

TITLE: EXPLORING THE BIOLOGICAL PROCESSES UNDERLYING MUTATIONAL SIGNATURES IDENTIFIED IN CANCER

PRINCIPLE INVESTIGATOR: DR SERENA NIK-ZAINAL MA, MBBCHIR, MRCP, PHD

SITE: WELLCOME TRUST SANGER INSTITUTE

ABSTRACT

Cancer is the ultimate genetic disease. Cancer is driven by genetic changes, or mutations, acquired throughout life. These mutations provide clues regarding the genetic damage that occurred through the lifetime of the patient and the development of the cancer. Recently, improvements in technologies that read the genetic code have made it possible for the entire genetic code of a human being to be "sequenced", and therefore for every mutation to be detected. Using this sequencing technology, scientists have shown that cigarette-smoke and ultraviolet light damage for example, can cause direct harm to the DNA in our cells, causing specific mutation patterns which we call *Mutational Signatures*, making it more likely to develop into particular cancers such as lung cancers or malignant melanomas respectively.

More recently, my colleagues and I searched for every mutation in the DNA of breast cancers from 21 women. Using computer-based tools we showed that many patterns of mutational signatures exist in breast cancers which were not appreciable before. However, our knowledge and understanding of these mutational signatures is limited. In this project, we aim to understand how and why the mutation patterns are generated and whether these patterns occur because of internal or externally arising DNA damage or because the essential DNA repair toolkit in the cell is awry.

Although it is possible to study mutational signatures in animal model systems, there are well-acknowledged differences between repair pathways in model organisms and human beings. In this project, we hope to recruit patients who are born with naturally-occurring genetic defects in genes involved in the essential DNA repair toolkit. These extremely rare conditions have recognizable clinical syndromes characterised by symptoms such as early-onset cancer, accelerated aging, growth abnormalities, birth defects and neurological abnormalities including learning disorders and/or neurodegeneration. Exploring mutational signatures in these patients will provide a very valuable framework of the mutational signatures that can arise in human cells when repair pathways go awry in humans. Furthermore, patients who have suffered from exposure to known environmental chemicals that can cause cancer will also provide a unique insight into the accumulation of genetic damage.

BACKGROUND OR RATIONALE OF THE PROJECT

1. A copy of the human genome is present in almost every cell in the human body

At the time of conception, we start life as a single cell, containing genetic material that we have inherited from our parents. This cell must divide many thousands of times to form a complex human baby. Each cell in the human body must therefore contain a copy of that human genome that each of us was originally conceived with.

2. Errors arise in DNA relatively frequently but are repaired by the DNA repair toolkit

The human genome is made up of building blocks called DNA. There are 6,000,000,000 building blocks in human DNA which has to be copied many thousands of times. The vast size of the human genome means that errors arise during the process of cell division. The DNA in human cells also comes under attack from a variety of everyday exposures whether these are external, environmental DNA damaging agents or internal DNA damaging chemicals. The human cell however, contains an essential DNA repair toolkit which involves many genes and is able to correct most of the errors that arise in DNA.

3. Errors that persist as mutations contain insights into past mutagenic activity

Sometimes, DNA errors persist because these are simply missed by the repair toolkit, the repair pathways have gone awry or there is simply an overwhelming amount of DNA damage. These errors that persist in human cells are called *somatic mutations* and typically, characterize cancer cells (1). By studying these somatic mutations, we can understand the nature of the DNA damaging agents and inefficiencies of repair pathways that have been operative during the development of a cancer or of any disease where there is damage occurring to DNA in cells, such as aging and neurodegeneration.

4. Mutational signatures are the imprints left behind by the mutagenic processes that have been active in cancers

The mutagenic processes that have generated somatic mutations may be exogenous in origin, for example, ultraviolet radiation damage or the by-products of tobacco-smoke. Alternatively, they may be injurious endogenous enzymes or reactive metabolites. Whatever the nature of the mutagenic damage, the effects of these mutagenic processes have been mitigated by the cellular DNA repair mechanisms or further contributed to if the latter have become defective. Each of these processes will leave an imprint, or *mutational signature*, on the cellular genome, characterized by distinct mutation patterns (2). Our knowledge and understanding of these signatures and their underlying mutational processes is still remarkably limited.

5. Next-generation sequencing technologies allow the inspection and analysis of thousands of mutations from whole genomes

The completion of the human genome sequence has allowed the development of systematic re-sequencing of human genomes to identify somatic mutations. Historic limitations in technology restricted early studies to (PCR-based) sequencing of exons of protein-coding genes. The recent advent of second-generation sequencing technology (3) has, however, permitted large-scale sequencing of whole genomes for identification of all classes of somatic mutation. Many previous studies have focused on specific single gene discovery of individual diseases whether in the germline or in cancer (4, 5). Now, detailed analyses of the

entire catalogue of somatic mutations in whole genomes (2, 6, 7) have laid the foundations for how genome-wide mutational signatures of mutagenic insults can be appreciated.

6. Multiple mutational signatures have been identified using this principle

In the course of my doctoral studies, I sequenced the whole genomes of twenty-one primary breast cancers, and complete catalogues of somatic mutation were archived from each. Using several mathematical approaches, multiple known and novel *mutational signatures* were extracted. At least five distinct signatures were identified. For example, a characteristic combination of mutational signatures distinguished breast cancers from women with germline abnormalities in *BRCA1* and *BRCA2*, proteins which are involved in DNA repair by homologous recombination, from other sporadic breast cancers (2, 8). These novel findings demonstrate how the power of second-generation sequencing technology can be harnessed to generate profound and novel insights into tumour development.

7. The biological basis underlying novel mutational signatures is unknown and warrants investigation

The biological mechanisms underlying these novel mutational signatures are, however, in large part unknown. In this study, I aim to explore the biological basis of these and other mutational signatures that have emerged and will continue to emerge from sequencing cancer genomes. My overall strategy will be to apply large-scale sequencing technologies in cells derived from individuals with known inherited DNA repair defects to characterize the resulting mutational signatures, or individuals with a known history of an excess of environmental exposure to mutagenic insults. Ultimately, we would aim to compare these experimentally generated signatures to cryptic signatures extracted from cancer genomes in order to understand the signatures that arise in cancers.

AIMS

Broadly, this project seeks to explore the biological reasons underlying the mutational signatures that have emerged from sequencing cancer genomes. The overall aims are to:

1. Identify people who have naturally occurring defects in genes involved in DNA repair, DNA copying mechanisms (replication) or any genes involved in maintaining the integrity of the human genome.
2. Collect tissue samples in order to establish a resource of induced pluripotent stem cells (iPSC) or EBV-transformed lymphoblastoid cell lines (LCLs) derived from these patients
3. Experiment with these cell lines, including cell populations isolated during iPSC or LCLs derivation and explore the ensuing mutational signatures present in these cells in order to understand the genesis of mutational signatures and dependencies on intact repair/replication pathways in humans

EXPERIMENTAL DESIGN

- 1. Identify people who have naturally occurring defects in genes involved in DNA repair, DNA copying mechanisms (replication) or any genes involved in maintaining the integrity of the human genome.**

Patients who have naturally occurring inherited genetic defects in DNA repair/replication pathways can have a multitude of clinical symptoms including early-onset cancers, progeria, congenital malformations and neurodisability/neurodegeneration. Drawing on my Clinical Genetics background and in collaboration with other Clinical Genetics colleagues, I hope to recruit patients with these clinical syndromes. These are extremely rare disorders and Clinical Geneticists would usually be involved in the care and management of such patients. They would also be accustomed to recruiting patients/their families into studies of rare disorders. A set of people who are healthy volunteers will also be identified to act as normal controls for this study. In addition to recruiting patients with a history of strong exposure to DNA damaging agents directly through clinical genetics services, I hope to recruit these research participants by contacting the membership secretaries of relevant charities and societies.

A clinician would highlight a potential patient for recruitment and the Principle Investigator (PI) would supply the clinician/patient with consent forms allowing time for consideration of participation. The recruiting clinician will take consent, file the documentation in the patient's notes and send the PI a copy for storage. The PI will assign a linked-anonymised ID to the patient and inform the Chief Investigator (CI). The CI will provide another level of anonymisation and will send anonymised barcoded sample tubes for sample collection by the clinician (Figure 1).

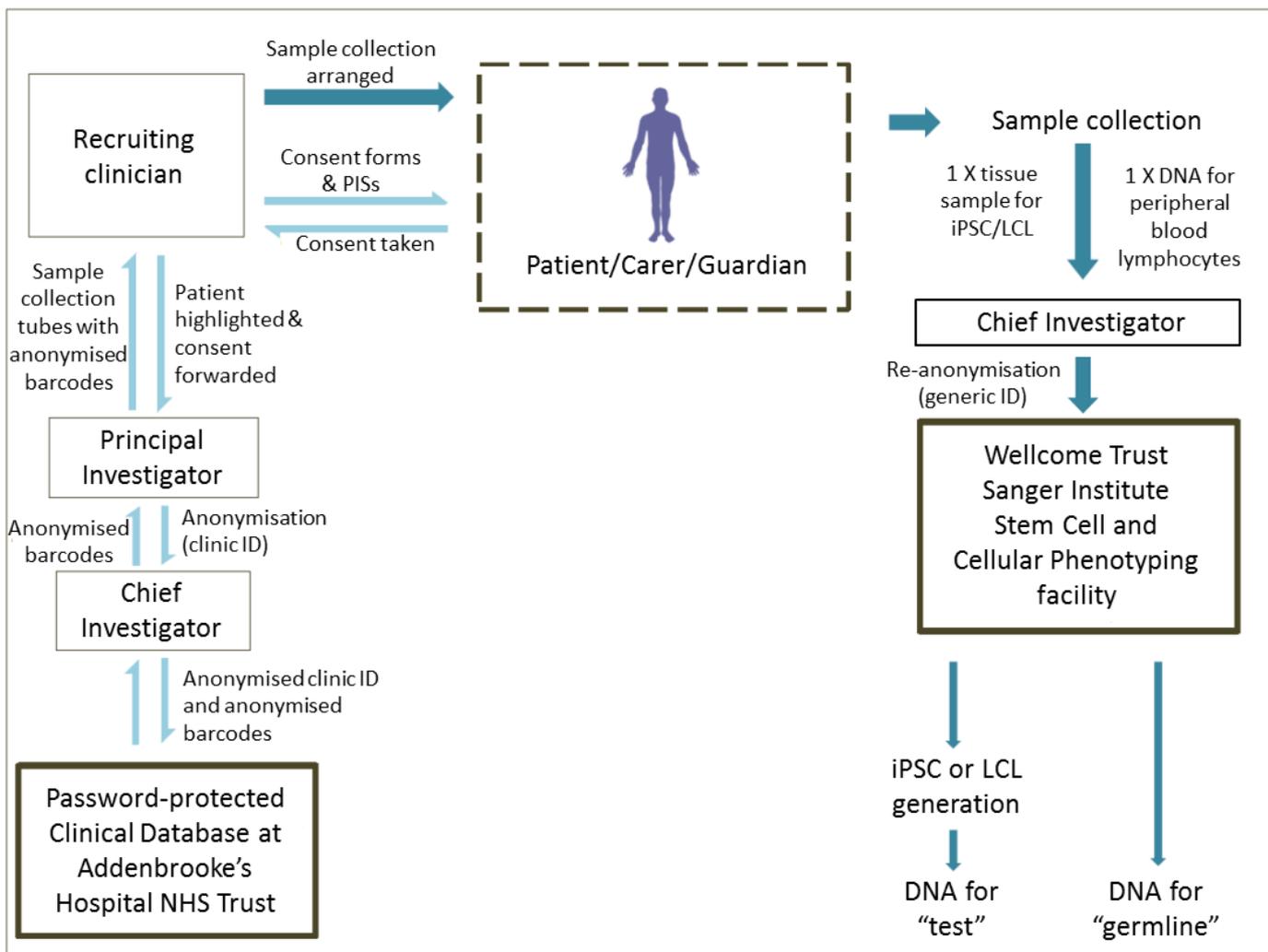


Figure 1: Workflow of patient recruitment and sample collection. Recruiting clinicians will highlight patients who are suitable for recruitment, consent the patients and arrange sample collections. People with a history of strong exposure to DNA damaging agents may be recruited through charities/societies. The Principal Investigator will anonymise each patient, inform the Chief Investigator who will store linked-anonymised information within a password-protected Clinical Database at Addenbrookes's Hospital NHS Trust. Double-anonymised samples will be sent by courier to the Wellcome Trust Sanger Institute Stem Cell and Cellular Phenotyping facility. The Chief Investigator will further anonymise the samples prior to iPSCs (induced pluripotent stem cells) and/or LCLs (lymphoblastoid cell lines) being generated. DNA will be extracted from the peripheral blood lymphocytes as "germline" DNA and DNA extracted from the iPSCs/LCLs will be "test" DNA. PIS = Patient Information Sheet.

2. Collect tissue samples in order to establish a resource of induced pluripotent stem cell systems (iPSC) or EBV-transformed lymphoblastoid cell lines (LCLs)

The expected date of sample collection will be agreed. Samples will be sent to the Sample Management Facility at the Wellcome Trust Sanger Institute by courier, and will be registered and processed as per sample management protocol.

A DNA sample extracted from peripheral blood lymphocytes ("germline" DNA) will be stored. A DNA sample from cell populations generated during iPS cell derivation and/or from iPS cells generated from the patient will also be extracted ("test" DNA) (Figure 1). Samples will be whole genome sequenced and aligned back to the Human Reference Genome. All germline changes will be subtracted away from the dataset to generate a catalogue of acquired or *somatic* mutations in human iPS cells from each patient. Therefore, germline or inherited changes will not become a part of the standard dataset for somatic analysis by researchers at the

Sanger Institute (Figure 2). For a small subset of patients, a *clinical* diagnosis of a rare DNA repair defect may be suspected but not confirmed by molecular genetic testing. In these individuals, an analysis of germline mutations will be offered in order to assist in confirmation of the diagnosis for the patient and their families. It will be stressed that this analysis will be entirely research-based and the responsibility of confirmation/validation of the result will have to be carried out by the collaborating clinician in a certified clinical laboratory.

3. Experiment with these cells and explore the ensuing mutational signatures present in these lines in order to understand the genesis of mutational signatures and dependencies on intact repair/replication pathways in humans

LCLs and iPSCs will be established from samples collected from patients mentioned above within the Stem Cell and Cellular Phenotyping programme at the Wellcome Trust Sanger Institute. These anonymised cells will be available indefinitely for further experimentation and analysis in years to come.

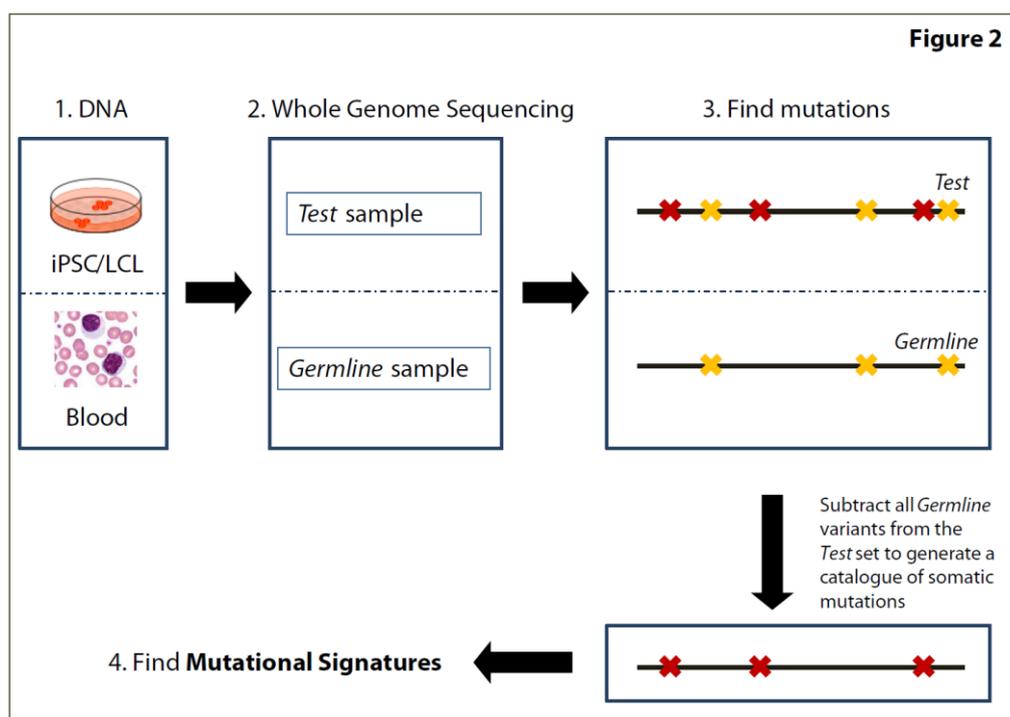


Figure 2: Data processing to generate somatic mutations for Mutational Signatures analysis. DNA will be extracted from the iPSCs/LCLs (*test*) as well as peripheral blood lymphocytes (*germline*). Both samples will be sent for whole genome sequencing. Both will be aligned back to the reference human genome and variants/mutations called in each sample separately. *Germline* variants will be subtracted to reveal the catalogue of somatic mutations for Mutational Signatures analysis.

ETHICAL CONSIDERATIONS

1. **Consent:** Consent will be required from all patients recruited to this study. Syndromes associated with disorders in genes involved in DNA repair/replication may frequently involve children or adults with learning/neurodegenerative disease. Consent may have to be taken from guardians of these individuals who need to be as fully informed regarding this study as possible. I have drafted information sheets, consent and assent forms for research participants with capacity, children and carers. Experienced staff in clinical genetics centres will be best placed to judge research participant capacity and obtain consent. We will need to ensure that patients recruited by clinicians from outside the UK have been correctly

consented. The Wellcome Trust Sanger Institute (WTSI) has an established Human Materials Data Management Committee (HMDMC), which oversees project compliance with all human tissue regulatory and ethical policy. To ensure that consent has been given for all samples from outside the UK that we wish to use, we will seek signed assurances that these are in place from the sample provider (sample custodian) through use of the WTSI HMDMC Sample Registration Form. Signed assurances from the sample custodian will include assurances that the samples have been collected in accordance with the Declaration of Helsinki, that the samples may be imported and they may be used in our study. An example sample registration form has been sent with this application.

2. **Anonymity of samples:** Tissue samples used in this study will require anonymisation in order to protect patient identity. Sample providers at clinical genetics centres will have anonymised the patients and then will be assigned another anonymised ID by the Chief Investigator (CI). Sample providers will be issued with barcoded tubes prior to sending these samples to the Wellcome Trust Sanger Institute. Once at the Institute, these samples will be handled and managed according to this anonymised ID only. The key linking the initial/original code to participant identifiable data will be held by site-specific PIs. Research participants will not be identifiable to researchers at the Wellcome Trust Sanger Institute.
3. **Whole genome sequencing:** In theory, genome sequencing of the “germline” genome could reveal incidental findings which are genetic changes unrelated to the disease being studied, but which may have health implications for the research participants in the future. However, the nature of the processing of data in this study (Figure 2) means that the germline genetic changes which are present in the patient (the inherited genetic variation present from birth) is not detected because it is subtracted away from the dataset. In this study, we are only interested in the new or acquired genetic changes called *somatic mutations*, which have arisen in the patient’s cells which do not comprise inherited genetic changes. However, in a small subset of patients where a DNA repair defect is clinically suspected but not confirmed, at the request of the patient and recruiting clinician germline mutations in DNA repair genes can be sought. Only germline mutations which are relevant to the disease being studied (i.e. provides confirmation of the clinical diagnosis) will be fed back to the recruiting clinician.
4. **Anonymised DNA sequence data** from the study will be stored indefinitely in the European Genome-Phenome Archive (EGA) which is an electronic archive at the European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge and made accessible via managed access to bona fide researchers worldwide. Researchers may access the data once they have signed a Data Access Agreement which specifies that research participants must not be identified. There is an extremely small risk that an individual's unique DNA sequence data may be matched to sequence in a data archive which also lists that individual's personal, identifiable data. There is a very small risk of identification of a participant if this happens, but participants will have this risk explained to them prior to consent.
5. A **secure clinical database** will be required to maintain all the linked anonymised data which will be password protected and stored on password-protected/encrypted devices and only accessible by the CI and a research assistant.
6. Clinical teams in Clinical Genetics Services will undertake research participant recruitment. There must be **no risk of coercion during recruitment**.

BENEFITS OF THE STUDY

Mutational signatures generated via experimental manipulation of iPSC/LCL lines of humans with defects in genes associated with DNA repair and replication will be catalogued in a database of experimentally-generated mutational signatures. Ultimately, mutational signatures extracted from the vast numbers of impending cancer genomes will be compared to those obtained from experimental data in order to gain biological insights into the mutational processes driving the development of cancers.

Understanding why mutation patterns arise in humans (what DNA damage and DNA repair pathways are largely responsible), may lead to learning how to interfere with them and ultimately, provide us with some way of effectively halting or preventing the disease processes that result in cancer, premature aging and neurodevelopmental problems in these patients and others in the population. It is envisaged that this resource of iPSC lines derived from patients with germline defects in DNA repair/replication pathways will be valuable for experimental manipulation in the future.

RESOURCES AND COSTS

Infrastructure and facilities

The manipulation and analysis of large whole genome datasets requires access to high-level computational facilities and expertise. The large-scale iPSC/LCL generation, sequencing efforts and data generated from this scientific proposal necessitate the sequencing, computational power and management to which the Wellcome Trust Sanger Institute (WTSI) is accustomed. Similar data generation, storage and analysis have been initiated in the sequencing and analysis of the twenty-one primary breast cancer genomes leading up to this proposal. Therefore, the current proposal will be building on that experience and expertise.

Personal funding and consumables

Personal funding for the Principal Investigator, Dr Serena Nik-Zainal and a research assistant has been secured via the Wellcome Trust Intermediate Clinical Fellows funding scheme. Funding for consumables include costs for iPS cell lines / LCL generation, library-making and sequencing, compute, storage, backup, characterisation of cells and experimental manipulation of these cells.

A summary of financial costs is provided below:

	TOTAL COST
(a) Salaries	£540,336.42
(b) Materials and consumables	£334,504.00
(c) Animals	0
(d) Equipment	0
(e) Miscellaneous	£5,000.00

(f) Work abroad	0
GRAND TOTAL	£879,840.42

REFERENCES

1. Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009) The cancer genome, *Nature* 458, 719-724.
2. Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L. A., Menzies, A., Martin, S., Leung, K., Chen, L., Leroy, C., Ramakrishna, M., Rance, R., Lau, K. W., Mudie, L. J., Varela, I., McBride, D. J., Bignell, G. R., Cooke, S. L., Shlien, A., Gamble, J., Whitmore, I., Maddison, M., Tarpey, P. S., Davies, H. R., Papaemmanuil, E., Stephens, P. J., McLaren, S., Butler, A. P., Teague, J. W., Jonsson, G., Garber, J. E., Silver, D., Miron, P., Fatima, A., Boyault, S., Langerod, A., Tutt, A., Martens, J. W., Aparicio, S. A., Borg, A., Salomon, A. V., Thomas, G., Borresen-Dale, A. L., Richardson, A. L., Neuberger, M. S., Futreal, P. A., Campbell, P. J., and Stratton, M. R. (2012) Mutational Processes Molding the Genomes of 21 Breast Cancers, *Cell* 149, 979-993.
3. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E. C. M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., Hoschler, K., Hurwitz, S., Ivanov, D. V., Johnson, M. Q., James, T., Huw Jones, T. A., Kang, G. D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ling Ng, B., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Chris Pinkard, D., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Chiva Rodriguez, A., Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Ernest Sohna Sohna, J., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele, S., Verhovskiy, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Yan, J., Yau, L., Zuerlein, M., Mullikin, J. C., Hurler, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R., and Smith, A. J. (2008) Accurate whole human genome sequencing using reversible terminator chemistry, *Nature* 456, 53-59.
4. Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A., and Shendure, J. (2009) Targeted capture and massively parallel sequencing of 12 human exomes, *Nature* 461, 272-276.
5. Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B. A., Cooper, C., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G. J., Bigner, D. D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J. W., Leung, S. Y., Yuen, S. T., Weber, B. L., Seigler, H. F., Darrow, T. L., Paterson, H., Marais, R., Marshall, C. J., Wooster, R., Stratton, M. R., and Futreal, P. A. (2002) Mutations of the BRAF gene in human cancer, *Nature* 417, 949-954.

6. Pleasance, E. D., Cheetham, R. K., Stephens, P. J., McBride, D. J., Humphray, S. J., Greenman, C. D., Varela, I., Lin, M. L., Ordonez, G. R., Bignell, G. R., Ye, K., Alipaz, J., Bauer, M. J., Beare, D., Butler, A., Carter, R. J., Chen, L., Cox, A. J., Edkins, S., Kokko-Gonzales, P. I., Gormley, N. A., Grocock, R. J., Haudenschild, C. D., Hims, M. M., James, T., Jia, M., Kingsbury, Z., Leroy, C., Marshall, J., Menzies, A., Mudie, L. J., Ning, Z., Royce, T., Schulz-Trieglaff, O. B., Spiridou, A., Stebbings, L. A., Szajkowski, L., Teague, J., Williamson, D., Chin, L., Ross, M. T., Campbell, P. J., Bentley, D. R., Futreal, P. A., and Stratton, M. R. (2010) A comprehensive catalogue of somatic mutations from a human cancer genome, *Nature* **463**, 191-196.
7. Pleasance, E. D., Stephens, P. J., O'Meara, S., McBride, D. J., Meynert, A., Jones, D., Lin, M. L., Beare, D., Lau, K. W., Greenman, C., Varela, I., Nik-Zainal, S., Davies, H. R., Ordonez, G. R., Mudie, L. J., Latimer, C., Edkins, S., Stebbings, L., Chen, L., Jia, M., Leroy, C., Marshall, J., Menzies, A., Butler, A., Teague, J. W., Mangion, J., Sun, Y. A., McLaughlin, S. F., Peckham, H. E., Tsung, E. F., Costa, G. L., Lee, C. C., Minna, J. D., Gazdar, A., Birney, E., Rhodes, M. D., McKernan, K. J., Stratton, M. R., Futreal, P. A., and Campbell, P. J. (2010) A small-cell lung cancer genome with complex signatures of tobacco exposure, *Nature* **463**, 184-190.
8. Nik-Zainal, S., Van Loo, P., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., Shlien, A., Cooke, S. L., Hinton, J., Menzies, A., Stebbings, L. A., Leroy, C., Jia, M., Rance, R., Mudie, L. J., Gamble, S. J., Stephens, P. J., McLaren, S., Tarpey, P. S., Papaemmanuil, E., Davies, H. R., Varela, I., McBride, D. J., Bignell, G. R., Leung, K., Butler, A. P., Teague, J. W., Martin, S., Jonsson, G., Mariani, O., Boyault, S., Miron, P., Fatima, A., Langerod, A., Aparicio, S. A., Tutt, A., Sieuwerts, A. M., Borg, A., Thomas, G., Salomon, A. V., Richardson, A. L., Borresen-Dale, A. L., Futreal, P. A., Stratton, M. R., and Campbell, P. J. (2012) The life history of 21 breast cancers, *Cell* **149**, 994-1007.