

BIBLIOGRAPHY

- Abnizova, I., Leonard, S., Skelly, T., Brown, A., Jackson, D., Gourtovaia, M., Qi, G., Te Boekhorst, R., Faruque, N., Lewis, K., *et al.* (2012). Analysis of context-dependent errors for illumina sequencing. *J Bioinform Comput Biol* *10*, 1241005.
- Aboussekhra, A., Biggerstaff, M., Shivji, M.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protic, M., Hubscher, U., Egly, J.M., and Wood, R.D. (1995). Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* *80*, 859-868.
- Al-Kuraya, K., Schraml, P., Torhorst, J., Tapia, C., Zaharieva, B., Novotny, H., Spichtin, H., Maurer, R., Mirlacher, M., Kochli, O., *et al.* (2004). Prognostic relevance of gene amplifications and coamplifications in breast cancer. *Cancer research* *64*, 8534-8540.
- Antoniou, A., Pharoah, P.D., Narod, S., Risch, H.A., Eyfjord, J.E., Hopper, J.L., Loman, N., Olsson, H., Johannsson, O., Borg, A., *et al.* (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet* *72*, 1117-1130.
- Armitage, P., and Doll, R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* *8*, 1-12.
- Avery, O.T., Macleod, C.M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine* *79*, 137-158.
- Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., Lawrence, M.S., Sivachenko, A.Y., Sougnez, C., Zou, L., *et al.* (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* *486*, 405-409.
- Barnes, D.E., Lindahl, T., and Sedgwick, B. (1993). DNA repair. *Curr Opin Cell Biol* *5*, 424-433.
- 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., *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* *456*, 53-59.
- Bergamaschi, A., Kim, Y.H., Wang, P., Sorlie, T., Hernandez-Boussard, T., Lonning, P.E., Tibshirani, R., Borresen-Dale, A.L., and Pollack, J.R. (2006). Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* *45*, 1033-1040.
- Berger, M.F., Lawrence, M.S., Demichelis, F., Drier, Y., Cibulskis, K., Sivachenko, A.Y., Sboner, A., Esgueva, R., Pflueger, D., Sougnez, C., *et al.* (2011). The genomic complexity of primary human prostate cancer. *Nature* *470*, 214-220.
- Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., and Plemmons, R.J. (2007). Algorithms and applications for approximate nonnegative matrix factorization. *Comput Stat Data An* *52*, 155-173.
- Bignell, G.R., Barfoot, R., Seal, S., Collins, N., Warren, W., and Stratton, M.R. (1998). Low frequency of somatic mutations in the LKB1/Peutz-Jeghers syndrome gene in sporadic breast cancer. *Cancer research* *58*, 1384-1386.
- Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C., *et al.* (2010). Signatures of mutation and selection in the cancer genome. *Nature* *463*, 893-898.
- Bignell, G.R., Santarius, T., Pole, J.C., Butler, A.P., Perry, J., Pleasance, E., Greenman, C., Menzies, A., Taylor, S., Edkins, S., *et al.* (2007). Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. *Genome research* *17*, 1296-1303.
- Birch, J.M., Alston, R.D., McNally, R.J., Evans, D.G., Kelsey, A.M., Harris, M., Eden, O.B., and Varley, J.M. (2001). Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* *20*, 4621-4628.
- Bishop, J.M. (1985). Viral oncogenes. *Cell* *42*, 23-38.
- Bishop, W.R., and Bell, R.M. (1985). Assembly of the endoplasmic reticulum phospholipid bilayer: the phosphatidylcholine transporter. *Cell* *42*, 51-60.

Bjursell, G., Gussander, E., and Lindahl, T. (1979). Long regions of single-stranded DNA in human cells. *Nature* *280*, 420-423.

Bonner, W.M., Redon, C.E., Dickey, J.S., Nakamura, A.J., Sedelnikova, O.A., Solier, S., and Pommier, Y. (2008). GammaH2AX and cancer. *Nat Rev Cancer* *8*, 957-967.

Brown, J.R., and Thornton, J.L. (1957). Percivall Pott (1714-1788) and chimney sweepers' cancer of the scrotum. *Br J Ind Med* *14*, 68-70.

Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America* *101*, 4164-4169.

Butlin, H.T. (1892). Three Lectures on Cancer of the Scrotum in Chimney-Sweeps and Others: Delivered at the Royal College of Surgeons of England. *Br Med J* *2*, 66-71.

Cadet, J., Ravanat, J.L., Tavernaporro, M., Menoni, H., and Angelov, D. (2012). Oxidatively generated complex DNA damage: Tandem and clustered lesions. *Cancer Lett*.

Campbell, P.J., Stephens, P.J., Pleasance, E.D., O'Meara, S., Li, H., Santarius, T., Stebbings, L.A., Leroy, C., Edkins, S., Hardy, C., *et al.* (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature genetics* *40*, 722-729.

Carpten, J.D., Faber, A.L., Horn, C., Donoho, G.P., Briggs, S.L., Robbins, C.M., Hostetter, G., Boguslawski, S., Moses, T.Y., Savage, S., *et al.* (2007). A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* *448*, 439-444.

Chaires, J.B. (1990). Biophysical chemistry of the daunomycin-DNA interaction. *Biophys Chem* *35*, 191-202.

Chaires, J.B. (1998). Drug--DNA interactions. *Curr Opin Struct Biol* *8*, 314-320.

Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.P., Ahmann, G.J., Adli, M., *et al.* (2011). Initial genome sequencing and analysis of multiple myeloma. *Nature* *471*, 467-472.

Chen, J., Lindblom, P., and Lindblom, A. (1998). A study of the PTEN/MMAC1 gene in 136 breast cancer families. *Hum Genet* *102*, 124-125.

Chen, J.M., Ferec, C., and Cooper, D.N. (2012a). Transient hypermutability, chromothripsis and replication-based mechanisms in the generation of concurrent clustered mutations. *Mutation research* *750*, 52-59.

Chen, S., Qiu, J., Chen, C., Liu, C., Liu, Y., An, L., Jia, J., Tang, J., Wu, L., and Hang, H. (2012b). Affinity maturation of anti-TNF-alpha scFv with somatic hypermutation in non-B cells. *Protein Cell*.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., *et al.* (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* *10*, 529-541.

Chin, S.F., Teschendorff, A.E., Marioni, J.C., Wang, Y., Barbosa-Morais, N.L., Thorne, N.P., Costa, J.L., Pinder, S.E., van de Wiel, M.A., Green, A.R., *et al.* (2007). High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol* *8*, R215.

Ching, H.C., Naidu, R., Seong, M.K., Har, Y.C., and Taib, N.A. (2011). Integrated analysis of copy number and loss of heterozygosity in primary breast carcinomas using high-density SNP array. *Int J Oncol* *39*, 621-633.

Chiou, C.C., and Yang, J.L. (1995). Mutagenicity and specific mutation spectrum induced by 8-methoxypsoralen plus a low dose of UVA in the hprt gene in diploid human fibroblasts. *Carcinogenesis* *16*, 1357-1362.

Coticello, S.G. (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biol* *9*, 229.

Cox, A., Dunning, A.M., Garcia-Closas, M., Balasubramanian, S., Reed, M.W., Pooley, K.A., Scollen, S., Baynes, C., Ponder, B.A., Chanock, S., *et al.* (2007). A common coding variant in CASP8 is associated with breast cancer risk. *Nature genetics* *39*, 352-358.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., *et al.* (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* *486*, 346-352.

Daley, J.M., and Wilson, T.E. (2005). Rejoining of DNA double-strand breaks as a function of overhang length. *Molecular and cellular biology* *25*, 896-906.

Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLoS biology* 9, e1000594.

DeMarini, D.M., Landi, S., Tian, D., Hanley, N.M., Li, X., Hu, F., Roop, B.C., Mass, M.J., Keohavong, P., Gao, W., *et al.* (2001). Lung tumor KRAS and TP53 mutations in nonsmokers reflect exposure to PAH-rich coal combustion emissions. *Cancer research* 61, 6679-6681.

Demple, B., and DeMott, M.S. (2002). Dynamics and diversions in base excision DNA repair of oxidized abasic lesions. *Oncogene* 21, 8926-8934.

Denissenko, M.F., Pao, A., Tang, M., and Pfeifer, G.P. (1996). Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science* 274, 430-432.

Denny, W.A. (2001). DNA minor groove alkylating agents. *Curr Med Chem* 8, 533-544.

Ding, L., Ley, T.J., Larson, D.E., Miller, C.A., Koboldt, D.C., Welch, J.S., Ritchey, J.K., Young, M.A., Lamprecht, T., McLellan, M.D., *et al.* (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481, 506-510.

Doane, A.S., Danso, M., Lal, P., Donaton, M., Zhang, L., Hudis, C., and Gerald, W.L. (2006). An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* 25, 3994-4008.

Duret, L., and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 96, 4482-4487.

Duval, A., and Hamelin, R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer research* 62, 2447-2454.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struwing, J.P., Morrison, J., Field, H., Luben, R., *et al.* (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447, 1087-1093.

Eckert, K.A., and Hile, S.E. (2009). Every microsatellite is different: Intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol Carcinog* 48, 379-388.

Eisenbrand, G., Muller, N., Denkel, E., and Sterzel, W. (1986). DNA adducts and DNA damage by antineoplastic and carcinogenic N-nitrosocompounds. *J Cancer Res Clin Oncol* 112, 196-204.

El-Bayoumy, K., Chae, Y.H., Rosa, J.G., Williams, L.K., Desai, D., Amin, S., and Fiala, E. (2000). The effects of 1-nitropyrene, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine and 7,12-dimethylbenz[a]anthracene on 8-hydroxy-2'-deoxyguanosine levels in the rat mammary gland and modulation by dietary 1,4-phenylenebis(methylene) selenocyanate. *Cancer Lett* 151, 7-13.

Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., Van Tine, B.A., Hoog, J., Goiffon, R.J., Goldstein, T.C., *et al.* (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature* 486, 353-360.

Eot-Houllier, G., Eon-Marchais, S., Gasparutto, D., and Sage, E. (2005). Processing of a complex multiply damaged DNA site by human cell extracts and purified repair proteins. *Nucleic Acids Res* 33, 260-271.

Evans, M.D., Dizdaroglu, M., and Cooke, M.S. (2004). Oxidative DNA damage and disease: induction, repair and significance. *Mutation research* 567, 1-61.

Fang, M., Toher, J., Morgan, M., Davison, J., Tannenbaum, S., and Claffey, K. (2011). Genomic differences between estrogen receptor (ER)-positive and ER-negative human breast carcinoma identified by single nucleotide polymorphism array comparative genome hybridization analysis. *Cancer* 117, 2024-2034.

Farber, E., and Cameron, R. (1980). The sequential analysis of cancer development. *Adv Cancer Res* 31, 125-226.

Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., Macgrogan, G., Bergh, J., Cameron, D., Goldstein, D., *et al.* (2005). Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 24, 4660-4671.

Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61, 759-767.

Feldmeyer, N., Schmeiser, H.H., Muehlbauer, K.R., Belharazem, D., Knyazev, Y., Nedelko, T., and Hollstein, M. (2006). Further studies with a cell immortalization assay to investigate the mutation signature of aristolochic acid in human p53 sequences. *Mutation research* 608, 163-168.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., and Parkin, D.M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International journal of cancer Journal international du cancer* 127, 2893-2917.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., *et al.* (2009). Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *The New England journal of medicine* 361, 123-134.

Foulds, L. (1958). The natural history of cancer. *J Chronic Dis* 8, 2-37.

Fresco, J.R., and Alberts, B.M. (1960). The Accommodation of Noncomplementary Bases in Helical Polyribonucleotides and Deoxyribonucleic Acids. *Proceedings of the National Academy of Sciences of the United States of America* 46, 311-321.

Frosina, G., Fortini, P., Rossi, O., Carrozzino, F., Raspaglio, G., Cox, L.S., Lane, D.P., Abbondandolo, A., and Dogliotti, E. (1996). Two pathways for base excision repair in mammalian cells. *J Biol Chem* 271, 9573-9578.

Giglia-Mari, G., and Sarasin, A. (2003). TP53 mutations in human skin cancers. *Hum Mutat* 21, 217-228.

Gouy, M., and Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* 10, 7055-7074.

Greene, C.N., and Jinks-Robertson, S. (1997). Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. *Molecular and cellular biology* 17, 2844-2850.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G.L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., *et al.* (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153-158.

Greinert, R., Volkmer, B., Henning, S., Breitbart, E.W., Greulich, K.O., Cardoso, M.C., and Rapp, A. (2012). UVA-induced DNA double-strand breaks result from the repair of clustered oxidative DNA damages. *Nucleic Acids Res* 40, 10263-10273.

Hahn, W.C., and Weinberg, R.A. (2002). Rules for making human tumor cells. *The New England journal of medicine* 347, 1593-1603.

Hainaut, P., and Pfeifer, G.P. (2001). Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis* 22, 367-374.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57-70.

Hanawalt, P.C., and Spivak, G. (2008). Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9, 958-970.

Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Molecular cell* 10, 1247-1253.

Hartley, J.A., Lown, J.W., Mattes, W.B., and Kohn, K.W. (1988). DNA sequence specificity of antitumor agents. Oncogenes as possible targets for cancer therapy. *Acta Oncol* 27, 503-510.

Hastings, P.J., Lupski, J.R., Rosenberg, S.M., and Ira, G. (2009). Mechanisms of change in gene copy number. *Nat Rev Genet* 10, 551-564.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.P., *et al.* (2001). Gene-expression profiles in hereditary breast cancer. *The New England journal of medicine* 344, 539-548.

Hefferin, M.L., and Tomkinson, A.E. (2005). Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair (Amst)* 4, 639-648.

Heikkinen, K., Rapakko, K., Karppinen, S.M., Erkkö, H., Knuutila, S., Lundan, T., Mannermaa, A., Borresen-Dale, A.L., Borg, A., Barkardottir, R.B., *et al.* (2006). RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. *Carcinogenesis* 27, 1593-1599.

Hicks, J., Krasnitz, A., Lakshmi, B., Navin, N.E., Riggs, M., Leib, E., Esposito, D., Alexander, J., Troge, J., Grubor, V., *et al.* (2006). Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome research* 16, 1465-1479.

Hicks, J., Muthuswamy, L., Krasnitz, A., Navin, N., Riggs, M., Grubor, V., Esposito, D., Alexander, J., Troge, J., Wigler, M., *et al.* (2005). High-resolution ROMA CGH and FISH analysis of aneuploid and diploid breast tumors. *Cold Spring Harb Symp Quant Biol* 70, 51-63.

Hicks, W.M., Kim, M., and Haber, J.E. (2010). Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329, 82-85.

Hofr, C., Farrell, N., and Brabec, V. (2001). Thermodynamic properties of duplex DNA containing a site-specific d(GpG) intrastrand crosslink formed by an antitumor dinuclear platinum complex. *Nucleic Acids Res* 29, 2034-2040.

Hori, M., Suzuki, T., Minakawa, N., Matsuda, A., Harashima, H., and Kamiya, H. (2011). Mutagenicity of secondary oxidation products of 8-oxo-7,8-dihydro-2'-deoxyguanosine 5'-triphosphate (8-hydroxy-2'-deoxyguanosine 5'-triphosphate). *Mutation research* 714, 11-16.

Hultquist, J.F., Lengyel, J.A., Refsland, E.W., LaRue, R.S., Lackey, L., Brown, W.L., and Harris, R.S. (2011). Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. *Journal of virology* 85, 11220-11234.

Hunter, D.J., Kraft, P., Jacobs, K.B., Cox, D.G., Yeager, M., Hankinson, S.E., Wacholder, S., Wang, Z., Welch, R., Hutchinson, A., *et al.* (2007). A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature genetics* 39, 870-874.

Hwang, D.G., and Green, P. (2004). Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13994-14001.

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363, 558-561.

Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol* 7, 335-346.

Jones, S., Zhang, X., Parsons, D.W., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., *et al.* (2008). Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science (New York, NY)* 321, 1801-1806.

Jovanic, T., Roche, B., Attal-Bonnefoy, G., Leclercq, O., and Rougeon, F. (2008). Ectopic expression of AID in a non-B cell line triggers A:T and G:C point mutations in non-replicating episomal vectors. *PLoS one* 3, e1480.

Kadyrov, F.A., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* 126, 297-308.

Kallioniemi, A. (2008). CGH microarrays and cancer. *Curr Opin Biotechnol* 19, 36-40.

Karran, P., and Lindahl, T. (1980). Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochemistry* 19, 6005-6011.

Katzen, A.L., Kornberg, T.B., and Bishop, J.M. (1985). Isolation of the proto-oncogene c-myc from *D. melanogaster*. *Cell* 41, 449-456.

Key, T.J., Verkasalo, P.K., and Banks, E. (2001). Epidemiology of breast cancer. *Lancet Oncol* 2, 133-140.

King, C.R., Kraus, M.H., and Aaronson, S.A. (1985). Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science* 229, 974-976.

Kinzler, K.W., and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell* 87, 159-170.

Klarer, A.C., and McGregor, W. (2011). Replication of damaged genomes. *Crit Rev Eukaryot Gene Expr* 21, 323-336.

Knobel, P.A., and Marti, T.M. (2011). Translesion DNA synthesis in the context of cancer research. *Cancer Cell Int* 11, 39.

Knox, R.J., Lydall, D.A., Friedlos, F., Basham, C., and Roberts, J.J. (1987). The effect of monofunctional or difunctional platinum adducts and of various other associated DNA damage on the expression of transfected DNA in mammalian cell lines sensitive or resistant to difunctional agents. *Biochim Biophys Acta* 908, 214-223.

Knuutila, S., Autio, K., and Aalto, Y. (2000). Online access to CGH data of DNA sequence copy number changes. *Am J Pathol* 157, 689.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. *Annu Rev Biochem* 74, 681-710.

Kunz, B.A., Straffon, A.F., and Vonarx, E.J. (2000). DNA damage-induced mutation: tolerance via translesion synthesis. *Mutation research* 451, 169-185.

Kuraguchi, M., Edelmann, W., Yang, K., Lipkin, M., Kucherlapati, R., and Brown, A.M. (2000). Tumor-associated Apc mutations in Mlh1^{-/-} Apc1638N mice reveal a mutational signature of Mlh1 deficiency. *Oncogene* 19, 5755-5763.

Landry, S., Narvaiza, I., Linfesty, D.C., and Weitzman, M.D. (2011). APOBEC3A can activate the DNA damage response and cause cell-cycle arrest. *EMBO reports* 12, 444-450.

Laquerbe, A., Guillouf, C., Moustacchi, E., and Papadopoulo, D. (1995). The mutagenic processing of psoralen photolesions leaves a highly specific signature at an endogenous human locus. *J Mol Biol* 254, 38-49.

Leary, R.J., Lin, J.C., Cummins, J., Boca, S., Wood, L.D., Parsons, D.W., Jones, S., Sjoblom, T., Park, B.H., Parsons, R., *et al.* (2008). Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America* 105, 16224-16229.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788-791.

Lee, J.A., Carvalho, C.M., and Lupski, J.R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131, 1235-1247.

Lee, W., Jiang, Z., Liu, J., Haverty, P.M., Guan, Y., Stinson, J., Yue, P., Zhang, Y., Pant, K.P., Bhatt, D., *et al.* (2010). The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* 465, 473-477.

Liao, W., Hong, S.H., Chan, B.H., Rudolph, F.B., Clark, S.C., and Chan, L. (1999). APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. *Biochem Biophys Res Commun* 260, 398-404.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709-715.

Litman, R., Peng, M., Jin, Z., Zhang, F., Zhang, J., Powell, S., Andreassen, P.R., and Cantor, S.B. (2005). BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCI. *Cancer Cell* 8, 255-265.

Longerich, S., Basu, U., Alt, F., and Storb, U. (2006). AID in somatic hypermutation and class switch recombination. *Current opinion in immunology* 18, 164-174.

Lutsenko, E., and Bhagwat, A.S. (1999). Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. A model, its experimental support and implications. *Mutation research* 437, 11-20.

Lynch, H.T., and de la Chapelle, A. (1999). Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* 36, 801-818.

Ma, W., Panduri, V., Sterling, J.F., Van Houten, B., Gordenin, D.A., and Resnick, M.A. (2009). The transition of closely opposed lesions to double-strand breaks during long-patch base excision repair is prevented by the coordinated action of DNA polymerase delta and Rad27/Fen1. *Molecular and cellular biology* 29, 1212-1221.

Mace, K., Aguilar, F., Wang, J.S., Vautravers, P., Gomez-Lechon, M., Gonzalez, F.J., Groopman, J., Harris, C.C., and Pfeifer, A.M. (1997). Aflatoxin B1-induced DNA adduct formation and p53 mutations in CYP450-expressing human liver cell lines. *Carcinogenesis* 18, 1291-1297.

Masciari, S., Larsson, N., Senz, J., Boyd, N., Kaurah, P., Kandel, M.J., Harris, L.N., Pinheiro, H.C., Troussard, A., Miron, P., *et al.* (2007). Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet* 44, 726-731.

McCulloch, S.D., and Kunkel, T.A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res* 18, 148-161.

Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., Hollestelle, A., Houben, M., Crepin, E., van Veghel-Plandsoen, M., *et al.* (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nature genetics* 31, 55-59.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* *266*, 66-71.

Modrich, P., and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem* *65*, 101-133.

Nassif, N., Penney, J., Pal, S., Engels, W.R., and Gloor, G.B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Molecular and cellular biology* *14*, 1613-1625.

Nedelko, T., Arlt, V.M., Phillips, D.H., and Hollstein, M. (2009). TP53 mutation signature supports involvement of aristolochic acid in the aetiology of endemic nephropathy-associated tumours. *International journal of cancer Journal international du cancer* *124*, 987-990.

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., *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* *461*, 272-276.

Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., *et al.* (2012a). Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* *149*, 979-993.

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., *et al.* (2012b). The life history of 21 breast cancers. *Cell* *149*, 994-1007.

Nospikel, T. (2009). DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. *Cell Mol Life Sci* *66*, 994-1009.

Nospikel, T., and Hanawalt, P.C. (2000). Terminally differentiated human neurons repair transcribed genes but display attenuated global DNA repair and modulation of repair gene expression. *Molecular and cellular biology* *20*, 1562-1570.

Nospikel, T.P., Hyka-Nospikel, N., and Hanawalt, P.C. (2006). Transcription domain-associated repair in human cells. *Molecular and cellular biology* *26*, 8722-8730.

Nowarski, R., Wilner, O.I., Cheshin, O., Shahar, O.D., Kenig, E., Baraz, L., Britan-Rosich, E., Nagler, A., Harris, R.S., Goldberg, M., *et al.* (2012). APOBEC3G enhances lymphoma cell radioresistance by promoting cytidine deaminase-dependent DNA repair. *Blood* *120*, 366-375.

Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. *Cell* *141*, 27-38.

Oikawa, S., and Kawanishi, S. (1999). Site-specific DNA damage at GGG sequence by oxidative stress may accelerate telomere shortening. *FEBS Lett* *453*, 365-368.

Oikawa, S., Tada-Oikawa, S., and Kawanishi, S. (2001). Site-specific DNA damage at the GGG sequence by UVA involves acceleration of telomere shortening. *Biochemistry* *40*, 4763-4768.

Okuyama, S., Marusawa, H., Matsumoto, T., Ueda, Y., Matsumoto, Y., Endo, Y., Takai, A., and Chiba, T. (2012). Excessive activity of apolipoprotein B mRNA editing enzyme catalytic polypeptide 2 (APOBEC2) contributes to liver and lung tumorigenesis. *International journal of cancer Journal international du cancer* *130*, 1294-1301.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., *et al.* (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine* *351*, 2817-2826.

Palacios, J., Robles-Frias, M.J., Castilla, M.A., Lopez-Garcia, M.A., and Benitez, J. (2008). The molecular pathology of hereditary breast cancer. *Pathobiology : journal of immunopathology, molecular and cellular biology* *75*, 85-94.

Papadopoulo, D., Laquerbe, A., Guillouf, C., and Moustacchi, E. (1993). Molecular spectrum of mutations induced at the HPRT locus by a cross-linking agent in human cell lines with different repair capacities. *Mutation research* *294*, 167-177.

Papaemmanuil, E., Cazzola, M., Boultonwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J.S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., *et al.* (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *The New England journal of medicine* *365*, 1384-1395.

Parker, R.C., Varmus, H.E., and Bishop, J.M. (1984). Expression of v-src and chicken c-src in rat cells demonstrates qualitative differences between pp60v-src and pp60c-src. *Cell* 37, 131-139.

Pathania, S., Nguyen, J., Hill, S.J., Scully, R., Adelmant, G.O., Marto, J.A., Feunteun, J., and Livingston, D.M. (2011). BRCA1 is required for postreplication repair after UV-induced DNA damage. *Molecular cell* 44, 235-251.

Pena-Diaz, J., and Jiricny, J. (2012). Mammalian mismatch repair: error-free or error-prone? *Trends Biochem Sci* 37, 206-214.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., *et al.* (2000). Molecular portraits of human breast tumours. *Nature* 406, 747-752.

Perucho, M., Goldfarb, M., Shimizu, K., Lama, C., Fogh, J., and Wigler, M. (1981). Human-tumor-derived cell lines contain common and different transforming genes. *Cell* 27, 467-476.

Petit, V., Guetard, D., Renard, M., Keriell, A., Sitbon, M., Wain-Hobson, S., and Vartanian, J.P. (2009). Murine APOBEC1 is a powerful mutator of retroviral and cellular RNA in vitro and in vivo. *J Mol Biol* 385, 65-78.

Pfeifer, G.P. (2000). p53 mutational spectra and the role of methylated CpG sequences. *Mutation research* 450, 155-166.

Pfeifer, G.P., Denissenko, M.F., Olivier, M., Tretyakova, N., Hecht, S.S., and Hainaut, P. (2002). Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* 21, 7435-7451.

Pfeifer, G.P., You, Y.H., and Besaratinia, A. (2005). Mutations induced by ultraviolet light. *Mutation research* 571, 19-31.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. *Nature* 424, 103-107.

Pickrell, J.K., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics* 27, 2144-2146.

Plesance, 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., *et al.* (2010a). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463, 191-196.

Plesance, 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., *et al.* (2010b). A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463, 184-190.

Polak, P., and Arndt, P.F. (2008). Transcription induces strand-specific mutations at the 5' end of human genes. *Genome research* 18, 1216-1223.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast cancer research : BCR* 12, R68.

Puente, X.S., Pinyol, M., Quesada, V., Conde, L., Ordonez, G.R., Villamor, N., Escaramis, G., Jares, P., Bea, S., Gonzalez-Diaz, M., *et al.* (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* 475, 101-105.

Pulciani, S., Santos, E., Lauver, A.V., Long, L.K., Robbins, K.C., and Barbacid, M. (1982). Oncogenes in human tumor cell lines: molecular cloning of a transforming gene from human bladder carcinoma cells. *Proceedings of the National Academy of Sciences of the United States of America* 79, 2845-2849.

Rahman, N., Seal, S., Thompson, D., Kelly, P., Renwick, A., Elliott, A., Reid, S., Spanova, K., Barfoot, R., Chagtai, T., *et al.* (2007). PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nature genetics* 39, 165-167.

Reddy, E.P., Reynolds, R.K., Santos, E., and Barbacid, M. (1982). A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* 300, 149-152.

Reid, S., Schindler, D., Hanenberg, H., Barker, K., Hanks, S., Kalb, R., Neveling, K., Kelly, P., Seal, S., Freund, M., *et al.* (2007). Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nature genetics* 39, 162-164.

Reis-Filho, J.S., and Pusztai, L. (2011). Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* 378, 1812-1823.

Renan, M.J. (1993). How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol Carcinog* 7, 139-146.

Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., Van Loo, P., Van Den Bossche, M., Catthoor, K., Sabbe, B., *et al.* (2012). Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nat Biotechnol* 30, 61-68.

Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M., *et al.* (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328, 636-639.

Roberts, J.D., and Kunkel, T.A. (1988). Fidelity of a human cell DNA replication complex. *Proceedings of the National Academy of Sciences of the United States of America* 85, 7064-7068.

Roberts, S.A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., Klimczak, L.J., Kryukov, G.V., Malc, E., Mieczkowski, P.A., *et al.* (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Molecular cell* 46, 424-435.

Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: Base excision repair: the long and short of it. *Cell Mol Life Sci* 66, 981-993.

Rogozin, I.B., Basu, M.K., Jordan, I.K., Pavlov, Y.I., and Koonin, E.V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. *Cell Cycle* 4, 1281-1285.

Rowley, J.D. (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 243, 290-293.

Sale, J.E., Lehmann, A.R., and Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol* 13, 141-152.

Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S.M., Riggins, G.J., *et al.* (2004). High frequency of mutations of the PIK3CA gene in human cancers. *Science* 304, 554.

Satoh, M.S., Jones, C.J., Wood, R.D., and Lindahl, T. (1993). DNA excision-repair defect of xeroderma pigmentosum prevents removal of a class of oxygen free radical-induced base lesions. *Proceedings of the National Academy of Sciences of the United States of America* 90, 6335-6339.

Seal, S., Thompson, D., Renwick, A., Elliott, A., Kelly, P., Barfoot, R., Chagtai, T., Jayatilake, H., Ahmed, M., Spanova, K., *et al.* (2006). Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nature genetics* 38, 1239-1241.

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

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418, 646-650.

Shen, J.C., Rideout, W.M., 3rd, and Jones, P.A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res* 22, 972-976.

Shibata, D., Peinado, M.A., Ionov, Y., Malkhosyan, S., and Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. *Nature genetics* 6, 273-281.

Shih, C., Padhy, L.C., Murray, M., and Weinberg, R.A. (1981). Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature* 290, 261-264.

Simon, M.A., Drees, B., Kornberg, T., and Bishop, J.M. (1985). The nucleotide sequence and the tissue-specific expression of Drosophila c-src. *Cell* 42, 831-840.

Sjoblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., *et al.* (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268-274.

Slupphaug, G., Kavli, B., and Krokan, H.E. (2003). The interacting pathways for prevention and repair of oxidative DNA damage. *Mutation research* 531, 231-251.

Smith, C.E., Llorente, B., and Symington, L.S. (2007). Template switching during break-induced replication. *Nature* 447, 102-105.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001a). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10869-10874.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., *et al.* (2001b). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10869-10874.

Spencer, W.A., Vadhanam, M.V., Jeyabalan, J., and Gupta, R.C. (2011). Oxidative DNA damage following microsomal/Cu(II)-mediated activation of the estrogens, 17beta-estradiol, equilenin and equilin: Role of reactive oxygen species. *Chemical research in toxicology*.

Spencer, W.A., Vadhanam, M.V., Jeyabalan, J., and Gupta, R.C. (2012). Oxidative DNA damage following microsomal/Cu(II)-mediated activation of the estrogens, 17beta-estradiol, equilenin, and equilin: role of reactive oxygen species. *Chemical research in toxicology* 25, 305-314.

Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., *et al.* (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature genetics* 39, 865-869.

Stacey, S.N., Manolescu, A., Sulem, P., Thorlacius, S., Gudjonsson, S.A., Jonsson, G.F., Jakobsdottir, M., Bergthorsson, J.T., Gudmundsson, J., Aben, K.K., *et al.* (2008). Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature genetics* 40, 703-706.

Staszewski, O., Baker, R.E., Ucher, A.J., Martier, R., Stavnezer, J., and Guikema, J.E. (2011). Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig loci in activated B cells. *Molecular cell* 41, 232-242.

Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J., and Harris, R.S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nature structural & molecular biology* 17, 222-229.

Stephens, P., Edkins, S., Davies, H., Greenman, C., Cox, C., Hunter, C., Bignell, G., Teague, J., Smith, R., Stevens, C., *et al.* (2005). A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nature genetics* 37, 590-592.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., *et al.* (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* 144, 27-40.

Stephens, P.J., McBride, D.J., Lin, M.L., Varela, I., Pleasance, E.D., Simpson, J.T., Stebbings, L.A., Leroy, C., Edkins, S., Mudie, L.J., *et al.* (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 462, 1005-1010.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., *et al.* (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature* 486, 400-404.

Stojic, L., Brun, R., and Jiricny, J. (2004). Mismatch repair and DNA damage signalling. *DNA Repair (Amst)* 3, 1091-1101.

Stransky, N., Egloff, A.M., Tward, A.D., Kostic, A.D., Cibulskis, K., Sivachenko, A., Kryukov, G.V., Lawrence, M.S., Sougnez, C., McKenna, A., *et al.* (2011). The mutational landscape of head and neck squamous cell carcinoma. *Science* 333, 1157-1160.

Strathern, J.N., Shafer, B.K., and McGill, C.B. (1995). DNA synthesis errors associated with double-strand-break repair. *Genetics* 140, 965-972.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. *Nature* 458, 719-724.

Strauss, B.S. (2002). The "A" rule revisited: polymerases as determinants of mutational specificity. *DNA Repair (Amst)* 1, 125-135.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., and Inouye, M. (1966). Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harb Symp Quant Biol* *31*, 77-84.

Streisinger, G., and Owen, J. (1985). Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. *Genetics* *109*, 633-659.

Suspene, R., Aynaud, M.M., Guetard, D., Henry, M., Eckhoff, G., Marchio, A., Pineau, P., Dejean, A., Vartanian, J.P., and Wain-Hobson, S. (2011). Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proceedings of the National Academy of Sciences of the United States of America* *108*, 4858-4863.

Teng, B., Burant, C.F., and Davidson, N.O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* *260*, 1816-1819.

Thibodeau, S.N., Bren, G., and Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* *260*, 816-819.

Thomas, N.E., Alexander, A., Edmiston, S.N., Parrish, E., Millikan, R.C., Berwick, M., Groben, P., Ollila, D.W., Mattingly, D., and Conway, K. (2004). Tandem BRAF mutations in primary invasive melanomas. *J Invest Dermatol* *122*, 1245-1250.

Thompson, D., Duedal, S., Kirner, J., McGuffog, L., Last, J., Reiman, A., Byrd, P., Taylor, M., and Easton, D.F. (2005). Cancer risks and mortality in heterozygous ATM mutation carriers. *J Natl Cancer Inst* *97*, 813-822.

Thompson, W.D. (1994). Genetic epidemiology of breast cancer. *Cancer* *74*, 279-287.

Tiraby, G., Fox, M.S., and Bernheimer, H. (1975). Marker discrimination in deoxyribonucleic acid-mediated transformation of various *Pneumococcus* strains. *J Bacteriol* *121*, 608-618.

Tornaletti, S., Maeda, L.S., and Hanawalt, P.C. (2006). Transcription arrest at an abasic site in the transcribed strand of template DNA. *Chemical research in toxicology* *19*, 1215-1220.

Tran, H.T., Keen, J.D., Krickler, M., Resnick, M.A., and Gordenin, D.A. (1997). Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Molecular and cellular biology* *17*, 2859-2865.

Turnbull, C., and Rahman, N. (2008). Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet* *9*, 321-345.

Usary, J., Llaca, V., Karaca, G., Presswala, S., Karaca, M., He, X., Langerod, A., Karesen, R., Oh, D.S., Dressler, L.G., *et al.* (2004). Mutation of GATA3 in human breast tumors. *Oncogene* *23*, 7669-7678.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., *et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* *415*, 530-536.

van Gent, D.C., and van der Burg, M. (2007). Non-homologous end-joining, a sticky affair. *Oncogene* *26*, 7731-7740.

Van Loo, P., Nordgard, S.H., Lingjaerde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., *et al.* (2010). Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences of the United States of America* *107*, 16910-16915.

Vincent-Salomon, A., Lucchesi, C., Gruel, N., Raynal, V., Pierron, G., Goudefroye, R., Reyat, F., Radvanyi, F., Salmon, R., Thiery, J.P., *et al.* (2008). Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. *Clinical cancer research : an official journal of the American Association for Cancer Research* *14*, 1956-1965.

Voulgaridou, G.P., Anastopoulos, I., Franco, R., Panayiotidis, M.I., and Pappa, A. (2011). DNA damage induced by endogenous aldehydes: current state of knowledge. *Mutation research* *711*, 13-27.

Vreeswijk, M.P., Overkamp, M.W., Westland, B.E., van Hees-Stuivenberg, S., Vrieling, H., Zdzienicka, M.Z., van Zeeland, A.A., and Mullenders, L.H. (1998). Enhanced UV-induced mutagenesis in the UV61 cell line, the Chinese hamster homologue of Cockayne's syndrome B, is associated with defective transcription coupled repair of cyclobutane pyrimidine dimers. *Mutation research* *409*, 49-56.

Walker, B., Jr., and Gerber, A. (1981). Occupational exposure to aromatic amines: benzidine and benzidine-based dyes. *Natl Cancer Inst Monogr*, 11-13.

Wang, D., Kreutzer, D.A., and Essigmann, J.M. (1998). Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. *Mutation research* *400*, 99-115.

Wang, J., Gonzalez, K.D., Scaringe, W.A., Tsai, K., Liu, N., Gu, D., Li, W., Hill, K.A., and Sommer, S.S. (2007). Evidence for mutation showers. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 8403-8408.

Wang, Y., Sheppard, T.L., Tornaletti, S., Maeda, L.S., and Hanawalt, P.C. (2006). Transcriptional inhibition by an oxidized abasic site in DNA. *Chemical research in toxicology* *19*, 234-241.

Waters, T.R., and Swann, P.F. (2000). Thymine-DNA glycosylase and G to A transition mutations at CpG sites. *Mutation research* *462*, 137-147.

Watson, J.D., and Crick, F.H. (1953a). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* *171*, 737-738.

Watson, J.D., and Crick, F.H. (1953b). The structure of DNA. *Cold Spring Harb Symp Quant Biol* *18*, 123-131.

Weigelt, B., Geyer, F.C., and Reis-Filho, J.S. (2010). Histological types of breast cancer: how special are they? *Mol Oncol* *4*, 192-208.

Weinberg, R.A. (1989). Oncogenes, antioncogenes, and the molecular bases of multistep carcinogenesis. *Cancer research* *49*, 3713-3721.

Weisburger, J.H., Dolan, L., and Pittman, B. (1998). Inhibition of PhIP mutagenicity by caffeine, lycopene, daidzein, and genistein. *Mutation research* *416*, 125-128.

Weterings, E., and Chen, D.J. (2008). The endless tale of non-homologous end-joining. *Cell Res* *18*, 114-124.

Witz, I.P., and Levy-Nissenbaum, O. (2006). The tumor microenvironment in the post-PAGET era. *Cancer Lett* *242*, 1-10.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjoblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J., *et al.* (2007). The genomic landscapes of human breast and colorectal cancers. *Science* *318*, 1108-1113.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., and Micklem, G. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* *378*, 789-792.

Wu, L., and Hickson, I.D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* *426*, 870-874.

Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M., and Innerarity, T.L. (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proceedings of the National Academy of Sciences of the United States of America* *92*, 8483-8487.

Yang, M. (2011). A current global view of environmental and occupational cancers. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev* *29*, 223-249.

Yang, S.C., Lin, J.G., Chiou, C.C., Chen, L.Y., and Yang, J.L. (1994). Mutation specificity of 8-methoxypsoralen plus two doses of UVA irradiation in the hprt gene in diploid human fibroblasts. *Carcinogenesis* *15*, 201-207.

Yang, Y., Sterling, J., Storici, F., Resnick, M.A., and Gordenin, D.A. (2008). Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. *PLoS Genet* *4*, e1000264.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* *25*, 2865-2871.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* *18*, 821-829.

Zhu, J., Yue, Z., and Wang, J. (1998). [Oxidative DNA damage of cartilage in osteoarthritis]. *Hunan Yi Ke Da Xue Xue Bao* *23*, 438-440.

LIST OF TABLES

Table 1.1:	Known mutational signatures of DNA damage and repair mechanisms	39
Table 1.2:	Known risk factors for breast cancer	42
Table 1.3:	Gene expression classification taken from Reis-Filho and Puztai, 2011	45
Table 1.4:	Summary of known genetic cancer-predisposing alleles obtained from review	48
Table 3.1:	The reasons and the definitions of post-processing filters	97
Table 3.2:	Summary of substitution variants	101
Table 3.3:	Fraction of the genome effectively excluded by relevant filters	103
Table 3.4:	Comparison between MuTect and CaVEMan	108
Table 4.1:	Demographic information regarding breast cancers	115
Table 4.2:	Final sequencing metrics of whole-genome sequenced breast cancers	116
Table 4.3:	Putative somatic substitution and insertion/deletion driver events	117
Table 4.4:	Breast cancer series and total number of substitutions	118
Table 4.5:	Breakdown of the different types of (predicted) substitution mutations	119
Table 5.1:	Junctional features of somatic structural variation in PD4107a	152
Table 5.2:	Junctional features of somatic structural variation in PD4103a	152
Table 5.3:	Regions of kataegis involving highest number of variants	154
Table 7.1:	The double substitutions identified in twenty-one breast cancers	183
Table 7.2:	Estimates of aberrant cell fraction and ploidy	193
Table 7.3:	Amplifications identified in 21 breast cancers	199
Table 7.4:	Homozygous deletions identified by ASCAT	200
Table 8.1:	Variant allele fractions of processive heterozygous mutations	209

LIST OF FIGURES

Figure 1.1:	Mutational signatures in breast cancers	10
Figure 1.2:	Base susceptibility to damage	12
Figure 1.3:	The range of mutagenic radiation in the electromagnetic spectrum	18
Figure 1.4:	DNA damaging agents and mutagenic effects	23
Figure 1.5:	An outline of short-patch versus long-patch base excision repair (BER)	26
Figure 1.6:	An outline of nucleotide excision repair	29
Figure 1.7:	The key steps involved in mismatch repair	31
Figure 1.8:	Repair of double-strand breaks: The options of non-homologous end-joining and homologous recombination	35
Figure 1.9:	Different modes of resolving double-strand breaks within homologous recombination	38
Figure 1.10:	Somatic mutations in breast cancer	51-52
Figure 2.1:	Flowchart of the whole-genome sequencing and analysis strategy for twenty-one breast cancers	58
Figure 2.2:	Flow diagram of the principles of the DNA preparation process	60
Figure 2.3:	Typical Agilent bioanalyser traces obtained during the DNA preparation process	64
Figure 2.4:	The principles of next-generation sequencing	66
Figure 2.5:	Library and sequencing QC metrics used within the Cancer Genome Project	68
Figure 2.6:	The fastq format	69
Figure 2.7:	The generation of BAM files	70
Figure 2.8:	The principle of calling somatic mutations	71
Figure 2.9:	The principle of indel detection	73
Figure 2.10:	The classification of somatic rearrangements	75
Figure 2.11:	Data handling of 454 pyrosequencing validation data	78
Figure 2.12:	Concordance in variant allele fractions between Illumina and 454 pyrosequencing data	79
Figure 2.13:	Validation of putative somatic rearrangements by custom-designed PCR	80
Figure 2.14:	Validation of somatic rearrangements by local reassembly	81

Figure 3.1:	Phred score and mapping qualities	85
Figure 3.2:	Differences in calling variants in a germline genome and a tumour genome	87
Figure 3.3:	The CaVEMan workflow	90
Figure 3.4:	Reads in G-browse, the genome browser used to view short read sequences	93
Figure 3.5:	Identifying false positive calls in substitution data	94
Figure 3.6:	The principle of developing post-processing filters	95
Figure 3.7:	Schematic of reduction of substitution variants following post-processing	100
Figure 3.8:	Variants removed exclusively by each filter	102
Figure 3.9:	Improvement in positive predictive value for each cancer genome	104
Figure 3.10:	A comparison of sensitivity and positive predictive value of PD3890a before and after development of post-processing filters	105
Figure 3.11:	Correlations between positive predictive value and sequence coverage and aberrant cell fraction	106
Figure 3.12:	Schematic of the final analysis pipeline	109
Figure 4.1:	Somatic mutation profiles of 21 breast cancers	121
Figure 4.2:	Selection of the optimal number of signatures via the NMF model selection framework	126
Figure 4.3:	Five mutational signatures extracted by NMF in 21 breast cancers	128
Figure 4.4:	Contrasting signatures and validation of signatures	130
Figure 4.5:	Cluster dendrogram generated by unsupervised hierarchical clustering	133
Figure 4.6:	The temporality of mutational processes	137
Figure 5.1:	The principle behind rainfall plots	145
Figure 5.2:	Rainfall plot for PD4107a	146
Figure 5.3:	Rainfall plots for chromosome 6 of PD4107a	149-150
Figure 5.4:	Rainfall plots for PD4103a	153
Figure 5.5:	Rainfall plots for 18 genomes	156
Figure 5.6:	Timing kataegis	158
Figure 6.1:	Transcriptional strands	165
Figure 6.2:	Cluster dendrogram of transcriptomic profiles of 17 breast cancers	166

Figure 6.3:	The relationship between mutation prevalence and transcription	169
Figure 6.4:	The effect of transcription start site and mutation prevalence	171
Figure 6.5:	Mutation rate with distance from transcription start site	172
Figure 7.1:	Relationship between total number of insertions/deletions and mutation burden of other classes of mutation	179
Figure 7.2:	Somatic mutation profile of indels	181
Figure 7.3:	Relationship between first and second substitution in double substitutions showing enrichment for CC>AA mutations	184
Figure 7.4:	Circos plots demonstrating rearrangements in 20 breast cancers	187
Figure 7.5:	Variation in rearrangement architecture in 21 breast cancers	188
Figure 7.6:	Patterns of microhomology and non-templated sequence at rearrangement breakpoints	190
Figure 7.7:	Copy number plots	195-197
Figure 8.1:	Models for APOBEC activity in the genesis of kataegis and Signature B	210

LIST OF ABBREVIATIONS

(6-4)PPs	6-4 pyrimidine-pyrimidone photoproducts
8-oxo-dG	8-oxo-2'-deoxyguanosine
AID	activation-induced cytidine deaminase
APC	adenomatous polyposis coli
AP site	apurinic/apyrimidinic site
APE1	AP-endonuclease
APOBEC	apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like
ASCAT	allele-specific copy number analysis of tumors
ATM	ataxia-telangiectasia mutated gene
B[a]DPE	benzo[a]pyrene diolepoxide
BIR	break-induced replication
BWA	Burrows-Wheeler Alignment
CaVEMan	Cancer Variants Through Expectation Maximization
CDH1	cadherin1
CGH	comparative genomic hybridisation
CNA	copy number aberration
CPD	cyclobutane pyrimidine dimers
DAR	domain-associated repair
DCIS	ductal carcinoma in-situ
DNA	deoxyribose nucleic acid
DNA-PK	DNA-dependent protein kinase
DSBR	double-strand break repair
ENU	ethyl nitrosurea
ER	oestrogen receptor
FEN1	flap structure-specific endonuclease 1
GGR	global genome repair

GWAS	genome-wide association study
HPRT	hypoxanthine guanine phosphoribosyltransferase 1
HR	homologous recombination
HRT	hormone replacement therapy
ICGC	International Cancer Genome Consortium
IHC	immunohistochemistry
K-S	Kolmogorov-Smirnov test
LCIS	lobular carcinoma in situ
LOH	loss of heterozygosity
MMR	mismatch repair
MMS	methyl methane sulfonate
MNNG	methylnitronitrosoguanidine
MNU	methyl nitrosurea
mRNA	messenger RNA
NER	nucleotide excision repair
NGS	next-generation sequencing technology
NHEJ	non-homologous end-joining
NBS1	Nijmegen breakage syndrome
NMF	non-negative matrix factorization
OCDL	oxidative clustered DNA lesion
OCP	oral contraceptive pill
PAH	polyaromatic hydrocarbon
PALB2	partner and localizer of BRCA2
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PhIP	amino-1-methyl-6-phenylimidazo [4,5-b] pyridine
Pol	polymerase
PPV	positive predictive value

PR	progesterone receptor
RFC	replication factor C
RNA	ribonucleic acid
RPM	revolutions per minute
SDSA	synthesis-dependent strand annealing
SNP	single-nucleotide polymorphism
SSA	single-strand annealing
TCR	transcription-coupled repair
TFIIH	transcription-factor IIH
TP53	tumour protein p53
UNG	uracil-N-glycosylase
UV	ultraviolet
XP	xeroderma pigmentosa
XRCC1	X-ray repair, complementing defective, in chinese hamster