



Published in final edited form as:

Clin Transl Oncol. 2010 March ; 12(3): 174–180.

Selection of extreme phenotypes: the role of clinical observation in translational research

José Luis Pérez-Gracia,

Medical Oncology Department, Clínica Universidad de Navarra, Universidad de Navarra, Avenida Pío XII, 36, ES-31008 Pamplona, Spain

Alfonso Gúrpide,

Medical Oncology Department, Clínica Universidad de Navarra, Universidad de Navarra, Avenida Pío XII, 36, ES-31008 Pamplona, Spain

María Gloria Ruiz-Ilundain,

Anesthesiology Department, Clínica Ubarmin, Pamplona, Spain

Carlos Alfaro Alegría,

Gene Therapy Unit CIMA, Universidad de Navarra Pamplona, Spain

Ramon Colomer,

MD Anderson International España, Madrid, Spain

Jesús García-Foncillas, and

Medical Oncology Department, Clínica Universidad de Navarra, Universidad de Navarra, Avenida Pío XII, 36, ES-31008 Pamplona, Spain

Ignacio Melero Bermejo

Gene Therapy Unit CIMA, Universidad de Navarra Pamplona, Spain. Internal Medicine Department, Clínica Universidad de Navarra, Universidad de Navarra, Pamplona, Spain

José Luis Pérez-Gracia: jlgracia@unav.es

Abstract

Systematic collection of phenotypes and their correlation with molecular data has been proposed as a useful method to advance in the study of disease. Although some databases for animal species are being developed, progress in humans is slow, probably due to the multifactorial origin of many human diseases and to the intricacy of accurately classifying phenotypes, among other factors. An alternative approach has been to identify and to study individuals or families with very characteristic, clinically relevant phenotypes. This strategy has shown increased efficiency to identify the molecular features underlying such phenotypes. While on most occasions the subjects selected for these studies presented harmful phenotypes, a few studies have been performed in individuals with very favourable phenotypes. The consistent results achieved suggest that it seems logical to further develop this strategy as a methodology to study human disease, including cancer. The identification and the study with high-throughput techniques of individuals showing a markedly decreased risk of developing cancer or of cancer patients presenting either an unusually favourable prognosis or striking responses following a specific treatment, might be promising ways to maximize the yield of this approach and to reveal the molecular causes that explain those phenotypes and thus highlight useful therapeutic targets. This manuscript reviews the current

Correspondence to: José Luis Pérez-Gracia, jlgracia@unav.es.

Conflict of interest The authors declare that they have no conflict of interest relating to the publication of this manuscript.

status of selection of extreme phenotypes in cancer research and provides directions for future development of this methodology.

Keywords

Cancer genetics; Protective genetic alterations; Single-nucleotide polymorphisms; Long-term cancer survivors; Familial cancer syndromes; Extreme phenotype selection

Introduction

Despite the fact that human beings share the vast majority of their genetic information, the few remaining variations account for an astonishingly wide range of different phenotypes. The importance of characterizing such differences is widely recognized by the scientific community and enormous efforts are being made to understand their role in disease [1]. Indeed, a useful method to advance in this path has been to correlate molecular data with phenotypes. The soundness of this strategy is straightforward and the development of phenotype databases has already been proposed [2,3]. Yet, although such databases are being generated for species such as yeasts [4] or rodents [5,6] and are under development for human beings [7,8], progress is proceeding at a relatively slow pace due to the enormous complexity of this task, which is at least in part caused by the multifactorial origin of many diseases and by the intricacy of accurately classifying phenotypes.

In the meanwhile, a frequent approach has been to correlate molecular features in groups of patients with their phenotypes, expressed as clinical variables, such as prognosis or treatment effects. Some relevant examples of the use of this strategy in oncology are the correlation between thymidylate synthase expression and efficacy of 5-fluorouracil in digestive tumours [9]; or the identification of gene-expression profiles of prognostic value in lymphoma [10] or breast cancer patients [11]. However, despite the significance of some results, the conclusions reached by many studies are of uncertain clinical significance or even contradictory [12]. This has led to the establishment of specific guidelines to validate conclusions before their publication [12–14]. Several factors may cause these biases, including methodological issues, such as retrospective data collection, limitations in laboratory techniques or the biology of complex diseases [12,15], such as cancer, which present multiplex phenotypes. In addition, classification of patients into subgroups with good or bad evolution that present moderate differences, such as subtle improvements in survival, from one to the other may lead to the identification of molecular features associated with modest differences of borderline clinical relevance.

A useful and intuitive approach to circumvent some of these problems has been to select individuals with very characteristic, clinically relevant phenotypes and to study the underlying causes. This strategy assumes that these patients are the most informative and thus should be studied separately, rather than being included in larger series of patients that might dissipate the information that they can provide. Even though this strategy has allowed the identification of relevant biological facts with great effectiveness, through the study of reduced numbers of subjects, and has been proposed as a methodology for the study of human disease [16–20], its use has not become widespread. This manuscript reviews the current status of extreme phenotype selection in cancer research and provides relevant examples that support its value, along with potential directions to further develop this strategy.

Selection of apparent phenotypes

Apparent phenotypes present characteristic attributes and therefore can be identified by observation. Sometimes the phenotype is readily recognized, because its characteristics are obvious. This is the case of the widely employed strategy of identifying gene mutations that cause genetically inherited diseases [21]. Paradigmatic examples in oncology include syndromes characterized by the development of multiple tumours, such as multiple endocrine neoplasia, type 1 (MEN-1). The description of parathyroid, pancreatic and pituitary tumours in autopsies of patients with acromegaly [22], and its familial association [23,24] preceded by decades the identification of the MEN-1 tumour-suppressor oncogene [25] and its mutations in affected individuals [26]. Another outstanding example is the detection of mutations in BRCA-1, a gene that was identified in families that presented a high incidence of early-onset breast carcinoma [27].

On other occasions the phenotypes are less evident and complex epidemiological studies are required to identify them. An illustrative example is the Li-Fraumeni syndrome, described through the identification of an increased incidence of rhabdomyosarcomas in siblings following the review of over 20,000 children's death certificates [28,29]. As in the previous example, the identification of the phenotype allowed the hypothesis to be formulated that eventually led to the detection of germ-line p53 mutations as the cause [30].

Finally, sometimes characteristic phenotypes are expressed only under certain circumstances, such as after treatment administration. For example, the description of severe toxicity after 5-fluorouracil administration [31] allowed the identification of the biochemical [32] as well as the genetic underlying causes [33]. Another relevant example is the discovery of the expression of epidermal growth factor receptor (EGFR) mutations in tumours of patients responding to EGFR tyrosine-kinase inhibitors [34,35]. Paradoxically, in this setting the study of a relatively low number of subjects yielded very relevant and clinically useful results, in contrast to the modest conclusions obtained after studying thousands of patients through conventional randomized trials [36–39], which assume that the benefit that the treatment produces only in a group of patients is large enough to administer it to the whole unselected population. Interestingly, the selection and study of patients with extreme responses (most vs. least sensitive), extreme drug metabolisms (high vs. low) or extreme toxic effects (no toxicity at high doses vs. high toxicity at low doses) following drug therapy has been described by Nebert as a well defined methodology in clinical pharmacology, in a manuscript that reviews studies in which this strategy has been successfully employed [16]. The same author and his collaborators have also described in detail the statistical rationale that supports this methodology [40].

The common factor underlying all these examples is that the initial step was the identification of the characteristic phenotypes. Subsequent study of the selected individuals led to the identification of the molecular causes. The effectiveness of this strategy is notable, because the number of subjects that need to be studied is relatively small. In addition, the clinical relevance is high, because the attributes of the selected phenotypes are significant. Therefore, we believe that this methodology should be further developed in cancer research. Relevant case-selection should include cancer patients with very characteristic and uncommon evolution. Currently, many advanced solid tumours are considered incurable and result in short survival. Nonetheless, clinical experience shows that exceptions exist even among the malignancies presenting the direst prognosis, and every oncologist treats patients that are unexpectedly cured or that live far beyond their estimated prognosis. Even though many of those cases are not published, some reports of long-term survivors of apparently incurable tumours such as pancreatic cancer [41,42], gastric cancer [43–46], colon cancer [47], small [48,49] and non-small-cell lung cancer [50,51] or multiple myeloma [52] can be

found in the medical literature. Assuming that the diagnostic and staging work-up has been correctly performed, these individuals may represent extreme phenotypes worthy of detailed study. Similarly, patients presenting early-stage cancer that receive adequate treatment and that either do not relapse despite presenting a high risk of recurrence or relapse despite a very low risk of recurrence, might be interesting groups to study. A clinical example of the former design would be to study patients with T1G3 bladder tumours that do not relapse and patients that present low-grade papillary tumours that recur following adequate local treatment. Patients showing extreme responses or toxic effects to a given treatment, as suggested by Nebert [16], also constitute interesting phenotypes to identify and to study. In fact, this methodological design deserves further consideration in the current scenario in which hundreds of drugs are being developed, but only a few obtain regulatory approval, based on conventional drug development methodology. Further development of multiple drugs might be of clinical interest, even if they just show clinical activity in a limited subgroup of patients, assuming that it would be possible to identify such patients. Studies in which patients receive a treatment in a non-randomized fashion and those patients that present marked benefit are intensively studied to elucidate specific markers of activity might become complementary to current randomized studies. The sample size of these studies should allow to identify a sufficient number of patients presenting clear benefit from the treatment. Such studies might become a useful tool to further develop personalized medicine, allowing identification of those patients that achieve a truly significant benefit from a specific treatment.

Multiple potential causes might explain these phenotypes, and they could be related to host and tumour factors as well as environmental causes (Fig. 1). Host factors might include regulation of immune response, angiogenesis, apoptosis or DNA repair mechanisms, ability to control metastasis or advantageous metabolism of anti-cancer treatments, among others. Tumour factors might comprise abnormalities in drug or immune resistance, cell cycle and apoptosis regulation, and so on. Since multiple hypotheses should be studied, the assessment of specific major molecular pathways and the use of high-throughput techniques, which allow the simultaneous assessment of multiple biological factors, seem necessary. Even though the interpretations of these techniques is somewhat cumbersome, due to the large amount of information they produce, the use of a reduced sample population that present marked and clinically relevant characteristics should clearly improve the efficiency of such studies.

This strategy has already been followed by some researchers. The study of melanoma patients with long-term survival has highlighted tumour immune escape as a mechanism of disease progression and shifting of T-cell responses as a response to this escape [53], as well as prolonged persistence of specific CD8+ cells as a potential cause of maintenance of long disease remissions [54,55]. Interestingly, all these observations, which support the concepts of cancer immunosurveillance and immunoediting [56], were obtained from the study of just 4 patients. We have used this strategy in renal cell carcinoma patients treated with the antiangiogenic drug sunitinib. Sera from 3 patients showing marked responses and from 3 patients that presented clear progressions were analyzed with a Human Cytokine Array, which evaluates 174 cytokines. We identified 27 cytokines, which varied significantly between both groups, and we further selected and assessed the most relevant cytokines by ELISA in 21 evaluable patients, concluding that TNF- α and MMP-9 baseline levels were predictive of response [57]. In a similar study performed in melanoma and renal cell-carcinoma patients, vascular endothelial growth factor and fibronectin were identified as predictors of response to high dose intravenous IL-2 [58]. Although these studies warrant confirmation, they support the relevance of selection of patients with extreme phenotypes and their study with high-throughput techniques as a valid method to identify candidate predictive factors of drug activity.

Selection of non-apparent phenotypes

Non-apparent phenotypes are those not associated with characteristic attributes and therefore they cannot be identified by mere observation. An example is the existence of protection against developing a disease. Since one specific disease does not develop in the majority of individuals, the existence of protection cannot be distinguished from the simple absence of disease by chance, unless the risk of developing that disease is taken into consideration.

It is well known that cancer incidence under similar environmental conditions is not uniform. If cancer risk followed a normal distribution, as most biological variables do, we could hypothesize that just as some individuals present an increased incidence, other subjects may have lower incidence than expected. If these individuals exist, it would be naïve to attribute their phenotype to chance, at least until other causes have been ruled out, and their identification and study could increase our knowledge about cancer and yield useful therapeutic targets. This protection could be secondary to many factors, including specific mechanisms of DNA repair, cell cycle regulation, metabolism of carcinogens, angiogenesis, apoptosis or immunological response among others (Fig. 1).

Although it seems reasonable that selection of individuals presenting decreased risk of developing cancer might have been favoured by evolution to a certain extent, the question remains if such subjects do really exist. Looking to other diseases, we can find relevant examples, some of which have been successfully identified through selection of extreme phenotypes. One outstanding paradigm is the identification of alterations in the gene encoding the chemokine coreceptor CCR5 that confer complete protection against certain strains of the human immunodeficiency virus (HIV) [59,60]. Since CCR5 mutations were not associated with phenotypic abnormalities, they were only identified after observing that some individuals highly exposed to HIV never developed the infection [61]. The study of those individuals allowed the discovery of a relevant target in HIV investigation.

Returning to cancer, some preclinical studies support the biological plausibility of the existence of protective mechanisms. A relevant example is the creation of a “super-p53” mouse, carrying supernumerary p53 copies, which shows a decreased risk of developing chemically induced tumours [62]. Although these mice and other similar models [63,64] have been artificially developed, their relatively normal phenotypes raise the question of whether similar phenomena could spontaneously occur and be selected for in nature. It is known that small malignant tumours identified in autopsies or by high-performance imaging tests outnumber the quantity of clinically overt cancers [65]. However, it is unknown whether this is caused by protective mechanisms or if it is due to other reasons, such as differences in tumour biology. The same questions arise when we analyse the different sensitivity of individuals to carcinogens such as tobacco or the variations in clinical aggressiveness of tumours in different patients: while large differences exist, it is difficult to establish their causes.

Even if we assume that individuals bearing protection against developing cancer may exist, the question of how to identify them remains. Several studies have assessed the protective role of enzymatic polymorphisms with inconclusive results, as reviewed elsewhere [18]. Most compared cancer patients with normal individuals. Such a design offers the disadvantage that, rather than true protection against cancer, normal subjects may just present absence of disease with a normal risk of developing it. Instead, protection can be expected in individuals not developing cancer despite presenting an increased risk. This approach has been successfully evaluated in the study of polymorphisms of detoxifying enzymes, using elderly individuals not presenting cancer as controls, sometimes even when they were smokers [66–69]. While these studies truly select extreme phenotypes, their

design could be improved by the use of high-throughput techniques, which study multiple potential causes, rather than just a few; and by selecting individuals with even more characteristic phenotypes, i.e., a markedly reduced individual or familial risk of developing cancer. Families presenting very low cancer incidence across several generations might show reduced familial risk. Subjects with high-risk cancer factors, such as extensive exposure to carcinogens or cancer familial syndromes that do not develop the disease, or in whom development is significantly delayed, may present reduced individual risk. Some sensitive models that have already been proposed might be familial adenomatous polyposis [18] or hereditary non-polyposis colorectal cancer [70], because they present high penetrance and, therefore, the likelihood that an affected individual will not develop cancer is low. The study of heavy smokers that do not develop cancer at an advanced age and of young smokers who develop the disease might also yield relevant information on cancer protective mechanisms and tumour development and growth. Different combinations of these strategies could be developed. Aggregation of similar phenotypes within one family would further support that their underlying cause is not random. Theoretically, the likelihood of finding clinically relevant results should be directly related to the magnitude of the discrepancy between the estimated risk of presenting cancer and the observed phenotype.

Conclusions and future directions

Extreme phenotype selection is a well defined biological concept that describes how environmental pressures favour, among traits exhibiting diverse phenotypes, the fittest to overcome the hazards encountered. Although it seems logical to study the favoured phenotypes to determine their mechanisms of success, the use of this strategy has often been restricted to obvious cases or to isolated observations made by discerning clinicians. The lack of a systematic approach to the identification of extreme phenotypes is what has probably precluded most of them from being studied. Therefore, a consistent methodology should be developed to maximize the potential benefit of this strategy. Specifically in the field of oncology, we propose the creation of databases compiling patient samples, together with clinical and epidemiological data from individuals presenting relevant phenotypes. The collection and intensive study of these extreme cases, rather than constituting a mere list of oddities, might provide excellent hunting grounds to discover Achilles heels of cancer.

This methodology does, however, raise some issues. Principal among these are the questions of how to classify phenotypes into quantitative groups [16] in order to define what constitutes an extreme phenotype, and how to identify them. Definition of phenotypes that might be clinically relevant and that can be found in clinical practice should probably be performed by consensus panels of experienced clinicians under the coordination of medical societies or cooperative groups. Case selection should be approached by training and creating awareness among medical specialists. Cancer patients with very favourable evolution, or with extreme responses or toxicities following therapy could be selected in oncology centres relatively easily, since their number is small and their characteristics are unusual. Individuals presenting cancer familial syndromes not developing cancer despite their high risk could be selected through genetic counselling units. In other cases, more complex epidemiological studies might be required to identify relevant discordances between the expected and the observed phenotypes. Since it is unlikely that an adequate number of subjects bearing extreme phenotypes can be detected in a limited number of institutions, this will require the collaboration of large cooperative groups, ideally at an international level.

Another issue is what variables should be studied in these subjects. Characteristic phenotypes may be caused by host or tumour factors, as well as by external causes. Therefore, all of these should be analyzed, and ideally samples from the tumour and from

the host's normal tissue should be collected, along with clinical and epidemiological data. The type and quantity of the samples should allow a wide variety of studies to be performed, including screening of cell genomes, epigenomic changes and transcriptomes through high-throughput techniques, as mentioned above, such as Genome Wide Association Studies (GWAS) or full genome sequencing. The samples should also permit additional studies to be performed in the future using techniques that are not yet available. At a minimum, whole blood, including serum and DNA and, in the case of cancer patients, fresh and paraffin-embedded tumor tissue, should be collected. The obtention of sequential samples, linked to relevant clinical events (e.g: start of a new treatment, evaluation of clinical response or observation of an unexpected toxicity) might provide relevant information about the evolution of different biomarkers in such situations. Even though this methodology does not avoid the inherent problems of sample collection, it does limit dramatically the number of samples to obtain and to study. Unavailability of samples might be an important limitation, which could be overcome by prospective collection of cases. The importance of an adequate infrastructure to collect and store samples and to manage the database cannot be overemphasized. Lastly, ethical issues may be another point of concern, since the use of biologic material is subject to strict regulations. Ethical boards must collaborate to make these studies feasible without compromising the rights of the participants. The possibility of contacting subjects in order to obtain more information or to perform functional studies should be taken into consideration. Well designed informed consent processes and prospective data collection should minimize these problems.

In conclusion, the selection and the study of extreme, clinically relevant phenotypes is an efficient strategy to identify their underlying causes. The creation of collaborative databases compiling biological samples and clinical information from such phenotypes might increase our knowledge of cancer and provide new therapeutic strategies. This will require close and continued collaboration between clinicians, who must identify appropriate cases, and basic scientists, who should perform adequate studies to identify and integrate the relevant targets. Even in the current age of modern molecular biology, clinical observation should remain a preferred strategy to generate hypothesis than intellectual speculation.

References

1. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–1320. [PubMed: 16255080]
2. Freimer N, Sabatti C. The human phenome project. *Nat Genet* 2003;34:15–21. [PubMed: 12721547]
3. Zbuk KM, Eng C. Cancer phenomics: RET and PTEN as illustrative models. *Nat Rev Cancer* 2007;7:35–45. [PubMed: 17167516]
4. Fernandez-Ricaud L, Warringer J, Ericson E, et al. PROPHECY: a yeast phenome database, update 2006. *Nucleic Acids Res* 2007;35:D463–467. [PubMed: 17148481]
5. Grubb SC, Maddatu TP, Bult CJ, Bogue MA. Mouse phenome database. *Nucleic Acids Res* 2009;37:D720–730. [PubMed: 18987003]
6. de la Cruz N, Bromberg S, Pasko D, et al. The Rat Genome Database (RGD): developments towards a phenome database. *Nucleic Acids Res* 2005;33:D485–491. [PubMed: 15608243]
7. Singer E. “Phenome” project set to pin down subgroups of autism. *Nat Med* 2005;11:583. [PubMed: 15937456]
8. Groth P, Pavlova N, Kalev I, et al. PhenomicDB: a new cross-species genotype/phenotype resource. *Nucleic Acids Res* 2007;35:D696–699. [PubMed: 16982638]
9. Johnston PG, Fisher ER, Rockette HE, et al. The role of thymidylate synthase expression in prognosis and outcome of adjuvant chemotherapy in patients with rectal cancer. *J Clin Oncol* 1994;12:2640–2647. [PubMed: 7989939]
10. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503–511. [PubMed: 10676951]

11. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–536. [PubMed: 11823860]
12. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;361:865–872. [PubMed: 12642066]
13. Editor. Freely associating. *Nat Genet* 1999;22:1–2. [PubMed: 10319845]
14. Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS Med* 2009;6:e22. [PubMed: 19192942]
15. Gambaro G, Anglani F, D'Angelo A. Association studies of genetic polymorphisms and complex disease. *Lancet* 2000;355:308–311. [PubMed: 10675088]
16. Nebert DW. Extreme discordant phenotype methodology: an intuitive approach to clinical pharmacogenetics. *Eur J Pharmacol* 2000;410:107–120. [PubMed: 11134663]
17. Perez-Gracia JL, Ruiz-Ilundain MG. Cancer protective mutations: looking for the needle in the haystack. *Clin Transl Oncol* 2001;3:169–171.
18. Perez-Gracia JL, Gloria Ruiz-Ilundain M, Garcia-Ribas I, Maria Carrasco E. The role of extreme phenotype selection studies in the identification of clinically relevant genotypes in cancer research. *Cancer* 2002;95:1605–1610. [PubMed: 12237932]
19. Reichardt JK, Mehrian-Shai R. The future of the human SNP identification: which individuals to sequence? *Proc Natl Acad Sci U S A* 2009;106:E50. [PubMed: 19416875]
20. Lin DW, Porter M, Montgomery B. Treatment and survival outcomes in young men diagnosed with prostate cancer: a population-based cohort study. *Cancer* 2009;115:2863–2871. [PubMed: 19466697]
21. Online Mendelian Inheritance in Man, OMIM™. Available at <http://www.ncbi.nlm.nih.gov/omim/>
22. Erdheim J. Zur normalen und pathologischen Histologie der glandula Thyroidea, Parathyroidea und Hypophysis. *Beitr Pathol Anat* 1903;33:1–234.
23. Rossier P, Dressler M. Familiäre Erkrankung innersekretorischer drüsen kombiniert mit Ulcuskrankheit. *Schweiz Med Wochenschr* 1939;69:985–990.
24. Schmid JR, Labhart A, Rossier PH. Relationship of multiple endocrine adenomas to the syndrome of ulcerogenic islet cell adenomas (Zollinger-Ellison). Occurrence of both syndromes in one family. *Am J Med* 1961;31:343–353. [PubMed: 13747921]
25. Larsson C, Skogseid B, Oberg K, et al. Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. *Nature* 1988;332:85–87. [PubMed: 2894610]
26. Chandrasekharappa SC, Guru SC, Manickam P, et al. Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* 1997;276:404–407. [PubMed: 9103196]
27. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 1990;250:1684–1689. [PubMed: 2270482]
28. Miller RW. Deaths from childhood cancer in sibs. *N Engl J Med* 1968;279:122–126. [PubMed: 5655197]
29. Li FP, Fraumeni JF Jr. Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. *J Natl Cancer Inst* 1969;43:1365–1373. [PubMed: 5396222]
30. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 1990;250:1233–1238. [PubMed: 1978757]
31. Tuchman M, Stoeckeler JS, Kiang DT, et al. Familial pyrimidinemia and pyrimidinuria associated with severe fluorouracil toxicity. *N Engl J Med* 1985;313:245–249. [PubMed: 2989687]
32. Diasio RB, Beavers TL, Carpenter JT. Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity. *J Clin Invest* 1988;81:47–51. [PubMed: 3335642]
33. Van Kuilenburg AB, Vreken P, Beex LV, et al. Heterozygosity for a point mutation in an invariant splice donor site of dihydropyrimidine dehydrogenase and severe 5-fluorouracil related toxicity. *Eur J Cancer* 1997;33:2258–2264. [PubMed: 9470816]
34. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500. [PubMed: 15118125]

35. Lynch TJ, Bell DW, Sordella R, et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004;350:2129–2139. [PubMed: 15118073]
36. Herbst RS, Prager D, Hermann R, et al. TRIBUTE: a phase III trial of erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer. *J Clin Oncol* 2005;23:5892–5899. [PubMed: 16043829]
37. Gatzemeier U, Pluzanska A, Szczesna A, et al. Phase III study of erlotinib in combination with cisplatin and gemcitabine in advanced non-small-cell lung cancer: the Tarceva Lung Cancer Investigation Trial. *J Clin Oncol* 2007;25:1545–1552. [PubMed: 17442998]
38. Giaccone G, Herbst RS, Manegold C, et al. Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial – INTACT 1. *J Clin Oncol* 2004;22:777–784. [PubMed: 14990632]
39. Herbst RS, Giaccone G, Schiller JH, et al. Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial – INTACT 2. *J Clin Oncol* 2004;22:785–794. [PubMed: 14990633]
40. Zhang G, Nebert DW, Chakraborty R, Jin L. Statistical power of association using the extreme discordant phenotype design. *Pharmacogenet Genomics* 2006;16:401–413. [PubMed: 16708049]
41. Silberstein E, Walfisch S, Lupu L, Sztarkier I. Twelve-year survival after the diagnosis of locally advanced carcinoma of the pancreas: a case report. *J Surg Oncol* 2000;75:142–145. [PubMed: 11064396]
42. Spinelli GP, Zullo A, Romiti A, et al. Long-term survival in metastatic pancreatic cancer. A case report and review of the literature. *JOP* 2006;7:486–491. [PubMed: 16998247]
43. Miyaji M, Ogoshi K, Kajiura Y, et al. A case of advanced gastric cancer with liver metastasis with no recurrence and long survival. *Gan To Kagaku Ryoho* 1996;23:915–918. [PubMed: 8678542]
44. Shibata N, Tamai M, Ikeda K. Case report of long-term survivor of advanced gastric cancer associated with peritoneal dissemination successfully treated with cancer chemotherapy. *Gan To Kagaku Ryoho* 1995;22:119–122. [PubMed: 7826066]
45. Fukasawa T, Shoji T, Gotoh H, Taniwaka K. A long-term survivor with stage IV gastric cancer due to postoperative weekly paclitaxel and 5'-DFUR combination therapy. *Gan To Kagaku Ryoho* 2006;33:235–238. [PubMed: 16484863]
46. Wada T, Iwanaga T, Koyama H, et al. A case report of long-term survival in advanced gastric cancer with multiple liver metastases treated with mitomycin C and tegafur therapy after palliative gastrectomy. *Gan To Kagaku Ryoho* 1988;15:353–356. [PubMed: 2449135]
47. Mukai M, Tokunaga N, Yasuda S, et al. Long-term survival after immunochemotherapy for juvenile colon cancer with peritoneal dissemination: a case report. *Oncol Rep* 2000;7:1343–1347. [PubMed: 11032941]
48. Seto T, Seto M, Ono K, et al. A long-term survivor of extensive-stage small cell lung cancer. *Gan To Kagaku Ryoho* 2000;27:135–138. [PubMed: 10660747]
49. Mall JW, Philipp AW, Mall W, Pollmann C. Long-term survival of a patient with small-cell lung cancer (SCLC) following treatment with thalidomide and combination chemotherapy. *Angiogenesis* 2002;5:11–13. [PubMed: 12549855]
50. Izumi Y, Mukai M, Kikuchi K, Kobayashi K. Long-term survival after incomplete resection of immunohistochemically diagnosed TON1 lung cancer: report of a case. *Surg Today* 2006;36:270–273. [PubMed: 16493539]
51. Mac Manus MP, Matthews JP, Wada M, et al. Unexpected long-term survival after low-dose palliative radiotherapy for non-small cell lung cancer. *Cancer* 2006;106:1110–1116. [PubMed: 16432830]
52. Dutcher JP, Wiernik PH. Long-term survival of a patient with multiple myeloma: a cure? A case report. *Cancer* 1984;53:2069–2072. [PubMed: 6704893]
53. Yamshchikov GV, Mullins DW, Chang CC, et al. Sequential immune escape and shifting of T cell responses in a long-term survivor of melanoma. *J Immunol* 2005;174:6863–6871. [PubMed: 15905528]

54. Le Gal FA, Widmer VM, Dutoit V, et al. Tissue homing and persistence of defined antigen-specific CD8+ tumor-reactive T-cell clones in long-term melanoma survivors. *J Invest Dermatol* 2007;127:622–629. [PubMed: 17039243]
55. Yamshchikov G, Thompson L, Ross WG, et al. Analysis of a natural immune response against tumor antigens in a melanoma survivor: lessons applicable to clinical trial evaluations. *Clin Cancer Res* 2001;7:909s–916s. [PubMed: 11300491]
56. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity* 2004;21:137–148. [PubMed: 15308095]
57. Perez-Gracia JL, Prior C, Guillen-Grima F, et al. Identification of TNF-alpha and MMP-9 as potential baseline predictive serum markers of sunitinib activity in patients with renal cell carcinoma using a human cytokine array. *Br J Cancer* 2009;101:1876–1883. [PubMed: 19904265]
58. Sabatino M, Kim-Schulze S, Panelli MC, et al. Serum vascular endothelial growth factor and fibronectin predict clinical response to high-dose interleukin-2 therapy. *J Clin Oncol* 2009;27:2645–2652. [PubMed: 19364969]
59. Liu R, Paxton WA, Choe S, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* 1996;86:367–377. [PubMed: 8756719]
60. Quillent C, Oberlin E, Braun J, et al. HIV-1-resistance phenotype conferred by combination of two separate inherited mutations of CCR5 gene. *Lancet* 1998;351:14–18. [PubMed: 9433423]
61. Rowland-Jones S, Sutton J, Ariyoshi K, et al. HIV-specific cytotoxic T-cells in HIV-exposed but uninfected Gambian women. *Nat Med* 1995;1:59–64. [PubMed: 7584954]
62. Garcia-Cao I, Garcia-Cao M, Martin-Caballero J, et al. “Super p53” mice exhibit enhanced DNA damage response, are tumor resistant and age normally. *EMBO J* 2002;21:6225–6235. [PubMed: 12426394]
63. Yu Q, Geng Y, Sicinski P. Specific protection against breast cancers by cyclin D1 ablation. *Nature* 2001;411:1017–1021. [PubMed: 11429595]
64. Cui Z, Willingham MC, Hicks AM, et al. Spontaneous regression of advanced cancer: identification of a unique genetically determined, age-dependent trait in mice. *Proc Natl Acad Sci U S A* 2003;100:6682–6687. [PubMed: 12724523]
65. Black WC, Welch HG. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med* 1993;328:1237–1243. [PubMed: 8464435]
66. Belogubova EV, Togo AV, Karpova MB, et al. A novel approach for assessment of cancer predisposing roles of GSTM1 and GSTT1 genes: use of putatively cancer resistant elderly tumor-free smokers as the referents. *Lung Cancer* 2004;43:259–266. [PubMed: 15165083]
67. Belogubova EV, Kuligina E, Togo AV, et al. ‘Comparison of extremes’ approach provides evidence against the modifying role of NAT2 polymorphism in lung cancer susceptibility. *Cancer Lett* 2005;221:177–183. [PubMed: 15808403]
68. Belogubova EV, Ulibina YM, Suvorova IK, et al. Combined CYP1A1/GSTM1 at-risk genotypes are overrepresented in squamous cell lung carcinoma patients but underrepresented in elderly tumor-free subjects. *J Cancer Res Clin Oncol* 2006;132:327–331. [PubMed: 16416283]
69. Belogubova EV, Togo AV, Kondratieva TV, et al. GSTM1 genotypes in elderly tumour-free smokers and non-smokers. *Lung Cancer* 2000;29:189–195. [PubMed: 10996421]
70. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. *Nature* 2003;422:835–847. [PubMed: 12695777]



Fig. 1.

Flow chart representing the process from selection of extreme phenotypes to determination of their molecular causes. Although the observation of extreme phenotypes clearly suggests that there is indeed a needle in the haystack to find, it offers limited clues about the nature of the causative mechanisms. Nonetheless, limiting the number of cases and tissue samples to analyze provides a considerable degree of simplicity that to date has been largely neglected. Initial screening of factors that might explain the phenotypes might be performed either by studying specific major molecular pathways or applying high throughput techniques. Once these potential clues have been identified, the specific underlying mechanisms should be addressed in detail