Supplementary Figure 1. Comparison of ARCS vs sequence-specific real time PCR for DNAs of differing quality/quantity. DNAs “A” and “B” were extracted and quantified by different laboratories using different protocols. ARCS plotted points are means of 2 replicates. Real time data was obtained using 4 replicates for target sequence and 4 replicates for reference sequence.
Supplementary Figure 2. Location of HP and TP53 amplicons.
Supplementary Figure 3.
Upper: Plot of ratio and yield for \textit{CCL3L1}.
Lower: Plot of ratio and yield for \textit{CCL3L1}, with ratio linearly corrected for yield-associated skewing.
Supplementary Figure 4. Ratio/ratio plot of $CCL3L1/TP53$, corrected for yield-associated skewing.
Supplementary Figure 5. Comparison of both ARCS and sequence-specific PCR (each run for 33 versus 53 cycles) for *CCL3L1*. 
Supplementary Figure 6. Comparison of ARCS with sequence-specific real time PCR. Regression of ARCS against real time PCR for a CCL3L1 CNV assay.

\[ y = 0.9658x + 0.0714 \]

\[ R^2 = 0.8275 \]
Supplementary Figure 7. Comparison between two runs (run ‘A’ and ‘B’) over 53 cycles of an ARCS assay for *CCL3L1* (*CCL3L1/TP53* peak height ratio shown).
Supplementary Figure 8. Real time amplification profiles for \textit{CCL3L1}, amplified by ARCS (upper panel) and sequence-specific PCR (lower panel). Note: for the ARCS plot, the two low temperature and one intermediate temperature initial annealing cycles are not shown in the real time output.
Supplementary Table 1. Comparison of ARCS vs other CNV techniques.

|                      | ARCS                         | Fibre FISH                   | Q-PCR           | QMPSF          | MAPH           | MLGA           | MLPA           | PRT           |
|----------------------|------------------------------|------------------------------|-----------------|----------------|----------------|----------------|----------------|---------------|
| **Accuracy**         | High at low copy numbers     | Extremely high               | High at low copy numbers | High           | High           | High           | High           | High          |
| **Throughput**       | High                         | Extremely low               | Medium          | Low (but multiple loci) | Low (but multiple loci) | Low (but multiple loci) | Low (but multiple loci) | High          |
| **Cost per CNV genotype** | Low                         | High                         | Low/Medium      | Medium         | Medium         | Medium         | Medium         | Low           |
| **Probes/Labelling required?** | No. Generic backbone-binding dye | FISH probes                  | Either generic dye or labelled probes | FAM fluorochrome | $^{32}$P label | Probes         | Probes         | Labelled primers |
| **Stages required**  | Generic design PCR. Melt profile. | Very labour intensive       | Create standard curves. Amplify target | PCR. Separate amplicons by size and quantify | Immobilise genomic DNA. O/N probe hyb. Wash filters. PCR (2 stages). PAGE + image analysis. | Digest genomic DNA. Circularise with probes. PCR. Size and quantify amplicons. | Prepare probes. Hybridise probe pairs with genomic DNA. Ligate. PCR. Size and quantify amplicons. | Identify appropriate paralogue & design specific assay. PCR. Size and quantify amplicons. |
| **Advantage(s)**    | Speed. Low cost. High throughput. Wide range of copy numbers. Generic. | Extreme sensitivity         | Wide range of copy numbers. Simultaneous PCR and quantitation | Multiplex loci. | Multiplex loci. | Multiplex loci. | Multiplex loci. | Speed. Low cost. High throughput. |
| **Discrete vs quantitative.** | Discrete for low copy numbers | Discrete for low copy numbers | Quantitative     | Quantitative   | Quantitative   | Quantitative   | Quantitative   | Discrete for low copy numbers |
| **Amount of DNA required** | 10 ng                        | 10-30 ng                     | 10 ng           | 100-150 ng     | 0.5-1.0 µg    | 40 ng          | 20 ng          | 10 ng         |
Supplementary Table 2. Primer sequences.

| Primer                        | Sequence                                                |
|-------------------------------|---------------------------------------------------------|
| UNIVERSAL                     | 5’gcgtactagctaccgtagctacggaggg                        |
| IMPROVED UNIVERSAL            | 5’gcgggcgtactagctaccgtagctacggaggg                    |
| HP forward (ARCS)             | 5’ gcgtactagctaccgtacggagggTTAGTGAAATCCTTTATTGGGAT     |
| HP reverse (ARCS)             | 5’ gcgtactagctaccgtacggagggGACCTCTGCACATCAATCTCCTT     |
| HP forward (non-ARCS)         | 5’GGGAATTTGGGAAATTCCCTTTATTGGGAT                      |
| HP reverse (non_ARCS)         | 5’GCTGCCTCTGACATCAATCTCCTT                             |
| CCL3L1 forward (ARCS)         | 5’gcgtactagctaccgtacggagggTTACAAAATTCCATGAAGAGAGCTA    |
| CCL3L1 reverse (ARCS)         | 5’gcgtactagctaccgtacggagggGATGTCTCTTTTCTGTCTATATT     |
| TP53 forward (ARCS)           | 5’ gcgtactagctaccgtacggagggATAACTTCGTGCAAGATGCT       |
| TP53 reverse (ARCS)           | 5’ gcgtactagctaccgtacggagggGCCTGTTATATGAGGAGGTTA      |
| CCL3L1 forward (Real-time PCR)| 5’GACAAAAATTCCATGAAGAGAGCTAAGAG                   |
| CCL3L1 reverse (Real-time PCR)| 5’CAGAGAAATGTCTCTTTGTTTCTGTCTATATT                 |
Supplementary Table 3. Genotype and allele frequencies for BWHHS cohort.

| Genotype / Allele | Number (%) [EXPECTED NUMBER] |
|-------------------|------------------------------|
| Hp1/1             | 359 (12.3) [370.8]           |
| Hp1/2             | 1361 (46.7) [1337.4]         |
| Hp2/2             | 1194 (41.0) [1205.8]         |
| Hp1               | 2079 (35.7)                  |
| Hp2               | 3749 (64.3)                  |

BWHHS        HWE chi² (1df) = 0.912  p = 0.34
Supplementary Table 4. Dimer-minimising dinucleotides. The listed combinations of dinucleotides, if included between universal and gene-specific sequences, offer the maximum protection against dimer formation. For each gene-specific primer pair (CNV or reference) the corresponding column contains the dinucleotide sequences that should be included between universal and gene-specific sequence for the forward and reverse primers. The subsequent columns list all possible combinations of those four dinucleotides in terms of whether they will match or mismatch in a “panhandle” structure. Perfect matches effectively prevent strand invasion (see Figure 2 c,d) and thus dimer amplification, whilst imperfect mismatches reduce strand invasion efficiency, compromising dimer amplification in favour of the target-specific products.

| Primer pair 1 | Primer pair 2 | Perfect match | Match at +1 only | Match at +2 only | Perfect mismatch |
|---------------|---------------|---------------|-----------------|-----------------|-----------------|
| aa,cc         | ac,ca         | 4             | 2               | 2               | 2               |
| aa,gc         | ac,ga         | 4             | 2               | 2               | 2               |
| aa,tc         | ac,ta         | 4             | 2               | 2               | 2               |
| aa,gc         | ag,ca         | 4             | 2               | 2               | 2               |
| aa,gg         | ag,ga         | 4             | 2               | 2               | 2               |
| aa,tg         | ag,ta         | 4             | 2               | 2               | 2               |
| aa,ct         | at,ca         | 4             | 2               | 2               | 2               |
| aa,gt         | at,ga         | 4             | 2               | 2               | 2               |
| aa,tt         | at,ta         | 4             | 2               | 2               | 2               |
| ac,gc         | ag,cc         | 4             | 2               | 2               | 2               |
| ac,gg         | ag,gc         | 4             | 2               | 2               | 2               |
| ac,tc         | at,cc         | 4             | 2               | 2               | 2               |
| ac,gt         | at,gc         | 4             | 2               | 2               | 2               |
| ac,tt         | at,tc         | 4             | 2               | 2               | 2               |
| ag,ct         | at,gc         | 4             | 2               | 2               | 2               |
| ag,gt         | at,gg         | 4             | 2               | 2               | 2               |
| ag,tt         | at,tt         | 4             | 2               | 2               | 2               |
| ca,gc         | cc,ga         | 4             | 2               | 2               | 2               |
| ca,tc         | cc,ta         | 4             | 2               | 2               | 2               |
| ca,gg         | cg,ga         | 4             | 2               | 2               | 2               |
| ca,tg         | cg,ta         | 4             | 2               | 2               | 2               |
| ca,gt         | ct,ga         | 4             | 2               | 2               | 2               |
| ca,tt         | ct,ta         | 4             | 2               | 2               | 2               |
| cc,gg         | cg,gc         | 4             | 2               | 2               | 2               |
| cc,tc         | cg,tc         | 4             | 2               | 2               | 2               |
| cc,gt         | ct,gc         | 4             | 2               | 2               | 2               |
| cc,tt         | ct,tc         | 4             | 2               | 2               | 2               |
| cg,gt         | ct,gg         | 4             | 2               | 2               | 2               |
| cg,tt         | ct,tg         | 4             | 2               | 2               | 2               |
| ga,tc         | gc,ta         | 4             | 2               | 2               | 2               |
| ga,tg         | gg,ta         | 4             | 2               | 2               | 2               |
| ga,tt         | gt,ta         | 4             | 2               | 2               | 2               |
| gc,tg         | gg,tc         | 4             | 2               | 2               | 2               |
| gc,tt         | gt,tc         | 4             | 2               | 2               | 2               |
| gg,tt         | gt,tg         | 4             | 2               | 2               | 2               |
Supplementary Table 5. Primer sequences for examining the effects of base pairing in the ARCS stem-loop structure, adjacent to the universal primer stem, on ARCS-PCR. Bases involved in pairing (for all except the “ZIP” primers) are highlighted in bold.

|               | Primer Sequence                                    |
|---------------|----------------------------------------------------|
| TP53 reverse  | 5’gcgtactagtgactacgtgacgacggGACGCCGTTATATGAGAGGTTA|
| TP53_0 forward| 5’gcgtactagtgactacgtgacgacggTAAATAACCTTCGTGCAAGATGCT|
| TP53_1 forward| 5’gcgtactagtgactacgtgacgacggGTAAATAACCTTCGTGCAAGATGCT|
| TP53_2 forward| 5’gcgtactagtgactacgtgacgacggTAATAACCTTCGTGCAAGATGCT|
| TP53_3 forward| 5’gcgtactagtgactacgtgacgacggTTCAATAACCTTCGTGCAAGATGCT|
| TP53_12 forward| 5’gcgtactagtgactacgtgacgacggGAATAACCTTCGTGCAAGATGCT|
| TP53_13 forward| 5’gcgtactagtgactacgtgacgacggGTAATAACCTTCGTGCAAGATGCT|
| TP53_23 forward| 5’gcgtactagtgactacgtgacgacggTACATAACCTTCGTGCAAGATGCT|
| TP53_123 forward| 5’gcgtactagtgactacgtgacgacggGACATAACCTTCGTGCAAGATGCT|
| TP53 ZIP_F    | 5’gcgtactagtgactacgtgacgacggTTGTAACCGACGGCCAGTATAACCTTCGTGCAAGATGCT|
| TP53 ZIP_R    | 5’gcgtactagtgactacgtgacgacggTTGTAACCGACGGCCAGTATAACCTTCGTGCAAGATGCT|
Supplementary Table 6. *CCL3L1* copy number calling – sequence-specific real time PCR (LC480) versus ARCS.

| CCL3L1 Copy Number call | ARCS | ARCS | ARCS |
|--------------------------|------|------|------|
| LC480                    | 1    | 4    | 0    |
| LC480                    | 2    | 0    | 14   |
| LC480                    | 3    | 0    | 0    | 4    |
Supplementary Note 1.

Python code to find ARCS exclusion doublets for primer design.

def revcomp(seq):
    subst = {"a":"t","c":"g","g":"c","t":"a"}
    return subst[seq[1]]+subst[seq[0]]

def testseq(seq,rseq):
    if seq == rseq:
        return "A"
    elif seq[0] == rseq[0] and seq[1] != rseq[1]:
        return "B"
    elif seq[0] != rseq[0] and seq[1] == rseq[1]:
        return "C"
    elif seq[0] != rseq[0] and seq[1] != rseq[1]:
        return "D"
    else:
        return "Err"

def testcombination(seqset):
    sa,sb,sc,sd = seqset
    se,sf,sg,sh = sa,sb,sc,sd#revcomp(sa),revcomp(sb),revcomp(sc),revcomp(sd)
    rseqset = [se,sf,sg,sh]
    mismatches = {}
    for seq in seqset:
        mismatches[seq] = 0
        for rseq in rseqset:
            if testseq(seq,rseq) == "D":
                mismatches[seq] += 1
            if mismatches[sa] > 0 and mismatches[sb] > 0 and mismatches[sc] > 0 and
                mismatches[sd] > 0:
                scores = [0,0,0,0]
                result = []
                #print seqset,rseqset
                for i in range(0,len(seqset)):
                    for j in range(i,len(rseqset)):
                        score = testseq(seqset[i],rseqset[j])
                        #print score,seqset[i],rseqset[j]
                        if score == "A":
                            scores[0] += 1
                        elif score == "B":
                            scores[1] += 1
                        elif score == "C":
                            scores[2] += 1
                        elif score == "D":
                            scores[3] += 1
                            result.append(seqset[i]+","+rseqset[j])
                        if scores[0] >= 2:
                            for item in scores:
                                result.append(str(item))
                            return (",".join(result))
                else:
                    return ("rejected")

nucleotides = ["a","c","g","t"]
doublets = []
for nucleotidea in nucleotides:
for nucleotideb in nucleotides:
    doublets.append(nucleotidea+nucleotideb)
combinations = []
w = 0
for a in range(w,len(doublets)):
x = w
    for b in range(x,len(doublets)):
y = x
        for c in range(y,len(doublets)):
z = y
            for d in range(z,len(doublets)):
combinations.append(doublets[a]+"\t"+doublets[b]+"\t"+doublets[c]+"\t"+doublets[d])
z+=1
    y+=1
x+=1
w+=1

doubletfile = open("doublets.txt",'w')
combinationfile = open("doubletcombinations.txt","w")
for doublet in doublets:
doubletfile.write(doublet + "\n")
for combination in combinations:
    combination2 = combination.split("\t")
    scores = testcombination(combination2)
    output = (combination + "\t" + scores + "\n")
    if scores != "rejected":
        combinationfile.write(output)
        if scores[-2:] == "t2":
            print output.strip()
doubletfile.close()
combinationfile.close()
Supplementary Note 2.

Consideration of ARCS intramolecular versus primer/template intermolecular distances during an ARCS reaction.

PRIMER: We use 10pmoles of UP per 10µl PCR per microlitre, which represents \((10 \times 10^{12})\) moles x (6 x \(10^{23}\)) molecules/mole [Avogadro’s number] x \(10^1\) molecules per 1µl i.e. \(60 \times 10^{10}\) or \(0.6 \times 10^{12}\) primer molecules per µl. Taking the cube root of \(1/(0.6 \times 10^{12})\) to determine the cube side dimension of each primer molecule’s typical space, gives 1.186 x \(10^{-4}\)mm, or 119 nm box sides for that primer molecule.

AMPLICON: If the final yield in 10µl reaction is 200 ng of a 100-mer product then, using an average molecular mass e.g 300g/mole/base, gives the following conversion of mass to molecules of product in 1µl:

\[
\frac{(200 \times 10^{-9}) \text{g} \times 1/(300 \times 100 \times 2) \text{g}^{-1} \text{[for moles of duplex]} \times 10^{-1} \text{[1µl not 10µl]} \times 6 \times 10^{23} \text{molecules/mole}}{2 \times 10^{11} \text{strands per µl at end of the reaction, or } 0.2 \times 10^{12}}.
\]

Cube root of \(1/(0.2 \times 10^{12})\) = 171nm box sides for that duplex or its constituent strands.

STRAND LENGTH/BOX SIZE: B DNA contains ~ 10 bases per turn. One turn ~3.4nm, so 100-mer duplex would be 34nm long (within 171nm typical space). A single strand will not tend to be extended to the same rigidity as duplex, so the two ends will tend to be considerably closer than this. Therefore even towards the end of PCR, strand ends are much nearer to each other than to primer molecules or other strands. Thus, while diffusibility/mobility of long strands is less than that of primers, most newly formed strands are likely to form a hairpin after the denaturation step.