numerous questions being asked by citizens and patient support groups: on information and consent, on the proper use of health data, on the anonymization of data vis-à-vis third parties, on the communication of how secondary discoveries and unwanted incidents inevitable when the entire genomes of patients and their family members are sequenced.

F. Lethimonnier* & Y. Levy
Department of Health Technology, INSERM, Paris, France
(*E-mail: franck.lethimonnier@inserm.fr)

Funding

None declared.

Disclosure

The authors have declared no conflicts of interest.

doi:10.1093/annonc/mdy027
Published online 19 January 2018

Introducing whole-genome sequencing into routine cancer care: the Genomics England 100 000 Genomes Project

Large-scale sequencing studies such as the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) have begun to catalogue the spectra of somatic mutations present in different solid tumour types [1, 2]. However, to date there has been minimal traction for solid tumours in alignment of large-scale sequencing data to longitudinal data on therapy and outcome. Such data are essential if we are to better target conventional cytotoxics, as well as emerging drugs such as immunotherapeutics. Stratification using molecular markers could improve benefit against cost and side-effects, especially critical in the adjuvant setting. Molecular heterogeneity of tumours and confounding patient factors mean datasets of daunting size and depth will be required. Arguably these will only be achieved through molecular analyses as standard for cancer patients entering clinical trials and, by adopting a population level approach to molecular analysis of patients undergoing routine cancer treatments.

The Genomics England 100 000 Genomes Project (100 000GP) cancer programme was initiated in 2012 to develop infrastructure for routine high throughput tumour sequencing [in particular whole-genome sequencing (WGS)] for NHS cancer patients to (i) establish a national research platform of molecular data with linkage to longitudinal clinical data and (ii) transform delivery of molecular testing in NHS clinical cancer care [3, 4] (Figure 1).

National coverage for the programme was critical, achieved through designation of thirteen established centres as 'NHS Genomics Medicine Centres' (GMCs), hub hospitals which each link out to up to 18 local recruiting hospitals. Tissue collection and preparation, DNA extraction and quantitation are undertaken locally, followed by transfer of DNA to the central national biorepository. WGS [of paired DNA, tumour (∼80×) and germline (∼40×)] is delivered by Illumina at the new national 100 000 Genomes Sequencing Centre in Hinxton, Cambridgeshire. Processed sequencing files (BAMs, FASTQs) are then passed back to Genomics England. Using standardised automated pipelines, Genomics England undertakes processing, calling, quality checking, prioritising (tiering), storing and presentation of the whole-genome analysis. Applying established knowledge-bases, the acquired somatic variants are tiered and annotated for potential diagnostic, predictive or prognostic 'actionability'. Results are presented back to the clinical users at GMCs via decision-support tools, which have been co-developed with commercial providers. A 'supplementary' analysis is also supplied, providing full analysis of tumour structural and copy number variation as well as other 'research' content including pan-genomic analyses of tumour signatures and mutational burden. Results are reviewed at multidisciplinary 'Tumour Sequencing Boards', which have been set up in federated networks across each NHS GMC. National standards for reporting of somatic cancer variants, technical validation and return of pertinent and secondary germline findings have been established through an NHS England-led national working group in 'Validation and Reporting'.

Researchers can access de-identified 100 000GP genomic and linked clinical data within the secure research data environment (Figure 2) via collaboration within a GeCIP (Genomics England Clinical Interpretation Partnership) domain. Although currently immature, these will develop into rich longitudinal datasets through lifelong linkage to National Cancer Registry (NCRAS) datasets including cancer registrations, subsequent cancer episodes, chemotherapy and radiotherapy (ENCORE, COSD, SACT and RTDS datasets), hospital episode statistics (HES) and other national datasets [5–8].

Procurement of tumour DNA of sufficient quantity, quality and purity has been a universal limitation in clinical and research tumour sequencing endeavours to date. Formalin fixation and paraffin embedding (FFPE) has been the clinical standard in tissue preparation for >100 years, with fresh tumour tissue being largely the preserve of tissue-banking and research projects. Following multisite pilot collection of FFPE samples, multivariate analysis of tissue handling metrics, controlled experimental work and national rollout to NHS GMCs of guidance for 'optimised' FFPE, from analysis of ensuing data, achieving FFPE-derived WGS nationally of sufficient and consistent quality was concluded to be infeasible.

To catalyse local pathways for collection of fresh tissue, 100 000GP eligibility was extended to include both (i) surgical resection and (ii) diagnostic biopsy from a wide range of early stage and advanced solid tumours [9]. Vacuum-packing, tissue refrigeration, novel coolants, diverse transport media and cytology sampling approaches have all been trialled within NHS GMCs under the 100 000GP. The Royal College of Pathologists, the Human Tissue Authority, the Health Research Authority, NHS England and Genomics England have come together in issuing a joint statement that collection of fresh tissue is a standard of care for cancer diagnostics, thus simplifying consenting for 100 000GP and molecular oncology studies more broadly [10].

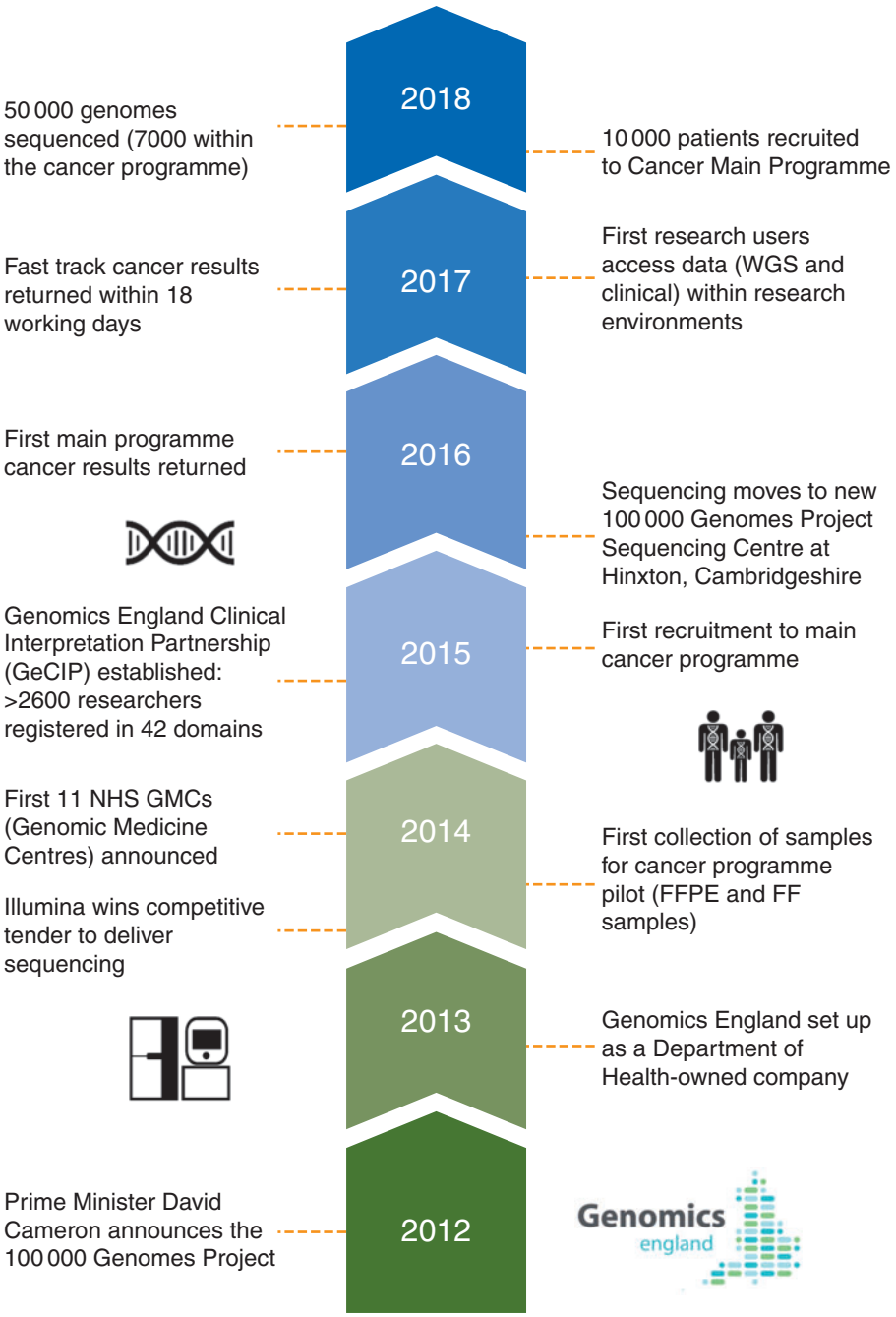

Figure 1. The Genomics England 100 000 Genomes Project Cancer Programme: a timeline

Aside from long term national transformation of tissue collection pathways, other key challenges during initial rollout of 100 000GP have included the high requirements for DNA, timescale of return of results and engagement of clinicians. Requirement for tumour-derived DNA has been reduced by >50%, easing inclusion of many tumour types. A 'fast-tracked' sample pipeline has been established through which turnaround time from receipt of DNA to issuing of a WGS analysis has now been reduced to <18 working days.

Nevertheless, in contrast to the parallel programme in undiagnosed rare disease, the perception across recruiting clinicians is unsurprisingly mixed around WGS being of value to individual patient management versus just being of relevance to research [11]. The analysed genome highlights *potentially* 'actionable' variants (typically within recognised cancer-associated genes) which indicate (i) established prognostic and therapeutic associations including eligibility to NICE-approved targeted drugs, (ii) potential eligibility to the UK trials (on-tumour or phase I) and (iii) off-tumour putative or established therapeutic associations. In practice, in most tumour types the *newly identified* and *truly* 'actionable' variants are limited in number on account of (i) previous standard of care testing, (ii) geographical variability in access to

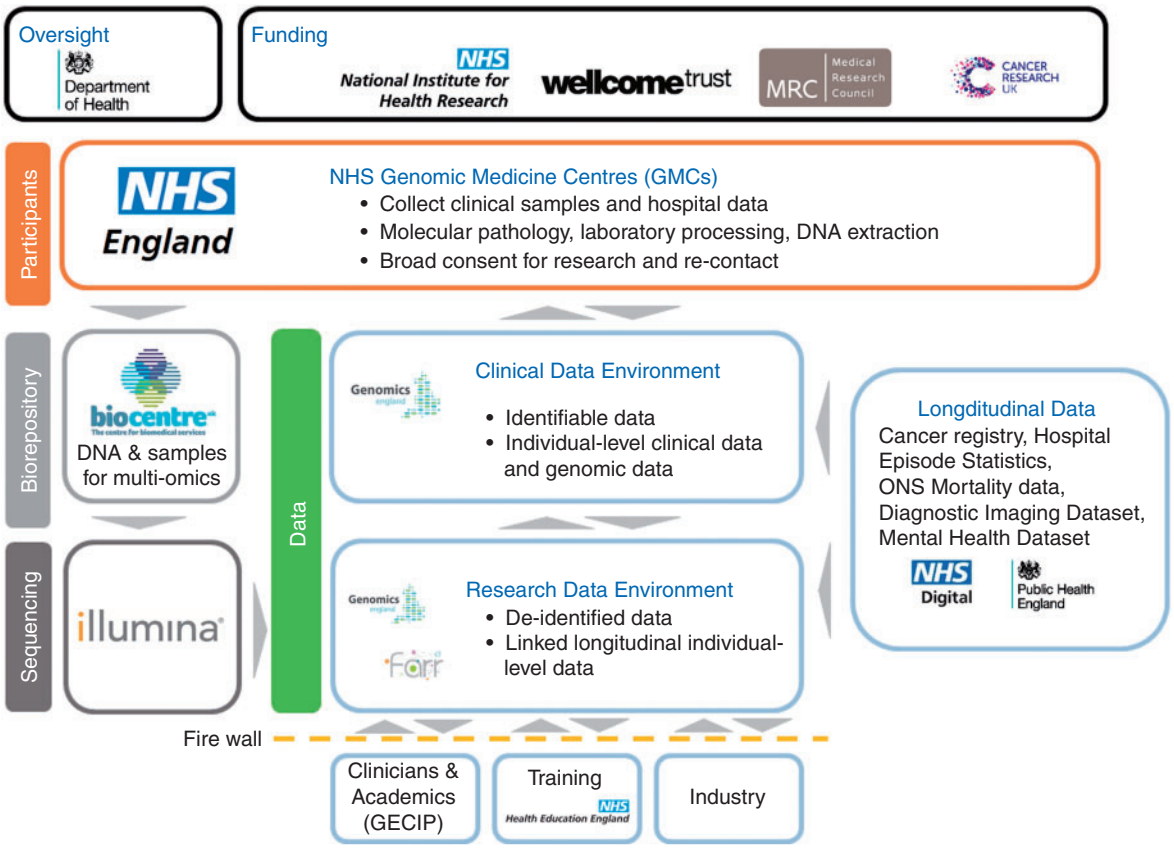

Figure 2. Flow of data and samples within the Genomics England 100 000 Genomes Project

clinical trials, (iii) highly restricted funding for targeted drugs used off-indication and (iv) patient's clinical status (absence of active disease post-surgery).

However, in certain groups, including sarcomas and subtypes of leukaemia, the recommended set of 'actionable' diagnostic, prognostic and therapeutic markers has become extensive and 'standard-of-care' testing has fallen behind [12–14]. For these tumour types, WGS is delivering the multiple emerging and established 'standard-of-care' structural and small variants in a single 'all-encompassing' test. Indeed, as cost of WGS falls, for these tumour types the economic tipping point towards WGS may be imminent, with opportunity to replace multiple platforms, complex workflows and perpetual redesign of panels. For the majority of solid tumours, this 'tipping-point' remains some way off, although it will likely be accelerated if there is emergence of therapeutic opportunities predicated upon biomarkers only tractable from analysis of the whole genome (e.g. 'pan-genomic' mutational signatures, patterns of novel structural variants) [15].

In recognition by the UK government and NHS leadership of this dynamic landscape, there has been commitment within the Government's Industrial Strategy Life Sciences Sector Deal towards ongoing non-NHS-commissioned tumour WGS beyond completion of 100 000GP, with particular focus on alignment to trials and collaboration with industry [16].

From late 2018, delivery of all NHS clinical sequencing will be through a new NHS Genomic Medicine Service comprising (i) a national centralised end-to-end ISO-accredited infrastructure for clinical WGS developed by Genomics England and (ii) a consolidated network of NHS England Genomic Laboratory Regional Genetics Hubs, re-procurement of which is underway [17]. A national testing directory is being established to advance consistency of testing indications and technologies. In the initial phase, the main focus for NHS central commissioned WGS will be in rare disease; in the majority of tumour types panel testing (comprising specified marker sets) will be indicated. However, as the number/complexity of 'actionable' markers increases, there is expectation of 'tipping' successive tumour types over to WGS. Underpinning emergence of a truly modernised UK genomics service will be not just the sequencing and informatics infrastructure but the local transformation, expansion in skills and new professional networks that have evolved through delivery of the 100 000 Genomes project.

C. Turnbull
Professor of Genomic Medicine, Queen Mary University of London; Consultant in Clinical Genetics, Guy's and St Thomas' NHS Trust, UK
(E-mail: clare.turnbull@genomicsengland.co.uk)

Acknowledgements

The author thanks Mark Caulfield, Angela Hamblin and Nirupa Murugaesu for their review of this editorial and to Adam Toms and Lisa Dinh for generation of figures.

Funding

None declared.

Disclosure

The author has declared no conflicts of interest.

References

1. Hudson TJ, Anderson W, Artez A et al. International network of cancer genome projects. Nature 2010; 464(7291): 993–998.
2. Weinstein JN, Collisson EA, Mills GB et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013; 45(10): 1113–1120.
3. Improving Outcomes: A Strategy for Cancer, Department of Health, 2011.
4. Achieving World-class Cancer Outcomes: A Strategy for England 2015–2020: Independent Cancer Taskforce, 2015.
5. Cancer Outcomes and Services Dataset (COSD): Public Health England, 2018.
6. National Radiotherapy Dataset (RTDS): Public Health England, 2018.
7. Systemic Anti-Cancer Therapy Dataset (SACT): Public Health England, 2018.
8. Hospital Episode Statistics: NHS Digital, 2018.
9. The 100, 000 Genomes Project Protocol: Genomics England, 2015.
10. Consensus Statement Diagnostic Pathways for NHS Cancer Genomic Sampling and Analysis, 2017.
11. Hill S. Introducing genomics into cancer care. Br J Surg 2018; 105(2): e14–e15.
12. Arber DA, Orazi A, Hasserjian R et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016; 127(20): 2391–2405.
13. Papaemmanuil E, Gerstung M, Bullinger L et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016; 374(23): 2209–2221.
14. Doyle LA. Sarcoma classification: an update based on the 2013 World Health Organization Classification of Tumors of Soft Tissue and Bone. Cancer 2014; 120(12): 1763–1774.
15. Davies H, Glodzik D, Morganella S et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med 2017; 23(4): 517–525.
16. Bell J, Life Sciences Industrial Strategy – A report to the Government from the Life Sciences Sector, 2017.
17. Creating a Genomic Medicine Service to Lay the Foundations to Deliver Personalised Interventions and Treatments. NHS England Board, 2017.

doi:10.1093/annonc/mdy054
Published online 15 February 2018

ESR1 and endocrine therapy resistance: more than just mutations

Estrogen receptor (ER)-positive breast cancer accounts for 70%–80% of all diagnosed breast cancers [1]. The adoption of endocrine therapies, including ER modulators/degraders (SERMs/SERDs), which antagonize ER, and aromatase inhibitors (AIs), which suppress estrogen synthesis, as the mainstay of treatment of ER-positive breast cancer patients has resulted in substantial survival benefit for patients with early stage disease [2]. Treating ER-positive metastatic breast cancer (MBC), however, remains a significant clinical challenge, due to the development of secondary resistance to all modalities of endocrine therapy [3]. Recently, studies have identified recurrent somatic mutations within the ligand-binding domain (LBD) of *ESR1* (encoding ER) in >30% of ER-positive MBC [4–8]. These mutations alter the conformation of ER and produce a constitutively active form of the protein. Mutations at residues 536–538, in particular, promote ER activity in the absence of ligand, resulting in resistance to AIs and reduced sensitivity to SERMs/SERDs [4, 5]. *ESR1* fusion genes have also been reported in ER-positive MBCs; however, a detailed description of their manifestations and clinical prevalence is lacking [9]. In this issue of *Annals of Oncology*, Hartmaier et al. reported the identification of recurrent hyperactive *ESR1* fusion genes in breast cancers resistant to endocrine therapy [10], adding to the diversity of reported *ESR1* alterations.

Hartmaier et al. carried out a retrospective study to discover genomic rearrangements involved in the acquired resistance to ER-targeted therapies [10]. Using mate-pair DNA sequencing and/or RNA sequencing of matched primary-metastasis-normal samples from 6 patients, the authors identified an *ESR1-DAB2* in-frame fusion transcript that fused exons 1–6 of *ESR1* to exons 3–15 of *DAB2*. This fusion was found only in the lymph node metastasis but not in the primary. RNA sequencing analysis of an additional 51 breast cancer metastases revealed an *ESR1-GYG1* fusion gene in a bone metastasis, comprising the *ESR1* exons 1–6, the same involved in the *ESR1-DAB2* fusion gene, and the 3′ end of *GYG1*. Importantly, both fusions were also detectable at the protein level [10].

Prompted by the discovery of recurrent *ESR1* fusion breakpoints, the authors analyzed 9542 breast cancers (including 5216 from metastases) and 254 circulating tumor DNA (ctDNA) samples from advanced breast cancer patients, and identified 7 additional *ESR1* fusion genes. Including the initial cohorts subjected to mate-pair and/or RNA sequencing, 5 fusions were identified in metastatic disease (5/5, 272, 0.09%), 1 in local recurrence after endocrine therapy (1/4, 329, 0.02% of primary tumors) and 3 in ctDNA (3/254, 1.2%). For the 4/9 patients with available clinical histories, all had been treated extensively with AIs. Of note, the *ESR1* breakpoints were all in or between exons 6 and 7, disrupting the LBD of *ESR1*. *In vitro* analysis of 3 of the fusions identified (*ESR1-DAB2*, *ESR1-GYG1*, and *ESR1-SOX9*) demonstrated that all had ligand-independent activity and two were hyperactive [10]. These observations suggest a potential role for the distinct 3′ gene partners in determining resultant ER activity.

With genomic breakpoints frequently located in intronic regions, capture-based targeted sequencing of exons does not always detect fusion genes. Structural rearrangements, however, are frequently associated with copy number alterations. Based on this notion, the authors devised a novel algorithm *copyshift* to detect intra-genic fusion junctions associated with copy number changes in targeted sequencing. When tested in a cohort of lung cancer with known *ALK* rearrangements, the authors showed that *copyshift* was specific (>89% positive predictive value), albeit with limited sensitivity (~85% false negative rate). Applying *copyshift* to the cohort of 9542 breast cancers to interrogate the recurrent *ESR1* breakpoint region, the authors found 83 *copyshift*-positive tumors. These tumors were enriched for ER-positive and