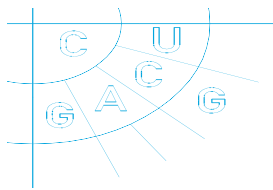


**WP2: Review of advances in toxicogenomics with relevance for the reduction of uncertainties in EMF related cancer risk assessment**

**Top-down approach: Overall assessment of the potential contribution of toxicogenomics (genomics, transcriptomics) to close knowledge gaps in current EMF cancer risk assessment**

February 2008



**COLOGNE  
CENTER FOR  
GENOMICS**

Prof. Dr. Peter Nuernberg  
Zentrum fuer Funktionelle Genomforschung (ZFG)  
Universität zu Köln  
Zùlpicher Strasse 47  
50674 Köln  
Germany

**Contents**

<b>Preface</b>	<b>3</b>
<b>Epidemiologic Studies of Extremely Low-Frequency Electromagnetic Fields</b>	<b>3</b>
<b>Quantitative and country-specific estimate of public health impact of extremely low-frequency electromagnetic fields</b>	<b>9</b>
<b>Uncertainty</b>	<b>11</b>
<b>Closing the Knowledge Gaps</b>	<b>12</b>
<b>Next-Generation Sequencing Technologies</b>	<b>16</b>
<b>References</b>	<b>30</b>

## Preface

Electric power is an essential commodity of the developed world, and is critical to the continuing progress of our technology-based society, as well as to the growth of less privileged societies. In contrast to its overwhelming benefits, there is a suspicion that the magnetic component of the electromagnetic fields (EMFs) associated with power distribution and electrical appliances have adverse health effects, especially a small increased incidence of childhood leukaemia. The possibility that environmental EMFs represent a health hazard has serious economic implications for government, the electricity industry and society, as well as raising several profound scientific challenges, including, in particular, biophysical mechanisms, experimental replication and scientific uncertainty. These challenges have to be explored by basic research on the biological effects of extremely low-frequency electromagnetic fields (ELF-EMFs).

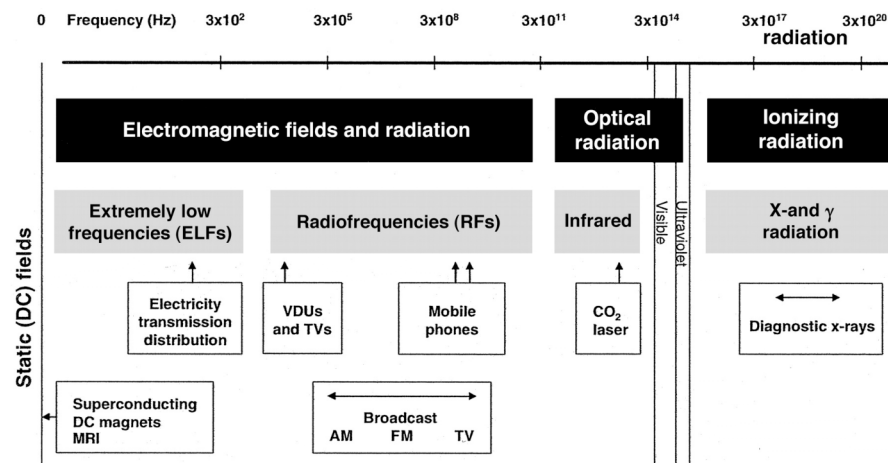


Figure 1: Electromagnetic spectrum

Exposure to electric and magnetic fields from 0 to 300 GHz has been increasing greatly as countries increase their capacity to generate and distribute electricity and take advantage of the many new technologies, such as telecommunications, to improve lifestyle and work efficiency (figure 1). Extremely low frequency (ELF) magnetic fields span the range from 3 to 300 Hz.

## Epidemiologic Studies of Extremely Low-Frequency Electromagnetic Fields

Given the ubiquitous nature of extremely low-frequency (ELF) electromagnetic fields (EMFs), there is concern regarding their potential to adversely affect the health. Numerous health effects have been studied in relation to the EMF exposure: cancer, reproductive disorders, as well as neurodegenerative and cardiovascular diseases. Cancer, especially childhood cancer, has received the most attention.

A number of reviews on the potential on EMF to cause damage to health have been published (IARC 2002; National Institute of Environmental Health Sciences [NIEHS] 1999). The general consensus is that cellular effects do not occur with exposure below 100  $\mu\text{T}$ . Also, expect for a very few animal studies that suggest adverse effect of EMF, these studies have been largely negative.

The EMFs are imperceptible, ubiquitous, have multiple sources, and can vary greatly over time and short distances (Bracken et al. 1993). In the absence of (a) biological mechanism to implicate one or more specific field parameters, the exposure assessment of EMFs has varied over the years. The epidemiologic studies in the last decade have employed improved exposure assessment methods. Most of the epidemiological studies discussed below use the time weighted average (TWA) measurements to characterize the exposure. Furthermore, with the technological advances and increased samples size studies, the higher exposures, i.e.,  $>0.4\mu\text{T}$ , are being explored. Although the epidemiologic evidence is not conclusive, it is generally agreed that the possibility of a causal association between the EMF and the adverse health outcomes cannot be excluded and that the epidemiologic studies of childhood leukemia provide the strongest evidence of an association.

The epidemiologic evidence is a major contributor to the understanding of the potential effects of EMF on health. The International Agency for Research on Cancer (IARC) classified EMF as a Group 2B or possible human carcinogen (IARC 2002); this classification was mostly based on the consistent epidemiologic evidence of an association between exposure of these fields and childhood leukemia and of laboratory studies in animals and cells which were not supportive of exposure to EMF causing cancer. Although the body of evidences always considered as a whole, on the basis on the weight of evidence approach and incorporating different lines of scientific enquiry, epidemiologic evidence, as the most relevant, is given the maximum weight.

The epidemiologic data is routinely critically assessed as shed light on the potential of EMF to cause harm to health. The next chapter provides an up-to-date review of the epidemiologic evidence and accompanying methodological concepts.

## Leukemia

The leukemias are cancers of the blood and bone marrow. The classification of leukemias is conventionally based on the origin of the cell types (lymphocytes, myelocytes, monocytes)

and rate at which the disease progresses (acute and chronic). The age distribution of leukemias is bimodal, with a first peak occurring at 4y, a decline at 15-29 y, and then a slow rise throughout the rest of life. Leukemia is the most common childhood malignancy, constituting more than half of all childhood cancers. Acute lymphocytic leukemias (ALLs) account for 75% of all the cases of childhood leukemia; the common types in adults are acute myeloid leukemia (AML) and chronic lymphocytic leukemia (CLL). The age-standardized rate of leukemia for children younger than 15 y has been estimated to be 3.5 per 100,000 per year for females and 4.2 per 100,000 per year for males in the developed world (IARC 2000). For adults, the rates for each type of leukemia have changed very little over the past 20 years. For children however, incidence rates have risen slightly; similar or more pronounced increases have occurred for other childhood cancers (Steliarova-Fouchier et al. 2004). Some of these increases might be attributable to the improved diagnosis. The environmental, occupational, and genetics factors have been associated with one or more types of leukemia. Increased risks have been associated with radiation exposure (all types of leukemia except CLL), cigarette smoking (parental smoking for childhood leukemia), and exposure to the human T-cell leukemia virus type one (HTLV-1) (associated with adult T-cell leukemia and lymphoma). Occupational exposures associated with increased risk are exposure to benzene (AML), manufacturing processes of styrene and butadiene production (lymphoid leukemia), and petroleum refining (AML and to a lesser extent CLL). The genetic risk factors include the chromosomal and congenital abnormalities Faconi's anemia, along with Down's, Bloom's, Klinefelter's, and trisomy G syndromes (Mezei et al. 2002). In general, the relationship between socioeconomic status (SES) and childhood leukemia is weak and inconsistent (Poole et al. 2006). The causes of leukemia, especially childhood leukemia, are not well understood. In addition, the presently known or suspected risk factors are likely to account for only a small proportion of all cases (Kheifets et al. 2005).

## Childhood Leukemia

### - Residual Exposures / Wire Codes

In the first study by Wertheimer and Leeper in 1979, each subjects exposure was categorized on the basis of (a) type of electric utility wiring adjacent to the residence, i.e., transmission lines, primary and secondary distribution lines of various types, and (b) the distance from that wiring to the residence. This exposure is referred to as the "wire code". Wertheimer and Leeper initially used a dichotomous exposure classification wherein subjects were classified as living in either high- or low-current configuration (HCC or LCC) homes. Later they

introduced a finer degree of resolution into their wire code classification system: very low-current configuration (VLCC), ordinary low-current configuration (OLCC), ordinary high-current configuration (OHCC), and very high-current configuration (VHCC). Savitz et al. (1988) further refined the original wire code configuration classification scheme. London et al. (1991) conducted an incidence study of childhood leukaemia similar in its basic design by Savitz et al., but with 24 h bedroom measurements added to the exposure assessment. Similar results emerged: (a) the residential magnetic field strength correlated weakly with wire code and (b) a trend of increased leukaemia risk across wire code, with elevated leukaemia risk among the VHCC sample relative to the referent.

- Exposure based on Distance and Calculated Fields

There are a handful of studies on childhood leukaemia using distance from residence to power lines as the exposure of interest. The results of these studies have not been wholly consistent. The studies using calculated field measurements seek to capture historical exposure levels. The exposure to EMF is most often assessed by using power line load data and specifications for power lines that are specific to the time period of interest. Both Feychting and Ahlbom (1993) and Tynes and Halderson (1997) used a nested case-control study approach to identify leukaemia cases within defined cohorts during specific time periods and calculated the field measurements. Feychting and Ahlbom (1993) showed that the risk of childhood leukaemia was elevated for children exposed to 0.3 $\mu$ T or greater; however, Tynes and Halderson (1997) showed that there was no elevated risk of leukaemia among those exposed to 0.14  $\mu$ T and greater. No analysis for higher cut points was presented. A study from Finland using similar design showed a nonsignificant elevated risk of childhood leukaemia among those exposed to 0,2  $\mu$ T and greater (Verkasalo 1993).

- Household and Personal Exposure

The majority of epidemiologic studies on EMF and childhood leukemia use residential EMF exposure measurements; they have been either spot measurements (Savitz et al. 1988) or area (mostly bedroom) magnetic field measurements of 24 h and longer (Green et al. 1999). Although most of these studies observed elevated odds ratios (OR), for many, the confidence intervals (CIs) around these estimates of effect are large due to small numbers of highly exposed subjects. Given the rarity of both the disease and the high exposure, securing enough highly exposed cases, needed for statistically stable estimates, is quite difficult. Only two studies employed the use of personal EMF meters to measure exposure (Green et Al. 1999; McBride et al. 1999). In both of these studies, there was no indication of elevated risk of leukemia due to high personal exposure. Although there is evidence that there is a close

correlation between household exposure and personal measurements (UK Childhood Cancer Study Investigators 1999), use of personal measurements in case-control studies might be biased as disease status might effect personal exposure measurements. Personal measurements are even more problematic in case-control studies of children, as exposure at young age is quite age dependent, and measurements made few years after the etiologically relevant time period might different than during the time period of interest.

#### - Pooled Analyses

Two pooled analyses represent the most powerful attempt so far to provide a cohesive assessment of the epidemiologic data of EMF and childhood leukemia (Ahlbom et al. 2000; Greenland et al. 2000). These analyses, although focusing on a largely overlapping but distinct set of studies, come to similar conclusions. In the pooled analyses by Greenland et al. (2000), 12 studies using measured or calculated fields were identified; the study included a total of 2656 cases and 7084 controls. For this analysis, the metric of choice was the time-weighted average. The estimated OR for childhood leukemia was 1.68 (95% CI 1.231-2.31) for exposures greater than 0.3  $\mu\text{T}$  as compared with exposures less than 0.1  $\mu\text{T}$ , controlling for age, sex, and study. Using more stringent inclusion criteria, Album et al. (2000) included nine studies using measured and calculated fields. There were a total of 3203 cases and 10338 controls in the pooled sample. Using the geometric mean as the metric of choice, the estimated OR for childhood leukemia was 2.00 (95% CI 1.27-3.13) for exposures greater than or equal to 0.4  $\mu\text{T}$  as compared with exposures less than 0.1  $\mu\text{T}$ , controlling for age, sex, and East and West (in Germany only).

#### - Electric Appliance Exposure

Studies that have evaluated the risks of childhood leukemia (London et al. 1991; Greenland et al. 2000) associated with the use electric appliance, used interviews of mothers to assess exposure information. Positive findings were observed in only a couple of these studies. There was an elevated risk of leukemia, but no evidence of dose-response relationship, among children watching black-and-white television in one study; another study observed a rise in leukemia with increasing number of hours children watched television, regardless of the child's distance from the television set (Hatch et al. 1998). Risks were increased for postnatal exposure to a few other appliances in one study, with no dose-response relationship. Postnatal use of electric blankets and their dryers (London et al. 1991; HATH et al. 1998) was linked with modestly elevated risks in more than one study, but there was no evidence of dose-response relationships. Overall, the small number of studies, evaluation of one appliance at a

time leading to a very large misclassification, and the absence of good measurement data within the studies preclude the straightforward interpretation of results.

#### - Parental Exposure

Of the studies that have examined parental occupation and childhood cancer (Savitz 1990) none have found any association with parental exposure to EMF. London et al. (1991) reported an association between a mother's exposure to nonionizing radiation during pregnancy and the risk of childhood leukemia. However, the exposure question did not focus on EMF and is likely to reflect exposures to higher than power frequencies. Two studies (Hatch et al. 1990; Savitz et al. 1990) observed small increase in the risk of childhood leukemia with the prenatal use of blankets. In general, on the basis of these studies, there is little evidence of electrical risk of leukemia in offspring associated with mothers' prenatal use of other types of electrical appliances.

#### Adult Leukemia

##### - Residential Exposures

The first study of residential exposures and adult cancer was also conducted by Wertheimer and Leeper (1982). Although the focus was on total cancers, leukemia did not contribute to the elevated risk found for all cancers combined. Several other studies have attempted to examine the risk of adult leukemia in populations residing near power lines (Coleman et al. 1989; Schreiber 1993). These studies are based on small numbers, low potential exposures, and very crude exposure estimation methods. The overall results are negative, with a few hints of a small, nonsignificant elevation of risk in some studies. Since then, other studies on adult leukemia and residential EMF exposure have shed more light on the plausible relationship, with a follow-up study by Feychting et al. (1997) and studies by Verkasalo and coworkers (Verkasalo 1996). A small increased risk for all leukemia was seen in some of these studies.

##### - Electric appliance Exposure

In a case-control study of adult AML and CML in Los Angeles County that used cases and matched neighborhood controls (Preston-Martin 1988), use of electric blankets was not related to leukemia risk. Lovely et al. (1994) examined adult leukemia risk and personal appliance use, on the basis of the study of Severson et al. (1988). Most noteworthy in this analysis were a small elevation in risk for ever/never use of electric razors (relative risk (RR) 1.3, 95% CI 0.8-2.2), and a larger increase in risk with duration of daily use (RR 2.4, 95% CI 1.1-5.5 for highest relative to the lowest category). However, it appears that this finding is due to bias in the proxy-reported information and that there is in actuality no association between leukemia risk and either use versus no use or duration of use of electric razors.

Studies on occupational exposure to EMF and adult leukemia have job titles mostly linked with cancer incidence or mortality. In a meta-analysis of this literature by Kheifets et al. (1997), there was a small increased risk of leukemia associated with work in electric occupations. However, jobs thought to have higher exposure (welders, electricians, linemen, and power plant operators) did not have higher risks than electric workers who generally have lower exposures (installers, engineers, and television or radio repairmen). Several occupational cohort and case-control studies have been published since this meta-analysis.

In summary, there is enough consistency in the results from the large body of high quality research on childhood leukemia to reach a general agreement that postnatal exposures above  $0.4\mu\text{T}$  are associated with an elevated risk of childhood leukemia as compared with exposures less than  $0.1\mu\text{T}$ ; this RR has been estimated at around 2, on the basis of two large pooled analyses. This is unlikely to be due to chance but may be partly due to bias. For other exposure measures and for adult leukemia the evidence is considerably weaker.

#### Brain Cancer

There are few clearly recognized risk factors for brain cancer. In addition to recognized risk factors for brain cancer. In addition to ionizing radiation, occupational exposures that have been linked to adult brain cancer risk in some studies including, organic solvents, and pesticides. Excess risk have been found among farmers and painters, but the specific agents responsible for these risks have not been identified. Brain cancers have also been linked to exposure to N-nitroso compounds and halomethanes. People with a family history of cancer, certain genetic diseases, or head injury may have an elevated risk of brain tumors. Tobacco and alcohol use have not been linked to brain cancer risk. Thus, the cause of brain cancer are not well understood. Presently known or suspected factors appear to account for only a small proportion of all the cases. Several of the childhood cancer studies described earlier examined childhood brain cancer in addition to leukemia and other childhood tumors.

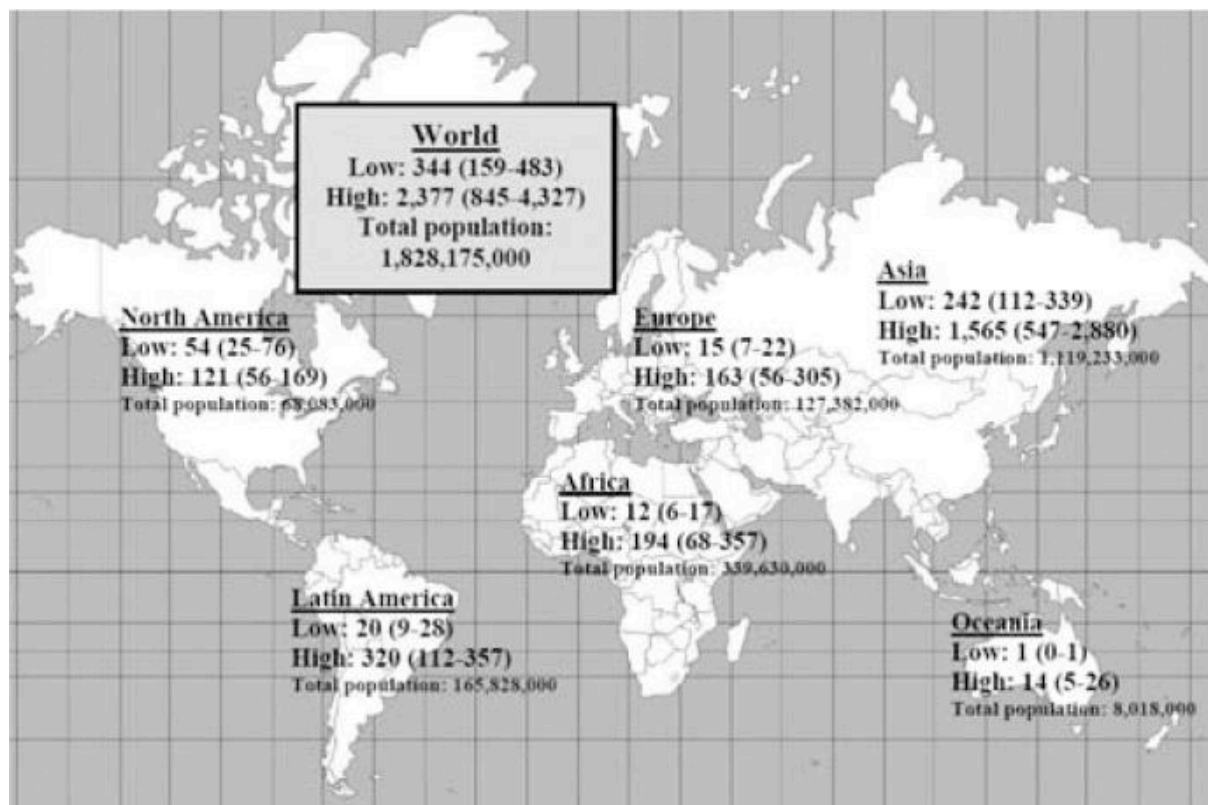
#### **Quantitative and country-specific estimate of public health impact of extremely low-frequency electromagnetic fields**

Kheifets et al. (2006) have quantitatively estimated the public health impact of extremely low-frequency electromagnetic fields. Five extensive surveys have been conducted to evaluate ELF exposures of the general population (Brix et al. 2001; Decat et al. 2005; Electric Power Research Institute (EPRI) 1993; EMF Rapid Program 1998; Yang et al. 2004). Estimating

exposures using the case exposures from case-control studies allowed them to look at a broader spectrum of countries and results.

In addition to assuming no significant difference in the exposure distributions based on exposure surveys and case-control studies, they have assumed that exposures obtained using personal measures are equivalent to those from household measurements, regardless of the length of time of measurement and regardless of whether they were for children or adults. This last assumption was tested using average and geometric mean household measurements (bedroom and home) from the U.S. Rapid Survey (EMF Rapid Program 1998), which included a sample of both children and adults. A two-sample *t*-test with equal variances comparing the distributions of the log of the average measurements of total home exposure showed no significant difference between the distributions for children (< 15 years) and adults ( $\geq 15$  years). Also, no significant difference was found between the adult and child distributions of the log of the geometric mean of total household exposure. The tests described above assumed that the log of the exposure is normally distributed, an assumption that was tested and found to fit the data well. Additional tests were conducted using either the Pearson chi-square or Fisher's exact tests to compare the percent of people  $> 0.3$  or  $0.4 \mu\text{T}$  for children and adults; these tests showed no significant difference in the percent of children and adults above these exposure levels for total household ELF exposure.

Worldwide estimates were performed. The attributable numbers (ANs) of leukemia cases were calculated for the scenarios of eliminating arithmetic mean exposure  $> 0.3 \mu\text{T}$  and of eliminating geometric mean exposure  $\geq 0.4 \mu\text{T}$ . This computation was made for regions around the world, then added to obtain a global estimate. To compute these regional estimates, they used the lowest level and highest exposure levels estimated from the countries in that region to obtain a regional range. Where no information was available from any countries in the region, the lowest and highest exposure levels overall have been used. Making certain assumptions about possible exposure reduction scenarios, they provide a range of attributable numbers thought to be most useful for administrations. Reducing exposures to  $< 0.3 \mu\text{T}$  (arithmetic mean) or  $0.4 \mu\text{T}$  (geometric mean) results in a number of averted cases (assuming causality) ranging from 100 to 2,400 cases annually worldwide.



**Figure 2:** Estimated number and range of worldwide and regional cases of childhood leukemia among children < 14 years of age that are possibly attributable to EMF arithmetic mean exposure > 0.3  $\mu$ T (and the corresponding derived 95% CI).

To estimate the potential public health impact, it is necessary to calculate a range of attributable fractions (AFs) under different scenarios. Even given a wide range of assumptions, the AF remains low, with point estimates ranging from < 1% to about 4%. As the AF is highly dependent on the exposure distribution, more data are needed on exposure levels worldwide, which should be collected in a large systematic survey of an appropriately selected sample. The fraction of childhood leukemia cases possibly attributable to ELF exposure across the globe appears to be small. There remain, however, a number of uncertainties in these AF estimates, particularly in the exposure distributions.

### Uncertainty

Despite of these results there are some uncertainties. Random error has many components, including temporal and geographic variation, inter- and intraindividual variability, as well as errors in measurement. Intraindividual, temporal, and geographic variability are usually accounted for by averaging measurements made via personal monitors and/or measurements

made in several locations over a specified period of time; this is essential and many ELF studies include extensive measurements, often with frequent sampling during measurement periods. Methods are available to account for errors in measurement (Carrol et al. 1995), but all studies on ELF effects ignore this aspect. Consequently, these studies most likely underestimate the true degree of uncertainty of their conclusions. In childhood leukemia, both pooled analyses performed in 2000 were based on large numbers and hence resulted in relative risk (RR) estimates with tight confidence intervals. When analyses are compared, they demonstrate consistency in the size of their effect estimates and range of confidence intervals. It appears unlikely that random variability (or chance) played a significant role in the observed effect estimates of both pooled analyses. However, this does not exclude the possibility that exposure was assessed with a large degree of random error, which could bias the observed relative risk (RR) toward the null and introduce a lot of uncertainty into the potential dose response. All attempts to examine potential confounding have not changed the risk estimates and substantial confounding from factors that do not represent an aspect of the electric or magnetic fields is unlikely. Selection bias may be partially responsible for the consistently described epidemiologic association between ELF and childhood leukemia (Mezei and Kheifets 2005). A large study by Linet et al. (1997) drives the overall risk estimate in both pooled analyses and may have had the greatest potential for selection bias, thereby potentially inflating the risk estimate associated with EMF exposure.

### **Closing the Knowledge Gaps**

- enlarge the number of persons who need to be incorporated into the study

One scene-setting perspective concerns the evidence in support of a link between exposure to ELF-EMFs and an adverse affect upon human health. The most satisfying evidence has been provided by epidemiological studies. The most direct approach to determine whether an environmental factor impacts upon human health is to look within the population for an association of the factor with a disease. The quality of such epidemiological studies depends critically upon two variables; first, the strength of the association of the factor with the condition and, second, the number of people who were studied in order to determine whether or not an association exists. The weaker the association of the factor with the condition, the larger the number of persons must be who need to be incorporated into the study in order to reveal the association. During the last 25 years, multiple epidemiological studies have been conducted in order to explore whether exposure to ELF-EMFs is linked to a human condition.

A diverse spectrum of conditions was targeted, including suicide, miscarriage and mental disease, although the majority of these studies focused on cancer and, in particular, leukaemia and brain cancers. In Crompton's (Crompton 2004) judgement, up to the year 2000, these multiple studies provided no satisfying positive or negative outcomes. However, this lack of clarity delivered one clear message, namely, that if ELF-EMF exposure is associated with an adverse human condition, then this association must be weak. In the year 2000, the publication of two independent studies caused a significant shift in opinion (Ahlbom et al. 2000; Greenland et al. 2000). The very important difference between these and the previous studies was the very much greater number of persons studied. This was achieved by pooling a number of previous studies and re-analysing the corporate data. Such an approach is known as a meta-analysis and is now commonplace in epidemiology; it has the notable benefit of revealing weak associations. The two independent meta-analyses delivered essentially the same outcome. This shows that continuous exposure over a long period to ELF-EMFs of average field strength greater than  $0.4 \mu\text{T}$  is associated with a doubling of the relative risk of childhood leukaemia.

To return to the perspective of the results of the two meta-analyses: First,  $0.4 \mu\text{T}$  represents a very weak field strength, especially when compared with the Earth's 100 fold stronger static magnetic field, to which all of us are exposed all the time. Second, a doubling of the relative risk represents only a weak association in epidemiological terms. For example: To express a doubling of the risk of childhood leukaemia in another way, this corresponds to an extra two cases each year in addition to the UK's annual average of 500 cases. Third, two scientific insights need to be factored into this discussion.

The first of these is that there is no known direct physical mechanism by which a field strength as low as  $0.4 \mu\text{T}$  can invoke a biological response. That is not to say that such a mechanism does not exist, but if it does, we currently have no knowledge of it. The second insight concerns the quality of the data and conclusions derived from a host of laboratory experiments that have sought to determine whether ELF-EMFs induce biological responses. That means these collective experiments do not provide any satisfying evidence in support of an affirmative answer.

To summarize, assuming the above perspectives are valid, then the epidemiological results are not underpinned by either a defined physical mechanism or prescribed biological responses. In other words, there is no established causal mechanism to support the results. When this

analysis is combined with the fact that the epidemiology revealed a weak association only, then it is apparent that the results are no more than indicative.

- Further investigation of biophysical mechanism

How should scientists respond to the challenge of this particular uncertainty?

The motivation is to explore whether there is indeed a direct physical mechanism by which weak magnetic fields could induce a biological response. Potentially, there are a number of possible mechanisms. One possibility is setting the focus on whether magnetic fields perturb the level of free radicals in biological systems. Free radicals represent an ideal biological candidate for transducing a physical force, because they occur naturally, are highly reactive and are mutagenic in living cells. The identification of the 'radical-pair mechanism' (Brocklehurst 2002) and the prediction of the so-called 'low-field effect' as a possible route to perturbing the concentration of free radicals in a biologically meaningful context have been reported (Brocklehurst & McLauchlan 1996; Timmel et al. 1998).

The starting point of radical-pair mechanism is a molecule that can be split by natural forces to form a pair of free radicals, which are in a 'singlet state' having opposite electron spins. If these radicals remain in close contact with one another, then they readily recombine to form the original molecule; whereas, if they diffuse away from each other, they remain as free radicals. Radicals in the singlet state can interconvert to the 'triplet state', having parallel spins. Triplet state radicals are not able to recombine; so in time they diffuse away from each other and are free to react with a variety of other molecules/radicals. The theory predicts that an applied magnetic field perturbs the interconversion of the singlet and triplet states, resulting in an increase in the proportion of the triplet state, and thus the free-radical concentration. While the description of the radical-pair mechanism represented a notable advance, there was a compelling need to prove the theory experimentally, as well as to confirm the predicted low-field effect. A simple chemical experiment in support of the theory and the prediction was performed. A mixture of pyrene and dimethylaniline exposed to a laser flash provided the free radicals, the yield of which was measured in the presence of an applied static magnetic field of increasing strength. (Note that the strength of the applied field was in the mT range.) The results showed that with the increasing strength of the field, the free-radical concentration (expressed in arbitrary units) initially increased to a maximal level before decreasing. The initial increase is commensurate with the predicted low-field effect.

While this is a very encouraging outcome, there is an absolute requirement to progress this experiment in a number of directions. First, the biological perspective demands studies with much weaker fields, and ideally in the 10  $\mu$ T range. However, if the radical-pair mechanism operates at such low fields, then the perturbation in free-radical concentration is unlikely to be sufficiently large per se to induce the magnitude of biological effects necessary to have an adverse effect upon human health. In other words, there is a need for an amplification step. Fortunately, a priori, there are a number of possible amplifiers, one of the most obvious being an enzyme system responsive to the increase in free-radical yield. A number of such potential systems exist. Second, and equally important, is to move forward from model chemical systems in solution to those which mimic biological systems, the ultimate endpoint being the biological cell.

#### - Replication of experiments

In the experimental sciences, independent replication of results represents the gold standard. In particular, in those cases where an experiment purports to show a positive result which borders on uncertainty, there is an absolute requirement for this result to be reproduced in an independent laboratory before it can be claimed to be fact.

During the past 20 years, the possibility that ELF-EMFs induce biological effects has been explored experimentally by a host of laboratories, using a large variety of model cellular systems, with different targets and various readouts. Within this collection of experiments, the majority claim to have shown a positive response.

#### - Database aided risk assessment

Quantitative risk assessment of the impact of ELF-EMFs exposure and the elucidation of mechanisms of its complexity requires computational infrastructure and innovative analysis approaches that systematically consider available data at all levels of biological organization. Thus a relational database is needed with associated data insertion, retrieval, and mining tools that manages all relevant physical, biological and, environmental data to facilitate comprehensive data integration, analysis, and sharing. Improvements in the quantitative risk assessment of chronic exposure in a complex environment require reducing uncertainties. Enterprise data management systems have proven to be indispensable in other fields, where they serve as the foundation for data integration across diverse sectors to support large data mining efforts. Similarly, risk assessment involves combining disparate data throughout the source-to-outcome continuum to identify diagnostic profiles and relationships in susceptible

populations and ecologically relevant species. These profiles may represent agglomerative biomarkers encompassing exposure, molecular responses, and adverse effects, which facilitate the re-evaluation of historical data in light of new information, or allow comparisons across complementary technologies, or species.

In addition, relational databases provide the infrastructure for the comprehensive management and integration of disparate data domains facilitating not only the phenotypic anchoring of toxicogenomic data, but also the development of advanced analysis methods, quality assurance protocols, predictive tools, and the systematic examination of mechanisms of the impact of electromagnetic fields in combination with other influences. Thus the database capabilities have to be extended to include metabonomic data, with proteomic and pathway information. Furthermore, such a system must support the integration of gene expression data, functional annotation, orthology, and genomic motif regulatory information. These efforts facilitate the elucidation of comprehensive mechanisms of the impact of ELF-EMF, the identification of mechanistically-based biomarkers, and will also engender more accurate mechanistically based quantitative risk assessments. Burgoon et al. (2005) have already developed a relational database (e.g. dbZach) for the appliance in the field of toxicological research.

- Using technology advances

Next-Generation Sequencing facilities may be used to analyse the changes in gene expression induced by several different stresses (e.g. ELF-EMF). (This exciting technology is described in the next chapter.)

### **Next-Generation Sequencing Technologies**

There is no doubt that a 1,000- to 10,000-fold decrease in the cost per base of reagents and supplies for DNA sequencing will foster unprecedented research and large-scale clinical studies. The rapid maturation of next-generation technologies promises to break through the limitations of current technologies, offering dramatically improved throughput, minimal use of reagents, ease of use, and, most importantly, a profound decrease in cost. Moreover, the advent of massively parallel sequencing provides researchers the prospect of directly interrogating entire DNA samples, rather than relying on the partial information provided by genotyping and single nucleotide polymorphism (SNP) analysis. Many rival technologies are vying to succeed Sanger sequencing, which has served the community so well for 3 decades

(and will probably still play an important role for years to come). The next-generation sequencing technologies can generally be grouped into 2 categories (Table 1).

The first category, including the first 4 instruments that have been (or are soon to be) commercially released, commonly use cycles of synthesis (or ligation) to read fragments of DNA sequence. One company (Helicos) measures single molecules, whereas the others look at polymerase chain reaction (PCR) pools of fragments (454, Illumina, ABI). Two other companies in this category are GE Healthcare and Intelligent Bio-Systems.

Category	Method	Facility or Enterprise
Cyclic: Grow then Read	1. Single molecule	Helicos
	2. Clonal Pools	454 (Pyrosequencing)
		Illumina/Solexa (sequence-by-synthesis)
		ABI (ligation)
		Intelligence Bio-Systems (Yu, Columbia University)
		GE Healthcare
Real-time Single Molecule	1. Nanopore	Branton/Golovchenko (Harvard) plus many academic groups
	2. Sequence-by-synthesis (read sequence like a movie)	Pacific Biosciences
		VisiGen

Table 1: Next-Generation Sequencing Technologies (Harris T. 2007)

The other category, further from commercialization, uses fluorescence to measure DNA single-molecule sequence in real time. This includes the theoretically promising approach of nanopore sequencing, as well as sequence-by-synthesis approaches of VisiGen and Pacific Biosciences. Here we examine the major technologies that make up the next generation sequencing portfolio.

#### - Sequencing-by-Synthesis/Pyrosequencing

The first commercial method to capture widespread attention was the single-molecule sequencing platform of 454 Life Sciences (Branford, CT). This method assumed that sequencing-by-synthesis had largely been discarded, but he envisioned an approach that coupled miniaturization with massive parallelization. After several years of development 454 published a landmark paper in Nature (Margulies et al. 2005), showcasing 454's debut instrument (Genome Sequencer 20) and its application to the sequencing of a small bacterial genome.

Technology	Company	Cost	Pros	Cons
------------	---------	------	------	------

Pyrosequencing	454 Life Sciences	\$500K	- >220-bp read lengths - First commercial system	- Lower output - Higher error rate - Relatively higher cost
Sequencing-by-synthesis	Illumina/Solexa	\$400K	- Affordability and flexibility - high bead density	- Shorter read lengths (~25bp) - Limited de novo applications
Single-molecule sequencing	Helicos	Cost of mass spectrometry	- No PCR bias or amplification errors - Reagent cost savings ~1000 times less than Sanger sequencing	- Not widely available until 2008 - Expensive
SOLiD (Supported Oligonucleotide Detection)	Applied Biosystems	~\$600K	- Based on ligation chemistry, not polymerisation - built-in error correction system distinguishes polymorphisms from sequencing errors - Novel color-coded readout	- Short read lengths (~25-35 bp) - Enormous data storage/processing requirements

Table 2: Features of Major Next-Generation Sequencing Platforms (Davies 2007)

Briefly, DNA is nebulized into tiny fragments and shaken in oil and water, so that each DNA fragment falls into a separate water droplet. (This technique is called emulsion polymerase chain reaction.) Each drop has some 10 million copies of the DNA fragment. The DNA is then attached to micron-sized beads, which are dropped into octagonal wells etched into microtiter plates. Next, each DNA base (A, C, T, G) is independently washed over the plate. A charge-coupled device (CCD) camera records each added base via the light emitted by a pyrosequencing reaction. (The pyrosequencing technique is licensed by the Swedish firm Biotage to 454.) The 454 Nature paper described sequencing individual strands up to 100 bases, as many as 400,000 wells at a time, with the bacterial genome sequenced at 50× coverage. The Achilles heel of the GS20 was the mediocre accuracy of each sequence read. The original read accuracy averaged approximately 96%. 454's second-generation machine, the FLX, was introduced in early 2007 and features several major upgrades over its predecessor, including greatly improved read length—up from an average of 100 bases to 250 bases. The new machine also doubles the number of reads per run from 200,000 to just over 400,000. Most importantly, the system has greatly improved accuracy—the error rate dropped to less than 0.5%. 454's success in the race to commercialization put it in pole position to forge numerous collaborations with a variety of academic research groups. Successes to date include the sequencing of the Neanderthal genome, assorted metagenomic and microbial projects, and, in early 2007, the sequencing of Jim Watson's complete genome. Compared to

other next-generation platforms under development, 454's instrument produces relatively little raw data. It is awaited that the data from a single run could be compressed sufficiently to be stored on a single DVD.

- Sequencing-by-Synthesis: Illumina

The Illumina (San Diego, CA) platform, the 1G Genome Analyzer (acquired from Solexa in 2006), is based on massively parallel sequencing of millions of fragments using novel reversible terminatorbased sequencing chemistry developed over the past 5 years (see Table 2). In this system, randomly fragmented DNA is attached to an optically transparent surface. Amplification produces an ultra-highdensity sequencing flow cell with more than 10 million clusters, each containing approximately 1,000 copies of template per square centimeter. These templates are sequenced using a 4-color DNA sequencing-by-synthesis technology that employs reversible terminators with removable fluorescence, which improves the sequencing of homopolymeric repeat sequences. Short sequence reads—initially in the 25-base range—are aligned against a reference genome, and genetic variants are highlighted using a unique data pipeline. These short fragment reads lend the technology to resequencing applications, gene expression, and small RNA analysis, less so to de novo sequencing of complex genomes. Although the read lengths are smaller than those of 454, Illumina compensates by using much smaller beads, producing a roughly 1,000- fold increase in bead density compared to 454's. Future refinements in chemistry and engineering are planned to increase individual fragment lengths, although this would not have a substantial impact on the assembly efficiency. By adding all 4 bases simultaneously, the polymerase can always select the correct base (as happens in nature). In 2006, Solexa began shipping its “groundbreaking” instrument, priced at around \$400,000. The 1G Analyzer can resequence 1 billion bases of human DNA in 1 to 2 days. One of the biggest potential applications for the system is in the area of haplotype mapping. Illumina is offering its customers the opportunity to store all of their images, betting that many clients will be reluctant to throw out their image data. However, this could push up to and beyond 1 TB data generated per run. Illumina is recommending that laboratories acquire a compute cluster that has something like 8 nodes and 10 TB of disk capacity.

- GE Healthcare

GE Healthcare involves a team that aims to use a novel combination of enzyme and dye-tagged nucleotide resources to simplify the front-end chemistry in massively parallel sequencing-by-synthesis. The method uses DNA polymerase to capture a single DNA base on an immobilized primer/template. The nucleotide is captured and tagged with a dye on the pyrophosphate group before the chemical reaction step of the polymerization cycle. The

tagged nucleotide is then identified using a fluorescence scanner that can scan hundreds of thousands of similar molecules at one time. When scanning is complete, a new cycle begins. During DNA polymerization, the detectable tag of the incoming nucleotide analog is associated with the growing DNA strand held by the DNA polymerase. By freezing the DNA polymerase in the state that tightly binds the nucleotide, before the nucleotide is chemically attached to the DNA strand and the tag is cleaved away, sequence information can be inferred from the tags. In contrast to other systems, however, the DNA strand is a perfectly normal, unlabeled fragment. GE Healthcare hopes to develop a machine that can determine the sequence of a person's genome in a couple of hours.

#### - Intelligent Bio-Systems

Intelligent Bio-Systems (IBS; Waltham, MA) is a spin-off from Columbia University, where, in 2006, Jingyue et al. published a variation on sequencing-by-synthesis using novel fluorescently tagged nucleotides (Ju et al. 2006). The IBS process begins by producing an array of millions of amplified DNA fragments (attached to a sequence primer) on a glass chip. By tagging the nucleotides in a different position than conventional modifications, IBS maintains that it improves the rate of incorporation into the elongating strands. Each nucleotide is also capped at the 3' end so that only 1 base is incorporated per cycle. After each round of incorporation, the array is scanned using a high-resolution electronic camera, which records the fluorescent output of each dye at each array position. Once the first round of nucleotides has been read, both the dye tag and a cap on the 3'-OH group are removed, before the next cycle. According to IBS, "The technology overcomes the limited throughput and high costs of conventional electrophoretic technologies and most deficiencies of other new-generation non-electrophoresisbased sequencing methods."

#### - SOLiD (Supported Oligonucleotide Detection)

In its effort to identify a high-throughput sequencing technology to complement or even succeed Sanger sequencing, Applied Biosystems investigated no fewer than 40 different technologies before acquiring Agencourt Personal Genomics (APG) from Beckman Coulter in 2006, for \$120 million. The system will be tested by early-access customers in summer 2007, with full commercial release planned for fall 2007. The SOLiD instrument produces more than a 500-megabases (MB) sequence per run, sequencing DNA off densely packed magnetic beads that have been put through emulsion PCR (see Table 2). Sheared DNA is prepared on 1-micron beads sitting in a microreactor of a water-in-oil emulsion. After the emulsion PCR, most of the reactors, and most of the beads, will be empty; this is by design so that, in most cases, each droplet has 1 bead and each bead 1 DNA fragment. An enrichment process, using

polystyrene beads to capture the beads with amplified template, boosts the proportion of beads with template from approximately 30% to 80%.

There are several distinguishing features of the SOLiD method. First, the system uses ligases rather than polymerases to detect new DNA strands. This facilitates a readout system that interrogates 2 adjacent bases at a time, resulting in a built-in error correction system—particularly useful for applications such as cancer resequencing. ABI says the system has 10-fold higher accuracy than rival systems, because it distinguishes naturally occurring polymorphisms from sequencing errors. The SOLiD system uses a series of overlapping ligation probes that hybridize to the template. Each ligation probe is an 8-mer that interrogates not 1 but 2 adjacent bases at a time. The 8-mers consist of 5 nucleotides and 3 universal bases, for a total pool of 1,024 (4<sup>5</sup>) probes. The 2 critical positions that are interrogated are positions 4 and 5 of the 8-mer. There are 16 (4 × 4) possible combinations of these 2 adjacent positions. Using 4 fluorescent dyes, each dye is thus tagged to 4 possible dinucleotide combinations (Figure 3). For example, blue dye corresponds to AA, CC, TT, or GG. Thus, any color by itself cannot reveal the identity of a particular base. Once the identity of the first base is known, however, the rest fall neatly into position. The SOLiD system uses a series of overlapping ligation probes that hybridize to the template.

For example, if the first base is an A, and the first fluorescent color detected is blue, then the first 2 bases must be A-A (because the other possibilities are, by definition, excluded). Let's suppose the next color is green. As we just deduced that the second base in the mystery sequence is an A, we deduce the identity of the third base by reading out which green-tagged dinucleotide pair starts with an A. Thus, the third base is C (Figure 4). This process continues for the read length of 25 to 35 bases (Davies 2007).

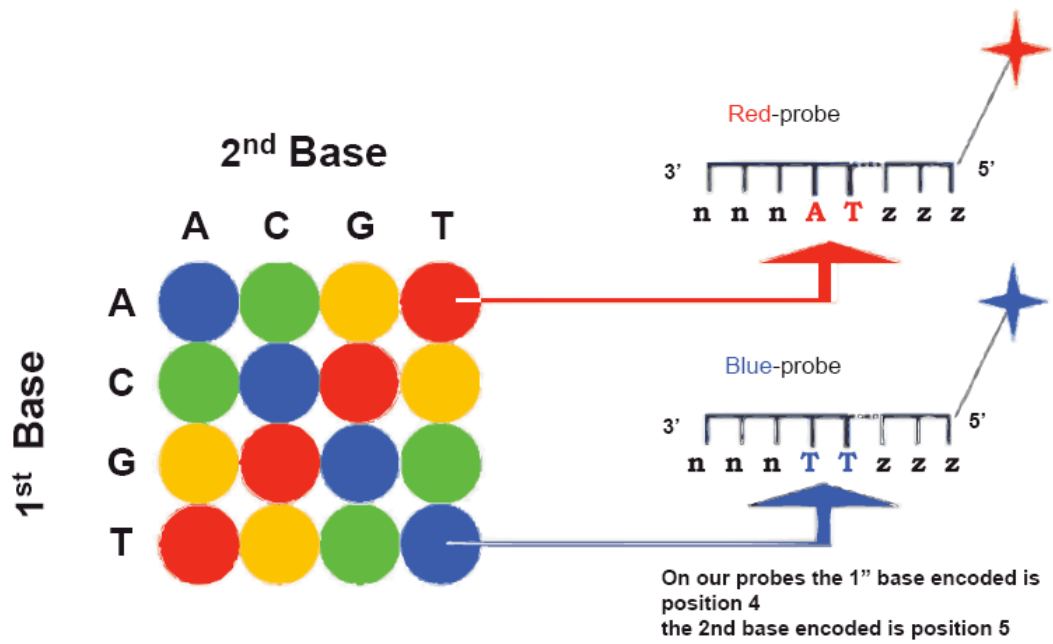


Figure 3: SOLiD Base Pair Encoding Using 4 Dyes (Davies 2007)

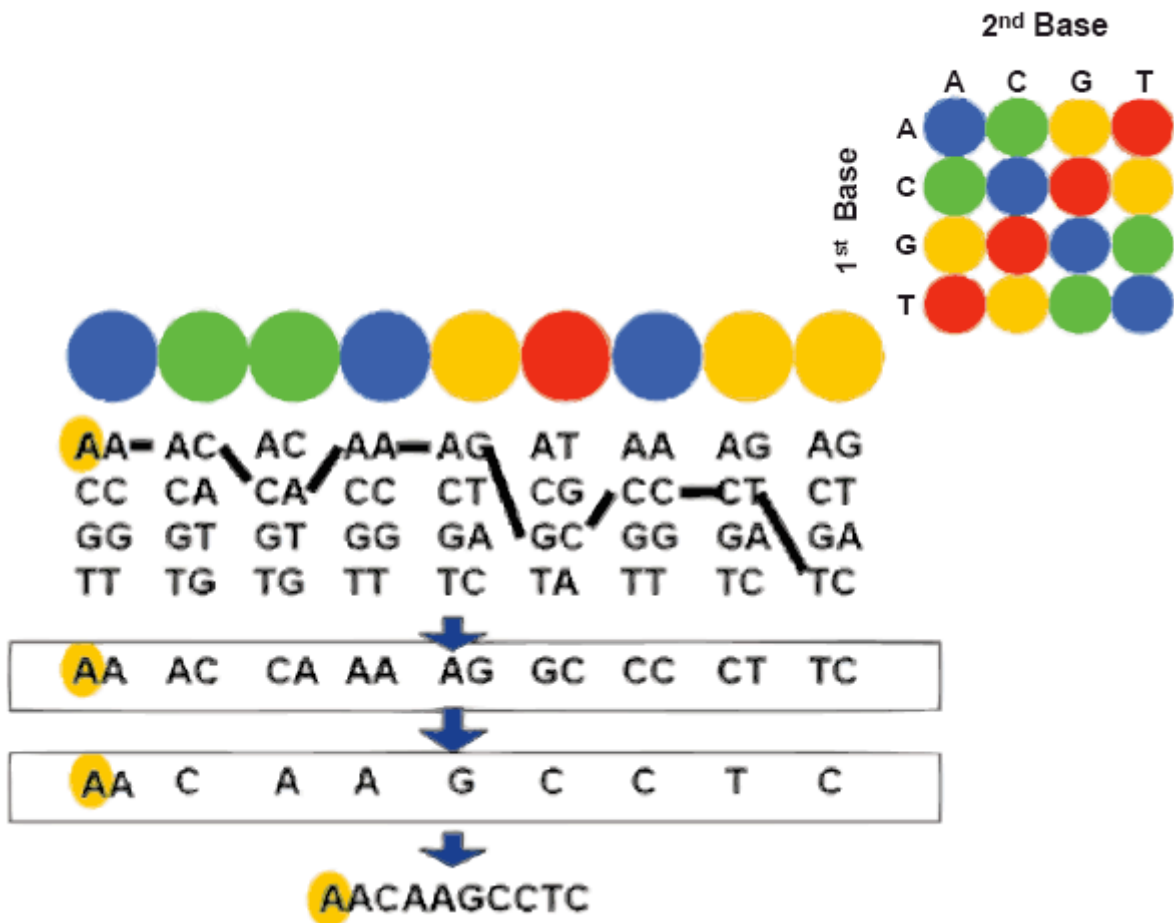


Figure 4: SOLiD: Example of Decoding (Davies 2007)

Although the individual read lengths are small, the SOLiD system neatly distinguishes between naturally occurring variants (SNPs) and random sequencing errors. In most sequencing systems, there is no way to discern a genuine SNP from a sequence error. However, in SOLiD, a random sequence error will only change a single colored tag. A genuine SNP, by contrast, is revealed as changes in consecutive color tags, because the altered base occurs in 2 successive nucleotide pairings. ABI is also putting a large effort into using paired-end sequencing to extend the range and coverage of its sequencing.

#### - Data Storage

A critical issue for users of SOLiD, and indeed most of the nextgeneration sequencing platforms going forward, is that of data storage and processing. A typical 454 FLX instrument run produces 12 GB data, with another 1 to 2 GB of metadata added. The SOLiD instrument, by contrast, ships with a large computing infrastructure to do real-time analysis of the images and sequence alignment. This consists of a 10-box, dual-core CPU Linux cluster (supplied initially by Dell) with 15 TB of storage. Those data can be offloaded via an Ethernet port.

#### - Polony Sequencing

The technique of polony sequencing was developed by George Church and colleagues at Harvard Medical School. The term polony comes from the parallel interrogation of “polymerase colonies.” Some of the fundamentals of this homegrown tag multiplex sequencing technique were laid down by Church in 1984, but making a viable cell-free, electrophoresis-free sequencing method took a lot of development. Aspects of the method have been licensed to Applied Biosystems. The polony method came into its own in 2005, when Church’s Shendure and coworkers published a seminal paper (Shendure 2005), applying the method to the sequencing of a laboratory strain of *Escherichia coli*. The group estimated its costs to be one-tenth those of traditional Sanger-based sequencing, with less than 1 error per million bases. In the paper, they described how they were able to adapt a standard-issue epifluorescence microscope for automated nonelectrophoretic DNA sequencing. The polony method amplifies DNA strands on millions of 1- $\mu$ m beads using emulsion PCR, which are then immobilized to an array. As fluorescently labeled DNA bases bind to the complementary DNA sequences via a novel ligation reaction, an epifluorescence microscope records each base in the growing fragment using 4-color imaging.

Subsequent efforts of the Church group have included a focus on the coding regions of the human genome—the exome. Church notes that personalized medicine focusing on the exome—the 1.0% to 1.5% portion of the genome that contains the majority of disease-causing

mutations—would be a logical and practical intermediate step before the cost of sequencing falls enough to truly permit a complete \$1,000 human genome.

#### - Single-Molecule Sequencing

Although 1 to 2 years behind in commercialization, Helicos BioSciences expects to release its debut instrument in 2007 (see Table 2). Helicos was founded in 2003 in Cambridge, Massachusetts, based on technology developed by Howard Hughes Medical Institute (HHMI)/Stanford biophysicist Stephen Quake. The Helicos platform is a form of sequencing-by-synthesis, but is arguably unique in that it is true “single-molecule” sequencing (tSMS), requiring no amplification steps, photographing the process of DNA replication. The process was first described by Quake and coworkers in a paper published in 2003 (Braslavsky et al. 2003). The HeliScope system consists of a high-speed mechanical stage, a laser, image acquisition, fluid handling, and computer subsystems that control and analyze the sequencing reactions. The platform uses a proprietary DNA polymerase, fluorescent bases, and imaging reagents. tSMS begins by shearing DNA and separating it into more than 1 billion single-stranded templates, each a few hundred bases in length, randomly attached to a glass slide. The DNA templates are washed with enzymes and fluorescently tagged nucleotides. One base is incorporated at a time; the newly incorporated bases are detected by fluorescence, which is then quenched; the slide is rinsed to remove all but the faintest traces of background fluorescence; and the cycle is repeated. After the introduction of each base, the flow cell is moved under a microscope lens; a laser illuminates the fluorescent tags and a camera images the flow cells through the microscope lens. (To stabilize the instrument, the HeliScope will be housed in 500-pound single blocks of granite.) The instrument’s computer system assembles and analyzes the images of the fluorescent bases to determine the sequence of the bases. It then pieces the short sequences together into a single complete one. The length of the templates grows asynchronously, but it is typically possible to read about 30 bases from each strand. tSMS allows direct interrogation of single molecules, as opposed to an amplified population of molecules, and holds great potential for elucidating the gamut of genetic aberrations in oncology, through the ability to serve as a universal detection system across a wide variety of applications for both DNA and RNA. In addition, Helicos views promising applications in transcriptional profiling, genome-wide methylation studies, and candidate region resequencing. Potential advantages of tSMS over rival systems include the absence of PCR bias or errors introduced by amplification and no dephasing issues, which are commonly present in amplified molecule sequencing. In addition, tSMS promises the highest possible throughput and enables reagent cost savings on the order of 1,000 times less than Sanger

sequencing. At the 2007 Advances in Genome Biology and Technology Conference (AGBT) in Marco Island, Florida, Helicos announced that it had developed novel nucleotide-polymerase formulations for the sequencing of homopolymeric regions of DNA. Helicos has predicted that the HeliScope could sequence a complete human genome in 3 days and for roughly \$5,000. The first version of the HeliScope is expected to sequence up to 90 million bases per hour. (By contrast, existing platforms sequence approximately 120,000 bases per hour.) Further improvements in the use of reagents, microfluidics, and image processing should eventually allow maximum throughput of close to 1 billion bases per hour. Helicos claims that it will improve on the current market-leading cost—\$1 per 1,000 bases— by generating sequence for about 1 cent per 1,000 bases before reducing it a further 100-fold. The debut instrument will take about 2 days to read a 25-base DNA fragment. As for ABI's SOLiD system, the HeliScope will require an impressive IT infrastructure. The machine in the process of being finished will generate 20 TB of image data/day. The HeliScope also generates approximately 1 TB of analyzed data (read strands).

#### - VisiGen

A variant on single-molecule sequencing is being developed by VisiGen (Houston, TX). The VisiGen system combines single-molecule detection, fluorescent molecule chemistry, computational biochemistry, and genetic engineering of biomolecules to create a novel sequencing system. Among the advantages touted by the company are the lack of amplification or cloning and the reuse of the template. However, the company's first instrument is not likely to be ready until at least 2009. VisiGen's approach involves engineering both polymerase and nucleotides to act together as direct molecular sensors of DNA base identity in real time. The goal is to use the massively parallel nature of this technology-generated DNA sequence at rates approaching 1 million bases per second per machine, or "developing hardware and software capabilities to improve sequencing rates by a factor approaching 100,000 compared to currently available technologies."

#### - Pacific Biosciences

Another promising real-time, single-molecule technique for sequencing DNA began in the laboratories of physicist Watt Webb, and his Cornell University colleague, Harold Craighead. The optical technique focuses on the elongation of a complementary DNA strand by DNA polymerase, using fluorescently labeled nucleotides (Figure 5) (Levene et al. 2003). The technique identifies each base as it is being incorporated into the growing DNA. The method uses a technique called zero-mode waveguides, which limits the observation volume to tens of zeptoliters, significantly enhancing the signal of the fluorescent dye tagged to the incoming

nucleotide against the fluorescent background. The zero-mode waveguides are generated by passing light through an array of micron-diameter pores in an 89-nm-thick sheet of aluminum overlaying a slide. (The holes are patterned using electron beam lithography.)

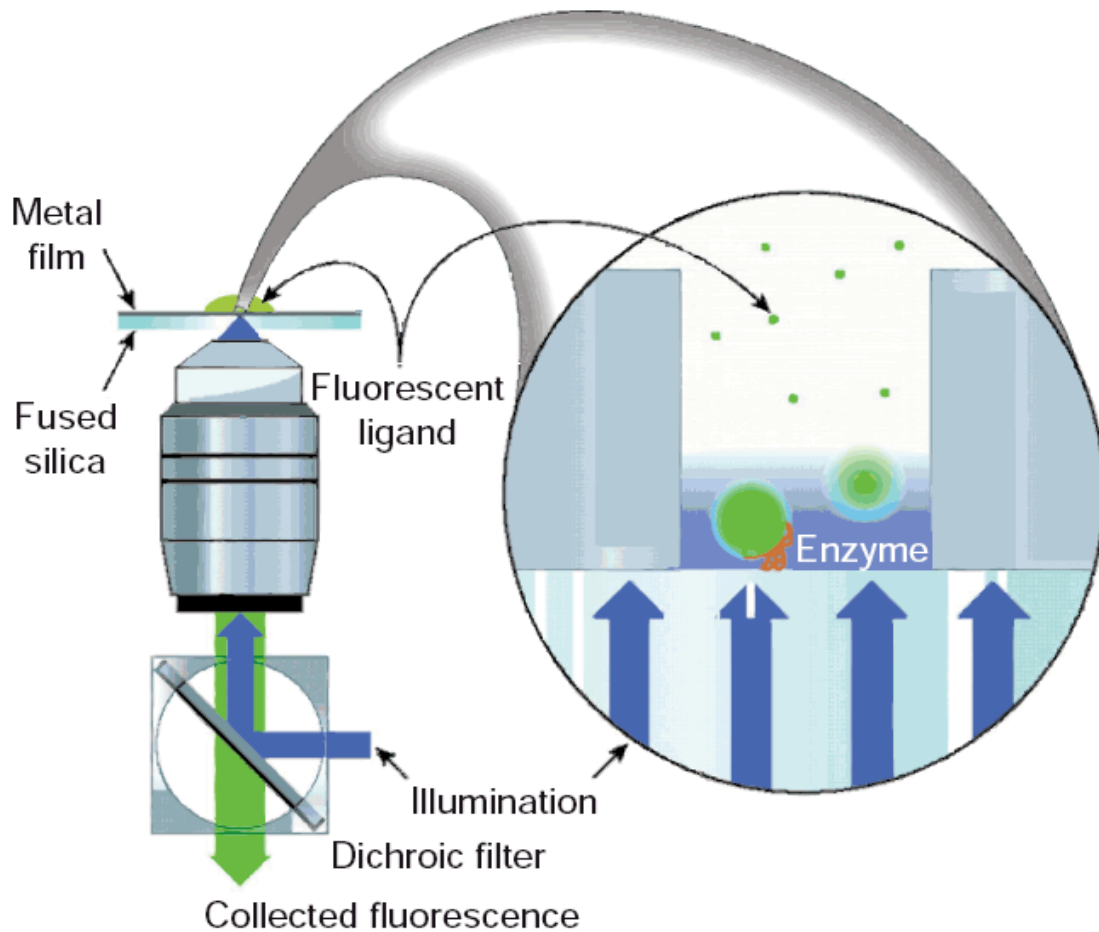


Figure 5. Apparatus for Single-Molecule Analysis Using Zero-Mode Waveguides (Levene et al. 2003)

#### - Nanopore Sequencing

Nanopore sequencing is one of the most exciting and direct methods for DNA sequencing—if the current practical limitations can be overcome. The process involves pushing linear, single-stranded DNA molecules through a minute pore, approximately 1.5 nm wide. These pores are modeled on naturally occurring membrane channels that shuttle ions or proteins in and out of cells. As the DNA strand passes through the pore, it changes the pore's electrical properties, which, in turn, can be measured by a shifting electrical signal. Improved sensitivity is obtained by using solid-state nanopores etched into silicon nitride, which Harvard University's Jene Golovchenko, and Daniel Branton demonstrated can measure single DNA molecules. Branton's strategy is to decorate both sides of the nanopores with tiny probes that serve as electrodes. The resulting current is modulated as the diameter of the pore shifts with

the moving bases in the DNA strand passing through the pore. In principle, this technology could provide true single-molecule sequence, although current success has been largely limited to long homopolymeric stretches of DNA (Rhee et al. 2006).

#### - Optical Trapping

Block, professor of applied physics at Stanford University, and Greenleaf published a paper in *Science* applying an optical tracking device to the single-molecule sequencing of DNA (Greenleaf et al. 2006). The method uses the motion of RNA polymerase molecules traversing a strand of DNA. Changes in the nucleotide base composition of the target molecule alter the rate of transcription. The method uses a pair of optical traps and 2 polystyrene beads, 1 tagged to the polymerase, the other to the end of the DNA template. Similar to Sanger sequencing, the process is carried out in quadruplicate, with each base present in limiting concentrations. This causes the polymerase to pause when the limiting base is to be introduced. The shifting motion of the polymerase is detected with “angstrom-level precision,” capable of single base pair resolution and allowing the sequence to be inferred from the pattern of pauses in the 4 parallel reactions. A flanking sequence is used to align each of the 4 reactions, analogous to the oligonucleotide primers in Sanger sequencing.

#### - Nanoscale Sanger Sequencing

Sanger sequencing has undergone a remarkable evolution since its radioactive slab-gel beginnings about 30 years ago (Blazej et al. 2006). The prevalence and longevity of the Sanger method stem from its ability to generate both long consecutive reads with high accuracy. The Berkeley group has developed “an efficient, nanoliter-scale microfabricated bioprocessor integrating all 3 Sanger sequencing steps, thermal cycling, sample purification, and capillary electrophoresis.” The authors state that their lab-on-a-chip integrated device performs complete Sanger sequencing from just 1 fmol (femtomole) of DNA template. The group was able to sequence more than 550 contiguous bases with 99% accuracy. The findings so far suggest that reagent requirements can be reduced 100-fold, with commensurate savings in infrastructure and labor costs. While the other novel methods described in this report will see a host of applications, there may yet be a role for Sanger sequencing in molecular analysis. As the Berkeley group states, “The ability to determine long-range genome structure and not simply local sequence uniqueness is critical in complex genome sequencing and is likely to be central in fully understanding the subtle effects of genome organization on development, speciation, and disease.”

#### - Applications of Next-Generation Sequencing

Although it is still early days for next-generation sequencing, a remarkable number of applications have already emerged, with many more on the horizon. Of surpassing interest is the promise for medical resequencing. Currently, genetic analysis, particularly for the common, complex disorders, relies on a combination of whole genome single nucleotide polymorphism (SNP) analysis and expression studies, with sequencing honing in on candidate genes identified by linkage or association studies.

- The \$1,000 Genome/Personal Genomics

Arguably the most exciting application of next-generation sequencing will be the ensuing revolution in personalized medicine. Although it may take several years to reach the \$1,000 plateau, even a genome sequence costing 5 to 10 times as much would transform the practice of medicine. At that point, genome sequencing becomes affordable not just for a handful of billionaires or celebrities. At the end of May 2007, researchers from 454 and Baylor College of Medicine presented Nobel laureate James Watson with his personal DNA sequence. 454 sequenced Watson's genome to 8× coverage in early 2007; Watson was not the first individual to have his sequence deposited in GenBank. A week earlier, Celera founder J. Craig Venter deposited his complete genome sequence (which was completed by scientists at the J. Craig Venter Institute earlier in 2007).

- The Cancer Genome Atlas

In 2006, senior investigators supported by the NIH led calls for a new cancer genome project: The Cancer Genome Atlas, or TCGA. The calls came 25 years after the discovery of oncogenes, the first cancer-causing mutations. In recent years, more systematic efforts to catalog mutations have borne fruit, with the discovery of hundreds of cancer-related genes. Another comprehensive study, published in 2006, from a group at Johns Hopkins concluded that mutations in more than 190 genes (out of 13,000 surveyed) conspire to cause breast and colon cancer (Sjöblom et al. 2006). Following the lead of the Human Genome Project and other recent medical genomics efforts, all these data will be made swiftly and will be freely available to the worldwide research community. To further enhance its usefulness to both basic and clinical researchers and, ultimately, health care professionals, TCGA will link its sequence data and genome analyses with information about observable characteristics of the original tumors and the clinical outcomes of the sample donors. The success of TCGA will hinge on the maturation of new sequencing technologies. And even if the full spectrum of mutations are identified, their role in cancer and metastasis will still have to be confirmed.

### - Evolutionary Genomics

In 2006, Svante Pääbo, PhD, and colleagues at the Max Planck Institute (Munich, Germany), in collaboration with 454 Life Sciences, published a breakthrough analysis of 1 million bases of the Neanderthal genome (Green et al. 2006). For the field of evolutionary genetics, this was one of the most exciting pieces of work since the discovery of the first Neanderthal skeleton 150 years ago in Germany. Extracting DNA from such aged skeletons is fraught with difficulties, notably the degradation of much of the DNA into tiny fragments (so small that many experts doubted that nuclear DNA could ever be analyzed). The real threat of bacterial contamination also complicates DNA analysis. However, both problems are minimized by the 454 platform, which sequences hundreds of thousands of individual fragments per run. The individual read lengths of 250 bases or so provide an excellent match for the length of the isolated Neanderthal DNA fragments. Pääbo hopes to go on to sequence an entire Neanderthal genome over the next 2 years. The results should clarify not only how Neanderthals and humans diverged, but also shed light on modern human evolution from the ancestor we shared with chimpanzees some 5 million years ago.

### - Environmental Genomics

Since the sequencing of the first microbial genome in 1995, close to 1,000 genomes have been sequenced, including hundreds of bacterial sequences. The advent of next-generation sequencing platforms will accelerate this production line of genome sequences, particularly of simpler organisms.

### - Gene Regulation

In addition to genome sequencing, high-throughput sequencing provides new opportunities for mapping sites of gene regulation. The new method—ChIP sequencing (ChIP-Seq)—combines chromatin immunoprecipitation with next-generation sequencing, and was performed with a speed and precision that go beyond what has been achieved with previous technologies (Fields 2007). ChIP is a well-established technique to identify specific sites where proteins latch onto the DNA. Until now, the most high-throughput application of this technique involved using microarrays to identify binding sites for transcription factors and the like. In another recent study published in *Cell*, researchers at the NIH demonstrated that direct sequencing of ChIP DNA is an efficient method for mapping genome-wide distributions of histone modifications and chromatin protein targets (Barski et al. 2007).

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. 1977;74:5463-5467

**References**

- Adey WR, Byus CV, Cain CD, et al. Spontaneous and nitrosourea-induced primary tumors of the central nervous system in Fischer 344 rats exposed to frequency-modulated microwave fields. *Cancer Res.* 2000;60 :1857 –1863
- Advisory Group on Non-ionising Radiation. *Health Effects Related to the Use of Visual Display Units. Report of and Advisory Group on Non-ionising Radiation.* Vol 5. Chilton, United Kingdom: National Radiological Protection Board; 1994
- Advisory Group on Non-ionising Radiation. *Health Effects From Radiofrequency Electromagnetic Fields. Report of an Independent Advisory Group on Non-ionising Radiation.* Vol 14. Chilton, United Kingdom: National Radiological Protection Board; 2003
- Ahlbom A, Day N, Feychting M, et al. A pooled analysis of magnetic fields and childhood leukemia. *Br J Cancer.* 2000;83 :692 –698
- Anane R, Dulou PE, Taxile M, Geffard M, Crespeau FL, Veyret B. Effects of GSM-900 microwaves on DMBA-induced mammary gland tumors in female Sprague-Dawley rats. *Radiat Res.* 2003;160 :492 –497
- Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci USA* 2003; 100:3960-3964
- Barski A, Cuddapah S, Cui K et al. High-resolution profiling of histone methylations in the human genome. *Cell.* 2007;129:823-837
- Blazej RJ, Kumaresan, P, Mathies RA. Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. *Proc Natl Acad Sci USA* 2006;103:7240-7245
- Böhler E, Schüz J. Cellular telephone use among primary school children in Germany. *Eur J Epidemiol.* 2004;19 :1043 –1050
- Bracken T, Kheifets L, and Sussman S. Exposure assessment for power frequency electric and magnetic fields (EMF) and its application to epidemiologic studies. *J. Expo. Anal. Environ. Epidemiol.* 1993;3(1):1-22.
- Brix J, Wettemann H, Scheel O, Feiner F, Matthes R. Measurement of the individual exposure to 50 and 16 2/3 Hz magnetic fields within the Bavarian population. *Bioelectromagnetics.* 2001;22(5):323–332.
- Brocklehurst B. Magnetic fields and radical reactions: recent developments and their role in nature. *Chem. Soc. Rev.* 2002;31:301–311.
- Brocklehurst B, McLauchlan K.A. Free radical mechanism for the effects of environmental electromagnetic fields on biological systems. *Int. J. Radiat. Biol.* 1996;69:3–24.

- Burgoon LD, Zacharewski TR. dbZach toxicogenomic information management system. *Pharmacogenomics*. 2007 Mar;8(3):287-91.
- Cardis E, Kilkenny M. International case-control study of adult brain, head and neck tumours: results of the feasibility study. *Radiat Prot Dosimetry*. 1999;83 :179 –178
- Carrol RJ, Rupert D, Stefansi LA. Measurement Error in Nonlinear Models. New York:*Chapman & Hall*. 1995.
- Cassidy RA, Vorhees CV, Minnema DJ, Hastings L. The effects of chlordane exposure during pre- and postnatal periods at environmentally relevant levels on sex steroid-mediated behaviors and functions in the rat. *Toxicol Appl Pharmacol*. 1994;126 :326 –337
- Chen D, Toone W.M, Mata J, Lyne R, Burns G, Kivinen K, Brazma A, Jones N, Bähler J. Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell*. 2003;14:214–229.
- Christensen HC, Schuz J, Kosteljanetz M, Poulsen HS, Thomsen J, Johansen C. Cellular telephone use and risk of acoustic neuroma. *Am J Epidemiol*. 2004;159 :277 –283
- Coleman M, Bell C, Taylor H et al. Leukemia and residence near electricity transmission equipment: a case-control study. *Br. J. Cancer*; 1989; 60(5): 793-798
- Crumpton MJ. The Bernal Lecture 2004 Are low-frequency electromagnetic fields a health hazard? *Philos Trans R Soc Lond B Biol Sci*. 2005 Jun 29;360(1458):1223-30.
- Davies K. The drive for the \$1000 genome. *Bio-IT World* 2007; 5:1-88
- Dawson TW, Caputa K, Stuchly MA, Kavet R. Electric fields in the human body resulting from 60-Hz contact currents. *IEEE Trans Biomed Eng*. 2001;48 :1020 –1026
- Decat G, Van den Heuvel I, Mulpas L. Final Report of the BBEMG Research Contract. June 10, 2005. Erembodegem, Belgium:*Flemish Environmental Agency*.
- Effects and Risks of Ionizing Radiation; UNSCEAR 2000 Report to the General Assembly. New York, NY: United Nations Scientific Committee on the Effects of Atomic Radiation, *United Nations*; 2000
- Electric Power Research Institute (EPRI). Survey of Residential Field Magnetic Field Sources: September 1993. Palo Alto, CA:Electrical Power Research Institute.
- EMF Rapid Program 1998. Survey of Personal Magnetic Field Exposure, Phase II: 1000-Person Survey. Research Triangle Park, NC:National Institute of Environmental Health Sciences.
- Feychting M, Ahlbom A. Magnetic fields and cancer in children residing near Swedish high-voltage power lines. *Am. J. Epidemiol*. 1993; 138(7):467-481

- Feychting M, Forssen U, Floderus B. Occupational and residential magnetic field exposure and leukemia and central nervous system tumors. *Epidemiology* 1997; 8(4): 384-389
- Fields S. Molecular biology. Site-seeing by sequencing. *Science*. 2007;316:1441-1442
- Friedman DR, Hatch EE, Tarone R, et al. Childhood exposure to magnetic fields: residential area measurements compared to personal dosimetry. *Epidemiology*. 1996;7 :151 –155
- Greaves M. Childhood leukemia. *BMJ*. 2002;324 :283 –287
- Green L, Miller A., Villeneuve P et al. A case-control study of childhood leukemia in Southern Ontario, Canada, and exposure to magnetic fields in residences. *Int. J. Cancer* 1999; 82(2): 161-170
- Green RE, Krause J, Ptak SE et al. Analysis of one million base pairs of Neanderthal DNA. *Nature*. 2006;444:330-336.
- Greenland S, Sheppard A, Kaune W, Poole C, Kelsh M. A pooled analysis of magnetic fields, wire codes, and childhood leukemia. Childhood Leukemia-EMF Study Group. *Epidemiology*. 2000;11 :624 –634
- Greenleaf WJ, Block SM. Single-molecule, motion-based DNA sequencing using RNA polymerase. *Science* 2006; 313:801
- Hatch E, Linet M, Kleinerman R, et al. Association childhood acute lymphoblastic leukemia and use of electrical appliance during pregnancy and childhood. *Epidemiology* 1998; 9(3): 234-245
- Henshaw D, Reiter R. Do magnetic fields cause increased risk of childhood leukemia via melatonin disruption? *Bioelectromagnetics*. 2005;Suppl 7:S86-97
- Heynick LN, Merritt JH. Radiofrequency fields and teratogenesis. *Bioelectromagnetics*. 2003;(suppl 6):S174 –S186
- Hocking B, Gordon IR, Grain HL, Hatfield GE. Cancer incidence and mortality and proximity to TV towers [published correction appears in *Med J Aust*. 1997;166:80]. *Med J Aust*. 1996;165 :601 –605
- IARC, Globocan 2000: Cancer Incidence, Mortality and Prevalence Worldwide. 2000, International Agency for Research on Cancer, Lyon, France.
- IARC, Non-Ionizing Radiation, Part 1: Static and Extremely Low-Frequency (ELF) Electric and Magnetic Fields, IARC Monograph, Vol. 80, 2002, International Agency for Research on Cancer, Lyon, France.
- International Association of Cancer Registries. Cancer Incidence, Mortality and Prevalence Worldwide. Lyon, France: *Globocan*; 2000

- International Commission on Non-ionizing Radiation Protection. Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz) [published correction appears in *Health Phys.* 1998;75:442]. *Health Phys.* 1998;74 :494 –522
- International Commission on Non-Ionizing Radiation Protection. Exposure to Static and Low Frequency Electromagnetic Fields: Biological Effects and Health Consequences (0–100 kHz). Munich, Germany: ICNIRP; 2003
- Ju J, Kim DH, Bi L et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci USA* 2006; 103:19635-19640
- Juutilainen J. Developmental effects of electromagnetic fields. *Bioelectromagnetics.* 2005;Suppl 7:S107-15. Review.
- Kavet R. Contact current hypothesis: summary of results to date. *Bioelectromagnetics.* 2005;Suppl 7:S75-85.
- Kheifets L, Afifi A, Buffler P et al. Occupational electric and magnetic field exposure and leukemia. A meta-analysis. *J. Occup. Environ. Med.* 1997; 39(11): 1074-1091
- Kheifets L, Afifi A, Shimkhada R. Public Health Impact of Extremely Low-Frequency Electromagnetic Fields. *Environ Health Perspect.* 2006 October; 114(10): 1532–1537. Published online 2006 June 22.
- Kheifets L, Kavet R, Sussman S. Wire codes, magnetic fields, and childhood cancer. *Bioelectromagnetics.* 1997;18 :99 –110
- Kheifets L, Shimkhada R. Childhood leukemia and EMF: Review of the epidemiologic evidence. *Bioelectromagnetics.* 2005; Volume 26, Issue S7 , Pages S51 - S59
- Lee GM, Neutra RR, Hristova L, Yost M, Hiatt RA. A nested case-control study of residential and personal magnetic field measures and miscarriages [published correction appears in *Epidemiology.* 2003;14:255]. *Epidemiology.* 2002;13 :21 –31
- Levene MJ, Korlach J, Turner SW et al. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 2003; 299:682- 686.
- Li DK, Odouli R, Wi S, et al. A population-based prospective cohort study of personal exposure to magnetic fields during pregnancy and the risk of miscarriage. *Epidemiology.* 2002;13 :9 –20
- Lightfoot T. Aetiology of childhood leukemia. *Bioelectromagnetics.* 2005;Suppl 7:S5-S11. Review.
- Linet MS, Devesa SS. Descriptive epidemiology of childhood leukemia. *Br J Cancer.* 1991;63 :424 –429

- Linnet M, Hatch E, Kleinermann R, Robison L, Kaune W, Friedman D, et al. Residential exposure to magnetic fields and acute lymphoblastic leukemia in children. *N Engl J Med.* 1997;337:1–7.
- Linnet MS, Wacholder S, Zahm SH. Interpreting epidemiologic research: lessons from studies of childhood cancer. *Pediatrics.* 2003;112(1 pt 2) :218 –232
- Little J. *Epidemiology of Childhood Cancer.* Lyon, France: International Agency for Research on Cancer; 1999. *IARC Scientific Publication* No. 149.
- London S, Thomas D, Bowman J et al. Exposure to residential electric and magnetic fields and risk of childhood leukaemia. *Am. J. Epidemiol.* 1991; 134(9):923-937
- Lonn S, Ahlbom A, Hall P, Feychting M. Mobile phone use and the risk of acoustic neuroma. *Epidemiology.* 2004;15 :653 –659
- Lovely R, Buschbom R, Slavich A et al. Adult leukemia risk and personal appliance use: a preliminary study, *Am. J. Epidemiol.* 1994; 140(6): 510-517
- Margulies M, Egholm M, Altman WE et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005; 437:376-380
- Maskarinec G, Cooper J, Swygert L. Investigation of increased incidence in childhood leukemia near radio towers in Hawaii: preliminary observations. *J Environ Pathol Toxicol Oncol.* 1994;13 :33 –37
- McBride M, Callagher R, Thériault H et al. Power-frequency electric and magnetic fields and risk of childhood leukemia in Canada, *Am. J. Epidemiol.* 1999; 149(9): 831-842
- Mezei G, Kheifets L. Clues to the possible viral etiology of childhood leukemia. *Technology.* 2002;9 :3 –14
- Mezei G, Kheifets L. Selection bias and its implications for case-control studies: a case study of magnetic field exposure and childhood leukemia. *Int J Epidemiol.* 2005;35:397–406.
- National Institute of Environmental Health Sciences (NIEHS), NIEHS Report on Health Effects from Exposure to Power-Line Frequency and Magnetic Fields, NIH Publications No. 99-4493, 1999, National Institute of Environmental Health Sciences, National Institute of Health, Research Triangle Park, NC.
- National Radiological Protection Board. *Review of the Scientific Evidence for Limiting Exposure to Electromagnetic Fields (0–300 GHz).* Chilton, United Kingdom: National Radiological Protection Board; 2004
- Nole G, Johnson AW. An analysis of cumulative lifetime solar ultraviolet radiation exposure and the benefits of daily sun protection. *Dermatol Ther.* 2004; 17(suppl 1) :57 –62

- Parkin DM, Kramarova E, Draper GJ, et al. *International Incidence of Childhood Cancer*. Lyon, France: International Agency for Research on Cancer; 1998:2 . IARC Scientific Publication No. 144
- Poole C, Greenland S, Luetters C et al. Childhood leukemia and socioeconomic status: a systematic review. *Int. J. Epidemiol.* 2006; 35(2): 370-384
- Preston-Martin S, Peters J, Yu M et al. Myelogenous leukemia and electricblanket use. *Bioelectromagnetics* 1988; 9(3): 207-213
- Repacholi MH, Basten A, Gebiski V, Noonan D, Finnie J, Harris AW. Lymphomas in E mu-Pim1 transgenic mice exposed to pulsed 900 MHz electromagnetic fields. *Radiat Res.* 1997;147 :631 –640
- Reynolds, T. M. "Time for SI units." *Lancet* 352.9122 (1998): 152.
- Rhee M, Burns MA. Nanopore sequencing technology: research trends and applications. *Trends Biotech.* 2006; 24:580-586
- Savitz D, Chen J. Parental occupation and childhood cancer: review of epidemiologic studies. *Environ. Health Perspect.* 1990; 88: 325-337
- Savitz J, John E, Kleckner R. Magnetic field exposure from electric appliances and childhood cancer. *Am. J. Epidemiol.* 1990; 131(5): 763-773
- Savitz D, Wachtel H, Barnes F et al. Case-control study of childhood cancer and exposure to 60-Hz magnetic fields. *Am. J. Epidemiol.* 1988; 128(1):21-38
- Schreiber G, Swean G, Meijers J et al. Cancer mortality and residence near electricity transmission equipment: a retrospective cohort study. *Int. J. Epidemiol.* 1993; 22(1): 9-15
- Schuz J, Grigat JP, Stormer B, Rippin G, Brinkmann K, Michaelis J. Extremely low frequency magnetic fields in residences in Germany. Distribution of measurements, comparison of two methods for assessing exposure, and predictors for the occurrence of magnetic fields above background level. *Radiat Environ Biophys.* 2000;39 :233 –240
- Schuz J. Implications from epidemiologic studies on magnetic fields and the risk of childhood leukemia on protection guidelines. *Health Phys.* 2007 Jun;92(6):642-8. Review.
- Severson R, Stevens R, Kaune W et al. Acute nonlymphocytic leukemia and residential exposure to power frequency magnetic fields. *Am. J. Epidemiol.* 1988; 128(1): 10-20
- Shendure J, Porreca GJ, Reppas NB et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005; 309:1728-1732
- Sienkiewicz Z. Neurobehavioural effects of electromagnetic fields. *Bioelectromagnetics.* 2005; 26 (S7): S116 - S126

- Sjöblom T, Jones S, Wood LD et al. The consensus coding sequence of human breast and colorectal cancers. *Science*. 2006;314:268-274
- Steliarova-Fouchier E, Stiller C, Kaatsch P, et al. Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCIS project): an epidemiological study. *Lancet*, 2004, 364(9451): 2097-2105
- Tynes T, Halderson T. Electromagnetic fields and cancer in children residing near Norwegian high-voltage power lines. *Am. J. Epidemiol* 1997; 145(3):219-226
- UK Childhood Cancer Study Investigators, Exposure to power-frequency magnetic fields and the risk of childhood cancer. *Lancet* 1999; 354(9194): 1925-1931
- United Nations Scientific Committee on the Effects of Atomic Radiation. *Sources, Effects and Risks of Ionizing Radiation; UNSCEAR 2000 Report to the General Assembly*. New York, NY: United Nations Scientific Committee on the Effects of Atomic Radiation, United Nations; 2000
- Verkasalo P, Pukkala E, Hongisto M et al. Risk of cancer in Finish children living close to power lines. *Br. Med. J.* 1993; 307(6909):895-899.
- Verkasalo P. Magnetic fields and leukemia risk for adults living close to power lines. *Scand. J. Work Environ. Health* 1996; 22(Suppl. 2): 1-56
- Wertheimer N, Leeper E. Electrical wiring configurations and childhood cancer. *Am. J. Epidemiol.* 1979; 109(3):273-284
- Wertheimer N, Leeper E. Adult cancer related to electrical wires near the home. *Am. J. Epidemiol.* 1982; 11(4):345-355
- Wiencke JK, Thurston SW, Kelsey KT, et al. Early age at smoking initiation and tobacco carcinogen DNA damage in the lung. *J Natl Cancer Inst.* 1999;91 :614 –619
- Yang KH, Ju MN, Myung SH. Sample of Korean's occupational and residential exposures to ELF magnetic field over a 24-hour period. In: *Abstracts of 26th Annual Meeting of the Bioelectromagnetics Society*, 21–24 June 2004, Washington, DC. Washington, DC: Bioelectromagnetics Society, 188–189