What Next? The Next Transit from Biology to Diagnostics: Next Generation Sequencing for Immunogenetics

Christian Gabriel · Stephanie Stabentheiner · Martin Danzer · Johannes Pröll

Red Cross Transfusion Service for Upper Austria, Linz, Austria

Keywords

Histocompatibility · Massively parallel sequencing · HLA typing

Summary

The human genome project triggered the introduction of next generation sequencing (NGS) systems. Although originally developed for total genome sequencing, metagenomics and plant genetics, the ultra-deep sequencing feature of NGS was utilized for diagnostic purposes in HIV resistance and tropism as well in detecting new mutations and tumor clones in oncology. Recent publications exploited the feature of clonal sequencing for immunogenetics to dissolve the growing number of ambiguities. This concept is quite reliable if all exons of interest are tested and the amplification region includes flanking introns. Challenging questions on quality control, cost effectiveness, workflow, and management of enormous loads of data remain if NGS is considered as routine method in the immunogenetics laboratory. If solved, NGS has big potential to have a major impact on immunogenetics by way of providing ambiguity-free HLA-typing results faster, but will also have a great influence on how immunogenetics testing and workflows are organized.

Schlüsselwörter

Histokompatibilität · Massively-Parallel-Sequenzierung · HLA-Typisierung

Zusammenfassung

Das humane Genom-Projekt hat Entwicklungen ausgelöst, die zur nächsten Generation von Sequenziermethoden (next generation sequencing, NGS) geführt hat. Ursprünglich als Methode für das Sequenzieren gesamter Genome, die Metagenomik und die Pflanzengenetik eingerichtet, wurde die Möglichkeit zum «Ultra-Deep»-Sequenzieren auch für die Diagnose von HIV-Resistenzen und -Tropismen, aber auch für die Erkennung neuer Mutationen und Tumorklone eingesetzt. Neuere Literatur hat die Möglichkeit zur klonalen Sequenzierung aufgezeigt, mit der die wachsende Anzahl von Ambiguitäten in der Immunogenetik aufgelöst werden kann. Dieses Konzept ist dann zuverlässig, wenn möglichst viele Exons in die Analyse mit einbezogen werden (insbesondere jene, die für die Polymorphismen in Frage kommen) und wenn die Amplifikationsregion auch die flankierenden Introns umfasst. Herausragende Fragen verbleiben noch für die Qualitätskontrolle, die Kosteneffizienz, den Arbeitsablauf, und die Bewältigung der enormen Datenflut bevor Überlegungen zur Einführung dieser Methode in der Routine angestellt werden können. Wenn diese Fragen gelöst sind, hat NGS ein hohes Potenzial, die Patienten schneller und sicherer mit eindeutigen HLA-Typisierungsergebnissen zu versorgen, aber auch einen großen Einfluss auf die zukünftige Organisation von Abläufen und Tests in der Immunogenetik.

Introduction

After 13 years, the multinational human genome project was completed in 2003. Fundamental goals, like mapping and identifying human genes and determining sequences of all chromosomes, have been achieved [1, 2]. A load of data was generated and made available for further analysis. Moreover, this project has demonstrated that concerted efforts and enormous investments have to be accomplished to reach these goals [3]. This led to the perception that new technologies are essential for high throughput sequencing, and it sparked the foundation of new companies and risky ventures directed towards new sequencing methods and instruments.

New Technologies with Different Approaches

Next generation sequencing (NGS) also dubbed as massively parallel sequencing or sometimes total genome sequencing bears no resemblance to chain-termination (Sanger) sequencing. In all NGS methods, genetic material is extracted by customary methods, then a DNA library is prepared, usually by physical fragmentation and ligation of specific adapters. In the case of amplicon-based sequencing, pre-amplified DNA is used instead. Adapters, composed of signal sequence oligonucleotides, attach to both ends (paired end library) or circularize the DNA fragment (mate pair library). A mate pair library is an option if long ranges of unknown DNA up to 2 kb have to be explored, because the sequence reaction runs from the site of the adapter in both directions. DNA fragments with their identical adaptors immobilize to solid surfaces or beads. This step is a prerequisite for clonal sequencing since in any case single DNA strands are separated. To increase the amount of identical DNA and ensure an appropriate signal output in the sequencing reaction which is strong enough for sensitive detection, DNA amplification follows, either by emulsion polymerase chain reaction (PCR) or solid phase amplification. This step requires specific attention because polymerases with low fidelity may become an additional source of error, if there is a tendency to incorporate erroneous nucleotides. This can result in high exclusion rates of reads and consequently lead to low coverage, dropouts, and bad data quality. Meanwhile, newer sequencing methods avoiding amplification are on the verge of introduction. These methods, sometimes amusingly named as 'next-next generation', are capable of detecting DNA sequencing reactions at the single molecule level.

Most common is the use of emulsion PCR protocols in which single DNA fragments are bound to beads. Combined with PCR reagents, a clonal PCR reaction is carried out in a droplet of an oil-water emulsion. To discriminate between DNA beads carrying DNA and those which remain unloaded, the REM e (Robotic Enrichment Module; Roche 454 Life Sciences, Branford, CT, USA) can be used where liquid han-

dling robots are available. This so-called recovery and bead-enrichment protocol can also be performed manually; the time needed is 4 h. Utilizing millions of beads in 1 reaction step, zillions of specific amplicons (for amplicon-based sequencing) or random amplicons (for shotgun sequencing) are produced. Solid phase amplification is characterized by covalently attached primers on a slide with high density. Bridge amplification takes place on a spot, where clusters of amplicons grow and are ready for sequencing. Ultimately, after amplification, DNA strands are attached to beads and immobilized to PicoTiterPlates™ (Roche 454 Life Sciences) or glass slides ready for sequencing, which is in principle a sequencing by synthesis reaction but varies in terms of the assay system.

Currently there are 3 NGS systems in broad use. Among the many other NGS systems, acting mainly on academic level, Ion Torrent's system named Personal Genome Machine (PGM; Ion Torrent, Guilford, CT, USA), which senses directly the ions produced by polymerase activity during the sequencing process, is observed with high interest. In a recently published Nature paper, bacterial and human genomes sequenced by PGM were presented [4]. The number of overall sequenced bases per run (250 Mb) as well as the mean read length of 120 bp are in an acceptable range, but base-quality information does currently not meet expectations. Whether or not this system will improve its performance similar to the competitors in a year's time, will be of great importance. The so-called next-NGS systems are not mentioned yet, as the level of experience and scientific communication is not substantial enough to generate an overview. An excellent review of the different assay platforms was published by Metzker [4].

The SOLiD™ instrument (Applied Biosystems, Foster City, CA, USA) uses a process in which double-nucleotide DNA templates begin to hybridize after a primer binds to the DNA strand. If complementary to the DNA strand, the double-nucleotide template is ligated to the primer or respective former DNA template. 4 different fluorescent labels specify a set of these double-nucleotide templates. After hybridization, the fluorescent dye is cleaved off and detected, enabling the next hybridization reaction. Several ligation cycles generate a DNA strand with approximately 35–70 base pairs in length. In the next reaction step, the initiating primer includes 1 extended base, and the same reaction is started producing different signals due to the shift of 1 base. The combination of signals enables the base call.

The Illumina system (Illumina Inc., San Diego, CA, USA), which is the most widely used NGS system, incorporates complementary terminator-based nucleotides to the immobilized template strand [5]. The DNA synthesis then stops, and the remaining nucleotides are washed away. The fluorescent terminator nucleotide is detected and the fluorescent molecule cleaved afterwards to enable incorporation of the next nucleotide. After a washing step, the next one of the 4 fluorescent labeled nucleotides is added. The reaction starts again if the nucleotide is matching the template. The 4 colors are detected

by total internal reflection fluorescence (TIRF). Misincorporation errors are the most common error, especially when a new nucleotide follows a G.

The 454 Sequencing™ system (Roche 454 Life Sciences), represented now by 3 instrument types, the Genome Sequencer (GS)-FLX (400 bp read length), the GS-FLX+ (800 bp), and the GS Junior (400 bp), is a pyrosequencing reaction in which in each step 1 of 4 nucleotides (dNTPs) is added to the reaction [6]. If complementary, the dNTP is incorporated by PCR, subsequently releasing pyrophosphate. Using pyrophosphate, ATP sulfurylase converts ADP to ATP. Luciferase oxidizes luciferin to oxiluciferin in the presence of ATP. This reaction is accompanied by the emission of light. The reaction chain is stoichiometric, so 1 incorporated nucleotide is equivalent to a batch of emitted photons. All 4 nucleotides are washed in every step of elongation, and light emission is detected by CCD optical systems. Millions of sequences are read by the sensitive optical systems, incorrect reads are deleted, and available ones are transferred. Hence, NGS requires sophisticated data management systems, high performance computing, and dedicated bioinformatic tools to process the high amount of data into results [7].

Next Generation Sequencing in Biology: A Leverage for Medicine

Nevertheless, NGS were designed primarily for biology, metagenomics, and plant genomics, where genomic complexity, long ranges of repetitive sequences, and multiple overlaying genomes cause problems whenever conventional sequencing methods are used. This adds to the fact, that chain-termination sequencing in genomics requires vast resources, although the speed of research is limited and by common sense it is nearly impossible to manage the sheer amount of workload and data in one institution. It was for instance possible with NGS to decipher the genome of crop, tomatoes, and more importantly grapevine [8–10]. On the basis of these total genome sequencing approaches it will be feasible to mediate the ripening, drought endurance, yield, and disease resistance of crops and fruits [11]. Total genome sequencing has also had an impact on paleogenomics, demonstrating that the very young species *Homo sapiens* does not directly originate from Neanderthals. This is mind-blowing, especially considering the meticulous work applied to examining minute bone samples aged more than 30,000 years with highly fragmented DNA and potential contamination [11, 12]. Even the sequencing of a woolly mammoth highlighted how fast total genome sequencing may be applied after discovery [13]. Sequencing organisms, especially workhorses in biological research like *Caenorhabditis elegans*, its transcriptome, or the distribution of miRNAs in *Danio rerio* and the mapping of the *Escherichia coli* genome elucidated basic requirements in structural biology and led to further work on unknown organisms and viruses [14–18]. Using NGS for me-

tagenomics had a positive impact on marine biology ranging from biodiversity of deep sea to more shallow marine environs where the variety of microorganisms reflects the health of coral atolls [19, 20]. Somehow puzzling is the use of related methods to explore the human gut microbiome. As in atolls, there is a clear linkage of reduced bacterial diversity to energy consumption in humans. Furthermore, twin studies demonstrated that the unbalanced representation of bacterial genes and overrepresentation of specific metabolic pathways links to the development of obesity [21–23]. NGS plays also a pivotal role in the detection of new viruses in human disease and, again, in the grapevine [24, 25]. Rapidly diminishing honey bee colonies are a worry to bee keepers and scientists, as their loss is immediate and poses a threat to agriculture. Honey bees pollinate up to 70% of commodity crops. So far, no unique reason had been found, but NGS made the search for infectious agents possible indicating that *Nosema ceranae* or Israeli acute paralysis virus of bees play an important role subsequent to or in combination with environmental stressors [26, 27]. With comparative shotgun NGS of previous and modified or infected samples, there is a good chance to detect viral or microbial genomes without complex isolation and cultivation attempts. One of the most impressive findings was the detection of a new Arenavirus causing fatal outcome in transplanted patients and the detection of polyomavirus as a causative agent of Merkel cell carcinoma [28, 29]. Although these examples depict that various biological fields and application categories have had a wide influence on unresolved medical problems, there seems to be no broad application of NGS in medical diagnostics.

NGS in Medical Diagnostics: Not Yet Well Developed

The first medical diagnostic utilization of NGS transferred from ultra-deep sequencing of microbes. Inherent to all NGS methods is the fact that all DNA molecules in a chosen sample are amplified and sequenced. Therefore multiple sequence reads (so-called reads) with different lengths of up to 400 base pairs are generated, stacked on the basis of analog sequences, and classified by the discrepancy into groups, which may then represent a clone or allele. If minute amounts of related genetic variants regardless of the origin of the single nucleotide polymorphisms (SNPs), deletions, or insertions appear, they are then classified as different alleles or clones. On a specific segment of reads, depth of coverage (abbreviated as coverage) is represented by the amount of sequences with the same nucleotides in the same positions. The higher the coverage, the higher the probability of finding different clones with appropriate certainty. The sensitivity of NGS in comparison to Sanger sequencing is much higher and makes the detection and quantification of various clones possible if overall coverage is high enough.

Sensitive diagnosis of drug-resistant HIV strains and subsequent modification of treatment is imperative for the success-

ful management of HIV-infected patients. Antiretroviral treatment is associated with an increased probability to develop resistant strains, improper detection of these strains raises mortality [30]. Phenotypic assays and even approved genotyping assays by chain termination sequencing detect resistant strains only if a resistant clone surpasses the threshold 20% of total viral DNA [31, 32]. Even at a lower level, resistant strains may develop clinically relevant virologic failure. Detecting low-abundance variants down to a level of about 1% of the viral population shows better clinical results in terms of mortality and virological failure [33]. Furthermore, chemokine receptor 5 (CCR5) antagonists blocking the entry of HIV into the target cell by binding to the CCR5 molecule represent a different class of drugs. Therapy with these inhibitors requires the knowledge of the V3-loop of HIV-1 env which binds to the CCR5 receptor. Phenotypic assays currently in use take a long time, therefore genotyping is recommended in Europe because it is cheaper and faster [34]. NGS has nowadays a clear position in the use for tropism testing prior to the administration of CCR5 inhibitors [35, 36].

In oncology, the primary interest was to map genes of solid tumors and to find and quantify genes in heterogenic tissue [37–41]. Although the diagnostic benefit is quite low, mapping and epigenetic testing may select some genes of interest [42]. Additionally new markers in leukemia and specific patterns of RUNX1, RAS, CBL, and TET2 were defined by NGS [43]. It seems to be becoming one of the new methods in hematology, due to the possibility to quantify genes and their patterns, and resolve multiple clones harboring these genetic modifications [44]. Even cytogenetically normal leukemia can be detected by NGS [45]. But more interesting are the results of ultra-deep sequencing to verify different clones [46, 47]. This approach will add some valuable information to refine treatment protocols.

Headaches with High Resolution Typing in Immunogenetics

Stem cell transplantation requires exact and reliable results in the determination of HLA alleles located in the highly polygenic and polymorphic region of the short arm of chromosome 6. It is commonly accepted that low resolution typing ('2 digits') is not sufficient and results in higher mortality as well more acute graft versus host disease [49–54]. High resolution typing ('4 digits') of HLA-A, B, C, DRB1 and, more recently, DQB1, is performed either by sequence-specific oligonucleotide probes (SSOP), sequence-specific primer (SSP) PCR, or sequence-based typing (SBT) on exons 2 and 3 for class I and exon 2 for class II. These are the most relevant genetic regions coding the α -helices and β -sheets of the antigen-binding cleft of the exofacial HLA portion. Interestingly there is no method superior to the other, and testing strategies may vary from one laboratory to another. In general there are 2

pathways to generate high resolution typing results: One is by starting with a preliminary low resolution typing approach and then selecting specific probes and primers for high resolution typing. The other is the selection of SBT as the primary test and then adding on SSP-PCR or SSOP to distinguish ambiguities. Using a more selective and specific SBT approach to narrow the possible variations is an alternative. Ambiguities are the most prominent reason for this cascade of assays, because they may occur in 41% of HLA-A and 21% of HLA-B typing results mostly due to the fact that exon 4 is commonly not tested [55]. Allele ambiguities appear when polymorphisms, which distinguish alleles, are located outside of the regions of amplification, where primers or hybridization oligonucleotides determine the borders. Anyway, the required assay is not able to sort out alleles defined by these outlier polymorphisms. These ambiguities usually determine null alleles in which the HLA molecule is not expressed or truncated and therefore may generate an alloreactive response [56, 57]. Not identifying an HLA null allele in recipients is de facto a mismatch and therefore a hazard in which donor T cells are stimulated and further induce graft-versus-host disease. In the case of a truncated version, a minor histocompatibility antigen is produced and may induce rejection. Null alleles should be detected, and this is usually done by serology or null-allele-specific SSP-PCR, or in expanded forms of SBT in which flanking intron splicing sites are sequenced. Genotyping ambiguities occur when it is impossible to establish phase between closely linked polymorphisms. One of the most common forms of ambiguity is cis/trans ambiguity in which adjacent ambiguous nucleotide readouts create combinatorial chances to assign to different polymorphisms. These ambiguities are resolved by SSP-PCR or SSOP on the region of ambiguity or by family typing.

However, more and more laboratories consider implementing testing of exon 4 for class I and exon 3 for class II. The clinical relevance of other exons which are not tested, in terms of alloreactivity, expression, or structural conformation, remains unknown, but there is a general perception that they may be clinically irrelevant or pose a low impact on the course of treatment. However, the majority of known HLA alleles are not thoroughly sequenced in all exons, and therefore a silent question remains if they are clinically relevant. Furthermore, typing additional loci like HLA-C or HLA-DP is recommended either as a result of retrospective mortality analysis in bone marrow transplantation or for immunobiological reasons relating to the match or mismatch to KIR [58, 59]. To add some complexity, the growth of newly found alleles is ever increasing. The IMGT/HLA database contained at the time of inception 964 alleles and includes nowadays (April 2011) 6,534 alleles [60]. However, all approaches require time as they are incrementally designed, and there is still a problem especially with SSOP and SSP-PCR due to unidentified genotypes. Extended waiting time for the transplant should be avoided as toxicity of the therapy and possible recurrence of

disease may cause complications. However, all laboratory strategies lead to loss of valuable time at the cost of the patient's chances of survival. On the other hand many different assays and reagents have to be kept in the repertoire of the immunogenetics laboratory. Resolving ambiguities is not only cumbersome but also costly, and reduces the patient's chances. The search for new ways to cope with the growing number of alleles and the complexity of resolving these alleles resulted in recent publications showing the usefulness of NGS in immunogenetics.

Selection of the Appropriate NGS Platform for Immunogenetics

Some features of NGS shed light on the possible use for immunogenetics. In the following, some parameters are highlighted for selecting the appropriate platform. All criteria and, more so, the longer experience, make the 454 system the most appropriate to be explored for immunogenetics.

Read Lengths

The short read lengths of SOLiD and Illumina are ideal for expression analysis, miRNA quantification, and promoter binding studies but have a lower potential for de novo sequencing due to the short read length. The 454 system started with 120 bp on average (GS 20), and duplicated read lengths to 250 bp (GS FLX) and now up to 400 bp (GS FLX titanium). This gives the superb advantage of sequencing total exons.

Clonal Sequencing

All NGS systems apply emulsion PCR or solid phase amplification and therefore clonal sequencing. Next-NGS sequencing by single DNA samples has not been explored on the grounds of clonal sequencing.

Time to Result

NGS takes a long time overall. The major segments of the process are: sample preparation and amplification, DNA library preparation, emulsion PCR or solid phase amplification, normalization, sequence run, data acquisition, and bioinformatics. The first 3 steps are nearly identical in every platform mentioned. A difference in time usage can be found in the amplification and in the sequence run which varies from a few hours to many days. The enormous amount of data can further delay getting the result for many hours, even days.

Barcoding

NGS systems are primarily designed for total genome sequencing, so some constraints have to be taken into account if amplicon-based sequencing is considered. The most significant is the inability to distinguish different samples in 1 run, as there is no way to allocate them per se to 1 individual and

emulsion PCR is a pooled PCR process. One crude method is the division of PicoTiterPlates and slides into multiple sections by using screens to form up to 16 gaskets. However, this implies a loss of total sequencing capacity, and a high workload as samples have to be processed individually and viewed with caution since sample mix-up may easily happen. Barcoding individual amplicons prior to emulsion PCR helps to tag individual samples and run through the whole process without change [61, 62]. Primers used for sample amplification are connected to an oligonucleotide with a signature sequence which is unique to the individual sample. This sequence appears in the read and the bioinformatics system bins all reads with the same barcode signature sequence. One obstacle is relating to the length of the barcode which adds to the adapter and primer and may count up to 10 bases. This in fact reduces the use in all NGS platforms with short reads. A further restraint is the fact that amplification may be inhibited by the long tail attached to the primer. Various primer/barcode combinations must be validated before use and meticulous testing in combinations is needed to verify if preferential amplification or inhibition is avoided. Not all barcodes are useful and restrictions in the breadth of possible barcodes may appear. A new approach (Access Array™; Fluidigm Inc., South San Francisco, CA, USA) is the use of an array platform in which up to 48 different primers can be ligated with different barcodes and prior to performing the sample PCR reaction, which circumvents the problem of high costs with endless batteries of primer/barcode combinations.

Capacity

In terms of economics it is important to pack as much as possible into 1 run. The overall amount of reads is not crucial as is the length of reads and the coverage required to have reliable results. Illumina and SOLiD produce about 1 Gb per run, whereas 454 is in the range of 400–600 Mb. The lower amount of reads is easily compensated by longer reads in this platform.

Current Strategies in Using NGS for Immunogenetics

Recent publications indicate that there are 2 possible NGS strategies on the 454 system which is to date the only system in use in immunogenetics but also in molecular immunohematology [63]. The most common approach is the expansion of SBT to NGS in which exons are amplified and sequenced, the other one is transcript sequencing.

Exon-Based Approaches

Bentley et al. [64] demonstrated that clonal sequencing enables unambiguous typing results in terms of cis/trans or genotyping ambiguities. Barcoded primers for the amplification of

24 cell lines of 14 exons in total (HLA-A, -B, -C: exon 2, 3, 4; HLA-DQA1; DQB1: exon 2, 3, DBP1, DRB1: exon2) were constructed and all samples simultaneously sequenced. The assignment of genotypes was computed by the Conexio ATF software (Conexio Genomics, Fremantle, WA, Australia). If more than 1 genotype was assigned, the top genotype was selected. Manual editing was required to inactivate sequences which did not meet quality rules or in cases where sequences had to be trimmed or when low copy numbers of homozygous samples occurred. The concordance of the sequences with known results, typed by SSOP and SBT, was 99.4% for all loci. Furthermore the authors could demonstrate that co-amplified sequences of other HLA loci (DRB3/4/5) have to be filtered out by the software to avoid confusion in the assignment process.

In a very interesting case the authors demonstrated the ultra-deep sequencing capabilities of their assay. Maternally derived cells were detected in a severe combined immunodeficiency (SCIDS) patient after stem cell transplantation. An additional HLA-B and HLA-C allele, the non-transmitted HLA-allele from the mother, was detected in minute amounts demonstrating clearly that 1–2% of the cells were maternally derived.

In a subsequent publication with donor samples, a strategy with bidirectional sequencing primers set in flanking intronic regions for exons 1.2.3.4 in HLA-A and -B confirmed that clonal sequencing enables unambiguous results [65]. The authors mentioned that caution should be exercised with regard to the quality of reads. To check the overall quality, a set of internal quality sequences with a high fraction of homopolymers, which are challenging for the 454 system, is included in each run. 24% of these internal control beads had sequencing errors, indicating that 1 error in every 5,011 nucleotides may occur. Errors accounted for unidentified reads in up to 20% of all reads in 1 locus. Moreover, there is preferred amplification which may not only show differences in the reads counted in each allele but also between forward and reverse reads which are not evenly distributed as expected. 44% of 454,847 reads did not pass the quality filters set by the software. The major sources of rejection were a clonal emulsion PCR resulting in the amplification of more than 1 type of DNA on a bead, short read lengths, incomplete extension, and bad trimming of read ends. In situations with heterozygous hybrid alleles, the combinatorial problems may still reside and therefore additional testing may be needed.

Expanding the quest for unambiguous typing, class II loci were included in the testing of patient/donor samples [66]. 4 samples were tested for HLA-A, -B, -C, -DRB1, -DQB1, in which 17 of the 20 loci had combinations of well documented ambiguities. About 14% of the controls did not pass the quality filters and 95.5% of the reads aligned correctly. Also, the authors demonstrated that on some occasions the amount of reads is not equal depending on the allele in question and

therefore support the conclusion, that due to allelic imbalance a generally higher coverage is required to enable safe assignments if one of the alleles is weakly amplified or has low reads. It was shown that the depth of base calls depends on the position. Again, most errors in the reads occurred at the end and comprised of misincorporations and incorrect insertions or deletions. In general the results made clear that unambiguous results are easily obtained and even make NMDP codes and interim reports obsolete.

Developments in software have improved testing protocols and have made their distribution feasible. In a study with 8 different reference laboratories, which had either high experience with NGS in immunogenetics, general experience with NGS or were well-known HLA-typing laboratories but used NGS for the first time, a uniform protocol with a basic, already published set of primers was used to evaluate the robustness and reliability of NGS in immunogenetics [67]. 4 cell lines and 16 samples coming from 6 participants were sent out with the same set of reagents. The samples were selected as challenging ones, as they all were previously tested by standard SBT and SSP-PCR assays and had certain ambiguities or contained new and unknown alleles. The loci included were HLA-A, -B, -C, DRB1, DRB3/4/5, DQB1, DPB1, and DQA1. Some of them had truncated coverage by sample PCR as some primers were located within the exons and not at their respective boundaries or in flanking introns. As foreseeable, this was in some cases a reason for residual ambiguities. A barcode (MID tag) was tailed to the primers to allow simultaneous amplification of 10 samples in each of the 2 gas-kets. Amplicons were purified to remove primer-dimers and evaluated by electrophoresis. Quantification by spectrofluorimetry and precise normalization of DNA concentrations was one of the most crucial steps prior to emulsion PCR as indicated by general publications [68]. In this intermediate step, the quality of the laboratories was documented. The target value was set to 600,000 amplicons for further sequencing, but even laboratories with the lowest bead loads of about 79,000 had superb results. The Conexio ATF software displayed the consensus sequence read in comparison to the IMGT/HLA database. Sets of genotypes are binned to a zero mismatch group or others with the respective amount of mismatches. Some exons, like exon 4 in class I MHC, had no comparative sequence in the database, and the study clearly indicated that total sequencing of new genotypes is needed for the purpose of NGS. In some groups, occurring mismatches were clearly ascribed to lower fidelity problems of the polymerase, and PCR misincorporations and pyrosequencing errors resulted in rare sequence reads which are clearly distinguishable from the zero mismatch reads, as the amount of reads accounts for a very small population. However, this requires manual editing which was essential in 3.7% of the loci. Overall concordance with known genotypes was obtained in 97.2% on the genotype level and in 98.3% on the allele level but correlated with the experience of the labora-

tories and the loci. HLA-A had the lowest concordance with 91% and HLA DQB1/DQA1 the highest with 100%. However, the fact, that such a high concordance was achieved even in laboratories with a low experience level was astonishing and adding to this, 3 new genotypes were found by all participants. Some known weaknesses of the method were demonstrated too: Reliability toward the end of the read is lower and, more importantly, homopolymers in the last 20 positions of any sequence read may not be correctly reproduced. As the maximum length of reads is about 250 bp and some exons span over 300 bp, the forward and reverse read have a lower coverage in the middle portion of the exon, which again limits the ability to sort out any ambiguity. Longer reads up to 450 bp as provided by the next version, the titanium chemistry, are certainly better as they enable full coverage of exons and even adjacent intron sites. NGS clearly showed in this study that clonal sequencing abrogates the appearance of genotype and cis/trans ambiguities. More and longer amplicons will provide better typing results. The workload is cumbersome, automation is clearly needed, and still in 3.7% manual editing is unavoidable. Even less experienced laboratories were able to achieve very good concordance, but some experience is required if the overall coverage is low.

cDNA Sequencing

With a similar approach as performed in the MHC typing of macaques, in which a 190 bp amplicon spanning from the $\alpha 1$ to the $\alpha 2$ domain in the MHC class I was generated after cDNA transcription from RNA, Lank et al. [69] selected 48 samples of 47 human cell lines and 1 clinical sample. RNA was extracted, quantified, and reverse transcribed to cDNA. 2 sites with low polymorphism in exon 1 and 3 of class I were found and primers designed which were theoretically capable to align to 2,330 class I alleles. Although some primers had 1 base pair mismatches, less efficient amplification was clearly demonstrated, but it was also indicated that these primers may amplify even if not fitting well to the template. These universal primers tailed with barcodes amplified a section of 581 bp which was then applied for NGS. Analysis started with quality trims and inactivation of short reads. Using a BLAST-like software, sequences were aligned. Albeit the length of the amplicon was much longer than the maximum achievable sequence length, the coverage of overlapping forward and reverse reads was high enough to align correctly. In comparison to previous results, 41 of 48 samples were concordant. One of the discordant samples was a null allele which is clearly not identifiable by this approach of NGS and reflects the clinical relevance. However, the authors indicated that a longer amplicon with 927 bp may be generated covering the whole stretch down to the transmembranous region and resolving 95% of all ambiguities.

Conclusion

454 sequencing has the potential for unambiguous HLA typing if a high number of exons is tested to avoid allele ambiguities. It is obvious that primers should be set into flanking intronic regions to avoid ambiguities. This has been clearly shown by studies in which a high concordance with pre-tested samples was found and ambiguities needed no further resolution. Some caution should be exercised with regard to heterozygous hybrid alleles which appear very seldom and can be resolved by cDNA sequencing. Although the time to produce a result seems to be much longer than with Sanger sequencing, the overall time to achieve an unambiguous result is shorter [67]. Moreover, pre- and post-processing are not yet solved sufficiently, as workload is high and repetitive. For instance more than 14,000 pipetting steps are required to test 48 samples. Specifically, for amplicon-based sequencing approaches, separation of pre- and post-PCR laboratory space despite automation is one of the most prominent features needed in the future. Unavoidably, costs to operate the whole NGS laboratory are an issue, which may be seen primarily by laboratories that rely on high throughput HLA testing, like donor registries. To make the whole system cost effective, a high volume of simultaneously tested samples is important. Adding to the investments, the high reagent costs push the laboratory to fill as many samples into the platform as possible. Packing as much as possible into the test generates a major disadvantage: High volumes are creating low coverage and thus are conflicting with the minimum coverage needed. Cheap testing is in conflict with quality but also causes further costs. Barcode and adapter primers require HPLC cleaning and are expensive. A high amount of samples also induces a high number of validated barcoded primers, loss of sequencing capacity, and high software performance which is not standardized yet. Significant amounts of reads are not passing the quality filter. These lost reads may be seen as sunk costs. Further development is needed to get a clear view of how the overall process can be improved to curtail losses and how to optimize the amount of samples.

Errors occur most often at the end of the read. Homopolymers located at the end may be troublesome, as they are in general a weak spot of the 454 system. Trimming ends with bioinformatic tools is important to improve the quantity of reads useful for assignment. Low fidelity sample PCR may also decrease the coverage and lead to allele dropouts. Internal quality control sequences are added to ensure a system which is monitoring the general quality of emulsion PCR and sequencing. The fraction of wrong or rejected controls indicates the overall performance and reliability of the system and should be closely followed as the percentage of rejected reads has an influence on the overall rejection rate and further on the coverage. If the use of a NGS in diagnostics is considered, precautions should be taken to validate runs with quality reports derived from internal quality controls.

Manual editing is required but should be circumvented. To avoid this, quality and even distribution of DNA content after emulsion PCR are required to ensure high coverage and a clear assignment of the sequences. A big challenge is manual editing in which rare sequences have to be distinguished from erroneous ones that passed the quality filter. Typically, rare alleles harbor multiple polymorphisms, and erroneous sequences may occur with only 1 position not fitting into the correct allele.

Contrary to the economic driving force of high numbers of simultaneously tested samples is the demand to generate high coverage. This is needed to leverage allelic imbalance, different efficiencies of forward and reverse reads, and inherent amplification preferences or shortcomings of specific sequences. A high coverage also enables the detection of rare alleles and by statistical means also helps to use bioinformatic tools much more easily. High coverage also prevents allele dropouts as does PCR primer optimization. Sample PCR should be meticulously validated to avoid allele imbalance. Another problem is the difficulty to correctly assign homopolymer regions. This will remain as a general problem, as correctly assigning homopolymers seems to be an inherent problem of pyrosequencing.

The importance of software shall not be underestimated. Good software is essential. Further developments are awaited. Especially needed are software tools to validate the runs, create quality reports, and manage data to track any sample throughout the entire process, which is very complicated.

Overall, NGS has an enormous potential in immunogenetics, but further improvements in the process, software, and reagents are required before NGS can be implemented in routine diagnostics. The scientific community in immunogenetics has to fill gaps in published genotypes to enable the safe use of databases. This technology will add a twist to laboratories because high throughput is required to keep such systems efficient. It will possibly reduce the workload of many laboratories, as the cascading strategy to dissolve ambiguities will not be needed anymore. This method will have to demonstrate in the future that it can bring further improvements for routine testing and be cost-effective. Otherwise it will become yet another hype.

Disclosure Statement

The authors declared no conflict of interest.

References

- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huseon DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G: The sequence of the human genome. *Science* 2001; 291:1304–1351.
- Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tüzün E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M, Saranga D, Brand A, Tao W, Gustafson E, McKernan K, Chen L, Malig M, Smith JD, Korn JM, McCarroll SA, Altshuler DA, Peiffer DA, Dorschner M, Stamatoyanopoulos J, Schwartz D, Nickerson DA, Mullikin JC, Wilson RK, Bruhn L, Olson MV, Kaul R, Smith DR, Eichler EE: Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008;453:56–64.
- Collins FS, Morgan M, Patrinos A: The human genome project: lessons from large-scale biology. *Science* 2003;300:286–290.
- Metzker ML: Sequencing technologies – the next generation. *Nat Rev Genet* 2010;11:31–46.
- Oliphant A, Barker DL, Stuelpegel JR, Chee MS: BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* 2002;suppl:56–8, 60–61.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgeson S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:326–327.
- Trombetti GA, Bonnal RJ, Rizzi E, De Bellis G, Milanese L: Data handling strategies for high throughput pyrosequencers. *BMC Bioinformatics* 2007;8(suppl 1):S22.
- Wicker T, Schlagenhauf E, Graner A, Close TJ, Keller B, Stein N: 454 sequencing put to the test using the complex genome of barley. *BMC Genomics* 2006;7:275–286.
- Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Ozyerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R: A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* 2007;2:e1326.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchiatti A, Kater MM, Masiero S, Lasserre P, Lepinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R: The genome of the domesticated apple (*Malus X domestica* Borkh.). *Nat Genet* 2010;42:833–839.
- Moxon S, Jing R, Szittyá G, Schwach F, Rusholme Pilcher RL, Moulton V, Dalmay T: Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res* 2008;18:1602–1609.
- Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, Du L, Egholm M, Rothberg JM, Paunovic M, Pääbo S: Analysis of one million base pairs of Neanderthal DNA. *Nature* 2006;444:330–336.
- Noonan J: Neanderthal genomics and the evolution of modern humans. *Genome Res* 2010;20:547–553.
- Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, Tomsho LP, Packard MD, Zhao F, Sher A, Tikhonov A, Raney B, Patterson N, Lindblad-Toh K, Lander ES, Knight JR, Irzyk GP, Fredrikson KM, Harkins TT, Sheridan S, Pringle T, Schuster SC: Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 2008;456:387–390.

- 15 Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W, Magrini VJ, Richt RJ, Sander SN, Stewart DA, Stromberg M, Tsung EF, Wylie T, Schedl T, Wilson RK, Mardis ER: Whole-genome sequencing and variant discovery in *C. elegans*. *Nat Methods* 2008;5:183–188.
- 16 Shin H, Hirst M, Bainbridge MN, Magrini V, Mardis E, Moerman DG, Marra MA, Baillie DL, Jones SJ: Transcriptome analysis for *Caenorhabditis elegans* based on novel expressed sequence tags. *BMC Biol* 2008;6:30.
- 17 Mendoza-Vargas A, Olvera L, Olvera M, Grande R, Vega-Alvarado L, Taboada B, Jimenez-Jacinto V, Salgado H, Juárez K, Contreras-Moreira B, Huerta AM, Collado-Vides J, Morett E: Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One* 2009;4:e7526.
- 18 Soares AR, Pereira PM, Santos B, Egas C, Gomes AC, Arrais J, Oliveira JL, Moura GR, Santos MA: Parallel DNA pyrosequencing unveils new zebrafish microRNAs. *BMC Genomics* 2009;10:195.
- 19 Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML: Microbial population structures in the deep marine biosphere. *Science* 2007;318:97–100.
- 20 Dinsdale EA, Pantos O, Smruga S, Edwards RA, Angly F, Wegley L, Hatay M, Hall D, Brown E, Haynes M, Krause L, Sala E, Sandin SA, Thurber RV, Willis BL, Azam F, Knowlton N, Rohwer F: Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One* 2008;3:e1584.
- 21 Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Li S, Jian M, Zhou Y, Li Y, Zhang X, Li S, Qin N, Yang H, Wang J, Brunak S, Doré J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J; MetaHIT Consortium, Bork P, Ehrlich SD, Wang J: A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65.
- 22 Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;444:1027–1031.
- 23 Turnbaugh PJ, Quince C, Faith JJ, McHardy AC, Yatsunenko T, Niaz F, Affourtit J, Egholm M, Henrissat B, Knight R, Gordon JI: Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci U S A* 2010;107:7503–7508.
- 24 Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, Khristova ML, Weyer J, Swanepoel R, Egholm M, Nichol ST, Lipkin WI: Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog* 2009;5:e1000455.
- 25 Pantaleo V, Saldarelli P, Miozzi L, Giampetruzzi A, Gisel A, Moxon S, Dalmay T, Bisztray G, Burgyan J: Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. *Virology* 2010;408:49–56.
- 26 Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, Moran NA, Quan PL, Briese T, Hornig M, Geiser DM, Martinson V, vanEngelsdorp D, Kalkstein AL, Drysdale A, Hui J, Zhai J, Cui L, Hutchison SK, Simons JF, Egholm M, Pettis JS, Lipkin WI: A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 2007;318:283–287.
- 27 Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B, Egholm M, Hutchison S, Pettis JS, Lipkin WI, Evans JD: Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS Pathog* 2009;5:e1000466.
- 28 Palacios G, Druce J, Du L, Tran T, Birch C, Briese T, Conlan S, Quan PL, Hui J, Marshall J, Simons JF, Egholm M, Paddock CD, Shieh WJ, Goldsmith CS, Zaki SR, Catton M, Lipkin WI: A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* 2008;358:991–998.
- 29 Feng H, Shuda M, Chang Y, Moore PS: Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008;319:1096–1100.
- 30 Hogg RS, Bangsberg DR, Lima VD, Alexander C, Bonner S, Yip B, Wood E, Dong WW, Montaner JS, Harrigan PR: Emergence of drug resistance is associated with an increased risk of death among patients first starting HAART. *PLoS Med* 2006;3:e356.
- 31 Panel on Antiretroviral Guidelines for Adults and Adolescents: Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. US Department of Health and Human Services, 2008, pp. 1–128.
- 32 Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW: Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res* 2007;17:1195–1201.
- 33 Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, Baxter JD, Huang C, Lubeski C, Turenchalk GS, Braverman MS, Desany B, Rothberg JM, Egholm M, Kozal MJ; Terry Bein Community Programs for Clinical Research on AIDS: Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naïve patients significantly impact treatment outcomes. *J Infect Dis* 2009;199:693–701.
- 34 European guidance published on the use of tropism tests in routine HIV care. www.eatg.org/eatg/Global-HIV-News/Treatment/European-guidance-published-on-the-use-of-tropism-tests-in-routine-HIV-care.
- 35 Archer J, Braverman MS, Taillon BE, Desany B, James I, Harrigan PR, Lewis M, Robertson DL: Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing. *AIDS* 2009;23:1209–1218.
- 36 Tsibris AM, Korber B, Arnaout R, Russ C, Lo CC, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, Hughes MD, Gulick RM, Greaves W, Coakley E, Flexner C, Nusbaum C, Kuritzkes DR: Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. *PLoS One* 2009;4:e5683.
- 37 Smith DR, Quinlan AR, Peckham HE, Makowsky K, Tao W, Woolf B, Shen L, Donahue WF, Tusneem N, Stromberg MP, Stewart DA, Zhang L, Ranade SS, Warner JB, Lee CC, Coleman BE, Zhang Z, McLaughlin SF, Malek JA, Sorenson JM, Blanchard AP, Chapman J, Hillman D, Chen F, Rokhsar DS, McKernan KJ, Jeffries TW, Marth GT, Richardson PM: Rapid whole-genome mutational profiling using next-generation sequencing technologies. *Genome Res* 2008;18:1638–1642.
- 38 Thomas RK, Nickerson E, Simons JF, Jänne PA, Tengs T, Yuza Y, Garraway LA, LaFramboise T, Lee JC, Shah K, O'Neill K, Sasaki H, Lindeman N, Wong KK, Borras AM, Gutmann EJ, Dragnev KH, DeBiasi R, Chen TH, Glatt KA, Greulich H, Desany B, Lubeski CK, Brockman W, Alvarez P, Hutchison SK, Leamon JH, Ronan MT, Turenchalk GS, Egholm M, Sellers WR, Rothberg JM, Meyer-son M: Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat Med* 2006;12:852–855.
- 39 Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, Ballinger DG, Sparks AB, Hartigan J, Smith DR, Suh E, Papadopoulos N, Buckhaults P, Markowitz SD, Parmigiani G, Kinzler KW, Velculescu VE, Vogelstein B: The genomic landscapes of human breast and colorectal cancers. *Science* 2007;318:1108–1113.
- 40 Dahl F, Stenberg J, Fredriksson S, Welch K, Zhang M, Nilsson M, Bicknell D, Bodmer WF, Davis RW, Ji H: Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc Natl Acad Sci U S A* 2007;104:9387–9392.
- 41 Berger MF, Lawrence MS, Demicheli F, Drier Y, Cibulskis K, Sivachenko AY, Shonker A, Esgueva R, Pflueger D, Sougnez C, Onofrio R, Carter SL, Park K, Habegger L, Ambrogio L, Fennell T, Parkin M, Saksena G, Voet D, Ramos AH, Pugh TJ, Wilkinson J, Fisher S, Winckler W, Mahan S, Ardlie K, Baldwin J, Simons JW, Kitabayashi N, MacDonald TY, Kantoff PW, Chin L, Gabriel SB, Gerstein MB, Golub TR, Meyerson M, Tewari A, Lander ES, Getz G, Rubin MA, Garraway LA: The genomic complexity of primary human prostate cancer. *Nature* 2011;470:214–220.
- 42 O'Riain C, O'Shea DM, Yang Y, Le Dieu R, Gribben JG, Summers K, Yeboah-Afari J, Bhaw-Rosun L, Fleischmann C, Mein CA, Crook T, Smith P, Kelly G, Rosenwald A, Ott G, Campo E, Rimsza LM, Smeland EB, Chan WC, Johnson N, Gascoyne RD, Reimer S, Braziel RM, Wright GW, Staudt LM, Lister TA, Fitzgibbon J: Array-based DNA methylation profiling in follicular lymphoma. *Leukemia* 2009;23:1858–1866.
- 43 Kohlmann A, Grossmann V, Klein HU, Schindela S, Weiss T, Kazak B, Dicker F, Schnittger S, Dugas M, Kern W, Haeflrich C, Haeflrich T: Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1. *J Clin Oncol* 2010;28:3858–3865.
- 44 Bacher U, Haeflrich C, Schnittger S, Kohlmann A, Kern W, Haeflrich T: Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies. *Ann Hematol* 2010;89:643–652.
- 45 Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Riees RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK: DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;456:66–72.

- 46 Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, Simen BB, Hanczaruk B, Nguyen KD, Nadeau KC, Egholm M, Miklos DB, Zehnder JL, Fire AZ: Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med* 2009;1:12ra23.
- 47 Campbell PJ, Pleasance ED, Stephens PJ, Dicks E, Rance R, Goodhead I, Follows GA, Green AR, Futreal PA, Stratton MR: Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc Natl Acad Sci U S A* 2008;105:13081–13086.
- 48 Valcárcel D, Sierra J, Wang T, Kan F, Gupta V, Hale GA, Marks DI, McCarthy PL, Oudshoorn M, Petersdorf EW, Ringdén O, Setterholm M, Spellman SR, Waller EK, Gajewski JL, Marino SR, Senitzer D, Lee SJ: One-antigen mismatched related versus HLA-matched unrelated donor hematopoietic stem cell transplantation in adults with acute leukemia: Center for International Blood and Marrow Transplant research results in the era of molecular HLA typing. *Biol Blood Marrow Transplant* 2011;17:640–648.
- 49 Petersdorf EW: Optimal HLA matching in hematopoietic cell transplantation. *Curr Opin Immunol* 2008;20:588–593.
- 50 Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, Fernandez-Vina M, Flomenberg N, Horowitz M, Hurley CK, Noreen H, Oudshoorn M, Petersdorf E, Setterholm M, Spellman S, Weisdorf D, Williams TM, Anasetti C: High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. *Blood* 2007;110:4576–4583.
- 51 Hansen JA, Petersdorf EW, Lin MT, Wang S, Chien JW, Storer B, Martin PJ: Genetics of allogeneic hematopoietic cell transplantation. Role of HLA matching, functional variation in immune response genes. *Immunol Res* 2008;41:56–78.
- 52 Petersdorf EW, Anasetti C, Martin PJ, Gooley T, Radich J, Malkki M, Woolfrey A, Smith A, Mickelson E, Hansen JA: Limits of HLA mismatching in unrelated hematopoietic cell transplantation. *Blood* 2004;104:2976–2980.
- 53 Arora M, Weisdorf DJ, Spellman SR, Haagenson MD, Klein JP, Hurley CK, Selby GB, Antin JH, Kernan NA, Kollman C, Nademanee A, McGlave P, Horowitz MM, Petersdorf EW: HLA-identical sibling compared with 8/8 matched and mismatched unrelated donor bone marrow transplant for chronic phase chronic myeloid leukemia. *J Clin Oncol* 2009;27:1644–1652.
- 54 Spellman S, Setterholm M, Maiers M, Noreen H, Oudshoorn M, Fernandez-Viña M, Petersdorf E, Bray R, Hartzman RJ, Ng J, Hurley CK: Advances in the selection of HLA-compatible donors: refinements in HLA typing and matching over the first 20 years of the National Marrow Donor Program Registry. *Biol Blood Marrow Transplant* 2008;14(9 suppl):37–44.
- 55 Adams SD, Barracchini KC, Chen D, Robbins F, Wang L, Larsen P, Luhm R, Stroncek DF: Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: when precise nucleotide sequencing leads to imprecise allele identification. *J Transl Med* 2004;2:30.
- 56 Elsner HA, Blasczyk R: Immunogenetics of HLA null alleles: implications for blood stem cell transplantation. *Tissue Antigens* 2004;64:687–695. Review. Erratum in: *Tissue Antigens* 2006;68:191.
- 57 Poli F, Scalomogna M, Sirchia G: HLA null alleles: implications in stem-cell transplantation. *Cytotherapy* 1999;1:365–366.
- 58 Shaw BE, Gooley TA, Malkki M, Madrigal JA, Begovich AB, Horowitz MM, Gratwohl A, Ringdén O, Marsh SG, Petersdorf EW: The importance of HLA-DPB1 in unrelated donor hematopoietic cell transplantation. *Blood* 2007;110:4560–4566.
- 59 Flomenberg N, Baxter-Lowe LA, Confer D, Fernandez-Vina M, Filipovich A, Horowitz M, Hurley C, Kollman C, Anasetti C, Noreen H, Begovich A, Hildebrand W, Petersdorf E, Schmeckpeper B, Setterholm M, Trachtenberg E, Williams T, Yunis E, Weisdorf D: Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. *Blood* 2004;104:1923–1930.
- 60 www.ebi.ac.uk/imgt/hla/stats.html.
- 61 Binladen J, Gilbert MTP, Bollback JP, et al.: The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2007;2:e197.
- 62 Meyer M, Stenzel U, Hofreiter M: Parallel tagged sequencing on the 454 platform. *Nat Protoc* 2008;3:267–278.
- 62 Stabenheimer S, Danzer M, Niklas N, Atzmüller S, Pröll J, Hackl C, Polin H, Hofer K, Gabriel C: Overcoming methodical limits of standard RHD genotyping by next-generation sequencing. *Vox Sang* 2011;100:381–388.
- 63 Bentley G, Higuchi R, Hoglund B, Goodridge D, Sayer D, Trachtenberg EA, Erlich HA: High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens* 2009;74:393–403.
- 64 Gabriel C, Danzer M, Hackl C, Kopal G, Hufnagl P, Hofer K, Polin H, Stabenheimer S, Pröll J: Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. *Hum Immunol* 2009;70:960–964.
- 65 Lind C, Ferriola D, Mackiewicz K, Heron S, Rogers M, Slavich L, Walker R, Hsiao T, McLaughlin L, D'Arcy M, Gai X, Goodridge D, Sayer D, Monos D: Next-generation sequencing: the solution for high-resolution, unambiguous HLA typing. *Hum Immunol* 2010;71:1033–1042.
- 66 Holcomb CL, Höglund B, Anderson MW, Blake LA, Böhme I, Egholm M, Ferriola D, Gabriel C, Gelber SE, Goodridge D, Hawbecker S, Klein R, Ladner M, Lind C, Monos D, Pando MJ, Pröll J, Sayer DC, Schmitz-Agheguian G, Simen BB, Thiele B, Trachtenberg EA, Tyman DB, Wassmuth R, White S, Erlich HA: A multi-site study using high-resolution HLA genotyping by next generation sequencing. *Tissue Antigens* 2011;77:206–211.
- 67 Mashayekhi F, Ronaghi M: Analysis of read-length limiting factors in pyrosequencing chemistry. *Anal Biochem* 2007;363:275–287.
- 68 Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, Detmer AM, Bouffard P, Levenkova N, Turcotte CL, Szekeres E Jr, Wright C, Harkins T, O'Connor DH: Major histocompatibility complex genotyping with massively parallel pyrosequencing. *Nat Med* 2009;15:1322–1326.
- 69 Lank SM, Wiseman RW, Dudley DM, O'Connor DH: A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based HLA class I genotyping. *Hum Immunol* 2010;71:1011–1017.