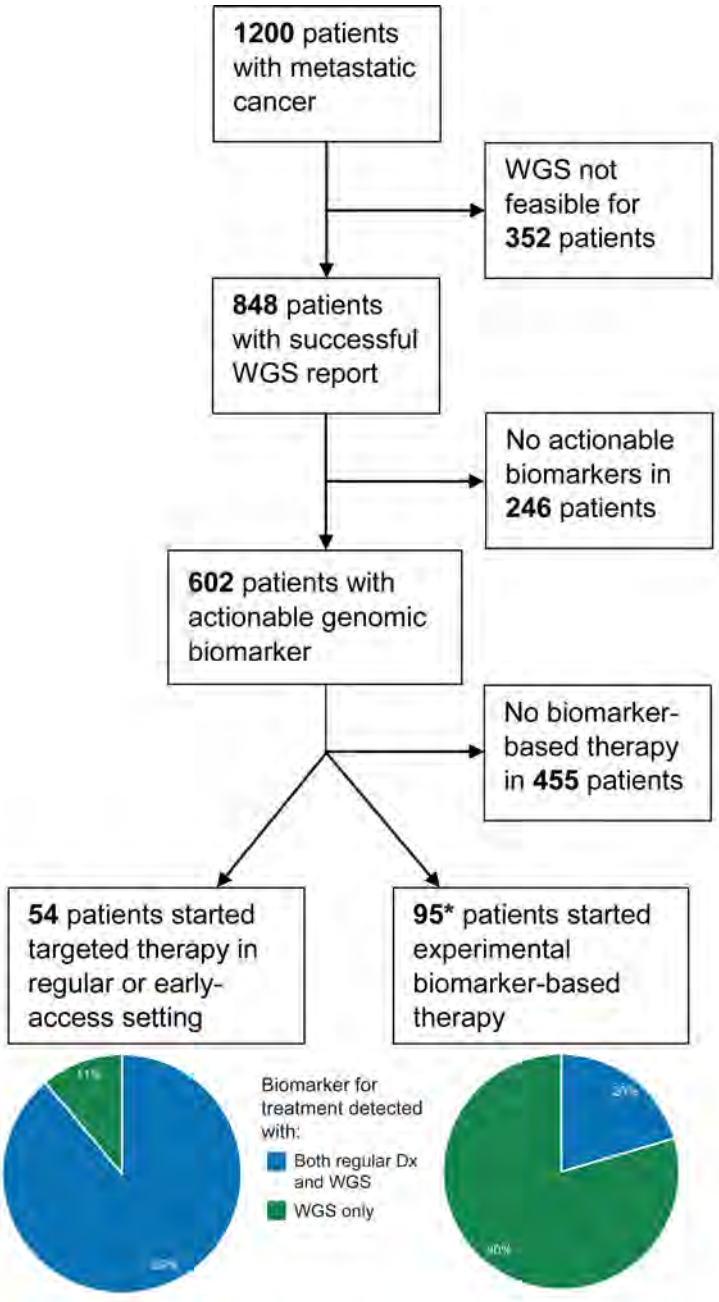


Figure 3. Clinical value of prospective WGS. Ultimately, 147 patients started biomarker-based therapy at a median follow-up of 14 months, of which 2 patients (*) received both biomarker-based therapy in a regular setting and an experimental setting after progression. Dx, diagnostics, WGS, whole-genome sequencing.

Table 3. Treatment options in patients (who initiated therapy) based on SOC + WGS or WGS only-results

Regular therapy	Number of patients
Detected with SOC + WGS	40
Detected with WGS only	3
Early access program	
Detected with SOC + WGS	8
Detected with WGS only	3
Clinical trials	
Detected with SOC + WGS	19
Detected with WGS only	76
Regular + early access program	
Detected with SOC + WGS	48
Detected with WGS only	6

SOC, standard of care; WGS, whole-genome sequencing

In 70 patients, 72 pathogenic germline variants (PGVs) were detected, 23 of which had been identified before with SOC germline diagnostics (Figure 4 and supplementary material, Table S8). Interestingly, somatic second hits (biallelic mutation or LOH) were present in only 41/72 (57%) patients, predominantly in cancer predisposition genes associated with the tumor type at hand (supplementary material, Figures S7, S8). Biallelic loss in the background of PGVs provided a rationale for tumor-directed therapy in 20/39 patients with PGVs in HRD genes, and 4/7 patients with PGVs in mismatch repair (MMR) genes (supplementary material, Figure S9). In all patients with biallelic loss of function of HRD or MMR genes, genome-wide signatures of HRD (15) and MSI, respectively, were present (supplementary material, Table S9).

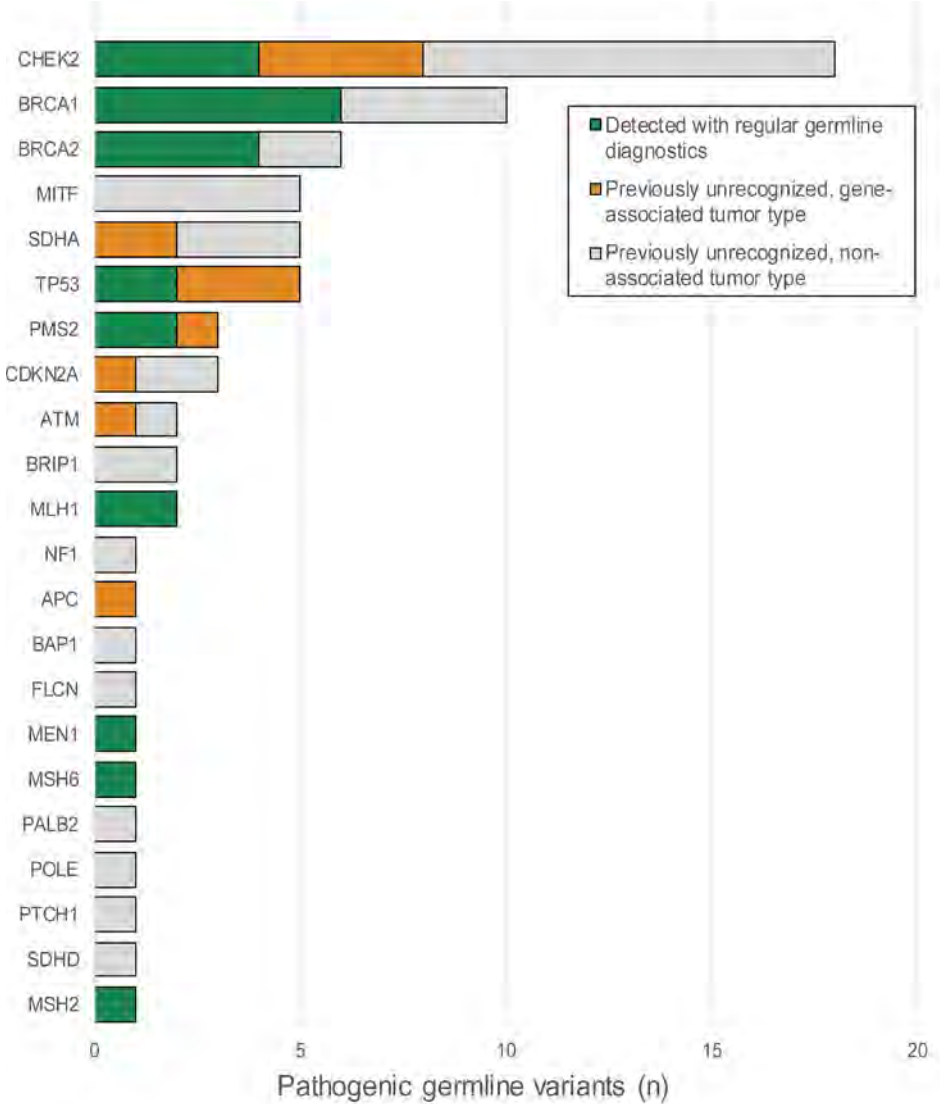


Figure 4. Germline variants detected by WGS. In total, 72 pathogenic germline variants (PGVs) were identified by WGS in 848 patients, of which 23 previously had been detected with SOC diagnostics. The figure shows the type and number of PGVs identified in these 848 patients and whether they were detected with SOC diagnostics or previously unrecognized.

Discussion

In modern precision oncology, we are facing the diagnostic challenge to identify all relevant genomic alterations for every individual cancer patient, the number of which increases with the rapidly expanding druggable genome. While this requires diagnostic pathology laboratories to continuously update their MolDx arsenal, in reality, infrastructural delays occur in implementing assays for new biomarkers, which translates into delayed access for patients to new treatments (31). As WGS allows for complete genomic characterization, any new DNA based biomarker is by definition already covered by WGS; it merely requires a small adaptation of the bioinformatics or reporting pipeline, thus providing a versatile solution to this challenge (32).

In the WIDE study we have demonstrated that implementation of WGS, including adapting to fresh frozen instead of FFPE sample logistics, is well feasible in routine pathology and clinical practice (25). While the current study was conducted in the setting of a dedicated comprehensive cancer center, laboratory procedures basically do not differ from other pathology laboratories operating under ISO15189 and according to professional guidelines. WGS succeeded in 71% of metastatic cancer patients within clinically acceptable timelines, even when only the second-best sample was used (33). Furthermore, with steadily decreasing sequencing costs, sequencing with a deeper coverage may further increase WGS feasibility in low TCP samples (28). Importantly, in the majority of cases in which a low TCP was limiting for WGS, SOC MolDx was still feasible using the isolated fresh frozen DNA or remaining biopsy material from the same procedure. A sensible strategy would therefore be to use WGS when possible and panels when needed, thus providing the most comprehensive MolDx possible for every patient (Supplementary material, Figure S10). The concordance between WGS and SOC MolDx of 98.8% demonstrated the clinical validity of WGS (Figure 2).

During the study, pathologists, clinical molecular biologists in pathology, medical oncologists and clinical geneticists became more familiar with the interpretation and additional diagnostic value of WGS, especially in the context of complex differential diagnoses (18, 34). In fact, WGS also appeared to have additional diagnostic value in germline diagnostics, with previously 49 unrecognized pathogenic germline variants in cancer susceptibility genes being identified. WGS thus encompasses valuable somatic (including genome-wide signatures) and germline information, allowing further optimization of therapeutic strategies.

In line with other reports (20, 35), here WGS identified actionable biomarkers for regular therapy options or clinical trial allocation in 71% of patients with a WGS result. The majority of these biomarkers were not detected with current SOC diagnostic approaches including targeted sequencing panels in selected patient populations. Moreover, comprehensive genomic characterization of tumors by WGS in combination with detailed clinical phenotyping provides a solid basis for a learning healthcare system, which is a crucial condition for deploying precision medicine to its full extent.

As an alternative to WGS, whole exome sequencing (WES) is occasionally proposed. Often this is a cost-based consideration since the laboratory logistics and bioinformatics pipeline are similar for both methods but less (costly) sequencing reagents are required. However, WGS allows the analysis of more complex tumor characteristics, including mutational signatures and MSI analysis. Reliable quantification of these characteristics also relies on the intronic and intergenic analysis, and these will not be analyzed by WES. Furthermore, detection of fusion genes, typically formed by fusions of intergenic breakpoint regions, is not possible with WES.

Evidently, cost is a crucial consideration for implementing WGS into routine pathology practice, and trade-offs may vary between different institutions/locations. While at an individual test level, direct costs of WGS are higher than of WES or NGS panels, a comprehensive cost versus benefit analysis is much more complex. This is being studied outside the scope of the present study and will be presented elsewhere.

In summary, the present prospective study has demonstrated that WGS-based diagnostics is feasible in routine pathology practice and adds value for clinical decision making. The required adjustments in laboratory logistics were well manageable and acceptable to the health care professionals involved which shows implementation hurdles in adopting WGS in routine pathology practice can be overcome. In fact, in immediate follow up to the present study, the Department of Pathology at the Netherlands Cancer Institute, in collaboration with Hartwig, has implemented WGS in routine clinical practice. This is further facilitated by the fact that recently a first provision for reimbursement of WGS in the Netherlands has been established (36).

References

1. Marcus L, Fashoyin-Aje LA, Donoghue M, Yuan M, Rodriguez L, Gallagher PS, et al. FDA Approval Summary: Pembrolizumab for the Treatment of Tumor Mutational Burden-High Solid Tumors. Clin Cancer Res. 2021;27(17):4685-9.
2. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA Approval Summary: Pembrolizumab for the Treatment of Microsatellite Instability-High Solid Tumors. Clin Cancer Res. 2019;25(13):3753-8.
3. Looney AM, Nawaz K, Webster RM. Tumour-agnostic therapies. Nat Rev Drug Discov. 2020;19(6):383-4.
4. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, et al. Efficacy of Larotrectinib in TRK Fusion–Positive Cancers in Adults and Children. New England Journal of Medicine. 2018;378(8):731-9.
5. Doebele RC, Drilon A, Paz-Ares L, Siena S, Shaw AT, Farago AF, et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. Lancet Oncol. 2020;21(2):271-82.
6. Yap TA, Johnson A, Meric-Bernstam F. Precision Medicine in Oncology-Toward the Integrated Targeting of Somatic and Germline Genomic Aberrations. JAMA Oncol. 2021;7(4):507-9.
7. Aitken M VP, Bennet K, Tewary V, Lech C. Optimizing Oncology Care Through Biomarker Adoption. IQVIA Institute. 2020.
8. Lynch JA, Berse B, Rabb M, Mosquin P, Chew R, West SL, et al. Underutilization and disparities in access to EGFR testing among Medicare patients with lung cancer from 2010 - 2013. BMC Cancer. 2018;18(1):306.
9. Mosele F, Remon J, Mateo J, Westphalen CB, Barlesi F, Lolkema MP, et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. Ann Oncol. 2020;31(11):1491-505.
10. Steeghs EMP, Groen HJM, Schuuring E, Aarts MJ, Damhuis RAM, Voorham QJM, et al. Mutation-tailored treatment selection in non-small cell lung cancer patients in daily clinical practice. Lung Cancer. 2022;167:87-97.
11. Horak P, Heining C, Kreutzfeldt S, Hutter B, Mock A, Hullein J, et al. Comprehensive Genomic and Transcriptomic Analysis for Guiding Therapeutic Decisions in Patients with Rare Cancers. Cancer Discov. 2021;11(11):2780-95.
12. van der Wijngaart H, Hoes LR, van Berge Henegouwen JM, van der Velden DL, Zeverijn LJ, Roepman P, et al. Patients with Biallelic BRCA1/2 Inactivation Respond to Olaparib Treatment Across Histologic Tumor Types. Clin Cancer Res. 2021;27(22):6106-14.
13. Sokol ES, Pavlick D, Khiabanian H, Frampton GM, Ross JS, Gregg JP, et al. Pan-Cancer Analysis of BRCA1 and BRCA2 Genomic Alterations and Their Association With Genomic Instability as Measured by Genome-Wide Loss of Heterozygosity. JCO precision oncology. 2020;4:442-65.
14. Van Hoeck A, Tjoonk NH, van Boxtel R, Cuppen E. Portrait of a cancer: mutational signature analyses for cancer diagnostics. BMC Cancer. 2019;19(1):457.

15. Nguyen L, J WMM, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. Nat Commun. 2020;11(1):5584.
16. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 2017;23(4):517-25.
17. Nguyen L vHA, Cuppen E. Machine learning-based tissue of origin classification for cancer of unknown primary diagnostics using genome-wide mutation features. BioRxiv. 2021;2021.10.05.463244.
18. Schipper LJ SP, Samsom KG, Bosch LJW, Lalezari F, Priestley P, et al. . #3448; Whole Genome Sequencing can classify diagnostically challenging tumors ESMO Congress 2021 Abstract Book, Annals of Oncology. 2021.
19. Campbell PJ GG, Korbel JO, Stuart JM, Jennings JL, Stein LD, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578(7793):82-93.
20. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575(7781):210-6.
21. Wong M, Mayoh C, Lau LMS, Khuong-Quang DA, Pinese M, Kumar A, et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. Nat Med. 2020;26(11):1742-53.
22. Nangalia J, Campbell PJ. Genome Sequencing during a Patient's Journey through Cancer. N Engl J Med. 2019;381(22):2145-56.
23. Byrjalsen A, Hansen TVO, Stoltze UK, Mehrjouy MM, Barnkob NM, Hjalgrim LL, et al. Nationwide germline whole genome sequencing of 198 consecutive pediatric cancer patients reveals a high incidence of cancer prone syndromes. PLoS Genet. 2020;16(12):e1009231.
24. Trotman J, Armstrong R, Firth H, Trayers C, Watkins J, Allinson K, et al. The NHS England 100,000 Genomes Project: feasibility and utility of centralised genome sequencing for children with cancer. Br J Cancer. 2022.
25. Samsom KG, Bosch LJW, Schipper LJ, Roepman P, de Bruijn E, Hoes LR, et al. Study protocol: Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient (WIDE). BMC Med Genomics. 2020;13(1):169.
26. van de Haar J, Hoes LR, Roepman P, Lolkema MP, Verheul HMW, Gelderblom H, et al. Limited evolution of the actionable metastatic cancer genome under therapeutic pressure. Nat Med. 2021;27(9):1553-63.
27. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
28. Roepman P, de Bruijn E, van Lieshout S, Schoenmaker L, Boelens MC, Dubbink HJ, et al. Clinical Validation of Whole Genome Sequencing for Cancer Diagnostics. J Mol Diagn. 2021;23(7):816-33.
29. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357-62.

30. Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007; 9(3):90-95.
31. van den Broek D, Hiltermann TJN, Biesma B, Dinjens WNM, t Hart NA, Hinrichs JWJ, et al. Implementation of Novel Molecular Biomarkers for Non-small Cell Lung Cancer in the Netherlands: How to Deal With Increasing Complexity. Front Oncol. 2019;9:1521.
32. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat Commun. 2019;10(1):3240.
33. Lindeman NI, Cagle PT, Aisner DL, Arcila ME, Beasley MB, Bernicker EH, et al. Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for Treatment With Targeted Tyrosine Kinase Inhibitors: Guideline From the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. Arch Pathol Lab Med. 2018;142(3):321-46.
34. Schipper LJ, Monkhorst K, Samsom K, Snaebjornsson P, Van Boven H, Roepman P, et al. Whole-genome sequencing to improve sarcoma diagnosis and patient care. Journal of Clinical Oncology. 2021;39(15_suppl):11540-.
35. Cobain EF, Wu YM, Vats P, Chugh R, Worden F, Smith DC, et al. Assessment of Clinical Benefit of Integrative Genomic Profiling in Advanced Solid Tumors. JAMA Oncol. 2021;7(4):525-33.
36. Zorginstituut. Moleculaire diagnostiek in de Oncologie. Nederland, Diemen. 2021.

CHAPTER 5

Complete genomic characterization in patients with cancer of unknown primary in routine diagnostics

Luuk J. Schipper*, Kris G. Samsom*, Petur Snaebjornsson, Thomas Battaglia, Linda J.W. Bosch, Ferry Lalezari, Peter Priestley, Charles Shale, Alexandra J. van den Broek, Nina Jacobs, Paul Roepman, Jacobus J.M. van der Hoeven, Neeltje Steeghs, Marieke A. Vollebergh, Serena Marchetti, Edwin Cuppen, Gerrit A. Meijer, Emile E. Voest & Kim Monkhorst

* *Equally contributing first authors*

ESMO Open, 2022

Abstract

Background

In approximately 3-5% of patients with metastatic disease, tumor origin remains unknown despite modern imaging techniques and extensive pathology work-up. With long diagnostic delays and limited and ineffective therapy options, the clinical outcome of patients with cancer of unknown primary (CUP) remains poor. Large-scale genome sequencing studies have revealed that tumor types can be predicted based on distinct patterns of somatic variants and other genomic characteristics. Moreover, actionable genomic events are present in almost half of CUP patients. This study investigated the clinical value of whole genome sequencing (WGS) in terms of primary tumor identification and detection of actionable events, in the routine diagnostic workup of CUP patients.

Patients and methods

A WGS-based tumor type 'cancer of unknown primary prediction algorithm' (CUPPA) was developed based on previously described principles and validated on a large pan-cancer WGS database of metastatic cancer patients (>4000 samples) and 254 independent patients, respectively. We assessed the clinical value of this prediction algorithm as part of routine WGS-based diagnostic work-up for 72 CUP patients.

Results

CUPPA correctly predicted the primary tumor type in 78% of samples in the independent validation cohort (194/254 patients). High-confidence predictions (>95% precision) were obtained for 162/254 patients (64%). When integrated in the diagnostic work-up of CUP patients, CUPPA could identify a primary tumor type for 49/72 patients (68%). Most common diagnoses included non-small cell lung (n=7), gastroesophageal (n=4), pancreatic (n=4), and colorectal cancer (n=3). Actionable events with matched therapy options in clinical trials were identified in 47% of patients.

Conclusion

Genome-based tumor type prediction can predict cancer diagnoses with high accuracy when integrated in the routine diagnostic work-up of patients with metastatic cancer. With identification of the primary tumor type in the majority of patients and detection of actionable events, WGS is a valuable diagnostic tool for patients with CUP.

Introduction

Cancer of Unknown Primary (CUP) accounts for approximately 3-5% of all metastatic cancers. Despite advancements in the diagnostic arsenal of pathologists and improved imaging modalities over the last decades, primary tumor type remains ambiguous or undetectable for this patient group. Consequently, effective therapeutic options remain very limited. Patients with CUP generally undergo a long diagnostic process, and clinical deterioration prohibits a timely start of treatment in more than 50% of cases (1). When CUP-directed treatment is initiated, patients are usually treated with standard platinum containing chemotherapy (combination) regimens with limited clinical benefit. Consequently, CUP patients continue to have a dismal prognosis, with a median overall survival of 3 months after diagnosis (2).

Genomic variation has long been recognized as a means for tumor type prediction (3). For example, the mutational landscape in driver genes across tumor types has been applied as a mean to delineate tumor type (4). Furthermore, mutational profiles of passenger mutations were identified as indicators of cancer etiology, which led to the development of widely used COSMIC signatures (5) that can also be used to distinguish tumor types with high accuracy (6). Next, the topological distribution of driver and passenger mutations varies considerably across cancer types due to chromatin organization and activity differences in the cells of origin (7). As such, regional mutational density across the genome has diagnostic power (8). Finally, combinatorial approaches, in which multiple genomic features are grouped within one prediction algorithm, have shown to further improve classifier accuracy (8-10). With the increasing use of comprehensive genomic profiling of cancer patients in daily clinical practice, genomics-based tumor type prediction algorithms would be readily implementable in clinical care.

A second advantage to using genomic profiling lies in simultaneous identification of actionable events for targeted treatments. CUP patients have a limited number of therapy options, only consisting of chemotherapy regimens. Almost half of CUP patients, however, harbor an actionable event with direct therapy options either in approved, off-label or clinical trial setting (11). Clinical efficacy of biomarker-guided targeted therapy based on these actionable events has been reported in several case reports. Moreover, based on a real-world cohort of 3,841 CUPs, patients with a targeted therapy approach demonstrated an improved overall survival compared to patients treated with chemotherapy (12).

Hence, genomic characterization could benefit CUP patients by (a) identification of primary tumor type with subsequent standard of care therapy options and (b) detection of actionable genomic events to increase the number of therapy options for these patients. For the purpose of primary tumor type prediction in CUPs in the clinical setting, we developed and trained a statistics-based diagnostic tumor tissue of origin prediction tool by combining tumor type-specific drivers, regional mutational density and mutational profile characteristics on a large pan-cancer WGS database (4,509 samples). Next, WGS and the Cancer of Unknown Primary Prediction Algorithm (CUPPA) was applied to 72 CUP patients who received a prospective whole-genome sequencing analysis in a routine diagnostic work-up, to analyze the value of WGS to identify primary tumor type and detect actionable events.

Methods

Sample collection and WGS procedure

Detailed information on sample collection and WGS procedure can be found elsewhere (13, 14). In summary, samples were collected as part of the CPCT-02 (NCT01855477), DRUP (NCT02925234), and WIDE (NL68609.031.18) clinical trials, approved by medical ethical committees of the University Medical Center Utrecht and the Netherlands Cancer Institute and conducted in concordance with the Declaration of Helsinki, Dutch law, and Good Clinical Practice. Fresh tumor samples were used for DNA isolation and sequenced at 90-100x coverage using uniform sample and data processing procedures by the Hartwig Medical Foundation (Hartwig) (15). A 10mL blood withdrawal was used to perform germline sequencing at 30x coverage to allow for somatic variant calling and excluding germline variation. After sequencing, genomic and clinical information including primary tumor type was stored in the Hartwig database. Samples with unknown or undocumented primary tumor type were excluded for development and validation of the tumor type prediction algorithm.

WGS-based tumor type prediction model

We developed CUPPA, a stastical model that weighs multiple genomic features, to find resemblance of a sample compared with different cohorts of samples based on their primary tumor origin (reference cohorts) (Fig. 1.). In case of a limited number of distinct samples of certain tumor origin, different primary tumor types were grouped into a single reference cohort based on clinicopathological similarities (Supplementary Table S1). Samples not fitting any reference cohort were excluded from analysis.

Three orthogonal DNA classifiers, each with predictive power for tissue of origin, were combined to reach an overall prediction. In more detail, independent classifiers for positional mutational distribution, relative contextual single nucleotide variants used for COSMIC signatures, and presence of cancer-type specific drivers and passenger mutational features, and the combined classifier are assigned a relative similarity likelihood to each primary cancer origin cohort with the sum of the likelihoods adding up to 1 across the 29 reference cohorts (figure 1). The similarity likelihood is provided with every CUPPA prediction as 'prediction likelihood'. Samples derived from males are excluded from matching 'Ovary' and 'Uterus' cancer reference cohorts and samples from females are excluded from matching the 'Prostate' reference cohorts in the combined classifier. A detailed description on the calculation of all classifiers can be found in the supplementary methods.

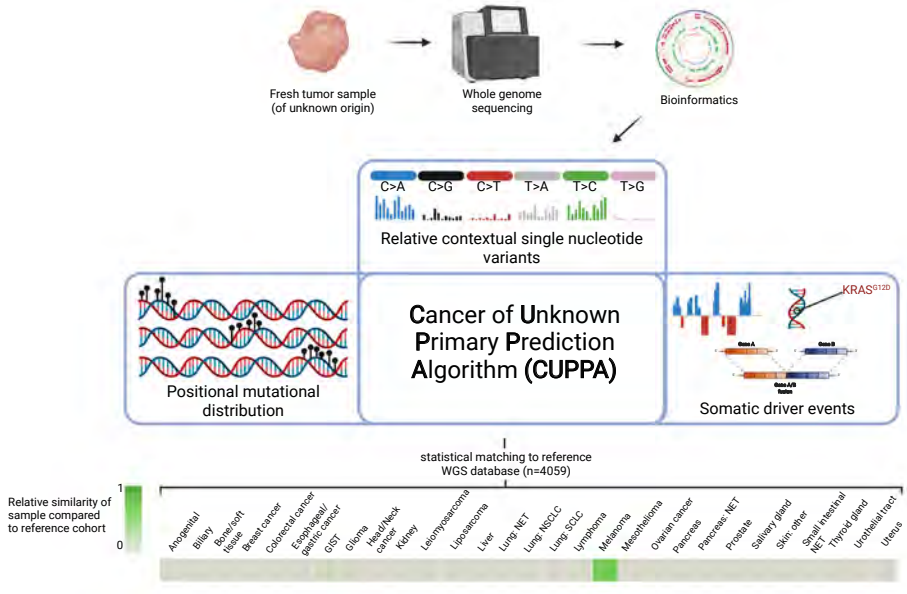


Figure 1. The cancer of unknown primary prediction algorithm (CUPPA). Created with BioRender.com. GIST, gastrointestinal stromal tumor; NET, neuroendocrine tumor; NSCLC, non-small-cell lung cancer; WGS, whole genome sequencing.

Internal validation

At time of analysis, the Hartwig database consisted of 4509 samples with known histopathological based primary origin. These samples were randomly divided in a reference set (90%, n = 4058) and a test set (10%, n = 451). CUPPA was applied to the test set with the reference set as base for reference cohort determination.

CUPPA overall prediction was compared to known histopathology based primary origin to determine predictive performance on tumors with known origin. Similarity likelihood scores were used to determine a cut-off for high-confidence and low-confidence CUPPA predictions.

Independent validation cohort

In the period January 2021 – September 2021, patients underwent WGS analysis as part of their regular diagnostic workup at the Netherlands Cancer Institute. WGS was either indicated for identification of therapy options, or for supporting classification of diagnostically challenging tumors or CUPs. All WGS analyses (data generation, data processing, actionable variant and CUPPA reporting) were performed in a centralized facility, operated by Hartwig. Use of data for this study was approved by the Institutional Review Board of the Netherlands Cancer Institute. All patients provided written informed consent for use of data for research purposes. Data were anonymized and handled in accordance with the Code for Proper Secondary Use of Human Tissue in the Netherlands.

Value in clinical setting

Treating oncologists and/or pathologists had the opportunity to request a CUPPA analysis for CUP patients and otherwise diagnostically complex tumors. Three patients have been described elsewhere (16). CUPs were defined as tumors with unknown origin or histological type. CUPPA was tested prospectively in all patients in the independent validation cohort. In a dedicated research meeting, with participation of expert pathologists and medical oncologists, CUPPA analyses were reviewed. Where needed, additional diagnostic tests were performed to adequately interpret clinical characteristics. For each CUPPA prediction, it was determined whether the prediction corresponded with the differential diagnosis of the pathologist prior to WGS analysis, or whether the prediction did not fit the clinical presentation of the patient.

Actionability

Previously described oncogenic driver likelihood scores were used to assess pathogenicity of variants, and if needed, additional diagnostic analyses were performed to determine gene and/or protein expression (13). For every oncogenic variant, it was determined whether the variant could elicit clinical trial participation in an ongoing trial within any Dutch hospital at time of the WGS analysis. For this, variants were correlated with the iClusion database, in which all clinical trials (phase I-III) running in the Netherlands and their eligibility criteria, including genomic indications, are stored. In case genomic variants were regarded as actionable at time

of WGS analysis, but compelling evidence for inefficacy led to discontinuation of the associated clinical trial (cohort) at a later time, these variants were retrospectively disregarded as viable clinical trial options.

Model evaluation

For the multiclass metastatic tumor origin classification task, an initial evaluation was performed using conventional definitions of sensitivity, specificity, F1-score and balanced accuracy on the argmax of the class probabilities (maximum cancer type probability). We next applied a one-versus-rest strategy to generate a single binary classification problem per class and then computed the mean area under the curve-receiver operator curve (AUC-ROC) and mean average precision (AP) through micro-, macro-, and weighted-averaging. Top-k accuracy estimates for $k \in 1, 3, 5$ in low-confidence tumor type predictions (no probability > 0.80), was computed by determining how often the true tumor type classification was included in the top-k model predictions, normalized by the number of samples.

Results

Model performance

Overall, CUPPA was able to correctly classify 84% and 78% of samples in the internal (n=451) and independent validation cohort (n=254), respectively (figure 1A, multimedia component 3) (Figure 2A, Supplementary Figure S1). Micro-averaged one-versus-rest ROCs showed an AUC of 0.993 and 0.986, and average precisions were 0.993 and 0.879 for the internal and independent validation cohort, respectively (figure 2B and C). Predictive performance of CUPPA was higher for common tumor types with a larger number of samples in the corresponding reference cohort (Supplementary Figures S1 and S2).

To achieve a predictive precision of 95% in the internal validation cohort, we used a similarity likelihood score of 0.8 as a cut-off for high-confidence predictions (Figure 2D). Within the internal validation cohort, 75% of samples (n=340/451) had a probability score of ≥0.8 (Supplementary Figure S3). At this cut-off, taking samples with a high-confidence prediction only, micro-averaged true positive rate and false positive rate were 0.961 and 0.0013, respectively. In the independent validation cohort, a lower percentage of samples reached a high-confidence prediction (162/254, 64%), possibly a result of the relative distribution of tumor types (Figure 2F, Supplementary Table S2). Subsequently, the lower overall precision for the external compared to the internal validation cohort (78% versus 84%) could be contributed to

tumor type distribution. All rare cancer samples without a matching reference cohort received a low-confidence prediction (Supplementary Table S3)

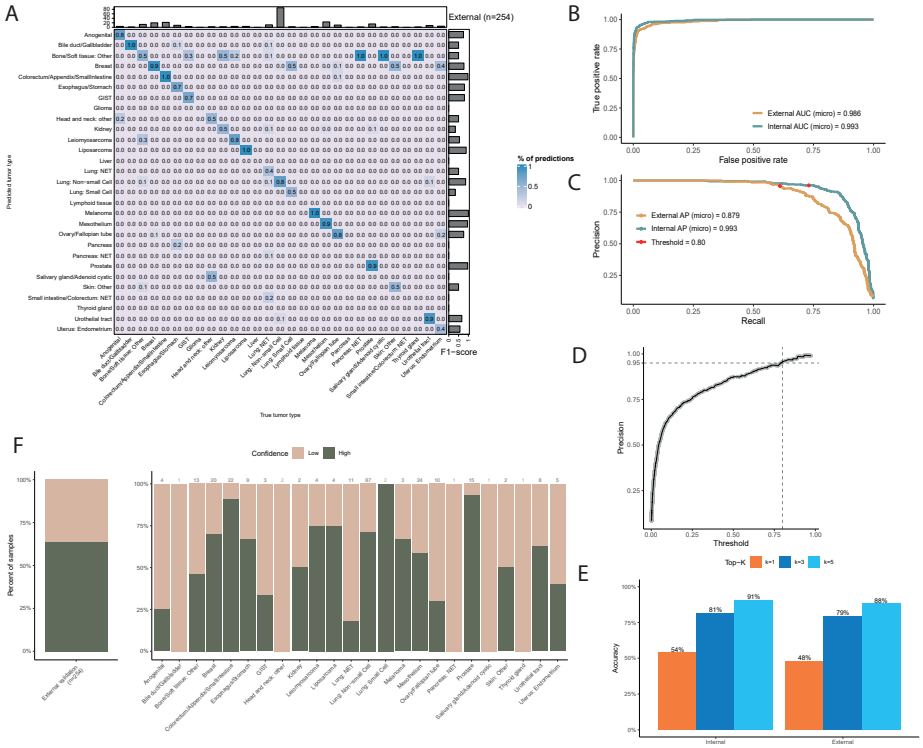


Figure 2. Predictive performance of CUPPA. (A) Confusion matrix showing model performance across 29 tumor types in the external validation cohort. The confusion matrices for the reference set and internal validation set can be found in the supplementary figures (Supplementary Figures S1 and S2). **(B)** Receiver operator curves and precision-recall curves **(C)** showing the overall model performance. **(D)** By using the probability score generated by the model as a cut-off, a high predictive precision (95%) was reached at a score of 0.8. The corresponding metrics at this cut-off are plotted in panel B and C. **(E)** In low-confidence predictions (<0.8), predictive precision was lower. The high top-3 and top-5 model accuracy demonstrated, however, that low-confidence predictions can be used as a mean to derive a differential diagnosis to correlate with clinicopathological differential diagnoses. **(F)** In total, 64% of samples reached a high-confidence prediction. Distribution of high- and low-confidence predictions varied across tumor types. AP, average precision; AUC, area under the curve; CUPPA, cancer of unknown primary prediction algorithm; GIST, gastrointestinal stromal tumor; NET, neuroendocrine tumor.

In samples with a low-confidence prediction (probability score <0.8), predictive precision was 48% (Figure 2E). To determine the value of low-confidence predictions in narrowing down differential diagnoses, we evaluated top-k accuracy, i.e. the predictive accuracy of the model taking the k highest confidence predictions.

CUPPA reached a top-3 accuracy of 79% and a top-5 accuracy of 88% in low-confidence predictions, indicating that top-5 predictions can serve as a genome-based differential diagnosis to correlate with the clinical, histopathological and immunohistochemistry derived differential diagnosis.

Clinical value in routine diagnostic work-up

The external validation cohort comprised of 254 patients with a variety of known tumor types that underwent a WGS analysis and prospective CUPPA analysis as part of their regular diagnostic work-up (Supplementary Figure S3). High-confidence predictions of CUPPA (64%) were prospectively incorporated within the diagnostic workup and correlated with the clinical presentation of patients. Initially, 7 of 162 (4.3%) high-confidence predictions did not match the presumed tumor type at time of diagnosis. Among these seven cases, three diagnostic revisions were made based on WGS and CUPPA analysis. This included one patient with a revision from a large-cell neuroendocrine carcinoma to a grade 3 neuroendocrine tumor of the lung, which was also more consistent with the clinical presentation of slow tumor progression. For a second patient, the diagnosis was revised from a sarcomatoid carcinoma of the lung to an undifferentiated sarcoma. Finally, one patient was diagnosed with small-cell lung cancer with widespread metastases in lymph nodes, pancreas, liver, skin, breast, and bones. Based on a high-confidence prediction for breast cancer, the lesion within the breast was regarded as the primary tumor, which fitted better with the clinical presentation: a young, female, non-smoking patient.

For two other discordant patients, CUPPA prediction did match with the expected tumor biology. First, a patient was diagnosed with a mature teratoma of the ovary with differentiation into intestinal type adenocarcinoma (Supplementary Figure S4). Consistent with tumor biology, this tumor was classified by CUPPA as a colorectal/small intestinal cancer sample with a high prediction (0.82). Second, a patient with non-small cell lung cancer (NSCLC) was biopsied after progression on chemoimmunotherapy. Although the case was initially not regarded as SCLC on morphological grounds, small cell transformation of NSCLC after immunotherapy has previously been associated with TP53 and RB1 loss (17).

When excluding these five cases, overall accuracy of CUPPA in high-confidence predictions was 155/157 (98.7%). Two cases with a high-confidence prediction were definitively misclassified; this includes a patient with diffuse-type gastric carcinoma that was classified as a pancreatic carcinoma and an undifferentiated pleiomorphic sarcoma that was classified as a leiomyosarcoma. Misclassifications of sarcoma subtypes are a known pitfall of CUPPA (Supplementary Table S2).

Tumor type predictions in CUP patients

Next, we applied WGS and CUPPA to 72 patients referred to the Netherlands Cancer Institute with a clinical diagnosis of CUP, for whom extensive pathological, radiological and endoscopic modalities failed to identify a primary tumor type (Table 1, multimedia component 9, multimedia component 10) (Table 1, Supplementary Tables S4 and S5). WGS was generally performed early in the disease course. 26 of these patients had a history of previous malignancy.

Table 1. Baseline characteristics of 72 CUP patients

	Total cohort (n=72)
Age at WGS analysis ***median (range)***	62 (18 – 81)
Gender ***n (%)***	
Female	31 (43%)
Male	41 (57%)
Disease duration in months ***median (range)***	2 (0 – 37)
Disease stage ***n (%)***	
Metastatic	67 (93%)
Locally advanced	5 (7%)
Number of previous systemic therapy lines ***n (%)***	
0	60
1	9
2	2
3	1
Oncological history ***n (%)***	26 (36%)
Tumor localization ***n (%)***	
Lymph nodes	55 (76%)
Lung	25 (35%)
Liver	24 (33%)
Bone	18 (25%)
Peritoneum	14 (19%)
Adrenal glands	8 (11%)
Skin/subcutis	5 (7%)
Other	15 (21%)
Morphology* ***n (%)***	32 (44%)
Adenocarcinoma	1 (1%)
Adenosquamous carcinoma	10 (14%)
Squamous cell carcinoma	2 (3%)
Neuroendocrine carcinoma	21 (29%)
Undifferentiated malignancy	6 (8%)
Other	

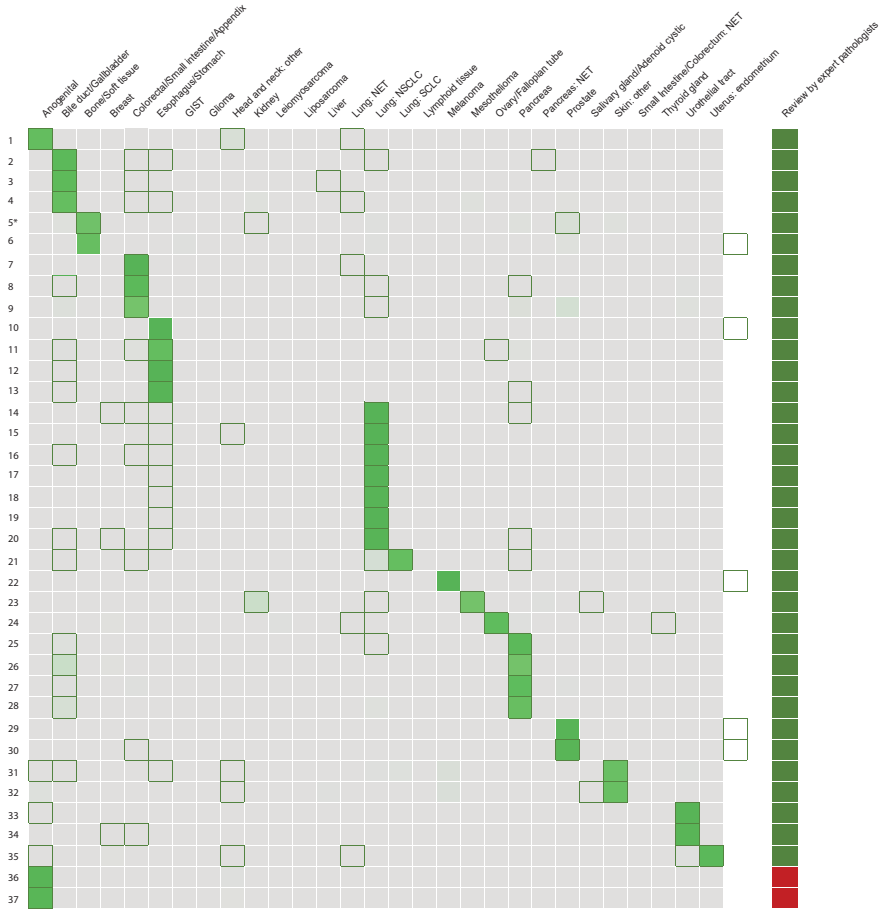
* Morphology as defined prior to reaching a diagnosis

WGS results and CUPPA predictions were correlated with the differential diagnosis that was composed based on clinical and pathological characteristics. For 37 patients (51%), CUPPA was able to provide a high-confidence prediction (probability score >0.8, Figure 3A). In all of these patients, the high-prediction score was consistent with one of the tumor types considered in the differential diagnosis prior to WGS analysis.

For 35 of the 37 patients, the high-confidence prediction led to a definitive diagnosis. The most common diagnoses included non-small cell lung cancer (n=7), gastroesophageal cancer (n=4), pancreatic (n=4), and colorectal/small intestinal cancer (n=3). For 2 of the 35 patients, the high-confidence tumor type prediction did not add to the clinicopathological considerations prior to WGS/CUPPA analysis. Based on clinicopathological diagnostics, these two patients were diagnosed with a squamous cell CUP, most likely from the anogenital area. Since anogenital carcinomas are grouped within one prediction category, these two high-confidence predictions (probability score of 98.4% and 98.3%, respectively) did not add any information regarding the primary site within the anogenital area. Of note, squamous cell carcinomas of the anogenital area are generally treated with platinum containing chemotherapy regimens.

In 35 CUP patients, a low-confidence prediction was reached (Figure 3B). By correlating the low-confidence predictions with the clinicopathological differential diagnosis of patients, however, these predictions proved to be informative in 12 patients where the differential diagnosis was narrowed down to a probable diagnosis as assessed by a panel of expert pathologists. A description on the diagnostic considerations and the added value of low-confidence predictions for these cases can be found in Supplemnetary Table S6. In two additional patients with a low-confidence prediction, WGS detected a disease-defining genomic event (Merkel cell polyoma virus and a SS18-SSX1 fusion). Taken together, with high-confidence (n=35) and low-confidence (n=12) CUPPA predictions and detection of diagnostic biomarkers (n=2), WGS was able to establish a diagnosis within the clinical context for 49/72 (68%) CUP patients.

A


Anogenital
Bile duct/Gallbladder
Bone/Soft tissue
Breast
Colorectal/Small intestine/Appendix
Esophagus/Stomach
GIST
Glioma
Head and neck: other
Kidney
Leiomyosarcoma
Liposarcoma
Liver
Lung: NET
Lung: NSCLC
Lung: SCLC
Lymphoid tissue
Melanoma
Mesothelioma
Ovary/Fallopian tube
Pancreas
Pancreas: NET
Prostate
Salivary gland/Adenoid cystic
Skin: other
Small Intestine/Colorectum: NET
Thyroid gland
Urothelial tract
Uterus: endometrium
Review by expert pathologists
1
2
3
4
5*
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

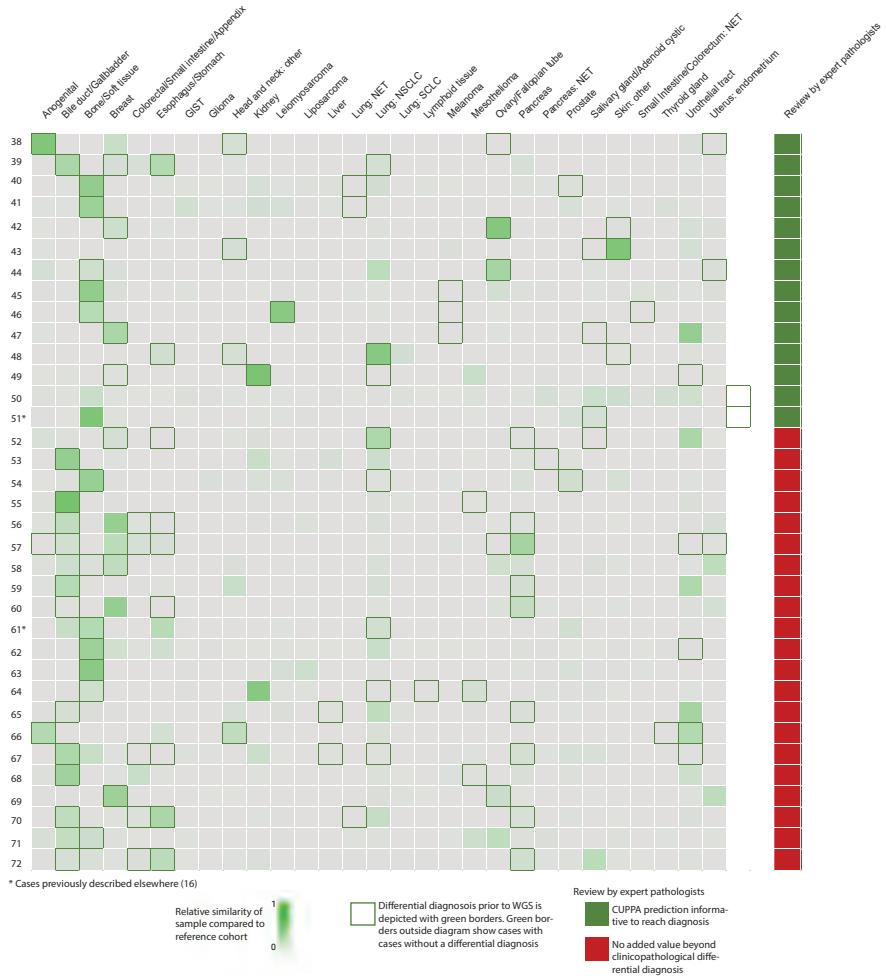


Figure 3. Application of CUPPA in diagnostic work-up of 72 CUP patients. For 37 (51%) patients, a high-confidence prediction was reached (panel A). All high-confidence predictions were consistent with the differential diagnosis (green borders) prior to WGS. For 35/37 cases, a final diagnosis could be reached (dark green boxes under 'Review by expert pathologists'). For two patients, the prediction did not provide additional information for the diagnosis (red boxes). For the remaining 35 cases, a low-confidence prediction was reached (panel B). When integrated with prior clinicopathological differential diagnoses, this prediction proved to be informative to reach a diagnosis in 12 patients (patients 38-49). In two additional patients (50 and 51), a disease-defining genomic event was detected with WGS. A description of the clinical value of CUPPA in low-confidence predictions can be found in Supplementary Table S6. CUPPA, cancer of unknown primary prediction algorithm; GIST, gastrointestinal stromal tumor; NET, neuroendocrine tumor; NSCLC, non-small-cell lung cancer; WGS, whole genome sequencing.

Therapeutic opportunities

With the identification of a primary tumor type, CUP-directed chemotherapy regimens can be substituted with tumor-type specific therapy regimens. In addition, genomic characterization allows for the detection of actionable variants, leading to biomarker-based therapeutic opportunities in clinical trial setting. We assessed genome-based actionability by assessing the number of experimental therapy options in ongoing Dutch clinical trials. For 34/72 CUP patients (47%), one or more therapy options were identified (Figure 4A, Supplementary Table S7). In 10 patients, multiple therapy options were identified, allowing for subsequent therapies or providing a rationale for combinatorial therapies. In most cases, this consisted of combinations with checkpoint inhibitors (Figure 4B). Actionable events were identified in 7/21 patients with a definite CUP (33%) for whom tumor-type directed therapy remained unavailable (Figure 4A). Taken together, adding up tumor-type directed regular therapy and detection of biomarker-based therapy options, WGS had potential therapeutic implications in the majority of CUP patients (56/72 patients, 78%).

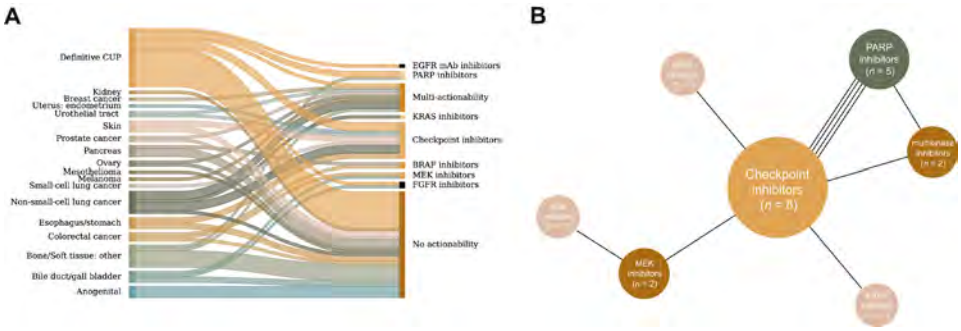


Figure 4. Biomarker-based therapy options detected with WGS in CUP patients. In 47% of patients, an actionable event was identified (panel A). In patients with a definitive CUP (n=21), an actionable event was identified in 33% (7 patients). For 10 patients, multiple therapy options were identified. In panel B, each line represents one of these 10 patients, showing the multiple therapy options identified in each patient. CUP, cancer of unknown primary; WGS, whole genome sequencing.

Discussion

We developed a tumor type prediction algorithm based on previously described genomic predictors that can be integrated in the regular diagnostic work-up with a short turnaround time of less than two weeks. As such, complete genomic characterization of CUP led to identification of a primary tumor type and detection of actionable events in 68% and 47% of patients (n=72), respectively. In all high-confidence predictions, the prediction was consistent with the clinicopathological differential diagnoses that were proposed prior to WGS analysis. Moreover, by integrating WGS with the clinicopathological differential diagnosis based on regular diagnostic tests, low-confidence predictions led to a diagnosis in 34% of patients. In our experience, clinical utility of CUPPA was optimal when used as an addition to the extensive diagnostic work-up that CUP patients generally receive, rather than using genome-based tumor type prediction as a stand-alone test. As such, CUPPA also holds the potential to recognize misdiagnoses, as 3/162 patients with a known diagnosis received a diagnostic revision after WGS analysis.

With this study, we focused on improving the diagnostic work-up among patients with CUP, both by identifying the primary tumor type and through identification of additional therapy options. The impact of providing more treatment opportunities on overall survival of these patients is beyond the scope of this study. CUPs are regarded as tumors that are relatively refractory to systemic treatment, and it remains to be seen whether tumor-type directed (chemo)therapy will benefit these patients in terms of survival. In a prospective clinical trial, 194 patients received a tumor type directed therapy based on a molecular gene expression classifier. Median overall survival time was higher compared to a historical cohort of CUP patients receiving CUP-directed chemotherapy (12.5 vs. 9.1 months) (18). In contrast, in two randomized clinical trials, site-specific treatment failed to improve median overall or progression free survival compared to empirical chemotherapy (19, 20). Results on efficacy of targeted therapy in this group have been more promising so far. Several case reports have demonstrated clinical benefit of molecularly guided targeted therapy, with durable partial responses or disease stabilization (21-25). Intriguingly, based on real-world data of 3841 CUP patients, a targeted therapy demonstrated improved survival outcomes compared with regular chemotherapy. Nonetheless, conclusive clinical evidence on the efficacy of targeted therapies in CUP patients is currently not available. To assess this systematically, the international multicenter CUPISCO trial is currently ongoing in which CUP patients are randomized (3:1) after three cycles of standard first-line chemotherapy and subsequently allocated to matched targeted therapy (26).

Of note, our approach has one major direct clinical advantage. All relevant diagnostic tests, including WGS analysis, were immediately requested at first suspicion of a CUP. WGS reports were routinely delivered within 2 weeks for the prospective part of this study (27). As a result , the time to diagnosis is shortened and a patient's tumor can be classified according to origin or as a definitive CUP. Shortening the diagnostic work-up could potentially improve clinical outcomes, as patients start systemic therapy with a better clinical condition. More importantly, the uncertainty regarding diagnosis poses a high psychological burden on patients.

Our tumor prediction algorithm is a statistical model that compares multiple genomic features of the sample of interest with a large reference database. An inherent limitation to this approach lies in the dependency on adequate sample size per reference cohort. Accurate tumor type classification is more challenging for cancer types with a limited number of samples in the reference cohort and low-confidence predictions should be interpreted with caution and regard for current pitfalls of the algorithm (Supplementary Table S2). Also, rare cancers are inevitably misclassified by the prediction algorithm, as they cannot be allocated to non-existing reference cohorts. To address this problem, our overarching ambition is to develop a learning healthcare system. Any new sample of known tumor origin that is being sequenced at Hartwig is automatically added to the reference database, providing patients have given consent for re-use of their data. Likewise, other WGS datasets of (metastatic) cancer patients generated outside Hartwig can be adjoined. Furthermore, enrichment strategies, in which samples of specific tumor types are sequenced and added to the reference database, can be implemented to reach a minimum threshold of samples for rare cancers or cancer types that are most relevant for differential and CUP diagnosis. Moreover, future discoveries in tumor type specific genomic disparities can be added in the algorithm as new predictors. For example, non-coding somatic drivers (28) or microbiome analyses (29) are possible with already available sequencing data. Finally, other modalities, like gene expression (30), methylation profiles (31), or digital whole-slide images (32), can be easily incorporated within the algorithm itself, although such strategies require additional analyses of the complete reference database and may be partially redundant with each other and/or with already used genomic features. The clinical value of incorporating multiple modalities has been demonstrated in a recently published cohort of 70 CUP patients, in which combined genomic, transcriptomic and methylome revealed a probable tumor type in 89% and treatment recommendations in 80% of patients (33). With the ability to continuously optimize the prediction algorithm, and increase the number of samples within the reference database, predictive accuracy of CUPPA is likely to improve over time. Adding other modalities will very likely improve the algorithm further.

In conclusion, complete genomic characterization with WGS was demonstrated to have a significant added value to the diagnostic arsenal for CUP patients. By integrating WGS into the diagnostic work-up, a primary tumor could be determined in 68% of CUP patients, and actionable events for matched therapy decisions in 47% of patients. Follow-up research on the efficacy of site-specific and matched targeted therapies is conducted internationally. Regardless, shortening of the diagnostic work-up allows for earlier treatment initiation, reduced diagnostic work up, and limited duration of CUP-associated psychological burden. With these considerations in mind, WGS is now being reimbursed for CUP patients in the Netherlands.

References

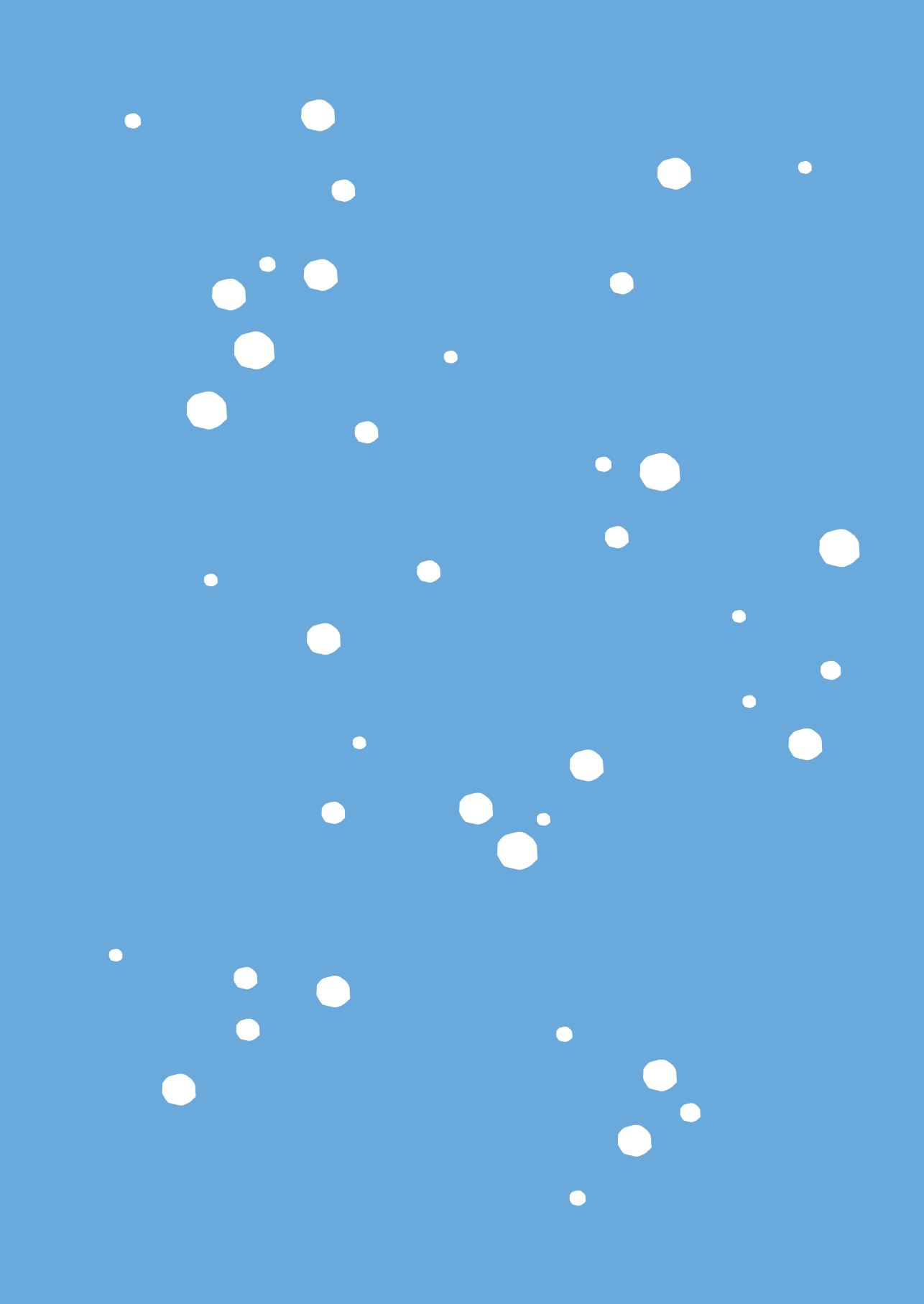
1. Meijer L, Verhoeven RHA, de Hingh IHJT, van de Wouw AJ, van Laarhoven HWM, Lemmens VEPP, et al. Extensive diagnostic work-up for patients with carcinoma of unknown primary. Clinical & Experimental Metastasis. 2021;38(2):231-8.
2. Hemminki K, Bevier M, Hemminki A, Sundquist J. Survival in cancer of unknown primary site: population-based analysis by site and histology. Annals of Oncology. 2012;23(7):1854-63.
3. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science. 1999;286(5439):531-7.
4. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502(7471):333-9.
5. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
6. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
7. Polak P, Karlić R, Koren A, Thurman R, Sandstrom R, Lawrence MS, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature. 2015;518(7539):360-4.
8. Jiao W, Atwal G, Polak P, Karlic R, Cuppen E, Al-Shahrour F, et al. A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns. Nature Communications. 2020;11(1):728.
9. Chakraborty S, Martin A, Guan Z, Begg CB, Shen R. Mining mutation contexts across the cancer genome to map tumor site of origin. Nature Communications. 2021;12(1):3051.
10. Nguyen L, van Hoeck A, Cuppen E. Machine learning-based tissue of origin classification for cancer of unknown primary diagnostics using genome-wide mutation features. bioRxiv. 2021:2021.10.05.463244.
11. Lombardo R, Tosi F, Nocerino A, Bencardino K, Gambi V, Ricotta R, et al. The Quest for Improving Treatment of Cancer of Unknown Primary (CUP) Through Molecularly-Driven Treatments: A Systematic Review. Front Oncol. 2020;10:533-.
12. Park C, Georlette D, Korn WM, Xiu J, Babiker H, Coelho Barata PM, et al. 1139P Carcinoma of unknown primary (CUP): The role of tumor genomic profiling. Annals of Oncology. 2021;32:S926.
13. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575(7781):210-6.
14. Samsom KG, Bosch LJW, Schipper LJ, Roepman P, de Bruijn E, Hoes LR, et al. Study protocol: Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient (WIDE). BMC Medical Genomics. 2020;13(1):169.
15. Roepman P, de Bruijn E, van Lieshout S, Schoenmaker L, Boelens MC, Dubbink HJ, et al. Clinical Validation of Whole Genome Sequencing for Cancer Diagnostics. J Mol Diagn. 2021;23(7):816-33.

16. Schipper LJ, Monkhorst K, Samsom KG, Bosch LJW, Snaebjornsson P, van Boven H, et al. Clinical Impact of Prospective Whole Genome Sequencing in Sarcoma Patients. Cancers (Basel). 2022;14(2).
17. Arakawa S, Yoshida T, Shirasawa M, Takayanagi D, Yagishita S, Motoi N, et al. RB1 loss induced small cell lung cancer transformation as acquired resistance to pembrolizumab in an advanced NSCLC patient. Lung Cancer. 2021;151:101-3.
18. Hainsworth JD, Rubin MS, Spigel DR, Boccia RV, Raby S, Quinn R, et al. Molecular Gene Expression Profiling to Predict the Tissue of Origin and Direct Site-Specific Therapy in Patients With Carcinoma of Unknown Primary Site: A Prospective Trial of the Sarah Cannon Research Institute. Journal of Clinical Oncology. 2012;31(2):217-23.
19. Hayashi H, Kurata T, Takiguchi Y, Arai M, Takeda K, Akiyoshi K, et al. Randomized Phase II Trial Comparing Site-Specific Treatment Based on Gene Expression Profiling With Carboplatin and Paclitaxel for Patients With Cancer of Unknown Primary Site. J Clin Oncol. 2019;37(7):570-9.
20. Fizazi K, Maillard A, Penel N, Baciarello G, Allouache D, Daugaard G, et al. A phase III trial of empiric chemotherapy with cisplatin and gemcitabine or systemic treatment tailored by molecular gene expression analysis in patients with carcinomas of an unknown primary (CUP) site (GEFCAPI 04). Annals of Oncology. 2019;30:v851.
21. Palma NA, Ali SM, O'Connor J, Dutta D, Wang K, Soman S, et al. Durable Response to Crizotinib in a MET-Amplified, KRAS-Mutated Carcinoma of Unknown Primary. Case Rep Oncol. 2014;7(2):503-8.
22. Kato S, Krishnamurthy N, Banks KC, De P, Williams K, Williams C, et al. Utility of Genomic Analysis In Circulating Tumor DNA from Patients with Carcinoma of Unknown Primary. Cancer Research. 2017;77(16):4238.
23. Subbiah IM, Tsimberidou A, Subbiah V, Janku F, Roy-Chowdhuri S, Hong DS. Next generation sequencing of carcinoma of unknown primary reveals novel combinatorial strategies in a heterogeneous mutational landscape. Oncoscience. 2017;4(5-6):47-56.
24. Tan DSW, Montoya J, Ng Q-S, Chan K-S, Lynette O, Sakktee Krisna S, et al. Molecular Profiling for Druggable Genetic Abnormalities in Carcinoma of Unknown Primary. Journal of Clinical Oncology. 2013;31(14):e237-e9.
25. Zhao P, Peng L, Wu W, Zheng Y, Jiang W, Zhang H, et al. Carcinoma of Unknown Primary with EML4-ALK Fusion Response to ALK Inhibitors. The Oncologist. 2019;24(4):449-54.
26. Pauli C, Bochtler T, Mileshkin L, Baciarello G, Losa F, Ross JS, et al. A Challenging Task: Identifying Patients with Cancer of Unknown Primary (CUP) According to ESMO Guidelines: The CUPISCO Trial Experience. Oncologist. 2021;26(5):e769-e79.
27. Samsom K, Monkhorst K, Schipper LJ, Roepman P, Bosch LJW, de Bruijn E, et al. Feasibility of whole-genome sequencing in routine clinical practice. Journal of Clinical Oncology. 2021;39(15_suppl):3013-.
28. Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature. 2020;578(7793):102-11.

29. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020;579(7800):567-74.
30. Penson A, Camacho N, Zheng Y, Varghese AM, Al-Ahmadie H, Razavi P, et al. Development of Genome-Derived Tumor Type Prediction to Inform Clinical Cancer Care. JAMA Oncol. 2020;6(1):84-91.
31. Koelsche C, Schrimpf D, Stichel D, Sill M, Sahm F, Reuss DE, et al. Sarcoma classification by DNA methylation profiling. Nature Communications. 2021;12(1):498.
32. Lu MY, Chen TY, Williamson DFK, Zhao M, Shady M, Lipkova J, et al. AI-based pathology predicts origins for cancers of unknown primary. Nature. 2021;594(7861):106-10.
33. Möhrmann L, Werner M, Oleś M, Mock A, Uhrig S, Jahn A, et al. Comprehensive genomic and epigenomic analysis in cancer of unknown primary guides molecularly-informed therapies despite heterogeneity. Nat Commun. 2022;13(1):4485.

PART 2

Use of Whole Genome Sequencing in rare tumor types or to detect rare genetic events



CHAPTER 6

Molecular prognostic factors in small-intestinal neuroendocrine tumours

Kris G. Samsom, Linde M. van Veenendaal, Gerlof D. Valk, Menno R. Vriens, Margot E.T. Tesselaar & José G. van den Berg

Endocrine Connections, 2019

Abstract

Background

Small intestinal neuroendocrine tumours (SI-NETs) represent a heterogeneous group of rare tumours. In recent years, basic research in SI-NETs has attempted to unravel the molecular events underlying SI- NET tumorigenesis.

Aim

We aim to provide an overview of the current literature regarding prognostic and predictive molecular factors in patients with SI-NETs.

Method

A PubMed search was conducted on (epi)genetic prognostic factors in SI-NETs from 2000 until 2019.

Results

The search yielded 1522 articles of which 20 reviews and 35 original studies were selected for further evaluation. SI-NETs are mutationally quiet tumours with a different genetic make-up compared to pancreatic NETs. Loss of heterozygosity at chromosome 18 is the most frequent genomic aberration (44-100%) followed by mutations of CDKN1B in 8%. Prognostic analyses were performed in 16 studies, of which 8 found a significant (epi)genetic association for survival or progression. Loss of heterozygosity at chromosome 18, gains of chromosome 4,5, 7, 14 and 20p, copy gain of the SRC gene and low expression of RASSF1A and P16 were associated with poorer survival. In comparison with genetic mutations, epigenetic alterations are significantly more common in SI-NETs and may represent more promising targets in the treatment of SI-NETs.

Conclusion

SI-NETs are mutationally silent tumours. No biomarkers have been identified yet that can easily be adopted into current clinical decision making. SI-NETs may represent a heterogeneous disease and larger international studies are warranted to translate molecular findings into precision oncology.

Introduction

Well-differentiated neuroendocrine tumours (NETs) represent a heterogeneous group of rare tumours, which have a relatively indolent disease course. Primary NETs can arise from neuroendocrine cells at various anatomic sites. They most commonly develop in the gastrointestinal tract and bronchopulmonary system[1-3]. NETs can be classified as functional or non-functional, based on whether they cause clinical symptoms as a result of hormone secretion or not. In patients with metastatic small intestinal NETs (SI-NETs) the carcinoid syndrome is common, which is characterized by diarrhoea, episodic flushing, bronchospasm and often carcinoid heart disease leading to right valvular dysfunction[4]. Patients with non-functional SI-NETs are often asymptomatic or experience non-specific symptoms resulting in metastatic disease at time of diagnosis in 27 to 73% of patients[1-3]. In contrast to pancreatic NETs, SI-NETs are not known to arise in the context of hereditary syndromes, e.g. multiple endocrine neoplasia (MEN) type 1 or 2 and Von Hippel Lindau disease. The reported incidence of SI-NETs has increased over the last four decades, from 0.2 per 100,000 individuals in 1973 to 1.25 per 100,000 individuals in 2012[5]. This progressive rise can mainly be contributed to more frequent use and improvements of diagnostic modalities or alterations in pathological disease definition[2,5]. In the group of gastroenteropancreatic NETs, SI-NETs are second most prevalent after rectum NETs and followed by pancreatic NETs[5]. Moreover, SI-NETs are the most frequent cancer type of the small intestine[6]. Currently, treatment for patients with SI-NETs is based on the availability of several treatment modalities, e.g. surgery, liver directed therapies, somatostatin receptor analogues and Peptide Receptor Radionuclide Therapy rather than on precision medicine. In case of non-functional, advanced and progressive SI-NETs, Everolimus, targeting the P13K/AKT/mTOR (mammalian target of rapamycin) pathway, has demonstrated anti-proliferative effects[7-9]. However, there is no biomarker available that predicts response to Everolimus. To conclude, personalized treatment based on molecular profiling has not yet entered the arena of treatment modalities in advanced SI-NETs. In order to move towards precision medicine, the genomic landscape of SI-NETs has been under increasing investigation over the past years in the hope of unravelling the molecular events underlying NET tumorigenesis, facilitating the identification of novel therapeutic targets, rational (targeted) therapy management strategies and to improve prognosis. Recently, whole-genome sequencing of primary pancreatic NETs revealed several genomic events which characterize their pathogenesis and are associated with tumour progression[10]. In general, gene-expression based subtyping has led to new classifications of multiple tumour types. In contrast, the genomic landscape of SI-NETs remains poorly elucidated and biomarkers have not yet been identified.

Moreover, the genetic constitution of SI-NETs has been shown to differ compared to pancreatic NETs[11]. With this review we aim to provide the clinician treating SI-NETs with an overview of the recent studies evaluating molecular characteristics of SI-NETs and their predictive and prognostic significance.

Methods

A literature search was performed in PubMed in March 2019. As our main objective was to provide an up to date overview of the current literature regarding prognostic molecular factors in SI-NETs for clinicians treating patients with SI-NETs, we did not aim to perform a formal systematic review. The domain of this search consisted of adult patients with sporadic SI-NETs, the determinant of genetic or epigenetic alterations and the outcomes of prognosis, survival or progression. Synonyms of SI-NETs and (epi)genetic alterations with the outcome described as prognosis, survival and progression were used for the search. Search terms and syntax are described in detail in Table 1. Screening based on title and abstract was conducted by one reviewer, in case of uncertainties a second reviewer was consulted. Citation search of the included articles was performed to identify additional original studies.

Table 1. Search terms and syntax

Syntax in PubMed	((((((carcinoid[Title/Abstract]) OR ((((((tumor*[Title/Abstract]) OR tumour*[Title/Abstract]) OR neoplas*[Title/Abstract]) OR malignan*[Title/Abstract])) AND ((neuroendocrin*[Title/Abstract]) OR ((((((small[Title/Abstract] AND bowel[Title/Abstract])) OR ileal*[Title/Abstract]) OR jejun*[Title/Abstract]) OR duoden*[Title/Abstract]) OR midgut[Title/Abstract])))))) AND ((((((((((genom*[Title/Abstract]) OR epigenetic*[Title/Abstract]) OR gene*[Title/Abstract]) OR exom*[Title/Abstract]) OR chromosom*[Title/Abstract]) OR molecular*[Title/Abstract]) OR allel*[Title/Abstract])) OR sequenc*[Title/Abstract]) OR (((((methylation*[Title/Abstract]) OR mutation*[Title/Abstract]) OR alteration*[Title/Abstract]) OR amplificat*[Title/Abstract]) OR loss[Title/Abstract]))))) AND (((prognos*[Title/Abstract]) OR survival*[Title/Abstract]) OR progressi*[Title/Abstract]))
Search terms	'carcinoid', 'tumor', 'tumour', 'neoplasia'. 'malignan*', 'neuroendocrin*', 'small bowel', 'ileal', 'jejun*', 'duoden*', 'midgut', 'genom', 'epigenetic*, 'gene*', 'exom*',chromosom*', 'molecular*', 'allel*', 'sequenc*' 'methylation*', mutation*', alteration*', 'amplificat*', 'loss' 'prognos*', 'survival', 'progressi*)

Inclusion criteria consisted of patient populations >18 years old, human, full text available in English, published between 01/01/2000 and 01/03/2019 and studies on gastroenteropancreatic NETs. Studies with a patient population with underlying genetic syndromes, no separate genome analysis for SI-NETs, using previously published results and on the taxonomy of SI-NETs were excluded.

Results

The PubMed search yielded 1522 hits, of which 1461 articles were excluded after screening of title and abstract (Figure 1). Following the full text screening of 61 articles, 14 articles were excluded. The citation search identified 22 additional articles of which 7 were excluded. In total 55 relevant articles were found, consisting of 20 reviews and 35 original studies. The results of the selected original studies are shown in Table 1 (Appendix 2). Our review will discuss the most relevant studies, with a special focus on the prognostic implications of the identified molecular alterations. The identified studies describe different genomic events and altered expression of several proteins which play a key role in various molecular pathways involved in SI-NET tumorigenesis. Events which have been described in multiple studies and are discussed in this paper are shown in Figure 2.

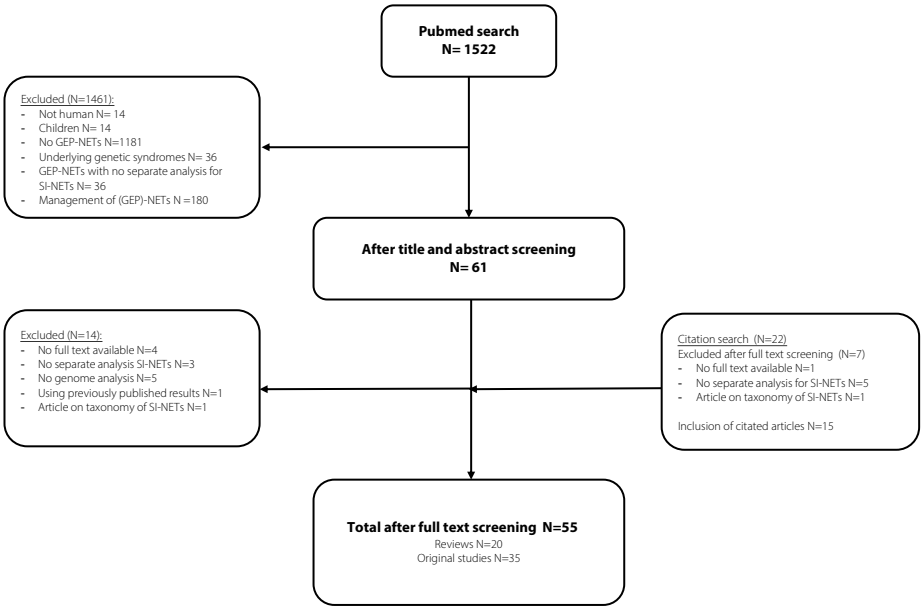


Figure 1. Flowchart of search and screening process in PubMed.

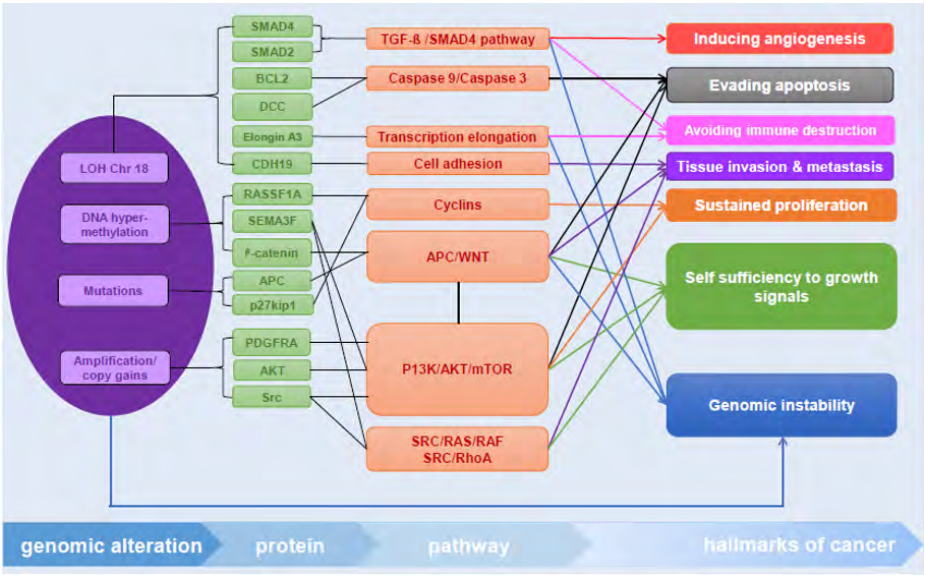


Figure 2. The studies presented in this review identified the deregulation of the expression of multiple genes in SI-NETs which are commonly associated with carcinogenesis in other tumors. In the figure above, only those molecular alterations that have been found in multiple studies are depicted, together with their presumed role as key regulators of different cell functions and their possible effect on tumor progression as defined by the hallmarks of cancer[55].

Table 2. Overview of studies.

Study	Publication year	No. of patients	Domain	Analysis technique
Löllgen et al.	2001	8	Metastatic midgut NETs (6 ileal, 1 ileocecal valve, 1 ascending colon)	Genome wide LOH* screening with microsatellite markers
Wang et al.	2005	47	Ileal NETs (n=16)	Micro-satellite markers, PCR amplification , sequencing of the BRAF gene
Zhang et al.	2006	33	SI-NETs with matched primary and metastatic tumours	Methylation-specific PCR, Western blot and immunochemistry
Kim et al.	2007	29	Well-differentiated ileal NETs (n=15)	Genome-wide high density single nucleotide polymorphism array analysis
Choi et al.	2007	35	Ileal NETs (n=15)	Pyrosequencing
Kulke et al.	2008	18	Primary and metastatic SI-NETs (n=24)	High resolution arrays of single nucleotide polymorphisms
Andersson et al.	2009	43	Ileal NETs	High resolution array based on comparative genomic hybridization
Cunningham et al.	2010	45 (37 sporadic and 8 familial)	Sporadic and familial ileal NETs (61 tumour samples)	High resolution genomic and gene expression profiling
Ruebel et al.	2010	8	Primary and metastatic ileal NETs	RT-PCR, miRNA expression assay, Northern blotting, in situ hybridization
Walsh et al.	2010	239 cases and 110 controls	Ileal NETs	Genome wide association study single nucleotide polymorphism genotyping

Molecular aberrations	Prognostic association	Remarks
Deletions on Chr* 18 in 88% of midgut NETs	-	Analysis included 1 colon NET
Allelic loss of both arms of chromosome 18 in 69%. No BRAF mutations were identified.	-	
Methylation of RASSF1A and CTNNB1 promoters more frequent in metastatic vs. primary tumors (p=0.013 and p=0.004, respectively)	-	-
Loss of Chr 18 in 67%, loss of Chr 21 or 21q in 13%.	No correlation between loss of Chr 18 and 21 with survival.	-
Hypomethylation of LINE-1 was greater in ileal NETs than in non-ileal and pNETs (p=0,047), and tumours with lymph node metastasis (p=0,02), Chr 18 loss (p=0,001) and RAS-association domain family I, isoform A gene methylation (p=0,02).	No survival differences were observed based on LINE-1 methylation levels.	-
Loss of Chr 18 in 61%, Chr 9 in 33% and Chr 16 in 22%. Gains on Chr 4 (33%) and Chr 5,7,20 and 14q (17%)	-	-
Loss of Chr 18 in 74%. Other frequent copy number alterations were gain of Chr 4 (30%),5 (28%),14 (23%) and 20 (33%), and loss of 11q22.1-q22.2, 11q22.3-q23.1 and 11q23.3 (21%), and loss of 16q12.2-q22.1 and 16q23.2-qter (16%).	Gain of Chr 14 was a strong predictor of poor survival (p<0,001; HR 8,39 (95% CI: 3.04-23.11)). Loss of 3p13 resulted in a reduced risk of death (p=0,028; HR 0.14 (95% CI: 0.02-0.8)). Significant correlation between gain of Chr 7,14 and 20, and loss of 18 and overall survival (p<0.05).	-
Chr 18 aberrations in both sporadic and familial ileal NETs (100% vs. 38%). Frequent gain of Chr 7 in metastasis vs. primary tumour (16% vs. 0%)	No difference in overall survival in patients with or without a gain of Chr 14.	-
Downregulation of miRNA-133a (ratio 0,27*), -145 (ratio 0,33*), -146 (ratio 0.36*), -222 (ratio 0.41*) and -10b (ratio 0.44*) in 100% of primary vs. matching metastasis, upregulation of miRNA-183 (ratio 1.99*), -488 (ratio 1.56*), and -19a+b (ratio 1.31*) in 75% metastatic ileal NETs compared to primary tumours.	-	*Mean metastatic/ primary tumours ratio <1 (upregulated in primary tumours) *Mean metastatic/ primary ratio tumours ratio >1 (downregulated in primary tumours)
No single nucleotide variants significantly associated with ileal NETs, rs2208059 in KIF16B had a trend towards statistical significance. 14/226 cases (6.19%) and 2/97 controls (2.06%) heterozygous copy number deletions at 18q22.1.	-	-

Table 2. Continued

Study	Publication year	No. of patients	Domain	Analysis technique
Edfeldt et al	2011	19	SI-NETs (n=18), lymph node metastases (n=17), liver metastases (n=7)	Gene expression arrays, qPCR
Stricker et al.	2012	58	SI-NETs (n=17)	Pyrosequencing
Banck et al.	2013	48	SI-NETs	Exome sequencing
Francis et al.	2013	180, including 48 from Banck et al.	SI-NETs	Exome and genome sequencing
Li et al.	2013	24	SI-NETs (5 primary tumours, 5 mesentery metastasis, 5 liver metastasis)	Affymetrix Genechip miRNA array, qRT-PCR, Northern Blot Analysis
Hashemi et al.	2013	30	SI-NETs (n=18) and metastases (n=12)	Comparative genome hybridization , qPCR
Bottarelli et al.	2013	30	Ileal NETs	DNA fragment analysis and sequencing of the mutation cluster region of the APC gene

Molecular aberrations	Prognostic association	Remarks
Three clusters of gene expression profiles were identified distinguishing primary tumors (11/18) from lymph node metastases (5/17) and a third group consisting of liver metastases (7/7), lymph node metastases (12/17) and primary tumours (7/8). The different profiles suggest changes in the development from primary tumour to metastases.	No association was found between group 1 and group 3 for indolent or progressive disease course (p=0.15)	-
LINE1 hypomethylation was detected in 82% of SI-NETs.	-	-
0,1 SNVs* per 10^6 nucleotides. No recurrent mutations in cancer genes. 197 protein altering SNVs affected multiple cancer genes, including FGFR2, MEN1, HOOK3, EZH2, MLF1, CARD11, VHL, NONO and SMAD1. Mutually exclusive amplification of AKT1 or AKT2 was the most common event in 16 patients with alteration of P13K/Akt/mTOR signaling.	-	-
Frameshift mutations of CDKN1B in 8% SI-NETs (8%; 95% CI 4.7-12.7), hemizygous deletions encompassing CDKN1B in 14% .	-	-
miRNA-96 ($p<0.01$ compared to mesentery metastasis (MM) and $p<0.001$ liver metastases (LM)), -182 (MM $p<0.05$, LM $p<0.001$), -183 (MM $p<0.001$, LM $p<0.01$), -196a (MM $p<0.001$, LM $p<0.001$) were upregulated during tumor progression. miRNA-31 (MM $p<0.05$, LM<0.05), -129-5p (MM $p<0.01$, LM $p<0.001$), -133a (MM $p<0.05$, LM <0.05) and -215 (MM $p<0.05$, LM $p<0.05$) were downregulated.	-	-
Loss of chromosome 18 in 70%. Copy number losses on chromosome 11 (23%), 16 (20%), and 9 (20%), with regions of recurrent copy number loss identified in 11q23.1-qter, 16q12.2-qter, 9pter-p13.2 and 9p13.1-11.2. Gains detected in chromosomes 14 (43%), 20 (37%), 4 (27%), and 5 (23%) with recurrent regions of copy number gain in 14q11.2, 14q32.2-32.31, 20pter-p11.21, 20q11.1-11.21, 20q12-qter, 4 and 5. Differences between primary tumours and metastases; loss of 16q (p=0.003) loss and gain of Chr 7 (p=0.016).	Gain in 20pter-p11.21 was associated with short survival (p=0.013). No other significant associations were observed between recurrent copy number alterations and survival.	-
APC gene mutations in 23%, of which missense (57%) and nonsense/frameshift (14%) mutations.	No association was found with tumour progression.	-

Table 2. Continued

Study	Publication year	No. of patients	Domain	Analysis technique
Edfeldt et al.	2013	43	SI-NETs	Gene copy number determination by PCR, real time quantitive RT-PCR, RNA interference, CpG methylation pyrosequencing
Fotouhi et al.	2014	33	SI-NETs (n=44)	Pyrosequencing, ELISA-based quantification of global DNA methylation, qRT-PCR
Verdugo et al.	2014	20	Matched primary SI-NETs (n=10) and their mesenteric lymph node metastases (n=10)	Human methylation 27 BeadChip array profiling
Norlen et al.	2014	15	Peritoneal carcinomatosis of SI-NETs (n=8) and controls (n=7)	Single nucleotide polymorphism array
Crona et al.	2015	200	SI-NETs (n=362)	Automated Sanger sequencing of the CDKN1B gene, immunohistochemistry
Maxwell et al.	2015	90	SI-NETs	Exome sequencing and CNV analysis by quantitive PCR

Molecular aberrations	Prognostic association	Remarks
One copy deletion in 89% SI-NETs with reduced Elongin A3 expression in 77%.	-	-
Methylation was seen in WIF1 (methylation index (MI) 50% , (16-92%)), RASSFA1 (MI 16% (1-69%)), CTNNB1 (MI 13% (4-34%)), CXCL14 (MI 14% (3-39%)), NKX2-3 (MI 10% (2-28%)), P16 (CDKN2A) (MI 4% (1-33%)), LAMA-1 (MI 10% (4-24%), and CDH1 (MI 8% (3-22%). APC (MI 3% (2-8%)), CDH3 (MI 6% (3-12%)),, HIC1 (MI 5% (1-12%)), P14 (CDKN2A) (MI 5% (2-17%)), SMAD2 (MI 4% (1-8%)), and SMAD4 (MI 3% (1-6%) had low levels of methylation. WIF1 methylation was significantly increased (p=0,001) and WIF1 expression was reduced in SI-NETs vs. normal references (p=0.003). WIF1, NKX2-3 and CXCL14 expression was reduced in metastases vs. primary tumors (p<0.02). Global methylation of LINE1 was reduced in tumors vs. normal references (65% vs. 75%), and was associated with loss of Chr18p and 18q (p=0.022, p=0.003, respectively)	Low expression of RASSF1A and P16 were associated with poor survival (p=0,045 and p=0,011, respectively). Gene-specific promoter methylation or global methylation did not influence survival.	
RUNX3, TP73 and CHFR were highly methylated (β value ≥ 0.9). At Chr 18q21-qter (β value > 0.7), SETBP1, ELAC1, MBD1, MAPK4, and TCEB3C were methylated including several members of the Serpin peptidase inhibitor family (SERPIN B3, SERPINB5).	SI-NETs with a higher methylation index had a more aggressive phenotype.	-
Two groups were identified, group A with a greater proportion of patients with PC (86%) than group B (25%), with LOH of the entire or major part of chromosome 18 in group A (75%) compared to limited LOH (75%) or no LOH (25%) in group B.		-
Mutations of CDKN1B in 8.5%. Inter- and intratumour heterogeneity at the CDKN1B locus was present (33% and 11% respectively). Expression of p27 did not correlate with CDKN1B mutation status. No differences in clinical characteristics between CDKN1B mutated and CDKN1B wild type tumor carriers were found.	No correlation was found between survival and CDKN1B mutation status (HR 0.76; 95% CI 0.36-1.57)	-
CDKN1B frameshift mutations in 3.5% of SI-NETs (95% CI: 1.1-9.8%), 1 patient had a hemizygous deletion of CDKN1B and 2 patients duplications (3.4%;95% CI 0.41-7.2%). Mutations of CDKN1B occurred in 6,9%.	-	-

Table 2. Continued

Study	Publication year	No. of patients	Domain	Analysis technique
Delgado Verdugo et al.	2015	7	SI-NETs	Whole exome capture, NGS*, high resolution SNP array, copy number variation analysis
Bollard et al.	2015	38	Ileal NETs	Immunochemistry, methylation specific PCR
Karpathakis et al.	2015	97	SI-NETs	Whole-genome or targeted CDKN1B sequencing, Human methylation 450 BeadChip array profiling, methylated DNA immunoprecipitation sequencing, CNV analysis, whole genome DASL* expression array profiling
Miller et al.	2016	90	Primary SI-NETs (n=28), adjacent normal small bowel (n=14), matched lymph node metastases (n=24), normal lymph node metastases (n=7), normal liver (n=2) and liver metastasis (n=15)	NanoString miRNA profiling, qRT-PCR, luciferase reporter assays and immunoblotting.
Andersson et al.	2016	33	Well-differentiated distal ileal NETs	Genome-wide sequencing
Dumanski et al.	2017	239	Sporadic (215) and familial (24) SI-NETs compared to three control cohorts with 35,688 subjects	NGS* of exome or whole genome DNA
Karpathakis et al.	2017	20	SI-NETs and matched liver metastasis	Human methylation 450 BeadChip array profiling, methylated DNA immunoprecipitation sequencing, whole genome DASL* expression array profiling

Molecular aberrations	Prognostic association	Remarks
Loss of Chr18 in 71% of SI-NETs. No tumor-specific somatic mutation was identified.	-	
SEMA3F expression was lost in 96% ileal NETs and all their metastases. SEMA3F loss of expression was associated with promoter gene methylation (no p-value provided).	-	-
Subgroup 1: chromosome 18 LOH, CDKN1B mutations, CIMP*, negativity. Subgroup 2: absence of arm-level CNVs, CIMP positivity Subgroup 3: multiple CNVs Epimutations were found at a recurrence rate up to 85%, and 21 epigenetically dysregulated genes were identified, including CDX1 (86%), CELSR (84%), FBP (84%), and GIPR (84%).	3 subgroups of SI-NETs with different PFS* (not reached at 10 years vs. 56 months vs. 21 months; p=0,04)	
miR-204-5p ($p=2.44x10^{-67}$), miR-7-5p ($p=2.57x10^{-144}$) and miR-375 ($p=6.30x10^{-67}$) were upregulated and miR-1 (p= 0.0004) and MiR-143-3 ($p=8.11x10^{-9}$) were downregulated in lymph node and liver metastases vs. primary tumours.	-	-
Loss of chromosome 18 in 65% and gains of chromosome 4,5,7,14 and 20 in 51%. Loss of CDKN1B in 8%. 3 subgroups were identified. The prostaglandin E receptor 2 (PTGER2) is the most activated in tumours of higher grade ($p=4.4x10^{-10}$) , whereas Forkhead box M1 (FOXM1) was the most activated regulator in tumors with gain of chromosome 14 ($p=2.5x10^{-4}$)	The largest subgroup (n=17) was characterized by longer survival (p<0,05) and higher expression of neuroendocrine markers, including SSTR2. Tumors with higher grade (G2/3) or gain of chromosome 14 were associated with shorter patient survival (p<0.05) and increased expression of cell cycle-promoting genes.	Analysis included 1 ileal NEC
A mutation in the MUTYH gene was significantly enriched in SI-NETs (both sporadic and familial) compared to controls (OR 5.09; 95% 1.56-14.74; p=0,0038)	-	-
SI-NET liver metastasis show Chr18 LOH in 79%. Amplification of Chr20 (42%), deletion of Chr19 (34%) and gain of 17q (21%) in liver metastasis. In liver metastasis enrichment of multiple cancer-related pathways was seen; P13K signaling events, ErbB1 downstream signaling, PDGFRβ signaling pathway and mTOR pathway (adjusted p<0.001). Using a previously defined panel of 21 epimutated genes, a trend of progressive dysregulation in liver metastasis compared to primary SI-NETs was observed.	-	-

Table 2. Continued

Study	Publication year	No. of patients	Domain	Analysis technique
Shi et al.	2017	267	SI-NETs (n=55)	Immunochemistry, CDKN1B sequencing
Nieser et al.	2017	148	SI-NETs	qRT-PCR, Western blot, immunochemistry, NGS*, SNP array analysis, miRNA analysis by qRT-PCR
Keck et al.	2018	12	Matched small bowel tissue, primary SI-NETs, liver metastases	RNA sequencing, Whole transcriptome microarrays, qPCR
Simbolo et al.	2018	52	SI-NETs	High-coverage target sequencing, qPCR, FISH, expression analysis of SRC gene, immunohistochemistry
Yao et al.	2019	89	SI-NETs (small intestine, jejunum, ileum, duodenum, cecum, CUP)	Whole exome and targeted sequencing

*Chr = chromosome, LOH = loss of heterozygosity, SNV = somatic single variants, PFS= progression free survival, CIMP = CpG island methylator phenotype, NGS= Next generation sequencing, DASL=cDNA-mediated Annealing, Selection, extension and Ligation), OR = odds ratio, CUP:= cancer of unknown primary, GI-NET = Gastrointestinal NET, WT=wildtype

Molecular aberrations	Prognostic association	Remarks
CDKN1B mutations in 10,9%. No clear association was found between CKDN1B mutation and protein expression.	A trend towards shorter overall survival associated with low expression of CDKN1B was observed (multivariate hazards ratio, 2.04;95% CI 1.06-3.93; p=0,03). CDKN1B mutation was not associated with survival.	-
Chr 18 LOH in 65%. Only DCC (deleted in colorectal cancer) revealed loss of/greatly reduced expression in 29%. No additional genetic or epigenetic alterations were present on Chr18.	Loss of CABLES did not correlate with survival.	-
Serial differential expression was validated in 7/10 genes, with several interacting members of the AKT, MYC, or MAPK3 pathways. Liver metastases had underexpression of PMP22 (p<0.001) High expression of SERPINA10 (primary p<0.001, liver metastases <0.001) and SYT13 (primary p<0.001, liver metastases <0.001) was characteristic of primary SI-NETs and liver metastases.	-	-
Mutations in CDKN1B (9.6%), APC and CDKN2C (each 7.7%), BRAF, KRAS, PIK3CA and TP53 (each 3.8%). Frequent allelic loss of 4 genes located on Chr 18 (BCL2, CDH19, DCC and SMAD4) in 44.2% and losses on chromosomes 11 (38%) and 16 (15%). Gains on chromosomes 4 (31%), 5 (27%), 14 (36%), and 20 (20%).	SRC gene copy number gains were associated with a poorer prognosis (p=0.0047)	
BCOR (5.6%) and CDKN1B (4.5%) most frequently mutated genes. LOH was present in approximately 50% and copy number gains of Chr 4,5,14 and 20 in >25%. Five distinct genomic clusters were identified (cluster 1: LOH of Chr 18, cluster 2: gain Chr 5 and 7, cluster 3: gain Chr 4,5,14 and 20, cluster 4: Chr 4,5,7,14 and 20, cluster 5: copy number gains across most chromosomes).	Lower generalized chromosomal instability in SI-NETs (n=55) was associated with longer survival compared to high CIN (n=38) with a PFS of 18.6 vs 9.2 months (HR 0,41; 95% CI 0.24-0.73; p=0.0021).	Analysis included CUP and cecum NETs.

Genetics of SI-NETs

Chromosomal aberrations

From genomic profile studies, two different groups of SI-NETs can be identified, one which is characterized by loss of heterozygosity (LOH) of chromosome 18 as an early event and the other group which has no alterations of chromosome 18 and shows clustered gains on chromosomes 4,5,7,14 and 20 [11-15]. Multiple studies reported loss of one copy of chromosome 18, with an incidence of 44 to 100% in primary SI-NETs[11-22]. Chromosome 18 harbours several candidate tumour suppressor genes, including DCC (deleted in colorectal cancer; involved in axon guidance), SMAD4 (Mothers against decapentaplegic homolog 4; TGF-B signal transduction), SMAD2 (Mothers against decapentaplegic homolog 2; TGF-B signal transduction) and TCEB3C (Transcriptional Elongation Factor B polypeptide 3C; encoding Elongin A3; RNA transcription). Banck et al., who performed whole-exome sequencing (WES) on 48 well-differentiated SI-NETs, found SMAD2 and SMAD4 monoallelic deletions in 21 tumours [23,24]. Edfeldt et al. (n=43) identified that in the majority of SI-NETs decreased expression of Elongin A3 (77%) was present and that the remaining TCEB3C gene was epigenetically silenced by DNA hypermethylation[25]. Nieser et al. (n=148) performed the first comprehensive study to identify chromosome 18 related events at genetic, epigenetic and gene/protein expression level, which only found DCC to be affected by the monoallelic loss of chromosome 18 [22]. In addition, Simbolo et al. (n=52) observed copy loss of multiple genes located on chromosome 18: CDH19 (Cadherin 19; cell adhesion; 46.2%), BLC2 (B-cell-lymphoma; regulation of cell death; 42.3%), DCC (42.3%) and SMAD4 (28.8%)[15]. The clinical significance of LOH of chromosome 18 has been evaluated in multiple studies, either focussing solely on LOH of chromosome 18 or as part of a molecular profile study. According to Andersson et al. (n=43) LOH of chromosome 18 is associated with worse overall survival[13]. In contrast, Kim et al. did not find a significant correlation between loss of chromosome 18 and survival[18]. Contrarily, Yao et al. (n=89) found that SI-NETs with low generalized chromosomal instability (CIN) (which consisted of a cluster with LOH of chromosome 18) displayed significantly longer median PFS than those with a high CIN (which consisted of 3 clusters with different combinations of gains of chromosome 4, 5,7,14, 20 and 1 cluster with copy number gains across most chromosomes). PFS in patients with a low CIN (n=55) was 18.6 vs. 9.2 months in high CIN (n=38) (HR; 0.41; 95% CI 0.24-0.73; p=0.0021)[11]. As described by the clusters of Yao et al., recurrent gains of chromosome 4,5,7, 14 and 20 are common in SI-NETs[11-15, 18, 19,21]. In two studies by Andersson et al., gain of chromosome 14 was seen in 6 of 32 well-differentiated SI-NETs and was associated with higher tumour grade and shorter survival (HR 8.39; 95% CI 3.04-23.11)[13,21]. However, Cunningham et al. (n=45) and Simbolo et al. (n=52)

could not corroborate these findings[15, 19]. Hashemi et al. (n=30) studied Copy Number Alterations (CNAs; gains and losses of areas of the chromosome) and reported an association between gain of 20pter-p11.21 and worse survival[14], which was also not confirmed by the findings of Simbolo et al.15. Generalized chromosomal instability seems to be a common feature of SI-NETs. This phenomenon could possibly be a reflection of diverse underlying defects in chromosomal maintenance that drive SI-NET development[11].

Mutational status

Banck et al. analysed 48 primary SI-NETs, predominantly grade 1, by massively parallel exome sequencing and detected a low mutation rate in the SI-NET genomes with an average of 0.1 somatic single-nucleotide variants (SNVs) per 10^6 nucleotides in the exome, suggesting that SI-NETs are mutationally quiet tumours[23]. No recurrent mutations in the 215 sequenced target genes were found. In the studied SI-NETS, 197 protein-altering SNVs were identified, affecting a multitude of cancer genes including FGFR2, MEN1, HOOK3, EZH2, MLF1, CARD11, VHL, NONO, FANCD2, SMAD1 and BRAF. In 29% of SI-NETs, there were genetic alterations in the P13K/AKT/mTOR pathway and mutually exclusive amplification of AKT1 or AKT2 were common. Amplifications were also observed at the PDFDR (Platelet Derived Growth Factor Receptor alpha) locus in 20.8% [23]. In a recent study by Simbolo et al. frequent copy gains were detected in AKT1 (30.8%) and PDGFRA (Platelet Derived Growth Factor Receptor alpha; 28.8%) as well. Furthermore, gains were present at the FOS gene (transcription factor subunit; 36.5%), KIT (involved in cell proliferation, survival, migration and differentiation; 28.8%) and KDR (Kinase Insert Domain Receptor; involved in VEGF signalling; 28.8%) genes[15]. Higher mutation rates in primary SI-NETs were associated with increased likelihood of recurrent liver metastases ($p<0.04$)[23]. In a study by Francis et al. (n=180) including 48 cases from Banck et al., heterozygous frame shift mutations of the cyclin-dependent kinase inhibitor 1B gene (CDKN1B) in 14 of 180 SI-NETs (8%; 95% CI 4.7-12.7%) were observed[24]. CDKN1B is located on chromosome 12 and encodes the protein $p27^{Kip1}$, a cyclin-dependent kinase inhibitor (CKI), whose main function is to control the progression from G1 to S phase in the cell cycle. The reported mutations in this putative tumour suppressor gene in SI-NETs are loss-of-function truncating mutations throughout the gene; no hotspot has been identified. A further study by Crona (n=200), confirmed the presence of CDKN1B mutations in 17 of 200 SI-NETs (8.5%); 95% CI 4.6-12.4%)[26]. Mutational status did not appear to correlate with protein expression of $p27^{Kip1}$ and no immediate detectable impact on clinical phenotype and survival was found[26]. Similarly, Shi et al. observed CDKN1B mutations in 10.9% of 55 SI-NETs and found no association between CKDN1B mutation, $p27^{kip1}$ expression and survival[27]. Only a trend towards shorter survival of patients with tumours exhibiting

low expression of p27^{kip1} (multivariate HR, 2.04;95% CI 1.06-3.93; p=0,03) was observed. Other studies found CDKN1B mutations in 4.5% to 9.6% of SI-NETs [11,15,28,29]. Furthermore, Yao et al., using whole exome and targeted panel sequencing on 89 SI-NETs from the RADIANT trials, found recurring mutations in BCOR (BCL-6-interacting corepressor) in 5.6% [11]. BCOR has interactions with histone deacetylases which are involved in regulation of gene expression through DNA methylation[11]. Another recent study, by Simbolo et al., performed targeted sequencing on 52 primary SI-NETs of which 34.6% showed somatic mutations[15]. APC (Adenomatous polyposis coli; WNT signalling pathway regulator) and CDKN2C (Cyclin Dependent Kinase Inhibitor 2C; cell growth regulator which controls cell cycle G1 progression) were found to be recurrently mutated in 7.7%. In addition, mutations were found in known oncogenes such as BRAF (involved in MAPK/ERK signalling pathway), KRAS (involved in RAS/MAPK signalling pathway), PIK3CA (Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha; involved in the P13K/AKT/mTOR pathway) and TP53 (tumour suppressor gene; regulator of cell proliferation and apoptosis) in 3.8% [15]. Previously, Bottarelli et al. (n=30) described APC gene alterations in 23% of ileal NETs[30]. A copy gain of the SRC gene (proto-oncogene; involved in cell signalling), which was present in 25% of SI-NETs, was associated with poorer prognosis (p=0.047)[15]. The SRC gene (cell signalling, cell cycle control and cell adhesion) was also the most commonly amplified oncogene (23%) in the study by Banck et al.[23]. Copy gains of the SRC gene could potentially be a novel prognostic biomarker, especially as whole genome sequencing becomes more widely adopted into clinical practice of SI-NETs.

Molecular alterations in primary tumours vs. metastases

Molecular differences between primary tumours vs. metastases can provide insight in the process of tumour progression. Cunningham et al. observed increased gains of chromosome 7 in metastases (30 mesenterial and 4 hepatic) vs. primary SI-NETs (16% vs. 0%)[19]. Correspondingly, Hashemi et al. reported frequent gain of 7q22.3-qter in metastases (12 regional and 7 distant; p=0.016) compared to primary tumours[14]. Loss of 16q12.2qter was more common in distant metastases vs. primary tumours (p=0.003)[14]. Karpathakis found LOH of chromosome 18 in 79% in liver metastases[31]. In the same study, amplification of chromosome 20 was found in 42%, deletion of chromosome 19 in 34% and gain of chromosome 17q in 21% of liver metastases[31]. Furthermore, at mRNA level, analysis of differentially expressed genes between liver metastases and primary tumours identified significant enrichment of multiple cancer related pathways overexpressed in liver metastasis, e.g. P13K signalling events, ErbB1 downstream signalling, PDGFRB signalling and the mTOR pathway (adjusted p<0.001)[31]. Keck et al demonstrated by RNA sequencing that liver metastases show underexpression of PMP22 (Peripheral Myelin Protein 22; integral membrane protein

involved in demyelinating disease and apoptosis) compared to the corresponding primary tumour (<0.001)[32]. Fotouhi et al. (n=33), found different expression levels of CXCL14 (chemokine CXC motif ligand 14; involved in cytokine activity and angiogenesis) mRNA in metastases compared to primary tumours (p=0.0016) which correlated with methylation status of the respective genes[33]. Furthermore, increased expression was found for mRNA encoding beta-catenin (involved in Wnt signaling pathway) in metastases compared with primary tumours (p=0.041); for mRNA encoding P16 (regulates entry into S-phase) in distant metastasis compared to primary tumours and regional metastases (p=0.015) and for mRNA encoding RASSF1A (involved in cell cycle regulation, apoptosis and migration), in regional metastases compared to primary tumours and distant metastases (p=0.008). Low mRNA expression of RASSF1A and P16 were each associated with short survival (p=0.045 and p=0.011, respectively)[33]. Using gene expression arrays, Edfeldt et al. were able to identify differentially expressed mRNA in SI NET metastases compared to primary tumours which resulted in the identification of 3 different gene expression clusters. However, these clusters did not correlate with tumour progression[34]. To conclude, dissemination of SI-NETs is associated with genomic events; yet the way in which these events contribute to tumour progression remains unclear.

Prognostic stratification based on loss of heterozygosity of chromosome 18, CDKN1B mutations, CpG island methylator phenotype and copy number variations

Karpathakis et al. identified different prognostic subgroups using hierarchical clustering. In a sophisticated large-scale integrated genomic analysis, including DNA methylation, gene expression and copy number variance (CNV) of 97 SI-NETs from a cohort of 85 patients they identified three molecular subtypes of SI-NETs using an integrated genome analysis[29]. The largest subgroup (55%) was defined by chromosome 18 LOH and is associated with the presence of CDKN1B mutations, and CpG island methylator phenotype (CIMP) negativity. The CpG island methylator phenotype refers to the DNA hypermethylation of promoter-associated CpG islands of tumour suppressor and DNA repair genes, which leads to transcriptional silencing of these genes. These patients had the most favourable PFS (not reached at 10-year follow-up) after resection and a median age of 67 years at diagnosis. A second subgroup (18%) was characterized by the absence of arm-level CNVs (copy number variations that span the chromosomal arm) and a high degree of CIMP positivity. This group had an intermediate PFS (56 months) and a younger median age at diagnosis (60 years). The third subgroup consisted of 26% of SI-NETs and was characterized by multiple CNVs; these patients had a significantly poorer PFS (21 months) and were youngest at onset (54 years), suggesting a more aggressive clinical

phenotype. In accordance with Karpathakis et al., Yao et al. (n=89), identified similar prognostic groups regarding LOH of chromosome 18 and alterations in chromosome 4,5 and 20 [11]. However, Simbolo et al. classified their cohort (n=52) into the three molecular groups of Karpathakis et al. and did not observe any statistically significant correlation with prognosis (p=0.73)[15]. These results of Simbolo et al. could be due to the relatively small cohort in comparison to the cohort of Karpathakis et al. and do not necessarily weaken the findings of Karpathakis and Yao et al. Based on the findings of Karpathakis et al. and considering the relatively low frequency of somatic mutations in SI-NETs, it seems unlikely that mutations of the CDKN1B gene or LOH of chromosome 18 alone are driving the SI-NET tumorigenesis and suggests a greater role for epigenetic dysregulation[11,15,20,25,29,35].

Germline mutations in SI-NETs

A germline mutation is defined as a mutation which occurs in reproductive cells and therefore is incorporated in every cell of the offspring. A study by Dumanski et al. (n=239), sequenced germline DNA from 24 patients from 15 families with a history of SI-NETs and from 215 sporadic SI-NET patients[36]. A mono-allelic mutation causing an amino-acid substitution p. (Gly396Asp) in MUTYH was found to be significantly enriched in both patients affected with familial SI-NETs and in sporadic SI-NETs, compared to controls (minor allele frequencies 0.013 and 0.03, respectively) with an odds ratio of 5.09 (95% CI 1.56-14.74; p=0,0038). MUTYH encodes a DNA glycosylase, involved in repair of oxidative DNA damage in order to prevent mutation accumulation leading to tumorigenesis. Biallelic germline MUTYH mutations lead to multiple colorectal adenomas and carcinomas, referred to as MUTYH-associated polyposis (MAP), a recessive hereditary colorectal polyposis syndrome. Interestingly, MUTYH germline mutations were also found in pancreatic NETs[10,37]. By defective DNA repair, carriers with MUTYH mutations thus seem to have a predisposition to develop NETs of the pancreas or small intestine.

Epigenetics in SI-NETs

Epigenetic modification can be defined as a change in gene expression without alterations to the gene's DNA sequence itself[38]. Since SI- NETs appear to have relatively few somatic mutations, epigenetic dysregulation could play an important role in the tumorigenesis of SI-NETs and may have important clinical implications[23,24,35]. Epigenetic changes include DNA methylation, histone modifications and the actions of microRNA. Hyper- and hypomethylation and histone modifications modify gene expression, whereas MiRNAs, small single-stranded RNA molecules, regulate gene expression post-transcriptionally. These processes can be pharmacologically modified by targeting enzymes involved in

DNA methylation and histone modifications, and by miRNA inhibitors, thereby representing an appealing target for therapy[35,39]. In comparison with genetic mutations, epigenetic alterations are significantly more common and recurrent in SI-NETs. Our search yielded studies ranging from 8 to 97 patients that showed epigenetic alterations in SI-NETs e.g. DNA methylation changes in 65-82% of SI-NETs and multiple miRNA deregulations[29,40-42]. Several studies reported differences in methylation and miRNA patterns between primary tumours and metastases, suggesting a possible role in tumour development or progression. A recurrent event is the epigenetic silencing of RASSFIA (RAS-association domain family 1, isoform A gene; tumour suppressor gene inducing cell cycle arrest) expression by hypermethylation of its promotor. This event was observed by Choi et al., Zhang et al., and Fotouhi et al. and was more prominent in metastases than in primary tumours [33, 43,44]. In addition, increased hypermethylation of the CTNNB1 promoter was observed in liver metastasis compared to the corresponding primary tumours[44]. Promotor gene methylation was also found in a study by Bollard et al. (n=38); in 96% of ileal NETs and their metastases. The expression of the axon guidance molecule SEMA3F (Semaphorin 3F) was lost due to hypermethylation[45]. SEMA3F expression is a negative regulator of MAPK and mTOR signaling pathways. The first genome-wide DNA methylation analysis of SI-NETs, performed by Verdugo et al. in 10 SI-NETs and 10 matched mesenteric lymph node metastasis observed a high level of methylation in another gene set located at chromosome 18q21-qter[46]. In these patients, high methylation index correlated with more malignant behaviour. Karpathakis et al. found hypermethylation of the promoter region of the gastric inhibitory polypeptide receptor (GIPR; inhibits gastric secretion and gastrin release and stimulates insulin release) gene body in 74% of primary SI-NETs. Of note, in this study DNA methylation in SI-NETs was compared to the methylation status of normal intestinal mucosa which normally expresses GIPR whereas the methylation status of enterochromaffin cells in the small intestine is unknown. Progressive hypermethylation of this gene was seen in liver metastases compared to primary tumours[29,31].

MicroRNAs in primary tumours vs. metastases

Two miRNA profiling studies (n=8, n=24, respectively) comparing primary SI-NETs to its respective metastases found multiple miRNAs to be deregulated during tumour progression[40,47]. A downregulation of miRNA-133a and upregulation of miR-183 was consistently found in metastases vs. primaries. A study by Miller et al. (n=28) confirmed downregulation of miRNA133a, and found differential expression of several other miRNAs in SI-NETs and their metastases[48].

Discussion

SI-NETs are rare tumours with a relatively indolent course. Unfortunately, treatment options are limited with minimal survival benefit. Therapies targeting somatostatin receptors, expressed by the majority of SI-NETs, are only able to stagnate disease progression temporarily. In an attempt to identify prognostic factors and new effective targets for precision medicine, the genomic landscape of SI-NETs has been under increasing investigation in recent years. LOH at chromosome 18 remains the most frequent genomic aberration (44 to 100%) found in SI-NETs[11-22]. The tumour suppressor gene, CDKN1B is mutated in approximately 8% SI-NET patients[11,15, 26-29]. Interestingly, CDKN1B is regulated by menin, the protein that is defect in the majority of patients with the inheritable MEN 1 syndrome (75-80%). Moreover, in MEN1 patients without mutations in the gene encoding menin (20-25%), CDKN1B was shown to be inactivated in some individuals (3.6%)[49]. Thus, several SI-NETs and MEN1 associated endocrine tumours may share a common oncogenic pathway. Genetic alterations in the P13K/AKT/mTOR were found in primary SI-NETs and liver metastasis, providing a rationale for the use of mTOR inhibitors[15,23,31,45]. However, a correlation between efficacy and mTOR mutational status prior to commencement of therapy with mTOR inhibitors has not yet been established. Daskalakis et al. (n=27) recently tested the ex-vivo activity of several targeted kinase inhibitors and found great variability in ex vivo sensitivity for most drugs, emphasizing the need for predictive biomarkers which could support clinical decision making[50]. Furthermore, mutations in APC, CDKN2C (both 7.7%) and BRAF, KRAS, PIK3CA and TP53 (each 3.8%) were recently identified in SI-NETs[15]. An association of (epi)genetic aberrations with prognosis was found in 16 of the 35 original studies reviewed. Karpathakis et al. (n=97) and Yao et al. (n=89) identified molecular subtypes of SI-NETs with significant difference in PFS[11,31]. However, validation of these subgroups in an independent and larger cohort is required before translation into clinical practice is possible. A gain of chromosome 14 and 20pter-p11.21 was associated with shorter survival in two studies ($p<0.001$, $p<0.013$ respectively)[13,14]. SRC copy number gains were associated with poorer prognosis ($p=0.047$)[11]. Epigenetic alterations such as specific promotor methylation and global methylation and their effect on prognosis are yet to be determined[33,41,43,44-46]. At present, predictive or prognostic biomarkers, which can be adopted into clinical practice, have not yet been established. Inactivated tumour suppressor genes ,which are found in SI-NETs, are generally unsuitable as targets since restoring the function of tumour suppressor genes is difficult to accomplish. Mutations in oncogenes, which should be easier to target, have only recently been described in small numbers in SI-NETs and thus far no clinical studies have been undertaken to target these mutations in SI-NETs. Of note, Alvarez et al. identified

the HDAC class inhibitor Etinostat as potent inhibitor of master regulatory activity for 42% of metastatic gastroenteropancreatic NETs, leading to the initiation of a clinical trial (NTC03211988)[9]. The low mutational burden found in SI-NETS may render these tumours less eligible for immunotherapy using immune checkpoint inhibitors because tumour mutational burden is an important determinant of clinical benefit to immune checkpoint blockade in most tumours. Additionally, the recently characterized tumour microenvironment in NETs, e.g. low expression for PD1 and PDL1 in SI-NETs, combined with a modest T-cell infiltrate, further tempers expectations regarding a response to the currently used PD1 and PDL-1 inhibitors, although this remains to be investigated[51]. More promising targets in SI-NETs may constitute the DNA methylation machinery. In comparison with genetic mutations, epigenetic alterations are significantly more common in SI-NETs. Specific genes such as RASSF1A, SEMA3F and CTNNB1 are hypermethylated in SI-NETs silencing their transcription[3,43-46]. RASSF1A hypermethylation is also observed in pancreatic NETs, lung NETs and thymic NETs whilst it is not found in appendiceal NETs[52]. During the last decades, an increasing number of drugs targeting DNA methylation and histone methylation have been developed and successfully tested pre-clinically which are currently in evaluation in phase I-III clinical trials[53]. Additionally, the more specific upregulation of miRNAs in SI-NETS as described above may provide actionable targets since multiple strategies for microRNA-based therapies are under investigation[54]. In this era of accumulating studies regarding the molecular background of SI-NETs, we felt there was an unmet need to provide the clinician with an overview of (epi)genetic alterations and explain their relevance in terms of prognosis and possible novel therapeutic options. Despite our efforts to perform an extensive and broad search, studies may have been missed due to its non-systemic character. Limitations of studies used in this review include relatively small and heterogeneous cohorts, different genomic analysis techniques and paucity of relation of (epi)genetic aberrations to clinical outcomes. The rarity of SI-NETs has hampered conducting sizable clinical trials involving large-scale integrated genomic analysis. In the coming years, hopefully international collaborations will enable larger studies to be performed which correlate (epi)genetic alterations to clinical outcomes and aim to identify targetable (epi)genetic alterations. Larger studies combined with evolving molecular technologies, might lead to a more effective treatment strategy in which patients with specific molecular tumour profiles will be selected for targeted pharmaceutical interventions.

Conclusion

SI-NETs have a low mutational burden, whereas epigenetic alterations are more prevalent. Mutations as described in pancreatic NETs are generally not observed in SI-NETs. Several studies identified (epi)genetic subtypes and molecular profiles of SI-NETs with significant difference in progression free survival (PFS) and overall survival (OS). More research should be conducted to identify prognostic and predictive biomarkers that can be adopted in clinical decision making.

References

1. Yao J, Hassan M, Phan A, Dagohoy C, Leary C, Mares J, Abdalla E, Fleming J, Vauthey J, Rashid A et al. One Hundred Years After 'Carcinoid': Epidemiology and Prognostic Factors for Neuroendocrine Tumors in 35,825 Cases in the United States. Journal of Clinical Oncology. 2008;26(18):3063-3072.
2. Hallet J, Law C, Cukier M, Saskin R, Liu N, Singh S. Exploring the rising incidence of neuroendocrine tumors: A population-based analysis of epidemiology, metastatic presentation, and outcomes. Cancer. 2014;121(4):589-597.
3. Garcia-Carbonera R, Jilmenez-Fonseca P, Teulé A, Barriuso J, Sevilla I. SEOM clinical guidelines for the diagnosis and treatment of gastroenteropancreatic neuroendocrine neoplasms (GEP-NENs) 2014. Clinical and Translation Oncology.2014;16(12):1025-1034.
4. Hassan S, Banchs J, Iliescu C, Dasari A, Lopez-Mattei J, Yusuf S. Carcinoid heart disease. Heart. 2017;103(19):1488-1495.
5. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, Shih T, Yao J. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. JAMA Oncology.2017;3(10):1335.
6. Pan S, Morrisson H. Epidemiology of cancer of the small intestine. World Journal of Gastrointestinal Oncology.2011;3(3):33-42.
7. Zatelli M, Fanciulli G, Malandrino P, Ramundo V, Faggiano A, Colao A et al. Predictive factors of response to mTOR inhibitors in neuroendocrine tumours. Endocrine-Related Cancer. 2015;23(3):R173-R183.
8. Singh S, Carnaghi C, Buzzoni R, Pommier R, Raderer M, Tomasek J, Lahner H, Valle J, Voi M, Bubutishvili-Pacaud L et al. Everolimus in Neuroendocrine Tumors of the Gastrointestinal Tract and Unknown Primary. Neuroendocrinology. 2017;106(3):211-220.
9. Alvarez M, Subramaniam P, Tang L, Grunn A, Aburi M, Rieckhof G, Komissarova E, Hagan A, Bodei L, Clemons P et al. A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. Nature Genetics. 2018;50(7):979-989.
10. Scarpa A, Chang D, Nones K, Corbo V, Patch A, Bailey P, Lawlor R, Johns A, Miller D, Mafficini A et al. Whole-genome landscape of pancreatic neuroendocrine tumours. Nature. 2017;543(7643):65-71.
11. Yao J, Garg A, Chen D, Capdevila J, Engstorm P, Pommier R, van Cutsem E, Singh S, Fazio N, He W et al. Genomic profiling of NETs: a comprehensive analysis of the RADIANT trials. Endocrine-Related Cancer. 2019;26(4):391-403.
12. Kulke M, Freed E, Chiang D, Philips J, Zahrieh D, Glickman J, Shivdasani R. High-resolution analysis of genetic alterations in small bowel carcinoid tumor reveals areas of recurrent amplification and loss. Genes, Chromosomes and Cancer. 2008;47(7):591-603.
13. Andersson E, Swärd C, Stenman G, Ahlman H, Nillson O. High-resolution genomic profiling reveals gain of chromosome 14 as a predictor of poor outcome in ileal carcinoids. Endocrine-Related Cancer. 2009;16(3):953-966.

14. Hashemi J, Fotouhi O, Sulaiman L, Kjellman M, Höög A, Zedenius J, Larsson C. Copy number alterations in small intestinal neuroendocrine tumors determined by array comparative genomic hybridization. BMC Cancer. 2013;13(1):505.
15. Simbolo M, Vincentini C, Maffacini A, Fassan M, Pedron S, Corbo V, Mastracci L, Rusev B, Pedrazzani C, Landoni L et al. Mutational and copy number asset of primary sporadic neuroendocrine tumors of the small intestine. Virchows Archiv. 2018;473(6):709-717.
16. Löllgen R, Hessman O, Szabo E, Westin G, Akerström G. Chromosome 18 deletions are common events in classical midgut carcinoid tumors. International Journal of Cancer. 2001;92(6):812-815.
17. Wang G, Yao J, Worah S, White J, Luna R, Wu T, Hamilton S, Rashid A. Comparison of genetic alterations in neuroendocrine tumors: frequent loss of chromosome 18 in ileal carcinoid tumors. Modern Pathology. 2005;18(8):1079-1087.
18. Kim D, Nagano Y, Choi I, White J, Yao J, Rashid A. Allelic alterations in well-differentiated neuroendocrine tumors (carcinoid tumors) identified by genome-wide single nucleotide polymorphism analysis and comparison with pancreatic endocrine tumors. Genes, Chromosomes and Cancer. 2007;47(1):84-92.
19. Cunningham J, Díaz de Ståhl T, Sjöblom T, Westin G, Dumanski J, Janson E. Common pathogenetic mechanism involving human chromosome 18 in familial and sporadic ileal carcinoid tumors. Genes, Chromosomes and Cancer. 2010;50(2):82-94.
20. Delgado Verdugo A, Crona J, Maharjan R, Hellman P, Westin G, Björklund P. Exome Sequencing and CNV Analysis on Chromosome 18 in Small Intestinal Neuroendocrine Tumors: Ruling Out a Suspect?. Hormone and Metabolic Research. 2014;47(06):452-455.
21. Andersson E, Arvidsson Y, Swärd C, Hofving T, Wängberg B, Kristiansson E, Nillson O. Expression profiling of small intestinal neuroendocrine tumors identifies subgroups with clinical relevance, prognostic markers and therapeutic targets. Modern Pathology. 2016;29(6):616-629.
22. Nieser M, Henopp T, Brix J, Stoß L, Sitek B, Naboulski W, Anlauf M, Schlitter A, Klöppel G, Gress T et al. Loss of Chromosome 18 in Neuroendocrine Tumors of the Small Intestine: The Enigma Remains. Neuroendocrinology. 2016;104(3):302-312.
23. Banck M, Kanwar R, Kulkarni A, Boora G, Metge F, Kipp B, Zhang L, Thorland E, Minn K, Tentu R et al. The genomic landscape of small intestine neuroendocrine tumors. Journal of Clinical Investigation. 2013;123(6):2502-2508.
24. Francis J, Kiezun A, Ramos A, Serra S, Pedamallu C, Qian Z, Banck M, Kanwar R, Kulkarni A, Karpathakis A et al. Somatic mutation of CKDN1B in small intestine neuroendocrine tumors. Nature Genetics. 2013;45(12):1483-1486.
25. Edfeldt K, Ahmad T, Akerström G, Janson E, Hellman P, Stålberg P, Björklund P, Westin G. TBEB3C a putative tumor suppressor gene of small intestinal neuroendocrine tumors. Endocrine-Related Cancer. 2013;21(2):275-284.
26. Crona J, Gustavsson T, Norlén O, Edfeldt K, Åkerström T, Westin G, Hellman P, Björklund P, Stålberg P. Somatic Mutations and Genetic Heterogeneity at the CDKN1B Locus in Small Intestinal Neuroendocrine Tumors. Annals of Surgical Oncology. 2015;22(S3):1428-1435.

27. Shi Y, Qian Z, Zhang S, Li W, Masugi Y, Li T, Chan J, Yang J, Da Silva A, Gu M et al. Cell cycle protein expression in neuroendocrine tumors: association of CDK4/CDK6, CCND1, and phosphorylated retinoblastoma proliferative index. Pancreas. 2017;46(10):1347-1353.
28. Maxwell J, Sherman S, Li G, Choi A, Bellizzi A, O'Dorisio T, Howe J. Somatic alterations of CDKN1B are associated with small bowel neuroendocrine tumors. Cancer Genetics. 2015;208(11):564-570.
29. Karpathakis A, Dibra H, Pipinikas C, Feber A, Morris T, Francis J, Oukrif D, Mandair D, Pericleous M, Mohmaduvesh M et al. Prognostic Impact of Novel Molecular Subtypes of Small Intestinal Neuroendocrine Tumor. Clinical Cancer Research. 2015;2(1):250-258.
30. Bottarelli L, Azzoni C, Pizzi S, D'Adda T, Silini E, Bordi C, Rindi G. Adenomatous polyposis coli gene involvement in ileal enterochromaffin cell neuroendocrine neoplasms. Human Pathology. 2013;44(12):2736-2742.
31. Karpathakis A, Dibra H, Pinnikas C, Feber A, Morris T, Francis J, Oukrif D, Mandair D, Pericleous M, Mohmaduvesh M et al. Progressive epigenetic dysregulation in neuroendocrine tumour liver metastases. Endocrine-Related Cancer. 2017;24(2):L21-L25.
32. Keck K, Breheny P, Braun T, Darbro B, Li G, Dillon J, Bellizzi A, O'Dorisio T, Howe J. Changes in gene expression in small bowel neuroendocrine tumors associated with progression to metastases. Surgery. 2018;163(1):232-239.
33. Fotouhi O, Abdel Fahmideh M, Kjellman M, Sulaiman L, Höög A, Zedenius J, Hashemi J, Larsson C. Global hypomethylation and promoter methylation in small intestinal neuroendocrine tumors. Epigenetics. 2014;9(7):987-997.
34. Edfeldt K, Björklund P, Akerström G, Westin G, Hellman P, Stålberg P. Different gene expression profiles in metastasizing midgut carcinoid tumours. Endocrine-Related Cancer. 2011;18(4):479-489.
35. Di Domenico A, Wiedmer T, Marinoni, Perren A. Genetic and epigenetic drivers of neuroendocrine tumors (NET). Endocrine-Related Cancer. 2017;24(9):R315-R334.
36. Dumanski J, Rasi C, Björklund P, Davies H, Ali A, Grönberg M, Wellin S, Sorbye H, Grønbæk H, Cunningham J et al. A MUTYH germline mutation is associated with small intestinal neuroendocrine tumours. Endocrine-Related Cancer. 2017;24(8):427-443.
37. Raj N, Shah R, Stadler Z, Mukherjee S, Chou J, Untch B, Li J, Kelly V, Saltz B, Mandelker D et al. Real-Time Genomic Characterization of Metastatic Pancreatic Neuroendocrine Tumors Has Prognostic Implications and Identifies Potential Germline Actionability. JCO Precision Oncology. 2018;(2):1-18.
38. Holliday R. The inheritance of epigenetic defects. Science. 1987;238(4824):163-170.
39. Stålberg, P., Westin, G. and Thirlwell, C. Genetics and epigenetics in small intestinal neuroendocrine tumours. Journal of Internal Medicine. 2016;280(6):584-594.
40. Ruebel K, Leontovich A, Stilling G, Zhang S, Righi A, Jin L, Lloyd R. MicroRNA expression in ileal carcinoid tumours: downregulation of microRNA-133a with tumor progression. Modern Pathology. 2009;23(3):367-375.
41. Stricker I, Tzivras D, Nambiar S, Wulf J, Liffers S, Vogt M, Verdoodt B, Tannapfel A, Mirmohammadsadegh A. Site- and grade-specific diversity of LINE1 methylation pattern in gastroenteropancreatic neuroendocrine tumours. Anticancer Research. 2012;32: 3699–3706.

42. Finnerty B, Gray K, Moore M, Zarnegar R, III T. Epigenetics of gastroenteropancreatic neuroendocrine tumors: A clinicopathologic perspective. World Journal of Gastrointestinal Oncology. 2017;9(9):341.
43. Choi I, Estecio M, Nagano Y, Kim D, White J, Yao J, Issa J, Rashid A. Hypomethylation of LINE-1 and Alu in well-differentiated neuroendocrine tumors (pancreatic endocrine tumors and carcinoid tumors). Modern Pathology. 2007;20(7):802-810.
44. Zhang H, Rumilla K, Jin L, Nakamura N, Stilling G, Ruebel K, Hobday T, Erlichman C, Erickson L, Lloyd R. Association of DNA methylation and epigenetic inactivation of RASSF1A and beta-catenin with metastases in small bowel carcinoid tumors. Endocrine. 2006;30(3):299-306.
45. Bollard J, Massoma P, Vercherat C, Blanc M, Lepinasse F, Gadot N, Couderc C, Poncet G, Walter T, Joly M et al. The axon guidance molecule semaphorin 3F is a negative regulator of tumor progression and proliferation in ileal neuroendocrine tumors. Oncotarget. 2015;6(34): 36731-36745.
46. Verdugo A, Crona J, Starker L, Stålberg P, Akerström G, Westin G, Hellman P, Björklund P. Global DNA methylation patterns through an array-based approach in small intestinal neuroendocrine tumors. Endocrine-Related Cancer. 2013;21(1):L5-L7.
47. Li S, Essaghir A, Martijn C, Lloyd R, Demoulin J, Öberg K, Giandomenico V. Global microRNA profiling of well-differentiated small intestinal neuroendocrine tumours. Modern Pathology. 2013;26(5):685-696.
48. Miller H, Frampton A, Malczewska A, Ottaviani S, Stronach E, Flora R, Kaemmerer D, Schwach G, Pfragner R, Faiz O et al. MicroRNAs associated with small bowel neuroendocrine tumours and their metastases. Endocrine-Related Cancer. 2016;23(9):711-726.
49. Georgitsi M, Raitila A, Karhu A, van der Luijt R, Aalfs C, Sane T, Vierimma O, Mäkinen M, Tuppurainen K, Paschke R et al. Germline CDKN1B/p27Kip1Mutation in Multiple Endocrine Neoplasia. The Journal of Clinical Endocrinology & Metabolism. 2007;92(8):3321-3325.
50. Daskalakis K, Norlén O, Karakatsanis A, Hellman P, Larsson R, Nygern P, Stålberg P. Ex vivo activity of cytotoxic drugs and targeted agents in small intestinal neuroendocrine tumors. Endocrine-Related Cancer. 2018;25(4):471-480.
51. Da Silva A, Bowden M, Zhang S, Masugi Y, Thorner A, Herbert Z, Zhou C, Brais L, Chan J, Hodi F, Rodig S et al. Characterization of the Neuroendocrine Tumor Immune Microenvironment. Pancreas. 2018;47(9):1123-1129.
52. Karpathakis A, Dibra H, Thirwell C. Neuroendocrine tumours: cracking the epigenetic code. Endocrine-Related Cancer. 2013;20(3):R65-R82.
53. Ahuja N, Sharma A, Baylin S. Epigenetic Therapeutics: A New Weapon in the War Against Cancer. Annual Review of Medicine. 2016;67(1):73-89.
54. Chakraborty C, Sharma A, Sharma G, Doss C, Lee S. Therapeutic miRNA and siRNA: Moving from Bench to Clinic as Next Generation Medicine. Molecular Therapy – Nucleic Acids. 2017;8:132-143.
55. Hanahan D, Weinberg R. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-674.
56. Walsh K, Choi M, Oberg K, Kulke M, Yao J, Wu C, Jurkiewicz M, Hsu L, Hooshmand S, Hassan M et al. A pilot genome-wide association study shows genomic variants enriched in the nontumor cells of patients with well differentiated neuroendocrine tumors of the ileum. Endocrine-Related Cancer. 2010; 18(1):171-180.

57. Norlén O, Edfeldt K, Akerström G, Westin G, Hellman P, Björklund P, Stålberg P. Peritoneal carcinomatosis from small intestinal neuroendocrine tumors: clinical course and genetic profiling. Surgery. 2014;156(6):1512-1522.

CHAPTER 7

Driver mutations occur frequently in metastases of well-differentiated small intestine neuroendocrine tumours

Kris G. Samsom, Sonja Levy*, Linde M. Veenendaal*, Paul Roepman, Liudmila L. Kodach, Neeltje Steeghs, Gerlof D. Valk, M. Wouter Dercksen, Koert F.D. Kuhlmann, Wieke H.M. Verbeek, Gerrit A. Meijer, Margot E. T. Tesselaar & José G. van den Berg

* *Equally contributing second authors*

Histopathology, 2021

Abstract

Aims

This study aims to investigate the clinicopathological significance of driver mutations in metastatic well-differentiated SI-NETs

Methods and results

Whole genome sequencing (WGS) of 35 and next generation sequencing (NGS) of 8 metastatic SI-NETs was performed. Biopsies were obtained between 2015 and 2019. Tumours were classified using the 2019 WHO classification. WGS included assessment of somatic mutations in all cancer related driver genes, tumour mutational burden (TMB) and microsatellite status. NGS entailed a cancer hotspot panel of 58 genes. Our cohort consisted of 21% G1, 60% G2 and 19% G3 SI-NETs. Driver mutations were identified in approximately 50% of SI-NETs. In total 27 driver mutations were identified, of which 74% in tumour suppressor genes (e.g. *TP53, RB1, CDKN1B*) and 22% in proto-oncogenes (e.g. *KRAS, NRAS, MET*). Allelic loss of chromosome 18 (63%), complete loss of *CDKN2A* and *CDKN1B* (both 6%) and CDKN1B mutations (9%) were most common. Potential targetable genetic alterations were detected in 21% of metastasized SI-NETs. All tumours were microsatellite stable and showed low TMB (median 1.10, IQR 0.87-1.35). Ki67 proliferation index was significantly associated with the presence of driver mutations (p=0.015).

Conclusion

Driver mutations occur in 50% of metastasized SI-NETs and their presence is associated with high Ki67 proliferation index. The identification of targetable mutations render these patients potentially eligible for targeted therapy.

Introduction

Well-differentiated neuroendocrine tumours (NETs) represent a group of rare tumours characterized by a relatively indolent disease course. Well differentiated NETs harbour relatively few genomic mutations and are often characterized by changes in the methylation machinery [1]. Neuroendocrine carcinomas (NECs) in contrast, have an aggressive clinical course and a dismal prognosis. Gastro-entero-pancreatic (GEP) NECs share oncogenic pathways with adenocarcinomas and have a relatively high mutational burden. To illustrate, the genetic make-up of GEP NECs includes loss of heterozygosity of *APC, TP53* and *DCC* tumour suppressor genes as well as mutations in *TP53, KRAS* and *BRAF* genes, which are typical for gastrointestinal adenocarcinomas [2].

Neuroendocrine neoplasia (NEN) are graded according to the World Health Organisation (WHO) grading system as grade 1, 2 or 3, based on proliferation rate, as quantified by mitotic and Ki67 proliferation index. Until 2017, all NEN of the digestive tract with a Ki67 proliferation index >20% were classified as NEC, regardless of clinical disease course or tumour morphology. In 2016, it was observed that a group of well-differentiated neuroendocrine tumours of the pancreas displayed a Ki67 proliferation index > 20% [3]. These tumours were classified as grade 3 well-differentiated neuroendocrine tumours. This term was adopted by both the WHO classification of neuroendocrine tumours as published in 2017 and subsequently by the WHO classification of tumours of the digestive system [4], concerning all NEN arising throughout the gastrointestinal tract and the hepatopancreaticobiliary organs. Mutational status of neuroendocrine neoplasms is currently not integrated in the clinicopathological classification. At initial presentation, histological grades of well-differentiated NETs can vary from grade 1 to 3. It is now assumed that NETs can progress from grade 1 to grade 2 to grade 3. The factors underlying such progression are currently unknown. In contrast, to our knowledge progression of well-differentiated NETs to NECs has not been reported. At time of diagnosis, 27-73% of patients with small intestinal NETs (SI-NETs) have metastatic disease [5-7]. For patients with metastatic disease, treatment is based on the availability of several treatment modalities, e.g. somatostatin analogues, peptide radionuclide receptor therapy (PRRT) and liver directed therapies. These treatment modalities generally slow down clinical progression but do not provide curation for the disease. However, no therapies are currently available which specifically target genetic alterations in NETs. The present study aims to investigate the presence of driver mutations in metastatic SI-NETs and to explore the clinicopathological significance of these mutations, by investigating whether they are related to tumour characteristics such as tumour grade and whether they provide a rationale for targeted therapy.

Materials and methods

Patient cohort and study procedures

For the analyses, patients with metastatic SI-NETs were selected, whom were included under the study protocol (NCT01855447) of the Center for Personalized Cancer Treatment (CPCT). The CPCT-02 protocol was approved on the first of August 2011 by the medical ethical committee of the University Medical Center of Utrecht (NL35781.041.11) and was conducted in accordance with the Declaration of Helsinki. Patients were eligible for inclusion if the following criteria were met: (1) age ≥ 18 years; (2) locally advanced or metastatic solid tumour; (3) indication for new line of systemic treatment with registered anti-cancer agents; (4) safe biopsy according to the intervening physician. The biopsies analysed for this study were taken between April 2016 and February 2019. All patients (n=35) provided informed consent. The study procedures consisted of the collection of matched peripheral blood samples for reference DNA and image-guided biopsy of a single metastatic lesions.

Whole genome sequencing data

WGS was performed on fresh frozen samples. One or two biopsies were selected with no visible necrotic tissue and freeze sections were cut to ensure a sufficient tumour cell percentage (>20%). Collection and whole genome sequencing of samples at Hartwig Medical Foundation (HMF) was performed according to the standard procedures as described in detail previously by Priestley et al. [8]. All procedures at HMF are automated as much as possible and the Illumina® HiSeqX and NovaSeq6000 platforms are used for sequencing. During the process, shallow whole-genome sequencing is first used to determine an accurate tumour purity of the received and processed tumour samples before continuing full sequencing of the samples with sufficient tumour content (molecular tumour cell percentage >20%). Sequencing data is analysed with an optimized in-house bio-informatic pipeline designed to detect all types of somatic alterations, including single and multiple nucleotide substitutions (SNV and MNV), insertions and deletions (indels), copy number alterations (amplifications and gene copy losses) and genomic rearrangements and structural variants (e.g. gene fusions) in 508 cancer related driver genes (Appendix 1) [9]. Furthermore, tumour mutational burden (TMB) and microsatellite stability score are provided. The tumour mutational burden score represents the number of all somatic variants across the whole genome of the tumour per Mega base (Mb). Tumour mutational load is the total number of somatic missense variants across the whole genome of the tumour. Patients with a mutational load over 140 could be eligible for immunotherapy. The microsatellite stability score represents the number of somatic inserts and deletes in (short) repeat sections across the whole genome of the tumour per Mb. The score is considered as

a marker for instability in microsatellite repeat regions. Tumours with a score greater than 4.0 are considered microsatellite unstable (MSI). A comparison between the tumour biopsy and blood sample is performed to out germline polymorphisms and in order to be able to report somatic variants only. All code and scripts used for analysis of the WGS data are available via Github [10]. HMF has established procedures for WGS under ISO17025 accreditation. The genomic data is presented in a detailed molecular patient report which describes all variants which are relevant for cancer treatment decision making and gives a visual overview of the genomic data using CIRCOS plots. Appendix 2 provides more information on the interpretation of CIRCOS plots.

Clinical and WGS data

WGS data and corresponding clinical data were obtained from HMF under data request number DR-070 on the 5th of June 2019. Both WGS and clinical data are freely available for academic use from the HMF (https://www.hartwigmedicalfoundation.nl/) through standardized procedures and after approval by the Data Access Board. Germline data was not included in the request.

Clinical and NGS data

In routine diagnostic practice, there was an opportunity to perform NGS on 8 liver biopsies of metastasized SI-NETs. Biopsies were received between August 2015 and November 2019. In all patients, NGS with a cancer hotspot mutation panel of 58 genes was performed. All patients consented for the use of their clinical information according to the opt-out consent procedure at the Netherlands Cancer Institute. By default specific clinical information may be used for research, unless a patient explicitly states he or she objects.

Histopathology

Of all patients (n=43), diagnosis was confirmed by histopathological revision of representative slides, consisting at least of hematoxylin and eosin slides and the following immunohistochemical stainings: Ki67, chromogranin and synaptophysin. The slides were revised by an experienced NET pathologist (JB) using the criteria of the WHO classification of tumours of the digestive system 2019 to ensure only well-differentiated SI-NETs were included in this study and neuroendocrine carcinomas were excluded. For 16 patients (37%), slides stained for Somatosatin Receptor 2A (SSTR2A) were available for assessment.

Immunohistochemical stainings

Formalin-fixed paraffin embedded (FFPE) sections were obtained from biopsies and from resection specimens. Four um FFPE slides were immunohistochemically

stained using the following antibodies: anti-Chromogranin A (LK2H10) primary antibody (Roche, ready to use), Synaptophysin (27G12) (Leica/Novocastra, 1:50 to 1:100), Ki67 Antigen, MIB 1 Concentrate (Agilent/Dako, 1:100) and Recombinant Anti-Somatostatin Receptor 2 antibody (UMB1)-C-terminal (ab134152) (Abcam, 1:400 to 1:800). Immunochemistry was performed on BenchMark Ultra equipment (Ventana Medical System Inc., Tucson, AZ). Positive SSTR2A staining was defined as moderate to strong staining, including circumferential staining, essentially as described by Körner et al. [11] and Mehta et al. [12]. The proportion of stained tumour cells was expressed in percentages with increments of 10.

Statistical analysis

Patient and tumour characteristics and DNA sequencing results were described using descriptive statistics. Association between Ki67 proliferation index and presence of driver mutations was assessed using the Mann-Whitney U test. Disease specific survival (DSS) was defined as time from biopsy to disease specific death or date of follow-up. Patients alive or lost to follow-up were censored. DSS was analyzed using the Kaplan-Meier method. IBM SPSS v25 (SPSS Inc., Chicago, IL) was used to perform all statistical analysis.

Results

Patients and tumour characteristics

Baseline patient characteristics are shown in Table 1. Of this cohort (n=43), the median age at diagnosis was 61 years (IQR 56-67). Fifty-three percent of patients were male. Of the total of 43 tumours, 9 were grade 1 (21%), 26 grade 2 (60%) and 8 were grade 3 (19%). All tumours, irrespective of grade were 100% Synapthophysin and Chromogranin positive. SSTR2A expression was positive in all grade 1 tumours and ranged from 50% to 100% in the grade 2 and 3 tumours. There was no significant correlation between SSTR2A expression and mutational status (p=0.840).

Table 1. Baseline table for all 43 included patients with metastasized SI-NETs.

Patient characteristics	N=43
Median age at diagnosis (IQR)	61 (56-67)
Sex n(%)	
Total	43
Male	53
Female	47
Grade n(%)	
Total	43
G1	9 (21)
G2	26 (60)
G3	8 (19)

IQR: interquartile range

* According WHO classification of tumors of the digestive system 2019.

Whole genome sequencing

WGS data on 35 metastatic NET samples obtained from HMF revealed a total of 23 driver mutations in 17 patients (49%). Of all driver mutations (n=23), 17 (74%) were present in tumour suppressor genes (*e.g. TP53, RB1, ATM, CDKN1B, SMAD2*) and 5 (22%) in proto-oncogenes (*KRAS, NRAS, CTNNB1*). All tumours were microsatellite stable (microsatellite stability score. The tumour mutational burden and load of all tumours was low with a median of 1.098 variants per Mb (IQR 0.870-1.350) and a median of 21 (IQR 10.5-28.0), respectively. In Figure 1, TMB of all samples is shown. The above mentioned WGS findings are shown in Table 2 and 3. Allelic loss of chromosome 18 was present in 63% of tumours. Other recurrent events were complete loss of *CDKN2A* and *CDKN1B* (both 6%) and *CDKN1B* mutations (9%).

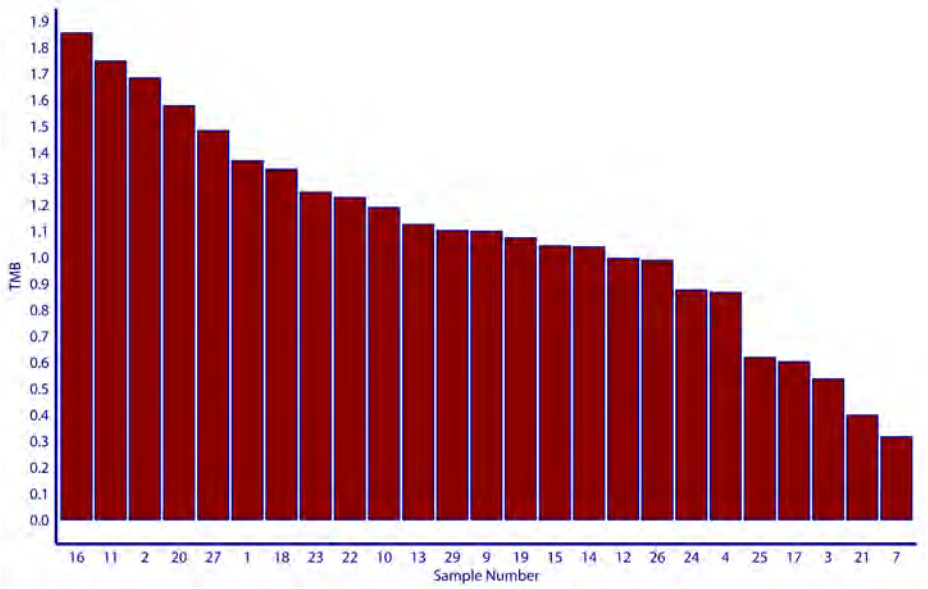


Figure 1. Tumour mutational burden per metastasized SI-NET sample. The TMB score represents the number of all somatic variants across the whole genome of the tumour per Mega base (Mb).

Table 2. Whole genome sequencing results for 35 patients with metastasised small-intestine neuroendocrine tumours

WGS results	N (%)/median (IQR)
Mutational status	
Total	35
Patients with driver mutations	17 (49)
Driver mutations	23
• Tumor suppressor genes	17 (74)
• Proto-oncogenes	5 (22)
• Unknown	1 (4)
No genomic aberrations	18(51)
Tumor mutational load (n=25)	21 (10.5-28.0)
Tumor mutational burden (variants per mb) (n=25)	1.098 (0.870-1.350)
Microsatellite status (n=25)	0.0311 (0.0233-0.0495)

IQR: interquartile range

Table 3. Mutations with high driver likelihood identified by whole genome sequencing

Samples	Driver mutations
Sample 4	SMAD2 c.1090C>T, p.Gln364 (TS) CDKN1B, c.92_03insCC, p.Leu32fs (TS)
Sample 9	TP53, c.19G>C, p.Asp7His (TS)
Sample 10	CDKN1B, c375_378delTGAG, p.GLu126fs (TS)
Sample 12	URB5, c.3622_3624delTGT, p.Cys1208del (TS)
Sample 14	KRAS, c.64C>A, p.Gln22Lys (PO)
Sample 15	SPEN, c.785C>A , p.Ala262Glu (TS)
Sample 16	DICER1, c.5113G>A, p.Glu1705Lys (TS)
Sample 17	PBRM1, c.4610A>G, p.Gln1537Arg (TS)
Sample 19	KMT2D, c.12667C>T, p.Gln4223 (TS) TCF7L2, c.1268A>G, p.Tyr423Cys (TS)
Sample 20	NRAS, c.37G>C, p.Gly13Arg (PO)
Sample 24	CTNNB1, c.110C>G, p.Ser37Cys (PO) PSIP1 (gene), c.283C>T (variant), p.Gln95 (impact) (?)
Sample 25	CDKN1B, c.280delC, p.Gly97fs (TS) ATM, c.5495_6496+2delAAGT, p.Glu1832fs (TS)
Sample 26	BCL9L, c.4283_4284dupTG, p.Thr1429fs (TS)
Sample 27	RB1, c.2357C>T, p.Arg787 (TS) PBRM1, c.2715_2718delGAGA, p.Glu908fs (TS)
Sample 28	GRIN2A, c.3321_3322insTTTTTTAATGATACGGC, p.Lys1107_ Thr1108insPhePheAsnASpThrAla (TS)
Sample 33	KRAS, A146V (PO) GNAS, R210H (PO)
Sample 35	CDKN1B, G97Vfs*22 (TS)

PO: proto-oncogene, TS: tumor suppressor gene

Next generation panel sequencing

In 8 patients NGS was conducted as part of routine diagnostic practice. NGS identified 4 tumours with driver mutations. The specific mutations of these tumours are shown in Table 4.

Table 4. Driver mutations identified by next generation panel sequencing

Samples	Driver mutations
Sample 40	CTNNB1 c.134C>T p.Ser45Phe (p.S45F) NM_001904.3 (PO)
Sample 41	TP53 [ENST00000269305.4]: codon 1-19, 21-257, 259-261, 263-394: c.1009C>T (p.Arg337Cys) (TS)
Sample 42	TP53 [ENST00000269305.4]: codon 1-19, 21-257, 259-261, 263-394: exon 5: c.404G>C (p.Cys135Ser) (TS) MET amplification of 6 amplicons
Sample 43	TP53 NM_000546.5 intron 4 c.376-1G>T p.? (p.?) (TS)

PO: proto-oncogene, TS: tumor suppressor gene

Association between driver mutations and Ki67 index

When comparing Ki67 proliferation index with the presence and absence of driver mutations, it was observed that patients with driver mutations had a significantly higher Ki67 index than those without driver mutations (p=0.015).

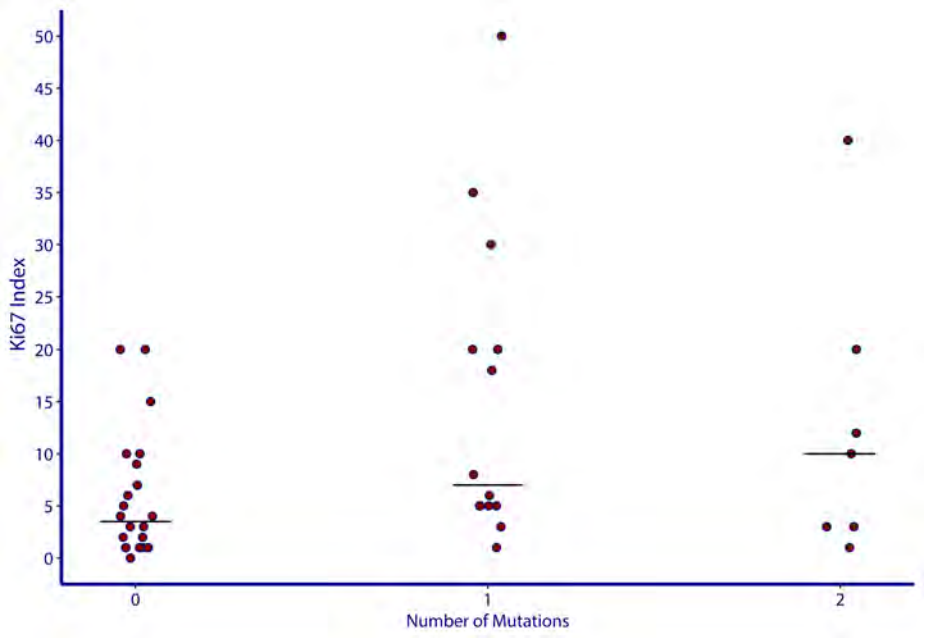


Figure 2. Number of mutations per patient compared to Ki67 proliferation index. The horizontal line represents the median Ki67 index per number of mutations. In case of 0 mutations the median Ki67 index was 3.5 (IQR 1.0-9.25), for 1 mutation the median Ki67 index was 7.0 (IQR 5.0-22.5) and for 2 mutations the median Ki67 index was 10.0 (IQR 3.0-20.0).

CIRCOS plots

In general, metastasized SI-NETs show little genomic aberrations resulting in a relatively empty CIRCOS plots. Figure 3 shows the histological features and CIRCOS plot of a grade 3 metastasized SI-NETs with driver mutations in KMT2D and TCG7KL2 and full loss of CDKN2A. In Figure 4, the histological features and a CIRCOS plot of a grade 2 metastasized SI-NET without driver mutations are shown. The CIRCOS plot of Figure 3 shows more genetic aberrations (e.g. somatic mutations, translocations and an amplification on chromosome 1) compared to the CIRCOS plot of Figure 4.

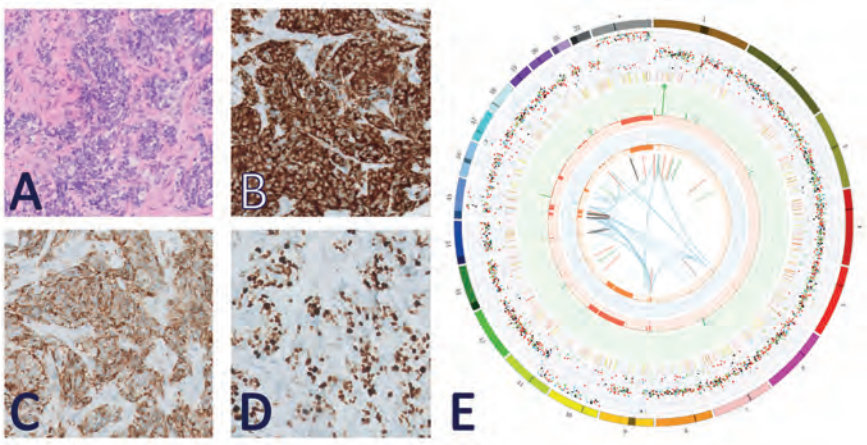


Figure 3. SI-NET grade 3 with driver mutations in KMT2D and TCF7L2 and CDKN2A loss. A: H&E staining, B: Synaptophysin staining: 100% positivity, C: Chromogranin staining: 100% positivity, D: Ki67 staining: 40%, E: CIRCOS plot.

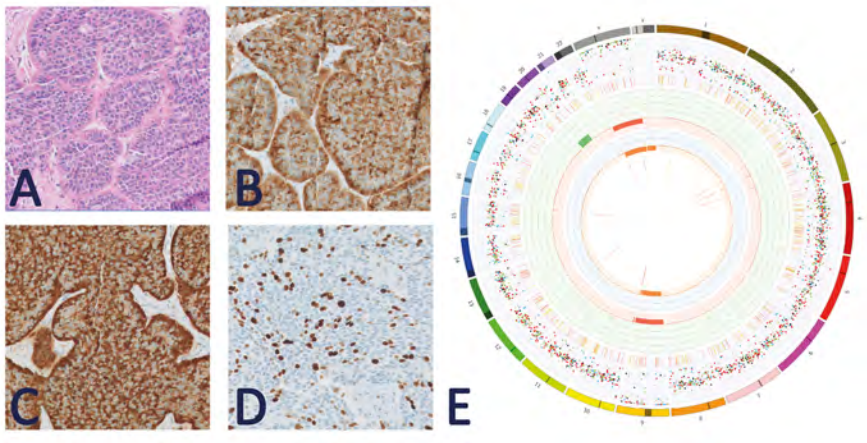


Figure 4. SI-NET grade 2 without driver mutations. A: H&E staining, B: Synaptophysin staining: 100% positivity, C: Chromogranin staining: 100% positivity, D: Ki67 staining: 10%, E: CIRCOS plot.

Targeted therapy

Potential actionable genetic alterations were detected in 9 (21%) patients in the BRCA pathway, the cyclin D/cyclin-dependent kinases 4-6 –retinoblastoma protein pathway, RAS/REF/MEK/ERK pathway and the HGF/MET pathway. These patients could be eligible for targeted therapy (off label). Table 5 shows potential actionable driver mutations and their potential targeted therapy.

Disease specific survival

After a median follow up of 25 (IQR 15-33) months, median disease specific survival was not reached as shown in Figure 5. Survival times did not differ significantly between patients with or without driver mutations (p=0.618) as is shown in Figure 6, nor a difference in DSS between tumour grades was seen (p=0.636).

Table 5. Potential actionable driver mutations and potential targeted therapy options

Actionable driver mutation	Potential precision drugs
CDKN1B	CDK4/6 inhibitors
KRAS	RAS/REF/MEK/ERK inhibitors
NRAS	RAS/REF/MEK/ERK inhibitors
GNAS	RAS/REF/MEK/ERK inhibitors
ATM	PARP inhibitors
MET amplification	MET inhibitors

ERK: Extracellular signal-regulated kinase, MEK: Mitogen-activated protein kinase, PARP: Poly(ADP-ribose) polymerase.

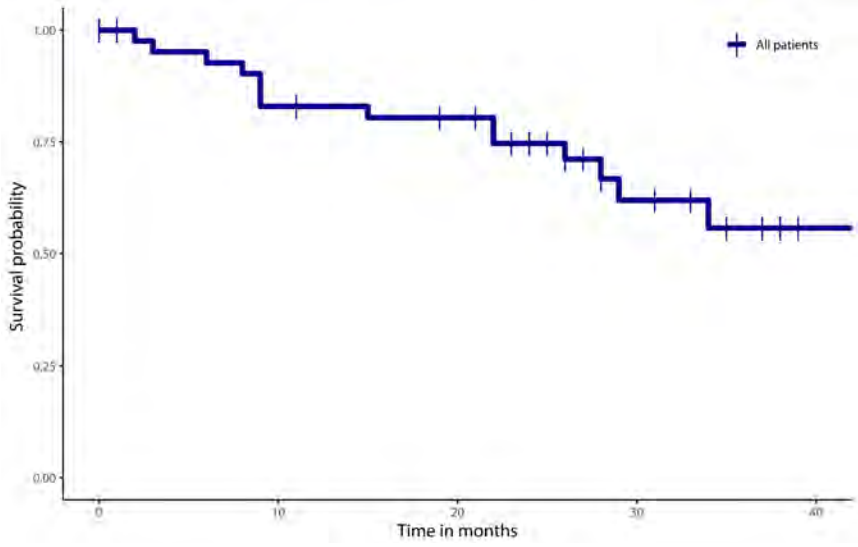


Figure 5. Disease specific survival for all patients in months, median DSS is not reached.

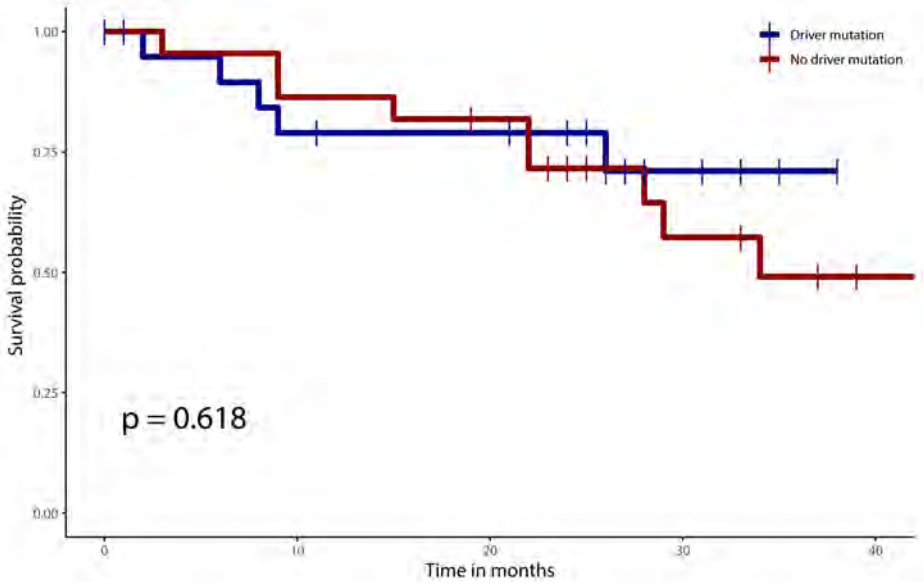


Figure 6. Disease specific survival in months of patients with driver mutations and without driver mutations.

Discussion

The recent WHO classification of NEN (2017) sharply distinguishes well-differentiated NET from poorly differentiated NEC. This distinction is based on pathologic (e.g. morphology and proliferation rate) and clinical features. At present, alterations in the genome of NEN do not contribute to the current classification, despite the paradigm shift in the classification of many other tumour types which has been caused by molecular subtyping in the past decade.

The distinction between NET and NEC has serious clinical implications in terms

of treatment and prognosis. To illustrate, advanced GEP NECs are treated with platinum based chemotherapy and have an overall survival of less than 12 months [13] whereas advanced GEP NETs are treated with multiple modalities and have an overall survival (largely dependent on primary tumour location) of approximately 33 months [14].

In this study, we aimed to investigate the presence of driver mutations in metastatic SI-NETs and to explore their clinicopathological significance. We show that well-

differentiated SI-NETs are mutationally quiet tumours with few genomic disruptions, which is in concordance with earlier studies (as reviewed in [15] in SI-NETs). Surprisingly, despite this low number of genomic disruptions, 50% of SI-NETs harbour driver mutations in cancer genes, including mutations in genes which are frequently affected in NEC, such as *TP53, RB1, KRAS* and *NRAS*. Our results are corroborated by WGS data of 25 well-differentiated SI-NETs of the MSK IMPACT study, which show complete loss of *CDKN2A* in 12% and driver mutations in 4.0% in *SMAD2, KRAS, RB1* and *TP53* [16]. This data was accessed through an open-access resource named cBioportal for Cancer Genomics (http://cbioportal.org) [17]. Of note, the biopsies which are included in this open-access resource are not reviewed by a pathologist whereas expert revision of all biopsies included in this study took place.

In our cohort, with a median follow up of 25 months, the presence of driver mutations did not affect disease specific survival which suggests that one or two driver mutations alone do not necessarily alter the clinical behaviour of metastasized SI-NETs. However, the identification of potential actionable genetic alterations in 21% of patients in our cohort is promising since it provides a rationale for the introduction of targeted therapy in the treatment of NET. For instance, in our study we found potential targets in the BRCA pathway, which would suggest that targeting DNA repair mechanisms may be effective in NET, e.g. through the use of Poly(ADP-ribose) Polymerase (PARP) inhibitors. Other targetable pathways included the cyclin D/cyclin-dependent kinases 4-6 –retinoblastoma protein, RAS/REF/MEK/ERK and the HGF/MET pathway. Furthermore, this study shows that SI-NETs have an invariably low tumour mutational load (median 21, IQR 10.5-28.0) and maintain chromosomal stability. In contrast to NECs, which have a high number of copy number alterations and a high mutational load [18,19]. Chromosomal stability and tumour mutational load therefore can be of practical aid in the distinction between NEC and NET. Loss of heterozygosity chromosome 18 was common in this cohort (63%), which is in accordance with earlier studies on primary SI-NETs (44-100%) [13,18-28]. Similarly, *CDKN1B* mutations occurred in 9%, which is also in accordance with earlier findings (4.5-11%) [20, 23, 30-35]. A complete loss of *CDKN2A* was found in 6% of SI-NETs. Loss of *CDKN2A* is an unspecific finding which is frequently encountered in metastasized solid tumours. In fact, *CDKN2A* has been identified in a pan cancer whole genome analysis of 2399 metastatic tumours as the most significantly deleted gene (n=415 (17%)) [8].

In conclusion, this study shows that well-differentiated metastasized SI-NETs do harbour driver mutations, which means that their presence is not exclusive to NECs. Consequently, the distinction between well-differentiated grade 3 NETs and poorly differentiated NECs should therefore not solely rely on the presence of driver

mutations, and rather be made on clinical and pathologic characteristics, such as a previous history of well-differentiated NET, a prolonged clinical course and well-differentiated morphology. The relationship between Ki67 proliferation index and the presence of driver mutations may suggest that these mutations may have contributed to tumour progression, i.e. progression from low to higher grade NET. However, this progression is not reflected in a decrease in disease specific survival and only in some patients by incomplete loss of SSTR2A expression. Our data support the notion that NET and NEC are two different disease entities and that progression of well-differentiated NET into poorly differentiated NEC is unlikely to occur.

References

1. Scarpa A. The landscape of molecular alterations in pancreatic and small intestinal neuroendocrine tumours. Annales d'Endocrinologie 2019;80(3):153-158.
2. Woischke C, Schaaf C, Yang H et al. In-depth mutational analyses of colorectal neuroendocrine carcinomas with adenoma or adenocarcinoma components. Modern Pathology 2016;30(1):95-103.
3. Tang L, Basturk O, Sue J, Klimstra D. A Practical Approach to the Classification of WHO Grade 3 (G3) Well-differentiated Neuroendocrine Tumor (WD-NET) and Poorly Differentiated Neuroendocrine Carcinoma (PD-NEC) of the Pancreas. The American Journal of Surgical Pathology 2016;40(9):1192-1202.
4. Klimstra D, Kloppel G, La Rosa S, Rindi G. Classification of neuroendocrine neoplasms of the digestive system. WHO Classification of Tumours: Digestive System Tumours, 5th ed, WHO Classification of Tumours Editorial Board (Ed), International Agency for Research on Cancer, Lyon 2019;16.
5. Yao J, Hassan M, Phan A et al. One hundred years after 'carcinoid' epidemiology and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. Journal of Clinical Oncology 2008;26:3063-3072.
6. Korse C, Taal B, van Velthuysen M, Visser O. Incidence and survival of neuroendocrine tumours in the Netherlands according to histological grade: Experience of two decades of cancer registry. European Journal of Cancer 2013;49(8):1975-1983.
7. Hallet J, Law C, Cukier M, Saskin R, Liu N, Singh S. Exploring the rising incidence of neuroendocrine tumours: a population-based analyses of epidemiology, metastatic presentation, and outcomes. Cancer 2015;121:589-597.
8. Priestly P, Baber J, Lolkema M et al. Pan-cancer whole genome analyses of metastatic solid tumors. Nature 2019;575(7781):210-216.
9. Cameron D, Di Stefano L, Papenfuss A. Comprehensive evaluation and characterization of short read general-purpose structural variant calling software. Nature Communcations 2019;10(1).
10. Github [Internet]. Hartwig Medical Foundation. 2020 [cited 9 July 2020]. Available from: https://github.com/hartwigmedical/
11. Körner M, Waser B, Schonbrunn A et al. Somatostatin receptor subtype 2A immunohistochemistry using a new monoclonal antibody selects tumors suitable for in vivo somatostatin receptor targeting. American Journal of Surgical Pathology 2012;36(2):242-52
12. Metha S, de Reuver P, Gill P, et al. Somatostatin Receptor SSTR-2a Expression Is a Stronger Predictor for Survival Than Ki-67 in Pancreatic Neuroendocrine Tumors. Medicine (Baltimore) 2015;94(40):e1281.
13. Sorbye H, Welin S, Langer S et al. Predictive and prognostic factors for treatment and survival in 305 patients with advanced gastrointestinal neuroendocrine carcinoma (WHO G3): the NORDIC NEC study. Annals of Oncology 2013;24:152-160.
14. Modlin I, Oberg K, Chung D et al. Gastroenteropancreatic neuroendocrine tumours. Lancet Oncology 2008;9(1):61-72.

15. Samsom K, van Veenendaal L, Valk G, Vriens M, Tesselaar M, van den Berg J. Molecular prognostic factors in small-intestinal neuroendocrine tumours. Endocrine Connections 2019;8:906-922.
16. Zehir A, Benayed R, Shah R et al. Mutational landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 patients. Nature Medicine 2017; 23(6):703-713.
17. Cbioportal.org.(2020). cBioPortal for Cancer Genomics. [online]. Available at: https://www.cbioportal.org/study/summary?id=msk_impact_2017%2Ccoadread_dfci_2016%2Ccoadread_genentech%2Ccoadread_tcga%2Ccoadread_tcga_pub%2Ccoadread_tcga_pan_can_atlas_2018%2Ccoadread_mskcc%2Crectal_msk_2019%2Ccrc_msk_2017%2Ccoad_caseccc_2015%2Ccoad_cptac_2019 . [Accessed 7 Feb.2020].
17. Furlan D. Different Molecular Profiles Characterize Well-Differentiated Endocrine Tumors and Poorly Differentiated Endocrine Carcinomas of the Gastroenteropancreatic Tract. Clinical Cancer Research 2004;10(3):947-957.
18. Vijayvergia N, Boland P, Handorf E, et al. Molecular profiling of neuroendocrine malignancies to identify prognostic and therapeutic markers: a Fox Chase Cancer Center Pilot Study. British Journal of Cancer 2016;115:564-570.
19. Simbolo M, Vincentini C, Maffacini A et al. Mutational and copy number asset of primary sporadic neuroendocrine tumors of the small intestine. Virchows Archiv 2018;473:709-717.
20. Kulke M, Freed E, Chiang D et al. High-resolution analysis of genetic alterations in small bowel carcinoid tumor reveals areas of recurrent amplification and loss. Genes, Chromosomes and Cancer 2008;47**:**591–603.
21. Andersson E, Swärd C, Stenman G, Ahlman H, Nillson O. High-resolution genomic profiling reveals gain of chromosome 14 as a predictor of poor outcome in ileal carcinoids. Endocrine-Related Cancer 2009;16**:**953–966.
22. Hashemi J, Fotouhi O, Sulaiman L et al. Copy number alterations in small intestinal neuroendocrine tumors determined by array comparative genomic hybridization. BMC Cancer 2013;13(1)**:** 505.
23. Löllgen M, Hessman O, Szabo E, Westin G, Äkerström G. Chromosome 18 deletions are common events in classical midgut carcinoid tumors. International Journal of Cancer 2001;92(6):812–815.
24. Wang G, Yao J, Worah S et al. Comparison of genetic alterations in neuroendocrine tumors: frequent loss of chromosome 18 in ileal carcinoid tumors. Modern Pathology 2005;18(8)**:**1079–1087.
25. Kim D, Nagano Y, Choi I, White J, Yao J, Rashid A. Allelic alterations in well-differentiated neuroendocrine tumors (carcinoid tumors) identified by genome-wide single nucleotide polymorphism analysis and comparison with pancreatic endocrine tumors. Genes, Chromosomes and Cancer 2008;47(1)**:**84–92.
26. Cunningham J, Díaz de Ståhl T, Sjöblom T, Westin G, Dumanski J, Janson E. Common pathogenetic mechanism involving human chromosome 18 in familial and sporadic ileal carcinoid tumors. Genes, Chromosomes and Cancer 2011;50(2)**:**82–94.

27. Delgado Verdugo A, Crona J, Maharjan R, Hellman P, Westin G, Björklund P. Exome sequencing and CNV analysis on chromosome 18 in small intestinal neuroendocrine tumors: ruling out a suspect? Hormone and Metabolic Research 2014; 47(6)**:**452–455.
28. Andersson E, Arvidsson Y, Swärd C et al. Expression profiling of small intestinal neuroendocrine tumors identifies subgroups with clinical relevance, prognostic markers and therapeutic targets. Modern Pathology 2016;29(6)**:**616–629.
29. Nieser M, Henopp T, Brix J et al. Loss of chromosome 18 in neuroendocrine tumors of the small intestine: the enigma remains. Neuroendocrinology 2017;104(3)**:**302–312.
30. Francis J, Kiezun A, Ramos A et al. Somatic mutation of CDKN1B in small intestine neuroendocrine tumors. Nature Genetics 2013;45(12):1483–1486.
31. Crona J, Gustavsson T, Norlén O et al. Somatic mutations and genetic heterogeneity at the CDKN1B locus in small intestinal neuroendocrine tumors. Annals of Surgical Oncology 2015;22(S3)**:**1428–S1435.
32. Shi Y, Qian Z, Zhang S et al. Cell cycle protein expression in neuroendocrine tumors. Pancreas 2017;46(10): 1347–1353.
33. Maxwell J, Sherman S, Li G et al. Somatic alterations of CDKN1B are associated with small bowel neuroendocrine tumors. Cancer Genetics 2015;208(11)**:**564–570.
34. Karpathakis A, Dibra H, Pipinikas C et al. Prognostic impact of novel molecular subtypes of small intestinal neuroendocrine tumor. Clinical Cancer Research 2015;22(1): 250–258.

CHAPTER 8

RET Fluorescence In Situ Hybridization Analysis Is a Sensitive but Highly Unspecific Screening Method for *RET* Fusions in Lung Cancer

Teodora Radonic*, W.R.R. Geurts-Riele*, Kris G. Samsom, Guido M.J.M. Roemen, Jan H. von der Thüsen, Erik Thunnissen, Isabelle C. Meijssen, Hein F.B. M. Sleddens, Winand N.M. Dinjens, Mirjam C. Boelens, Karin Weijers, Ernst Jan M. Speel, Stephen P. Finn, Cathal O'Brien, Tom van Wezel, Danielle Cohen, Kim Monkhorst, Paul Roepman & H.J. Dubbink

* *Equally contributing first authors*

Journal of Thoracic Oncology, 2021

Abstract

Introduction

RET gene fusions are established oncogenic drivers in 1% of non-small cell lung cancer (NSCLC). Accurate detection of advanced patients with *RET* fusions is essential to ensure optimal therapy choice. We investigated performance of fluorescence in situ hybridization (FISH) as a diagnostic test for detecting functional *RET* gene fusions.

Methods

Between January 2016 and November 2019, 4873 NSCLC patients from six European cancer centers were routinely screened for *RET* fusions using either FISH or targeted RNA next generation sequencing (NGS). If sufficient material was available, positive cases were analyzed by both methods and multiple FISH assays. In an independent cohort of 520 NSCLC patients, whole genome sequencing (WGS) data were investigated for disruptive structural variations and functional fusions in the *RET, ALK* and *ROS1* loci.

Results

FISH analysis of 2858 patients showed a *RET* rearrangement in 62 cases; 36 cases could be double tested with RNA NGS and only 9/36 cases (25%) had a functional RET fusion. All 9 RET fusions detected by RNA NGS screening and double tested with FISH showed RET locus rearrangement. Of these 18 cases, 16 showed a split signal and two showed a complex rearrangement in FISH. In WGS the prevalence of functional fusions compared to all disruptive events was lower in the *RET* (4/9, 44%) compared to the *ALK* (27/34, 79%) and *ROS1* (9/12, 75%) loci.

Conclusions

FISH is a sensitive but unspecific technique for *RET* screening always requiring a confirmation using an orthogonal technique, due to frequently occurring rearrangements not resulting in functional fusions in NSCLC.

Introduction

RET gene fusions emerged in 2011 as potentially actionable drivers in non-small cell lung cancer (NSCLC)[1,2]. *RET* encodes a tyrosine kinase, and on a genomic rearrangement, a fusion oncogene can be formed resulting in a chimeric *RET* receptor tyrosine kinase with constitutive, ligand independent, dimerization and *RET* kinase activation[3]. *RET* oncogenic fusions drive tumorigenesis by triggering oncogenic signaling pathways and increased cell survival and proliferation[4]. These rearrangements are mutually exclusive with other oncogenic alterations including *EGFR, KRAS, ERBB2*, and *BRAF* mutations, and *ALK* and *ROS1* fusions, indicating that *RET* fusions are independent oncogenic drivers in lung cancer[5,6].

RET fusions are predominantly found in younger patients with no smoking history, who generally have a good prognosis when treated with conventional cancer therapy [7], and in particular have a good response when treated with pemetrexed-based chemotherapy [8]. The incidence of *RET* gene fusions in the literature varies approximately 1% of NSCLC and depends on age, smoking history, histologic subtype, and racial origin (0.7% in non-Asian population [9] and 1,3% in Asian population[10]).

Targeted therapy with multikinase *RET* inhibitors had objective response rates in approximately 30% of patients with *RET* fusions [11,12] , but lower than the rates known from *ALK* and *ROS1* small molecule inhibitors (approximately 60-70%)[13,14]. However, novel selective *RET* inhibitors, such as BLU-667 and LOXO-292, in early phase clinical trials [15] were found to have higher response rates of 68% with a manageable toxicity profile [11,16]. LOXO-292 has recently been FDA approved for treatment of advanced *RET*-driven lung and thyroid cancers, and BLU-667 may follow soon [17]. *RET* rearrangements have been reported as a resistance mechanism in patients with *EGFR* mutated cancer treated with tyrosine kinase inhibitors [18]. This underlines the importance to implement robust and practical screening methods to identify patients who are likely to benefit from *RET*-targeted therapy.

In many pathology laboratories fluorescence in situ hybridization (FISH) is currently used for screening of patients with NSCLC, because RET immunohistochemistry had been found to have low sensitivity and specificity [6,7,19] . Break apart FISH detects the separation of two FISH probes hybridizing against the 3' and 5' side of the *RET* gene, which is indicative for *RET* rearrangement. It is a fast technique and requires little tissue, but it does not provide information on the *RET* fusion partner. FISH screening may, therefore, result in false positive results, because all rearrangements in the *RET* locus are detected, independent of whether these result in a functional oncogenic

fusion or not. Another diagnostic modality is reverse -transcriptase polymerase chain reaction, which has a high specificity. This method is potentially less sensitive, because it can only detect *RET* fusions with known partner genes for which the test has been designed, but it is not able to detect *RET* fusions with unknown fusion partners or those not present in the test[20]. Recently, targeted RNA next generation sequencing (NGS) has become a novel diagnostic test to concurrently screen for gene fusions of multiple genes without the knowledge of the fusion partners [21]. With the increasing number of potentially actionable driver fusions, this is an elegant technique for multiplex testing.

In this study, we present the *RET* fusion screening yields of 4873 NSCLC from daily clinical practice of six European cancer centers using either FISH or targeted RNA NGS. Where possible, positive cases were tested with both techniques allowing a direct comparison of *RET* FISH and targeted RNA NGS in a subset of 39 patients. Whole genome sequencing (WGS) data of an independent patient cohort with NSCLC were used to substantiate our observations on rearrangements in the *RET* locus and its specific properties when compared with ALK and ROS1 loci.

Method section

Case selection

As part of routine molecular diagnostics in six European centers, driver negative NSCLC samples were screened in the period between 1 January, 2015, and November 1, 2019, for the presence of *RET* fusions using two different screening techniques: *RET* FISH (Amsterdam University Medical Center, Erasmus MC, Maastricht University Medical Center and Netherlands Cancer Institute) and targeted RNA NGS (St. James's Hospital and Leiden University Medical Center). If a case had a rearranged signal with FISH or a *RET* fusion was detected with targeted RNA NGS, the case was, in addition, tested by the other testing method, provided that sufficient tissue was available.

FISH analysis

Four different *RET* break apart FISH assays were used in the four centers: Vysis (Abbott Molecular, Abbott Park, IL), SureFISH (Agilent, Santa Clara, CA), Kreatech (Leica Biosystems, Wetzlar, Germany) and Cytocell Aquarius (Cytocell Cambridge, United Kingdom). FISH was performed on 4-mm thick formalin-fixed, paraffin-embedded (FFPE) tumor sections according to the manufacturer's instructions. Signals were evaluated in at least 50 tumor nuclei per specimen. Cases were considered positive

for *RET* rearrangement if 15% or more of cells displayed a split of at least more than one signal diameter between the 5' and 3' signals or a single 3' signal. A complex pattern was defined as a rearrangement with any pattern that could not be classified using the usual split or single 3'patterns[22]. A complex pattern was defined as a rearrangement with any pattern that could not be classified using the usual split or single 3' patterns. Complex patterns were considered potentially positive, and, on their finding, a confirmatory test was always initiated if sufficient tissue was available. Single 5' patterns were reported but considered clinically negative because of loss of the *RET* kinase domain[23]. Whenever sufficient tissue was available, a case with a rearranged signal was tested using at least one other FISH assay.

Targeted RNA NGS

Total nucleic acid or RNA was extracted from FFPE-processed material using either a (1) tissue preparation system (TPS) robot (Siemens), (2) RNeasy FFPE kit (Qiagen), (3) AllPrep FFPE kit (Qiagen), (4) Maxwell RSC RNA FFPE kit (Promega), or (5) FormaPure Total FFPE kit (Beckman). RNA NGS by Archer: Library preparation was performed using the Archer FusionPlex Lung or CTL panel library preparation kit (ArcherDx, Boulder, CO). The resulting libraries were sequenced on Ion Torrent (Ion S5 or S5 XL) or Illumina MiSeq sequencers. Raw sequence data were analyzed by using the Archer Analysis software package (version 5 or 6 ArcherDx).

RNA NGS by Oncomine: Library preparation was performed according to the Oncomine Focus Gene fusion kit protocol (Thermo Fisher Scientific, Waltham, MA) using manual library preparation, templating on Ion Chef instrument and sequencing using Ion 530 sequencing chemistry on the Ion Torrent S5 instrument with Torrent Suite Software version 5.10.1. (Thermo Fisher Scientific) as described previously[24]. Data were then uploaded to Ion Reporter 5.10.3.0. and analyzed by using the Oncomine Focus w2.4 Fusion analysis.

RNA NGS by Asuragen Quantidex: Library preparation was performed using the Asuragen Quantidex NGS RNA Lung Cancer Kit (Asuragen, Austin, Texas)[21]. The resulting libraries were sequenced on an Illumina MiSeq sequencer. Raw sequence data were analyzed by using the Quantidex NGS Reporter software package 3.0.3. No cases were retested using another RNA NGS technique.

Analysis of RET, ROS1 and ALK loci in whole genome sequencing data

Available WGS data [25] of 520 lung cancer samples (stage IV, all lines of screening and therapy) were used for detailed analysis of the presence of genomic rearrangements [26], including functional gene fusions and disruptive structural variants in *RET*, including

ALK and *ROS1*. Further details and scripts are available at Github (https://github.com/hartwigmedical/). All data were obtained from routine diagnostic reports and were anonymized before processing.

Results

In total, 4873 NSCLC were screened for the presence of *RET* fusions using either FISH or targeted RNA NGS (Figure 1). The cases, either rearranged in FISH or positive by targeted RNA NGS and that could be tested using both techniques (n = 39), are found in Table 1. The distribution of cases per technique is presented subsequently.

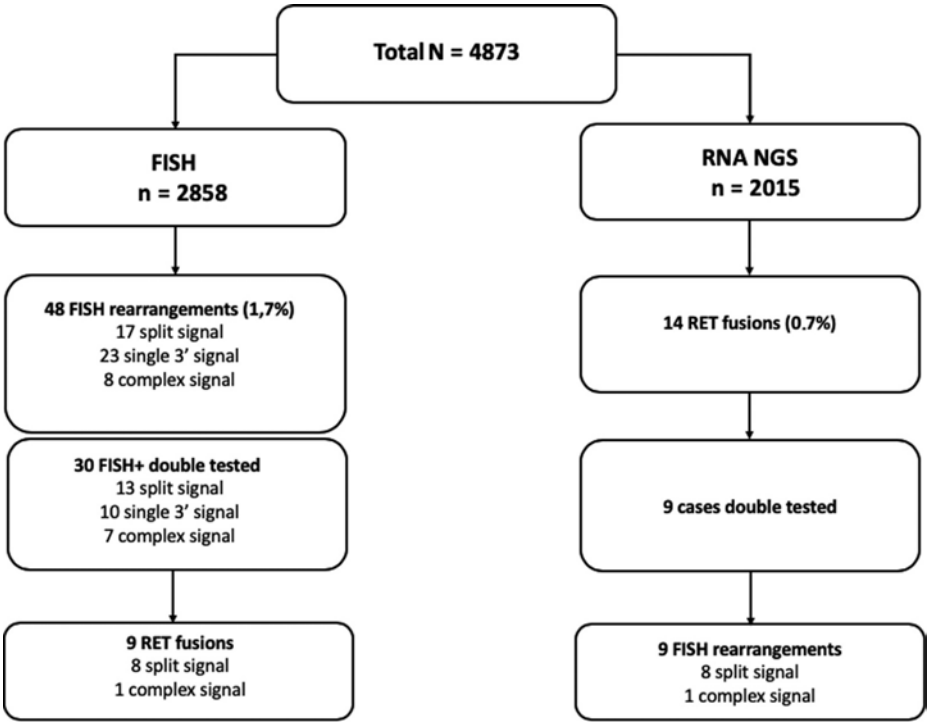


Figure 1. Total number of screened NSCLC. Of 30 cases with a rearrangement in FISH, only nine had a functional RET fusion in targeted RNA NGS. All nine cases detected by means of targeted RNA NGS were found to have a rearrangement in FISH. FISH, fluorescence in situ hybridization; NGS, next-generation sequencing.

Table 1. Cases tested using both FISH and targeted RNA NGS

Case	FISH	FISH pattern	Archer
1	Kreatech and Vysis	Split	KIF5B(ex15): RET(ex12)
2	Kreatech and Vysis	Split	KIF5B(ex15): RET(ex12)
3	Kreatech and Vysis	Split	KIF5B(ex15): RET(ex12)
4	Kreatech and Vysis	Split	CCDC6(ex1):RET(ex12)
5	Sure FISH	Split	CCDC6(ex1):RET(ex12)
6	Sure FISH	Split	KIF5B(ex15): RET(ex12)
7	Vysis	Split	KIF5B(ex15): RET(ex12)
8	Vysis	Split	KIF5B(ex15): RET(ex12)
9	Kreatech and Vysis	Split	KIF5B (ex15):RET(ex11)
10	Vysis	Split	KIF5B(ex15): RET(ex12)
11	Sure FISH	Split	KIF5B(ex15): RET(ex12)
12	Sure FISH	Split	CCDC6(ex1):RET(ex12)
13	Cytocell Aquarius	Split	KIF5B(ex15): RET(ex12)
14	Sure FISH and Cytocell Aquarius	Split	KIF5B(ex15): RET(ex12)
15	Cytocell Aquarius	Split	CCDC6(ex1):RET(ex12)
16	Cytocell Aquarius	Split	KIF5B(ex15): RET(ex12)
17	Cytocell Aquarius	Split	No fusion
18	Cytocell Aquarius	Split	No fusion
19	Cytocell Aquarius	Split	No fusion
20	Kreatech	Split	No fusion
21	Sure FISH and Cytocell Aquarius	Split	No fusion
22	Cytocell Aquarius	Single 3′	No fusion
23	Kreatech	Single 3′	No fusion
24	Sure FISH	Single 3′	No fusion
25	Sure FISH and Cytocell Aquarius	Single 3′	No fusion
26	Sure FISH, Kreatech and Vysis	Single 3′	No fusion
27	Sure FISH, Kreatech and Vysis	Single 3′	No fusion
28	Sure FISH, Kreatech and Vysis	Single 3′	No fusion
29	Vysis	Single 3′	No fusion
30	Vysis	Single 3′	No fusion
31	Vysis	Single 3′	No fusion
32	Sure FISH and Vysis	Complex	KIF5B(ex15): RET(ex12)
33	Sure FISH	Complex	KIF5B(ex15): RET(ex12)
34	Kreatech	Complex	No fusion
35	Kreatech	Complex	No fusion
36	Kreatech and Vysis	Complex	No fusion
37	Sure FISH, Kreatech and Vysis	Complex	No fusion
38	Vysis	Complex	No fusion
39	Vysis	Complex	No fusion

Note: All cases tested both by FISH and RNA NGS. RNA NGS was performed in almost all cases by Archer, except cases 32 and 33 (Oncomine) and cases 7, 8, 29, 31, 38, and 39 (Asuragen Quantidex). Unshaded cases were primarily identified by initial screening with FISH, and gray-shaded cases were identified by means of primary screening using targeted RNA NGS. Italic marked cases had different FISH patterns when different assays were used. ex, exon; FISH, fluorescence in situ hybridization; NGS, next-generation sequencing.

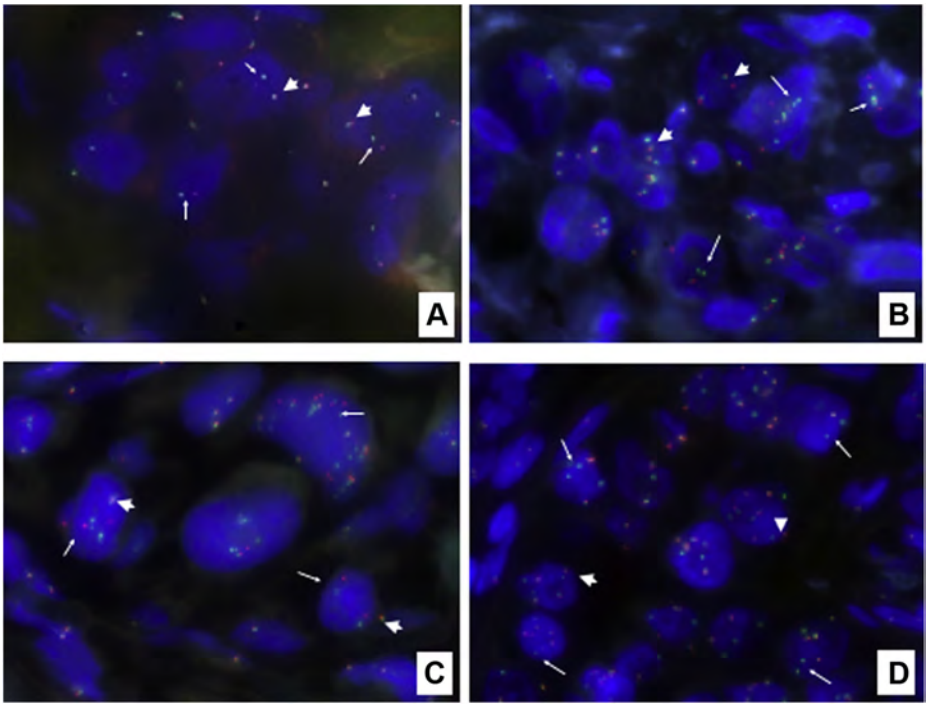


Figure 2. FISH patterns of four RET-positive cases, as confirmed by RNA NGS. (A) Example of a representative positive RET FISH with one fused (arrowhead) and one split signal (arrow) per nucleus. (B, C, D) Difficult-to-interpret positive RET FISH patterns. Case in panel B was detected by means of FISH screening that had a multiplication of the 3′ (green) probe (arrow) in the presence of the 5′ (red) signal and occasional true splits (arrowhead). (C) Case detected in FISH screening was found to have a prominent variability in nuclear size and polyploidy. Note the nuclei with a none-aberrant classic fused signal (arrow head) and multiple separate 3′ and 5′ signals (arrow), especially present in the larger nuclei. (D) Case detected using RNA NGS with confirmatory FISH. FISH revealed a difficult-to-interpret pattern with multiple single 3′ signals (arrow) next to the fused signals and occasional split signals (arrowhead) in the background of polysomy. FISH, fluorescence in situ hybridization; NGS, next-generation sequencing.

Primary screening using FISH

In total, 2858 lung adenocarcinoma were screened by FISH analysis (Figure 1). In 62 cases, FISH revealed a rearrangement: in 48 cases (1.7%), FISH was considered positive, and in 14 cases, it revealed a single 5′ signal. In 30 of 48 cases with a rearrangement, sufficient material was available for double testing using targeted RNA NGS. The presence of a *RET* fusion was confirmed in only nine cases (30%, Table 1). Eight of the confirmed cases had a split signal (example of typical split FISH signal in Figure 2A, case 5 in Table 1), and one had a complex pattern in FISH (Figure 2B, case 33). One of the confirmed cases with split signals showed a difficult-to-interpret pattern in the FISH owing to prominent polymorphism of the nuclei and polyploidy (Figure 2C, case 6) Unexpectedly, five cases with a split signal could not be confirmed

in targeted RNA NGS (Figure 3 A-E, cases 17-21). All 10 cases with a single 3′ pattern were negative for a RET fusion by targeted RNA NGS. An additional six cases with a single 5′ FISH pattern were tested using RNA NGS, and none had a functional fusion.

FISH patterns in cases detected via targeted RNA NGS

In total, 2015 NSCLC cases were screened initially by targeted RNA NGS, yielding 14 RET fusions (0.7%, Figure 1). Of these, nine could be tested by FISH and all had an aberrant FISH: eight had split signals, whereas one case was found to have a complex rearrangement pattern by FISH analysis (Figure 2D case 32)

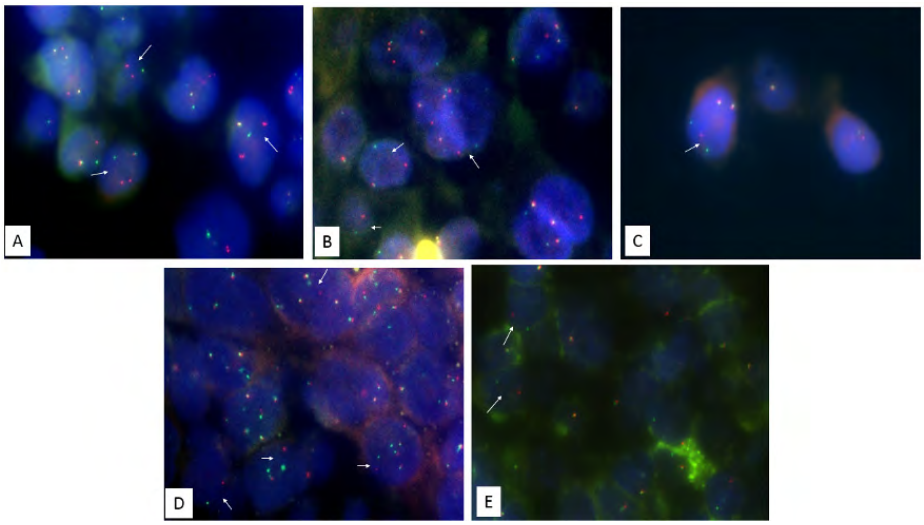


Figure 3. (A–E) Five cases with split FISH signals without a confirmed RET fusion in targeted RNA NGS (cases 17–22 from Table 1, respectively). Arrows denote split signals. FISH, fluorescence in situ hybridization; NGS, next-generation sequencing.

Cross testing using more than one FISH assay

Seventeen cases with rearrangement in FISH were retested using at least one other FISH assay than the one initially used (10 tested using two assays, seven tested using three assays). In five cases (27%), the FISH showed a different pattern in at least one other assay; all cases are found in Supplementary Figure 1 (cases 21, 25, 26, 37, and 44, respectively). FISH patterns as initially reported are presented in Table 1. None of these five cases had a functional RET fusion gene by targeted RNA NGS.

Whole-gene RET analysis

To further investigate a possible explanation for the high prevalence of rearrangements revealed by RET FISH, which do not result in functional oncogenic

RET fusions, we used available WGS data of an independent cohort[25] including 520 NSCLC samples. Structural variant analysis indicated that 9 samples harbored a disruptive RET event. However, only four of these (44%) showed a functional RET fusion (Figure 4), compared with 79% for ALK (27 of 34) and 75% for ROS1 (9 of 12).

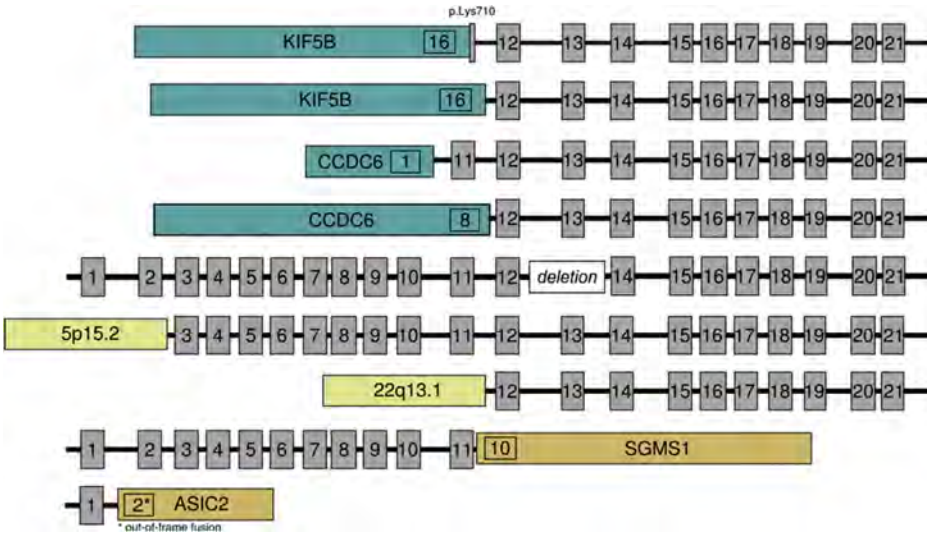


Figure 4. WGS analysis of RET locus for functional fusions and disruptive events. Analysis of disruptive RET events in WGS. RET SVs resulting in viable RET gene fusions (green), partial intragenic deletion (white), translocation to intergenic regions (yellow), or nonviable *RET* fusions (orange). SV, structural variation; WGS, whole-genome sequencing

Discussion

In this study, we report by parallel analysis of lung cancer tissues, selected from nearly 4873 patients with NSCLC in routine diagnostics, by RET FISH and targeted RNA NGS, that RET FISH frequently led to false-positive results. By contrast, NSCLC tissue with a functional RET fusion detected by means of targeted RNA NGS had an aberrant pattern by RET FISH in all nine cases with tissue available for double testing. Therefore, our data strongly indicate that RET FISH is a sensitive method to detect RET locus aberrations, but not very specific for detection of functional oncogenic RET fusions, and cannot be used as a stand-alone test method to screen for patients who might benefit from RET-targeted therapy. Our clinical data were substantiated by WGS analysis in an independent cohort of 520 patients with NSCLC, revealing that fewer than half of the disruptive RET locus rearrangements resulted in a functional

RET fusion. To the best of our knowledge, this is the first study systematically describing the drawbacks of RET fusion screening using FISH in a clinical setting.

Break-apart FISH has been the accepted standard assay for ROS1 and ALK rearrangement detection in the clinical trials for establishment of the criteria of eligibility for targeted therapy[13,27]. FISH is a technique available in most pathology laboratories and intuitively probably often the first choice for RET screening in NSCLC. Literature on RET FISH is limited to case series that describe, in line with our data, a consistently aberrant FISH pattern in reverse-transcriptase polymerase chain reaction-positive cases, underlining the high sensitivity of RET FISH[6,28]. Several limitations to the FISH technique are known from the ALK and ROS1 literature, some of which we also encountered, including false-positives due to rearrangements without a functional fusion. The percentage of functional fusions relative to all disruptive events in ALK and ROS1 loci in our presented WGS data was higher than we observed for RET, likely due to the lower frequency of genomic rearrangements not yielding a functional ALK or ROS1 fusion. Other limitations of ALK and ROS1 FISH from the literature are false-negatives, owing to small genomic deletions, unspecific probe hybridization leading to false-positives, and high levels of background noise[29–31]. However, in combination with ALK or ROS1-positive immunohistochemistry, sensitivities and specificities approaching 100% were described[23,32]. Unfortunately, RET immunohistochemistry has revealed variable performance with an overall low sensitivity and low specificity and is therefore not suitable for daily pathology practice[6,7,33].

RET FISH patterns of cases with a functional RET fusion in the limited available literature were almost always split signals, in line with our data, with an occasional case with a single 3' positive RET FISH pattern. In our cohort, none of the single 3' pattern cases had a functional RET fusion indicating that this pattern in *RET* FISH is very uncommon.

We present a relatively high percentage of cases with discordant results between different FISH assays. This variability between FISH assays may largely be attributed to interassay variation, including differences in probe DNA binding sites, digestion, incubation time and buffer conditions[34]. The exact DNA binding sites are often difficult to obtain from the commercial kits, prohibiting the evaluation of probe locations and genomic breakpoints. Independent of the FISH assay used, our data underscore that aberrant RE*T* FISH patterns should always be confirmed by another unambiguous method such as RNA NGS or WGS, given the lack of reliable RET immunohistochemistry; otherwise these FISH results should be interpreted with

great caution. Genomic profiling of lung cancer using multiplex NGS techniques was recommended in the most recent report of the European Society for Medical Oncology precision medicine working group[35], in line with our findings.

In our laboratories, targeted RNA NGS was used as a confirmatory test for FISH-rearranged cases. Targeted RNA NGS allows simultaneous testing of multiple genes, including, for example RET, ALK, and ROS1, in one test, and for the presence of both known and as yet unknown fusion partners[21]. In addition, targeted RNA NGS is a sensitive method for reliable detection of low copy numbers of a gene fusion transcripts in a background of normal RNA[36]. With an increasing number of targetable oncogenic gene fusions present as driver at diagnosis of NSCLC or as resistance mechanism to targeted therapy in EGFR driven cancers[18,37], RNA NGS is therefore a feasible approach. A drawback of RNA NGS might be its reliance on RNA, because RNA, like DNA quality[31], was described to be low in routine FFPE patient material. However, the most often encountered problem in our cohort was insufficient material.

Predictive testing in NSCLC is hampered by low DNA and RNA yields from small bronchoscopy or transthoracic biopsies and the increasing number of requested predictive tests. Tissue management is of essential importance and requires interaction of all the involved physicians and special cutting protocols for lung biopsies[38]. One-step DNA and RNA isolation and simultaneous NGS of both DNA and RNA (after complementary DNA synthesis) instead of a consecutive approach might reduce the required amount of tissue [39].

A limitation of our study is that, owing to the study design, no firm conclusions could be made regarding the prevalence of RET fusion genes in our cohort and, more importantly, the precise sensitivity and specificity of both FISH and targeted RNA NGS. That would require parallel screening of a large cohort of patients with NSCLC by both methods, which would be a huge effort of multiple institutes, because of the low prevalence of RET fusions. Nevertheless, our study is the largest and most detailed study so far on FISH as a screening method for RET fusion detection in clinical practice.

In conclusion, RET FISH seems to be a sensitive technique, but with a very low specificity, resulting in a large proportion of false-positive results. WGS analysis reveals that this is due to a high fraction of disruptive events in the RET locus that do not result in a functional RET fusion. RET FISH rearrangement always needs confirmation by an orthogonal technique in order to select the patients who might benefit from RET-targeted therapy.

References

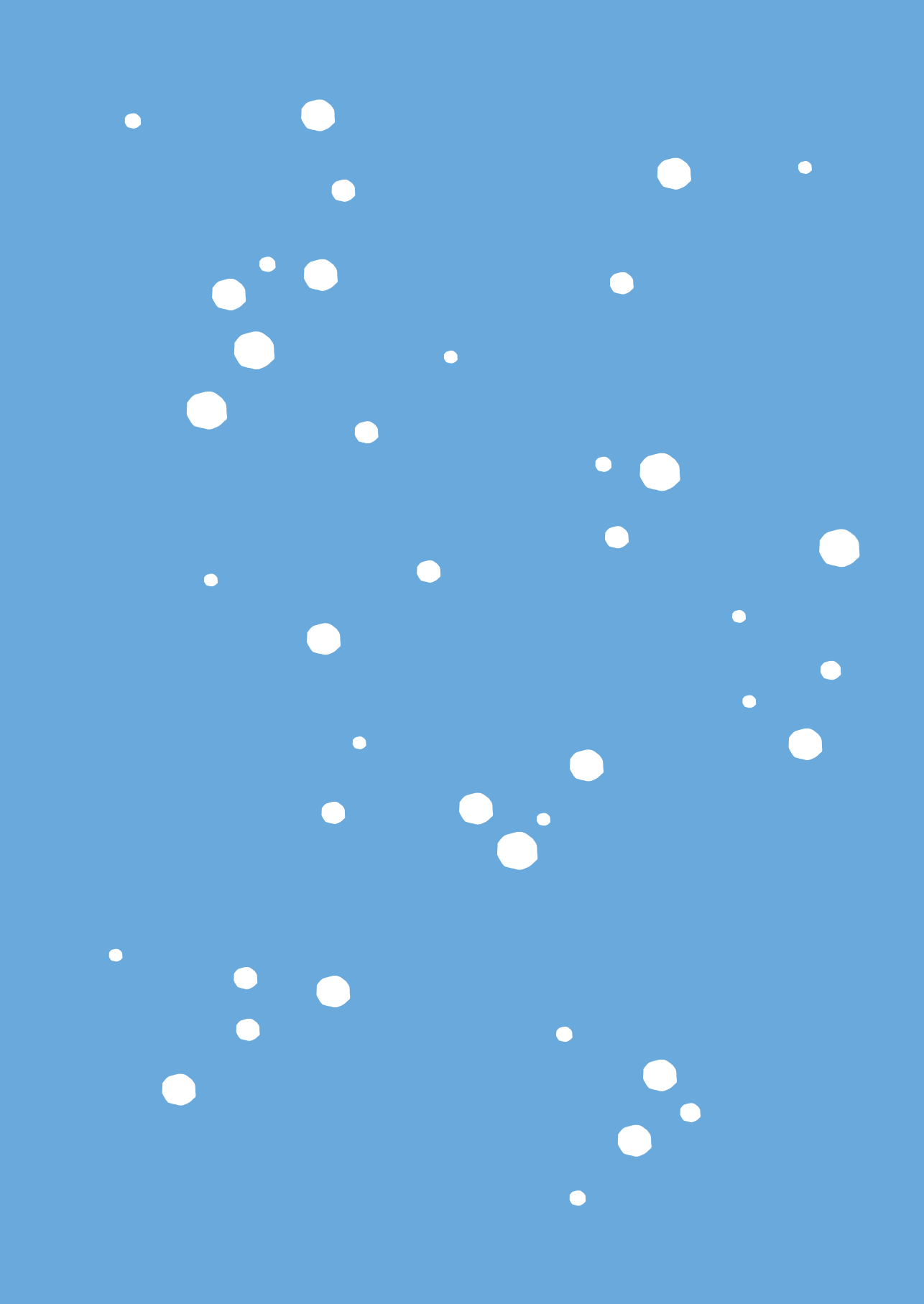
1. Ju YS, Lee WC, Shin JY, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res. 2012;22:436–445.
2. Lira ME, Choi YL, Lim SM, et al. A single-tube multiplexed assay for detecting ALK, ROS1, and RET fusions in lung cancer. J Mol Diagn. 2014;16:229–243.
3. Carlomagno F. Thyroid cancer: role of RET and beyond. Eur Thyroid J. 2012;1:15–23.
4. Mizukami T, Shiraishi K, Shimada Y, et al. Molecular mechanisms underlying oncogenic RET fusion in lung adenocarcinoma. J Thorac Oncol. 2014;9:622–630.
5. Pan Y, Zhang Y, Li Y, et al. ALK, ROS1 and RET fusions in 1139 lung adenocarcinomas: a comprehensive study of common and fusion pattern-specific clinicopathologic, histologic and cytologic features. Lung Cancer. 2014;84:121–126.
6. Lee SE, Lee B, Hong M, et al. Comprehensive analysis of RET and ROS1 rearrangement in lung adenocarcinoma. Mod Pathol. 2015;28:468–479.
7. Tsuta K, Kohno T, Yoshida A, et al. RET-rearranged nonsmall-cell lung carcinoma: a clinicopathological and molecular analysis 2014;. 2014;110, 1571–1478.
8. Shen T, Pu X, Wang L, et al. Association between RET fusions and efficacy of pemetrexed-based chemotherapy for patients with advanced NSCLC in China: a multicenter retrospective study. Clin Lung Cancer. 2020;21:e349–e354.
9. Platt A, Morten J, Ji Q, et al. A retrospective analysis of RET translocation, gene copy number gain and expression in NSCLC patients treated with vandetanib in four randomized phase III studies. BMC Cancer. 2015;15:171.
10. Wang R, Hu H, Pan Y, et al. RET fusions define a unique molecular and clinicopathologic subtype of non-smallcell lung cancer. J Clin Oncol. 2012;30:4352–4359.
11. Drilon A, Oxnard G, Wirth L, et al. PL02.08 registrational results of LIBRETTO-001: a phase 1/2 trial of LOXO-292 in patients with RET fusion-positive lung cancers. J Thorac Oncol. 2019;14(suppl):S6–S7.
12. Gautschi O, Milia J, Filleron T, et al. Targeting RET in patients with RET-rearranged lung cancers: results from the global, multicenter RET registry. J Clin Oncol. 2017;35:1403–1410.
13. Shaw AT, Felip E, Bauer TM, et al. Lorlatinib in nonsmall-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, singlearm first-in-man phase 1 trial. Lancet Oncol. 2017;18:1590–1599.
14. Shaw AT, Kim DW, Mehra R, et al. Ceritinib in ALKrearranged non-small-cell lung cancer. N Engl J Med. 2014;370:1189–1197.
15. Li AY, McCusker MG, Russo A, et al. RET fusions in solid tumors. Cancer Treat Rev. 2019;81:101911.
16. Gainor JF, Lee DH, Curigliano G, et al. Clinical activity and tolerability of BLU-667, a highly potent and selective RET inhibitor, in patients (pts) with advanced RETfusionþ non-small cell lung cancer (NSCLC). J Clin Oncol. 2019;37(suppl 15):9008–9008.
17. FDA approves selpercatinib; pralsetinib may soon follow. Cancer Discov. 2020;10:OF1.

18. Klempner SJ, Bazhenova LA, Braiteh FS, et al. Emergence of RET rearrangement co-existing with activated EGFR mutation in EGFR-mutated NSCLC patients who had progressed on first- or second-generation EGFR TKI. Lung Cancer. 2015;89:357–359.
19. Tachon G, Cortes U, Richard S, et al. Targeted RNAsequencing assays: a step forward compared to FISH and IHC techniques? Cancer Med. 2019;8:7556–7566.
20. Bustin SA, Benes V, Nolan T, Pfaffl MW. Quantitative realtime RT-PCR - a perspective. J Mol Endocrinol. 2005;34:597–601.
21. Blidner RA, Haynes BC, Hyter S, et al. Design, optimization, and multisite evaluation of a targeted nextgeneration sequencing assay system for chimeric RNAs from gene fusions and exon-skipping events in non–small cell lung cancer. J Mol Diagn. 2019;21:352–365.
22. Clark J, Attard G, Jhavar S, et al. Complex patterns of ETS gene alteration arise during cancer development in the human prostate. Oncogene. 2008;27:1993–2003.
23. Bubendorf L, Büttner R, Al-Dayel F, et al. Testing for ROS1 in non-small cell lung cancer: a review with recommendations. Virchows Arch. 2016;469:489–503. May 2021 RET FISH as Screening Method in Lung Cancer 805.
24. Letovanec I, Finn S, Zygoura P, et al. Evaluation of NGS and RT-PCR methods for ALK rearrangement in European NSCLC patients: results from the European thoracic oncology platform Lungscape project. J Thorac Oncol. 2018;13:413–425.
25. Priestley P, Baber J, Lolkema MP, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575:210–216.
26. Cameron D, Baber J, Shale C, et al. GRIDSS, PURPLE, LINX: unscrambling the tumor genome via integrated analysis of structural variation and copy number. https://www.biorxiv.org/content/10.1101/781013v1.full. pdf. Accessed November 26, 2020.
27. Shaw AT, Ou SHI, Bang YJ, et al. Crizotinib in ROS1- rearranged non-small-cell lung cancer. N Engl J Med. 2014;371:1963–1971.
28. Go H, Jung YJ, Kang HW, et al. Diagnostic method for the detection of KIF5B-RET transformation in lung adenocarcinoma. Lung Cancer. 2013;82:44–50.
29. Camidge DR, Kono SA, Flacco A, et al. Optimizing the detection of lung cancer patients harboring anaplastic lymphoma kinase (ALK) gene rearrangements potentially suitable for ALK inhibitor treatment. Clin Cancer Res. 2010;16:5581–5590.
30. Martin V, Bernasconi B, Merlo E, et al. ALK testing in lung adenocarcinoma: technical aspects to improve FISH evaluation in daily practice. J Thorac Oncol. 2015;10:595–602.
31. Davies KD, Le AT, Sheren J, et al. Comparison of molecular testing modalities for detection of ROS1 rearrangements in a cohort of positive patient samples. J Thorac Oncol. 2018;13:1474–1482.
32. Thunnissen E, Bubendorf L, Dietel M, et al. EML4-ALK testing in non-small cell carcinomas of the lung: a review with recommendations. Virchows Arch. 2012;461:245–257.
33. Sasaki H, Shimizu S, Tani Y, et al. RET expression and detection of KIF5B/RET gene rearrangements in Japanese lung cancer. Cancer Med. 2012;1:68–75.

34. Garimberti E, Tosi S. Fluorescence in situ hybridization (FISH), basic principles and methodology. Methods Mol Biol. 2010;659:3–20.
35. Mosele F, Remon J, Mateo J, et al. Recommendations for the use of next-generation sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. Ann Oncol. 2020;31:1491– 1505.
36. Barua S, Wang G, Mansukhani M, Hsiao S, Fernandes H. Key considerations for comprehensive validation of an RNA fusion NGS panel. Pract Lab Med. 2020;21:e00173.
37. Piotrowska Z, Isozaki H, Lennerz JK, et al. Landscape of acquired resistance to osimertinib in EGFR-mutant NSCLC and clinical validation of combined EGFR and RET inhibition with osimertinib and BLU-667 for acquired RET fusion. Cancer Discov. 2018;8:1529–1539.
38. Bubendorf L, Lantuejoul S, de Langen AJ, Thunnissen E. Nonsmall cell lung carcinoma: diagnostic difficulties in small biopsies and cytological specimens. Eur Respir Rev. 2017;26:170007.
39. Turashvili G, Yang W, McKinney S, et al. Nucleic acid quantity and quality from paraffin blocks: defining optimal fixation, processing and DNA/RNA extraction techniques. Exp Mol Pathol. 2012;92:33–43.

PART 3

Summary and future perspectives



CHAPTER 9

Summary and future perspectives

This thesis aimed to investigate the implementation and application of WGS in routine clinical care. The results of this thesis show that WGS is feasible and a clinically valid technique in routine clinical care with acceptable turnaround times. Importantly, the required adjustments to multiple logistic processes were perceived as manageable to the health care professionals involved, indicating that implementation hurdles in adopting WGS in routine clinical care can be overcome. WGS demonstrated clinical value in the identification of actionable biomarkers and integration of germline diagnostics. In addition, WGS-based diagnostics was of value in patients with cancer of unknown primary (CUP). This thesis is divided in the implementation process of WGS in routine clinical care and the use of (whole) genome sequencing in rare tumor types or to detect rare genetic events, both will be discussed in more detail in the following sections.

Implementation of Whole Genome Sequencing in routine clinical care

Until now, WGS has mainly been used for research purposes and has not yet entered clinical oncology. The appropriate use of genomic tests in clinical oncology, largely depends on evidence-based assessment of feasibility, clinical validity, clinical value and cost-effectiveness in routine clinical care (1). In addition, guidelines on how to implement genomic tests in clinical settings are warranted. The *Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient* (WIDE) study aimed to provide evidence on the performance of WGS in a clinical setting and to identify a sound clinical implementation strategy. **Chapter 2** of this thesis describes the study protocol of the WIDE study, a prospective observational cohort study, conducted at the Netherlands Cancer Institute (NKI). The study aimed to include 1200 consecutive patients with (a suspicion) of stage IV disease of solid tumors without pre-selection on tumor-type within a timeframe of 18-24 months. WGS was performed in parallel to standard of care (SOC) diagnostics on routinely obtained tumor samples. The primary endpoint feasibility was expressed as the percentage of patients for whom processing from biopsy to WGS report was successful and the turnaround time as the time of biopsy until WGS report in working days. Clinical validity, the second primary endpoint, was defined as the percentage of variants for which WGS detected (at minimum) the same clinically relevant variants as molecular SOC diagnostics. Secondary endpoints included health technology assessment, additional treatment options identified by WGS, better informed decision making and enrichment of the Hartwig database. Key performance indicators were evaluated after every 200 patients enrolled, and procedures optimized accordingly. **Chapter 3** describes a strategy for efficient implementation of WGS into routine pathology practice. As a result of the WIDE study, WGS was successfully implemented at the NKI as part of routine diagnostics by January 2021. The successful clinical implementation of WGS has relied on adhering to a comprehensive protocol by

all health care professionals involved, i.e., radiologists, pathologists, clinical DNA specialists, laboratory technicians, clinical geneticists and medical oncologists. In addition, important lessons learned from the implementation process during the WIDE study are outlined which include recommendations on increasing success rates of WGS by using ≤18 gauge needles, taking multiple (2-4) biopsies and maximizing the yield of tumor material by using a certain manual microdissection method called 4FME (fresh-frozen fine fixed MCC embedded).

In **Chapter 4** the results of the WIDE study on feasibility, clinical validity and clinical value of WGS in routine pathology practice are discussed. The results showed that WGS analysis was feasible for 71% of patients with metastatic cancer with a median turnaround time of 11 working days. Low tumor cell percentage was the main reason for WGS ineligibility. The clinical validity of WGS was demonstrated with a biomarker detection rate of 99.2% (two sided 95CI 98.4-99.6%). In accordance with previous studies (2, 3), WGS identified actionable biomarkers for regular therapy options or clinical trial allocation in 71% of patients. Of those patients, 24% started biomarker-based therapy at median follow-up of 14 months. WGS also proved its additional diagnostic value in germline diagnostics, with previously 49 unrecognized pathogenic germline variants being identified by WGS. The evidence emerging from this chapter indicates that WGS is a clinically valid technique which identifies additional actionable biomarkers for regular treatment options and clinical trial allocation. **Chapter 5** prospectively investigated the clinical value of WGS in the routine diagnostic work-up of 72 patients with CUP. CUP was defined as tumors with unknown origin or histology type in patients with metastatic cancer. Previous studies have shown that genomic features such as mutational signatures and topological distribution of driver and passenger mutations can be used to accurately predict tumor type (4-8). This chapter describes the development and validation of a WGS-based 'cancer of unknown primary prediction algorithm' (CUPPA) which combines tumor-type specific drivers, regional mutational density and mutational profile characteristics. CUPPA was trained and validated on a large WGS database of metastatic cancer patients (4,509), which was randomly divided in a reference set (90%, n = 4058) and a test set (10%, n = 451), and on 254 independent patients. CUPPA was able to correctly classify 84% and 78% of samples in the internal (n = 451) and independent validation cohort (n = 254), respectively. In the 72 patients with CUP who received WGS as part of their routine diagnostic work-up, WGS and CUPPA could determine primary tumor type for 49 patients (68%) and detect actionable events in 34 patients (47%). Common diagnoses included non-small cell lung (n=7), gastroesophageal (n=4), and pancreatic cancer (n=4). WGS as part of the routine diagnostic work-up in CUP patients therefore proved to be valuable diagnostic tool.

Use of (Whole) Genome Sequencing in rare tumor types or to detect rare genetic events

Rare cancers have a more dismal prognosis than common cancers, potentially due to poor elucidation of the genomic alterations underlying tumorigenesis and lack of effective therapeutic options (9). **Chapter 6** provides an overview of literature with the aim to gain more insight into the genomic landscape of SI-NETs and identify prognostic molecular factors. SI-NETs have a low mutational burden, with loss of heterozygosity of chromosome 18 (44-100%) as the most frequent genomic aberration, followed by mutations of *CDKN1B* (8%). Other mutations identified in SI-NETs include mutations in *APC*, *CDKN2C* (both 7.7%) and *BRAF*, *KRAS*, *PIK3CA* and *TP53* (each 3.8%). In comparison with genetic mutations, epigenetic alterations were significantly more common in SI-NETs. In SI-NETs, *RASSF1A*, *SEMA3F* and *CTNNB1* are hypermethylated, which silences their transcription. Loss of heterozygosity at chromosome 18, gains of chromosome 4, 5, 7, 14 and 20p, copy gain of the *SRC* gene and low expression of RASSF1A and P16 were associated with poorer survival. No biomarkers have been identified yet that can easily be adopted as prognostic factors into current clinical decision making. As epigenetic dysregulation is more common in SI-NETs, more promising targets for precision oncology may constitute the DNA methylation machinery. **Chapter 7** further investigates the presence and significance of driver mutations in metastases of well-differentiated SI-NETs. WGS was performed on 35 metastatic SI-NETs and NGS on eight metastatic SI-NETs. The cohort consisted of 21% grade 1, 60% grade 2 and 19% grade 3 SI-NETs. Driver mutations were identified in approximately 50% of SI-NETs. In total, 27 driver mutations were identified, of which 74% were in tumor suppressor genes (i.e. *TP53, RB1*, and *CDKN1B*) and 22% were in proto-oncogenes (i.e. *KRAS*, *NRAS*, and *MET*). Allelic loss of chromosome 18 (63%), complete loss of *CDKN2A* and *CDKN1B* (both 6%) and CDKN1B mutations (9%) were most common. All tumors were microsatellite-stable and showed low TMBs (median 1.10; interquartile range 0.87-1.35). The Ki67 proliferation index was significantly associated with the presence of driver mutations ($P = 0.015$). There was no significant correlation between mutational status and SSTR2a expression or disease specific survival. Potential targetable genetic alterations were detected in 21% of metastasized SI-NETs which provides a rationale for the introduction of targeted therapy in the treatment of SI-NETs. **Chapter 8** investigates the performance of fluorescence in situ hybridization (FISH) as a diagnostic test for detecting functional *RET* fusions. The incidence of *RET* fusions is approximately 1% in non-small cell lung cancers (NSCLCs) (10, 11). Accurate detection of *RET* fusions is important to ensure optimal therapy selection. In total, 4873 patients with NSCLC were routinely screened in six European cancer centers for *RET* fusions using either FISH (n=2858) or targeted RNA NGS fusion analysis (n=2015). When possible, positive

cases were tested with both techniques (n=39). FISH revealed *RET* rearrangement in 48 of 2858 cases. Of 30 cases also tested with NGS fusion analysis, only nine had a confirmed functional *RET* fusion. In total, NGS fusion analysis detected *RET* fusions in 14 of 2015 cases, of which the nine cases double tested with FISH all showed *RET* locus rearrangement. These results were substantiated by WGS analysis data in an independent cohort of 520 patients with NSCLC, revealing that fewer than half of the disruptive *RET* locus rearrangements shown by FISH resulted in a functional *RET* fusion. To conclude, FISH is a sensitive but unspecific technique resulting in large proportion of false-positive results. A demonstrated *RET* FISH rearrangement always needs confirmation by an orthogonal technique (i.e. RNA NGS or WGS) to select patients who might benefit from RET-targeted therapy.

Future perspectives

Over the past two years, WGS has increasingly found its way into clinical care in multiple pediatric cancer centers (12-15). Similar to adult cancer patients, WGS has demonstrated clinical benefit in pediatric cancer patients due to detection of treatment relevant biomarkers and integration of germline diagnostics. However, the setting of pediatric cancer centers differs substantially in volume and scale compared to adult cancer centers and therefore cannot be easily translated to adult oncology. The success of the WIDE study and subsequent implementation of WGS as part of clinical care at the NKI, has demonstrated its feasibility in the adult oncology setting. The introduction of WGS as part of routine diagnostics has yielded valuable lessons and key considerations for future clinical implementation of WGS in other hospitals worldwide which will be elaborated upon below.

WGS logistics

WGS analysis requires fresh or fresh frozen samples for generating high-quality, accurate genome-wide variant calls, however most, if not all, pathology departments work with formalin-fixed paraffin embedded (FFPE) tissue samples in their routine diagnostic workflow. The transition from a FFPE to a fresh frozen workflow is perceived as one of the major hurdles in the broad clinical implementation of WGS. Although implementing a fresh frozen workflow can be more complex than using readily available FFPE samples, the results of this thesis show that this hurdle can rather easily be overcome if a comprehensive protocol is adhered to by qualified staff and appropriate equipment (i.e. PrestoCHILL device) is available. Other considerations for WGS logistics concern choices of which part of the procedures to perform in house and which at an external WGS provider site. For example, blood and/or tissue can either be directly shipped to an external WGS provider site or DNA from blood and/or tissue can be isolated first in house and subsequently shipped

to an external WGS provider for sequencing. Depending on the external WGS provider, data can be returned as a processed patient report or as raw data. A clinical laboratory can also opt to build its own sequencing analysis pipeline. The fully open source in-house bio-informatic pipeline used by Hartwig is available through github.com/hartwigmedical and can serve as a template. Lastly, WGS results need to be interpreted in the clinical context of the patient by a trained clinical DNA specialist and a clinical geneticist needs to be involved for counselling patients in case of pathogenic germline findings.

Centralized approach to WGS-based diagnostics

It might not be feasible for every healthcare institution to adopt WGS-based diagnostics and employ clinical DNA specialists for sequencing data interpretation. In order to ensure all eligible patients have access to WGS, it may be necessary to join forces and concentrate specialist staff and equipment in centers of excellence. A centralized approach, i.e. concentration of WGS-based diagnostics in expertise centers, seems to be a realistic future scenario. Moreover, in expertise centers diagnostic expertise is available if needed, access to clinical studies is provided and therapeutic guidance can be safeguarded. Furthermore, the rapidly changing scientific, technological, pharmaceutical and clinical fields can constantly be taken into account. Molecular Tumor Boards (MTBs) consisting of clinicians from different medical divisions, pathologists, clinical DNA specialists and clinical geneticists are essential for the deployment of WGS-based diagnostics. In MTBs, molecular alterations can be discussed within the clinical and diagnostic context of the patient, including the possibilities for available matching therapies which can be offered as part of standard care, in clinical trial setting, compassionate use or early access programs. Importantly, interdisciplinary knowledge is transferred at MTBs (16). In order to ensure equal access to WGS-based diagnostics and precision oncology for every cancer patient, it is pivotal that, when indicated, cancer patients are referred to centers where WGS-based diagnostics is available or that samples and data are efficiently transferred between centers of excellence and collaborating institutions. Furthermore, access to MTBs by external health care providers should be encouraged (17).

Costs

Although WGS has demonstrated added clinical value and implementation in routine clinical care is feasible, costs are an important factor in its implementation. While at an individual test level, direct costs of WGS are higher than of WES or NGS panels, a comprehensive cost versus benefit analysis is complex. Challenges in such an analysis include incorporation of all other potential benefits of WGS-based diagnostics and translation of WGS results into long term patient outcomes. On an organizational

level for example, WGS eliminates the need to redesign and validate NGS panels which results in a future-proof approach with potential cost-effective effects on the long term. Moreover, implementation of WGS allows for simplification of laboratory logistics by abandoning certain tests which results in less costs. Similarly, the incorporation of germline diagnostics and pharmacogenomics, improved uniformity between centers, less repetitive analyses and secondary biopsies, harbor cost-saving potential. On a patient-related level, WGS could be beneficial and improve quality of life by providing accurate and relatively quick information on diagnosis (particularly important in CUP patients), possible treatment options and pathogenic germline findings with direct clinical implications to patients and their families. In addition, accurate diagnosis and characterization of actionable biomarkers prevents under- or overtreatment i.e. by refraining from immunotherapy in RET rearranged lung cancers. From a scientific perspective, storing WGS results in a central database enables identification of patients eligible for experimental treatments. In addition, it allows for discovery of biomarkers to identify patients who do not respond to (costly) targeted treatments, which could be highly cost-effective in the future. The results of a comprehensive cost-benefit analysis of the WIDE-study will soon be finalized and published. In the coming months to years, growing uptake of WGS will increase efficiency and hence reduce costs. In fact, recently Ultima Genomics has introduced a new approach to sequencing that drives down costs and increases throughput which renders the costs of WGS only 100 dollars (18). Increased competition on the price of WGS, for example by competing parties offering WGS will potentially further lower WGS costs. Reduction in costs will allow for higher depth sequencing and a further increase in WGS sensitivity (19).

Indications for WGS

In 2021, a first provision for reimbursement of WGS was established in the Netherlands (20). This reimbursement pertains to patients with CUP. However, other patient populations might also clinically benefit from WGS-based diagnostics i.e. patients with rare cancers (including sarcomas) and with diagnostically challenging tumors (21-23). Recently, it has been demonstrated that there is a significant overlap in genomic targets between rare and common cancers when comprehensive genomic profiling is performed. Moreover, patients with rare cancers appear to have similar benefit from off-label targeted agents as patients with common cancers (24). Another patient population with high unmet needs are adolescents and young adults (AYAs; 18 up to 39 years of age). Currently, the benefit of WGS in this patient group is studied within the GENAYA (a national database of GENome data of Adolescent and Young Adult cancers) project (25).

In case of small intestinal neuroendocrine tumors, future research endeavors should focus on genomic alterations beyond the >500 variants with high driver likelihood which are described on the WGS report. Of note, although each tumor sample is analyzed for its whole genome characteristics, including all genes (exons and introns) and intergenic regions, the WGS report is limited to variants with high driver likelihood in order to provide clinically manageable reports. As high driver likelihood, a score for each mutation being a potential driver event, is determined based on a driver mutation catalogue in known cancer genes, it might be worthwhile to explore beyond those limits to potentially increase the yield of WGS in (small intestinal) neuroendocrine tumors (3).

Another important aspect is the timing of WGS during a patient's disease course. Studies show that there is limited evolution of the actionable metastatic cancer genome under therapeutic pressure (26, 27). A single WGS of a metastatic biopsy is therefore generally sufficient to identify biomarkers for on-label treatments and for clinical trial enrollment. The limited evolution of the actionable genome of treated metastases implies that WGS could be performed early in the patient's disease course. This approach could provide clinicians with a complete genomic characterization at the time a patient presents with metastatic disease. As a result, clinical management could be improved by identifying (1) genomic alterations for which on-label treatments are available, (2) identifying patients with rare genomic aberrations (i.e. MSI or NTRK fusions) which can be targeted effectively and (3) identifying patients with pathogenic germline variants with direct clinical implications (i.e. pharmacogenomic variants or BRCA1/2 mutations).

In the near future, in the current Dutch situation, an extension of the indications for reimbursement for WGS for rare cancers, diagnostically challenging tumors and for adolescents and young adults with (metastatic) cancer, especially patients with poor prognosis, should be considered. If in the short term costs of WGS fall below the costs of current complete molecular diagnostics (NGS-based fusion analysis and NGS panels), the way should be paved for every metastatic cancer patient to at least undergo WGS once at the appropriate time in their disease course.

Clinical utility

With implementation of WGS, the ability to interpret WGS data will improve and will reveal new diagnostic possibilities (i.e. CUPPA), and by extension, the clinical utility of WGS will expand. Due the comprehensive nature of WGS data, the generated data will always be comparable in time and place. Moreover, collection of comprehensive genomic data combined with detailed clinical phenotyping in a central database

provides a solid basis for a learning health care system. Learning health care systems, are systems in which knowledge generation processes are embedded in daily practice to produce continual optimizations in health care. Collection of data in a central database also facilitates artificial intelligence such as the use of machine-learning algorithms and software. To illustrate, such a learning health care system could identify biomarkers that select patients who do not respond to a certain therapy and thereby prevent under- and overtreatment, increase the use of algorithms for the molecular classification and stratification of tumors (as opposed to morphological classification) and accelerate development of other precision oncology strategies. In the future, this central database could also be enriched by other data sources such as transcriptomic, proteomic and immune profiling data. Altogether, collection, storage and analysis of clinicogenomic data is a critical resource to deploy precision oncology to its full extent.

Conclusion

In conclusion, WGS is a clinically valid technique and WGS-based diagnostics is feasible in routine pathology practice. The implementation of WGS at the NKI has demonstrated that implementation hurdles can be tackled and lessons drawn from this implementation process can guide WGS uptake in other hospitals. At present, WGS has become part of routine diagnostics in 5 other hospitals in the Netherlands. This was facilitated by the fact that in 2021 a first provision for reimbursement of WGS in the Netherlands was established for patients with CUP. In the short term, an extension of this reimbursement could be considered for patients with rare cancers, diagnostically challenging tumors and adolescents and young adults with (metastatic) cancer with poor prognosis. As costs of WGS will steadily decrease in the foreseeable future, the path is smoothed for implementation of WGS in routine clinical care worldwide, thereby, optimally deploying precision oncology and supporting learning health care systems.

References

1. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med. 2009;11(1):3-14.
2. Cobain EF, Wu YM, Vats P, Chugh R, Worden F, Smith DC, et al. Assessment of Clinical Benefit of Integrative Genomic Profiling in Advanced Solid Tumors. JAMA Oncol. 2021;7(4):525-33.
3. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. Nature. 2019;575(7781):210-6.
4. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. Science. 1999;286(5439):531-7.
5. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502(7471):333-9.
6. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21.
7. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94-101.
8. Polak P, Karlic R, Koren A, Thurman R, Sandstrom R, Lawrence M, et al. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature. 2015;518(7539):360-4.
9. de Heus E, Duijts SFA, van der Zwan JM, Kapiteijn E, Nieveen van Dijkum EJM, van Herpen CML, et al. The gap between rare and common cancers still exists: Results from a population-based study in the Netherlands. Eur J Cancer. 2022;167:103-11.
10. Platt A, Morten J, Ji Q, Elvin P, Womack C, Su X, et al. A retrospective analysis of RET translocation, gene copy number gain and expression in NSCLC patients treated with vandetanib in four randomized Phase III studies. BMC Cancer. 2015;15:171.
11. Wang R, Hu H, Pan Y, Li Y, Ye T, Li C, et al. RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. J Clin Oncol. 2012;30(35):4352-9.
12. Wong M, Mayoh C, Lau LMS, Khuong-Quang DA, Pinese M, Kumar A, et al. Whole genome, transcriptome and methylome profiling enhances actionable target discovery in high-risk pediatric cancer. Nat Med. 2020;26(11):1742-53.
13. Byrjalsen A, Hansen TVO, Stoltze UK, Mehrjouy MM, Barnkob NM, Hjalgrim LL, et al. Nationwide germline whole genome sequencing of 198 consecutive pediatric cancer patients reveals a high incidence of cancer prone syndromes. PLoS Genet. 2020;16(12):e1009231.
14. Trotman J, Armstrong R, Firth H, Trayers C, Watkins J, Allinson K, et al. The NHS England 100,000 Genomes Project: feasibility and utility of centralised genome sequencing for children with cancer. Br J Cancer. 2022.
15. Shukla N, Levine MF, Gundem G, Domenico D, Spitzer B, Bouvier N, et al. Feasibility of whole genome and transcriptome profiling in pediatric and young adult cancers. Nat Commun. 2022;13(1):2485.

16. Willemsen A, Krausz S, Ligtenberg MJL, Grunberg K, Groen HJM, Voest EE, et al. Molecular tumour boards and molecular diagnostics for patients with cancer in the Netherlands: experiences, challenges, and aspirations. Br J Cancer. 2019;121(1):34-6.
17. Mateo J, Steuten L, Aftimos P, Andre F, Davies M, Garralda E, et al. Delivering precision oncology to patients with cancer. Nat Med. 2022;28(4):658-65.
18. Ultima Genomics I. Unleash the power of genomics at scale [https://www.ultimagenomics.com/about-ultima-genomics]. California2022 [cited 2022 25th of July].
19. Roepman P, de Bruijn E, van Lieshout S, Schoenmaker L, Boelens MC, Dubbink HJ, et al. Clinical Validation of Whole Genome Sequencing for Cancer Diagnostics. J Mol Diagn. 2021;23(7):816-33.
20. Zorginstituut. Moleculaire diagnostiek in de Oncologie. Nederland, Diemen2021.
21. Schipper LJ, Monkhorst K, Samsom KG, Bosch LJW, Snaebjornsson P, van Boven H, et al. Clinical Impact of Prospective Whole Genome Sequencing in Sarcoma Patients. Cancers (Basel). 2022;14(2).
22. Horak P, Heining C, Kreutzfeldt S, Hutter B, Mock A, Hullein J, et al. Comprehensive Genomic and Transcriptomic Analysis for Guiding Therapeutic Decisions in Patients with Rare Cancers. Cancer Discov. 2021;11(11):2780-95.
23. Schipper LJ SP, Samsom KG, Bosch LJW, Lalezari F, Priestley P, Shale C, Jacobs N, van den Broek AJ, Roepman P, van der Hoeven JJM, Steeghs N, Cuppen E, Meijer G, Voest EE and Monkhorst K. Whole genome sequencing can classify diagnostically challenging tumors. Annals of Oncology. 2021;32 S921-S30.
24. Hoes LR, van Berge Henegouwen JM, van der Wijngaart H, Zeverijn LJ, van der Velden DL, van de Haar J, et al. Patients with Rare Cancers in the Drug Rediscovery Protocol (DRUP) Benefit from Genomics-Guided Treatment. Clin Cancer Res. 2022;28(7):1402-11.
25. Foundation HM. Complete DNA analysis for young adults with cancer Amsterdam2022 [updated 06-01-2022; cited 2022 08-08]. hartwigmedicalfoundation.nl/en/complete-dna-analyse-voor-jongvolwassenen-met-kanker/].
26. van de Haar J, Hoes LR, Roepman P, Lolkema MP, Verheul HMW, Gelderblom H, et al. Limited evolution of the actionable metastatic cancer genome under therapeutic pressure. Nat Med. 2021;27(9):1553-63.
27. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science. 2018;361(6406):1033-7.

HOOFDSTUK 10

Nederlandse samenvatting

Achtergrond

Kanker is de meest voorkomende doodsoorzaak in Nederland en jaarlijkse stijgt het aantal patiënten die de diagnose kanker krijgen. In 2021, werd bij 124.000 Nederlanders kanker vast gesteld, wat een stijging betrof van 10% ten op zichte van het voorgaande jaar. De overlevingskans van mensen met kanker neemt toe onder andere door de ontwikkeling van nieuwe behandelingen. De laatste jaren maakt de algemene behandeling van gemetastaseerde kanker met chemotherapie en bestraling, plaats voor 'precision oncology', ofwel een op maat gemaakte behandeling. Hierbij wordt gebruik gemaakt van de sterk in opkomst zijnde doelgerichte medicijnen die alleen effectief zijn tegen tumorcellen met specifieke afwijkingen in het DNA. De meest voorkomende oorzaak van kanker zijn afwijkingen (mutaties) in het DNA van een cel met als gevolg ongecontroleerde en abnormale celgroei. In een minderheid van de gevallen is het ontstaan van kanker geassocieerd met erfelijke DNA afwijkingen (kiembaanmutaties). Het detecteren van deze DNA afwijkingen gebeurt met behulp van tumorweefsel dat verkregen kan worden via een operatie, biopsie of punctie. Uit het tumor weefsel wordt vervolgens DNA geïsoleerd waarop moleculaire diagnostiek wordt toegepast. Echter stijgt door de toenemende ontwikkeling van doelgerichte medicijnen, het aantal en de complexiteit van de DNA afwijkingen waarop getest kan worden (biomarkers) in rap tempo waardoor de logistiek en houdbaarheid van de huidige moleculaire diagnostiek structureel onder druk komt te staan. Hierdoor duurt het lang voordat een nieuwe biomarker kan worden toegevoegd aan reeds bestaande testen in de routine diagnostiek. Dit heeft als gevolg dat er vertragingen optreden in de klinische implementatie van nieuwe biomarkers en derhalve suboptimale toegang tot doelgerichte therapie ontstaat en op termijn ongelijke toegang tot zorg. Bovendien, wordt doelgerichte medicijnen steeds vaker ingezet voor tumor-agnostische indicaties. Dit betekent dat de selectie voor doelgerichte medicijnen enkel berust op de aanwezigheid van bepaalde DNA afwijkingen en ongeacht is van waar de tumor zijn oorsprong vindt. De huidige tumor type afhankelijke moleculaire diagnostische aanpak is minder wenselijk aangezien het patiënten beperkt bij het vinden van alle potentiele behandelingsmogelijkheden. Er is dus een grote behoefte aan één techniek die alle DNA afwijkingen van de tumor (ongeacht het tumor type) kan identificeren. Een allesomvattende techniek, Whole Genome Sequencing (WGS), is de laatste jaren door de snel dalende kosten toegankelijk geworden voor implementatie in de routine diagnostiek. Dit proefschrift onderzocht de klinische implementatie van WGS in de routine klinische zorg (**deel 1**) en de toepassing van (whole) genome sequencing op zeldzame kankers of om zeldzame DNA afwijkingen te detecteren (**deel 2**).

Deel 1: Implementatie van Whole Genome Sequencing in de routine zorg

Tot op heden, was de rol van WGS met name beperkt tot wetenschappelijke doeleinden en had deze techniek nog geen plaats binnen de klinische oncologie. Het toetreden van genomische testen tot de klinische oncologie is grotendeels afhankelijk van een evidence-based beoordeling van de haalbaarheid, kwaliteit, klinische waarde en kosteneffectiviteit van de betreffende test in de routine zorg. Het doel van de *Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient* (WIDE) studie was daarom om bewijs te genereren met betrekking tot het functioneren van WGS in de routine zorg en om een generieke klinische implementatie strategie te ontwikkelen. **Hoofdstuk 2** van dit proefschrift beschrijft het studie protocol van de WIDE studie, een prospectieve observationele cohort studie, welke plaats vond in het Nederlands Kanker Instituut (NKI). De studie beoogde om 1200 opeenvolgende patiënten te includeren met (een verdenking) op uitgezaaide ziekte van solide tumoren zonder preselectie op tumor type en binnen een tijdsperiode van 18-24 maanden. WGS vond plaats in parallel met standard of care (SOC) diagnostiek op tumorweefsel dat was verkregen tijdens routine diagnostische procedures. Het primaire eindpunt, haalbaarheid, werd gedefinieerd als het percentage patiënten voor wie verwerking van biopt tot WGS rapport succesvol was en de doorlooptijd als de tijd van biopt tot WGS rapport in werkdagen. Klinische validiteit, het tweede primaire eindpunt, werd gedefinieerd als het percentage van genetische varianten waarvoor WGS (minimaal) dezelfde klinische relevante varianten detecteerde als moleculaire SOC diagnostiek. Secundaire eindpunten betroffen kosteneffectiviteit, additionele behandelingsopties geïdentificeerd door WGS, beter geïnformeerde klinische besluitvorming en verrijking van de Hartwig database. Kritieke prestatie indicatoren werden geëvalueerd na elke inclusie van 200 patiënten en procedures geoptimaliseerd indien nodig. **Hoofdstuk 3** beschrijft een strategie voor efficiënte implementatie van WGS in de routine pathologie. Als resultaat van de WIDE studie, is WGS succesvol geïmplementeerd in het NKI als onderdeel van de routine diagnostiek sinds januari 2021. De succesvolle klinische implementatie van WGS was sterk afhankelijk van het volgen van een uitgebreid protocol door alle betrokken zorgprofessionals bestaande uit radiologen, pathologen, klinische moleculair biologen in de pathologie, analisten, klinische genetici en medisch oncologen. Bovendien worden belangrijke lessen die getrokken kunnen worden uit het implementatie proces gedurende de WIDE studie besproken, waaronder aanbevelingen om de slagingspercentages van WGS te vergroten d.m.v. gebruik van ≤18 gauge naalden, afnemen van meerdere (2-4) biopten en maximalisatie van de opbrengst van tumor materiaal door een bepaalde manuele microdissectie techniek genaamd 4FME (fresh-frozen fine fixed MCC embedded) toe

te passen. In **hoofdstuk 4** worden de resultaten van de WIDE studie met betrekking tot de haalbaarheid, klinische validiteit en klinische waarde van WGS in de routine pathologie gepresenteerd. De resultaten laten zien dat WGS analyse haalbaar is voor 71% van de patiënten met gemetastaseerde kanker met een mediane doorlooptijd van 11 werkdagen. De voornaamste reden voor ongeschiktheid voor WGS analyse was een te laag tumor cel percentage. De klinische validiteit van WGS bleek uit een biomarker detectie percentage van 99.2% (tweezijdige 95CI 98.4-99.6%). In lijn met voorgaande studies, identificeerde WGS klinisch relevante biomarkers voor reguliere behandelingsmogelijkheden of klinische trials voor 71% van de patiënten. Van deze patiënten, was 24% gestart met biomarker gebaseerde therapie bij een mediane follow-up van 14 maanden. WGS liet ook additionele diagnostische waarde zien in de kiembaan diagnostiek (erfelijkheidsonderzoek) door 49 pathogene kiembaan varianten te identificeren die eerder onbekend waren. Uit de resultaten van de WIDE-studie volgt dat WGS een klinische valide techniek is die additionele klinisch relevante biomarkers identificeert voor reguliere behandelingsmogelijkheden of klinische trials. **Hoofdstuk 5** onderzocht prospectief de klinische waarde van WGS in de routine diagnostiek van 72 patiënten met een primaire tumor onbekend (PTO). PTO was gedefinieerd als een tumor met een onbekende origine of type histologie bij patiënten met gemetastaseerde kanker. Uit eerder onderzoek blijkt dat genomische eigenschappen zoals mutational signatures en topologische distributie van driver en passenger mutaties gebruikt kunnen worden om nauwkeuring het tumor type te voorspellen. Dit hoofdstuk beschrijft de ontwikkeling en validatie van een WGS gebaseerd 'cancer of unknown primary prediction algorithm' (CUPPA). Dit algoritme combineert tumor type specifieke driver mutaties, regionale mutatie dichtheid en mutatie profiel karakteristieken. CUPPA werd getraind en gevalideerd op een grote WGS database van patiënten met gemetastaseerde kanker (4.509), welke willekeurig werd verdeeld in een referentie set (90%, n=4.058) en een test set (10%, n=451), en op 254 onafhankelijke patiënten. CUPPA was in staat om respectievelijk 84% en 78% van de monsters in het interne (n=451) en het onafhankelijke validatie cohort (n=254) correct te classificeren. Voor de 72 patiënten met PTO die WGS ondergingen als onderdeel van de routine diagnostiek, kon WGS en CUPPA het primaire tumor type bepalen voor 49 patiënten (68%) en klinisch relevante biomarkers detecteren voor 34 patiënten (47%). Veelvoorkomende diagnosen bestonden uit niet kleincellige longkanker (n=7), maag/slokdarm kanker (n=4) en alvleesklier kanker (n=4). Concluderend, bleek WGS als onderdeel van de routine diagnostiek van patiënten met PTO een waardevol diagnostisch hulpmiddel te zijn.

Deel 2: Toepassing van (whole) genome sequencing op zeldzame tumor typen of om zeldzame genetische afwijkingen te detecteren

Een kankersoort is zeldzaam wanneer deze bij minder dan 6 op de 100.000 mensen per jaar gevonden wordt. In Nederland, betreft 18% van alle gediagnosticeerde kankers een zeldzame kanker. De vijfjaarsoverleving van zeldzame kankers is slechter dan voor veelvoorkomende kankers (52.0% vs. 68.7%). Dit overlevingsverschil kan deels verklaard worden doordat de DNA afwijkingen die ten grondslag liggen aan het ontstaan van deze tumoren onvoldoende opgehelderd zijn en er een gebrek is aan effectieve behandelingsmogelijkheden. Dunne darm neuroendocriene tumoren (SI-NETs) zijn zeldzame tumoren met een jaarlijkse incidentie van 1.05 per 100.000 personen. **Hoofdstuk 6** biedt een overzicht van de literatuur met als doel om meer inzicht te verkrijgen in het genomische landschap van SI-NETs en om prognostische moleculaire factoren te identificeren. SI-NETs hebben een lage mutational burden, met een verlies van heterozygositeit van chromosoom 18 (44-100%) als meest voorkomende genomische afwijking, gevolgd door mutaties van *CKDN1B* (8%). Andere mutaties die zijn geïdentificeerd in SI-NETs zijn mutaties in *APC*, *CDKN2C* (beiden 7.7%) en *BRAF*, *KRAS*, *PIK3CA* en *TP53* (elk 3.8%). In vergelijking met genetische mutaties, worden epigenetische veranderingen vaker gezien in SI-NETs. Door epigenetische veranderingen wordt de functie van een gen anders, zonder dat de DNA-code verandert. Bij SI-NETs, werd hypermethylering gezien van *RASSF1A*, *SEMA3F* en *CTNNB1* waardoor de transcriptie wordt gesilenced. Verlies van heterozygositeit van chromosoom 18, gain van chromosoom 4, 5, 4, 5, 7, 14 en 20p, copy gain van het *SRC* gene en verminderde expressie van RASSF1A en P16 waren geassocieerd met een slechtere overleving. Tot op heden zijn er nog geen prognostische biomarkers geïdentificeerd die gemakkelijk vertaald zouden kunnen worden naar de huidige klinische besluitvorming. Gezien epigenetische dysregulatie vaker lijkt voor te komen in SI-NETs, vormt de DNA methylerings machinerie mogelijk een meer veelbelovend doelwit voor precision oncology. Voortbouwend op hoofdstuk 6, wordt in **hoofdstuk 7** de aanwezigheid en betekenis van driver mutaties in metastasen van goed gedifferentieerde SI-NETs verder onderzocht. Er werd WGS uitgevoerd op 35 gemetastaseerde SI-NETs en next generation sequencing (NGS) op 8 gemetastaseerde SI-NETs. Het cohort bestond uit 21% graad 1, 60% graad 2 en 19% graad 3 SI-NETs. Driver mutaties werden geïdentificeerd in ongeveer 50% van de SI-NETs. In totaal, werden 27 driver mutaties geïdentificeerd, waarvan 74% aanwezig waren in tumor suppressor genen (bv. *TP53, RB1* en *CDKN1B*) en 22% in proto-oncogenen (bv. *KRAS*, *NRAS*, en *MET*). Allelisch verlies van chromosoom 18 (63%), compleet verlies van *CDKN2A* en *CDKN1B* (beiden 6%) en *CDKN1B* mutaties (9%) waren het meest voorkomend. Alle tumoren waren microsatelliet stabiel en hadden lage tumor mutational burdens (mediaan 1.10; interkwartiel afstand 0.87-1.35). De

Ki67 proliferatie index was significant geassocieerd met de aanwezigheid van driver mutaties (P = 0.015). Er was geen significante correlatie tussen mutatie status en SSTR2a expressie of ziekte specifieke overleving. DNA veranderingen die potentieel kunnen worden behandeld met doelgerichte therapie, werden gedetecteerd in 21% van de gemetastaseerde SI-NETs wat de introductie van doelgerichte therapie in de behandeling van SI-NETs ondersteunt. **Hoofdstuk 8** onderzocht het functioneren van fluorescentie in situ hybridisatie (FISH) als een diagnostische test voor het detecteren van functionele *RET* fusies. De incidentie van *RET* fusies is ongeveer 1% in niet kleincellige long kankers (NSCLCs). Nauwkeurige detectie van *RET* fusies is belangrijk om optimale therapie selectie te bewerkstelligen. In totaal werden er 4.873 patiënten met NSCLC routinematig gescreend op *RET* fusies in zes Europese kanker centra gebruikmakend van FISH (n=2.858) of targeted RNA NGS fusie analyse (n=2.015). Indien mogelijk werden positieve gevallen getest met beide technieken (n=39). FISH analyse toonde *RET* herschikkingen aan in 48 van de 2.858 gevallen. Van de 30 gevallen die ook werden getest met NGS fusie analyse, hadden enkel 9 een functionele *RET* fusie. In totaal, detecteerde NGS fusie analyse *RET* fusies in 14 van de 2.015 gevallen, alle negen gevallen die dubbel werden getest met FISH hadden een herschikking van het *RET* locus. Deze resultaten werden onderbouwd door WGS data van een onafhankelijk cohort van 520 patiënten met NSCLC, waarbij werd gezien dat minder dan de helft van de disruptieve *RET* locus herschikkingen leidt tot een functionele *RET* fusie. Ten slotte, FISH is een gevoelige maar niet specifieke techniek met als gevolg een groot aantal vals-positieve resultaten. *RET* FISH herschikkingen moeten derhalve altijd bevestigd worden met een orthogonale techniek (bv. RNA NGS of WGS) voor optimale selectie van patiënten die mogelijk baat hebben bij RET doelgerichte therapie.

APPENDICES

Table of contents – Data supplement

Data supplement Chapter 2

Study protocol: Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient (WIDE)

Supplementary material

WIDE project - Survey clinicians – Baseline

Introduction

As part of the WIDE-project, we would like to (in terms of a qualitative analysis) evaluate how treating clinicians experience the value of molecular diagnostic (reporting) based on Whole Genome Sequencing (WGS) with regards to clinical decision making compared to the current reporting of standard of care (SOC) molecular diagnostics. This survey will therefore be performed at the start, midterm and at the end of the WIDE-project based on the questionnaire below. The questionnaires are anonymous.

We want to kindly request you to fill out this first survey *within 2 weeks*.

The second and last survey will follow after 6 months and at the end of the project (after 1 year). You will receive an automatic message as a reminder.

Filling out the questionnaire will take 5 minutes at maximum.

On behalf of the entire team of the WIDE-project, we want to thank you in advance for your contribution!

Explanatory word list:

- *WGS* = Whole Genome Sequencing
- *SOC:* standard molecular diagnostics (for example targeted panel sequencing, Fluorescent in Situ Hybridization (FISH), immunohistochemistry (IHC))

*

Part 1 – Multiple choice and open questions

You can find several multiple choice and open questions below. You only need to fill out one answer, unless otherwise is stated.

1. What type of clinician are you?

- ❑ Medical oncologist
- ❑ Pulmonologist
- ❑ Neurologist
- ❑ Surgeon
- ❑ Radiotherapist
- ❑ Otherwise, ...

2. Why are you performing molecular diagnostics in general?
(Multiple answers are possible)

- ❑ Diagnosis
- ❑ Prognoses
- ❑ Prediction of response to treatment
- ❑ Monitoring of response to treatment
- ❑ Pharmacokinetics
- ❑ Genetic counseling
- ❑ Research purposes
- ❑ Otherwise, ...

3. Are you content with the current offer of standard of care (SOC) molecular diagnostics?

- ❑ Yes
- ❑ No

4. Are you missing anything in the current SOC (in terms of content or reporting)?

- ❑ Yes
- ❑ No

5. If so, can you specify what you are missing in the current SOC (in terms of content or reporting? *(Multiple answers are possible).*

- ❑ Reporting of all findings in one comprehensive report
- ❑ Feedback when the usual turnaround time is exceeded
- ❑ Information on loss of heterozygosity (LOH)
- ❑ Information on (treatable) germline variants

- ❑ Information on tumour mutational load / tumour mutational burden
- ❑ Information on 'gene disruptions' (for example inversions, deletions, insertions, etc.)
- ❑ Fast implementation of new biomarkers
- ❑ Identification of new therapeutic opportunities (off label / clinical trials)
- ❑ Otherwise, ...

6. Do you have prior experience with reporting based on Whole Genome Sequencing (WGS)?

- ❑ Yes
- ❑ No

7. Do you think molecular diagnostics by means of WGS has added value for clinical decision making, compared to SOC?

- ❑ Yes
- ❑ No

8. If so, can you indicate what the added value would be for you (in terms of content and reporting)? *(Multiple answers are possible)*

- ❑ All genomic information and potential actionable targets are summarized in one report
- ❑ I can immediately apply all new biomarkers clinically (without the need for validation of a new technique)
- ❑ It can help me identify new therapeutic possibilities for patients for whom regular therapeutic options are no longer available
- ❑ I receive information on tumour mutational load/ tumour mutational burden.
- ❑ I receive information on gene disruptions (for example inversions, deletions, insertions)
- ❑ I receive information on gene amplifications and deletions
- ❑ I receive information on treatable germline variants
- ❑ I receive information on loss of heterozygosity (LOH)
- ❑ Otherwise, ...
- ❑ Not applicable

9. If no, can you indicate why you think that WGS will have no added value for you (in terms of content and reporting)? *(Multiple answers are possible).*

- ❑ I think the turn around time from biopsy to result takes too long
- ❑ I think the extensive report based on WGS information is too complex to interpret
- ❑ I receive too much information, also information of which the clinical relevance is unknown to me (for example 'variants of unknown significance')

*

- ❑ I think the WGS information is not presented clear enough
- ❑ I don't use new biomarkers in general (as long as these are not adopted into the guidelines)
- ❑ Otherwise, ...
- ❑ Not applicable

10. To what extent according to you does the clinical utility of a new biomarker (found with WGS) has to be proven, prior to implementation into routine practice?

- ❑ At least with evidence of preclinical research
- ❑ At least with evidence of retrospective clinical research
- ❑ At least with evidence of prospective/observational clinical research
- ❑ At least with evidence of randomized clinical research
- ❑ Otherwise, ...

11. To what extent you think that *the profession* needs the clinical utility of a new biomarker (found with WGS) to be proven, in order for the biomarker to be adopted in the guidelines?

- ❑ At least with evidence of preclinical research
- ❑ At least with evidence of retrospective clinical research
- ❑ At least with evidence of prospective/observational clinical research
- ❑ At least with evidence of randomized clinical research
- ❑ Otherwise, ...

12. Do you think a cost-effectiveness analysis with regards to WGS will be of added value?

- ❑ I'm neutral
- ❑ I think this information is important for the adoption of WGS in case of relevant medical indications in the basic insurance package
- ❑ I think this has no added value, to prove clinical utility is sufficient
- ❑ Otherwise, ...

Part 2 – Statements

The statements below concern the added value in terms of content and reporting of molecular diagnostics based on WGS compared to SOC.

Indicate below to what extent you agree with the following statements:

1 = I very much disagree, 2 = I disagree, 3 = neutral, 4 = I agree,
5 = I very much agree

1. I think that the turn around time from biopsy to the result of sequencing is too long
2. I think that the WGS report will be too complex to interpret
3. I think that the WGS report will be too complex to explain to my patient
4. I think that I will have insufficient time to discuss the full WGS report with my patient
5. I don't know how to deal with possible (germline) incidental findings (for example BRCA 1 / 2 mutations, MSI of non-oncological findings)
6. I think that WGS is better than SOC because all possible molecular diagnostic tests are combined in one, therefore I receive one *report* in which all results are summarized
7. I think WGS is better than SOC because new biomarkers can be immediately applied clinically (without the need for validation of a new technique)
8. I think that WGS report will help me identify new therapeutic possibilities for patients for whom regular therapeutic options are no longer available (either off-label, or in clinical trials)
9. I think WGS will have *no* added value for me in my current clinical decision making (as compared to SOC)

Part 3 – Additional comments / reactions

Below there are four last open questions.

1. **Can you indicate shortly, and if so how, the reporting of molecular diagnostics (SOC and/or WGS based) is being discussed and interpreted within your organization?**

   ……..
   ……..
   ……..
   ……..

2. **If there was no barrier whatsoever to use WGS, when and for which indications would you use WGS?**

   ……..
   ……..
   ……..
   ……..

3. **Do you have additional comments or suggestions concerning the WIDE project you would like to share?**

   ……..
   ……..
   ……..
   ……..

4. **Are there questions or statements you have missed in this questionnaire?**

   ……..
   ……..
   ……..
   ……..

WIDE project - Survey clinicians – Midterm

Introduction

As part of the WIDE-project, we would like to (in terms of a qualitative analysis) evaluate how treating clinicians experience the value of molecular diagnostic (reporting) based on Whole Genome Sequencing (WGS) with regards to clinical decision making compared to the current reporting of standard of care (SOC) molecular diagnostics. This survey will therefore be performed at the start, midterm and at the end of the WIDE-project based on the questionnaire below. The questionnaires are anonymous. The first survey you have filled out at the start of the project. This is the second survey. The last survey will follow at the end of the project (after 1 year). You will receive an automatic message as a reminder.

We want to kindly request you to fill out this first survey *within 2 weeks*.

Filling out the questionnaire will take 5 minutes at maximum.

On behalf of the entire team of the WIDE-project, we want to thank you in advance for your contribution!

Explanatory word list:

- *WGS* = Whole Genome Sequencing
- *SOC:* standard molecular diagnostics (for example targeted panel sequencing, Fluorescent in Situ Hybridization (FISH), immunohistochemistry (IHC))

*

Part 1 – Multiple choice and open questions

You can find several multiple choice and open questions below. You only need to fill out one answer, unless otherwise is stated.

1. What type of clinician are you?

- ❑ Medical oncologist
- ❑ Pulmonologist
- ❑ Neurologist
- ❑ Surgeon
- ❑ Radiotherapist
- ❑ Otherwise, ...

2. Are you content with the current offer of standard of care (SOC) molecular diagnostics?

- ❑ Yes
- ❑ No

3. Are you missing anything in the current SOC (in terms of content or reporting)?

- ❑ Yes
- ❑ No

4. If so, can you specify what you are missing in the current SOC (in terms of content or reporting? *(Multiple answers are possible).*

- ❑ Reporting of all findings in one comprehensive report
- ❑ Feedback when the usual turnaround time is exceeded
- ❑ Information on loss of heterozygosity (LOH)
- ❑ Information on (treatable) germline variants
- ❑ Information on tumour mutational load / tumour mutational burden
- ❑ Information on 'gene disruptions' (for example inversions, deletions, insertions, etc.)
- ❑ Fast implementation of new biomarkers
- ❑ Identification of new therapeutic opportunities (off label / clinical trials)
- ❑ Otherwise, ...

5. Do you think molecular diagnostics by means of WGS has added value for clinical decision making, compared to SOC?

- ❑ Yes
- ❑ No

6. If so, can you indicate what the added value would be for you (in terms of content and reporting)? (*Multiple answers are possible.)*

- ❑ All genomic information and potential actionable targets are summarized in one report
- ❑ I can immediately apply all new biomarkers clinically (without the need for validation of a new technique)
- ❑ It can help me identify new therapeutic possibilities for patients for whom regular therapeutic options are no longer available
- ❑ I receive information on tumour mutational load/ tumour mutational burden
- ❑ I receive information on gene disruptions (for example inversions, deletions, insertions)
- ❑ I receive information on gene amplifications and deletions
- ❑ I receive information on treatable germline variants
- ❑ I receive information on loss of heterozygosity (LOH).
- ❑ Otherwise,
- ❑ Not applicable

7. If no, can you indicate why you think that WGS will have no added value for you (in terms of content and reporting)? *(Multiple answers are possible).*

- ❑ I think the turn around time from biopsy to result takes too long
- ❑ I think the extensive report based on WGS information is too complex to interpret
- ❑ I receive too much information, also information of which the clinical relevance is unknown to me (for example 'variants of unknown significance')
- ❑ I think the WGS information is not presented clear enough
- ❑ I don't use new biomarkers in general (as long as these are not adopted into the guidelines)
- ❑ Otherwise,
- ❑ Not applicable

8. To what extent according to you does the clinical utility of a new biomarker (found with WGS) has to be proven, prior to implementation into routine practice?

- ❑ At least with evidence of preclinical research
- ❑ At least with evidence of retrospective clinical research
- ❑ At least with evidence of prospective/observational clinical research
- ❑ At least with evidence of randomized clinical research
- ❑ Otherwise,

9. To what extent you think that *the profession* needs the clinical utility of a new biomarker (found with WGS) to be proven, in order for the biomarker to be adopted in the guidelines?

- ❑ At least with evidence of preclinical research
- ❑ At least with evidence of retrospective clinical research
- ❑ At least with evidence of prospective/observational clinical research
- ❑ At least with evidence of randomized clinical research
- ❑ Otherwise,

10. Do you think a cost-effectiveness analysis with regards to WGS will be of added value?

- ❑ I'm neutral
- ❑ I think this information is important for the adoption of WGS in case of relevant medical indications in the basic insurance package
- ❑ I think this has no added value, to prove clinical utility is sufficient
- ❑ Otherwise,

11. How many WGS reports have you received in the last months from the WIDE project?

- ❑ 0 – 5
- ❑ 5 – 15
- ❑ 15 – 30
- ❑ > 30

12. Has the WGS report aided you to identify new therapeutic possibilities for patients for whom regular treatment options were no longer available (either off-label, or clinical trials)?

- ❑ No, never
- ❑ Yes, for about 10% of my patients who participated
- ❑ Yes, for about 25% of my patients who participated
- ❑ Yes, for half of my patients who participated
- ❑ Yes, for the majority of my patients who participated
- ❑ Yes, for *all* of my patients who participated

13. Have you used the WGS report to immediately apply new biomarkers into clinical practice?

- ❑ Yes
- ❑ No

14. Did you find it difficult to independently interpret the WGS report?

❑ Yes

❑ No

If so, can you indicate what needs to be improved to make it easier for you to interpret the WGS report?

.....

.....

.....

.....

*

Part 2 – Statements

The statements below concern the added value in terms of content and reporting of molecular diagnostics based on WGS compared to SOC.

Indicate below to what extent you agree with the following statements:

1 = I very much disagree, 2 = I disagree, 3 = neutral, 4 = I agree,
5 = I very much agree

1. The turn aorund time from biopsy to results of WGS has to be within 2 weeks
2. I think the WGS report is mostly too complex to independently interpret
3. I think the WGS report is mostly too complex to explain to my patient
4. I have insufficient time to discuss the full WGS report with my patient
5. I do not know how to deal with possible (germline) incidental findings (for example BRCA 1 / 2 mutations, MSI of nom-oncological findings)
6. I prefer WGS over SOC because all possible molecular diagnostic tests are combined in one, therefore I receive one *report* in which all results are summarized
7. I prefer WGS over SOC because new biomarkers can be immediately applied clinically (without the need for validation of a new technique)
8. The WGS rapport has helped me identify new therapeutic possibilities for patients for whom regular therapeutic options are no longer available (either off-label, or in clinical trials)
9. WGS has *no* added value for me in my current clinical decision making (as compared to SOC)

Part 3 – Additional comments/reactions

Below there are three last open questions.

1. **If there was no barrier whatsoever to use WGS, *when* and *for which* indications would you use WGS?**

   ……..
   ……..
   ……..
   ……..

2. **Do you have additional comments or suggestions concerning the WIDE project you would like to share?**

   ……..
   ……..
   ……..
   ……..

3. **Are there questions or statements you have missed in this questionnaire?**

   ……..
   ……..
   ……..
   ……..

*

WIDE project - Survey clinicians – End of study

Introduction

As part of the WIDE-project, we would like to (in terms of a qualitative analysis) evaluate how treating clinicians experience the value of molecular diagnostic (reporting) based on Whole Genome Sequencing (WGS) with regards to clinical decision making compared to the current reporting of standard of care (SOC) molecular diagnostics. This survey will therefore be performed at the start, midterm and at the end of the WIDE-project based on the questionnaire below. The questionnaires are anonymous. The first two surveys you have filled out at the start and midterm of the project. This will be the last survey.

We want to kindly request you to fill out this first survey *within 2 weeks*.

Filling out the questionnaire will take 5 minutes at maximum.

On behalf of the entire team of the WIDE-project, we want to thank you in advance for your contribution!

Explanatory word list:

- *WGS* = Whole Genome Sequencing
- *SOC:* standard molecular diagnostics (for example targeted panel sequencing, Fluorescent in Situ Hybridization (FISH), immunohistochemistry (IHC))

Part 1 – Multiple choice and open questions

You can find several multiple choice and open questions below. You only need to fill out one answer, unless otherwise is stated.

1. What type of clinician are you?

- ❑ Medical oncologist
- ❑ Pulmonologist
- ❑ Neurologist
- ❑ Surgeon
- ❑ Radiotherapist
- ❑ Otherwise, ...

2. Are you content with the current offer of standard of care (SOC) molecular diagnostics?

- ❑ Yes
- ❑ No

3. Are you missing anything in the current SOC (in terms of content or reporting)?

- ❑ Yes
- ❑ No

4. If so, can you specify what you are missing in the current SOC (in terms of content or reporting? *(Multiple answers are possible).*

- ❑ Reporting of all findings in one comprehensive report
- ❑ Feedback when the usual turnaround time is exceeded
- ❑ Information on loss of heterozygosity (LOH)
- ❑ Information on (treatable) germline variants
- ❑ Information on tumour mutational load / tumour mutational burden
- ❑ Information on 'gene disruptions' (for example inversions, deletions, insertions, etc.)
- ❑ Fast implementation of new biomarkers
- ❑ Identification of new therapeutic opportunities (off label / clinical trials)
- ❑ Otherwise, ...

5. Do you think molecular diagnostics by means of WGS has added value for clinical decision making, compared to SOC?

- ❑ Yes
- ❑ No

*

6. If so, can you indicate what the added value would be for you (in terms of content and reporting)? (*Multiple answers are possible.)*

- ❑ All genomic information and potential actionable targets are summarized in one report
- ❑ I can immediately apply all new biomarkers clinically (without the need for validation of a new technique)
- ❑ It can help me identify new therapeutic possibilities for patients for whom regular therapeutic options are no longer available
- ❑ I receive information on tumour mutational load/ tumour mutational burden.
- ❑ I receive information on gene disruptions (for example inversions, deletions, insertions)
- ❑ I receive information on gene amplifications and deletions
- ❑ I receive information on treatable germline variants
- ❑ I receive information on loss of heterozygosity (LOH)
- ❑ Otherwise, ...
- ❑ Not applicable

7. If no, can you indicate why you think that WGS will have no added value for you (in terms of content and reporting)? *(Multiple answers are possible).*

- ❑ I think the turn around time from biopsy to result takes too long
- ❑ I think the extensive report based on WGS information is too complex to interpret
- ❑ I receive too much information, also information of which the clinical relevance is unknown to me (for example 'variants of unknown significance')
- ❑ I think the WGS information is not presented clear enough
- ❑ I don't use new biomarkers in general (as long as these are not adopted into the guidelines)
- ❑ Otherwise, ...
- ❑ Not applicable

8. To what extent according to you does the clinical utility of a new biomarker (found with WGS) has to be proven, prior to implementation into routine practice?

- ❑ At least with evidence of preclinical research
- ❑ At least with evidence of retrospective clinical research
- ❑ At least with evidence of prospective/observational clinical research
- ❑ At least with evidence of randomized clinical research
- ❑ Otherwise

9. To what extent you think that *the profession* needs the clinical utility of a new biomarker (found with WGS) to be proven, in order for the biomarker to be adopted in the guidelines?

- ❑ At least with evidence of preclinical research
- ❑ At least with evidence of retrospective clinical research
- ❑ At least with evidence of prospective/observational clinical research
- ❑ At least with evidence of randomized clinical research
- ❑ Otherwise

10. Do you think a cost-effectiveness analysis with regards to WGS will be of added value?

- ❑ I'm neutral
- ❑ I think this information is important for the adoption of WGS in case of relevant medical indications in the basic insurance package
- ❑ I think this has no added value, to prove clinical utility is sufficient
- ❑ Otherwise, ...

11. How many WGS reports have you received in the last months from the WIDE project?

- ❑ 0 – 5
- ❑ 5 – 15
- ❑ 15 – 30
- ❑ > 30

12. Has the WGS report aided you to identify new therapeutic possibilities for patients for whom regular treatment options were no longer available (either off-label, or clinical trials)?

- ❑ No, never
- ❑ Yes, for about 10% of my patients who participated
- ❑ Yes, for about 25% of my patients who participated
- ❑ Yes, for half of my patients who participated
- ❑ Yes, for the majority of my patients who participated
- ❑ Yes, for *all* of my patients who participated

13. Have you used the WGS report to immediately apply new biomarkers into clinical practice?

- ❑ Yes
- ❑ No

*

14. Did you find it difficult to independently interpret the WGS report?

❑ Yes

❑ No

If so, can you indicate what needs to be improved to make it easier for you to interpret the WGS report?

.....

.....

.....

.....

Part 2 – Statements

The statements below concern the added value in terms of content and reporting of molecular diagnostics based on WGS compared to SOC.

Indicate below to what extent you agree with the following statements:

1 = I very much disagree, 2 = I disagree, 3 = neutral, 4 = I agree, 5 = I very much agree

1. The turn aorund time from biopsy to results of WGS has to be within 2 weeks
2. I think the WGS report is mostly too complex to independently interpret
3. I think the WGS report is mostly too complex to explain to my patient
4. I have insufficient time to discuss the full WGS report with my patient
5. I do not know how to deal with possible (germline) incidental findings (for example BRCA 1 / 2 mutations, MSI of nom-oncological findings)
6. I prefer WGS over SOC because all possible molecular diagnostic tests are combined in one, therefore I receive one *report* in which all results are summarized
7. I prefer WGS over SOC because new biomarkers can be immediately applied clinically (without the need for validation of a new technique)
8. The WGS rapport has helped me identify new therapeutic possibilities for patients for whom regular therapeutic options are no longer available (either off-label, or in clinical trials)
9. WGS has *no* added value for me in my current clinical decision making (as compared to SOC)

Part 3 – Additional comments/reactions

Below there are three last open questions.

1. **If there was no barrier whatsoever to use WGS, *when* and *for which* indications would you use WGS?**

   ……..
   ……..
   ……..
   ……..

2. **Do you have additional comments or suggestions concerning the WIDE project you would like to share?**

   ……..
   ……..
   ……..
   ……..

3. **Are there questions or statements you have missed in this questionnaire?**

   ……..
   ……..
   ……..
   ……..

Data supplement Chapter 3

Optimization of the workflow for Whole Genome Sequencing in routine pathology practice

Supplementary 1. Hartwig registration form (PDF)

Supplementary 2. Example of WGS report (PDF)

Data supplement Chapter 4

Feasibility of Whole Genome Sequencing based diagnostics in routine pathology practice

Supplementary materials and methods

MolDx portfolio at the NKI

The SOC molecular diagnostic arsenal at NKI comprises targeted DNA next generation panel sequencing (NGS) (Ampliseq, Cancer hotspot panel V2, Illumina Inc, San Diego, CA, USA, RNA-based NGS fusion analysis (Archer Fusionplex, Lung and Sarcoma panels, Archer DX Inc, Boulder, CO, USA), as well as Sanger sequencing, reverse transcriptase polymerase chain reaction analysis, multiplex fragment analysis polymerase chain reaction, high resolution melting and fluorescence in situ hybridization. The genes covered by NGS modified Ampliseq panel and Fusionplex Lung and Sarcoma panel can be found in supplementary material, Tables S2–S4. SOC germline diagnostics were performed using dedicated assays for specific genes in selected patients based on clinicopathological characteristics (e.g. gender, tumor type, family history) in line with national clinical guidelines and were not influenced by WIDE study participation.

WGS bioinformatics

Initially, shallow WGS (8 – 12 x average depth WGS) was used to determine purity of tumor samples before continuing full sequencing at an average depth of >90x. This approach was abandoned during the study to optimize turn around time (TAT. Platinum is completely open source and free to use and consists of open source tools optimized for diagnostic use (code available through github.com/hartwigmedical). Platinum was designed to detect all types of somatic alterations, including single and multiple nucleotide substitutions (SNV and MNV), insertions and deletions (indels), copy number alterations (amplifications and gene copy losses), genomic rearrangements and structural variants (e.g. gene fusions) and includes advanced analytics including tumor adjusted variant allele frequency, bi-allelic loss of tumor suppressors, gene inactivation due to copy neutral structural variants and complex biomarkers annotation like DNA repair status (HRD, MSI) and mutation load [24]. WGS from germline DNA was included in the bioinformatic data analyses to discriminate inherited germline variants from somatic variants, which enabled accurate reporting of mutation acquired in the tumor, but also to identify relevant Supplementary actionable germline variants.

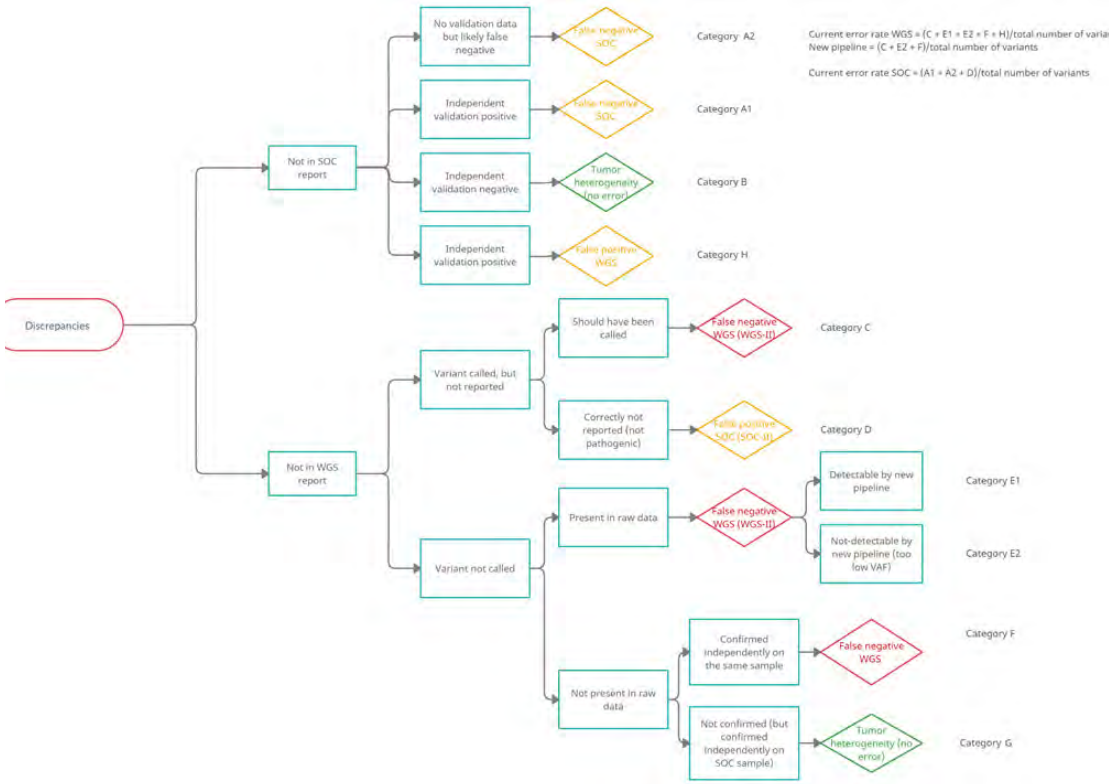


Figure S1. Predefined workflow for resolving discordant results between WGS and SOC molecular diagnostics

In case of discrepancies, appropriate additional verification tests were performed using the original input samples for WGS and SOC according to the predefined workflow below. Subsequently, the outcomes were discussed in a dedicated molecular tumor board.: SOC = Standard of Care, WGS = Whole Genome Sequencing, VAF = Variant Allele Frequency

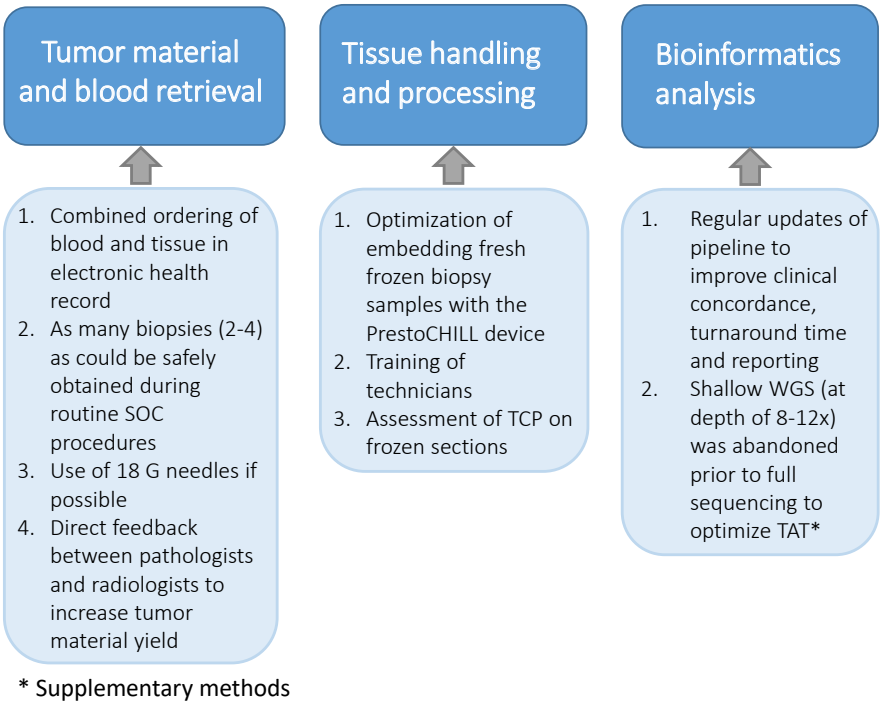


Figure S2. Optimizations of the workflow during the course of the study

The design of the study allowed for continuous evaluation and improvement of procedures, in line with ISO15189:2012. Study progress was evaluated bi-weekly in a multidisciplinary team involving study coordinators, pathologists, radiologists, medical oncologists, clinical geneticists, and support staff. As a result, multiple stages of the process underwent optimizations as shown below. SOC = Standard Of Care, TCP = Tumor Cell percentage, TAT = Turn Around Time

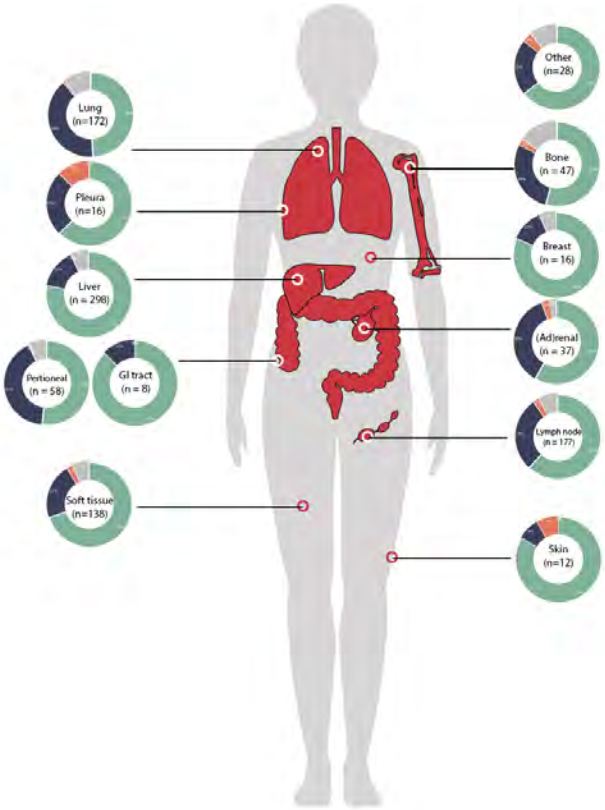


Figure S3. Feasibility of WGS on biopsies per localization.

The feasibility of WGS was largely dependent on biopsy location. The liver was the most common biopsy location, with a relatively high feasibility (78% of all biopsies). Other biopsy locations with a high feasibility included lymph nodes (n=177, 62%) and soft tissue (n=138, 70%). Contrastingly, transthoracic (n=168), peritoneal (n=58), and osseous (n=47) biopsies had low success rates (49%, 52%, and 54%, respectively). GI = Gastrointestinal.

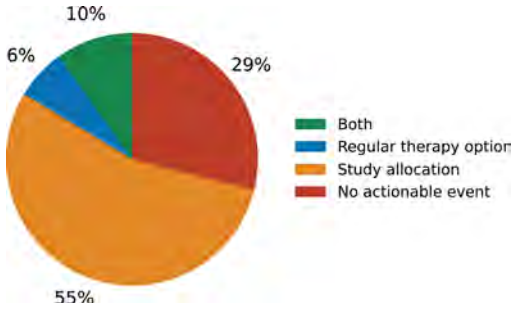


Figure S4. Clinical value of prospective WGS: actionable events.

Clinical value of prospective WGS expressed in actionable events. In 848 patients, 603 (71%) had ≥1 actionable events making them eligible for biomarker–based therapy either in regular or in experimental setting. In 29% of patients no actionable biomarker was identified.

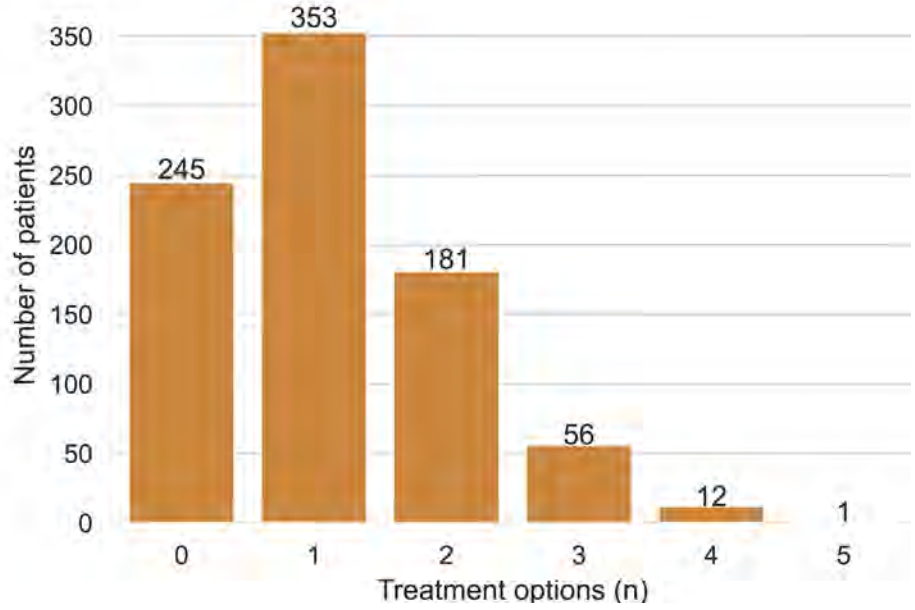


Figure S5. Clinical value of prospective WGS: number of therapy options per patient.

Clinical value of prospective WGS expressed in number of therapy options per patient. 250 patients had more than one therapy option identified by WGS and/or SOC molecular diagnostics.

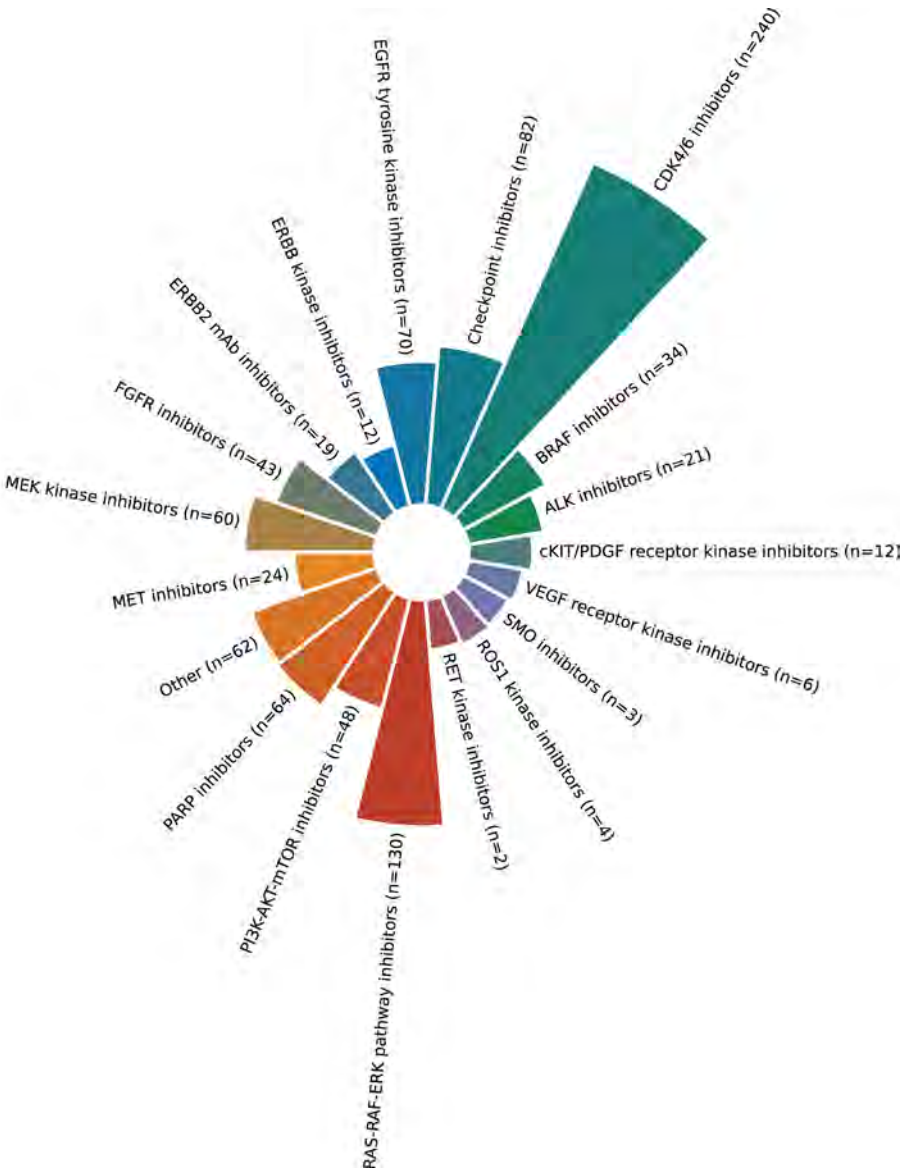


Figure S6. Clinical value of prospective WGS: therapy options

Clinical value of prospective WGS expressed in therapy options. The actionable events identified by WGS and/or SOC molecular diagnostics amounted to a total of 936 therapy options with multiple targeted therapies.

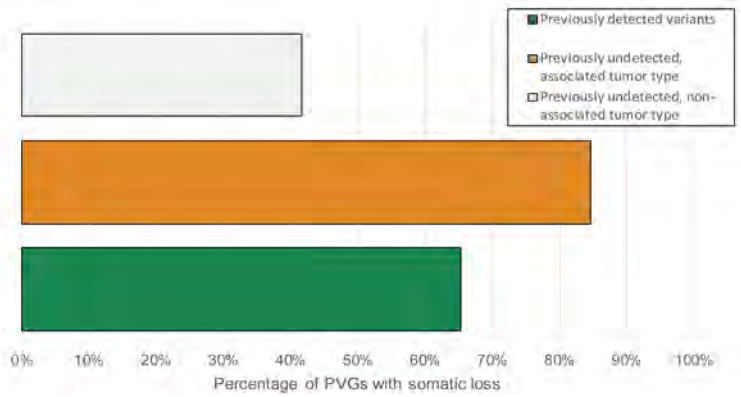


Figure S7. Pathogenic germline variants with somatic loss.

Somatic losses were frequently observed in prototypic hereditary cancer tumor types (e.g. BRCA loss in breast or ovarian cancer) as is shown in orange. PGV = Pathogenic Germline Variant

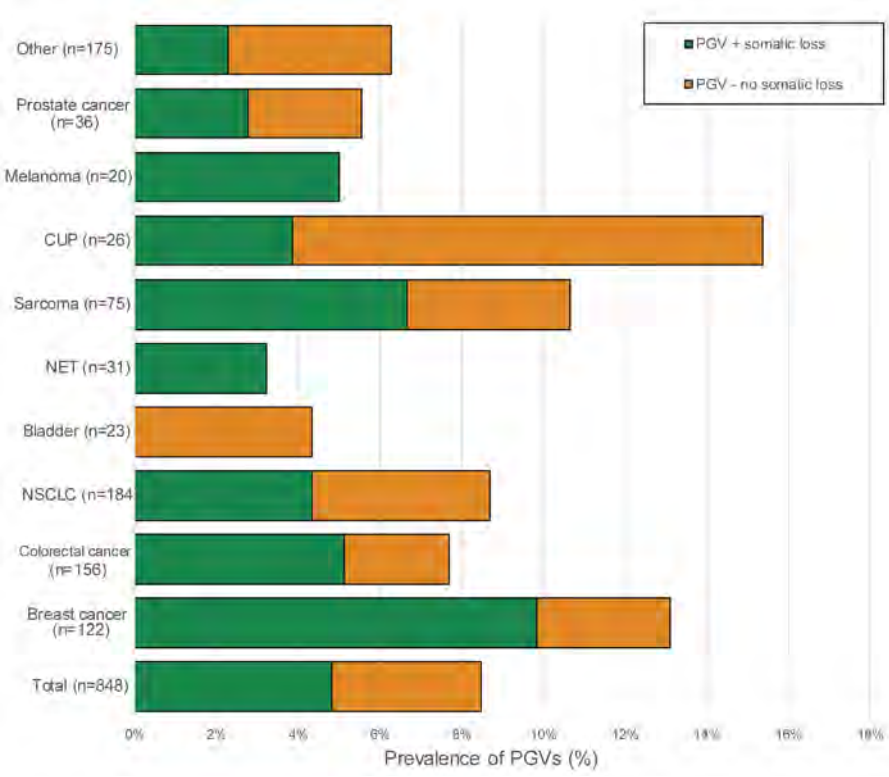


Figure S8. Pathogenic germline variants per tumor type

Pathogenic germline variants with/without somatic loss were present across all tumor types. PGV = Pathogenic Germline Variant

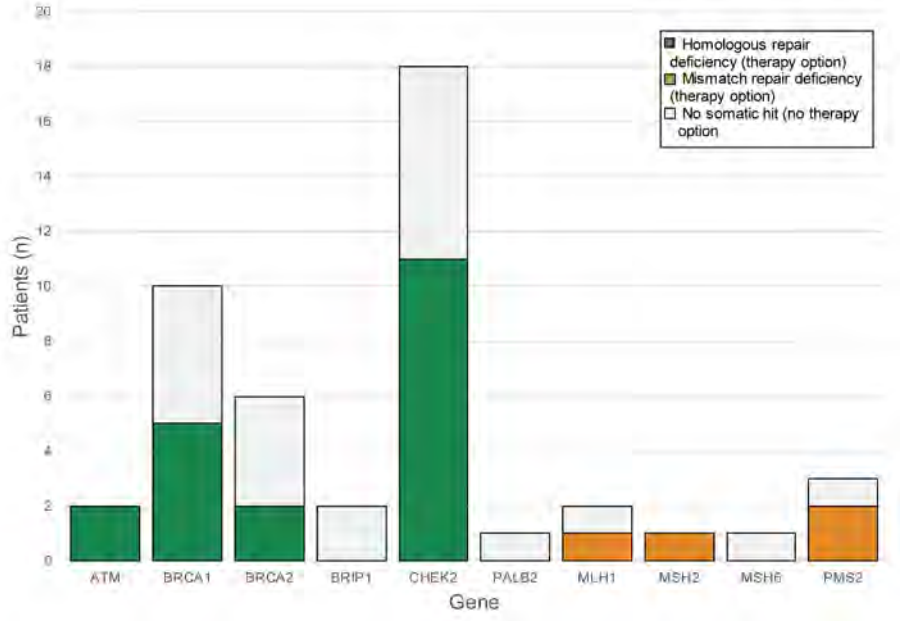


Figure S9. Pathogenic germline variants with therapy options.

Second hits in pathogenic germline variants with somatic loss provided a rationale for therapy options in 24 patients.

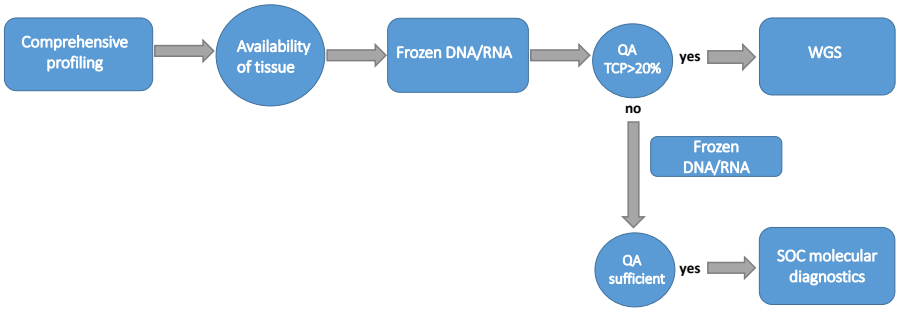


Figure S10. Cascade strategy for molecular diagnostics.

Based on the availability and quality of tissue, the diagnostic modality with the highest information return can be selected. With this strategy, it is guaranteed to retrieve as much information out of the available material as possible. QA = Quality Assessment, TCP = Tumor Cell Percentage, WGS = Whole Genome Sequencing, SOC = Standard Of Care

Supplementary Tables S1–S9

Table S1. Genes for which pathogenic germlines were analyzed.

Gene	
APC	NTHL1
ATM*	PALB2*
BAP1	PMS2
BMPR1A	POLD1
BRCA1*	POLE
BRCA2*	PTCH1
BRIP1*	PTEN
CDC73	RAD51C*
CDH1	RAD51D*
CDK4	RB1
CDKN2A*	RET
CHEK2*	SDHA
EPCAM	SDHAF2
FH	SDHB
FLCN	SDHC
KIT*	SDHD
MEN1	SMAD4
MET	SMARCB1
MITF	STK11
MLH1	TP53
MSH2	TSC1
MSH6	TSC2
MUTYH	VHL
NF1	WT1
NF2	

* Pathogenic variants in these genes were reported back as germline findings in case of opt-in for germline reporting

Genes were selected by selecting cancer associated genes on two criteria: 1) diagnostic value and 2) actionability in terms of tumor-directed therapy. Only genes with direct implications for tumor-directed therapy were reported to the patient in case of opt-in for germline reporting.

Table S2. NGS modified Ampliseq panel

Gene	Exon	Codon
ALB1	Exon 4,5,6,7	Codon 252–278, 294–298, 334–378, 399–431
AKT1	Exon 3,6	Codon 17–51, 155–182
ALK	Exon 21,22,23,24,25	Codon 1121–1215, 1242–1278
APC	Exon 17	Codon 861–890, 1090–1125, 1285–1325, 1343–1384, 1427–1470, 1484–1523, 1544–1581
ATM	Exon 8,9,12,17,26,34,35,36,39, 50,54,55,56,59,61,63	Codon 327–355, 408–411, 602–625, 835–864, 1293–1324, 1675–1707, 1727–1756, 1791–1814, 1927–1945, 2437–2453, 2651–2666, 2683–2710, 2718–2736, 2866–2890, 2934–2950, 2997–3025, 3042–3057
BRAF	Exon 11,15	Codon 439–472, 582–610
CDH1	Exon 3,8,9	Codon 66–96, 337–374, 380–408
CDKN2A	Exon 2	Codon 51–89, 98–139
CSF1R	Exon 7,22	Codon 298–318, 954–973
CTNNB1	Exon 3	Codon 10–48
EGFR	Exon 3,7,15,18,19,20,21	Codon 97–123, 280–296, 575–601, 696–725, 729–823, 856–875
ERBB2	Exon 8,16,18,19,20,21	Codon 302–321, 634–648, 695–710, 753–796, 840–881
ERBB4	Exon 3,4,6,7,8,9,15,23	Codon 137–140, 168–185, 226–247, 255–289, 296–322, 334–366, 581–622, 920–947
EZH2	Exon 16	Codon 625–649
FBXW7	Exon 5,8,9,10,11	Codon 265–287, 379–402, 435–472, 479–508, 567–593
FGFR1	Exon 4,6,7,8, 11	Codon 121–149, 218–248, 250–274, 323–360, 484–517
FGFR2	Exon 7,9,12	Codon 251–274, 297–313, 363–398, 546–557
FGFR3	Exon 7,9,14,16,18	Codon 248–277, 368–402, 632–653, 691–719, 772–807
FLT3	Exon 11,14,16,20	Codon 438–465, 570–609, 664–684, 828–847
GNA11	Exon 4,5	Codon 166–179, 203–219
GNAQ	Exon 4,5	Codon 164–201, 206–245
GNAS	Exon 8,9	Codon 196–219, 221–239
HNF1A	Exon 3,4	Codon 193–221, 254–282
HRAS	Exon 2,3	Codon 6–34, 43–81
IDH1	Exon 4	Codon 101–134
IDH2	Exon 4	Codon 134–176
JAK2	Exon 14	Codon 604–621
JAK3	Exon 4,13,16	Codon 129–140, 568–579, 710–733
KDR	Exon 6,7,11,19,21,26,27,30	Codon 245–290, 472–479, 873–893, 961–987, 1136–1155, 1193–1220, 1284–1309, 1325–1357
KIT	Exon 2,9,10,11,13,14,15,17,18	Codon 24–58, 450–487, 495–513, 526–591, 628–660, 665–684, 715–724, 803–828, 833–857

Table S2. Continued

Gene	Exon	Codon
KRAS	Exon 2,3,4	Codon 6–65, 114–149
MET	Exon 2,11,14,16,17,19	Codon 160–187, 817–855, 982–1027, 1106–1131, 1148–1189, 1230–1273
MLH1	Exon 12	Codon 374–414
MPL	Exon 10	Codon 502–521
NOTCH1	Exon 26,27,34	Codon 1566–1601, 1674–1679, 2436–2475
NPM1	Exon 11	Codon 283–295
NRAS	Exon 2,3,4	Codon 4–30, 43–68, 125–150
PDGFRA	Exon 12,14,15,18	Codon 552–583, 645–667, 672–709, 820–854
PIK3CA	Exon 2,5,7,8,10,14,19,21	Codon 55–89, 107–117, 317–350, 391–421, 450–468, 523–549, 678–720, 899–923, 1018–1050, 1066–1069
POLE	Exon 9,10,11,12,13,14	Codon 268–403, 410–433, 460–491
PTEN	Exon 1,3,5,6,7,8	Codon 1–24, 56–69, 100–134, 165–183, 213–215, 232–267, 283–299, 313–342
PTPN11	Exon 3,13	Codon 47–81, 486–527
RB1	Exon 4,6,10,11,14,17	Codon 131–159,197–202,314–344, 351–366, 453–463, 548–565
RET	Exon 10,11,13,15,16	Codon 609–654, 763–785, 876–923
ROS1	Exon 36,37,38	Codon 1936–1975, 1981–2001, 2013–2045
SMAD4	Exon 3,4,5,6,8,9,10,11,12	Codon 99–135, 143–145, 166–202, 243–262, 308–318, 327–364, 385–423, 444–473, 495–532
SMARCB1	Exon 2,4,5,9	Codon 36–71, 145–205, 374–386
SMO	Exon 3,5,6,9,11	Codon 187–227, 308–331, 392–418, 512–542, 609–645
SRC	Exon 14	Codon 500–532
STK11	Exon 1,4,5,6,8	Codon 23–64, 156–181, 192–207, 254–285, 318–360
TP53	Exon 2,3,4,5,6,7,8,9,10,11	Codon 1–19, 26–30, 33–59, 68–260, 263–393
VHL	Exon 1–3	Codon 79–107, 115–149, 156–173

Table S3. Archer Fusionplex – Lung v1.0

Gene	Exon	Assay type	Description
ALK	Exon 2,4,6,10,16,17,18,19,20,21, 22,23,26	Fusion	5′
ALK	Exon 22,23, 25	Mutation	T1151–C1156, F1174, L1196–S1206, G1269
BRAF	Exon 2,7,8,9,10,11,12,15,16	Fusion	5′
BRAF	Exon 1,3,7,8,10,13	Fusion	3′
BRAF	Exon 15	Mutation	V600
EGFR	Exon 7,8,9,16,19,20	Fusion	5′
EGFR	Exon 1,24,25	Fusion	3′
EGFR	Exon 8	Exon 2–7 skipping	5′
EGFR	Exon 1	Exon 2–7 skipping	3′
EGFR	Exon 18,19,20,21	Mutation	E709–G719, E746–L760, V774–C797, L858–861
FGFR1	Exon 2,3,4,5,6,7,8,9,10,11,17	Fusion	5′
FGFR1	Exon 12,17	Fusion	3′
FGFR2	Exon 2,5,7,8,9,10	Fusion	5′
FGFR2	Exon 16,17	Fusion	3′
FGFR3	Exon 3,5,8,9,10	Fusion	5′
FGFR3	Exon 16,17,18	Fusion	3′
KRAS	Exon 2,3	Mutation	G12–G13, Q61
MET	Exon 2,4,5,6,13,14,15,16,17,21	Fusion	5′
MET	Exon 2,13	Fusion	3′
MET	Exon 15	Exon 14 skipping	5′
MET	Exon 13	Exon 14 skipping	3′
NRG1	Exon 1,8	Fusion	5′
NRG1	Exon 1,2,3,4,6	Fusion	5′
NRG1	Exon 1	Fusion	3′
NTRK1	Exon 2,4,6,8,10,11,12,13	Fusion	5′
NTRK2	Exon 5,7,9,11,12,13,14,15,16,17	Fusion	5′
NTRK3	Exon 4,7,10,12,13,14,15,16	Fusion	5′
NTRK3	Exon 15	Fusion	5′
NTRK3	Exon 13,14,15	Fusion	3′
RET	Exon 2,4,6,8,9,10,11,12,13,14	Fusion	5′
RET	Exon 15,16	Mutation	A883, M918
ROS1	Exon 2,4,7,31,32,33,34,35,36,37	Fusion	5′
ROS1	Exon 38	Mutation	G2032

Table S4. Archer Fusionplex – (Expanded) Sarcoma v2.0

Gene	Exon	Assay type	Description
ALK	2,4,6,10,16,17,18,19, intron19, 20, mid-exon 20,21,22,23,26	Fusion	5′
ALK	23	Mutation	p.T1151–p.C1156, p.F1174, p.L1196–p. S1206, p.G1269
BCOR	2, intron2,3, mid-exon4,5,7,mid-exon7,10,12,14,15	Fusion, Internal Tandem, Duplication	3′
BCOR	3,4,mid-exon4,5,6,7,8,9,11,15	Fusion, Internal Tandem, Duplication	5′
BCOR	8	Fusion	5′
BRAF	1,2,3,7,8,10,13,18	Fusion, Kinase Domain Duplication, BRAFΔ2–10, BRAFΔ4–10, BRAFΔ2–8, BRAFΔ3–8, BRAFΔ4–8	3′
BRAF	2,3,4,5,7,8,9,10,11,12,15,16	Fusion, Kinase Domain Duplication, BRAFΔ2–10, BRAFΔ4–10, BRAFΔ2–8, BRAFΔ3–8, BRAFΔ4–8	5′
BRAF	15	Mutation	p.V600
CAMTA1	3	Fusion	3′
CAMTA1	8,9,mid-exon9,10	Fusion	5′
CIC	12,14,15,16,17,18,mid-exon19,19,mid-exon20,3′UTR	Fusion	3′
CIC	12	Fusion	5′
CSF1	5,6,7,8,3′UTR	Fusion	3′
CSF1	2,3,4,5,6	Fusion	3′
EGFR	1,24,25	Fusion, Exon 2–7 Skipping (EGFRvIII), Kinase Domain Duplication	3′
EGFR	7,8,9,16,17,18,19,20	Fusion, Exon 2–7 Skipping (EGFRvIII), Kinase Domain Duplication	5′
EGFR	18,19,20,21	Mutation	pE709–p.G719,p.E746–p. L760,p.V774–p.C797, p.L858–p.L861
EPC1	9,10,11	Fusion	3′
ERG	2,3,4,5,6,7,8,9,10,11	Fusion	5′
ESR1	6,7	Fusion	3′
ESR1	7,8	Fusion	5′
ESR1	3,4,5,6,7	Fusion	3′
ESR1	7,8	Fusion	5′
EWSR1	4,5,6,7,8,9,10,11,12,13,14	Fusion	3′
FGFR1	12,17	Fusion, Kinase Duplication	3′
FGFR1	2,3,4,5,6,7,8,9,10,11,17	Fusion, Kinase Domain Duplication	5′
FGFR1	4,14,12	Mutation	p.T141,p.V561,p.K656
FGFR2	16,17,18	Fusion	3′

Table S4. Continued

Gene	Exon	Assay type	Description
FGFR2	5'UTR,3,5,6,7,8,9,10	Fusion	5'
FGFR2	7,9,12,13,14	Mutation	p.R248–p.S249,p.G370–p.R399,p.V555,p.D641–p.K650,p.G697–p.K715
FGFR3	6,17,intron17,mid-exon18	Fusion	3'
FGFR3	3,5,8,10,12,13,14	Fusion	5'
FGFR3	7,9,13,14	Mutation	p.R248–p.S249,p.G370–p.R399, p.V555, p.D641–p.K650, p.G697–p.K715
FOS	Mid-exon4	Fusion	3'
FOSB	5'UTR,1,2	Fusion	5'
FOXO1	1,2,3	Fusion	3'
FOXO1	1,2,3	Fusion	5'
FUS	3,4,5,mid-exon6,6,7,8,9,10,11,13,14	Fusion	3'
GLI1	4,5,6,7	Fusion	3'
GLI1	4,5,6,7	Fusion	5'
HMGA2	1,2,3,4,5,3'UTR	Fusion	3'
JAZF1	2,3,4	Fusion	3'
MDM2	2,4,6,8,10	Expression, Fusion	3'
MDM2	5,9	Expression, Fusion	5'
MEAF6	4,5	Fusion	3'
MET	2,13	Exon 14 Skipping	3'
MET	2,4,5,6,13,14,15,16,17,21	Exon 14 Skipping	5'
MGEA5	4,5,6,7,8,9,12,13,14,15	Fusion, Expression	5'
MKL2	11,12,13	Fusion	5'
MYOD1	1	Mutation	L122R
NCOA1	11,12,13,14,15	Fusion	5'
NCOA2	11,12,13,14,intron14,15,16,22	Fusion	5'
NCOA2	14	Fusion	3'
NR4A3	2,3,4,5,7,9	Expression, Fusion	5'
NR4A3	8	Expression, Fusion	3'
NTRK1	1,2,3,4,5,6,7,8,9,10,11,12,13,14	Fusion	5'
NTRK1	1,2	Fusion	5'
NTRK1	13,14,15,16,17	Mutation	Full Kinase Domain coverage for resistance mutations including p.G595
NTRK2	4,5,6,7,8,9,10,11,12,13,14,15,16,17,18	Fusion	5'
NTRK2	11,14	Fusion	3'
NTRK2	16,17,18,19,20,21	Mutation	Full Kinase Domain coverage for resistance mutations
NTRK3	15	Fusion	5'
NTRK3	13,14,15,17	Fusion	3'

Table S4. Continued

Gene	Exon	Assay type	Description
NTRK3	3,4,5,6,7,8,9,10,12,13,14,15,16	Fusion	5′
NTRK3	15,16,17,18,19	Mutation	Full Kinase Domain coverage for resistance mutation detection including p.F617, p.G623, p.G696
NUTM1	5′UTR,mid-exon3,4,5,mid-exon6,6	Fusion	5′
PAX3	3,5,6,7,8	Fusion, Expression	3′
PAX3	2,4,8	Fusion, Expression	5′
PDGFB	2,3	Fusion	5′
PHF1	1,2	Fusion	5′
PHF1	10,11,12	Fusion	3′
PLAG1	1,2,3,4	Fusion	5′
PRKCA	4,5,6,9	Fusion	5′
PRKCB	1,3,8	Fusion	5′
RAF1	4,5,6,7,9	Fusion	3′
RAF1	4,5,6,7,8,9,10,11,12	Fusion	5′
RET	2,4,6	Fusion	5′
RET	8,9,10,11,mid-exon11,12,13,14	Fusion	5′
RET	15,16	Mutation	p.A883,p.M918
ROS1	2,4,7,31,32,33,34,35,36,37	Fusion	5′
ROS1	38	Mutation	p.G2032
SS18	4,5,6,8,9,10	Fusion	3′
SS18	2,3,4,6,10,11	Fusion	5′
STAT6	1,2,3,4,5,6,7,15,16,17,18,19,20	Fusion	5′
TAF15	5,6,7,9	Fusion	3′
TAF15	6,7	Fusion	5′
TCF12	4,5,6	Fusion	3′
TFE3	2,3,4,5,6	Fusion	3′
TFE3	2,3,4,5,6,7,8	Fusion	5′
TFG	3,4,5,6,7,mid-exon8	Fusion	3′
TFG	6	Fusion	5′
USP6	1,mid-exon1,2,3	Fusion	5′
VGLL2	1,2,3,intron3,4	Fusion	3′
YAP1	1,2,3,4,5,6,7	Fusion	3′
YAP1	1,mid exon1,2,3,4,8,9	Fusion	5′
YWHAE	5	Fusion	3′

Table S5. Information per WGS attempt. (Provided separately as an Excel file)

Table S6. Additional variants detected by WGS. (Provided separately as an Excel file)

Table S7. Clinical validity. (Provided separately as an Excel file)

Table S8. Pathogenic germline variants detected by WGS. (Provided separately as an Excel file)

Table S9. Correlation signatures in BRCA and Lynch patients

Lynch patients (n=7)			
		MSI score	
		MSI high	MSS
Biallelic loss of MLH1, MSH2, MSH6 or PMS2 on genetic level	Present	4	0
	Absent	0	3
BRCA germline patients (n=16)			
		CHORD score	
		HRD	HRP
Biallelic BRCA loss on genetic level	Present	6	0
	Absent	1*	8

* One patient received previous treatment with PARP inhibitors, and as a result, the mutated *BRCA2* allele was lost as a resistance mechanism to PARP inhibitors. The genome-wide mutational damage initially caused by homologous repair deficiency was still present, resulting in a high CHORD score.

Reference numbers refer to the main text list

Data supplement Chapter 5

Complete genomic characterization in patients with cancer of unknown primary in routine diagnostics

Supplementary material

Supplementary methods: CUPPA classifier

Determination of cancer type cohorts

Cohorts for training the algorithm were constructed from the HMF database by selecting the highest purity sample from each unique patient from our database with qcStatus = 'PASS'. 36 tumor categories were defined based on the clinical annotations of primaryTumorLocation, primaryTumorSubLocation, primaryTumorType and primaryTumorSubType (table S1). Certain cancers such as Esophagus and Stomach were combined for the categorisation as we found empirically that the CUPPA classifiers had little ability to distinguish between them. For other cancers including Lung, Bone/Soft tissue, Skin, Uterus & Pancreatic cancers we have broken into subtypes where histological information allows. All cancers not in one of these 36 cohorts was deemed as "Other" and were excluded from the reference cohorts for analysis. Samples with 'unknown' tumor type are also excluded. Finally, 45 samples were also explicitly excluded from the reference cohort where our analysis strongly suggested the clinical configured cancer type may be incorrect for these samples.

DNA classifier logic

CUPPA includes 3 orthogonal DNA classifiers based on positional mutational distribution, SNV mutational profile and feature prevalence, and a 4th classifier which combines the 3 together to make an overall prediction. Each classifier assigns a likelihood to each cancer type with the sum of the likelihoods adding up to 1 across the cancer types.

1. GENOMIC_POSITION_SIMILARITY

This classifier solely relies on the mutational distribution of tumors of genomic position, which has been shown previously to have strong predictive power for tissue of origin. CUPPA calculates a consensus mutation distribution for each cohort by counting SNV TMB by bucketed genomic position across each cohort. High TMB samples are downsampled to 20k mutations in this consensus so that individual samples cannot dominate a cohort. CUPPA counts mutations using a window size of

500kb bases (chosen after testing various sizes from 100kb to 10Mb). The genomic position similarity likelihood for a given sample is determined by first calculating the cosine similarity (CSS) of a sample to each cohort consensus distribution and then weighing using the following algorithm:

Score(sample=s,cancerType=i) = 8^[100(CSS(i,s)-BestCSS(s))]*

CUPPA sums the scores across each tumor type to estimate a likelihood for each cancer type:

Likelihood(tumorType=i) = Score(i) / SUM(all tumors) [Score]

2. SNV_96_PAIRWISE_SIMILARITY

This classifier relies solely on relative SNV counts via the 96 trinucleotide buckets frequently used for cosmic signatures. The cosmic signatures are not used directly, but the classifier is designed to capture the obvious similarities that can also be observed via signatures capturing known cancer specific mutagenic effects such as UV & Smoking and also background signatures per cancer type. Unlike the genomic position similarity which determines a consensus view of mutational distribution, the SNV_96_PAIRWISE classifier does not create a consensus view per tumor type as tumor types may have a diverse range of mutational profiles. Instead the classifier calculates a pairwise cosine similarity between the sample in question and every other sample in the Hartwig cohort. Once a pairwise CSS has been determined, a score is calculated for each pair using the following formula:

Score(i,j) =
8^[-100(1-CSS)]^[maxCSS^8]*mutationCountWeightFactor**
cohortSizeWeightFactor

Where:

- MaxCSS is the maximum pairwise CSS for any sample in the cohort. This factor reduces confidences in general for samples that have no close pairwise match.
- mutationCountWeightFactor penalises pairs with large differences in SNV TMB. This is implemented as: *mutationCountWeightFactor = min(SNV_TMB(i)/SNV_TMB(j),SNV_TMB(j)/SNV_TMB(i))*
- cohortSizeWeightFactor penalises larger cohorts which will have more similar tumors just by chance (eg. Breast cohort =~ 750 samples vs Thyroid cohort =~ 20 samples), implemented as: *cohortSizeWeightFactor = sqrt(# of samples of*

tumor type) / SUM(i)[sqrt(# of samples of tumor type i)]

As for genomic position similarity, CUPPA sums the scores across each tumor type to estimate the likelihood:

Likelihood(tumorType=i) = SUM(tumorType=i)[Score] / SUM(all tumors) [Score]

3. FEATURE

The FEATURE classifier uses observed prevalence of both cancer type specific drivers as well as certain passenger mutational features that may be significantly enriched or depleted in certain types to predict the cancer type of a sample.

Driver Prevalence

Driver (or driver like) features used include all driver point mutation, high amplification, homozygous deletion and homozygous disruptions in the driver catalog as well as viral insertions & fusions. For fusions, known pathogenic fusion pairs, IG rearrangement pairs and exon deletions/duplications configured in the HMF fusion knowledge base are all considered as features as are fusions with highly promiscuous exons such as ALK exon 20-21. For Sarcomas specifically, we override the prevalence for a list of 56 pathognomonic fusions which are highly diagnostic but may not be prevalent enough to be present in our database to the appropriate cancer type with the maximal allowed feature weight.

Indels in repeat contexts of 6 or less bases in 3 lineage defining genes: ALB (highly specific to Liver cancer) and SFTPB & SLC34A2 (highly specific to Lung cancer) are also treated as additional features (note though that they are ignored for MSI samples). A set of Lung cancer specific EGFR hotspots (including T790M, L858R and exon 19 and 20 inframe deletions) are also treated as a single feature.

Features are weighted by driver likelihood. For point mutations the driver likelihood (the dnds calculated probability between 0 and 1 that the mutation is a driver) is used to weight the mutations, whilst other mutations, virus insertions and fusions are assumed to have probability of 1.

The prevalence of each feature in each cancer type is calculated

Prevalence = minPrevalence + sum (driverLikelihood) / COUNT(samples)]

Where minPrevalence is a fixed notional background rate of observing a passenger set to 0.15 / count of cancer types for drivers or indels in lineage defining genes and 0.01 / count of cancer types for fusions and viral insertions which are rarely passengers.

A combined driver score for each cancer type is calculated by taking the product of the observed prevalence of each of the drivers from the sample in the cancer type cohort, discounted by the driver likelihood in the cancer itself. ie:

DriverScore = weightFactor(cohort) PRODUCT*
[Prevalence(d)^driverLikelihood(d,s)]

Where the *weight factor = meanDriverLoad(pan-cancer) / meanDriverLikelihood(cohort)* and is intended to reduce the tendency for cancer types with higher average rates of drivers such as Urinary Tract and Esophagus to have higher driver scores

Passenger Prevalence

In addition to drivers, mutational burdens of certain types of events can vary widely across different cancer types. For example LINE insertions are universally observed in Esophagus and certain other cancers but almost non-existent in other cancers. Depending on the feature it may be useful to test that the rate observed is either higher or lower than what is expected of the cancer type.

Since different cancers may have different characteristic frequencies, this is modeled for this classifier as a prevalence with a dynamic cutoff based on the rate observed in the sample itself. Specifically if testing for an enriched rate, the cutoff is set to 25% below the observed rate limited to a maximum value of the highest observed 95th percentile rate of any cancer cohort. Conversely if testing for a depleted rate, the cutoff is set to 25% below the observed rate limited to a maximum value of the highest observed 95th percentile rate of any cancer cohort.

The following features are tested for enrichment and/or depletion:

Feature	Enrichment	Depletion
SNV_TMB	True	True
MS_INDEL_TMB	True	True
LINE_COUNT	True	True
TELOMERIC_SG_BE_COUNT	True	NA
MAX_COMPLEX_SIZE	True	NA
SIMPLE_DEP_32B_200B	True	NA

As for drivers the prevalence in each cancer type is added to a minPrevalence set to 0.15 / count of cancer types. The passenger score is simply the product of all the passenger prevalence rates

PassengerScore = PRODUCT [max(Passenger Prevalence,minPrevalence)]

Combining scores to a likelihood

The passenger and driver scores are multiplied together to get a single score:

*Score = PassengerScore * Driver Score*

And finally CUPPA sums the scores across each tumor type to estimate the likelihood:

Likelihood(tumorType=i) = Score(i)^correlationDampenFactor / SUM(all tumors) [Score^correlationDampenFactor]

The correlationDampenFactor is introduced to reduce the confidence of the classifier and set at 0.8 to empirically match the observed accuracy. This is required as some of the driver or passenger features may be correlated with each other - for example same arm amplifications are highly correlated and TMB might be positively correlated with more drivers in general.

4. DNA_COMBINED CLASSIFIER

A combined score is calculated by multiplying the 3 likelihoods together with an absolute floor set at 1% per likelihood. The likelihood is then calculated as

Likelihood(tumorType=i) =
PRODUCT(max(0.01,Classifier(i,j)))^correlationDampenFactor /
SUM(all tumors)[
PRODUCT(max(0.01,Classifier(j)))]^correlationDampenFactor

As for the feature classifier, a correlationDampenFactor is introduced to reduce the confidence of the classifier and reflect the fact that the individual classifiers are not completely independent. A value of 0.65 is chosen to empirically match the confidence to the observed accuracy. For the DNA_COMBINED classifier, males are excluded from matching 'Ovary' and 'Uterus' cancer cohorts and females are excluded from matching the 'Prostate' cohort. 'Breast' cancer scores for male cancer cohorts are penalised but not excluded.

5. NEAREST NEIGHBOR ANALYSIS

In addition to the classifiers, the 20 nearest neighbour samples by pairwise cosine similarity are reported for 3 different features:

- Count of SNV TMB per 500k genomic position buckets
- Count of SNV TMB by 96 mutational context bucket
- Log(TPM+1) RNA expression by gene

Note that all samples are used for this analysis including rare cancer types that are not one of the CUPPA categorizations used in the classifiers.

Table S1. Reference cohorts

Reference cohort	Weight female	Weight male	Includes
Anogenital	1.0	1.0	Penis, Vulva, Vagina, Anus (<exl. Melanoma>), Uterus:Cervix
Bile duct/Gallbladder	1.0	1.0	Bile duct; Hepatobiliary system; Gallbladder
Bone/Soft tissue: Other	1.0	1.0	Bone/Soft tissue (<other or unspecified>)
Breast	1.0	0.1	Breast
Colorectum/Appendix/ Small intestine	1.0	1.0	Colorectum (<other or unspecified>); Appendix; Small intestine(<other or unspecified>)
Esophagus/Stomach	1.0	1.0	Esophagus (<excluding Nueroendocrine tumor); Stomach (<excluding Nueroendocrine tumor); Gastroesophageal
GIST	1.0	1.0	Bone/Soft tissue (Gastrointestinal stromal tumor)
Head and neck	1.0	1.0	Head and neck (<other>)
Kidney	1.0	1.0	Kidney
Leiomyosarcoma	1.0	1.0	Bone/Soft tissue (Leiomyosarcoma)
Liposarcoma	1.0	1.0	Bone/Soft tissue (Liposarcoma)
Liver	1.0	1.0	Liver (<excluding Neuroendocrine tumor>)
Lung: NET	1.0	1.0	Lung(Neuroendocrine tumor)
Lung: Non-Small Cell	1.0	1.0	Lung(Carcinoma:Non-small cell carcinoma); Lung(Carcinoma:Adenocarcinoma); Lung(<other>)
Lung: Small Cell	1.0	1.0	Lung(Carcinoma:Small cell carcinoma); Lung(Carcinoma:Small cell carcinoma combined type)
Lymphoid tissue	1.0	1.0	Lymphoid tissue
Mesothelium	1.0	1.0	Mesothelium
Ovary/Fallopian tube	1.0	0.0	Ovary; Fallopian tube
Pancreas	1.0	1.0	Pancreas (<other or unspecified>)
Pancreas: NET	1.0	1.0	Pancreas (Neuroendocrine tumor)
Prostate	0.0	1.0	Prostate
Salivary gland/Adenoid cystic	1.0	1.0	Neck:Parotid gland, Head and Neck:Sublingual gland, <any>(Carcinoma:Adenoid cystic carcinoma), Trachea
Melanoma	1.0	1.0	<Any, excluding Eye> (Melanoma)
Skin: other	1.0	1.0	Skin (<other or unspecified>)
Small intestine/ Colorectum: NET	1.0	1.0	Small intestine(Neuroendocrine tumor); Colorectum(Neuroendocrine tumor)
Thyroid gland	1.0	1.0	Thyroid gland
Urothelial tract	1.0	1.0	Urothelial tract
Uterus: Endometrium	1.0	0.0	Uterus: endometrium

Table S1. Continued

Reference cohort	Weight female	Weight male	Includes
Other	1.0	1.0	Gastrointestinal tract, Eye, Bone marrow, Nervous system(<other), Adrenal Gland, Thymus, Testis, Esophagus (Neuroendocrine tumor),Stomach (Neuroendocrine tumor)

Reference cohorts were determined based on the prevalence of samples in the reference set. In case of a limited number of distinct samples of certain tumor origin, different primary tumors were grouped into a single reference cohort based on clinicopathological similarities. Males are excluded from matching 'Ovary' and 'Uterus' cancer cohorts and females are excluded from matching the 'Prostate' cohort. 'Breast' cancer scores for male cancer cohorts are penalised but not excluded.

Table S2. CUPPA predictions for rare cancer samples without matching reference cohort

Sample ID	True tumor type	Predicted tumor type	Similarity likelihood score
HMF002649A	Adrenal cortical carcinoma	Breast	0.202
HMF004708A	Thymoma	Head and neck: other	0.645
HMF005182A	Germ cell tumor	Ovary/Fallopian tube	0.403
HMF005394A	Esthesioneuroblastoma	Skin: other	0.191
HMF005489A	Thymoma	Liposarcoma	0.362
HMF005521A	Urachal carcinoma	Bile duct/Gallbladder	0.396

Table S3. CUPPA prediction pitfalls

Bias	Relevant for	Classifier	Description	Considerations for clinical application in low-confidence predictions
AID_APOBEC	>2%	SNV_96/ GENOMIC_ POSITION	Signature shared across 5-6 cohorts, but strongest in Urothelial/Breast. The genomic position signature for AID_ APOBEC seems to be very different. Lung and Eso/Stomach samples in particular get low GEN_POSITION. Other cancer types such as Anogenital & Head & neck perform ok on GEN_POS, but poorly on other classifiers.	Samples with a high mutational load based on the APOBEC mutational signature tend to match Urothelial/Breast reference cohorts
Small cohort size	2%	ALL DNA	Rounding issues and noise dominate all classifiers where cohort size is small (<25 samples), prevents us from small cohorts such as Testis, and diminishes performance even > 25 samples. Also true for pairwise classifiers even though we adjust for it.	Rare tumor types are challenging to classify correctly with the algorithm
TMBPerMb < .7	0.5%	ALL DNA	Generally low confidence. Often mismatch to Pancreas:NET, likely due to 'Low TMBPerMB' feature	Samples with a low number of genetic variants often mismatch
High driver load	1%	FEATURE	Samples with a high number of drivers tend to match Urothelial Tract cancers (these have the highest rate of drivers)	Samples with a high number of drivers are challenging to classify and might be misclassified as urothelial cell cancers
MSI	0.3%	GENOMIC_POSITION	Samples with MSI typically have very low GENOMIC_POSITION scores to the correct cancer type. Similar to AID_APOBEC effect	The algorithm is less suited for samples with a MSI profile
Pathognomonic events	1%	FEATURE	Rare pathognomonic events (eg. Sarcoma fusions) may not be found previously in our cohort or may not be weighed highly enough due to 'min_prevalence'. For some drivers the mechanism may be diagnostic whereas we only calculate features at a gene level, eg: • EGFR hotspots and ecdna (glioma) vs bfb (lung) • SPOP amp (breast) vs mutation (prostate) • KIT amp (lung) vs mutation (sarcoma) • FOXA1 amp (lung) vs mutation (breast/prostate) • KRAS amp (esophagus) vs mutation (CRC/pancreas) • Hypermutations in BCL2 & other genes (Lymphoid)	A manual check should be performed for pathognomonic events, as they might not always be incorporated within the algorithm. Pathognomonic events are separately reported as somatic variants.

Table S3. Continued

Bias	Relevant for	Classifier	Description	Considerations for clinical application in low-confidence predictions
Treatment signatures	0.1%	SNV_96	Samples with strong treatment signatures (eg SYD985) will match each other with high certainty	Previous therapies should be taken into account
Lung: Small cell vs. non-small cell	0.5%	GENOMIC_POSITION	'Lung: non-small cell' can strongly match the genomic position profile of Lung: small cell with high confidence. Possibly due to timing of transformation to small cell?	Low-confidence predictions should be matched with morphology
Bile Duct / Gallbladder	1%	ALL	Can be mistaken for Liver or Pancreas with high confidence	Hepatopancreaticobiliary cancers are challenging to classify. The reference database might contain misclassified tumors due to the difficult distinction between these tumors. Predictions should be thoroughly correlated with imaging and morphological results
Non-smoking lung cancer	0.2%	GENOMIC_POSITION	Performance is weaker, but can mostly be explained by AID_APOBEC / pathognomonic events	NSCLC with a known driver event (e.g. EGFR mutations, ALK mutations, etc.) generally receive a lower prediction confidence. Presence of these driver events is indicative of a NSCLC in these samples.
Anogenital vs Head & Neck: Other	0.4%	ALL	Can often be mistaken for each other.	Low-confidence predictions in squamous cell carcinomas should be interpreted with caution
Sarcoma	1.5%	ALL	Frequent mismatches between Leiomyosarcoma, Liposarcoma, Osteosarcoma and 'other'. Multiple causes: • Pathognomonic fusions not taken into account • Larger cohorts would help make clearer cohorts and could allow distinct groups for Rhabdomyosarcoma and others. • Some samples marked as 'Sarcoma' and matched to Leiomyosarcoma are reported as match=F, but may be TP • Spindle cell sarcoma appear to group better with Leiomyosarcoma but are marked as 'other'	Classification of sarcoma subtypes should be performed with caution

Several pitfalls of the current algorithm were identified. These pitfalls generally lead to a low-confidence prediction, and the results should be interpreted with caution. Identification of these pitfalls allow for further optimization of the algorithm and can guide future strategies for improvement.

Table S4. Per patient overview of immunohistochemical stains.

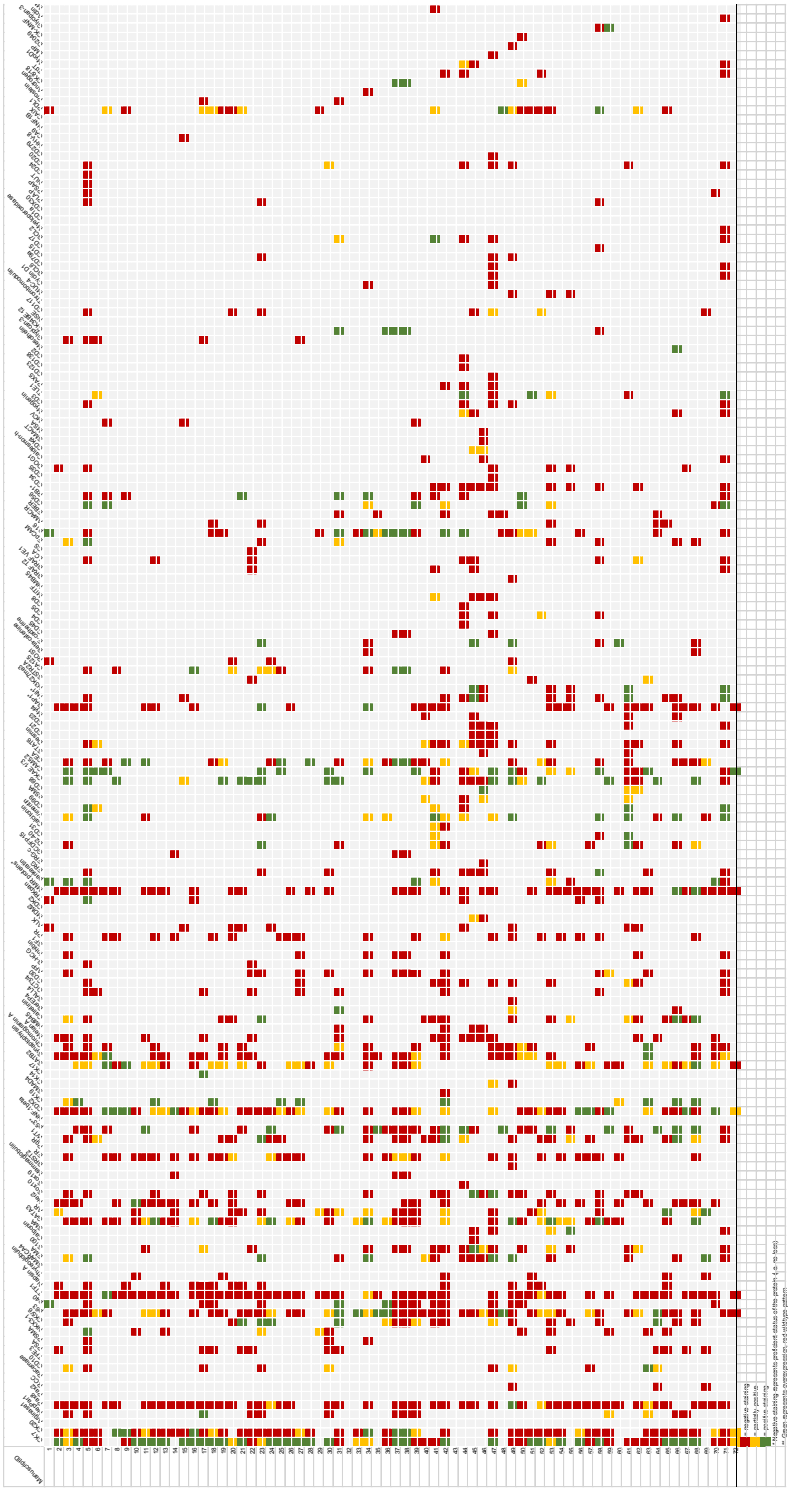


Table S5.Description of informative low-confidence predictions. Case numbers correspond with cases in figure 3.

Case	Clinical information	Top-5 CUPPA predictions (similarity likelihood scores)	Diagnosis after WGS
38	Gender: female Age: 51 Oncological history: basal cell carcinoma (2020) Metastatic pattern: inguinal, abdominal, and para-aortal/-iliacal lymph nodes Gastro-/colonoscopy: colonoscopy negative Morphology: poorly differentiated carcinoma Other: cervical cytology negative; anal biopsies negative Prior to WGS: metastasis of poorly differentiated carcinoma with a non-specific immunohistochemical profile	1. Anogenital (0,698) 2. Breast (0,156) 3. Head and neck (0,0584) 4. Urothelial tract (0,04) 5. Uterus (0,01) NB. HPV type 16 detected with WGS	Endocervical carcinoma Immunoprofile (CEA and CK7 positive) Presence of HPV CUPPA prediction Metastatic pattern
39	Gender: female Age: 76 Oncological history: DCIS (1995, 2005) Metastatic pattern: liver, bone Gastro-/coloscopy: gastroscopy and colonoscopy negative Morphology: poorly differentiated carcinoma Prior to WGS: The immunohistochemical profile was non-specific in terms of primary origin. Differential diagnosis of proximal digestive tract (esophagus, stomach, pancreas, biliary tract/gall bladder and small intestine), lung or breast cancer	1. Bile duct/Gallbladder (0,352) 2. Esophagus/Stomach (0,303) 3. Lung: Non-small Cell (0,0791) 4. Breast (0,0674) 5. Colorectum (0,063)	Gall bladder carcinoma Negative colonoscopy and gastroscopy Abdominal mass in/around liver, no tumor in breast or lung Pancreas not in top-5 predictions
40	Gender: male Age: 61 Oncological history: prostate cancer (2019), non-small cell lung cancer (2020) Metastatic pattern: lung, brain, inguinal lymph nodes Gastro-/coloscopy: not performed Morphology: sarcomatoid carcinoma Prior to WGS: Inguinal biopsy not fitting profile of previously diagnosed prostate cancer and non-small cell lung cancer due to sarcomatoid differentiation. Doubt about initial NSCLC diagnosis (sarcoma?)	1. Bone/Soft tissue (0,499) 2. Prostate (0,096) 3. Skin: Other (0,0641) 4. Pancreas (0,0505) 5. Kidney (0,0492)	Sarcomatoid carcinoma of the prostate CUPPA prediction Metastasis of NSCLC excluded Double tumor: NSCLC (lung and probably brain lesions), and sarcomatoid differentiation of prostate cancer (inguinal lymph nodes)

Table S5. Continued

Case	Clinical information	Top-5 CUPPA predictions (similarity likelihood scores)	Diagnosis after WGS
41	Gender: male Age: 80 Oncological history: - Metastatic pattern: lung Gastro-/coloscopy: not performed Morphology: undifferentiated Prior to WGS: Large pleural lesion that could fit lung cancer, mesothelioma or sarcoma	1. Bone/Soft tissue (0,471) 2. GIST (0,0977) 3. Kidney (0,0842) 4. Leiomyosarcoma (0,0524) 5. Prostate (0,0492)	Pleural sarcoma NOS Lung cancer and mesothelioma not in top-5 predictions Further differentiation in sarcoma subtype not possible
42	Gender: female Age: 68 Oncological history: - Metastatic pattern: lung, bone, abdominal and lymph nodes (axillary, abdominal, neck, supra/infraclavicular) Gastro-/coloscopy: not performed Morphology: undifferentiated Other: mammography negative Prior to WGS: Large pleural lesion that could fit lung cancer, mesothelioma or sarcoma	1. Ovary/Fallopian tube (0,668) 2. Breast (0,132) 3. Urothelial tract (0,0525) 4. Uterus: Endometrium (0,0364) 5. Lung: Non-small Cell (0,0249) NB. BRCA2 germline mutation detected	Ovarian cancer CUPPA prediction Negative mammography BRCA2 germline variant
43	Gender: male Age: 72 Oncological history: - Metastatic pattern: lymph nodes (neck) Gastro-/coloscopy: not performed Morphology: squamous cell carcinoma Other: laryngoscopy negative Prior to WGS: Squamous cell carcinoma without lesion on dermatological assessment. Differential diagnosis: upper respiratory tract, parotis, skin	1. Skin: Other (0,711) 2. Urothelial tract (0,0777) 3. Head and neck: other (0,0687) 4. Melanoma (0,0361) 5. Anogenital (0,0166)	Cutaneous squamous cell carcinoma CUPPA prediction High UV signature, indicating sun exposure

Table S5. Continued

Case	Clinical information	Top-5 CUPPA predictions (similarity likelihood scores)	Diagnosis after WGS
44	Gender: female Age: 73 Oncological history: uterine carcinosarcoma (2020) Metastatic pattern: lymph nodes (mediastinal/hiliary, neck, supra/infraclavicular) Gastro-/coloscopy: not performed Morphology: undifferentiated Other: laryngoscopy negative Prior to WGS: Undifferentiated tumor with uncertain distinction between metastasis of known carcinosarcoma, or second primary sarcoma (DSCRT or rhabdomyosarcoma were considered)	1. Ovary/Fallopian tube (0,384) 2. Lung: Non-small Cell (0,232) 3. Bone/Soft tissue: Other (0,112) 4. Anogenital (0,0682) 5. Breast (0,0399)	Carcinosarcoma of the uterus Medical history CUPPA prediction with ovary and sarcoma in top 5 predictions
45	Gender: female Age: 69 Oncological history: - Metastatic pattern: thoracic wall Gastro-/coloscopy: not performed Morphology: epithelioid Other: - Prior to WGS: Epithelioid tumor with uncertain distinction between epithelioid sarcoma or melanoma based on morphology and immunohistochemistry	1. Bone/Soft tissue: Other (0,554) 2. Ovary/Fallopian tube (0,0885) 3. Breast (0,0445) 4. Uterus: Endometrium (0,0349) 5. Thyroid gland (0,0334)	Epithelioid MPNST Top prediction: sarcoma Melanoma not in top-5 predictions
46	Gender: female Age: 77 Oncological history: melanoma (2003) Metastatic pattern: lung, axillary lymph nodes Gastro-/coloscopy: not performed Morphology: undifferentiated Prior to WGS: Undifferentiated malignant ulcerating tumour. Based on the morphology and immunohistochemical profile, an undifferentiated sarcoma or histiocytic sarcoma were considered. However, given the clinical presentation (an ulcerating skin tumor and lymph node metastases) and prior history of melanoma, a melanoma with marker loss could not be excluded	1. Leiomyosarcoma (0,641) 2. Bone/Soft tissue: Other (0,284) 3. Lung: Non-small Cell (0,0196) 4. Breast (0,00888) 5. Ovary/Fallopian tube (0,00684)	Sarcoma NOS Two sarcoma subtypes as highest predictions (combined score 0.925) Melanoma not in top-5 predictions

*

Table S5. Continued

Case	Clinical information	Top-5 CUPPA predictions (similarity likelihood scores)	Diagnosis after WGS
47	Gender: female Age: 72 Oncological history: - Metastatic pattern: axillary lymph nodes Gastro-/coloscopy: not performed Morphology: poorly differentiated adenocarcinoma Other: tumor marker S100 slightly elevated Prior to WGS: Tumor localization indicative of breast cancer. However, immunohistochemical profile was aspecific, and recurrent radiological assessments (ultrasound, mammography, PET-CT scan) failed to identify a primary (breast) tumor.	1. Urothelial tract (0,517) 2. Breast (0,364) 3. Uterus: Endometrium (0,0246) 4. Ovary/Fallopian tube (0,0212) 5. Anogenital (0,0151)	Triple negative breast cancer Urothelial tract carcinoma excluded based on immunohistochemical profile and clinical presentation Second prediction fits clinical differential diagnosis NB. Six months after WGS, a primary lesion in the breast was detected
48	Gender: male Age: 70 Oncological history: - Metastatic pattern: lymph nodes (neck) Gastro-/coloscopy: not performed Morphology: squamous cell carcinoma Other: laryngoscopy negative Prior to WGS: Squamous cell carcinoma of unknown origin with a differential diagnosis of head and neck, lung, esophagus and skin cancer.	1. Lung: Non-small Cell (0,645) 2. Esophagus/Stomach (0,108) 3. Lung: Small Cell (0,0824) 4. Head and neck: other (0,0649) 5. Urothelial tract (0,0269)	Lung cancer High probability score for lung cancer Esophageal cancer was excluded based on imaging, clinical presentation and a high smoking signature NB. A pulmonary lesion was detected four months after WGS analysis
49	Gender: female Age: 44 Oncological history: - Metastatic pattern: lung, adrenal gland, lymph nodes (abdominal, retroperitoneal, neck) Gastro-/coloscopy: not performed Morphology: undifferentiated Other: Prior to WGS: based on the clinical presentation, this patient was initially suspected for a lung carcinoma. Based on the immunohistological profile, the differential diagnosis was further extended with thymus and kidney cancer.	1. Kidney (0,758) 2. Mesothelium (0,149) 3. Bile duct/Gallbladder (0,0205) 4. Breast (0,0147) 5. Pancreas: NET (0,00774)	Renal cell carcinoma Prediction with a relative high similarity likelihood score fitting differential diagnosis

Table S6: Actionable events in CUP patients

ID	CHORD score	MSI score	Mutational load	Actionable biomarker (gene)	Actionability	In case of multi-actionability
1	0	0.17	92	-	No actionability	
2	0	0.08	22	FGFR2 fusion	FGFR inhibitors	
3	0	0.1	38	NRASQ61R activating mutation	MEK inhibitors	
4	0	0.07	22	-	No actionability	
5	0	0.04	26	-	No actionability	
6	0	0.03	16	-	No actionability	
7	0	0.79	133	-	No actionability	
8	0	0.27	102	BRAFV600E mutation	BRAF inhibitors	
9	0	0.14	31	NRASG12D activating mutation	MEK inhibitors	
10	0	0.67	156	-	Checkpoint inhibitors	
11	0	0.18	78	-	No actionability	
12	0	0.8	160	-	Checkpoint inhibitors	
13	0	0.64	136	-	No actionability	
14	0.08	0.2	134	-	No actionability	
15	0.01	0.01	73	KRASG12C activating mutation	KRAS inhibitors	
16	0.14	0.59	708	-	Checkpoint inhibitors	
17	0.03	0.09	228	BRAFG466A activating mutation	Multi-actionability	Checkpoint inhibitors, BRAF inhibitors
18	0.06	0.09	140	-	Checkpoint inhibitors	
19	0	0.15	267	CHEK2C286* inactivating mutation	Multi-actionability	Checkpoint inhibitors, PARP inhibitors
20	0.01	0.29	307	KRASG12C activating mutation	Multi-actionability	Checkpoint inhibitors, KRAS inhibitors
21	0.01	0.47	338	-	Checkpoint inhibitors	
22	0	0.07	394	NRASQ61R activating mutation	Multi-actionability	Checkpoint inhibitors, MEK inhibitors
23	0	0.01	12	-	No actionability	
24	0.02	0.08	57	-	No actionability	
25	0	0.01	19	-	No actionability	
26	0	0.09	38	-	No actionability	
27	0	0.15	35	BRAFN486_P490del activating mutation	BRAF inhibitors	
28	0	0.02	21	-	No actionability	
29	0	0.03	40	-	No actionability	

Table S6. Continued

ID	CHORD score	MSI score	Mutational load	Actionable biomarker (gene)	Actionability	In case of multi-actionability
30	0	0.26	43	PIK3CAV344G activating mutation, MAP3K1 deletion	Multi-actionability	PI3K-AKT-mTOR inhibitor, MEK inhibitors
31	0	0.02	515	-	Checkpoint inhibitors	
32	NA	0	113	-	No actionability	
33	0	0.17	260	-	Checkpoint inhibitors	
34	0	0.13	275	RAF1 amplification	Multi-actionability	Checkpoint inhibitors, multikinase inhibitor
35	NA	65.44	722	-	Checkpoint inhibitors	
36	0	0.16	105	-	No actionability	
37	0	0.09	79	-	No actionability	
38	0	0.06	89	-	No actionability	
39	0	0.23	47	-	No actionability	
40	0.18	0.1	34	-	No actionability	
41	0.85	0.01	36	KIT amplification	Multi-actionability	Multikinase inhibitors, PARP inhibitor
42				-	No actionability	
43	0.97	0.18	216	BRCA2c.8754+5G>A inactivating mutation	Multi-actionability	PARP inhibitors, checkpoint inhibitors
44	0.03	0	180	-	Checkpoint inhibitors	
45	0	0.27	77	-	No actionability	
46	0.03	0.04	16	CHEK2c.1100del inactivating mutation	PARP inhibitors	
47	0	0.03	33	-	No actionability	
48	0.94	0.2	151	-	Multi-actionability	Checkpoint inhibitors, PARP inhibitors
49	NA	0	5	-	No actionability	
50	0.01	0.63	341	-	Checkpoint inhibitors	
51	0.02	0.01	10	-	No actionability	
52	0.04	0.05	52	-	No actionability	
53	0.01	0.03	87	-	No actionability	
54	0	0.12	60	-	No actionability	
55	0	0.01	11	-	No actionability	
56	0.02	0.13	51	-	No actionability	
57	0	0.06	33	-	No actionability	
58	0	0.1	21	-	No actionability	

Table S6. Continued

ID	CHORD score	MSI score	Mutational load	Actionable biomarker (gene)	Actionability	In case of multi-actionability
59	0	0.2	89	-	No actionability	
60	0.89	0.01	28	-	PARP inhibitors	
61	0.01	0.1	36	-	No actionability	
62	0	0.04	30	-	No actionability	
63	0.02	0.18	62	-	No actionability	
64	0	0.26	172	-	Checkpoint inhibitors	
65	0	0.09	23	FGFR3Y375C activating mutation	FGFR inhibitors	
66	0	2	459	-	Checkpoint inhibitors	
67	0	0.21	49	ATMI2878R inactivating mutation	PARP inhibitors	
68	NA	11.17	288	-	Checkpoint inhibitors	
69	1	0.25	192	-	Multi-actionability	Checkpoint inhibitors, PARP inhibitors
70	0	0.73	92	EGFR amplification	EGFR mAb inhibitors	
71	0.21	0.04	26	-	No actionability	
72	NA	0	6	-	No actionability	

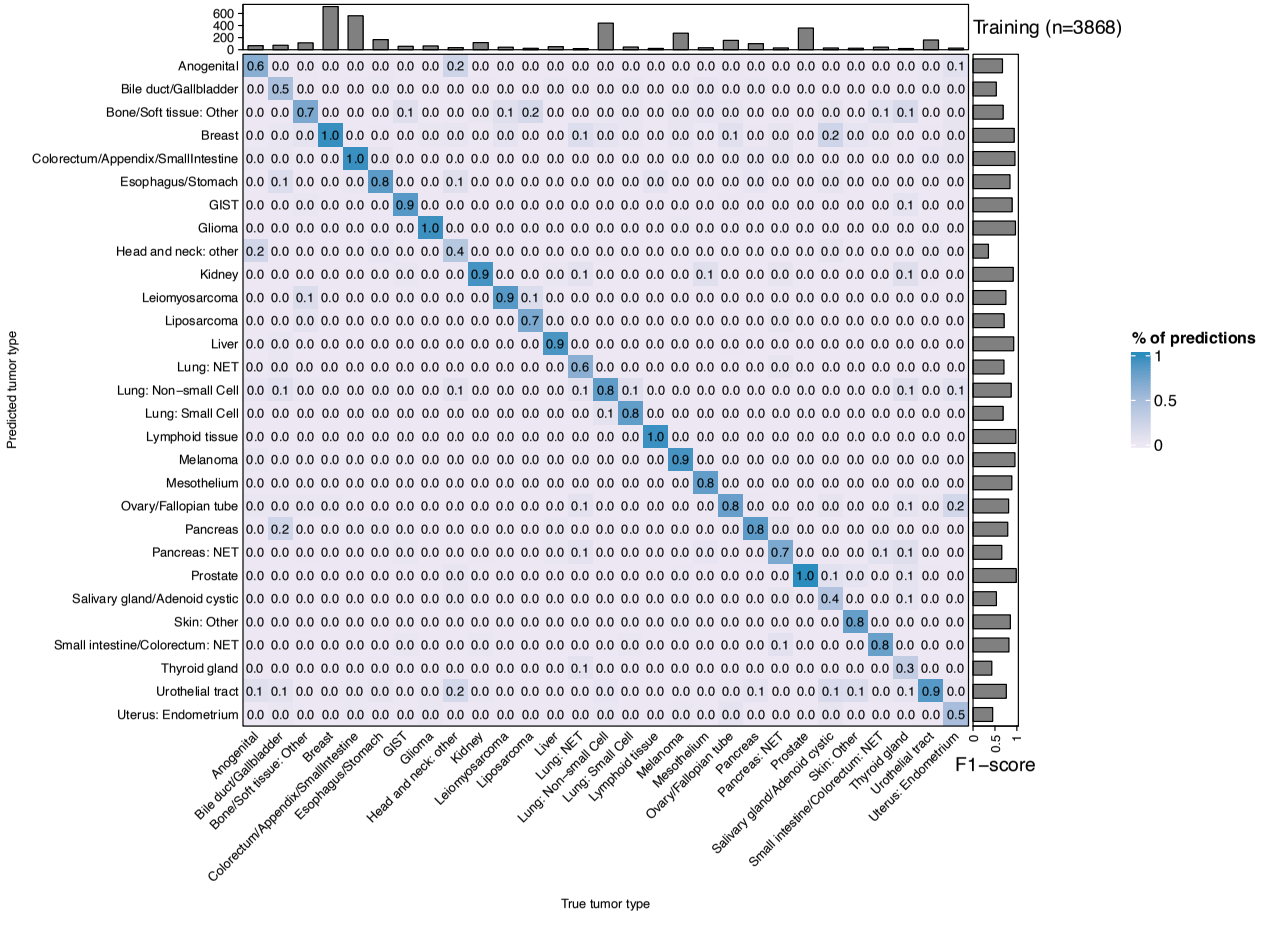


Figure S1. Confusion matrix internal validation set. Samples with a unknown primary origin (n=7), or with a primary tumor type not fitting any of the reference cohorts (n=2) are excluded from this analysis.

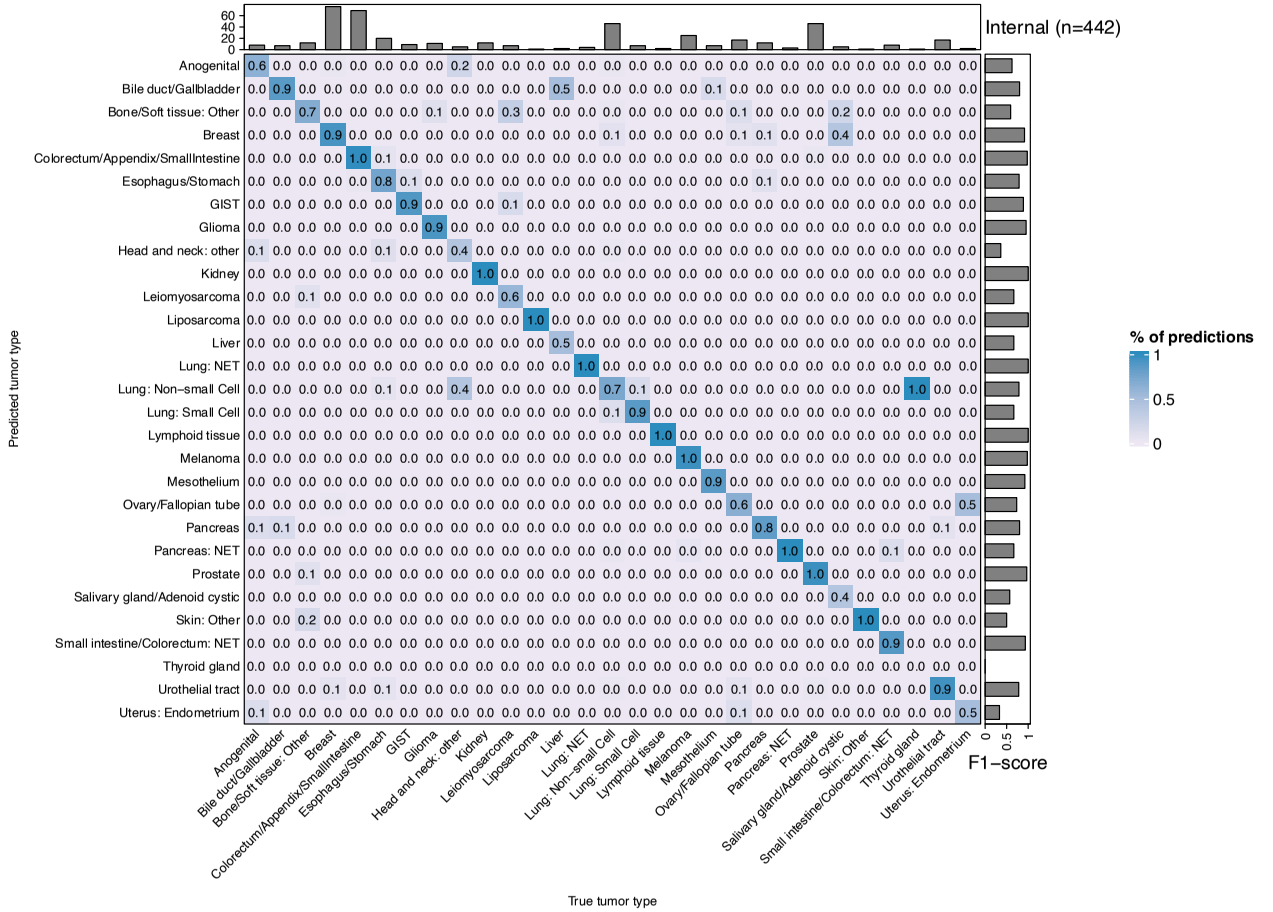


Figure S2. Confusion matrix reference set

*

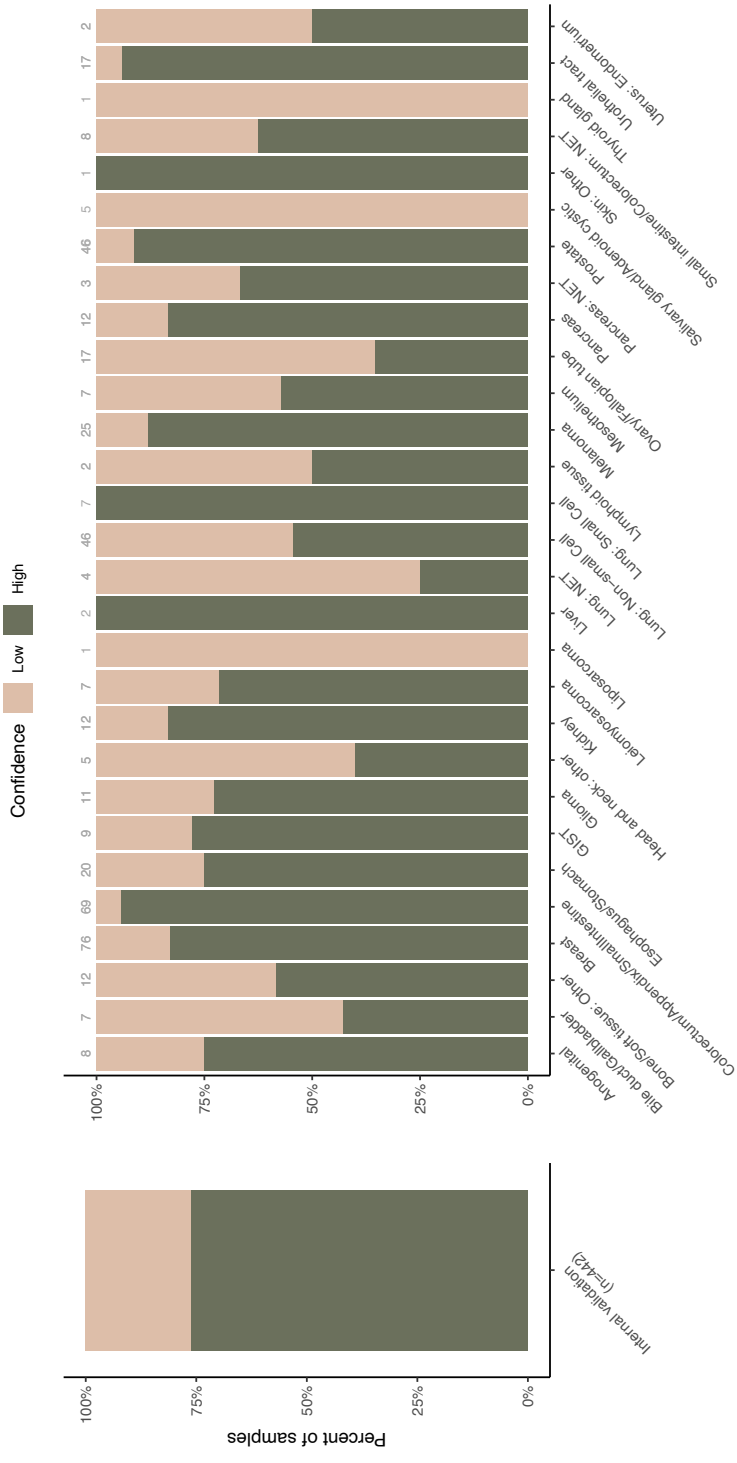


Figure S3. Distribution of high- and low-confidence predictions in the internal validation set

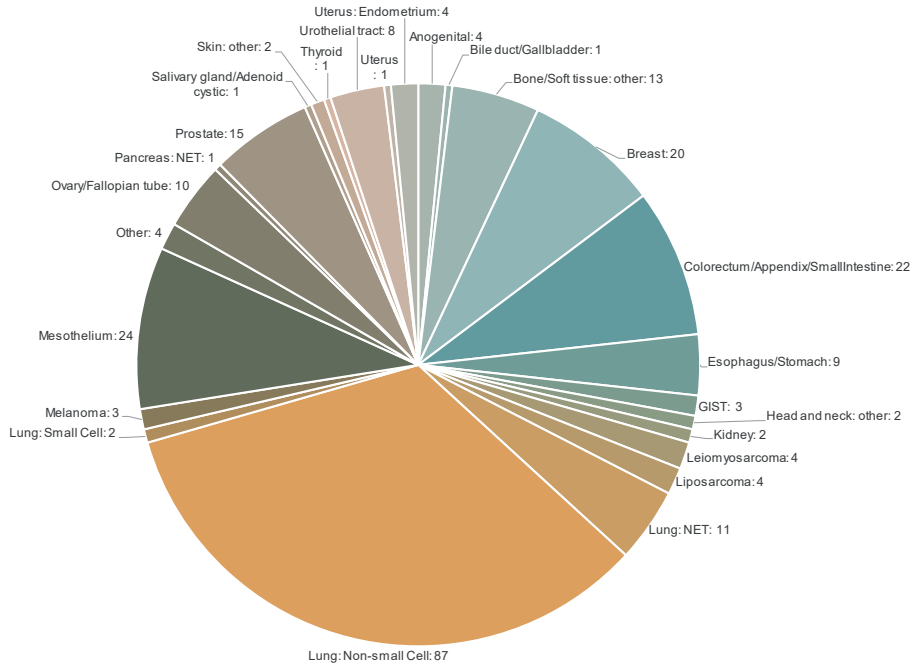


Figure S4. Tumor types in external validation cohort

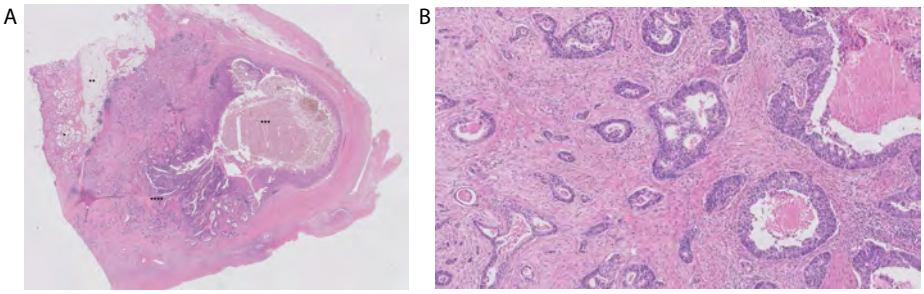


Figure S5. Morphology of malignant teratoma case

Histological figures of ovarian mature teratoma with malignant transformation into intestinal type adenocarcinoma, for which CUPPA predicted colorectal cancer with high confidence (0.82). In the overview figure (panel A) it can be appreciated that the mature teratoma contains on the left skin and adnexal elements (*) and adipose tissue (**). More on the right there is a cyst (***) representing a well ordinated intestinal structure. Such structures have previously been described as a very rare finding in mature cystic teratomas (33). The central part of panel A is occupied by a malignant transformation into an intestinal type adenocarcinoma (****). In more detail (panel B), the adenocarcinoma demonstrates colonic type morphology with tubular and cribriform glands containing luminal dirty necrosis.

Data supplement Chapter 7

Driver mutations occur frequently in metastases of well-differentiated small intestinal neuroendocrine tumors

Appendix 1. 508 cancer driver genes

Gene,chromosome,chromosomeBand,transcriptId,transcriptStart,transcriptEnd

ABCB1,7,q21.12,ENST00000265724,87133175,87342564
ABCC3,17,q21.33,ENST00000285238,48712218,48769613
ABL1,9,q34.12,ENST00000372348,133589333,133761070
AC093642.5,2,q37.3,ENST00000456398,243030784,243082789
AC144568.2,8,p23.3,ENST00000522481,22601,29775
ACVR1,2,q24.1,ENST00000263640,158592958,158731623
ACVR1B,12,q13.13,ENST00000541224,52345485,52388001
ACVR2A,2,q22.3,ENST00000241416,148602086,148688393
ADAM30,1,p12,ENST00000369400,120436156,120439118
ADNP2,18,q23,ENST00000262198,77866915,77898234
AGK,7,q34,ENST00000355413,141250989,141355044
AHCYL1,1,p13.3,ENST00000369799,110527308,110566357
AJUBA,14,q11.2,ENST00000262713,23440383,23451851
AKT1,14,q32.33,ENST00000554581,105235686,105260461
AKT2,19,q13.2,ENST00000392038,40736224,40791302
AKT3,1,q43-q44,ENST00000366539,243665065,244013430
AL356215.1,11,p13,ENST00000598940,35150091,35150215
ALB,4,q13.3,ENST00000295897,74269956,74287129
ALK,2,p23.1-p23.2,ENST00000389048,29415640,30144432
AMBRA1,11,p11.2,ENST00000314845,46417964,46612914
AMER1,X,q11.2,ENST00000330258,63404997,63425624
ANKRD11,16,q24.3,ENST00000301030,89334038,89556969
AP001464.4,21,q11.2,ENST00000457565,14371997,14389013
APC,5,q22.2,ENST00000457016,112043218,112181936
APOOL,X,q21.1,ENST00000373173,84258832,84343069
AR,X,q12,ENST00000374690,66764465,66950461
ARAF,X,p11.23,ENST00000377045,47420516,47431299
AREG,4,q13.3,ENST00000395748,75310851,75320726
ARHGAP35,19,q13.32,ENST00000404338,47421933,47508334
ARID1A,1,p36.11,ENST00000324856,27022524,27108595
ARID1B,6,q25.3,ENST00000346085,157099063,157531913
ARID2,12,q12,ENST00000334344,46123448,46301823
ARID5B,10,q21.2,ENST00000279873,63661059,63856703
ASNS,7,q21.3,ENST00000175506,97481430,97501854
ASXL1,20,q11.21,ENST00000375687,30946155,31027122
ASXL2,2,p23.3,ENST00000435504,25956622,26101385
ATG7,3,p25.3,ENST00000354449,11314102,11599139
ATM,11,q22.3,ENST00000278616,108093559,108239826
ATP1A1,1,p13.1,ENST00000537345,116916489,116947394
ATP2B3,X,q28,ENST00000263519,152801580,152848387
ATR,3,q23,ENST00000350721,142168079,142297668
ATRX,X,q21.1,ENST00000373344,76760356,77041702
AURKA,20,q13.2,ENST00000395909,54944445,54967351
AXIN1,16,p13.3,ENST00000262320,337440,402659
AXIN2,17,q24.1,ENST00000307078,63524681,63557765
AXL,19,q13.2,ENST00000301178,41725108,41767671
B2M,15,q21.1,ENST00000558401,45003675,45011075
BAIAP2L1,7,q21.3-q22.1,ENST00000005260,97920963,98030380
BAP1,3,p21.1,ENST00000460680,52435029,52444366
BARD1,2,q35,ENST00000260947,215590370,215674428
BCL2,18,q21.33,ENST00000398117,60790579,60987361
BCL9L,11,q23.3,ENST00000334801,118764584,118781613
BCOR,X,p11.4,ENST00000378444,39910501,39956656
BCR,22,q11.23,ENST00000305877,23522397,23660224
BICC1,10,q21.1,ENST00000373886,60272900,60591195
BIRC3,11,q22.2,ENST00000263464,102188226,102208465
BIRC7,20,q13.33,ENST00000217169,61867235,61871859
BMPR2,2,q33.1-q33.2,ENST00000374580,203241659,203432474
BRAF,7,q34,ENST00000288602,140434279,140624564
BRCA1,17,q21.31,ENST00000471181,41197646,41277500
BRCA2,13,q13.1,ENST00000544455,32889617,32973805
BRD4,19,p13.12,ENST00000263377,15347647,15391262
BRD7,16,q12.1,ENST00000394689,50352941,50402690
BRIP1,17,q23.2,ENST00000259008,59758627,59940882
BTG3,21,q21.1,ENST00000339775,18965971,18985162
BTK,X,q22.1,ENST00000308731,100604438,100641183
CACNA1D,3,p21.1,ENST00000288139,53529076,53847760
CALR,19,p13.2,ENST00000316448,13049421,13055303
CARD11,7,p22.2,ENST00000396946,2945775,3083579
CASP8,2,q33.1,ENST00000358485,202122759,202152434
CASZ1,1,p36.22,ENST00000377022,10696661,10856705
CBFB,16,q22.1,ENST00000412916,67063148,67134925
CBL,11,q23.3,ENST00000264033,119076752,119178859
CBLB,3,q13.11,ENST00000264122,105374305,105587887
CCDC6,10,q21.2,ENST00000263102,61548521,61666414
CCND1,11,q13.3,ENST00000227507,69455855,69469242
CCND2,12,p13.32,ENST00000261254,4382938,4414516
CCND3,6,p21.1,ENST00000372991,41902671,41909586
CCNE1,19,q12,ENST00000262643,30302805,30315215
CCSER1,4,q22.1,ENST00000509176,91048736,92523064
CD274,9,p24.1,ENST00000381577,5450525,5470547
CD28,2,q33.2,ENST00000324106,204571271,204603635

CD58,1,p13.1,ENST00000369489,117057157,117113661
CD79A,19,q13.2,ENST00000221972,42381190,42385439
CD79B,17,q23.3,ENST00000392795,62006100,62009696
CDC73,1,q31.2,ENST00000367435,193091147,193223031
CDH1,16,q22.1,ENST00000261769,68771128,68869451
CDH10,5,p14.1-p14.2,ENST00000264463,24487209,24645087
CDK12,17,q12,ENST00000447079,37618292,37691399
CDK4,12,q14.1,ENST00000257904,58141510,58146304
CDK6,7,q21.2,ENST00000265734,92234235,92463231
CDKN1A,6,p21.2,ENST00000405375,36646487,36655108
CDKN1B,12,p13.1,ENST00000228872,12870058,12875305
CDKN2A,9,p21.3,ENST00000498124,21968055,21974865
CDKN2B,9,p21.3,ENST00000276925,22002902,22009362
CDKN2C,1,p32.3,ENST00000262662,51426417,51440305
CDX2,13,q12.2,ENST00000381020,28536274,28545276
CEBPA,19,q13.11,ENST00000498907,33790840,33793470
CHD1,5,q15-q21.1,ENST00000284049,98191449,98262240
CHEK1,11,q24.2,ENST00000534070,125496236,125527031
CHEK2,22,q12.1,ENST00000382580,29083751,29137826
CIC,19,q13.2,ENST00000575354,42788817,42799949
CLTC,17,q23.1,ENST00000269122,57697219,57773671
CNOT3,19,q13.42,ENST00000406403,54645112,54659418
COL1A1,17,q21.33,ENST00000225964,48260650,48278993
COL2A1,12,q13.11,ENST00000380518,48366748,48398269
CRBN,3,p26.2,ENST00000231948,3191695,3221394
CREBBP,16,p13.3,ENST00000262367,3775055,3930727
CRLF2,X,p22.33,ENST00000381567,1314890,1331527
CSF1,1,p13.3,ENST00000329608,110453255,110473614
CSF1R,5,q32,ENST00000286301,149432854,149492935
CSF3R,1,p34.3,ENST00000373103,36931644,36948879
CSMD1,8,p23.2,ENST00000537824,2796107,4851938
CTCF,16,q22.1,ENST00000264010,67596310,67673086
CTLA4,2,q33.2,ENST00000302823,204732509,204738683
CTNNA1,5,q31.2,ENST00000302763,138089112,138270723
CTNNB1,3,p22.1,ENST00000349496,41240930,41281936
CUX1,7,q22.1,ENST00000360264,101459291,101901513
CXCR4,2,q22.1,ENST00000409817,136871919,136873813
CYLD,16,q12.1,ENST00000427738,50776671,50835846
DAXX,6,p21.32,ENST00000374542,33286335,33290791
DDR2,1,q23.3,ENST00000367922,162602255,162757190
DDX3X,X,p11.4,ENST00000399959,41192651,41209462
DEPDC5,22,q12.2-q12.3,ENST00000382112,32150838,32302991
DGCR8,22,q11.21,ENST00000351989,20067755,20099394
DIAPH2,X,q21.33,ENST00000324765,95939711,96859996
DICER1,14,q32.13,ENST00000526495,95552565,95624347
DMD,X,p21.1-p21.2,ENST00000357033,31137345,33229636
DNM2,19,p13.2,ENST00000389253,10828807,10942579
DNMT3A,2,p23.3,ENST00000264709,25455845,25565459
DPYD,1,p21.3,ENST00000370192,97543299,98386579
DROSHA,5,p13.3,ENST00000511367,31400604,31532168
EBF1,5,q33.3,ENST00000313708,158122928,158526769
EEF1A1,6,q13,ENST00000316292,74225473,74230741
EGFR,7,p11.2,ENST00000275493,55086794,55279321
ELF3,1,q32.1,ENST00000359651,201977073,201985137
EML4,2,p21,ENST00000318522,42396490,42559688
EP300,22,q13.2,ENST00000263253,41487790,41576081
EPAS1,2,p21,ENST00000263734,46524541,46613836
EPHA2,1,p36.13,ENST00000358432,16450832,16482582
EPHA3,3,p11.1,ENST00000336596,89156674,89531284
ERBB2,17,q12,ENST00000269571,37856333,37884915
ERBB3,12,q13.2,ENST00000267101,56473645,56497289
ERBB4,2,q34,ENST00000342788,212240446,213403565
ERCC1,19,q13.32,ENST00000013807,45916692,45926824
ERCC2,19,q13.32,ENST00000391945,45853095,45873876
ERCC4,16,p13.12,ENST00000311895,14014014,14046202
ERCC6,10,q11.23,ENST00000355832,50663414,50747072
EREG,4,q13.3,ENST00000244869,75230860,75254468
ERG,21,q22.2,ENST00000417133,39751949,40033618
ERRFI1,1,p36.23,ENST00000377482,8071779,8086368
ESR1,6,q25.1,ENST00000440973,152011631,152424409
ETNK1,12,p12.1,ENST00000266517,22778009,22843599
ETS2,21,q22.2,ENST00000360214,40177231,40196879
ETV6,12,p13.2,ENST00000396373,11802788,12048336
EWSR1,22,q12.2,ENST00000414183,29664305,29696333
EYS,6,q12,ENST00000503581,64429876,66417118
EZH2,7,q36.1,ENST00000320356,148504477,148581370
FANCA,16,q24.3,ENST00000389301,89803957,89883054
FANCC,9,q22.32,ENST00000289081,97861336,98079984
FANCL,2,p16.1,ENST00000402135,58386382,58468485
FAT1,4,q35.2,ENST00000441802,187508937,187645009
FAT4,4,q28.1,ENST00000394329,126237554,126414087
FBXL17,5,q21.3,ENST00000542267,107194746,107717799
FBXO11,2,p16.3,ENST00000403359,48034059,48132932
FBXW7,4,q31.3,ENST00000281708,153242410,153457253
FGF3,11,q13.3,ENST00000334134,69624992,69633792
FGF4,11,q13.3,ENST00000168712,69587797,69590171
FGFR1,8,p11.22-p11.23,ENST00000425967,38268656,38325363
FGFR2,10,q26.13,ENST00000457416,123239371,123357917
FGFR3,4,p16.3,ENST00000340107,1795039,1810599
FGFR4,5,q35.2,ENST00000292408,176513887,176525145
FHIT,3,p14.2,ENST00000468189,59737133,61237133
FHOD3,18,q12.2,ENST00000257209,33877677,34360018
FIP1L1,4,q12,ENST00000337488,54243812,54325835
FLCN,17,p11.2,ENST00000285071,17115526,17140453
FLI1,11,q24.3,ENST00000527786,128563665,128683162
FLT1,13,q12.2-q12.3,ENST00000282397,28874489,29069232
FLT3,13,q12.2,ENST00000241453,28577411,28674729
FLT4,5,q35.3,ENST00000261937,180028506,180076624
FNTB,14,q23.3,ENST00000246166,65453438,65529316
FOSL2,2,p23.2,ENST00000264716,28615725,28640179
FOXA1,14,q21.1,ENST00000250448,38059189,38064239
FOXA2,20,p11.21,ENST00000419308,22561643,22565101

*

FOXD4,9,p24.3,ENST00000382500,116237,118417
FOXL2,3,q22.3,ENST00000330315,138663066,138665982
FOXO3,6,q21,ENST00000406360,108882069,109005971
FOXP1,3,p13,ENST00000491238,71008342,71179988
FOXQ1,6,p25.3,ENST00000296839,1312675,1314992
FRG1,4,q35.2,ENST00000226798,190861943,190884359
FRS2,12,q15,ENST00000299293,69864129,69973559
FUBP1,1,p31.1,ENST00000370768,78414090,78444770
G6PD,X,q28,ENST00000393562,153759606,153775469
GAGE12J,X,p11.23,ENST00000442437,49178536,49185863
GATA1,X,p11.23,ENST00000376670,48644962,48652715
GATA2,3,q21.3,ENST00000341105,128198270,128212028
GATA3,10,p14,ENST00000379328,8096656,8117161
GLI2,2,q14.2,ENST00000452319,121549985,121750229
GLRX3,10,q26.3,ENST00000368644,131934663,131978640
GMDS,6,p25.3,ENST00000380815,1624041,2245926
GNA11,19,p13.3,ENST00000078429,3094408,3124002
GNAQ,9,q21.2,ENST00000286548,80331003,80646374
GNAS,20,q13.32,ENST00000371100,57427769,57486247
GOLGA6L6,15,q11.2,ENST00000427390,20737094,20747114
GPHN,14,q23.3,ENST00000478722,66974125,67648520
GPS2,17,p13.1,ENST00000380728,7215981,7218883
GRIN2A,16,p13.2,ENST00000396573,9852376,10276611
GSK3B,3,q13.33,ENST00000316626,119545533,119812513
GSTP1,11,q13.2,ENST00000398606,67351066,67354131
H3F3A,1,q42.12,ENST00000366813,226251678,226259702
H3F3B,17,q25.1,ENST00000254810,73772515,73775860
HDHD1,X,p22.31,ENST00000424830,6967804,7066199
HIF1A,14,q23.2,ENST00000539097,62164340,62214489
HIST1H1C,6,p22.2,ENST00000343677,26055968,26056699
HIST1H3B,6,p22.2,ENST00000244661,26031817,26032288
HIST2H3D,1,q21.2,ENST00000331491,149784826,149785236
HLA-A,6,p22.1,ENST00000396634,29909037,29913661
HLA-B,6,p21.33,ENST00000412585,31321649,31324964
HLA-C,6,p21.33,ENST00000376228,31236526,31239863
HNF1A,12,q24.31,ENST00000257555,121416346,121440315
HRAS,11,p15.5,ENST00000451590,532243,535550
IDH1,2,q34,ENST00000415913,209100989,209119046
IDH2,15,q26.1,ENST00000330062,90626277,90645736
IFNL2,19,q13.2,ENST00000331982,39759154,39760732
IGF1R,15,q26.3,ENST00000268035,99192200,99507759
IGF2,11,p15.5,ENST00000434045,2153730,2162468
IGLL5,22,q11.22,ENST00000526893,23229960,23238005
IKBKB,8,p11.21,ENST00000520810,42128820,42189126
IKZF1,7,p12.2,ENST00000439701,50348318,50470264
IL1RAPL1,X,p21.2-p21.3,ENST00000378993,28605516,29974840
IL6ST,5,q11.2,ENST00000381298,55230923,55290821
IL7R,5,p13.2,ENST00000303115,35856951,35879705
IL9R,X,q28,ENST00000244174,155227246,155240273
IMMP2L,7,q31.1,ENST00000405709,110303110,111202347
INPP4B,4,q31.21,ENST00000513000,142944313,143767443
IRF2,4,q35.1,ENST00000393593,185308867,185395734
ITGAV,2,q32.1,ENST00000261023,187454792,187545628
JAK1,1,p31.3,ENST00000342505,65298912,65432187
JAK2,9,p24.1,ENST00000381652,4985245,5128183
JAK3,19,p13.11,ENST00000458235,17935595,17958841
KANSL1,17,q21.31,ENST00000262419,44107322,44270166
KCNJ5,11,q24.3,ENST00000529694,128761251,128790930
KDM5C,X,p11.22,ENST00000375401,53221334,53254604
KDM6A,X,p11.3,ENST00000377967,44732757,44971847
KDR,4,q12,ENST00000263923,55944644,55991756
KEAP1,19,p13.2,ENST00000171111,10596796,10614417
KIAA1549,7,q34,ENST00000422774,138522270,138666064
KIF5B,10,p11.22,ENST00000302418,32297938,32345359
KIT,4,q12,ENST00000288135,55524085,55606881
KLF4,9,q31.2,ENST00000374672,110247133,110251927
KLF5,13,q22.1,ENST00000377687,73632930,73651676
KMT2A,11,q23.3,ENST00000534358,118307205,118397539
KMT2B,19,q13.12,ENST00000222270,36208921,36229779
KMT2C,7,q36.1,ENST00000262189,151832010,152133090
KMT2D,12,q13.12,ENST00000301067,49412758,49449107
KRAS,12,p12.1,ENST00000256078,25362365,25403737
KRT5,12,q13.13,ENST00000252242,52908359,52914471
KRTAP5-5,11,p15.5,ENST00000399676,1651033,1652160
LATS2,13,q12.11,ENST00000382592,21547171,21635686
LINC00221,14,q32.33,ENST00000334298,106938474,106951462
LINC00290,4,q34.3,ENST00000512487,181985242,182080302
LINC01001,11,p15.5,ENST00000540375,127115,131056
LMNA,1,q22,ENST00000368300,156084498,156109880
LRP1B,2,q22.1-q22.2,ENST00000389484,140988992,142889270
LRRN3,7,q31.1,ENST00000451085,110731062,110765507
LSAMP,3,q13.31-q13.32,ENST00000490035,115521235,116164378
LZTR1,22,q11.21,ENST00000215739,21336302,21353327
MACROD2,20,p12.1,ENST00000217246,13976015,16033842
MAP2K1,15,q22.31,ENST00000307102,66679155,66784650
MAP2K2,19,p13.3,ENST00000262948,4090319,4124126
MAP2K4,17,p12,ENST00000353533,11924141,12047140
MAP2K7,19,p13.2,ENST00000397979,7968776,7979363
MAP3K1,5,q11.2,ENST00000399503,56111401,56191979
MAP3K13,3,q27.2,ENST00000265026,185080908,185206885
MAPK1,22,q11.21-q11.22,ENST00000215832,22108789,22221919
MAX,14,q23.3,ENST00000358664,65542262,65569188
MCL1,1,q21.3,ENST00000369026,150547032,150552066
MDM2,12,q15,ENST00000462284,69201956,69239214
MDM4,1,q32.1,ENST00000367182,204485511,204527248
MED12,X,q13.1,ENST00000374080,70338573,70362297
MEF2D,1,q22,ENST00000348159,156433519,156470620
MEN1,11,q13.1,ENST00000337652,64570996,64578766
MET,7,q31.2,ENST00000318493,116312459,116436396
MGA,15,q15.1,ENST00000219905,41952610,42062141
MGEA5,10,q24.32,ENST00000361464,103544209,103578175
MGMT,10,q26.3,ENST00000306010,131265448,131566271

MITF,3,p13-p14.1,ENST00000352241,69788586,70017487
MLH1,3,p22.2,ENST00000231790,37034823,37092409
MLK4,1,q42.2,ENST00000366624,233463514,233520894
MLLT4,6,q27,ENST00000392108,168227671,168364976
MPL,1,p34.2,ENST00000372470,43803478,43818443
MSH2,2,p21,ENST00000233146,47630108,47710367
MSH3,5,q14.1,ENST00000265081,79950467,80172279
MSH6,2,p16.3,ENST00000234420,48010221,48037240
MSLN,16,p13.3,ENST00000382862,811089,818861
MST1R,3,p21.31,ENST00000296474,49924435,49941070
MTAP,9,p21.3,ENST00000380172,21802542,21867077
MTHFR,1,p36.22,ENST00000376592,11845780,11863302
MTOR,1,p36.22,ENST00000361445,11166592,11322564
MUC6,11,p15.5,ENST00000421673,1012821,1036706
MYC,8,q24.21,ENST00000377970,128748330,128753674
MYCL,1,p34.2,ENST00000397332,40361098,40367685
MYCN,2,p24.3,ENST00000281043,16080686,16087129
MYD88,3,p22.2,ENST00000417037,38179969,38184510
MYOD1,11,p15.1,ENST00000250003,17741115,17743678
NAALADL2,3,q26.31,ENST00000454872,174577070,175523428
NCOA2,8,q13.3,ENST00000452400,71021997,71316040
NCOR1,17,p11.2-p12,ENST00000268712,15932471,16118863
NEGR1,1,p31.1,ENST00000357731,71861623,72748417
NF1,17,q11.2,ENST00000358273,29421945,29704695
NF2,22,q12.2,ENST00000338641,29999547,30094587
NFE2L2,2,q31.2,ENST00000397062,178095033,178129859
NFKBIE,6,p21.1,ENST00000275015,44225903,44233500
NOTCH1,9,q34.3,ENST00000277541,139388896,139440314
NOTCH2,1,p11.2-p12,ENST00000256646,120454176,120612240
NPM1,5,q35.1,ENST00000296930,170814652,170838141
NQO1,16,q22.1,ENST00000320623,69743304,69760854
NRAS,1,p13.2,ENST00000369535,115247090,115259515
NRG1,8,p12,ENST00000356819,32405728,32622548
NSD1,5,q35.2-q35.3,ENST00000439151,176560926,176727216
NT5C2,10,q24.32-q24.33,EN
ST00000343289,104845940,104953056
NTRK1,1,q23.1,ENST00000524377,156830686,156851434
NTRK2,9,q21.33,ENST00000376214,87283466,87638505
NTRK3,15,q25.3,ENST00000360948,88420022,88799661
NUTM1,15,q14,ENST00000333756,34638066,34649933
OR11H1,22,q11.1,ENST00000252835,16448824,16449805
OR4N2,14,q11.2,ENST00000315947,20295608,20296531
OSBPL1A,18,q11.2,ENST00000319481,21742008,21977823
PABPC1,8,q22.3,ENST00000318607,101715144,101734940
PAK1,11,q13.5-q14.1,ENST00000278568,77033993,77185105
PALB2,16,p12.2,ENST00000261584,23614491,23652631
PAPSS1,4,q25,ENST00000265174,108534822,108641608
PARD3,10,p11.21-p11.22,ENST00000374789,34398489,35104249
PARK2,6,q26,ENST00000366898,161768452,163148803
PAX5,9,p13.2,ENST00000358127,36833272,37034103
PBRM1,3,p21.1,ENST00000394830,52581857,52719852
PCDH11X,X,q21.31-q21.32,EN
ST00000373094,91089659,91878226
PCM1,8,p22,ENST00000325083,17780349,17885477
PDE4D,5,q11.2-q12.1,ENST00000340635,58264865,59189625
PDGFB,22,q13.1,ENST00000331163,39619364,39640756
PDGFRA,4,q12,ENST00000257290,55095264,55164414
PDGFRB,5,q32,ENST00000261799,149493400,149535423
PDPK1,16,p13.3,ENST00000342085,2587965,2653189
PDYN,20,p13,ENST00000217305,1959405,1974703
PHF6,X,q26.2,ENST00000332070,133507342,133562819
PHOX2B,4,p13,ENST00000226382,41746099,41750987
PIK3CA,3,q26.32,ENST00000263967,178866311,178957881
PIK3CB,3,q22.3,ENST00000477593,138372860,138553594
PIK3R1,5,q13.1,ENST00000521381,67511548,67597649
PIK3R2,19,p13.11,ENST00000222254,18263928,18281350
PIK3R4,3,q22.1,ENST00000356763,130397779,130465673
PLCG1,20,q12,ENST00000373272,39765877,39804361
PLCG2,16,q23.3,ENST00000359376,81812863,81991899
PML,15,q24.1,ENST00000268058,74287058,74339112
PMS2,7,p22.1,ENST00000265849,6012870,6048756
POLE,12,q24.33,ENST00000320574,133200348,133263945
POT1,7,q31.33,ENST00000357628,124462440,124570035
PPFIBP2,11,p15.4,ENST00000299492,7535001,7674987
PPM1D,17,q23.2,ENST00000305921,58677544,58741849
PPP1R15A,19,q13.33,ENST00000200453,49375649,49379314
PPP2R1A,19,q13.41,ENST00000322088,52693292,52730687
PPP2R2A,8,p21.2,ENST00000315985,26150732,26228402
PPP2R3B,X,p22.33,ENST00000390665,294698,347445
PPP6C,9,q33.3,ENST00000451402,127908853,127952218
PRDM1,6,q21,ENST00000369096,106534195,106557813
PREX2,8,q13.2,ENST00000288368,68864353,69149265
PRKACA,19,p13.12,ENST00000308677,14202500,14228556
PRKAR1A,17,q24.2,ENST00000589228,66508568,66529572
PRKG1,10,q11.23-q21.1,ENST00000373980,52833934,54058110
PSIP1,9,p22.3,ENST00000380733,15464064,15511017
PTCH1,9,q22.32,ENST00000331920,98205262,98270943
PTEN,10,q23.31,ENST00000371953,89622870,89731687
PTK6,20,q13.33,ENST00000217185,62159778,62168695
PTPN11,12,q24.13,ENST00000351677,112856718,112947717
PTPN13,4,q21.3,ENST00000436978,87515468,87736324
PTPRB,12,q15,ENST00000334414,70910630,71031220
PTPRD,9,p23-p24.1,ENST00000381196,8314246,10033790
PTPRN2,7,q36.3,ENST00000389418,157331760,158380371
QRSL1,6,q21,ENST00000369046,107077453,107116292
RAC1,7,p22.1,ENST00000356142,6414170,6442151
RACGAP1,12,q13.12,ENST00000434422,50382945,50419307
RAD21,8,q24.11,ENST00000297338,117858174,117887105
RAD50,5,q31.1,ENST00000265335,131892630,131979752
RAD51,15,q15.1,ENST00000382643,40987372,41023791
RAD51B,14,q24.1,ENST00000487270,68286525,69062713
RAD51C,17,q22,ENST00000337432,56769934,56811703

*

RAD51D,17,q12,ENST00000590016,33427691,33446832
RAD54L,1,p34.1,ENST00000371975,46713404,46744144
RAF1,3,p25.2,ENST00000251849,12625100,12705725
RANBP2,2,q12.3,ENST00000283195,109335937,109402267
RARA,17,q21.2,ENST00000254066,38465446,38513048
RARG,12,q13.13,ENST00000425354,53604354,53626036
RASA1,5,q14.3,ENST00000274376,86563705,86686744
RB1,13,q14.2,ENST00000267163,48877911,49056122
RBM10,X,p11.23,ENST00000377604,47004268,47046210
RET,10,q11.21,ENST00000355710,43572475,43625799
RHOA,3,p21.31,ENST00000418115,49396578,49449635
RHOB,2,p24.1,ENST00000272233,20646835,20649200
RICTOR,5,p13.1,ENST00000357387,38938021,39074510
RIT1,1,q22,ENST00000368322,155870065,155880706
RNF111,15,q22.1-q22.2,ENST00000559209,59280151,59389253
RNF43,17,q22,ENST00000584437,56431037,56494894
ROS1,6,q22.1,ENST00000368508,117609463,117747018
RP11-436K8.1,1,p32.1,ENST00000423403,61125303,61291256
RP11-686G23.2,11,q14.1,ENST00000526840,80462243,80473869
RPL10,X,q28,ENST00000424325,153626571,153630680
RPL22,1,p36.31,ENST00000234875,6245080,6259672
RPL5,1,p22.1,ENST00000370321,93297582,93307481
RPS6KA3,X,p22.12,ENST00000379565,20168029,20284958
RSF1,11,q14.1,ENST00000308488,77371041,77532063
RSPO2,8,q23.1,ENST00000276659,108911545,109095876
RSPO3,6,q22.33,ENST00000356698,127439749,127518441
RUNX1,21,q22.12,ENST00000300305,36160098,36421641
RXRA,9,q34.2,ENST00000481739,137218426,137328865
SAV1,14,q22.1,ENST00000324679,51100354,51135049
SERPINB3,18,q21.33,ENST00000283752,61322431,61329197
SETBP1,18,q12.3,ENST00000282030,42260863,42648475
SETD2,3,p21.31,ENST00000409792,47058179,47205457
SF3B1,2,q33.1,ENST00000335508,198254508,198299815
SGCZ,8,p22,ENST00000382080,13947373,15095848
SH2B3,12,q24.12,ENST00000341259,111843752,111889427
SIX1,14,q23.1,ENST00000247182,61110133,61116180
SIX2,2,p21,ENST00000303077,45232300,45236569
SLC29A1,6,p21.1,ENST00000393841,44191363,44201888
SLCO1B1,12,p12.1-p12.2,ENST00000256958,21284136,21392180
SMAD2,18,q21.1,ENST00000402690,45357922,45457512
SMAD3,15,q22.33,ENST00000327367,67358183,67487533
SMAD4,18,q21.2,ENST00000342988,48556583,48611409
SMARCA1,X,q25,ENST00000371122,128580482,128657477
SMARCA4,19,p13.2,ENST00000429416,11071677,11172958
SMARCB1,22,q11.23,ENST00000263121,24129161,24176703
SMARCD1,12,q13.12,ENST00000394963,50478755,50494495
SMO,7,q32.1,ENST00000249373,128828713,128853386
SMTNL2,17,p13.2,ENST00000389313,4487834,4511609
SOCS1,16,p13.13,ENST00000332029,11348262,11350036
SOX4,6,p22.3,ENST00000244745,21593972,21598847
SOX9,17,q24.3,ENST00000245479,70117161,70122561
SPEN,1,p36.13-p36.21,ENST00000375759,16174359,16266955
SPOP,17,q21.33,ENST00000393331,47676248,47755525
SPTAN1,9,q34.11,ENST00000372739,131314869,131395941
SQSTM1,5,q35.3,ENST00000389805,179247759,179265078
SRC,20,q11.23,ENST00000373578,35974557,36034453
SRSF2,17,q25.1,ENST00000392485,74730197,74733415
STAG2,X,q25,ENST00000218089,123094369,123235348
STAT3,17,q21.2,ENST00000264657,40465342,40540586
STAT5B,17,q21.2,ENST00000293328,40351186,40428424
STK11,19,p13.3,ENST00000326873,1205740,1228428
STS,X,p22.31,ENST00000217961,7137497,7272851
SUFU,10,q24.32,ENST00000369902,104263744,104393292
SUZ12,17,q11.2,ENST00000322652,30264037,30328064
SYK,9,q22.2,ENST00000375754,93564069,93660831
TACC3,4,p16.3,ENST00000313288,1723266,1746898
TBL1XR1,3,q26.32,ENST00000430069,176737143,176914266
TBX3,12,q24.21,ENST00000257566,115108060,115121395
TCF7L2,10,q25.2-q25.3,ENST00000543371,114710009,114927437
TERT,5,p15.33,ENST00000310581,1253282,1295162
TET2,4,q24,ENST00000540549,106067450,106200973
TFG,3,q12.2,ENST00000240851,100428269,100467810
TGFBR2,3,p24.1,ENST00000359013,30648093,30735634
TGIF1,18,p11.31,ENST00000330513,3451675,3458409
TLK2,17,q23.2,ENST00000346027,60556386,60690851
TMPRSS2,21,q22.3,ENST00000398585,42836480,42879992
TNFAIP3,6,q23.3,ENST00000237289,138188581,138204449
TNFRSF10B,8,p21.3,ENST00000276431,22877646,22926692
TNFRSF14,1,p36.32,ENST00000355716,2487805,2495268
TNKS2,10,q23.32,ENST00000371627,93558069,93625033
TOP1,20,q12,ENST00000361337,39657458,39753127
TOP2A,17,q21.2,ENST00000423485,38544768,38574202
TP53,17,p13.1,ENST00000269305,7571720,7590856
TP63,3,q28,ENST00000264731,189349216,189615068
TPMT,6,p22.3,ENST00000309983,18128542,18155305
TRAF7,16,p13.3,ENST00000326181,2205766,2228130
TRIM24,7,q33-q34,ENST00000343526,138145079,138274738
TSC1,9,q34.13,ENST00000298552,135766735,135820008
TSC2,16,p13.3,ENST00000219476,2097466,2138713
TSHR,14,q31.1,ENST00000541158,81421387,81612646
TYMS,18,p11.32,ENST00000323274,657604,673578
U2AF1,21,q22.3,ENST00000291552,44513066,44527697
UBR5,8,q22.3,ENST00000520539,103265240,103425069
UGT1A1,2,q37.1,ENST00000305208,234668894,234681945
USP22,17,p11.2,ENST00000261497,20902910,20946352
USP28,11,q23.2,ENST00000003302,113668596,113746292
USP8,15,q21.2,ENST00000433963,50716579,50793276
VEGFA,6,p21.1,ENST00000372055,43738444,43752346
VHL,3,p25.3,ENST00000256474,10182692,10193904
VMP1,17,q23.1,ENST00000262291,57784826,57918453
WEE1,11,p15.4,ENST00000450114,9595228,9610754
WT1,11,p13,ENST00000332351,32409321,32457176

WWC3,X,p22.2,ENST00000380861,9983602,10112518
WWOX,16,q23.1-q23.2,ENST00000566780,78133310,79246564
XPO1,2,p15,ENST00000401558,61704984,61765418
XRCC1,19,q13.31,ENST00000262887,44047192,44080158
YAP1,11,q22.1-q22.2,ENST00000282441,101981192,102104154
ZBTB10,8,q21.13,ENST00000430430,81397902,81438500
ZFHX3,16,q22.2-q22.3,ENST00000268489,72816784,73082274
ZFP36L1,14,q24.1,ENST00000439696,69254379,69259957
ZFP36L2,2,p21,ENST00000282388,43449541,43453748
ZFPM1,16,q24.2,ENST00000319555,88519725,88603424
ZFX,X,p22.11,ENST00000379177,24167862,24234206
ZKSCAN1,7,q22.1,ENST00000324306,99613204,99639312
ZMIZ1,10,q22.3,ENST00000334512,80828792,81076276
ZMYM2,13,q12.11,ENST00000382869,20532810,20665968
ZNF217,20,q13.2,ENST00000371471,52183604,52210378
ZNF557,19,p13.2,ENST00000414706,7069471,7087979
ZNF703,8,p11.23,ENST00000331569,37553269,37557537
ZNF750,17,q25.3,ENST00000269394,80787311,80798454
ZNRF3,22,q12.1,ENST00000544604,29279580,29453475
ZRSR2,X,p22.2,ENST00000307771,15808595,15841383

Appendix 2: CIRCOS plot

The outer first circle shows the chromosomes. The darker shaded areas represent large gaps in the human reference genome: i.e. regions of centromeres, heterochromatin & missing short arms.

The second circle shows all tumour specific variants (including exon, intron and intergenic regions) and are divided into an outer ring of single nucleotide polymorphism (SNP) allele frequencies and an inner ring of short insertion/deletion (INDEL) locations. Variant allele frequencies have been corrected for tumour purity and scale from 0 to 100%. Each dot represents a single variant and are colored according to the type of base change (e.g. C>T/G/A in red) and are in concordance with the coloring used in Alexandrov et al. 2013 Nature paper which describes the use of mutational signatures. INDELs are colored yellow and red for insertions and deletions respectively.

The third circle shows all observed tumour purity adjusted copy number changes, including both focal and chromosomal events. Copy number losses are indicated in red, green shows regions of copy number gain. The scale ranges from 0 (complete loss) to 6 (high level gains). If the absolute copy number is >6 it is shown as 6 with a green dot on the diagram.

The fourth circle represents the observed 'minor allele copy numbers' across the chromosome. The range of the chart is from 0 to 3. The expected normal minor allele copy number is 1, and anything below 1 is shown as a loss and represents a LOH event (orange). Minor allele copy numbers above 1 indicate amplification events of both A and B alleles at the indicated locations (blue).

The innermost circle displays the observed structural variants within or between the chromosomes. Translocations are indicated in blue, deletions in red, insertions in yellow, tandem duplications in green and inversions in black.

Data supplement Chapter 8

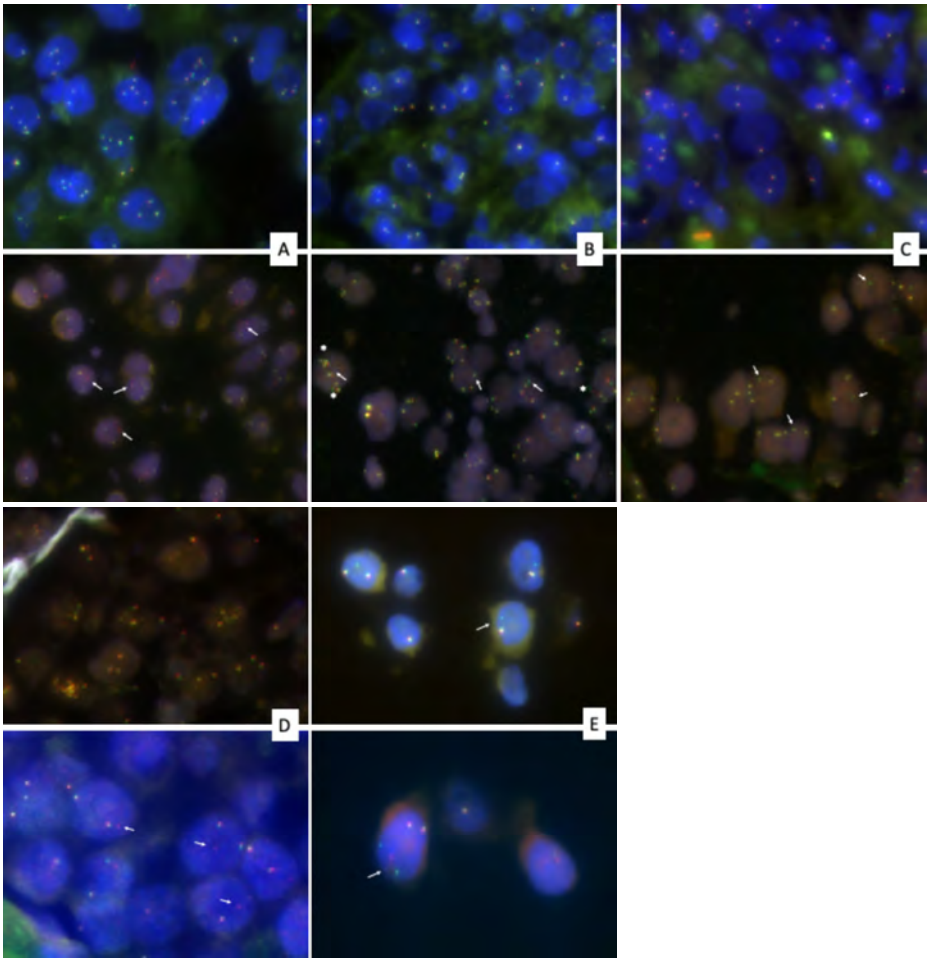


Figure S1. RET FISH analysis is a sensitive but highly unspecific screening method for RET fusions in lung cancer

Discordant results when different FISH assays were used. FISH of 5 cases with different patterns when tested with two different assays. In the upper panels cases were tested using probe 1 with 3' labeled green, lower panel probe 2 where 3' was labeled red. A. Case 1: note polyploidy and a single 3' pattern in the lower panel and fused signals in the upper panel. B. Case 2: note polyploidy (n=3-5) and a low abundant single 5' (white arrow, 16% of nuclei) in the lower panel. Next to that, in few cells complex mixed signals were noted (asterisk). Using another probe tumor cells showed polyploidy. C. Case 3: in the upper panel the tumor cells have clear fused signals. Note the single 5' pattern in the lower panel using another probe. D. Case 4: in the upper panel tumor cells have fused signals in the background of unspecific probe binding. Using another probe, a single 3' pattern was noted (lower panel, arrows). E Case 5 (cytology sample). The upper panel shows predominantly single green (3')signals, same case using another probe showed split signals.

Contributing authors and affiliations

Battaglia, Thomas, MSc
Department of Gastro-enterology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Boelens, Mirjam C, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Bosch, Linda JW, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Buffart, Tineke E, MD, PhD
Department of Gastrointestinal Oncology, Netherlands Cancer Institute, The Netherlands

Cohen, Danielle, MD, PhD
Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

Cuppen, Edwin, PhD
Hartwig Medical Foundation, Amsterdam, The Netherlands
Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands
Oncode Institute, Utrecht, The Netherlands

De Bruijn, Ewart,
Hartwig Medical Foundation, Amsterdam, The Netherlands

De Langen, Adrianus J, MD, PhD
Department of Thoracic Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Dercksen, M Wouter, MD, PhD
Department of Medical Oncology, Maxima Medical Center, Eindhoven, The Netherlands

Dinjens, Winand NM, MD, PhD
Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Dubbink, Hendrikus J, PhD
Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Frederix, Geert WJ, PhD
Department of Health Technology, University Medical Center Utrecht, Utrecht, the Netherlands

Finn, Stephen P, MD, PhD
Department of Histopathology, St. James's Hospital and Trinity College Dublin, Dublin, Ireland
Cancer Molecular Diagnostics, St. James's Hospital and Trinity College Dublin, Dublin, Ireland
Thoracic Oncology Research Group, Trinity Translational Medical Institute, St. James's Hospital and Trinity College Dublin, Dublin, Ireland

Geurts-Giele, Wilhelmina RR, PhD
Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Hoes, Louisa R, MD, PhD
Department of Molecular Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Jacobs, Nina, MSc
Hartwig Medical Foundation, Amsterdam, The Netherlands

Kodach, Liudmila L, MD, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Klompenhouwer, Elisabeth G, MD, PhD
Department of Radiology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Kuhlmann, Koert FD, MD, PhD
Department of Surgery, Netherlands Cancer Institute, Amsterdam, The Netherlands

Lalezari, Ferry, MD
Department of Radiology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Levy, Sonja, MD
Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Marchetti, Serena, MD, PhD
Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Meijer, Gerrit A, MD, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Meijsen, Isabelle C, MSc
Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Monkhorst, Kim, MD, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

O'Brien, Cathal, PhD
Department of Histopathology, St. James's Hospital and Trinity College Dublin, Dublin, Ireland
Cancer Molecular Diagnostics, St. James's Hospital and Trinity College Dublin, Dublin, Ireland
Thoracic Oncology Research Group, Trinity Translational Medical Institute, St. James's Hospital and Trinity College Dublin, Dublin, Ireland

Priestley, Peter
Hartwig Medical Foundation Australia, Sydney, Australia

Radonic, Teodora, MD, PhD
Department of Pathology, Cancer Center Amsterdam, Vrije University, Amsterdam University Medical Center, Amsterdam, The Netherlands

Retèl, Valesca R, PhD
Department of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands
Department of Healthcare Technology and Services Research, University of Twente, Twente, The Netherlands

Riethorst, Immy, MSc
Hartwig Medical Foundation, Amsterdam, The Netherlands

Roemen, Guido MJM, BSc
Department of Pathology, Maastricht University Medical Center, Maastricht, The Netherlands

Roepman, Paul, PhD
Hartwig Medical Foundation, Amsterdam, The Netherlands

Schipper, Luuk J, MD
Department of Molecular Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Schoenmaker, Lieke
Hartwig Medical Foundation, Amsterdam, The Netherlands

Schout, Daoin, BSc
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Shale, Charles, BSc
Hartwig Medical Foundation Australia, Sydney, Australia

Sleddens, Hein FBM, BSc
Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Snaebjornsson, Petur, MD, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Speel, Ernst Jan M, MD, PhD
Department of Pathology, Maastricht University Medical Center, Maastricht, The Netherlands
School for Oncology and Developmental Biology (GROW), Maastricht, The Netherlands

Steeghs, Neeltje, MD, PhD
Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam

Tesselaar, Margot ET, MD, PhD
Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Thunissen, Erik, MD, PhD
Department of Pathology, Cancer Center Amsterdam, Vrije University, Amsterdam University Medical Center, Amsterdam, The Netherlands

Van den Berg, José G, MD, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Van den Broek, Alexandra J, PhD
Hartwig Medical Foundation, Amsterdam, The Netherlands

Van der Hoeven, Jacobus JM, MD, PhD
Hartwig Medical Foundation, Amsterdam, The Netherlands

Van der Kolk, Lizet E, MD, PhD
Family Cancer Clinic, Netherlands Cancer Institute, Amsterdam, The Netherlands

Van der Noort, Vincent, PhD
Department of Biometrics, Netherlands Cancer Institute, Amsterdam, The Netherlands

Van Snellenberg, Hans, MSc
Hartwig Medical Foundation, Amsterdam, The Netherlands

Van Veenendaal, Linde M, MD, PhD
Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Von der Thüsen, Jan H, MD, PhD
Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Van Wezel, Tom, PhD
Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands
Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

Valk, Gerlof D, MD, PhD
Department of Endocrine Oncology, University Medical Center Utrecht, The Netherlands

Verbeek, Wieke HM, MD, PhD
Department of Gastro-enterology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Voest, Emile E, MD, PhD
Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, the Netherlands
Department of Molecular Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands
Oncode Institute, Utrecht, The Netherlands

Vriens, Menno R, MD, PhD
Department of Surgical Oncology and Endocrine Surgery, University Medical Centre Utrecht, Utrecht, The Netherlands.

Vollebergh, Marieke A, MD, PhD
Department of Gastro-enterology, Netherlands Cancer Institute, Amsterdam, The Netherlands

Weijers, Karin, MSc
Department of Pathology, Cancer Center Amsterdam, Vrije University, Amsterdam University Medical Center, Amsterdam, The Netherlands

List of publications

1. **Samsom KG*,** Schipper LJ*, Roepman P, Bosch LJW, Lalezari F, Klompenhouwer EG, de Langen AJ, Buffart TE, Riethorst I, Schoenmaker L, Schout D, van der Noort V, van den Berg JG, de Bruijn E, van der Hoeven JJM, van Snellenberg H, van der Kolk LE, Cuppen E, Voest EE, Meijer GA†, Monkhorst K†. (2022). Feasibility of whole genome sequencing based tumor diagnostics in routine pathology practice. *Journal of Pathology*. doi: 10.1002/path.5988.
2. Schipper LJ*, **Samsom KG***, Snaebjornsson P, Battaglia T, Bosch LJW, Lalezari F, Priestley P, Shale C, van den Broek AJ, Jacobs N, Roepman P, van der Hoeven JJM, Steeghs N, Vollebergh MA, Marchetti S, Cuppen E, Meijer GA, Voest EE, Monkhorst K. Complete genomic characterization in patients with cancer of unknown primary origin in routine diagnostics. (2022). *Accepted by ESMO open*.
3. **Samsom KG,** Bosch LJW, Schipper LJ, Schout D, Roepman P, Boelens MC, Lalezari F, Klompenhouwer EG, de Langen AJ, Buffart TE, Riethorst I, Schoenmaker L, van den Berg JG, de Bruijn E, van der Hoeven JJM, van Snellenberg H, van der Kolk LE, Cuppen E, Voest EE, Meijer GA, Monkhorst K. Optimization of the workflow for Whole Genome Sequencing in routine pathology practice. *Accepted by Nature Protocols for Clinical Use.*
4. Schipper LJ, Monkhorst K, **Samsom KG**, Bosch LJW, Snaebjornsson P, van Boven H, Roepman P, van der Kolk LE, van Houdt WJ, van der Graaf WTA, Meijer GA, Voest EE. Clinical Impact of Prospective Whole Genome Sequencing in Sarcoma Patients. *Cancers.* 2022. 14(2);436. https://doi.org/10.3390/cancers14020436.
5. Roepman P, de Bruijn E, van Lieshout S, Schoenmaker L, Boelens MC, Dubbink HJ, Geurts Giele WRR, Groenendijk FH, Huibers MMH, Kranendonk MEG, Roemer MGM, **Samsom KG**, Steehouwer M, de Leng WWJ, Hoischen A, Ylstra B, Monkhorst K, van der Hoeven JJM, Cuppen E. Clinical validation of Whole Genome Sequencing for cancer diagnostics. *The Journal of Molecular Diagnostics.* 2021.23;7;816-833. https://doi.org/10.1016/j.jmoldx.2021.04.011.
6. Radonic T*, Geurts-Giele W*, **Samsom KG**, Roemen GMJM, von der Thüsen JH, Thunissen E, Meijsen IC, Sleddens HFBM, Dinjens WNM, Boelens MC, Wijers K, Speel EM, Finn SP, O'Brein C, van Wezel T, Cohen D, Monkhorst K, Roepman P, Dubbink HJ. RET Fluorescence In Situ Hybridization Analysis Is a Sensitive but Highly Unspecific Screening Method for RET Fusions in Lung Cancer. *Journal of Thoracic Oncology.* 2021; S1556-0864(21)01666-X. doi: 10.1016/j.jtho.2021.01.1619
7. **Samsom KG,** Levy S*, van Veenendaal LM*, Roepman P, Kodach LL, Steeghs N, Valk GD, Dercksen MW, Kuhlmann KFD, Verbeek WHM, Meijer GA, Tesselaar MET, van den Berg JG. Driver mutations occur frequently in metastases of well-differentiated small intestinal neuroendocrine tumors. *Histopathology*. 2021; 78(4):566-566. Doi: 10.1111/his.14252.
8. **Samsom KG*,** Bosch LJW*, Schipper L*, Roepman P, de Bruijn E, Hoes LR, Riethorst I, Schoenmaker L, van der Kolk LE, Retel VP, Frederix GWJ, Buffart TE, van der Hoeven JJM, Voest EE, Cuppen E, Monkhorst K, Meijer GA. Study protocol: Whole genome sequencing Implementation in standard Diagnostics for Every Cancer patient (WIDE). *BMC Medical Genomics*. 2020; 13, 169. https://doi.org/10.1186/s12920-020-00814-w.

9. **Samsom KG,** van Veenendaal LM, Valk GD, Vriens MR, Tesselaar MET, van den Berg JG. Molecular prognostic factors in small-intestinal neuroendocrine tumours. *Endocrine Connections.*2019;8(7):907-922. https://doi.org/10.1530/EC-19-0206**.**

*Equal author contribution.

Dankwoord

De totstandkoming van dit proefschrift was niet mogelijk geweest zonder de hulp en steun, zowel direct als indirect, van velen. Graag wil ik iedereen bedanken voor zijn of haar bijdrage. Een aantal personen wil ik in het bijzonder bedanken.

Promotieteam

Prof. dr. **G.A. Meijer**, beste Gerrit, dank voor het in mij gestelde vertrouwen om als PhD-student onderdeel te mogen zijn van dit mooie project. Uw betrokken begeleiding, zorgvuldigheid en kritische blik hebben ervoor gezorgd dat mijn werk naar een hoger niveau werd getild. Ik heb geleerd dat het proces minstens zo belangrijk is als het resultaat.

Prof. dr. **E.E. Voest**, beste Emile, dank voor uw optimisme en de welgemeende interesse. Door uw snelle analyses, wetenschappelijke overzicht en stimulerende aanpak, heeft u dit project in een stroomversnelling gebracht.

Dr. **L. J.W. Bosch**, beste Linda, al vroeg gedurende mijn promotietraject moest je noodgedwongen een stap terug doen als mijn copromotor. Desalniettemin bleef je betrokken en bood je een luisterend oor wanneer ik dat nodig had. Ik waardeer het dat je hiervoor tijd wist vrij te maken. Je hebt je onwijs krachtig getoond gedurende de afgelopen drie jaar en deze kracht wist je ook over te brengen op mij. Je leerde mij om jouw taken en verantwoordelijkheden over te nemen binnen het team. Deze nieuwe rol paste initieel niet feilloos, het waren grote schoenen om te vullen, maar uiteindelijk maakte ik hierdoor een enorme groei door in korte tijd. Daarnaast leerde je mij voor mijzelf op te komen, de leiding te nemen en mijn grenzen te bewaken. Met je waardevolle lessen en inspirerende quotes ben je tot op heden een grote bron van support waar ik uit kan putten, waarvoor mijn dank.

Dr. **K. Monkhorst**, beste Kim, bedankt voor je intensieve begeleiding en toewijding de afgelopen jaren. Onze samenwerking stond in het teken van snel schakelen, efficiëntie en oplossingen. Hierdoor werden we al snel een geoliede machine binnen het WIDE-team. Ter illustratie: we begonnen onze mails te ondertekenen met K&K waarbij er verwarring ontstond bij de ontvangers wie nou Kim en wie Kris was. Je was altijd (maar ook echt altijd!) beschikbaar voor overleg. We hebben meermaals gebeld vanuit verre vakantieoorden (op de latten, vanaf boten of het strand) of tijdens borrels om de laatste puntjes op i te zetten van manuscripten. Bovendien hebben we veel gelachen de afgelopen jaren. Je gevleugelde uitspraken zoals '*If you can't beat them, confuse them*', '*kaliumcowboys*' en '*internerds*' blijven erin. De lessen die je mij bijbracht hadden weinig

met werk te maken en des te meer met het leven. Je leerde mij om een breed vizier te houden, om vrijheid te koesteren, kritisch te blijven (maar niet te) en het allemaal niet te serieus te nemen. Lessen die mij ongetwijfeld gaan helpen in de toekomst.

Leescommissie

Leden van de manuscriptcommissie, prof. dr. I.H.M. Borel Rinkes, prof. dr. R. Bernards, prof. dr. W.T.A. van der Graaf, prof. dr. P.J. van Diest en dr. E.G. Klompenhouwer, hartelijk dank voor de tijd die jullie hebben genomen om mijn proefschrift te lezen en dank voor het zitting nemen in deze commissie.

Neuroendocriene research groep

Prof. dr. **M.R. Vriens**, beste Menno, mijn wetenschappelijke avontuur begon in 2015 in het UMC Utrecht. Als derdejaars geneeskundestudent had ik wetenschappelijke ambities en je toegankelijkheid tijdens mijn heelkunde coschap, maakte dat ik jou als beoogde begeleider zag. Je reageerde op mijn mail met '*Let's do it!*', een reactie kenmerkend voor je enthousiasme. Je wist mijn fascinatie voor neuroendocriene tumoren en hun ontstaan aan te wakkeren met je aanstekelijke energie. Al snel volgde mijn introductie bij het Antoni van Leeuwenhoek ziekenhuis, tevens onderdeel van het 'ENETs centre of excellence'. Deze introductie mondde uit in 3 jaar wetenschappelijk onderzoek als student onderzoeker, gevolgd door 3 jaar als PhD-student. Een zeer succesvolle match waarvoor mijn grote dank! Gedurende de afgelopen jaren wist je ondanks je drukke agenda tijd te vinden om met mij te sparren over zowel de wetenschap als de richting van mijn carrière. Ik wil je bedanken voor je deze richtinggevende gesprekken, betrokkenheid en motiverende woorden.

Dr. **J.G. van den Berg**, beste José, onze eerste ontmoeting was in het kader van het manuscript van hoofdstuk 6. Het klikte meteen tussen ons en een keuzecoschap pathologie was al snel een feit. Je gaf mij de vrijheid om zelfstandig te werken en de pathologie (destijds onbekend terrein) eigenhandig te ontdekken. Toen mijn geneeskunde studie eindigde, stelde je voor om als ANIOS pathologie te komen werken en samen een interessant onderzoeksproject te zoeken. Eerlijk gezegd had ik nog nooit van de functie ANIOS pathologie gehoord maar ik had alle vertrouwen in onze samenwerking. Binnen 3 maanden diende zich al een prachtig project aan namelijk de WIDE studie. Deze aanpak is kenmerkend voor jou, je bent een ster in het creëren van een omgeving waarin mensen tot hun recht komen en kunnen groeien. Ik ben je dankbaar voor je vertrouwen en de geweldige kans die je mij hebt gegeven. Gedurende de afgelopen jaren was je mijn mentor. Je hebt mij veel geleerd op het gebied van sterk empathisch leiderschap, inclusiviteit en opkomen voor waar je in gelooft. Daarnaast hebben we genoten. We hebben alles gevierd wat er te vieren viel,

zelfs tijdens COVID-19 wist je kleine afdelingsborrels te organiseren. Het hoogtepunt was Barcelona waar ik mocht spreken op de EUGW en jij voor een paar uur kwam ingevlogen. We hadden zoveel zeevruchten gegeten dat we allebei misselijk in het vliegtuig terug zaten. Dank voor je niet aflatende steun en *'for just being there'*.

Overige leden van het NET-team: **prof. dr. G.D. Valk, dr. M.E.T. Tesselaar, dr. W.A. Buikhuisen, dr. K.F.D. Kuhlmann, dr. W.H.M. Verbeek, dr. R.S. van Leeuwaarde, dr. M.J.C. van Treijen** en **dr. M.W. Dercksen**, bedankt voor de prettige samenwerking de afgelopen jaren en jullie visie op de manuscripten. **Linde van Veenendaal**, bedankt voor het begeleiden van mijn eerste stappen binnen de wetenschap. **Sonja Levy**, als opvolger van Linde wist je de prettige samenwerking voort te zetten. Dank voor het teamwork van de afgelopen jaren.

WIDE-team

De WIDE studie en vervolgens de implementatie van WGS in het AvL was niet mogelijk geweest zonder de inzet van velen, ieder met zijn of haar eigen waardevolle expertise.

Sandra Visser en Judith Westra, bedankt voor het in recordtempo (18 maanden!) includeren van 1200 patiënten. Dit geldt ook voor alle enthousiaste clinici die hebben geholpen bij het voorlichten en includeren van patiënten. **Dr. L.E. van der Kolk**, Lizet, dank voor je onmisbare kennis op het gebied van de klinische genetica en je vooruitstrevendheid. Daarnaast bedankt voor je persoonlijke betrokkenheid en advies. **Dr. T.E. Buffart en dr. Adrianus J. de Langen**, Tineke en Joop; dank voor jullie uitgebreide klinische ervaring en enthousiasme. **F. Lalezari en dr. E.G. Klompenhouwer**, Ferry en Lisa; dank voor het RECIST beoordelen van de WIDE-patiënten, het was een monsterklus! **Luuk Schipper**, ik wil je bedanken voor het onderlinge vertrouwen, de prettige afstemming op elkaar en je humor. Ik ben trots op wat we samen bereikt hebben. **Dr. V. P. Retèl, dr. G.W.J. Frederix, dr. S.N. Koole en I. Eekhout**, Valesca, Geert, Simone en Inge, dank voor alle energie die jullie gestoken hebben in de Health Technology Assessment van de WIDE studie. En **Vincent van der Noort**, bedankt voor de statistische ondersteuning.

Ik wil de collega's van de **Hartwig Medical Foundation** bedanken voor de fijne en intensieve samenwerking de afgelopen jaren. Het was inspirerend om samen te werken aan de missie om de zorg voor patiënten met kanker te verbeteren. **Hans van Snellenberg**, dank voor je scherpe visie, pragmatisme en oplossingsgericht denken. **Prof. dr. ir. E.P.J.G. Cuppen**, Edwin, dank voor je cruciale wetenschappelijke inzichten en technologische kennis. **Prof. dr. ir. J.J.M. van der Hoeven**, Koos, dank voor je jarenlange oncologische ervaring en kennis waar ik als jonge arts-onderzoeker uit

mocht putten. **Dr. P. Roepman**, Paul, dank voor het samen invullen van de database en je duidingen van de gevonden DNA-afwijkingen, ik heb er veel van geleerd.

Daarnaast ben ik de mensen in het AvL van de klinische studies **(Jan-Nico Ridderbos, Kelly van Deventer en Saphira van Diest)**, het Algemeen Klinisch Laboratorium **(Daan van den Broek)** en de overige analisten zeer erkentelijk voor het faciliteren van dit project door het verwerken van het weefsel en bloed. Ik wil in het bijzonder **Daoin Schout** bedanken voor de ontwikkeling van de 4FME methode. Tevens wil ik de medewerkers van de wetenschappelijke administratie **(Danny Baars, Tony van de Velde, Michelle de Haan en Lisette Al-van Wijck)** bedanken voor de ICT en database ondersteuning van dit project. Bovendien gaat mijn dank uit naar de laboratorium managers **(Immy Riethorst en Ewart de Bruijn**), bio-informatici en IT developers **(Lieke Schoenmaker, Nina Jacobs, Luan Nguyen, Peter Priestley, Charles Shale en dr. Alexandra J. van den Broek)** van de Hartwig Medical Foundation.

Studie-deelnemers

Grote dank gaat uit naar de **studie-deelnemers** voor deelname aan de WIDE studie en afgifte van bloed en weefsel. Zonder deze bijdrage was dit onderzoek niet mogelijk geweest.

Funding

Tevens wil ik **ZonMw** en **Hartwig Medical Foundation** bedanken voor het financieel mogelijk maken van de WIDE-studie.

Co-auteurs

Beste co-auteurs, dank voor de prettige samenwerking en jullie bijdragen aan de manuscripten van dit proefschrift.

Pathologen en KMPB-ers

In de eerste plaats wil ik de pathologen en KMPB-ers bedanken voor hun bijdrage aan de WIDE-studie door middel van tumor cel percentage beoordelingen, het slim inzetten van weefsel zodat er het maximale uit gehaald kon worden en de analyse, interpretatie en rapportage van de moleculaire diagnostiek. In de tweede plaats, minstens zo belangrijk, wil ik jullie, **Claudie Flohil, Elise Bekers, Liudmila Kodach, Petur Snaebjornsson, Maurits van Montfoort, Laura Smit, Hester van Boven, Hugo Horlings, Bart van de Wiel, Joyce Sanders, Emilie Groen, Mijke Bol, Jacqueline van der Wal, Mirjam Boelens, Frans Hogervorst en Tom van Wezel**, bedanken voor alle gezelligheid, fashion tips, juicy verhalen en jullie oprechte betrokkenheid bij mijn PhD traject. Jullie waren de reden dat ik altijd op de C3 rondhing!

*

Translational Gastrointestinal Oncology (TGO) groep

We hebben elkaar grotendeels virtueel moeten aanschouwen gedurende de afgelopen jaren vanwege COVID-19. Echter maakte dit de schaarse real life uitjes (kajakken, sushi eten, borrels bij Radion en picknicken) des te gezelliger. **Pien, Anne, Marianne en Margriet,** bedankt dat ik altijd met praktische vragen bij jullie terecht kon en bij de hulp bij het verspreiden van dit proefschrift. **Beatriz en Remond,** bedankt voor de inhoudelijke bijdrages. **Meike**, dankzij jou waren de financiën van de WIDE-studie op orde. **Lana, Noriko en Mariska,** dank voor jullie bio-informatische ondersteuning en het managen van de data, jullie waren mijn gids in het data doolhof. **Iris, Pieter, Willemijn, Soufyan, Carmen en Astrid,** dank voor de gezelligheid en het delen van PhD tips en tricks, nog veel succes met jullie boekjes! **Pauline,** bedankt voor je organisatorische hulp bij het afstemmen van overvolle agenda's tot de zaken rondom mijn promotie.

NKI PhD Council

Ik wil alle leden van de PhD council bedanken voor de leuke tijd. In het bijzonder wil ik **Chavelli, Mees, Robin, Renee, Sanket, Ziva en Iris** bedanken voor de prettige samenwerking als onderdeel van de mental health task force. Samen hebben we veel kunnen betekenen voor de mentale gezondheid van PhD studenten. Keep up the good work!

Nieuwe collega's

Chirurgen, opleiders (dr. **M. van Heijl** en dr. **A.H.W. Schiphorst**), arts-assistenten, PA-ers, verpleegkundigen en poliassistenten van de afdeling Heelkunde van het Diakonessenhuis, bedankt voor de zachte landing als ANIOS.

Paranimfen

Elsa Soons en Maxime Verkijk, bedankt dat jullie vandaag aan mijn zijde willen staan. Elsa, wij begonnen rond dezelfde tijd aan ons PhD traject en hebben alle diepte- en hoogtepunten samen meegemaakt. Hoewel de wereld stilstond tijdens de COVID-19 periode, werd er hard gewerkt aan onze manuscripten aan afwisselend jouw en mijn keukentafel. Mogelijk heeft het sociale isolement tijdens deze periode er juist aan bijgedragen dat we onze proefschriften succesvol hebben afgerond. Gelukkig wisten we inspanning ook goed af te wisselen met de nodige ontspanning. De keuze voor elkaar als paranimf was vanzelfsprekend. Desalniettemin ben ik trots dat het ons allebei is gelukt en dat ik naast jou mag staan tijdens jouw verdediging op 18 april 2023. Maxime, sinds onze studietijd zijn we bij elk belangrijk moment in elkaars leven aanwezig geweest. En dat waren nogal wat momenten; onze buluitreiking, mijn verhuizing naar jouw huis op de Vinkenburgstraat, jouw verhuizing, je toelating

als huisarts in opleiding, mijn start als PhD-student, je verloving (ik had je zelfs aan de telefoon op het moment suprême) en je zwangerschap en de komst van Florine. Het is dan ook niet meer logisch dat jij op dit belangrijke moment mijn paranimf bent. Ik verheug mij erop om op 17 juni 2023 ceremoniemeester van jou en Sieb te mogen zijn op jullie huwelijk.

Buiten het werk waren er vele mensen die altijd voor mij klaar stonden met een luisterend oor en welkome afleiding. Ook hen wil ik bedanken voor de hulp en ondersteuning.

Lieve **Janneke, Mauritia en Carlijn**, we kennen elkaar al sinds de basisschool. Onze vriendschap is vertrouwd, ongedwongen en wordt gekenmerkt door onze gezamenlijke humor. Ik krijg energie van jullie aanwezigheid, gekke verhalen en spontane acties. Het mooie is dat er na al die jaren weinig veranderd is. Bedankt voor het delen van al jullie lief en leed (het blijft een grote bron van vermaak) en jullie onvoorwaardelijke steun. Lieve **Anne**, al sinds de middelbare school ben je met je oprechte interesse, vrolijkheid en zorgzaamheid een geweldige vriendin. Ik verheug mij op alle mooie momenten die in onze (nabije) toekomst liggen!

Lieve **Floor en Merel**, ik leerde Floor kennen in de brugklas en kreeg Merel als tweelingzus cadeau. Jullie authenticiteit en creativiteit is inspirerend. Genetisch hetzelfde maar verschillend in vele opzichten. Met onze verschillende achtergronden (kunstmatige intelligentie, theater en medisch) ontstaan tijdens etentjes de meest interessante gesprekken en ontkomen we niet aan een vorm van theater (zang, dans of toneel).

Lieve **jaarclub**, inmiddels zijn we al 10 jaar verder en ik ben er trots op dat we nog steeds steevast 1 keer per week samen eten. Dank voor alle gezellige avonden, clubweekenden en vakanties met als klap op de vuurpijl de lustrumreis naar Guatemala & Belize. Ik kijk nu al uit naar onze volgende reis in 2024! **Suus, Dido en Doesjka**, het was bijzonder om deels gelijktijdig een PhD traject te volgen als clubgenootjes en elkaar hierin in te ondersteunen. Gedeelde smart is echt halve smart!

Lieve **oud-huisgenoten, Nick, Anouk en Manon**, de basis voor onze vriendschap werd gelegd op de Nobelstraat. Nick Justin, a hairdresser is truly a girl's best friend. Dank dat je altijd mijn haren weer weet te temmen. Het is bijzonder om te zien hoe je je dromen waarmaakt. Anouk, ook wij begonnen gelijktijdig aan ons PhD avontuur. We hebben eindeloos gespard over onze statistische tests, figuren en manuscripten met welkome afleiding in de vorm van sportieve activiteiten, borrels, huisfeestjes

en dagjes naar het strand. Ik kijk nu al uit naar jouw proefschrift en verdediging! Manon, bedankt voor de vrolijke noot die je bent, het komische element te midden van het serieuze materiaal, en dat je nog steeds niet gek wordt van al die medische gespreksonderwerpen.

Lieve **Maaike**, toen ik bij jou en Maxime in de werkgroep kwam in het derde jaar van de studie vonden jullie mij maar een nerd. Ik deed een aantal werkgroepen mijn best om niet in lachen uit te barsten maar uiteindelijk ging ik ook voor de bijl. Sindsdien hebben we veel gelachen en genoten van borrels, feestjes en vakanties. Dank voor je humor en joie de vivre!

Lieve **familie Verkijk**, ik wil jullie bedanken voor jullie gastvrijheid en betrokkenheid. Gedurende mijn semiartsstage woonde ik zelfs 3 maanden bij jullie in huis, een ontzettend leuke tijd. Om er vervolgens achter te komen dat we verre familie van elkaar zijn, wellicht is dat de reden dat het zo vertrouwd voelt bij jullie. **Kiki**, gedurende mijn PhD traject was jij bezig met de afronding van de studie geneeskunde. Je belde mij vaak om te overleggen of kwam gezellig langs. Je gedroeg je als klein zusje, altijd ontwapenend en eerlijk. Nu kom ik je vaak tegen in de kliniek en hoewel ik tracht mijn professionele houding te behouden, krijg ik altijd een dikke knuffel of een belletje tijdens mijn nachtdienst.

Lieve **familie Cohen-Vos**, een groot deel van mijn jeugd bracht ik door bij jullie als 'pleegdochter' en nog steeds voel ik jullie warmte en oprechte interesse waarvoor mijn dank. Lieve **Celeste**, inmiddels zijn we al 25 jaar hechte vriendinnen. We zaten op dezelfde basisschool, gingen samen naar de middelbare school en vertrokken samen naar Utrecht om te studeren. Nog steeds wonen we dicht bij elkaar en zijn we een constante factor in elkaars leven. De afgelopen jaren waren extra bijzonder vanwege jullie huwelijk waar ik ceremoniemeester mocht zijn, de komst van Hugo en op het moment dat dit proefschrift gedrukt wordt waarschijnlijk nog een klein meisje. Ik wil je bedanken voor je niet aflatende steun, warmte en loyaliteit.

Lieve ouders, **Herberth-Jan en Kerry**, ik wil jullie bedanken voor de vrijheid die jullie mij altijd hebben gegeven om mijn eigen keuzes te maken en dat jullie mij de wereld hebben laten zien. **Herberth**, je hebt mij geleerd in oplossingen te denken, hard te werken en geen concessies te doen. Met je verfrissende kijk op het leven maak je complexe situaties in één keer overzichtelijk. Ik waardeer je hulpvaardigheid en logistieke ondersteuning op vele fronten (onder andere de (H)uber). Je hebt van het leven genieten tot een kunst verheven en deelt dit graag. Dank hiervoor en voor je onnavolgbare provocerende humor. **Kerry**, dankjewel voor je onvoorwaardelijke

steun en voor alle zorg en liefde die je in mijn opvoeding hebt gestoken. Je open- en eerlijkheid maken dat ik alles tegen je kan zeggen. We delen een drang naar avontuur wat ons op vele onverwachte locaties heeft gebracht. Je onbevangenheid maakt dat we in de meest gekke situaties terecht komen. Bedankt voor deze bijzondere herinneringen en dat je er altijd voor zorgt dat ik er fashionable uit zie. Lieve **Tess**, met maar 15 maanden verschil zijn we sinds jongs af aan al zeer close. Toen we het huis uit gingen, besloten we samen op kamers te gaan. Inmiddels woon jij in Amsterdam. Desalniettemin ben je nog altijd dichtbij en zien we elkaar frequent. Zo zwaaiden we iedere dag naar elkaar wanneer onze tramlijnen elkaar kruisten 's ochtends als ik naar het AvL ging en jij naar je werk. Dankjewel dat je er altijd voor mij bent. Lieve **Mick**, dankjewel voor je humoristische kijk op het leven en je onbevangen vrolijkheid. Je maakt mijn leven een stuk grappiger en bent mijn grootste fan. Ik kan er altijd op rekenen dat je met of om mij lacht. Bovendien herinner jij mij er dagelijks aan dat het echte goede leven zich niet in Nederland maar in Barcelona afspeelt!

Curriculum vitae auctoris

Kris Samsom

Kris Samsom was born on the 15th of June 1993 in Utrecht, the Netherlands. In 2011, she graduated from the Christelijk Lyceum Zeist. After a gap year in Montpellier, France, she started to study Medicine at the University of Utrecht in 2012 and obtained her Master's degree cum laude in 2018. During her study, she got the opportunity to perform internships at the Ndlovu care group, South Africa, and at the Royal Melbourne Hospital, Australia. From 2015 to 2017, she also participated in an honours program which involved conducting research on neuroendocrine tumors (NETs), under supervision of prof. dr. M.R Vriens in collaboration with the Netherlands Cancer Institute (NKI). Afterwards, she continued her research, under supervision of prof dr. M.R. Vriens, dr. M.E.T. Tesselaar and dr. J.G. van den Berg, on the mutational landscape of small-intestinal NETs as a medical student at the NKI. In 2019, Kris started her PhD on the WIDE (*Whole genome sequencing Implementation in standard Diagnostics for Every cancer patient)* project at the pathology department of the NKI. This project, supervised by professor Gerrit Meijer (Pathology department, NKI) and professor Emile Voest (Medical Oncology department, NKI), with dr. Kim Monkhorst and dr. Linda Bosch as co-promotors (Pathology department, NKI), aimed to investigate the feasibility, clinical validity and clinical value of WGS-based diagnostics in routine clinical care. In immediate follow-up of this project, the department of pathology at the NKI, in collaboration with Hartwig, has implemented WGS in routine clinical care and a first provision for reimbursement of WGS in the Netherlands was established. During her PhD, Kris was a part of the mental health taskforce of the NKI PhD student council.

Currently, Kris lives in Utrecht and has started working as a resident not in training (ANIOS) at the Department of Surgery of the Diakonessenhuis Utrecht since September 2022.